ECON30580 Economics of Betting Markets
10. Efficiencies and Favourite-Longshot Bias

Karl Whelan

School of Economics, UCD

Spring 2026

Part I

Efficiencies in Bookmakers' Odds

Efficiency in Fixed-Odds Betting Markets

- Our monopoly model predicts odds will be set as

$$D = \sqrt{\frac{1-\mu}{p\,(p+\delta)}}$$

- At these odds, the bookmaker should make a profit: Weak efficiency should hold, meaning there is no easy strategy to make money from bookmakers.
- But the model predicts expected payouts are

$$pD = \sqrt{\frac{(1-\mu)\,p}{p+\delta}}$$

  which rises with p.

- So strong efficiency should not hold. The expected returns are not the same on all bets.
- There should be a **favourite-longshot bias** with lower returns on longshot bets.
- Here we check evidence on both weak and strong efficiency.

The Football-Data.co.uk Dataset

- Gambling expert Joseph Buchdahl maintains two websites that are commonly used by academic researchers examining betting markets.
  1. **www.football-data.co.uk** contains historical data on results of matches and betting odds for many soccer leagues.
  2. **www.tennis-data.co.uk** does the same for professional tennis
- From this source, I have put together a dataset from 11 European countries (Belgium, England, France, Germany, Greece, Italy, Netherlands, Portugal, Scotland, Spain, Turkey) providing 221,738 matches with results from 1993 onwards and 151,254 matches with results and average betting odds from 2005 onwards.
- Suppose you had to use this dataset to figure out which team was more likely to win a soccer match.
- You would need to develop a model of some sort.
- Next, we will describe one of the most popular models for this purpose: The Elo model.

The Elo Model

- In 1960, physicist Arpad Elo came up with a new rating system for assessing chess players.
- Elo's model worked as follows: If player A with rating score R_A played player B with rating score of R_B, then the probability of A beating B was

$$P_A = \frac{1}{1 + 10^{(R_B - R_A)/400}}$$

- Technically, Elo was assuming that the probability of A winning was determined by the cumulative distribution function (CDF) of the logistic distribution
- The number 400 was set so that if one player had an Elo rating that was 400 greater than the other player, then they had a 91% chance of winning.
- The Elo system assigned each player a rating of 1500 to begin and would update the rating after each game so that if the player won the game, their rating improved and if they lost the game, their rating fell.
- Let's describe how this could be applied to soccer.

Applying the Elo Model to Soccer

- Let's define the outcome from team A playing team B as
  - $S_A = 1$: Team A wins.
  - $S_A = 0.5$: Draw.
  - $S_A = 0$: Team A loses.
- Define the expected outcome for team A as

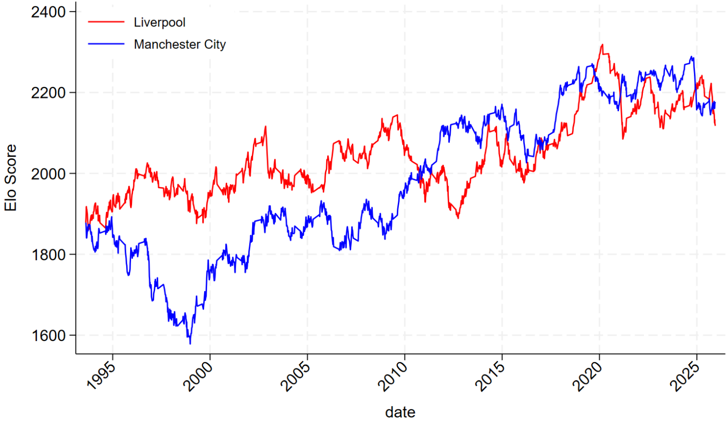
$$E_A = \frac{1}{1 + 10^{(R_B - R_A)/400}}$$

- The Elo rating for team A is updated as

$$R'_A = R_A + K(S_A - E_A)$$

  where K is an arbitrary parameter defining how sensitive the ratings are to recent form. The more unexpected the result, the bigger the change in rating.
- I applied the Elo model to the Football-Data dataset and found $K = 27$ produced the ratings with the best power for predicting match outcomes.
- The next page shows the Elo ratings I calculated from this dataset for Liverpool and Man City from the mid-1990s onwards.

Elo Ratings for Liverpool and Man City



Forecasting Soccer Matches: Elo versus Betting Odds

- We have built a pretty sophisticated model for use in forecasting the outcomes of soccer matches.
- How do Elo ratings compare with the information from bookmakers' odds in forecasting match outcomes?
- On the next page, we show output from a regression of the match outcome (1 if a home win, 0.5 if a draw, 0 if an away win) on the home and away Elo ratings. It has an R^2 of 0.0911.
- The page after shows the regression of the match outcome on the normalised probabilities of a home win and an away win (using the average odds). This has an R^2 of 0.1213. Bookmakers' odds far out-perform Elo scores.
- The page after shows Elo scores add almost no additional forecasting power when added to the bookmaker-derived probabilities, increasing the R^2 by only 0.0003.
- This shows bookmakers are factoring in all sorts of additional information beyond the results data (injuries, specific match-ups suiting some teams and not others and so on). They may also be modelling the results data with a more sophisticated model.

Regression of Outcome on Elo Ratings

```
. reg outcome elo_home elo_away
```

Source	SS	df	MS
Model	2452.39398	2	1226.19699
Residual	24474.8026	151,251	.161815807
Total	26927.1966	151,253	.178027521

Number of obs	=	151,254
F(2, 151251)	=	7577.73
Prob > F	=	0.0000
R-squared	=	0.0911
Adj R-squared	=	0.0911
Root MSE	=	.40226

outcome	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
elo_home	.0009715	8.37e-06	116.03	0.000	.0009551	.0009879
elo_away	-.0009669	8.37e-06	-115.58	0.000	-.0009833	-.0009505
_cons	.5689198	.0087051	65.35	0.000	.5518579	.5859816

Regression of Outcome on Normalised Probabilities

```
. reg outcome probhome probaway

      Source |       SS           df       MS      Number of obs   =   151,254
-------------+----------------------------------   F(2, 151251)    =  10443.70
       Model |  3267.36446         2  1633.68223   Prob > F        =    0.0000
    Residual |  23659.8322   151,251  .156427608   R-squared       =    0.1213
-------------+----------------------------------   Adj R-squared   =    0.1213
       Total |  26927.1966   151,253  .178027521   Root MSE        =    .39551

------------------------------------------------------------------------------
     outcome | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    probhome |    .554967   .0263278    21.08   0.000     .5033651    .6065689
    probaway |  -.5357599   .0291717   -18.37   0.000    -.5929358    -.478584
       _cons |    .492488   .0199654    24.67   0.000     .4533563    .5316197
------------------------------------------------------------------------------
```

Regression of Outcome on Elo and Normalised Probabilities

```
. reg outcome probhome probaway  elo_home elo_away
```

Source	SS	df	MS
Model	3273.62026	4	818.405066
Residual	23653.5764	151,249	.156388316
Total	26927.1966	151,253	.178027521

Number of obs	=	151,254
F(4, 151249)	=	5233.16
Prob > F	=	0.0000
R-squared	=	0.1216
Adj R-squared	=	0.1215
Root MSE	=	.39546

outcome	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
probhome	.613578	.0292986	20.94	0.000	.5561534	.6710027
probaway	-.5655797	.0302719	-18.68	0.000	-.6249119	-.5062474
elo_home	-.0001005	.0000172	-5.85	0.000	-.0001341	-.0000668
elo_away	.0001075	.000017	6.32	0.000	.0000741	.0001408
_cons	.4650899	.0206147	22.56	0.000	.4246855	.5054943

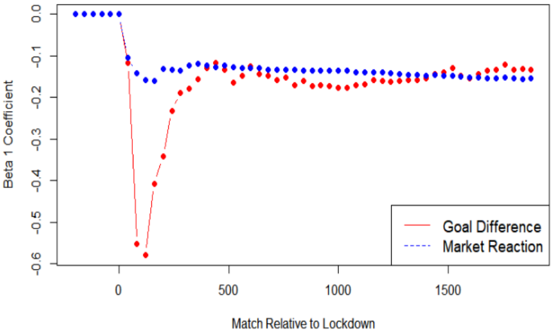
Reacting to News

- The Elo example shows that bookmakers are very good at figuring out how likely participants in sporting events are to win.
- There are also plenty of examples of where bookmakers can be seen to respond efficiently to news that a competitor is likely to get stronger or weaker.
- Studies focusing on things like injuries to star players or the response of in-game odds to goals or sendings off have shown that bookmakers adjust odds very quickly in line with the fundamental changes in probabilities implied by these events.
- See the paper in Brightspace by Ferguson and Pinnuck on how betting markets change odds when there are injuries to "superstar" Aussie rules players.
- Bookmakers even turn out to be very good at figuring out the implications of events where there is almost no historical data on which to build a model.
- See the discussion on the next page of how bookmakers responded to soccer games being played behind closed doors during the Covid pandemic. The evidence comes from my first paper on betting odds, co-written with Tadgh Hegarty.

Covid-19 and Home Advantage in Soccer

- Prior to Covid, home advantage was worth an average of 0.25 goals per game to the home team.
- Asian Handicap betting adds a "spread" to the weaker team's score. When bookmakers offering these bets started taking bets on the games with no crowds, their handicaps/spreads implied they thought the home advantage effect had been cut in half.
- The next page shows the average change in the "home advantage" effect on goal difference (estimated from regressions adding more data each week for new matches) and the size of the change priced by bookmakers.
- The bookmakers nailed the size of the change in home advantage effect that emerged once we had enough data to be confident in it.
- Some economists argued that the early discrepancy between red and blue lines meant there was an inefficiency in betting markets but I don't think so.
- The averages shown by the red line for the first few regressions are based on very small samples. The final numbers on the red line are based on a large enough sample to be reliable and we can see that (without having seen all this data) the bookmakers got the size of the effect pretty much perfectly.

Actual Average Change in Goal Difference Advantage for Home Teams (Red) and the Estimated Change in Bookmakers' Pricing of This Effect (Blue)



The x-axis shows the number of matches in the dataset that had been played. So we are adding more data as we move to the right.

Odds Get More Accurate as the Game Approaches

- I gathered a dataset of real-time odds from different bookmakers from Odds-API.com. Quotes from
  - UK-based bookmakers on 1,120 English Premier League games
  - US-licensed bookmakers on 3,669 NBA games.
  - Three seasons: 2022/23 to 2024/25
- I calculated normalised probabilities for each quote and checked the probabilities against actual outcomes, calling the outcome 1 if it happened and zero otherwise. I then calculated the correlation coefficient between normalised probabilities and the outcomes.
- The table below shows the correlations between the normalised probabilities and outcomes. They show the predictive power of forecasts getting gradually better as we get closer to the game.
- The gains are not huge–most of the useful information is available by a day or two before a game starts–but it shows that new information is absorbed and the "wisdom of crowds" effect of aggregating information from the public is still at work even in the final hour before a game starts.

Normalised probabilities for EPL and NBA are better forecasts of outcomes as we get closer to the game starting

Correlations between normalised probabilities and outcomes for 1,120 English Premier League matches and 3,669 NBA games sorted by time prior the start of games

	EPL	*NBA*
48 hours before	0.3951	
24 hours before	0.3959	0.3905
18 hours before	0.3974	0.3927
12 hours before	0.3986	0.3987
6 hours before	0.3992	0.4012
1 hours before	0.4021	0.4058
10 minutes before	0.4024	0.4073

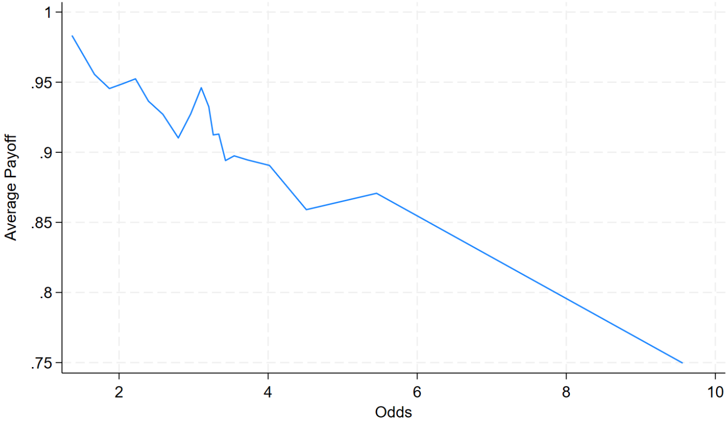
Part II

Favourite-Longshot Bias

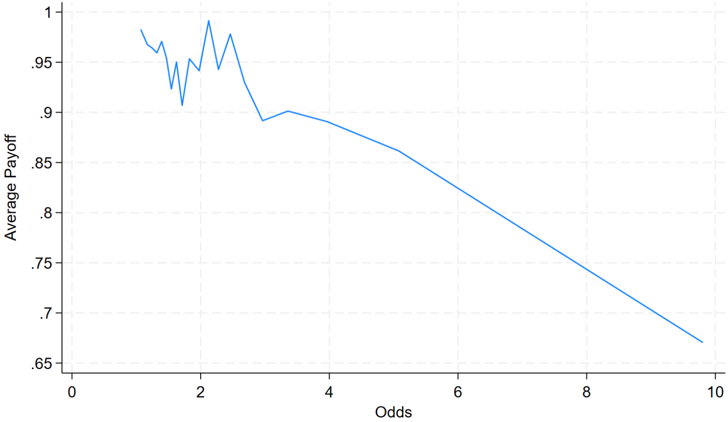
A Big Inefficiency: Favourite-Longshot Bias

- So fixed-odds betting markets are pretty efficient in the sense of being good at reacting to news and it being difficult to make a profit from them based on publicly available information.
- But recall that our definition of efficiency in betting markets was that the expected return on each bet on a contest should be the same.
- Most fixed-odds betting markets fail this test for the same reason that pari-mutuel markets did: There is a **favourite-longshot bias**.
- Let's illustrate this with the datasets on European soccer from 2005-2025 and professional tennis from 2010-2025 that we described earlier.
- The charts on the next two pages sort 453,765 bets on soccer and 262,638 bets on tennis into 20 "quantiles" for odds, going from low to high, and show the average payout on €1 bets for each quantile. Average payouts of less than 1 imply bettors losing on average. There is a very clear pattern of average payouts being lower on high-odds bets.
- The charts on the following two pages illustrate the pattern with the bets sorted by normalised probabilities: Just as our monopoly model predicts, loss rates accelerate as the probability of bet success falls.

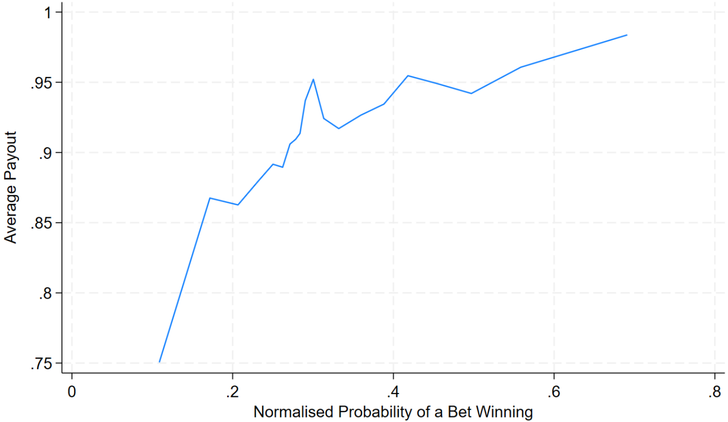
Average Payout Rates on €1 Bets on Soccer Sorted by Odds



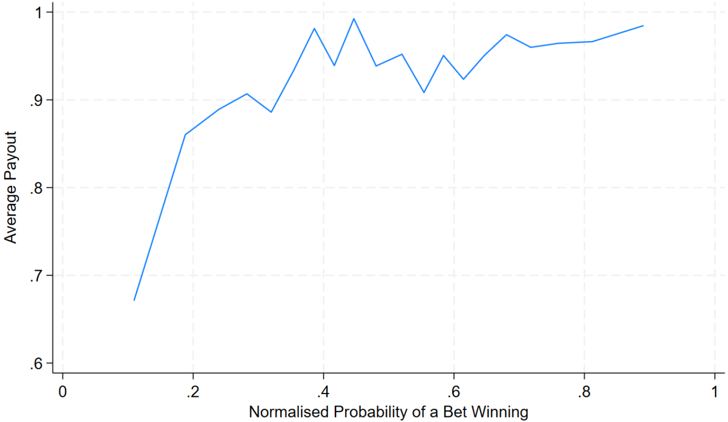
Average Payout Rates on €1 Bets on Tennis Sorted by Odds



Average Payout Rates on €1 Bets on Soccer Sorted by Normalised Probabilities



Average Payout Rates on €1 Bets on Tennis Sorted by Normalised Probabilities



Favourite-Longshot bias in horse racing and golf

Average loss rates on horse racing and golf bets

Horse Racing (4,107,315 quotes)		*Golf Majors (31,921 quotes)*	
Odds Range	*Loss Rate*	*Odds Range*	*Loss Rate*
1.01 to 5.5	10.8%	1.1 to 41	33%
5.55 to 9.5	17.3%	43 to 101	69%
9.6 to 15.0	27.3%	102 to 201	66%
15.1 to 26.0	39.7%	202 to 501	93%
Over 26	59.2%	Over 501	100%

Favourite-Longshot bias in ice hockey and American football with US bookmakers. "Moneyline" means backing a team to win.

Average loss rates on NHL and NFL moneyline bets with US-licensed bookmakers, 2022/23 to 2024/25 seasons

NHL (121,518 quotes)		*NFL (28,236 quotes)*	
Odds Range	*Loss Rate*	*Odds Range*	*Loss Rate*
1.00 to 1.57	0.8%	1 to 1.41	3.5%
1.58 to 1.80	0.8%	1.42 to 1.69	4.2%
1.81 to 2.08	6.0%	1.7 to 2.22	4.8%
2.09 to 2.48	6.4%	2.23 to 3	2.7%
Over 2.48	10.1%	Over 3	11.0%

An exception: No favourite-Longshot bias in NBA or MLB with US bookmakers. "Moneyline" means backing a team to win.

Average loss rates on NBA and MLB moneyline bets with US-licensed bookmakers, 2022/23 to 2024/25 seasons

Basketball (205,238 quotes)		*Baseball (219,290 quotes)*	
Odds Range	*Loss Rate*	*Odds Range*	*Loss Rate*
1.00 to 1.38	5.1%	1.00 to 1.62	4.6%
1.39 to 1.70	2.7%	1.63 to 1.81	3.0%
1.71 to 2.20	4.3%	1.82 to 2.05	4.2%
2.21 to 3.15	5.2%	2.06 to 2.36	3.7%
Over 3.15	0.9%	Over 2.36	3.4%

Model Fit and Explaining the Exception

- The results generally fit well with our monopoly pricing model's predictions.
- With realistic calibrations of the cost parameter *m* (about 0.02 or 0.03), the model fits the data well with relatively modest amounts of disagreement/over-optimisim.
- For example, the model with $\delta = 0.06$ (meaning the biggest optimist is 6 percentage points too high in their estimates relative to the truth) matches the soccer evidence well.
- What about the exceptions?
  - Favourite-longshot bias can only prevail when markets are not highly competitive. But there is a good reason for the new US sportsbooks to price bets in the NBA and MLB competitively.
  - NBA and baseball seasons generate a relentless flow of games, and for many customers they are the sports that keep the app open night after night. That makes them unusually valuable to bookmakers.
  - When a product is that central to customer attraction and retention, it becomes a natural candidate for more competitive pricing: give up some margin in order to attract and retain high-volume bettors.

Part III

Implications for Normalised Probabilities and Margins

Bad News for Normalised Probabilities

- Previously, we showed that in a strongly efficient betting market, where each bet in a contest had the same expected payout, the normalised probabilities equalled the true probabilities.
- But if bookmakers set odds so loss rates increase as the probability of bet success falls, then the normalised probabilities are no longer good estimates.
- To see this, assume that odds are determined by bookmakers according to

$$D_i = \frac{1 - m_i}{P_i} \qquad i = 1, ..., N$$

  where the average margins m_i depend negatively on the P_i.
- We showed before that in efficient markets with an equal expected payout m on all bets, we can calculate m as

$$\hat{m} = 1 - \frac{1}{\sum_{i=1}^{N} \frac{1}{D_i}}$$

  and the probabilities are estimated as

$$\hat{P}_i = \frac{1 - \hat{m}}{D_i}$$

Bad News for Normalised Probabilities

- Normalised probabilities are measured as

$$\hat{P}_i = \frac{1-\hat{m}}{D_i}$$

- But when the bookmaker is setting the odds with a favourite-longshot bias

$$D_i = \frac{1-m_i}{P_i} \qquad i = 1, ..., N$$

then the normalised probability calculation gives

$$\hat{P}_i = \left(\frac{1-\hat{m}}{1-m_i}\right) P_i$$

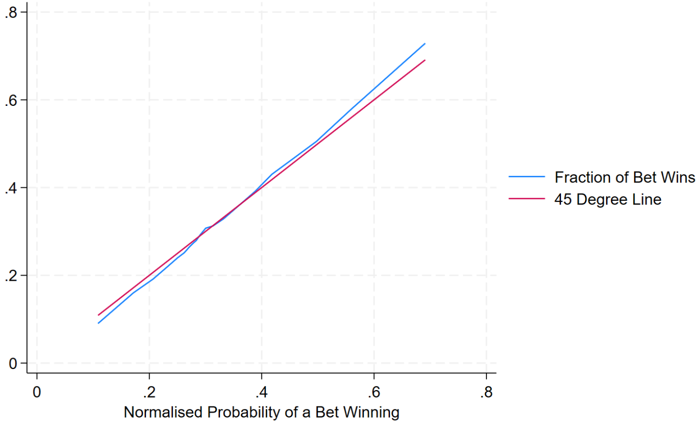
- Implications

1. For teams with low values of P_i and thus relatively high values of m_i, $\frac{1-\hat{m}}{1-m_i} > 1$ so the normalised probabilities are too high.
2. For teams with high values of P_i and thus relatively low values of m_i, $\frac{1-\hat{m}}{1-m_i} < 1$ so the normalised probabilities are too low.

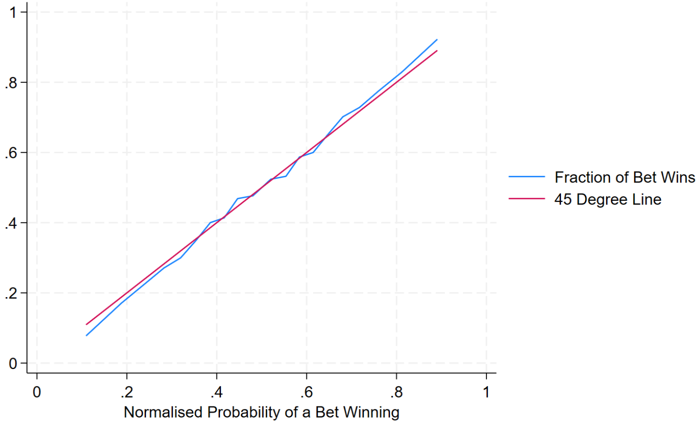
Confirming the Bad News

- The charts on the next two pages again use the data on bets on soccer and tennis.
- Again, the bets were sorted into 20 quantiles, this time based on the normalised probability of the bet winning.
- Our dataset includes the results of each game, so we know which bets won and lost.
- So we calculated the actual win percentage for each of the 20 groups of bets.
- If the probabilities are good estimates, then we should see the data showing a 45-degree line: The actual win rates should rise one-for-one with the normalised probabilities.
- We don't see this.
- Normalised probabilities overestimate win percentages for longshots and underestimate them for favourites. You can see this from the blue line (representing the actual fraction of bets that won) being below the red line (representing the normalised probability) for low normalised probabilities and above it for high normalised probabilities.

Actual Fraction of Wins on Soccer Bets Sorted by Normalised Probability that the Bet Wins



Actual Fraction of Wins on Tennis Bets Sorted by Normalised Probability that the Bet Wins



Bad News for Estimates of Expected Payouts

- We showed before that when markets were efficient and there were N possible outcomes of an event, you could estimate the expected payout on all bets as

$$1 - \hat{m} = \frac{1}{\frac{1}{D_1} + \frac{1}{D_2} + \ldots + \frac{1}{D_N}}$$

- But now the bookmaker is setting the odds with a favourite-longshot bias

$$D_i = \frac{1 - m_i}{P_i} \qquad i = 1, \ldots, N$$

so the $1 - \hat{m}$ estimate of the expected payout will be too high for some bets (longshots) and too low for others (favourites).

Even Worse News for Estimates of Expected Payouts

- There is even worse news than this. You might hope that the standard $\hat{m}$ calculation at least will give us the average value of m_i across the different bets but it does not do this.
- It turns out that, if bookmakers are pricing with a favourite-longshot bias, then the typical calculation of the expected payout (which has a complicated formula) can be approximated with the simpler formula

$$\hat{m} \approx \sum_{i=1}^{N} P_i m_i$$

  This weights the expected payouts on bets by their probability of success.
- The average margin across the N available bets will be

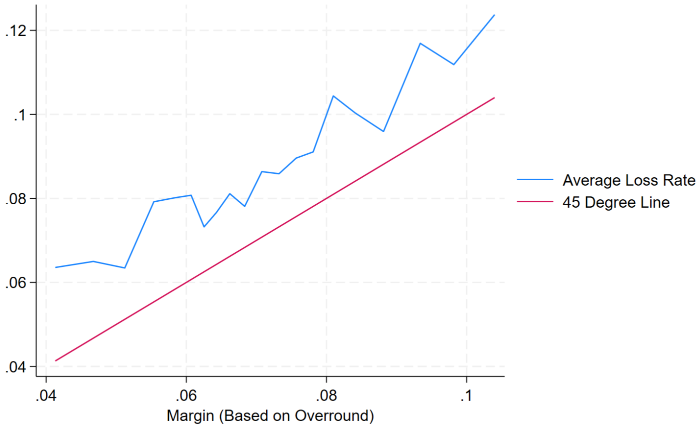
$$\frac{1}{N}\sum_{i=1}^{N} m_i > \sum_{i=1}^{N} P_i m_i = \hat{m}$$

- $\hat{m}$ puts more weight on the low margin for the favourites than on the high margin for the longshots. So our usual estimate of the profit margin **under-predicts average losses made across all N bets**.

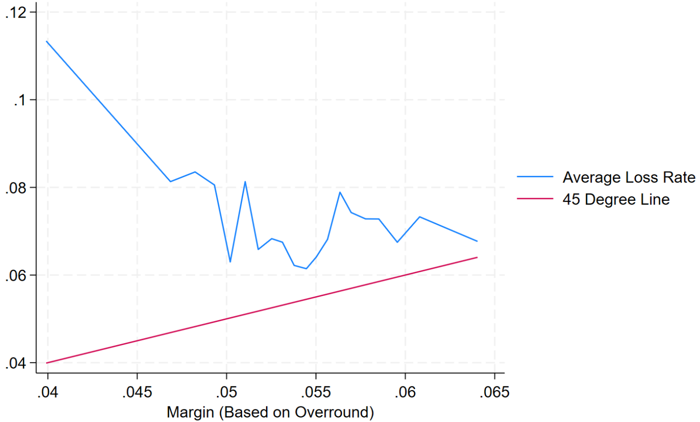
Confirming the Worse News

- Going to our soccer and tennis datasets, we can do the same thing for payout rates as we did before for win rates.
- In this case, we sorted all the bets into 20 quantiles based on the predicted average loss rate as measured by the bookmaker's margin as estimated using the overround.
- Then the blue line representing actual loss rates for each group of bets (the average loss on a €1 bet) is charted as well as the red 45-degree line (representing the predicted loss rate).
- Again, if the estimated payout rates are good estimates, then we should see the actual loss rates being approximately equal to the predicted rates.
- We don't see this: Actual average loss rates across bets are larger than predicted. In some cases, they are considerably larger.

Actual Average Loss Rates for Bets on Soccer Sorted by Overround-Based Estimate



Actual Average Loss Rates for Bets on Tennis Sorted by Overround-Based Estimate



Supplementary Material in the Draft Book

- Chapter 14: Evidence on Odds and Information
- Chapter 15: The Favourite-Longshot Bias and its Implications