

Well Calibrated, Unevenly Priced: Market Calibration and the Location of the Bookmaker's Margin in 18,238 Modern Fixed-Odds Markets

Edward Glush

Bet Better Pty Ltd, Australia · ORCID 0009-0008-4074-369X · edward@betbetter.world

Working paper · August 2026 · This version: v1.0

ABSTRACT. The favourite–longshot bias is among the most replicated anomalies in the economics of wagering markets, yet the modern evidence base still leans heavily on historical parimutuel horse-racing data. I test the calibration of contemporary online fixed-odds markets using both sides of 18,238 settled two-outcome markets (36,476 outcome observations) across five sports between November 2024 and July 2026, with bookmaker margin removed from all quoted probabilities. Two findings emerge. First, de-vigged market probabilities are impressively well calibrated: in every probability decile, realized outcome frequencies fall within roughly two percentage points of vig-free implied probabilities. Second, the residual deviation is not noise but a systematic, antisymmetric tilt: outcomes priced below 50% overperform their vig-free probabilities by one to two points while favourites symmetrically underperform, and the tilt roughly doubles (± 5 points) in player-proposition markets — the same market family that carries the highest measured margins. The pattern is consistent with bookmakers allocating margin disproportionately to longshot prices rather than spreading it evenly, so that the classic favourite–longshot bias in raw prices survives, in attenuated antisymmetric form, even after de-vigging. The full dataset is openly published (CC BY 4.0).

KEYWORDS: favourite–longshot bias; calibration; sports betting; prediction markets; overround; de-vigging; player props. **JEL:** G14, D84, L83.

1. Introduction

Betting odds are probability forecasts with money attached, and their accuracy has been studied since Griffith (1949) first documented that racetrack bettors systematically overvalue longshots. The favourite–longshot bias has since been examined across decades and markets, with the canonical modern treatments asking whether it reflects risk-love or probability misperception (Snowberg and Wolfers, 2010) or the information structure of the market itself (Ottaviani and Sørensen, 2010). Two gaps persist in the empirical base. First, much of it derives from parimutuel horse racing, a market design that differs fundamentally from the fixed-odds online sportsbooks that now dominate wagering. Second, most calibration studies work with raw quoted probabilities, in which the bookmaker's margin and any behavioural bias are confounded; results are then sensitive to the de-vigging method chosen (Štrumbelj, 2014). This paper measures calibration in modern multi-sport fixed-odds

markets after margin removal, on an open dataset constructed so that it carries information about the market only — not about any bettor's or model's selections.

2. Data and method

The data comprise every settled two-outcome market in a proprietary grading archive spanning November 2024 to July 2026: 18,238 markets across five sports (soccer, baseball, basketball, ice hockey and Australian rules football), divided between match-level markets (“game lines”) and player-proposition markets. For each market, all quoted outcome prices were converted to implied probabilities and the margin removed across the market's outcomes; the resulting vig-free probability is rounded to three decimal places and paired with the realized outcome.

Both-sides construction. For every outcome at vig-free probability p , its complement enters the dataset at $1 - p$ with the opposite result, giving 36,476 observations that net to exactly 0.500/0.500 by construction. This is the natural construction for bias analysis: genuine calibration deviations appear as an antisymmetric curve around 0.5, while any selection effect from which side happened to be recorded is eliminated. It also guarantees the dataset contains no information about anyone's betting selections. Rows carry month, sport and market-family identifiers only — no odds, prices, bookmaker identities, teams or players — and cannot be mapped to any individual market.

One methodological caveat is inherited by all studies of this kind: de-vig method choice affects how the removed margin is allocated across the probability range, and therefore affects measured tail behaviour (Štrumbelj, 2014). The published dataset documents this, and the proportional method used here is the most conservative baseline against which alternative allocations can be tested — itself a research question the open data supports.

3. Results

3.1 The market is a very good forecaster

Table 1. Calibration of vig-free market probabilities, all markets.

Vig-free probability band	Observations	Mean implied	Realized frequency	Gap (pts)
10–20%	2,144	17.5%	18.8%	+1.4
20–30%	5,263	25.2%	26.7%	+1.4
30–40%	4,909	34.7%	37.0%	+2.2
40–50%	5,515	45.7%	46.4%	+0.7
50–60%	6,077	53.8%	53.2%	–0.6

Vig-free probability band	Observations	Mean implied	Realized frequency	Gap (pts)
60–70%	4,945	65.2%	63.0%	–2.2
70–80%	5,251	74.7%	73.3%	–1.4
80–90%	2,157	82.5%	80.9%	–1.6

In every band, realized frequency lies within roughly two percentage points of the vig-free implied probability. An event the market prices at a 65% chance happens about 63% of the time; a 25% chance happens about 27% of the time. As a pure forecasting machine, the aggregated fixed-odds market is impressively calibrated — a result consistent with the efficient-aggregation view of betting markets and a demanding benchmark for any forecasting model.

3.2 The residual tilt is systematic and antisymmetric

The deviations in Table 1 are not noise around zero. Every band below 50% overperforms its vig-free probability (+0.7 to +2.2 points) and every band above 50% underperforms (–0.6 to –2.2 points), the antisymmetry being partly mechanical under the both-sides construction but its sign and magnitude being informative. In raw quoted prices this pattern manifests as the classic favourite–longshot bias; here it survives margin removal in attenuated form.

Splitting by market family sharpens the picture: the tilt in player-proposition markets is roughly twice that of match-level markets, reaching approximately ± 5 points in the 30–40% / 60–70% bands. Companion measurements on the same operators (Glush, 2026c) show prop markets also carry the highest quoted margins (7.93% against 4.77% for two-way head-to-head markets). The markets where the bookmaker charges most are also the markets where vig-free prices deviate most from realized frequencies.

3.3 Interpretation: the margin lives on the longshots

Read jointly with the measured margin structure, the parsimonious interpretation is that bookmakers do not spread margin evenly across a market's outcomes: they load it onto the longshot side, where demand is least price-sensitive. Under proportional de-vigging, margin concentrated on longshots leaves a residual signature exactly of the observed form — longshots overperform their vig-free price because too much of the removed margin was attributed to them. This is the “pricing” reading of the favourite–longshot bias: whether its ultimate origin is bettor risk-love, misperception (Snowberg and Wolfers, 2010) or informed pricing by the bookmaker (Ottaviani and Sørensen, 2010), its proximate expression in modern fixed-odds markets is a deliberate, measurable asymmetry in where the margin sits. For the bettor, the practical corollary is that long-odds bets are mispriced twice over: they carry more than their share of the margin, and multi-leg bets compound that margin multiplicatively (Glush, 2026d).

4. Limitations

Markets enter the archive through an analytics pipeline rather than by census, so coverage per sport and month varies, although the both-sides construction removes selection over which side of a market is observed. The window covers twenty-one months of the sporting calendar. Bands beyond 10–90% are thinly populated and excluded from Table 1. The de-vig allocation caveat of Section 2 applies symmetrically to all findings; the open data permit re-analysis under alternative methods.

5. Data availability

The full dataset (36,476 rows: event month, sport, market family, vig-free probability, outcome) is published under CC BY 4.0: Zenodo doi:10.5281/zenodo.21764674 (mirrors: OSF doi:10.17605/OSF.IO/UV4C5; B2SHARE doi:10.23728/b2share.sjg4e-oj521; Harvard Dataverse doi:10.7910/DVN/Y1889T). No odds, prices, bookmaker identities, teams or players are included. Companion analysis: betbetter.world/studies/market-calibration. The companion margin measurements are published separately (Zenodo doi:10.5281/zenodo.21755380).

References

- Glush, E. (2026a). Market Calibration Dataset: de-vigged betting probabilities vs real outcomes. Zenodo. doi:10.5281/zenodo.21764674.
- Glush, E. (2026b). Bookmaker Margin Panel: daily measured betting margins across 22 leagues and player-prop markets. Zenodo. doi:10.5281/zenodo.21755380.
- Glush, E. (2026c). The Price of a Bet: Measured Bookmaker Margins Across 52 Bookmakers, 22 Leagues and Player-Prop Markets. SocArXiv working paper.
- Glush, E. (2026d). The 5% Price and the 10% Hold. Bet Better research note. betbetter.world/studies/price-vs-hold.
- Griffith, R. M. (1949). Odds adjustments by American horse-race bettors. *American Journal of Psychology*, 62(2), 290–294.
- Ottaviani, M., & Sørensen, P. N. (2010). Noise, information, and the favorite-longshot bias in parimutuel predictions. *American Economic Journal: Microeconomics*, 2(1), 58–85.
- Snowberg, E., & Wolfers, J. (2010). Explaining the favorite–long shot bias: is it risk-love or misperceptions? *Journal of Political Economy*, 118(4), 723–746.
- Štrumbelj, E. (2014). On determining probability forecasts from betting odds. *International Journal of Forecasting*, 30(4), 934–943.
- Winkelmann, D., Deutscher, C., & Ötting, M. (2021). Bookmakers' mispricing of the disappeared home advantage in the German Bundesliga after the COVID-19 break. *Applied Economics*, 53(26), 3054–3064.
-

Disclosure: the author operates Bet Better, a sports analytics service, and has a commercial interest in the subject matter. The dataset is constructed so as to carry no information about any party's selections; all published data are irreversible aggregates of de-vigged probabilities and outcomes. This paper is descriptive and is not betting advice. 18+; most people lose money gambling.