

7429198-the-profit-bias-identity-in-sports-betting-bookmaker-profit-as-the-public-s-prediction-error

AVTORJI: Jacek P. Dmochowski

POVZETEK: Raziskava obravnava, kako se zasliženka trgovca v sportnem stazanju razdeli na tri kanale: zasliženka, katera je rezultat odstotka zasebnega prepoznavanja, težavnost prepoznavanja in kovarianco med delom staze in rezultatom. Natančno je pokazano, da je zasliženka vendar nekaj, če je prepoznavanje nekaj drugo. Raziskava testira predpisi na podatkih 1.139 MLB igre in najde, da je videti odvisnost med delom staze in rezultatom enotni primer Simpsonove paradokse.

METODA: Raziskava je založena na modelu, ki razdeli zasliženke na tri kanale: zasliženke, težavnosti prepoznavanja in kovarianco med delom staze in rezultatom. Metodologija je opisana zasliženke, prepoznavanja in kovarianco. Podatki so testirani na 1.139 MLB igrah.

KLJUČNE UGOTOVITVE:

- Zasliženka je affine in narašča z pričakovanim delom staze na strani porazitelja.
- Zasliženka je ena trih kanalov, ki jih razdelijo zasliženka, težavnost prepoznavanja in kovarianco med delom staze in rezultatom.
- Zasliženka je nekaj, če je prepoznavanje nekaj drugo.
- V skupini igre je videti odvisnost med delom staze in rezultatom, vendar je ta odvisnost nevidna, ko so igre razdeljene glede na stran, ki jo je trgovec preferiral.

POMEN ZA TRGOVANJE: Raziskava pokaže, da je zasliženka vendar nekaj, če je prepoznavanje nekaj drugo. Trgovci morajo razumeti, da je zasliženka odvisna od prepoznavanja in težavnosti prepoznavanja, kar lahko pomaga pri izboljšanju strategij za trgovljanje.

THE PROFIT–BIAS IDENTITY IN SPORTS BETTING: BOOKMAKER PROFIT AS THE PUBLIC’S PREDICTION ERROR

Jacek P. Dmochowski
Department of Biomedical Engineering
City College of New York
New York, NY 10031
jdmochowski@ccny.cuny.edu

September 6, 2026

ABSTRACT

Sports betting moves money continuously from a large public to a small number of firms. The most influential account of that flow, due to Levitt (2004), holds that books price away from the market-clearing point to exploit predictable public biases. It computes profit from two numbers (the probability that a side wins the proposition and the fraction of handle it attracts), treating the share on a side as independent of the outcome. Here we relax that assumption and derive a *profit–bias identity*: profit is affine and increasing in the expected share of handle on the losing side, with Levitt’s expression as the special case of independence. The identity resolves the book’s margin into exactly three channels: its hold, the product of its price shading and the public’s lean, and the covariance between bet share and outcome. A lean is thus worthless without shading, and shading is worthless without a lean. Under a public-belief model, the profit driver is the public’s Bayes error: the probability of a representative bettor selecting the losing side. We provide necessary and sufficient conditions for a “Goldilocks Zone”: prices at which book *and* bettor both profit. Testing these predictions on 1,139 Major League Baseball games, we find that the apparent dependence between bet share and outcome is a Simpson’s paradox: present when games are pooled, but absent once they are separated by which side the book favored. The public leans heavily toward favorites, but we detect no matching shading, and the realized margin is indistinguishable from the hold.

Introduction

Every account of how a sportsbook makes money rests on a claim about where the betting public’s money goes relative to the price at which it was offered. The most influential, due to Levitt [2004], posits that books deliberately price away from the market-clearing value to exploit predictable biases in public betting as opposed to balancing their exposure. That model, and the substantial literature testing it [Paul and Weinbach, 2007, 2008, 2009, 2011, 2012, Humphreys, 2010, 2011], computes the book’s expected profit from two numbers: the probability that the favorite covers, and the fraction of the handle it attracts. Two numbers suffice, as we show, only if the money and the outcome are statistically independent—only if the side drawing the handle is neither more nor less likely to cover than its posted price implies. To the best of our knowledge, this assumption has gone unexamined.

Treating the public’s money as unconnected to the outcome discards precisely the information that separates an informed betting public from an uninformed one, namely whether the side that attracts the money is more or less likely to win. This distinction is not incidental: a competing account attributes the bookmaker’s margin not to public bias but to adverse selection against privately informed bettors [Shin, 1991, 1992, 1993], a mechanism recently re-examined and weighed against ordinary bettor disagreement [Whelan, 2024, 2025]. By assuming independence between bet share and outcome, Levitt’s model cannot distinguish a *sharp* public, whose money follows the eventual winner (as the insider account would predict), from a *square* one, whose money follows the eventual loser.

Here we revise the Levitt model to allow a statistical dependence between betting behavior and match outcomes. Relaxing the independence restriction yields a *profit-bias identity*: the sportsbook’s expected profit is an affine function of the expected share of handle on the *losing* side, and it reduces to Levitt’s expression when bet shares are independent of the outcome. Because the identity presumes neither account, the sharp and square cases can be distinguished empirically rather than assumed away [Woodland and Woodland, 1994, Gandar et al., 2002, Ottaviani and Sørensen, 2008]. The identity further separates the book’s expected margin into three parts: the hold that a balanced book collects, a term in which the book’s price shading multiplies the public’s lean, and the covariance between bet share and outcome. Because the second term is the product of shading and lean, neither factor generates profit on its own: a lean earns nothing for a book that does not shade, and shading earns nothing when the public does not lean. Under a public-belief model in which bettors act on a noisy, biased signal of the game [Snowberg and Wolfers, 2010, Jullien and Salanié, 2000, Avery and Chevalier, 1999], the profit-driving quantity is the probability that the public selects the losing side, a misclassification (Bayes-error) probability. The same formulation characterizes a “Goldilocks Zone” of prices at which the sportsbook and an infinitesimal bettor can simultaneously profit [Shleifer and Vishny, 1997, Hubáček et al., 2019].

We tested these predictions with data from 1,139 Major League Baseball (MLB) games, finding that the apparent bias of public money onto the eventual losing side is a Simpson’s paradox: present when games are pooled, and absent once they are separated by which side the book has designated as the favorite. The public is square in the aggregate but not in either of its halves. The data do exhibit a strong favorite bias in public betting, but *without* an accompanying increase in book profit over its hold, which we connect to the absence of detectable price shading.

Results

Terminology

The forthcoming treatment employs the terminology of sports betting, which is reviewed in Table 1. Of note, the terms *lean* and *shading* refer to skews in public betting and sportsbook pricing, respectively, and are often conflated.

Setup and notation

We consider a generic two-sided sportsbook proposition, encompassing point spreads, totals, and player props. Let the realized outcome be a real-valued random variable $M \in \mathbb{R}$. For spreads, M can be the realized margin of victory (home score minus visitor score); for totals, M can be total points; for props, M can be a player performance statistic.

The sportsbook posts a line $s \in \mathbb{R}$. Relative to s , we write the possible outcomes as:

$$\text{home loses} \iff M < s, \quad \text{visitor loses} \iff M > s, \quad \text{push} \iff M = s.$$

The outcome CDF at s and push probability are given by:

$$F_m(s) \equiv \Pr(M < s), \quad \pi_{\text{push}}(s) \equiv \Pr(M = s), \quad (1)$$

such that the complementary CDF is $\bar{F}_m(s) \equiv \Pr(M > s) = 1 - F_m(s) - \pi_{\text{push}}(s)$.

Let the (random) fraction of total stake placed on the home side at price s be $B_h(s) \in [0, 1]$, and let the fraction placed on the visitor side be $B_v(s) \equiv 1 - B_h(s)$. We explicitly allow statistical dependence between betting behavior and the realized outcome (i.e., $B_h(s)$ may be informative about M). We define the conditional expected bet shares on the *eventual losing side* by:

$$b_{h,L}(s) \equiv \mathbb{E}[B_h(s) \mid M < s], \quad b_{v,L}(s) \equiv \mathbb{E}[B_v(s) \mid M > s]. \quad (2)$$

These quantities capture the fraction of the handle that (in expectation) ends up on the losing side, conditional on which side actually loses.

The profit-bias identity

We derive the per-game expected profit in the symmetric-vig ($\phi_h = \phi_v = \phi$), no-push case ($\pi_{\text{push}}(s) = 0$); the general treatment allowing asymmetric vigs $\phi_h, \phi_v \in [0, 1]$ and pushes is deferred to the SI (Sec. S1). Throughout, total stake is normalized to $B_h(s) + B_v(s) = 1$. The book’s single-game profit is $\pi(s) = B_h(s) - \phi B_v(s)$ when the home side loses, $\pi(s) = B_v(s) - \phi B_h(s)$ when the visitor loses, and $\pi(s) = 0$ in the event of a push. Conditioning on the realized outcome and taking total expectation gives:

$$\mathbb{E}[\pi(s)] = F_m(s) \mathbb{E}[B_h(s) - \phi B_v(s) \mid M < s] + \bar{F}_m(s) \mathbb{E}[B_v(s) - \phi B_h(s) \mid M > s]. \quad (3)$$

Table 1: **Betting terminology.** The terms are defined in the point-spread setting, where M is the realized outcome, b_h the public’s home share of handle, p^* the price-implied probability that the home side wins the bet (with the book’s commission removed), and $\bar{F}_m = \Pr(M > s)$ is the corresponding true probability.

Term	Symbol	Meaning
Spread	s	A number posted by the book. The spread is a handicap applied to one side’s score so that a wager on either is closer to a 50/50 proposition. The bettor’s task is to predict whether the outcome will exceed the spread: $M > s$.
Price	ϕ_h, ϕ_v	The profit on a winning unit wager on the home and away sides, respectively. At the standard -110 price $\phi_h = \phi_v = 100/110 = 0.909$: the bettor risks 1.10 to win 1. The special case $\phi_h = \phi_v \equiv \phi$ is referred to herein as <i>symmetric vig</i> .
Cover	$M > s$	A side “covers” when it wins the match taking into account the spread – also called “beating the spread”. We take the home side to cover when $M > s$ and to lose the bet when $M < s$.
Push	$M = s$	A “push” occurs when the realized outcome is equal to the spread. In this event, all wagers are returned.
Handle	—	The total amount staked. Book profit is reported per unit of handle, such that a margin of 4% is four cents kept per dollar wagered.
Vig	—	The book’s commission, embedded in the prices rather than charged separately. The vig is the amount by which the sum of the two price-implied probabilities exceeds one: $(1 + \phi_h)^{-1} + (1 + \phi_v)^{-1} - 1$. For $\phi_h = \phi_v = \phi$, the vig is $(1 - \phi)/(1 + \phi)$, or 4.8% at -110 . Note that “the vig” is often used inconsistently in practice: sometimes as this excess, sometimes as the 10% surcharge on the stake, or sometimes as the hold below.
Hold	h	The share of handle that the book keeps when its liabilities are balanced: its return from the vig alone, $h = (1 - \phi_h \phi_v)/(2 + \phi_h + \phi_v)$, or $(1 - \phi)/2$ under a symmetric vig, giving 4.5% at -110 .
Shading	θ	The deviation of the book’s posted price from the truth, $\theta = p^* - \bar{F}_m$. Positive when the home side is priced as more likely to win than it truly is.
Lean	λ	The public’s displacement from balanced betting: $\lambda = b_h - p^*$. Positive when public betting favors the home side beyond what is implied by the book price.
Square	$\delta > 0$	A public whose money tends to be wagered on the eventual <i>loser</i> , so that the book profits beyond its hold.
Sharp	$\delta < 0$	A public whose money tends to be wagered on the eventual <i>winner</i> (book profits less than the hold, or takes a loss).

We note also that the term “line” is the generalization of the spread s to markets such as the point total or player props. In all cases it is a number set by the book that dictates the threshold determining which side wins the wager.

Using $B_v = 1 - B_h$ and the conditional losing-side shares (2), the two conditional expectations equal $(1 + \phi) b_{h,L}(s) - \phi$ and $(1 + \phi) b_{v,L}(s) - \phi$, respectively, so that:

$$\mathbb{E}[\pi(s)] = F_m(s) [(1 + \phi) b_{h,L}(s) - \phi] + \bar{F}_m(s) [(1 + \phi) b_{v,L}(s) - \phi]. \quad (4)$$

Collecting the $(1 + \phi)$ terms and using $F_m(s) + \bar{F}_m(s) = 1$ in the no-push case yields the *profit-bias identity*:

$$\boxed{\mathbb{E}[\pi(s)] = (1 + \phi) Q(s) - \phi}, \quad (5)$$

where:

$$Q(s) \equiv b_{h,L}(s) \Pr(M < s) + b_{v,L}(s) \Pr(M > s) \quad (6)$$

is the expected fraction of handle wagered on the losing side. Because $b_{h,L}(s)$ and $b_{v,L}(s)$ are conditional losing-side shares, $Q(s)$ admits arbitrary dependence between betting and outcome, and the book’s expected profit is affine and strictly increasing in it.

The profit-bias identity generalizes the bookmaker-profit expression of Levitt [2004]. Namely, when bet shares are independent of the outcome, the two coincide. However, whenever public betting covaries with the realized outcome, the expressions deviate. Levitt’s book profit underestimates the true value whenever the public allocates a larger stake on the eventual losing side, while overestimating when the stake favors the winning side. The mathematical relationship between the profit-bias identity and Levitt’s bookmaker profit is detailed in SI, Sec. S2.

Independence of bet shares from the outcome is the pair of conditions: $b_{h,L}(s) = b_h(s)$ and $b_{v,L}(s) = b_v(s)$, where $b_h(s) \equiv \mathbb{E}[B_h(s)]$ and $b_v(s) \equiv 1 - b_h(s)$ are the unconditional shares. Denoting the winning-side home

share by $b_{h,W}(s) \equiv \mathbb{E}[B_h(s) \mid M > s]$, the unconditional share is a convex combination of the two: $b_h(s) = F_m(s) b_{h,L}(s) + \bar{F}_m(s) b_{h,W}(s)$. Note that $b_{h,L}(s) = b_h(s)$ can only be satisfied when $b_{h,L}(s) = b_{h,W}(s)$. Similarly, $b_{v,L}(s) = b_v(s)$ requires $b_{v,L}(s) = b_{v,W}(s)$. The two conditions therefore collapse to:

$$\delta(s) \equiv b_{h,L}(s) - b_{h,W}(s) = 0, \quad (7)$$

which is the discrepancy between the expected home stake share in games the home side loses versus wins. Thus, in order to test whether bet shares covary with realized outcomes, one evaluates $\delta(s)$ against zero.

The identity (5) attributes the book's entire margin to $Q(s)$, but does not afford insight into the sources of its variation. To that end, we write the book's *shading* and the public's *lean* as signed deviations:

$$\theta(s) \equiv F_m(s) - \frac{1}{2}, \quad \lambda(s) \equiv b_h(s) - \frac{1}{2}, \quad (8)$$

such that $\theta(s) > 0$ when the spread exceeds the median outcome and $\lambda(s) > 0$ when public betting favors the home side. This allows us to decompose the losing-side share into three distinct and interpretable components.

Proposition 1 (Anatomy of the book's margin). *Under a symmetric vig ϕ and no pushes:*

$$Q(s) = \frac{1}{2} + 2\theta(s)\lambda(s) + 2F_m(s)\bar{F}_m(s)\delta(s), \quad (9)$$

and therefore:

$$\mathbb{E}[\pi(s)] = \underbrace{\frac{1-\phi}{2}}_{\text{hold}} + 2(1+\phi) \left[\underbrace{\theta(s)\lambda(s)}_{\text{shading} \times \text{lean}} + \underbrace{F_m(s)\bar{F}_m(s)\delta(s)}_{\text{outcome covariance}} \right]. \quad (10)$$

The book's expected margin exceeds its hold if and only if the bracket is strictly positive.

The general expression allowing asymmetric vigs and pushes is developed in the SI Sec. S1. Two consequences follow from (10).

Corollary 2 (Public lean only increases book profit if matched with a shaded line.). *If the book prices at the median, then $F_m(s) = \bar{F}_m(s) = \frac{1}{2}$ and $\theta(s) = 0$, yielding:*

$$Q(s) = \frac{1}{2}(1 + \delta(s)), \quad \mathbb{E}[\pi(s)] = \frac{1-\phi}{2} + \frac{1+\phi}{2}\delta(s),$$

regardless of the public's lean $\lambda(s)$. Thus, when pricing accurately, the book exceeds its hold if and only if $\delta(s) > 0$.

Corollary 3 (Price shading only increases book profit if paired with a public lean.). *If bet shares are independent of the outcome, $\delta(s) = 0$, then $\mathbb{E}[\pi(s)] = (1-\phi)/2 + 2(1+\phi)\theta(s)\lambda(s)$, which exceeds the hold if and only if $\theta(s)$ and $\lambda(s)$ are non-zero and have the same sign. If public betting is balanced, price shading does not increase book profit.*

By making use of the identity $b_h(s) = F_m(s) b_{h,L}(s) + \bar{F}_m(s) b_{h,W}(s)$, one can show that $\text{Cov}(B_h(s), \mathbf{1}\{M < s\}) = F_m(s)\bar{F}_m(s)\delta(s)$, meaning that the third term in 9 is equal (up to a constant) to the covariance between the home bet share and whether the home side loses:

$$Q(s) = \frac{1}{2} + 2\theta(s)\lambda(s) + 2\text{Cov}(B_h(s), \mathbf{1}\{M < s\}). \quad (11)$$

Note that the second term of the right-hand side is a product of marginal expectations while the third term is a covariance. In other words, the book profit's excess factors into a product of marginals and a covariance term, which is standard for a joint expectation of random variables. Although both terms may be viewed as "public bias", the two terms have very distinct interpretations: $\theta(s)\lambda(s)$ is a directional bias that is theoretically observable before the match takes place, while $\text{Cov}(B_h(s), \mathbf{1}\{M < s\})$ is outcome-covariant and can only be resolved after.

It is also instructive to consider the extreme case of a public that always backs the home side. Here the home bet share is clearly independent of the outcome, and $b_h(s) = b_{h,L}(s) = b_{h,W}(s) = 1$. Despite the fact that $\delta(s) = 0$, the book can still profit above its hold if $F_m(s) > \frac{1}{2}$. This is precisely the mechanism presented by Levitt [2004] and the second term of 11. Because Levitt implicitly assumes independence between bet shares and outcomes, $\delta(s) = 0$ and the product $\theta\lambda$ is the only mechanism by which the book can profit above its hold in his account.

The profit driver as a misclassification probability

Note that the profit-bias identity (5) makes no assumptions about how the public chooses sides beyond the existence of the conditional expectations in (2). All market- and behavior-specific structure enters through $Q(s)$, which depends on the joint distribution of $(B_h(s), M)$. Here we employ a simple public-belief model and show that it leads to a decision-theoretic interpretation of $Q(s)$.

Consider the point spread setting where M models the difference between home and away score, s is the sportsbook's posted spread, and assume that the public forms a noisy, biased belief Y about the (latent) expected outcome X :

$$Y = X + \varepsilon + V, \quad M = X + U, \quad (12)$$

where ε is the public's systematic bias, V is belief noise, and U captures noise on the outcome that is assumed to be independent of V . We further assume that the public backs the home side whenever $Y > s$, such that the home team's bet share is captured by a representative agent: $B_h(s) = \mathbf{1}\{Y > s\}$. In other words, the public acts on the aggregate signal Y and places its stake on one side $-B_h(s) \in \{0, 1\}$. With this selection rule, the conditional losing-side shares $b_{h,L}(s)$ and $b_{v,L}(s)$ of (2) become conditional misclassification probabilities, and $Q(s)$ takes on a decision-theoretic interpretation.

Proposition 4 (Profit driver as a misclassification probability). *Under the public-belief model with selection rule $B_h(s) = \mathbf{1}\{Y > s\}$, the profit driver of the profit-bias identity satisfies:*

$$Q(s) = \Pr(\text{public backs the losing side}).$$

$Q(s)$ is thus the misclassification probability of the binary predictor $\mathbf{1}\{Y > s\}$ for the event $M > s$. As a result, the sportsbook's expected profit is an affine, strictly increasing function of the public's classification error in predicting the sign of $M - s$.

Consequently, the book profits to the extent that the public's belief signal fails to classify the outcome. The proof of Proposition 4, as well as a closed-form evaluation of $Q(s)$ in terms of the belief and outcome distributions, are given in the SI (Sec. S4).

Conditions for the Existence of a “Goldilocks Zone”

We ask when a posted spread s is simultaneously attractive to the sportsbook and to a small bettor: that is, when the book earns a profit in expectation *and* an individual bettor (whose wager is negligible relative to the public aggregate) can find at least one side with positive expected value at the posted price. We assume fractional spreads ($\pi_{\text{push}}(s) = 0$) and symmetric vigs ($\phi_h = \phi_v = \phi$ with $\phi \in (0, 1)$). For notational convenience, we define:

$$\tau \equiv \frac{\phi}{1 + \phi} \in (0, \frac{1}{2}), \quad (13)$$

such that τ and $1 - \tau = \frac{1}{1 + \phi}$ are symmetric about $\frac{1}{2}$.

From (5), the sportsbook's expected profit is positive if and only if:

$$Q(s) > \tau. \quad (14)$$

Note from (6) that $Q(s)$ is a convex combination of the two conditional losing-side shares, weighted by the outcome CDF:

$$Q(s) = (1 - F_m(s)) b_{v,L}(s) + F_m(s) b_{h,L}(s). \quad (15)$$

Because $F_m(s) \in [0, 1]$, it follows that $Q(s)$ is bounded between the two losing bet shares:

$$\min(b_{h,L}(s), b_{v,L}(s)) \leq Q(s) \leq \max(b_{h,L}(s), b_{v,L}(s)). \quad (16)$$

Thus, whether the book profits depends on where the threshold τ is situated relative to $b_{h,L}(s)$ and $b_{v,L}(s)$.

Proposition 5 (Book-profit regimes). *Fix a spread s and let $b_{\min}(s) = \min(b_{h,L}, b_{v,L})$ and $b_{\max}(s) = \max(b_{h,L}, b_{v,L})$. The book-profit condition (14) obeys exactly one of three regimes:*

- (i) **Always** ($b_{\min}(s) > \tau$): $Q(s) > \tau$ for every value of $F_m(s)$.
- (ii) **Never** ($b_{\max}(s) \leq \tau$): $Q(s) \leq \tau$ for every value of $F_m(s)$.
- (iii) **Threshold** ($b_{\min}(s) \leq \tau < b_{\max}(s)$): as $F_m(s)$ varies from 0 to 1, $Q(s)$ increases monotonically from $b_{v,L}(s)$ to $b_{h,L}(s)$, crossing τ only once at:

$$F^*(s) = \frac{\tau - b_{v,L}(s)}{b_{h,L}(s) - b_{v,L}(s)} \in [0, 1].$$

The book profits on whichever side of this crossing puts more of the weight on the larger of the two shares:

$$Q(s) > \tau \iff \begin{cases} F_m(s) > F^*(s), & \text{if } b_{h,L}(s) > b_{v,L}(s), \\ F_m(s) < F^*(s), & \text{if } b_{h,L}(s) < b_{v,L}(s). \end{cases}$$

In particular, $b_{\min}(s) > \tau$ is sufficient and $b_{\max}(s) > \tau$ is necessary for the book to profit at s .

The proof of Proposition 5 is in the SI (Sec. S5). The key idea is that book profitability is governed by the values of the conditional losing-side shares relative to the threshold τ : the book always profits whenever the public’s losing-side stake exceeds τ on *both* outcomes ($b_{\min} > \tau$), while never profiting when it fails to exceed τ on *either* ($b_{\max} \leq \tau$). In the remaining case, book profitability is determined by the value of F_m relative to F^* at the posted spread s .

We define the “Goldilocks Zone” as the set of spreads at which the book *and* a small bettor can profit:

$$\mathcal{G} \equiv \{s : Q(s) > \tau\} \cap \mathcal{B}, \quad \mathcal{B} \equiv \{s : \text{a negligible bettor has a positive expected value wager at } s\}. \quad (17)$$

The conditions under which a bettor may find a profitable wager have been previously derived [Dmochowski, 2023]. Briefly, a unit wager nets ϕ on a win and -1 on a loss, such that a bet has positive expected profit only when its win probability exceeds $1/(1 + \phi) = 1 - \tau$. The win probabilities of the home and away sides are given by $\bar{F}_m(s)$ and $F_m(s)$, respectively, leading to:

$$\mathbb{E}[\text{home bet}] > 0 \iff F_m(s) < \tau, \quad \mathbb{E}[\text{visitor bet}] > 0 \iff F_m(s) > 1 - \tau. \quad (18)$$

These conditions amount to s lying in a *tail* of the outcome distribution F_m . In other words, the bettor may profit only when the outcome is extreme enough to overcome the vig:

$$\mathcal{B} = \{s : F_m(s) < \tau\} \cup \{s : F_m(s) > 1 - \tau\}. \quad (19)$$

Existence of a Goldilocks Zone. Combining the book profitability conditions of Proposition 5 with (19) yields the following (distribution-free) conditions on the existence of a non-empty Goldilocks Zone:

- **Sufficient.** If, for at least one s , the bet shares satisfy $b_{\min}(s) > \tau$ and $F_m(s)$ lies in a tail ($F_m(s) < \tau$ or $F_m(s) > 1 - \tau$), then $s \in \mathcal{G}$ and the Goldilocks Zone is non-empty.
- **Necessary.** Any $s \in \mathcal{G}$ must satisfy $b_{\max}(s) > \tau$ and $F_m(s)$ in a tail; where $b_{\max}(s) \leq \tau$, the book cannot profit and $\mathcal{G}$ is empty regardless of F_m .

For the threshold regime $b_{\min}(s) \leq \tau < b_{\max}(s)$, non-emptiness of $\mathcal{G}$ requires that the tail constraint on $F_m(s)$ and the “crossover” constraint ($F_m(s)$ vs. $F^*(s)$) hold at the *same* spread. Importantly, the quantities $F_m(s)$, $b_{h,L}(s)$, and $b_{v,L}(s)$ all covary with s , such that existence is determined by the joint trajectory $s \mapsto (F_m(s), b_{h,L}(s), b_{v,L}(s))$, the object that we estimate in the forthcoming empirical section.

Visualizing the profit driver and the Goldilocks Zone

To visualize the theoretical results, we modeled the outcome distribution $F_m(s)$ and the public’s home bet share $b_h(s)$ with toy Gaussian models (see *Methods* for details). In the schematics of Fig. 1, the horizontal axis is a *candidate* point spread that the sportsbook may post for a match whose median outcome is taken to be zero (without loss of generality); moving the spread to the right (left) increases the handicap on the home (away) team. The vertical axis is the profit driver $Q(s)$ – the expected share of money on the eventual losing side – and the book profits when $Q(s)$ exceeds the vig-defined threshold τ . We examine how book profit responds to four types of public betting behaviors: a *calibrated* public whose beliefs are aligned with the outcome distribution (panel a); a *favorite-leaning* public whose bias toward the stronger team (the side the sportsbook favors) places a preponderance of money on the favorite (panel b); a *square* public that allocates a higher bet share to the eventual losing side (panel c); and a *sharp* public that allocates a higher bet share to the eventual winning side (panel d). It is important to note that the first two behaviors are independent of the realized outcome, whereas the last two convey a statistical dependence between betting and outcome.

A calibrated public confines the book’s profitable region (i.e., the region of the curve where $Q(s)$ exceeds τ) to an interior band around the median ($|s| < \epsilon$; Fig. 1a). Here the book profits *only by pricing accurately*: its profitability requires that the posted spread be near the true median. The corresponding Goldilocks regions (green bands) are situated on either side of the median, where the posted spread differs from the median by a minimum amount without exceeding ϵ .

A favorite-leaning public – one that backs the favorite in excess of its fair probability – produces a qualitative change in book profit (Fig. 1b): the book now profits at the tails of the candidate spread where the public’s bias is largest. Note that here the profit curve opens upward (U-shaped). This arises without dependence between betting and realized outcome: it is a structural phenomenon, rooted in the the public’s tendency to overvalue favorites. The Goldilocks region widens correspondingly, extending across the tails.

Introducing a dependence between betting and outcome produces vertical shifts of the profit curve; we model it with a constant δ between the conditional bet shares (solid line, home loses; dashed line, home wins). A *square* public, whose

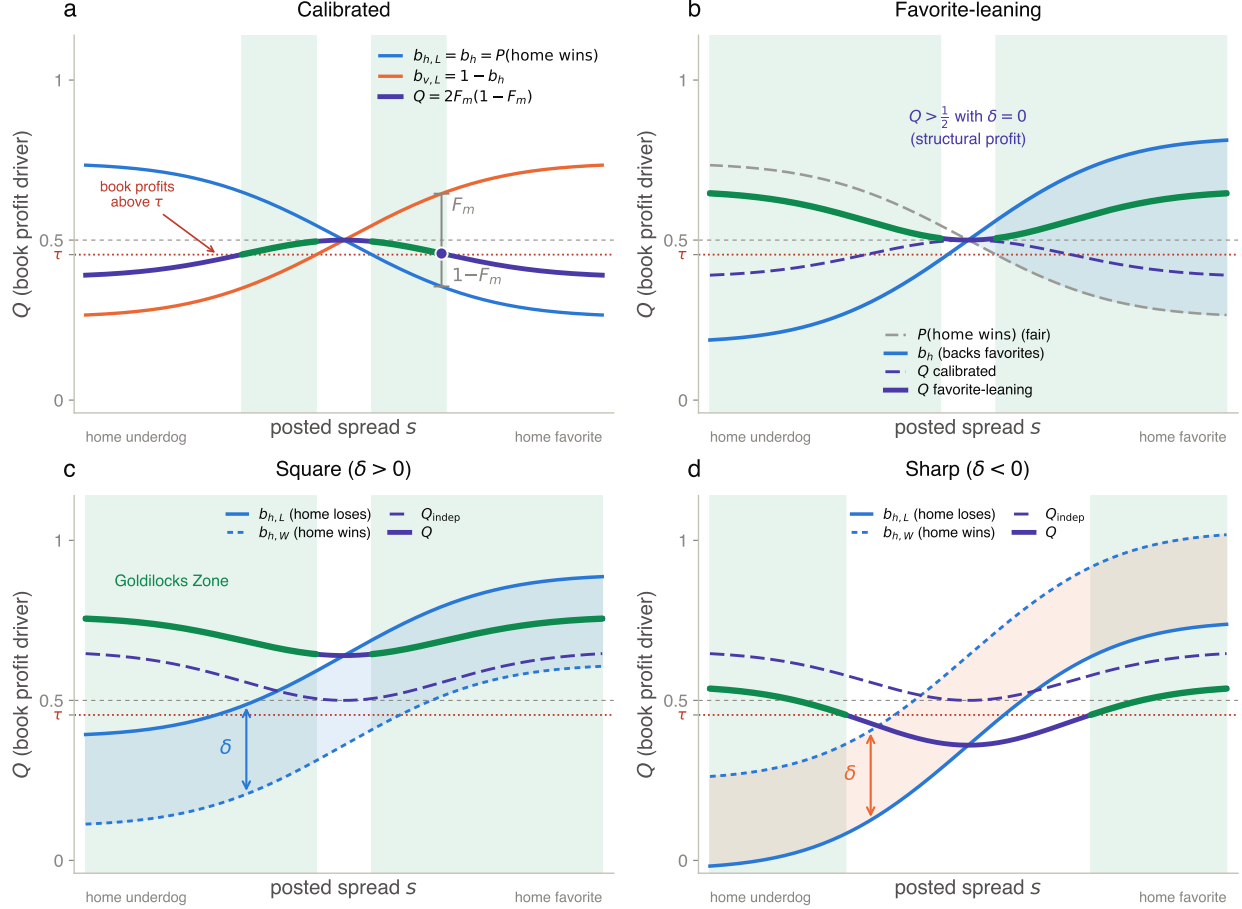


Figure 1: **The sportsbook profit driver $Q(s)$ across four public behaviours.** In every panel $Q(s) = F_m(s) b_{h,L}(s) + \bar{F}_m(s) b_{v,L}(s)$ is the F_m -weighted average of the home losing-side share $b_{h,L} = \mathbb{E}[B_h | M < s]$ and the visitor losing-side share $b_{v,L} = \mathbb{E}[B_v | M > s]$; the dotted line marks the threshold τ , above which the book profits. **(a)** A *calibrated* public bets each side at its actual win probability. The bracketed connector shows Q dividing the gap between the two shares at a ratio of $F_m : (1 - F_m)$. The inverted-U curve is upper bounded by $\frac{1}{2}$ and falls below the profit threshold at the tails of the posted spread. **(b)** A *favorite-leaning* public backs favorites in excess of their fair cover probability. Even with the betting fully independent of outcome ($\delta = 0$), Q now opens up and is markedly increased at the tails of the candidate spread. **(c)** A *square* betting public systematically allocates a larger bet share on the eventual losing side ($\delta = b_{h,L} - b_{h,W} > 0$), increasing book profit above a purely favorite-leaning public. **(d)** A *sharp* betting public allocates a greater bet share on the eventual winning outcome ($\delta < 0$), driving the book profit towards the threshold τ . In all panels, the *Goldilocks* region (green) marks the set of candidate spreads where both an individual bettor and the book profit. For a calibrated public (a), it is limited to two interior bands; a favorite-leaning public (b) yields Q above τ across the distribution tails and thus widens the Goldilocks Zone. Square dependence (c) increases the magnitude of the book's profit without altering the extent of the Goldilocks Zone; with a sharp betting public (d), the zone is relegated to more extreme spreads. All curves are schematics generated with Gaussian models for both outcome and public belief.

money skews toward the eventual loser (Fig. 1c), quantitatively increases the magnitude of the book's profit without greatly changing the extent of the Goldilocks Zone. Conversely, a *sharp* public reduces book profit across the range of spreads, and in the schematic shown drives the book to a loss over a substantial interior region (Fig. 1d); the Goldilocks region is correspondingly limited to the extreme tails of the candidate spread.

Taken together, these panels relate book profitability to three mechanisms: (i) accurate forecasting of the outcome distribution; (ii) structural biases in public betting, such as over-valuing favorites; and (iii) any conditional dependencies between betting and outcome. The empirical section below attempts to shed light onto which of these mechanisms may manifest in real betting markets.

Empirical results

To investigate the relationship between sportsbook profit, public betting, and game outcomes, we collected Major League Baseball (MLB) market prices and bet shares at a major North American sportsbook over an eleven-week duration. The specific market analyzed here, termed the “run line,” fixes the favorite’s handicap at 1.5 runs and instead expresses the teams’ relative strength through the payouts ϕ_h and ϕ_v . Note that this is unlike a standard point spread market where payouts on the two sides are approximately equal. The bettor’s task is thus to predict whether the home team will win by more than (or lose by less than) 1.5 runs at a price reflecting the perceived difference in team strengths. Prior tests of Levitt’s model have employed empirical bet shares (sometimes termed betting “splits”) [Paul and Weinbach, 2007, 2008, 2012, Humphreys et al., 2013]. Here we add a temporal dimension, observing each line and its associated bet shares from posting to game onset, as in previous high-frequency line-movement studies [Simon, 2024, Gandar et al., 1998].

We asked the following questions: (i) Do bet shares covary with realized outcomes? (ii) Does the bet share-outcome covariance evolve across time? (iii) Do bet shares drive prices, or do prices drive bet shares, or both? (iv) What are the contributions of price shading, public lean, and outcome covariance to the book’s profit?

Bet share-outcome dependence in MLB run lines: a Simpson’s paradox Across $n = 1,139$ games, the betting public displayed a striking preference for the match favorite: the home bet share varies sigmoidally with the sportsbook’s implied win probability (Fig. 2a), increasing from below 25% when the away team is strongly favored to above 80% when the home team is. In contrast, the sportsbook’s implied home win probabilities span a compressed central range (0.39–0.67).

We computed the home team’s bet share separately for games that the home team won versus lost against the spread. This revealed an apparent discrepancy: in games where the home side lost the proposition, the home bet share was 5.8 percentage points higher (56.4% versus 50.6%; game-level bootstrap $p = 0.0004$; $n_L = 584$, $n_W = 555$). This finding may be interpreted as evidence for a square betting public. However, stratifying the analysis by the identity of the favorite (home versus away favorite) removed the effect: the home bet share showed no significant dependence on the realized outcome in either stratum, and both point estimates were small (home favorites, $\hat{\delta} = -1.5$ pp, $p = 0.51$, $n = 608$; away favorites, $\hat{\delta} = +0.6$ pp, $p = 0.79$, $n = 531$).

The apparent dependence is thus a Simpson’s paradox: a correlation in pooled data that is removed upon conditioning on a confounding variable. Here the confounder is the identity of the favorite: when the sportsbook designates the home team as the favorite, both the home bet share b_h and the probability of the home team losing against the spread $F_m(s)$ are increased. The latter is due to the fraction of home team losses being slightly larger than that implied by book prices. We elaborate on the implications of this structural form of book profit in the *Discussion*.

Dynamics of book profit and public belief Public betting and sportsbook price setting operate as a closed loop, producing fluctuations in both quantities prior to the start of the game. Focusing on the MLB run line market, we collected hourly samples of the point spread s , payouts (ϕ_h , ϕ_v), and bet shares (b_h , b_v) in the 24 hours leading up to game onset. Within each stratum (home favorite versus home underdog), we calculated the book’s profit driver $Q(s)$ as well as the public’s outcome dependency $\delta(s)$ in non-overlapping 6 hour windows.

The book’s profit driver evolves distinctly across the two strata. For matches with a home underdog, $\hat{Q}$ increases from $46.9\% \pm 2.6$ at posting to $52.0\% \pm 2.1$ at game onset – the initial “sharp” betting is not sustained. In matches with a home favorite, the losing side share is reliably above 50% throughout ($54.4\% \pm 2.4$ to $53.8\% \pm 2.0$, mean $\pm$ SEM).

We asked whether the outcome-conditioned bias δ exhibits significant movement across time, employing a linear mixed model with time and outcome as factors and the home bet share as the dependent variable. A significant interaction between time and outcome indicates that public belief carries information about the eventual outcome that is not constant over the pre-game window – a dynamic departure from Levitt’s implicit independence. For home underdogs, we found a significant decreasing slope, indicating that the home bet share decreases for games in which the home team subsequently wins against the spread ($\hat{b}_{h,W}$ falling from $\approx 44\%$ a day out to $\approx 36\%$ at first pitch; game-clustered bootstrap $p = 0.038$, cluster-robust $p = 0.038$, $n = 256$ games). This finding is suggestive of a square betting pattern that emerges closer to the start of the game. The corresponding test for home favorites was non-significant (game-clustered bootstrap $p = 0.85$, $n = 288$ games). We note the two strata’s slopes do not significantly differ (three-way outcome $\times$ time $\times$ stratum interaction, game-clustered bootstrap $p = 0.12$) – the dynamic departure from independence is specific to home underdogs.

Feedback between book pricing and public betting The public-belief model treats the posted price as an input to the public’s wager rather than a response to it – this ordering can be tested directly. We estimated Jordà local projections

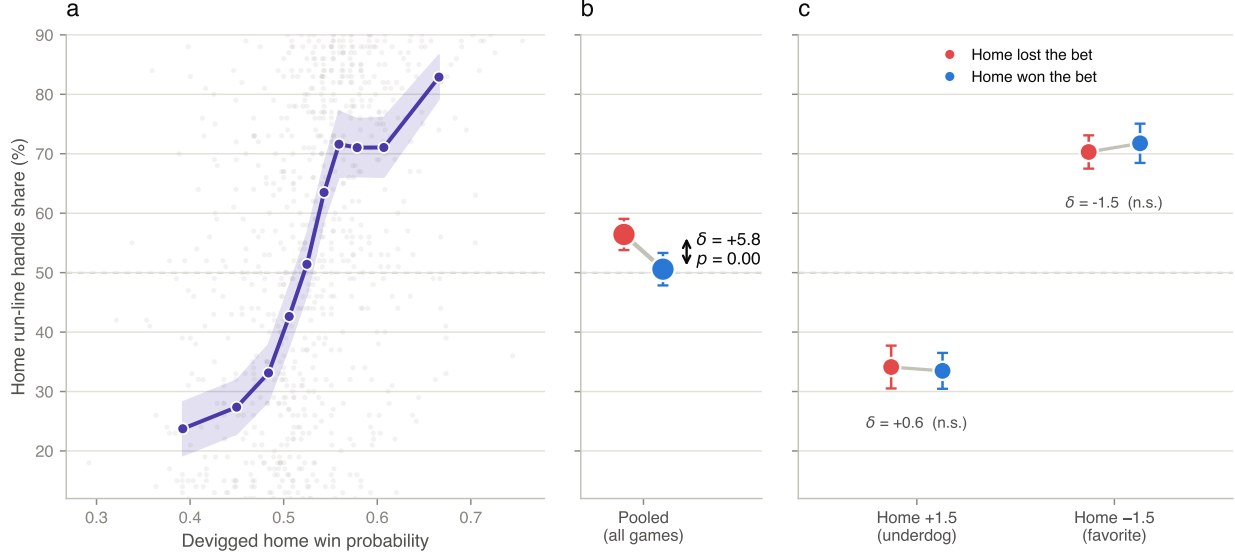


Figure 2: A Simpson’s paradox explains the apparent bet share-outcome covariance in the MLB run line market. We collected market prices and bet shares from the MLB run line market across $N = 1,139$ games. **(a)** The public shows a striking preference for favorites: the home bet share varies sigmoidally with the price-implied home win probability; light grey markers denote individual games, while the curve shows the binned mean with 95% confidence band. **(b)** Computed over all games, the home bet share appears to covary with the realized outcome: $\hat{\delta} = \hat{b}_{h,L} - \hat{b}_{h,W} = 5.8$ points ($p = 0.0004$): more money is wagered on the home side in games that the home team eventually lost ($\hat{b}_{h,L}$, red) than won ($\hat{b}_{h,W}$, blue). **(c)** However, stratifying the analysis by the identity of the favorite eliminates the effect: $\hat{\delta}$ is -1.5 for home favorites and $+0.6$ for home underdogs (both n.s.). Because a home favorite attracts a majority of the bet share *and* wins the proposition less often than the price-implied win probability, the pooled bet share-outcome covariance is spurious.

[Jordà, 2005] of each series on a one-step innovation in the other ((22); see *Methods*). A local projection regresses the level of one variable at horizon h on an innovation in the other, yielding the impulse response horizon by horizon rather than extrapolating it from a fitted dynamic model. A game fixed effect absorbs the latent state that sets both the opening price and the public’s prior, and a pre-move momentum control blocks the reverse channel, so that under sequential ignorability the horizon- h coefficient is the response to the move itself.

On the present dataset, the two directions were markedly asymmetric. A move in the book’s line was followed by a small “fade” of the public: a one-percentage-point increase in the devigged home win probability lowered the home bet share by roughly 0.2 points, significant at horizons of one to five hours and peaking at $h = 4$ ($\hat{\beta} = -0.20$, $t = -3.11$, $p = 0.002$, $n = 1090$ games; Fig. 4a). As expected, when wagering on the home side becomes more expensive, the public’s tendency to do so lowers. The converse channel, however, was absent: the line’s response to a one-point move in the home share was flat at zero and insignificant at every horizon ($|\hat{\beta}| \leq 0.010$ points, all $p > 0.27$, $n = 1103$ games; Fig. 4b). The price thus leads the money rather than the reverse, matching the ordering assumed by the belief model.

The book’s profit margin is indistinguishable from its hold. Propositions 1 and S1 dictate that a book profits beyond its hold if public betting favors a side that is overpriced, or if bet shares covary negatively with the realized outcome. We asked whether either of these two conditions may be present in our dataset. The book earned a mean profit 5.20% of handle per game (95% CI [0.88, 9.35], $N = 1,139$; Fig. 5a), compared to the 4.39% hold that even a perfectly balanced book would collect – on this sample, the book’s profit is thus indistinguishable from the hold. Moreover, as is evident in the figure, the per-game profit exhibits very high dispersion (s.d. 73.8% of handle).

Stratifying games by the identity of the favorite, the bet share on the favored side is +29.6 pp higher when the home side is the favorite and -24.5 pp higher when the visitor is favored (Fig. 5b). Public betting on the favorite thus exceeds the book’s implied price by 24–30 percentage points, reflecting a bias that has been previously well documented [Levitt, 2004, Paul and Weinbach, 2007, Franck et al., 2011]. Note from Corollary 3 that a public lean requires a corresponding

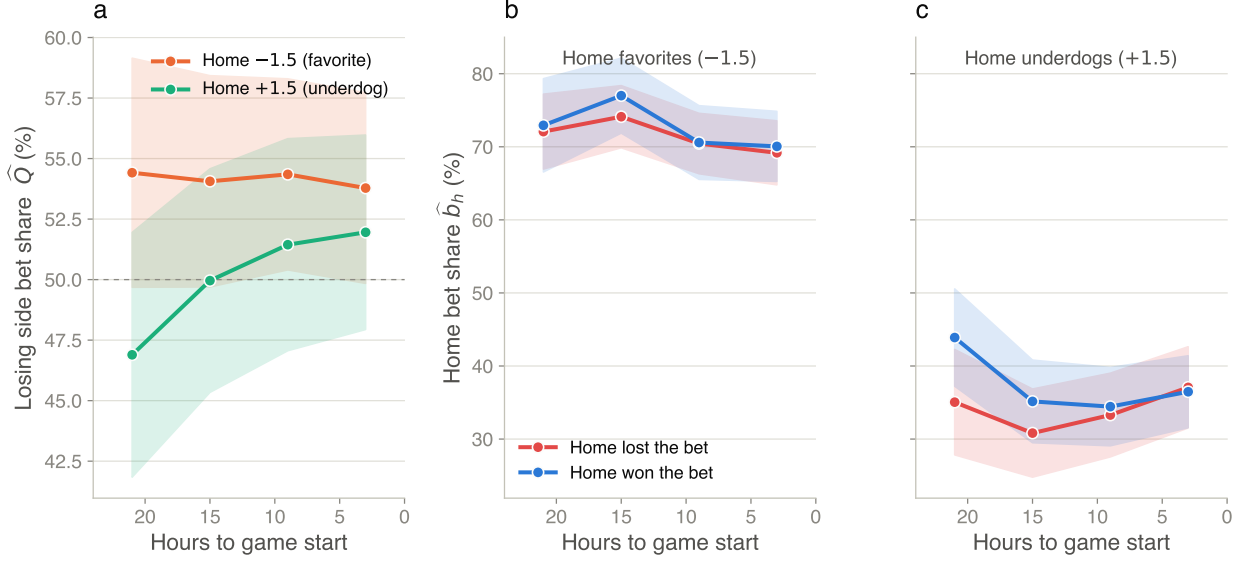


Figure 3: **Temporal evolution of public bias and book profit during the pre-game window.** We collected hourly samples of public bet shares and prices for the MLB run line market. To gain insight into the temporal dynamics of book profit and public belief, we then estimated the fraction of share on the losing-side $\hat{Q}$ in non-overlapping six hour windows terminating at game onset. The analysis was stratified into home-favorite ($N = 288$) and home-underdog ($N = 256$) games. **(a)** The share of money on the eventual losing side $\hat{Q}(t)$ by stratum: for home underdogs (green), the losing share increases from $46.9\% \pm 2.6$ to $52.0\% \pm 2.1$, while for home favorites (orange) it remains stable and above 50% ($54.4\% \pm 2.4$ to $53.8\% \pm 2.0$). **(b,c)** The home bet share in games that the home side eventually *lost* ($\hat{b}_{h,L}$, red) versus *won* ($\hat{b}_{h,W}$, blue) the closing bet, for home favorites (b) and home underdogs (c); the gap between the two curves is the outcome-conditioned bias $\hat{\delta}(t)$. For games with a home underdog, the home bet share separates by outcome early ($t = -24$ to $t = -18$ h: $\hat{\delta} = -8.9$ points; more home money on the eventual winner) but converges by the start of the game. The observed change in $\hat{\delta}$ is significant (linear mixed model with time and outcome as factors, time $\times$ outcome interaction: bootstrap $p = 0.038$, $n = 256$), suggesting an increase in square betting leading up to the start of the game. Games with a home favorite show little separation at any time ($p = 0.85$); the difference between the two strata’s slopes is itself not significant ($p = 0.12$).

book shading to yield excess profit. Thus, the empirical finding that the book does not profit above its hold despite the strong favorite-bias suggests that the book’s prices are calibrated.

To test this, we estimated the book’s mispricing $\hat{\theta} = p^* - \bar{F}_m$ across the price range and contrasted it with the case of a perfectly calibrated book. We found no departure from calibration: the average mispricing is +0.48 pp (95% CI $[-2.42, +3.38]$), and the largest local departure, 5.8 pp, is within the 9.4 pp that a calibrated book would produce by chance at this sample size ($p = 0.54$). The posted prices are indeed consistent with being calibrated (but see the Discussion for limitations).

For each excess-profit-generating mechanism (shading $\times$ lean, outcome covariance), we asked what the largest level of that mechanism is that remains consistent with the observed data. For shading, we assumed that the book shades in the maximally profitable direction – the side that the public backs is priced as more likely to cover than it truly is – and calculated the largest shading whose implied profit margin still falls inside the book profit confidence interval. Because the public’s lean is large, (counterfactual) shading would be highly profitable: 1.35% of handle per percentage point. Consequently, the upper bound on price shading is tight: at most 3.7 pp (Fig. 5c), meaning that the devigged book price departs from the true cover probability by no more than 0.037 in the direction of the public lean.

We repeated the analysis for outcome-covariance, calculating the largest δ compatible with the observed book profit. The bound is again modest: 5.2 pp (Fig. 5d), implying that the public’s home bet share differs by at most 5.2 percentage points between games won versus lost by the home side.

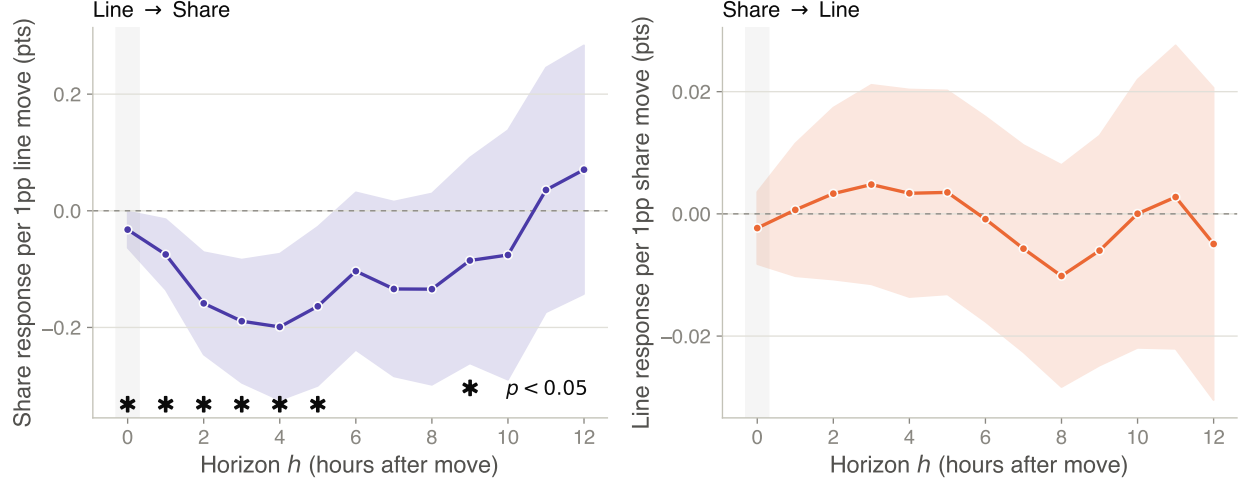


Figure 4: **Book pricing drives public betting, but not the reverse.** To investigate causal coupling between sportsbook pricing and public betting, we computed Jordà local projections of the book price and bet share time series across on an hourly grid ($n = 1103$ games; game fixed effect, pre-move momentum control). The response of each series at horizon h (number hours after the change in the other signal) is shown with 95% CI, and the shaded strip indicates the contemporaneous $h = 0$. **(a) Line $\rightarrow$ share:** a 1-percentage-point increase in the devigged home win probability is followed by a small but significant reduction of the home bet share ($\hat{\beta} \approx -0.20$ points, $t = -3.11$, $p = 0.002$, $n = 1090$ at the $h = 4$ peak; significant at $h = 1-5$; dark-ringed markers). **(b) Share $\rightarrow$ line:** in contrast, the line does not significantly respond to a 1-point change in home bet share ($|\hat{\beta}| \leq 0.010$ points, $p > 0.27$ at every horizon, $n = 1103$). Thus, at least in the present data, the book does not reprice to changes in observed bet shares; the price drives the public’s belief, not the reverse.

Discussion

We argue that Levitt’s influential model of bookmaker profit [Levitt, 2004] assumes that public betting is independent of the realized outcome, and then generalize the model of sportsbook profit to allow betting-outcome dependence. Importantly, this leads to an expression for bookmaker profit that is affine in the public’s bet share on the eventual losing side: the profit-bias identity. That central object of the identity may be understood in the framework of decision theory, such that bookmaker profit is synonymous with the public’s misclassification error of the outcome relative to the book’s line. This places the book’s accounting and the bettor’s estimation problem on a common footing, since the same quantiles of the outcome distribution that govern one govern the other [Dmochowski, 2023]. There is thus a direct link between how much the book profits and how biased the public’s belief of the outcome is – at the posted spread. A profit-maximizing book seeks to identify the price which maximizes the share of betting on the losing side but weighted by the outcomes of the corresponding loss. This is inherently a three-dimensional optimization problem – for each candidate point spread, the profit-optimizing book must forecast (i) the share of bets on the losing side if home wins, (ii) the share of bets on the losing side if visitor wins, and (iii) the probability of the home side covering. Critically, (i) and (ii) do *not* generally sum to one: their sum is $1 + \delta(s)$, and equals one precisely when the public’s money is statistically independent of the outcome. That independence is the scenario treated by Levitt, whose profit expression follows as a special case of the profit-bias identity. Generalizing the model does not, however, overturn his conclusion: that books shade prices to exploit predictable public bias remains consistent with the profit-maximizing view of fixed-odds pricing [Kuypers, 2000] and with evidence that books do not move lines to equalize the money on the two sides [Paul and Weinbach, 2007]. What the present work adds is that the proximate driver of book profit is the public’s misclassification error of the outcome relative to the posted line.

A surprising finding is that the book can profit above its hold even in the absence of statistical dependence between outcomes and betting. The bookmaker profit decomposes into three “channels”: (i) its hold, (ii) a structural profit term that is positive when the public leans towards the same side that the book’s prices are shaded, and (iii) the covariance between betting and outcomes. Perhaps counterintuitively, we show that public lean and book shading may align even if the public has no information pertaining to the realized outcome. As long as the side with the higher bet share has a lower probability of winning than that implied by its price, the book earns a structural profit in addition to the hold. The lean that drives this channel is itself well documented, and is usually attributed to the royalties and sentiment

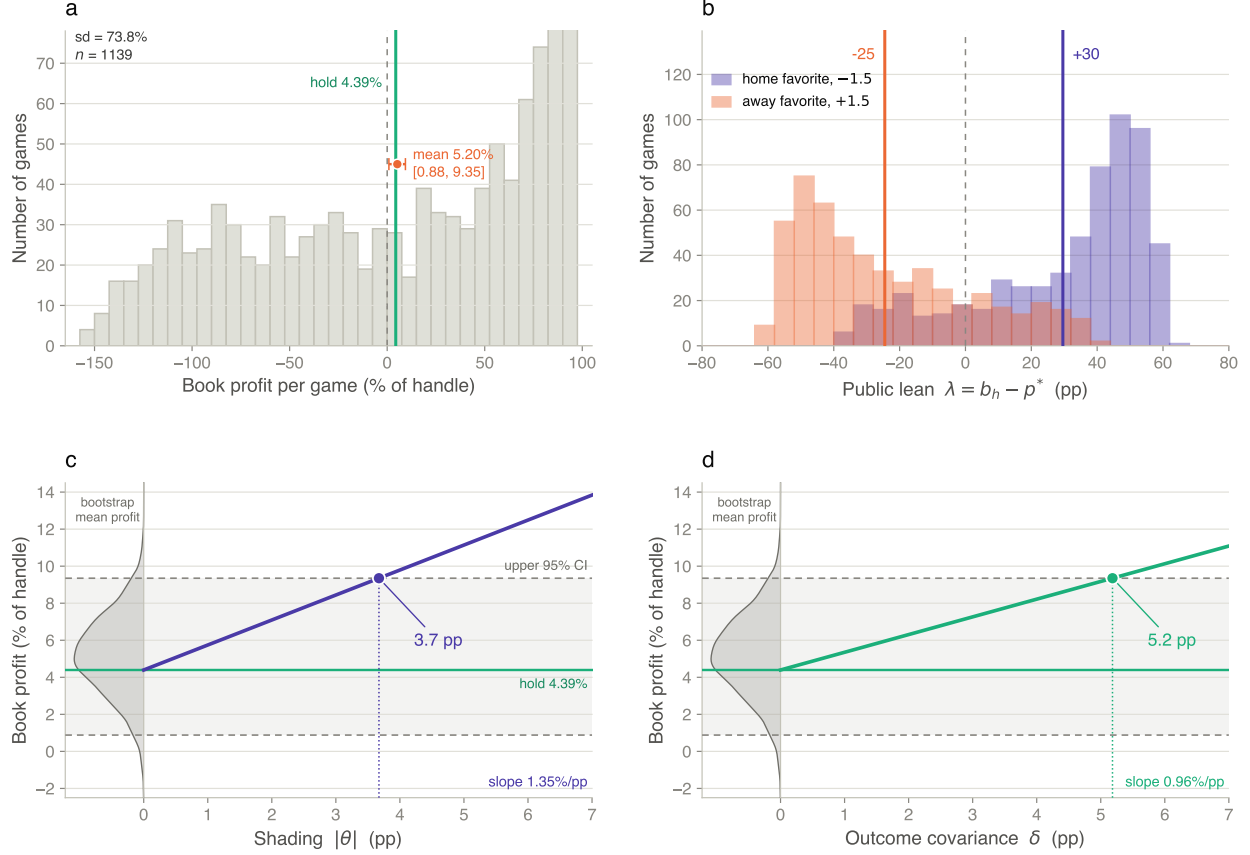


Figure 5: The book’s profit margin is indistinguishable from its hold, despite a large public lean toward favorites. (a) Reconstructed book margin per game, shown as a percentage of that game’s handle ($N = 1,139$). The mean profit margin of 5.20% (95% CI [0.88, 9.35], orange) is statistically indistinguishable from the 4.39% margin that a perfectly balanced book would collect from the hold alone (green), though individual games are widely dispersed (s.d. 73.8%). (b) The public lean $\lambda = b_h - p^*$, split by the identity of the favored side. Betting is highly skewed to the favorite: +29.6 pp where the home side is the favorite, −24.5 pp where the visitor is favored. (c,d) Upper bounds on the two mechanisms that the book may exploit to profit beyond the hold. The grey density is the bootstrap distribution of the mean profit margin, with its 95% CI interval marked; the sloped line denotes the excess profit margin that the book earns as a function of the level of the corresponding mechanism (either price shading or outcome-covariance). The largest values of shading and outcome covariance that are compatible with the empirical sample are thus marked by the lines’ crossing of the upper CI: 3.7 pp of shading aligned with the lean, and 5.2 pp of outcome covariance.

of a recreational public rather than to information [Avery and Chevalier, 1999, Franck et al., 2011, Staněk, 2017]. The mechanism is complementary to the classical account of bookmaker margins, in which both the margin and the favorite–longshot bias arise from the threat of privately informed insiders [Shin, 1991, 1992]; here no informed trader is required, only a marginal taste for favorites meeting a handicap that makes that taste costly. That bias takes the opposite sign in parimutuel racetrack betting, where longshots are overbet [Thaler and Ziemba, 1988, Snowberg and Wolfers, 2010, Ottaviani and Sørensen, 2008]; the favorite preference we observe is its fixed-odds, team-sport counterpart, of a piece with earlier study of the baseball betting market [Woodland and Woodland, 1994]. Note that the conditions for this second channel of book profit are known before the actual outcome is observed: the collected bet shares and likelihood of each side winning the bet. In order for the book to profit from the third channel, it must elicit a preponderance of money on the side that loses the match (presumably by identifying a price at which this occurs systematically). This third channel cannot be resolved until after the game has concluded.

Another surprising aspect of our work pertains to our empirical investigation of the MLB run-line market at a major North American sportsbook, where we discovered that a Simpson’s paradox can spuriously indicate betting–outcome dependence. When pooling all matches, we observed a significant majority of the public’s money on the eventual losing side (Fig. 2b). However, this effect is completely eliminated once stratifying the data by the identity of the favorite: in

games with a home underdog, there was no significant difference between the bet share on the losing versus winning side (Fig. 2c). The same finding was observed in matches where the home side was favored (Fig. 2c). Yet when pooling the two strata, a seeming “square” public betting pattern emerges. The resolution to this Simpson’s paradox is that the identity of the favorite drives both public betting tendencies (because the public prefers favorites) *and* the eventual outcome (because the favorites win less often than implied by the book’s prices). The finding of more money on the eventual losing side when aggregating all games is thus a composition artifact. Importantly, this result is a caution to tests of market efficiency and betting-outcome dependence, which are often performed on pooled samples [Paul and Weinbach, 2002, 2011, Humphreys et al., 2013]. The excess-share term $2F_m\bar{F}_m\delta$ of (S15) makes the separation explicit: what a pooled test reads as behavioral dependence is the sum of a within-stratum term and a composition term, and only the former carries the interpretation usually assigned to it. Our findings point to the importance of conditioning such analyses on match attributes.

Our empirical analysis reconstructed book profit over 1,139 games and found that the resulting profit margin does not vary from that of the book’s guaranteed hold (Fig. 5a). This is despite very strong evidence for the public’s bias towards the match favorite (Fig. 5b). From the second channel of book profit (10), a margin no larger than the hold despite so pronounced a lean implies that the book’s price shading must be small: the data bound it to at most 3.7 pp in the direction of the lean, and a direct check of the posted prices shows no detectable departure from calibration. We read this as a null on shading rather than as a demonstration of accurate pricing. Expected profit depends on the mispricing only through its product with the lean, so a margin test carries power only against mispricing that aligns with where the money sits: in simulation on our own design, 5 pp of aligned mispricing is detected 88% of the time, against 6% when the same mispricing is oriented independently of the lean. The direct check is limited in a complementary way, since at this sample size a perfectly calibrated book would itself generate local departures of up to 9.4 pp by chance. What we bound is therefore Levitt’s mechanism specifically; a book could misprice substantially in ways unrelated to the public’s lean and leave no trace in its margin. Because the book can only profit from a public lean towards favorites if the prices are also shaded towards favorites (Cor. 3), the absence of detectable shading nullifies the potential additional profit. Interestingly, this scenario finds the book taking on additional risk – the variability in profit margin is enormously higher than that of the hold itself – without “reward”. The expected profit is no larger than the risk-free hold, but comes with a much wider confidence interval.

One limitation of the present study is that the treatment is limited to two-outcome markets, whereas several prominent markets are three-way (e.g. European football home-draw-away). As such, extending our framework to multi-way markets such that the profit-bias relationship can be quantified represents an avenue for future research. Our empirical results are best understood as demonstrations of the proposed theoretical framework rather than inferential claims about the economics of modern-day sportsbooks. We analyzed data from one sport and one market (albeit both major) and the empirical findings serve to demonstrate that key objects such as $Q(s)$, δ may be estimated from data and that they convey insight into that specific market. The number and size of sports betting markets is vast, and we do not aim to characterize how those objects vary across the spectrum of sports and bet types. Relatedly, public bet-share data are scarce. Here we collected hourly data from a single book over several months, but quantities such as the extent of the Goldilocks Zone require the joint distribution of bet share and outcome over many games and remain hard to estimate. Prediction markets – whose order books are an analogue of the bet share – may generate richer data [Wolfers and Zitzewitz, 2004, Snowberg et al., 2013], but it is unclear how neatly this may map to sportsbook markets. A further caveat is that the book’s margin is reconstructed rather than observed: the splits record shares rather than dollars, so each game enters per unit of its own handle, and the aggregate depends on a volume distribution we do not see. Finally, with the exception of the local-projection analysis (Fig. 4), our treatment is non-causal. Whether pre-game movement of the kind we observe reflects the arrival of informed money is the subject of an extensive line-movement literature [Gandar et al., 1998, Krieger and Fodor, 2013, Simon, 2024, Croxson and Reade, 2014]. On the other hand, betting markets likely exhibit a deep causal structure linking information arrival, book prices, and public betting tendencies; as such, a causal account of how book prices and public beliefs evolve is an important direction for future work.

The profit-bias identity naturally organizes betting markets by the properties of the conditional losing shares. In particular, we define the Goldilocks Zone as the set of prices at which the sportsbook *and* a negligible bettor are simultaneously positive in expectation. This requires that the public’s losing-side share exceeds the book’s break-even threshold while the outcome distribution is sufficiently removed from the median. A Goldilocks Zone represents an inefficiency that still allows the book to profit, and thus connects the present work to the long-standing debate over betting market efficiency [Sauer, 1998, Gray and Gray, 1997, Vandenbruene et al., 2022]. Whether such a zone can be exploited in practice depends on the same constraints that limit arbitrage in financial markets [Shleifer and Vishny, 1997, Moskowitz, 2021]: betting limits, transaction costs, and the difficulty of consistently identifying the profitable side.

The view that sportsbooks profit by exploiting a predictably biased public has organized the literature for two decades. Legalized online betting is now expanding that market and shifting the composition of the betting population. It remains

to be seen whether the prevalent view outlasts this change; the profit-bias identity provides the means to answer that question empirically.

Methods

Construction of schematic figure (Fig. 1)

Figure 1 is an illustrative schematic; as such, its parameters are chosen for clarity rather real-world plausibility. Outcomes follow a Gaussian model where the median outcome (home score - away score) is assumed to be 0 (center of the horizontal axis) and the horizontal axis represents the sportsbook’s posted point spread, which may be viewed as their estimate of the outcome. We define $z(s) = 2\Phi(s/2) - 1$, $z \in (-1, 1)$, where Φ is the standard normal CDF, to capture the strength of the implied favorite. The outcome distribution is then modeled as $F_m(s) = \frac{1}{2} + 0.24 z(s)$, such that favorites (larger $|s|$) win the proposition less than half the time. The sportsbook’s break-even threshold is set to a realistic vig, $\phi = 100/120$ (−120 American odds), yielding $\tau = \phi/(1 + \phi) \approx 0.45$. The four public behaviours differ only in the home bet share $b_h(s)$ and the outcome-covariance $\delta(s)$: **(a) calibrated**, $b_h = \bar{F}_m$ (each side backed at its actual win probability), so that $Q = 2F_m(1 - F_m)$; **(b) favorite-leaning**, $b_h = \frac{1}{2} + 0.32 z$ (favorites backed beyond their actual win probability) with $\delta = 0$, such that $Q = \frac{1}{2} + 2(F_m - \frac{1}{2})(b_h - \frac{1}{2})$; and **(c) square-****(d) sharp**, which add a constant $\delta = \pm 0.28$ such that $Q = \frac{1}{2} + 2(F_m - \frac{1}{2})(b_h - \frac{1}{2}) + 2F_m\bar{F}_m\delta$. We clip δ to the range $\min(b_h/F_m, (1 - b_h)/\bar{F}_m)$ to ensure that the conditional shares $b_{h,L} = b_h + \bar{F}_m\delta$ and $b_{h,W} = b_h - F_m\delta$ remain in $[0, 1]$. The shaded Goldilocks region shows the set of candidate spreads where the book profits ($Q > \tau$) while still allowing a negligible bettor to find a profitable wager ($F_m < \tau$ or $F_m > 1 - \tau$).

Data collection

We obtain public betting data from the DraftKings Network betting-splits website [DraftKings Network, 2026], which reports, for each upcoming event and market, the percentage of handle and tickets wagered on each side. The splits are collected each hour, yielding a series of samples from posting to game start: we record the event, market, selection, posted odds, and the handle and ticket shares. We focus on the two-sided market to which (5) most directly applies: the baseball *run line*. The run line is a variant of conventional point spread but with an important modification: whereas a standard point spread varies the handicap and holds the payouts (ϕ_h, ϕ_v) roughly symmetric (e.g. $\phi_h = \phi_v = 0.91$), the run line fixes the handicap at ± 1.5 and expresses the teams’ relative strength through (generally asymmetric) payouts. As an example, a home favorite with a -1.5 handicap may pay $\phi_h = 1.5$ with the corresponding visiting underdog ($+1.5$) paying $\phi_v = 0.56$. The book thus moves the price (i.e., not the magnitude of the handicap) and it is the combination of the payouts and identity of the handicapped side that conveys the favored side.

We obtain final scores from the public ESPN scoreboard API [ESPN, 2026]. For each event, we determine whether it won, lost, or pushed against the line *in effect at that snapshot*. A run-line wager wins when the sum of the selected team’s score and its ± 1.5 handicap exceeds that of its opponent’s. The dataset analysed here spans games from 2026-05-31 through 2026-08-28

Empirical testing of the bet share-outcome dependence (Fig. 2)

To empirically probe the statistical dependence between public betting and realized outcomes, we analyzed MLB run-line bet shares measured at game onset. Specifically, we calculated the sample estimate $\hat{\delta} = \hat{b}_{h,L} - \hat{b}_{h,W}$ – the difference in home bet share between matches where the home team won versus lost against the spread. The quantity $\hat{\delta}$ was first computed over all games, and then on two individual strata – the set of games with a home favorite, and those with a home underdog. The rationale for stratifying the data (i.e., conditioning on the identity of the favorite) was to control for hidden variables that produce spurious dependencies between bet share and realized outcome.

Figure 2(a) plots the home bet share against the home win probability, which we computed from the sportsbook’s “moneyline” payouts (the moneyline is a bet on which team wins the contest, regardless of margin). The reason for this choice, which affects only the display item, was to allow for sorting the matches along the “strength of favorite” dimension. For home and visiting moneyline payouts of ϕ_h and ϕ_v , respectively, we remove the book’s vig (“devig”) by normalizing the corresponding implied win probabilities $1/(1 + \phi_h)$ and $1/(1 + \phi_v)$ to sum to one:

$$\text{dv}(\phi_h, \phi_v) = \frac{1/(1 + \phi_h)}{1/(1 + \phi_h) + 1/(1 + \phi_v)}. \quad (20)$$

We use the resulting devigged home-win probability $\text{dv}(\phi_h, \phi_v)$ to order games by favorite strength. In the figure, games are grouped into deciles, markers are decile means, and the band denotes bootstrapped 95% CI. Panels (b) and (c)

report $\widehat{b}_{h,L}$ (home lost the bet) and $\widehat{b}_{h,W}$ (home won), pooled and within strata respectively, along with $\widehat{\delta} = \widehat{b}_{h,L} - \widehat{b}_{h,W}$ and its two-sided bootstrapped p -value.

Dynamics of book profit and bet share-outcome dependence (Fig. 3)

To study the temporal dynamics of book profit and associated bet share-outcome dependence, we measured the losing-side share $\widehat{Q}$ and the home bet share $B_{g,t}$ across four, non-overlapping 6-hour windows terminating at game onset. Only games with at least one sample in each window were included. Within each game, we averaged all samples from the same window. We then fit, separately for home favorites and home underdogs, the following mixed model:

$$B_{g,t} = \beta_0 + \beta_h h_{g,t} + \beta_W W_g + \beta_{hW} (h_{g,t} \times W_g) + u_g + \varepsilon_{g,t}, \quad u_g \sim \mathcal{N}(0, \sigma_u^2), \quad (21)$$

where $B_{g,t}$ is the observed home bet share at sample t of game g , $h_{g,t}$ is the number of hours before onset of game g , $W_g \in \{0, 1\}$ is a binary indicator of whether the home side won the closing bet, and u_g is a per-game random intercept. Since $\delta(h) = b_{h,L}(h) - b_{h,W}(h) = -\beta_W - \beta_{hW} h$, statistical significance of the closing share is assessed with the regression coefficient β_W (at $h = 0$). The *movement* of the dependence is assessed by the outcome $\times$ time interaction β_{hW} . The reported p -values were obtained with a game-level cluster bootstrap (resampling whole games with replacement and refitting the model). Figure 3 displays, for each stratum, the losing-side share $\widehat{Q}(h)$ and the conditional bet shares $\widehat{b}_{h,L}(h)$ and $\widehat{b}_{h,W}(h)$. The shares are first averaged within each game (across samples in a 6-hour bin) and then across games. Shading denotes across-game SEM.

Line-share feedback (Fig. 4)

We investigate the direction of the line-share feedback: whether the book's price drives the public's share, the reverse, or both. Since the run-line handicap is fixed to ± 1.5 , line movement is reflected in the payouts. We represent the momentary line by $\ell_{g,t} = \text{dv}(\phi_{h,g,t}, \phi_{v,g,t})$, obtained by applying the devig transform (20) to the *run-line* payouts at sample t . In other words, we operationalize "line" to the devigged probability that the home side covers ± 1.5 . We then fit Jordà local projections in both directions, for horizons $h = 0, \dots, 12$ hours:

$$\begin{aligned} B_{g,t+h} - B_{g,t-1} &= \beta_h^{\ell \rightarrow B} \Delta \ell_{g,t} + \gamma_h \Delta B_{g,t-1} + a_g + \varepsilon_{g,t}, \\ \ell_{g,t+h} - \ell_{g,t-1} &= \beta_h^{B \rightarrow \ell} \Delta B_{g,t} + \gamma'_h \Delta \ell_{g,t-1} + a'_g + \varepsilon'_{g,t}, \end{aligned} \quad (22)$$

where $\Delta \ell_{g,t} = \ell_{g,t} - \ell_{g,t-1}$ and $\Delta B_{g,t} = B_{g,t} - B_{g,t-1}$ are the one-step line and share deviations, the outcome is measured from the pre-move level $t - 1$, the game fixed effect a_g absorbs the latent state, and the lagged one-step move of the responding series (i.e., $\Delta B_{g,t-1}$ and $\Delta \ell_{g,t-1}$, which we term the *pre-move momentum*) prevents the book from "chasing" the money already wagered. Standard errors are clustered by game. Under sequential ignorability (i.e., the current move is as-good-as-random given the game state and the money so far), β_h is the impulse response of one series to a move in the other. We denoise the integer-rounded bet shares with a 3-hour moving average.

The book's profit margin is indistinguishable from its hold. (Fig. 5). For each game (indexed here by i), we reconstruct the book's profit margin per unit of that game's handle from three observables: the closing home bet share $b_{h,i}$, the home and away payouts $\phi_{h,i}, \phi_{v,i}$, and the realized outcome $C_i \in \{0, 1\}$ representing whether the home side covered:

$$\pi_i = C_i [(1 - b_{h,i}) - \phi_{h,i} b_{h,i}] + (1 - C_i) [b_{h,i} - \phi_{v,i} (1 - b_{h,i})]. \quad (23)$$

Here π_i is a profit margin per unit of betting volume, which we have assumed to be equivalent for all games in the dataset. Denoting the raw implied win probability with $a = 1/(1 + \phi)$, the corresponding *devigged* win probability is calculated as $p_i^* = a_{h,i}/(a_{h,i} + a_{v,i})$. We then tabulate $D_i = 2 + \phi_{h,i} + \phi_{v,i}$, the per-game hold $h_i = (1 - \phi_{h,i} \phi_{v,i})/D_i$, and the public lean $\lambda_i = b_{h,i} - p_i^*$. Evaluating Prop. S1 at the realized outcome (rather than the cover probability) yields the following expression for per-game profit:

$$\pi_i = h_i + D_i \lambda_i (p_i^* - C_i). \quad (24)$$

Confidence intervals for the mean across-game profit margin $\bar{\pi} = \frac{1}{N} \sum_i \pi_i$ are percentile intervals from a nonparametric bootstrap over games (4,000 resamples).

To upper bound the level of price shading or outcome-covariance that is consistent with the observed data, we calculated the largest levels that yield profit margins within the observed confidence intervals. For price shading that is maximally aligned with the public lean, $\theta = c \text{sgn}(\lambda)$, where $c > 0$ is the magnitude of price shading in probability units. Assuming $\delta = 0$, the expected profit margin follows as $h + c \mathbb{E}[D|\lambda]$. For a constant outcome covariance δ with $\theta = 0$ and $\bar{F}_m = p^*$, the profit margin is $h + \delta \mathbb{E}[D p^* (1 - p^*)]$. In both cases, the profit margins are linear in the level and have a

slope fixed by the observed lean and price distribution; the largest compatible value is thus given by where line crosses the upper limit of the bootstrap interval.

To assess overall calibration of book prices, we pool games and calculate the mean deviation between devigged price and realized cover rate, employing a binomial standard error. To assess calibration across the price range, we employed Gaussian kernel regression: on a grid of 200 points spanning the observed range of p^* , we estimate $\Pr(\text{cover} | p^*)$ as a Gaussian-weighted average of the realized cover indicators C_i , each game weighted by its distance in price from that point (Nadaraya–Watson regression, bandwidth 0.05). We then form the miscalibration curve as $\hat{\theta}(p^*) = p^* - \widehat{\Pr}(\text{cover} | p^*)$, which we summarize with its largest absolute value $\sup |\hat{\theta}|$.

To generate a null distribution for $\sup |\hat{\theta}|$, we hold the observed prices fixed and draw counterfactual outcomes according to the book’s implied probabilities: $C_i \sim \text{Bernoulli}(p_i^*)$. We then recompute the supremum over 4,000 draws – the p -value is the fraction of simulated suprema that reach the observed one.

References

- Christopher Avery and Judith Chevalier. Identifying investor sentiment from price paths: The case of football betting. *The Journal of Business*, 72(4):493–521, 1999. doi: 10.1086/209625.
- Karen Croxson and J. James Reade. Information and efficiency: Goal arrival in soccer betting. *The Economic Journal*, 124(575):62–91, 2014. doi: 10.1111/eoj.12033.
- Jacek P Dmochowski. A statistical theory of optimal decision-making in sports betting. *Plos one*, 18(6):e0287601, 2023.
- DraftKings Network. Betting splits. <https://dknetwork.draftkings.com/draftkings-sportsbook-betting-splits/>, 2026. Public handle and ticket percentages by market; accessed 2026.
- ESPN. Scoreboard api. <https://site.api.espn.com/apis/site/v2/sports/>, 2026. Final scores and game status; accessed 2026.
- Egon P. Franck, Erwin Verbeek, and Stephan Nüesch. Sentimental preferences and the organizational regime of betting markets. *Southern Economic Journal*, 78(2):502–518, 2011. doi: 10.4284/0038-4038-78.2.502.
- John M. Gandar, William H. Dare, Craig R. Brown, and Richard A. Zuber. Informed traders and price variations in the betting market for professional basketball games. *The Journal of Finance*, 53(1):385–401, 1998. doi: 10.1111/0022-1082.155346.
- John M. Gandar, Richard A. Zuber, R. Stafford Johnson, and William Dare. Re-examining the betting market on major league baseball games: Is there a reverse favourite-longshot bias? *Applied Economics*, 34(10):1309–1317, 2002. doi: 10.1080/00036840110095427.
- Philip K. Gray and Stephen F. Gray. Testing market efficiency: Evidence from the NFL sports betting market. *The Journal of Finance*, 52(4):1725–1737, 1997. doi: 10.1111/j.1540-6261.1997.tb01129.x.
- Ondřej Hubáček, Gustav Šourek, and Filip Železný. Exploiting sports-betting market using machine learning. *International Journal of Forecasting*, 35(2):783–796, 2019. doi: 10.1016/j.ijforecast.2019.01.001.
- Brad R. Humphreys. Point spread shading and behavioral biases in NBA betting markets. *Rivista di Diritto ed Economia dello Sport*, 6(1):13–26, 2010.
- Brad R. Humphreys. The financial consequences of unbalanced betting on NFL games. *International Journal of Sport Finance*, 6(1):60–71, 2011.
- Brad R. Humphreys, Rodney J. Paul, and Andrew P. Weinbach. Bettor biases and the “home-underdog” bias in the NFL. *International Journal of Sport Finance*, 8(4):294–311, 2013.
- Òscar Jordà. Estimation and inference of impulse responses by local projections. *American Economic Review*, 95(1):161–182, 2005.
- Bruno Jullien and Bernard Salanié. Estimating preferences under risk: The case of racetrack bettors. *Journal of Political Economy*, 108(3):503–530, 2000. doi: 10.1086/262127.
- Kevin Krieger and Andy Fodor. Price movements and the prevalence of informed traders: The case of line movement in college basketball. *Journal of Economics and Business*, 68:70–82, 2013. doi: 10.1016/j.jeconbus.2013.04.001.
- Tim Kuypers. Information and efficiency: An empirical study of a fixed odds betting market. *Applied Economics*, 32(11):1353–1363, 2000. doi: 10.1080/00036840050151449.

- Steven D. Levitt. Why are gambling markets organised so differently from financial markets? *The Economic Journal*, 114(495):223–246, 2004. doi: 10.1111/j.1468-0297.2004.00207.x. URL <https://pricetheory.uchicago.edu/levitt/Papers/LevittWhyAreGamblingMarkets2004.pdf>.
- Tobias J. Moskowitz. Asset pricing and sports betting. *The Journal of Finance*, 76(6):3153–3209, 2021. doi: 10.1111/jofi.13082.
- Marco Ottaviani and Peter Norman Sørensen. The favorite-longshot bias: An overview of the main explanations. In Donald B. Hausch and William T. Ziemba, editors, *Handbook of Sports and Lottery Markets*, pages 83–101. North-Holland, 2008.
- Rodney J. Paul and Andrew P. Weinbach. Market efficiency and a profitable betting rule: Evidence from totals on professional football. *Journal of Sports Economics*, 3(3):256–263, 2002. doi: 10.1177/1527002502003003003.
- Rodney J. Paul and Andrew P. Weinbach. Does sportsbook.com set pointspreads to maximize profits? tests of the levitt model of sportsbook behavior. *Journal of Prediction Markets*, 1(3):209–218, 2007. URL <https://www.ubplj.org/index.php/jpm/article/download/429/461>.
- Rodney J. Paul and Andrew P. Weinbach. Price setting in the nba gambling market: Tests of the levitt model of sportsbook behavior. *International Journal of Sport Finance*, 3(3):137–145, 2008. URL <https://fitpublishing.com/content/price-setting-nba-gambling-market-tests-levitt-model-sportsbook-behavior-pp-137-145>.
- Rodney J. Paul and Andrew P. Weinbach. Sportsbook behavior in the NCAA football betting market: Tests of the traditional and Levitt models of sportsbook behavior. *The Journal of Prediction Markets*, 3(2):21–37, 2009.
- Rodney J. Paul and Andrew P. Weinbach. NFL bettor biases and price setting: Further tests of the Levitt hypothesis of sportsbook behaviour. *Applied Economics Letters*, 18(2):193–197, 2011. doi: 10.1080/13504850903508242.
- Rodney J. Paul and Andrew P. Weinbach. Sportsbook pricing and the behavioral biases of bettors in the NHL. *Journal of Economics and Finance*, 36(1):123–135, 2012. doi: 10.1007/s12197-009-9112-4.
- Raymond D. Sauer. The economics of wagering markets. *Journal of Economic Literature*, 36(4):2021–2064, 1998.
- Hyun Song Shin. Optimal betting odds against insider traders. *The Economic Journal*, 101(408):1179–1185, 1991. doi: 10.2307/2234434.
- Hyun Song Shin. Prices of state contingent claims with insider traders, and the favourite-longshot bias. *The Economic Journal*, 102(411):426–435, 1992. doi: 10.2307/2234526.
- Hyun Song Shin. Measuring the incidence of insider trading in a market for state-contingent claims. *The Economic Journal*, 103(420):1141–1153, 1993. doi: 10.2307/2234240.
- Andrei Shleifer and Robert W. Vishny. The limits of arbitrage. *The Journal of Finance*, 52(1):35–55, 1997. doi: 10.1111/j.1540-6261.1997.tb03807.x.
- Jay Simon. Inefficient forecasts at the sportsbook: An analysis of real-time betting line movement. *Management Science*, 70(12):8583–8611, 2024. doi: 10.1287/mnsc.2022.00456.
- Erik Snowberg and Justin Wolfers. Explaining the favorite-long shot bias: Is it risk-love or misperceptions? *Journal of Political Economy*, 118(4):723–746, 2010. doi: 10.1086/655844.
- Erik Snowberg, Justin Wolfers, and Eric Zitzewitz. Prediction markets for economic forecasting. In Graham Elliott and Allan Timmermann, editors, *Handbook of Economic Forecasting*, volume 2, pages 657–687. Elsevier, 2013.
- Rostislav Staněk. Home bias in sport betting: Evidence from Czech betting market. *Judgment and Decision Making*, 12(2):168–172, 2017.
- Richard H. Thaler and William T. Ziemba. Anomalies: Parimutuel betting markets: Racetracks and lotteries. *Journal of Economic Perspectives*, 2(2):161–174, 1988. doi: 10.1257/jep.2.2.161.
- Jonas Vandenbruaene, Marc De Ceuster, and Jan Annaert. Efficient spread betting markets: A literature review. *Journal of Sports Economics*, 23(7):907–949, 2022. doi: 10.1177/15270025211071042.
- Karl Whelan. Risk aversion and favourite–longshot bias in a competitive fixed-odds betting market. *Economica*, 91(361):188–209, 2024. doi: 10.1111/ecca.12500.
- Karl Whelan. On estimates of insider trading in sports betting. *The Manchester School*, 2025. doi: 10.1111/manc.12505.
- Justin Wolfers and Eric Zitzewitz. Prediction markets. *Journal of Economic Perspectives*, 18(2):107–126, 2004. doi: 10.1257/0895330041371321.
- Linda M. Woodland and Bill M. Woodland. Market efficiency and the favorite-longshot bias: The baseball betting market. *The Journal of Finance*, 49(1):269–279, 1994. doi: 10.1111/j.1540-6261.1994.tb04429.x.

Supplementary Information

Notation

Table 2 defines all symbols used in the paper, where we indicate the section in which each variable first appears.

Table 2: **Symbols used in the paper.** All bet shares are fractions of the total handle (i.e., normalized to 1).

Symbol	Meaning	Defined
<i>Outcome and line</i>		
M	Realized outcome: margin of victory, total, or player statistic	Setup
s	Sportsbook’s posted line (e.g. point spread)	Setup
$F_m(s)$	$\Pr(M < s)$, the outcome CDF at the line	(1)
$\bar{F}_m(s)$	$\Pr(M > s)$; the home side wins the proposition	(1)
$\pi_{\text{push}}(s)$	$\Pr(M = s)$, the probability of a “push” where all bets are cancelled	(1)
<i>Bet shares</i>		
$B_h(s), B_v(s)$	Public’s random bet share on the home / visitor side, $B_v = 1 - B_h$	Setup
$b_h(s), b_v(s)$	Unconditional expected bet shares, $b_h = \mathbb{E}[B_h]$	(7)
$b_{h,L}(s)$	$\mathbb{E}[B_h \mid M < s]$: home bet share when the home side loses	(2)
$b_{v,L}(s)$	$\mathbb{E}[B_v \mid M > s]$: visitor bet share when the visitor loses	(2)
$b_{h,W}(s)$	$\mathbb{E}[B_h \mid M > s]$: home bet share when the home side wins	(7)
$b_{\min}, b_{\max}$	min and max of $b_{h,L}$ and $b_{v,L}$	Prop. 5
<i>Prices and the vig</i>		
ϕ	Net payout per unit wager on a winner (symmetric vig)	(13)
ϕ_h, ϕ_v	Net payouts on the home / visitor side (asymmetric vig)	S1
τ	$\phi/(1 + \phi)$, the book’s break-even losing-side share	(13)
p^*	Devigged (overround-removed) price of the home side	Prop. S1
O	Overround: the two raw implied probabilities summed	Prop. S1
h	Hold: share of handle that a balanced book keeps, $h = 1 - 1/O$	Prop. S1
D	$2 + \phi_h + \phi_v$	Prop. S1
<i>Profit and its drivers</i>		
$\pi(s)$	Book profit per unit of handle	(5)
$Q(s)$	Expected share of handle on the losing side	(6)
$Q_{\text{indep}}(s)$	Value of Q under Levitt independence	(S12)
$\delta(s)$	$b_{h,L} - b_{h,W}$: outcome covariance of the bet share	(7)
$\theta(s)$	Shading: deviation of the posted price from the truth	(8)
$\lambda(s)$	Lean: deviation of the public’s bet share from balance	(8)
$\mathcal{G}$	The Goldilocks Zone	(17)
<i>Public-belief model</i>		
X	Latent expected outcome	(12)
Y	Public’s noisy, biased belief about X	(12)
ε	Systematic bias in the public’s belief	(12)
V, U	Belief noise and game randomness	(12)
<i>Estimation</i>		
C_i	Realized outcome: 1 if the home side won covered in game i	(23)
π_i	Book’s profit margin for game i	(23)

S1. Asymmetric vigs, pushes, and the general profit–bias identity

We allow for asymmetric vigs: if a unit stake is placed on the winning home side, the bettor receives a gross return of $1 + \phi_h$, and if a unit stake is placed on the winning visitor side, the bettor receives a gross return of $1 + \phi_v$, where $\phi_h, \phi_v > 0$ represent the bettor’s profit on a winning unit bet and satisfy $\phi_h \phi_v \leq 1$. The latter indicates that the book’s overround $O \equiv (1 + \phi_h)^{-1} + (1 + \phi_v)^{-1}$ is at least one, and is exactly one only for a fair book. If the outcome is a push ($M = s$), all bets are returned. With total stake normalized so that $B_h(s) + B_v(s) = 1$, the sportsbook profit for a single game is:

$$\pi(M; s) = \begin{cases} B_v(s) - \phi_h B_h(s), & M > s, \\ B_h(s) - \phi_v B_v(s), & M < s, \\ 0, & M = s. \end{cases} \quad (\text{S1})$$

Expected profit is computed by conditioning on the three outcomes $\{M < s\}$, $\{M > s\}$, $\{M = s\}$. If $M < s$ (home loses), then using (S1) and the definition of $b_{h,L}(s)$:

$$\mathbb{E}[\pi(M; s) \mid M < s] = \mathbb{E}[B_h(s) - \phi_v(1 - B_h(s)) \mid M < s] = (1 + \phi_v) b_{h,L}(s) - \phi_v.$$

Similarly, if $M > s$, $\mathbb{E}[\pi(M; s) \mid M > s] = (1 + \phi_h) b_{v,L}(s) - \phi_h$. If $M = s$, $\pi = 0$. Taking expectations over the three cases yields:

$$\mathbb{E}[\pi(M; s)] = F_m(s) [(1 + \phi_v) b_{h,L}(s) - \phi_v] + \bar{F}_m(s) [(1 + \phi_h) b_{v,L}(s) - \phi_h]. \quad (\text{S2})$$

Define $Q_h(s) \equiv F_m(s) b_{h,L}(s)$, $Q_v(s) \equiv \bar{F}_m(s) b_{v,L}(s)$, $Q(s) \equiv Q_h(s) + Q_v(s)$, and $\Delta\phi \equiv \phi_v - \phi_h$. Substituting into (S2) gives:

$$\mathbb{E}[\pi(M; s)] = (1 + \phi_h) Q(s) + \Delta\phi Q_h(s) - \phi_v F_m(s) - \phi_h \bar{F}_m(s).$$

Using $F_m(s) + \pi_{\text{push}}(s) + \bar{F}_m(s) = 1$ and $\phi_v F_m(s) + \phi_h \bar{F}_m(s) = \phi_h(1 - \pi_{\text{push}}(s)) + \Delta\phi F_m(s)$, we obtain the decomposition:

$$\mathbb{E}[\pi(M; s)] = (1 + \phi_h) Q(s) + \Delta\phi (Q_h(s) - F_m(s)) - \phi_h (1 - \pi_{\text{push}}(s)). \quad (\text{S3})$$

Special cases. If $\phi_h = \phi_v = \phi$ then $\Delta\phi = 0$ and (S3) simplifies to:

$$\mathbb{E}[\pi(M; s)] = (1 + \phi) Q(s) - \phi (1 - \pi_{\text{push}}(s)), \quad (\text{S4})$$

which implies the uniform bound $\mathbb{E}[\pi(M; s)] \leq (1 + \phi) Q(s)$. Setting $\pi_{\text{push}}(s) = 0$ in (S4) recovers the profit-bias identity (5) of the main text.

Anatomy of the book's margin with asymmetric vigs. The three-term decomposition of the main text (Prop. 1) can be generalized to asymmetric vigs by replacing the balanced-book reference $\frac{1}{2}$ with the devigged price, and the symmetric hold $(1 - \phi)/2$ by the overround-implied hold.

Proposition S1 (Anatomy of the book's margin, asymmetric vigs). *Assume $\pi_{\text{push}}(s) \equiv 0$ and let $D \equiv 2 + \phi_h + \phi_v$. Define:*

$$p^* \equiv \frac{1 + \phi_v}{D}, \quad h \equiv \frac{1 - \phi_h \phi_v}{D}, \quad \theta(s) \equiv p^* - \bar{F}_m(s), \quad \lambda(s) \equiv b_h(s) - p^*,$$

where p^* is the devigged implied win probability of the home side and h is the book's hold. Then:

$$\mathbb{E}[\pi(s)] = h + D \theta(s) \lambda(s) + D F_m(s) \bar{F}_m(s) \delta(s). \quad (\text{S5})$$

Setting $\phi_h = \phi_v = \phi$ gives $p^* = \frac{1}{2}$, $h = (1 - \phi)/2$ and $D = 2(1 + \phi)$, recovering (10).

Proof. Start from (S2) with $\pi_{\text{push}}(s) \equiv 0$. By the law of total expectation, $b_h(s) = F_m(s) b_{h,L}(s) + \bar{F}_m(s) b_{h,W}(s)$; together with $\delta(s) = b_{h,L}(s) - b_{h,W}(s)$, this yields:

$$b_{h,L}(s) = b_h(s) + \bar{F}_m(s) \delta(s), \quad b_{h,W}(s) = b_h(s) - F_m(s) \delta(s),$$

and since $B_v = 1 - B_h$, $b_{v,L}(s) = 1 - b_{h,W}(s) = (1 - b_h(s)) + F_m(s) \delta(s)$. Substituting both into (S2) and collecting the δ terms, whose coefficient is $F_m \bar{F}_m [(1 + \phi_v) + (1 + \phi_h)] = D F_m \bar{F}_m$, leads to:

$$\mathbb{E}[\pi(s)] = F_m [(1 + \phi_v) b_h - \phi_v] + \bar{F}_m [(1 + \phi_h)(1 - b_h) - \phi_h] + D F_m \bar{F}_m \delta.$$

Expanding the first two terms and factoring b_h yields:

$$\mathbb{E}[\pi(s)] = [F_m(1 + \phi_v) - \bar{F}_m(1 + \phi_h)] b_h + [\bar{F}_m - F_m \phi_v] + D F_m \bar{F}_m \delta.$$

We evaluate the two brackets in turn. Substituting $F_m = 1 - \bar{F}_m$ into the first and using $(1 + \phi_v) + (1 + \phi_h) = D$ and $(1 + \phi_v) = D p^*$:

$$F_m(1 + \phi_v) - \bar{F}_m(1 + \phi_h) = (1 + \phi_v) - \bar{F}_m [(1 + \phi_v) + (1 + \phi_h)] = D p^* - D \bar{F}_m = D \theta.$$

Performing the same substitution in the second bracket gives $\bar{F}_m - F_m \phi_v = \bar{F}_m(1 + \phi_v) - \phi_v$. Writing $\bar{F}_m = p^* - \theta$, we have:

$$\bar{F}_m(1 + \phi_v) - \phi_v = [p^*(1 + \phi_v) - \phi_v] - \theta(1 + \phi_v) = [p^*(1 + \phi_v) - \phi_v] - D \theta p^*.$$

The bracketed constant evaluates to:

$$p^*(1 + \phi_v) - \phi_v = \frac{(1 + \phi_v)^2}{D} - \phi_v = \frac{(1 + \phi_v)^2 - \phi_v D}{D} = \frac{1 - \phi_h \phi_v}{D},$$

because $(1 + \phi_v)^2 - \phi_v(2 + \phi_h + \phi_v) = 1 - \phi_h\phi_v$. Substituting both brackets back:

$$\mathbb{E}[\pi(s)] = D\theta(s)b_h(s) + \frac{1 - \phi_h\phi_v}{D} - D\theta(s)p^* + D F_m(s)\bar{F}_m(s)\delta(s),$$

and collecting the two θ terms through $\lambda(s) = b_h(s) - p^*$:

$$\mathbb{E}[\pi(s)] = \frac{1 - \phi_h\phi_v}{D} + D\theta(s)\lambda(s) + D F_m(s)\bar{F}_m(s)\delta(s).$$

It remains to show that the constant $\frac{1 - \phi_h\phi_v}{D}$ is equivalent to the book's hold h . Putting the two raw implied probabilities in the definition $O \equiv (1 + \phi_h)^{-1} + (1 + \phi_v)^{-1}$ over a common denominator results in the numerator $(1 + \phi_v) + (1 + \phi_h) = D$, such that:

$$O = \frac{D}{(1 + \phi_h)(1 + \phi_v)},$$

and hence:

$$1 - \frac{1}{O} = \frac{D - (1 + \phi_h)(1 + \phi_v)}{D} = \frac{1 - \phi_h\phi_v}{D} = h,$$

where we have used $D - (1 + \phi_h)(1 + \phi_v) = (2 + \phi_h + \phi_v) - (1 + \phi_h + \phi_v + \phi_h\phi_v) = 1 - \phi_h\phi_v$. The constant term is therefore the share of handle that a balanced book keeps (i.e., h), giving (S5). $\square$

S2. Recovering Levitt (2004) under independence

The profit-bias identity (5) is a strict generalization of the bookmaker-profit expression in Levitt [2004]. In Levitt's notation, p is the probability that the favorite wins, f is the fraction of total dollars wagered on the favorite, and v is the vig; his expected gross profit per unit bet is:

$$\mathbb{E}[\pi_{\text{Levitt}}] = [(1 - p)f + p(1 - f)](1 + v) - [(1 - p)(1 - f) + pf]. \quad (\text{S6})$$

The first bracketed term is the expected fraction wagered on the eventual *losing* side, from which the book collects $1 + v$ per unit; the second bracketed term is the expected fraction on the eventual *winning* side, on which it pays out 1 per unit. Levitt's expression is therefore of the form losing share $\times (1 + v)$ – winning share, which follows as:

$$\mathbb{E}[\pi_{\text{Levitt}}] = [(1 - p)f + p(1 - f)](1 + v) - \left(1 - [(1 - p)f + p(1 - f)]\right) = (2 + v)(f + p - 2pf) - 1,$$

which is his Eq. (2). His profit is thus already an affine, increasing function of a single losing-side share — the same structural form as (5).

An important distinction between Levitt's and our expressions pertains to the composition of the losing-side share. In (S6), f is a single *unconditional* number, such that forming $(1 - p)f$ implicitly treats the favorite's bet share as statistically independent of the realized outcome. Identifying the favorite with the home side gives $p = \bar{F}_m(s)$ and $1 - p = F_m(s)$, and imposing that independence gives:

$$b_{h,L}(s) = b_h(s) = f, \quad b_{v,L}(s) = b_v(s) = 1 - f,$$

such that our outcome-conditioned share (6) equates to Levitt's bracket:

$$Q(s) = F_m(s)f + \bar{F}_m(s)(1 - f) = (1 - p)f + p(1 - f) = f + p - 2pf \equiv Q_{\text{indep}}.$$

This shows alignment of bet shares, but not yet of profits. The two profit expressions are denominated in different units and use different vig conventions – we reconcile the different conventions next, and then verify that the identity reproduces Levitt's Eq. (2) exactly. We note also that Levitt's later step of writing $f = f(p)$ with $\partial f / \partial p > 0$ does not affect his independence assumption: it describes how aggregate betting responds to the *posted* line, not how bet shares covary with the *realized* outcome.

Reconciling the two conventions. To allow a direct comparison between (5) and (S6), one must first reconcile two bookkeeping conventions relating to (i) the vig and (ii) the unit of measurement of profit.

(i) *The vig.* Levitt's v is defined relative to the *win* amount: a losing bettor pays $1 + v$ and a winning bettor collects 1 (at the standard –110 price, 110 is risked and 100 is earned, so $v = 1/10$). Our ϕ is the net payout per unit *wagered*: a bettor risks 1 to earn ϕ . The relationship between Levitt's v and our ϕ is thus given by:

$$\phi = \frac{1}{1 + v}. \quad (\text{S7})$$

For example, -110 in American odds corresponds to $\phi = 10/11$ and $v = 1/10$.

(ii) *Normalization of profit.* Our $\pi(s)$ is the profit per unit of *handle*: total stake is normalized to $B_h(s) + B_v(s) = 1$. Levitt’s “gross profit per unit bet” is denominated in the win amount, such that a single unit bet in his accounting puts $1 + v$ of handle at risk. One Levitt unit therefore carries $1 + v = 1/\phi$ units of handle, and the two profit measures are related by:

$$\mathbb{E}[\pi_{\text{Levitt}}] = (1 + v) \mathbb{E}[\pi(s)] = \frac{\mathbb{E}[\pi(s)]}{\phi}. \quad (\text{S8})$$

Exact recovery of Levitt’s Eq. (2). Applying (S7) and (S8) to the identity (5) yields:

$$\frac{\mathbb{E}[\pi(s)]}{\phi} = \frac{(1 + \phi) Q(s) - \phi}{\phi} = \left(1 + \frac{1}{\phi}\right) Q(s) - 1 = (2 + v) Q(s) - 1. \quad (\text{S9})$$

If we assume Levitt’s independence and designate the favorite as the home side, (S9) follows as:

$$\mathbb{E}[\pi_{\text{Levitt}}] = (2 + v) (f + p - 2pf) - 1,$$

which is Levitt’s Eq. (2) exactly.

As will be derived in the next section, the difference between Levitt’s $f + p - 2pf$ and our $Q(s)$ is precisely the covariance between bet shares and realized outcomes. Levitt’s expression thus understates (or overstates) the book’s margin whenever the public’s money tends toward the eventual loser (winner).

S3. Derivation of the independence test

Here we derive the statistical test of bet share–outcome independence (7), which implies equality between conditional and unconditional bet shares:

$$b_{h,L}(s) = b_h(s), \quad b_{v,L}(s) = b_v(s). \quad (\text{S10})$$

We show below that these two conditions collapse into a single equality, and thus bet share–outcome independence may be assayed with a single test.

Consider the home condition first. Excluding pushes, the law of total expectation gives:

$$b_h(s) = F_m(s) b_{h,L}(s) + \bar{F}_m(s) b_{h,W}(s). \quad (\text{S11})$$

The unconditional share is therefore a weighted average of two conditional shares, with the weights given by the probabilities of the two outcomes. Enforcing $b_{h,L}(s) = b_h(s)$ leads to:

$$\bar{F}_m(s) [b_{h,L}(s) - b_{h,W}(s)] = 0.$$

Assuming that $\bar{F}_m(s) > 0$, this implies:

$$b_{h,L}(s) = b_{h,W}(s).$$

Note that the visitor loses the proposition precisely when the home side wins, such that:

$$b_{v,L}(s) = 1 - b_{h,W}(s).$$

Combined with $b_v(s) = 1 - b_h(s)$, the requirement $b_{v,L}(s) = b_v(s)$ is identical to $b_{h,W}(s) = b_h(s)$. In turn, $b_{h,W}(s) = b_h(s)$ implies:

$$F_m(s) [b_{h,L}(s) - b_{h,W}(s)] = 0.$$

Both conditions in (S10) thus reduce to:

$$\delta(s) = b_{h,L}(s) - b_{h,W}(s) = 0.$$

which is (7).

Excess profit from bet-share outcome dependence Here we derive an expression for the change in the book profit that is brought about by a non-zero δ . Substituting (S10) into (6) yields the profit driver under Levitt’s independence:

$$Q_{\text{indep}}(s) \equiv F_m(s) b_h(s) + \bar{F}_m(s) b_v(s). \quad (\text{S12})$$

To relate $Q_{\text{indep}}(s)$ to the realized losing-side share $Q(s)$, it is helpful to write both in terms of home bet shares:

$$Q(s) = F_m(s) b_{h,L}(s) + \bar{F}_m(s) (1 - b_{h,W}(s)), \quad Q_{\text{indep}}(s) = F_m(s) b_h(s) + \bar{F}_m(s) (1 - b_h(s)).$$

Subtracting Q_{indep} from $Q(s)$, one obtains:

$$Q(s) - Q_{\text{indep}}(s) = F_m(s)[b_{h,L}(s) - b_h(s)] + \bar{F}_m(s)[b_h(s) - b_{h,W}(s)]. \quad (\text{S13})$$

Employing (S11), the two bracketed expressions above may be written as:

$$b_{h,L}(s) - b_h(s) = \bar{F}_m(s) \delta(s), \quad b_h(s) - b_{h,W}(s) = F_m(s) \delta(s). \quad (\text{S14})$$

Substituting (S14) back into (S13) leads to:

$$Q(s) - Q_{\text{indep}}(s) = 2 F_m(s) \bar{F}_m(s) \delta(s).$$

To connect the excess profit to the covariance between bet shares and realized outcomes, let $I = \mathbf{1}\{M < s\}$ represent the indicator that the home side fails to cover, such that $\mathbb{E}[I] = F_m(s)$. Note that one can write $\mathbb{E}[B_h I] = F_m(s) b_{h,L}(s)$. The covariance between B_h and I is then given by:

$$\text{Cov}(B_h(s), I) = \mathbb{E}[B_h I] - \mathbb{E}[B_h] \mathbb{E}[I] = F_m(s)[b_{h,L}(s) - b_h(s)] = F_m(s) \bar{F}_m(s) \delta(s),$$

where in the last step we have used (S14). Putting all of this together, the excess profit can be written in two equivalent forms:

$$Q(s) - Q_{\text{indep}}(s) = 2 F_m(s) \bar{F}_m(s) \delta(s) = 2 \text{Cov}(B_h(s), \mathbf{1}\{M < s\}). \quad (\text{S15})$$

The significance of (S15) is that the deviation of our profit driver Q from Levitt's Q_{indep} is twice the covariance between bet shares and realized outcomes.

S4. Public-belief model: Q as a misclassification probability

Here we provide the proof of Proposition 4 in the no-push case where $\Pr(M = s) = 0$. We model the public's belief about the latent expected margin X with:

$$Y = X + \varepsilon + V, \quad M = X + U, \quad (\text{S16})$$

where ε is a systematic bias and U (game randomness) is independent of V (belief noise). Let F_U denote the CDF of U and let F_V denote the CDF of V .

Let the public bet the home side whenever $Y > s$. Define the Bernoulli indicator:

$$B_h(s) \equiv \mathbf{1}\{Y > s\}, \quad B_v(s) \equiv 1 - B_h(s) = \mathbf{1}\{Y \leq s\}.$$

This constrains the public's bet share to the two extrema $\{0, 1\}$, such that the public acts on the signal Y and stakes one side exclusively. This "specialization" converts the losing-side shares into misclassification probabilities but is not restrictive for $Q(s)$: with a continuum of bettors holding independent signals $Y_i = X + \varepsilon + V_i$, the home bet share $B_h(s) = \Pr(Y_i > s | X) = 1 - F_V(s - X - \varepsilon) \in (0, 1)$ yields the same $Q(s)$ derived below, since integrating the fraction over X reproduces (S20) term by term.

Recall the definition of $b_{h,L}(s)$ and $b_{v,L}(s)$ as conditional expected bet shares on the losing side:

$$b_{h,L}(s) \equiv \mathbb{E}[B_h(s) | M < s], \quad b_{v,L}(s) \equiv \mathbb{E}[B_v(s) | M > s]. \quad (\text{S17})$$

Because $B_h(s)$ and $B_v(s)$ are Bernoulli, their conditional expectations may be written as conditional probabilities:

$$b_{h,L}(s) = \Pr(Y > s | M < s), \quad b_{v,L}(s) = \Pr(Y \leq s | M > s). \quad (\text{S18})$$

Substituting (S18) into the profit-bias identity (6) yields:

$$Q(s) = \Pr(Y > s, M < s) + \Pr(Y \leq s, M > s). \quad (\text{S19})$$

Proof of Proposition 4. The claim follows immediately from (S19): the two events $\{Y > s, M < s\}$ (public backs home, home loses) and $\{Y \leq s, M > s\}$ (public backs visitor, visitor loses) are exactly the two ways that the public's signal Y misorders the realized margin of victory M relative to the line s . The union of these disjoint events is the event that the public backs the losing side, such that $Q(s) = \Pr(\text{public backs the losing side})$, which is the probability of misclassifying the target $M > s$ using the predictor $\mathbf{1}\{Y > s\}$. $\square$

Evaluating $Q(s)$ under the public-belief model. From (S16), the dependence between Y and M stems from the shared latent variable X . Even if $U \perp V$, the events $\{Y > s\}$ and $\{M < s\}$ are not generally independent; however, they are independent conditional on X because Y depends on (X, V) and M depends on (X, U) with $U \perp V$. We can therefore write:

$$\Pr(Y > s, M < s) = \mathbb{E}_X[\Pr(Y > s | X) \Pr(M < s | X)],$$

and similarly for $\Pr(Y \leq s, M > s)$. Conditioning on X gives:

$$\Pr(Y > s | X) = \Pr(V > s - X - \varepsilon) = 1 - F_V(s - X - \varepsilon),$$

and:

$$\Pr(M < s | X) = \Pr(U < s - X) = F_U(s - X).$$

Therefore, combining with (S19):

$$Q(s) = \mathbb{E}_X[(1 - F_V(s - X - \varepsilon)) F_U(s - X)] + \mathbb{E}_X[F_V(s - X - \varepsilon) (1 - F_U(s - X))]. \quad (\text{S20})$$

Equation (S20) is the expression for $Q(s)$ under the public-belief model.

To connect this to the losing-side bet shares, we write (S18) as:

$$b_{h,L}(s) = \Pr(Y > s | M < s) = \frac{\Pr(Y > s, M < s)}{\Pr(M < s)} = \frac{\mathbb{E}_X[(1 - F_V(s - X - \varepsilon)) F_U(s - X)]}{F_m(s)}, \quad (\text{S21})$$

and

$$b_{v,L}(s) = \Pr(Y \leq s | M > s) = \frac{\Pr(Y \leq s, M > s)}{\Pr(M > s)} = \frac{\mathbb{E}_X[F_V(s - X - \varepsilon) (1 - F_U(s - X))]}{1 - F_m(s)}. \quad (\text{S22})$$

Substituting (S21) and (S22) into $Q(s) = b_{h,L}(s)F_m(s) + b_{v,L}(s)(1 - F_m(s))$, we recover (S20) identically.

S5. The Goldilocks book-profit region

We define $\tau \equiv \phi/(1 + \phi)$. The book is profitable if (14) is $Q(s) > \tau$, and $Q(s)$ is affine in $F_m(s)$:

$$Q(s) = F_m(s) b_{h,L}(s) + \bar{F}_m(s) b_{v,L}(s) = b_{v,L}(s) + F_m(s)(b_{h,L}(s) - b_{v,L}(s)), \quad (\text{S23})$$

such that (14) is the linear inequality:

$$F_m(s)(b_{h,L}(s) - b_{v,L}(s)) > \tau - b_{v,L}(s). \quad (\text{S24})$$

Below we prove Proposition 5.

Proof of Proposition 5. Write $a = b_{h,L}(s)$, $b = b_{v,L}(s)$, $F = F_m(s) \in [0, 1]$, so that $Q = (1 - F)b + Fa$ is a convex combination of a and b ; consequently, $\min(a, b) \leq Q \leq \max(a, b)$, which is (16).

(i) *Always.* If $b_{\min}(s) = \min(a, b) > \tau$, then $Q \geq \min(a, b) > \tau$ for every $F \in [0, 1]$.

(ii) *Never.* If $b_{\max}(s) = \max(a, b) \leq \tau$, then $Q \leq \max(a, b) \leq \tau$ for every $F \in [0, 1]$, so (14) fails.

(iii) *Threshold.* Suppose $\min(a, b) \leq \tau < \max(a, b)$, such that $a \neq b$ and $F^* \equiv (\tau - b)/(a - b)$ is well defined. If $a > b$, then $b = \min(a, b) \leq \tau < a = \max(a, b)$, so that $0 \leq \tau - b < a - b$; dividing by $a - b > 0$ gives $F^* \in [0, 1]$ and (S24) becomes the condition $F > F^*$. If instead $a < b$, then $a \leq \tau < b$, so that $a - b \leq \tau - b < 0$; dividing by $a - b < 0$ gives $F^* \in (0, 1]$, and the direction of the inequality (S24) is reversed to yield $F < F^*$. In both cases, $F^* \in [0, 1]$ and (S24) places $F_m(s)$ on the side of F^* that assigns greater weight to $b_{\max}(s)$.

The sufficiency of $b_{\min}(s) > \tau$ is case (i); the necessity of $b_{\max}(s) > \tau$ follows from case (ii). $\square$

When the shares coincide ($a = b = b_L$), (S23) reduces to $Q(s) = b_L(s)$ and the book profit is positive if $b_L(s) > \tau$. Intersecting the book-profit region with the better set $\mathcal{B}$ gives the Goldilocks Zone $\mathcal{G} = \{s : Q(s) > \tau\} \cap \mathcal{B}$, with $Q(s)$ given by (S23). If F_m is strictly increasing, better profitability (19) may be written in terms of quantiles:

$$\mathcal{B} = \left(-\infty, F_m^{-1}(\tau)\right) \cup \left(F_m^{-1}(1 - \tau), \infty\right). \quad (\text{S25})$$

The sufficient and necessary conditions of the main text follow by intersecting cases (i) and (ii) with membership of $F_m(s)$ in one of these tails; only the Threshold regime (iii) requires the joint values of F_m , $b_{h,L}$, and $b_{v,L}$ at the same s , and is therefore evaluated empirically.