



Cards, Fouls and the Referee

A booking-market model built on what the data says about officials, which is not what the tipping industry says.

A shrunk multiplicative count model for cards and fouls -- league and season, both clubs, the official, and how competitive the match is priced -- with a negative binomial for the over/under lines. It prices the markets a goals model cannot reach, and it is the only model in the library whose headline finding is that a popular angle does not exist.

SPORT

Football

FAMILY

Match pricing

LICENCE

£399

one-off, perpetual

RUNNING IN THE PLATFORM

<https://edgewortharena.com/arena/api/discipline/preview>

Inputs

NAME	TYPE	UNITS	WHERE IT COMES FROM
matches	list of (date, competition, season, home, away)	-	the platform's fixture ingest
match_stats	list of (fixture, home/away yellows, reds, fouls, corners)	counts	the football-data ingest
referee (optional)	string	-	the fixture record · absent for about seven matches in ten; an unknown official is priced as an average one
odds_1x2 (optional)	three decimal prices	decimal	any book, pre-match · used only to measure how competitive the match is, never as a target
lines (optional)	list of floats	goals/cards	the lines you want priced

Outputs

NAME	TYPE	UNITS	NOTES
expected count	float	cards, fouls or corners	
over/under probabilities and fair prices	dict per line	probability and decimal	
referee factor	float	multiplier	with the number of matches it rests on, so a thin estimate is visible as one
contest factor	float	multiplier	which decile of competitiveness the match falls in



Method

- 1 Model the count as a product: a league-season base, a multiplier for each club home and away, one for the official, and one for how competitive the market priced the match.

THIS IS THE SAMPLE

The method runs to 5 steps; one is shown. The assumptions, the limits, the module layout and the reference implementation come with the licence. Everything on the page below this -- what the model was measured at -- is the same in both versions, because that is the part you should not have to take on trust.

Assumptions

What has to be true for the numbers to mean what they say.

- A club's and an official's tendencies carry forward from the matches already seen. Measured by splitting each official's matches in half by date, they do: fouls correlate at 0.55 between halves and yellows at 0.43.

Limits

Where it is known to be wrong. Published because a model without stated limits is one nobody has pushed on.

- The referee is a much smaller term than the industry implies, and on the result it is nothing at all. On 14,002 refereed matches the split-half correlation of an official's effect is 0.05 on goals and 0.09 on home wins -- indistinguishable from noise. This model therefore makes no referee claim about the 1x2, and neither should anyone else.

LICENCE

Issued to Evaluation copy. One perpetual licence to use and modify this model inside your own organisation. Redistribution of the specification or the reference implementation is not included. No warranty is given: the validation above is the whole of the claim being made.