Data and code for: Fooled by performance randomness: over-rewarding luck

Romain Gauriot and Lionel Page

August 15, 2018

We bought the data from the company Opta.[1] Our license does not allow us to make the entire dataset public. Researchers interested by this dataset can acquire from Opta. Our contact in 2017 was Tonderai Kanhukamwe (Tonderai.Kanhukamwe@performgroup.com), commercial manager at Perform Group, the company owning the Opta brand.

We give here a small extract of our data points in the file data.dta (N=100). This extract shows the structure of the data and the variables we used.

The Stata code used to perform the estimations is provided in do files in this folder. Each do file reproduces one figure or table of the paper. Due to the smaller sample size available, the code may not always run. For instance, when we perform an exact match there might not be such exact match in the dataset. We included comments in table3.do to explain how the do file works. Other do files follow the same format.

Here is a brief description of the variables contained in data.dta:

- periodminute: Minute in the period at which the post occurred.
- periodid: 1 if the post occurred in the first period, 2 if it occurred in the second period.
- mv: Market value of the player given by Transfermarkt.
- note: rating given by the journalists.
- absy2 absx2: x-y co-ordinates of the shots in meters.
- postin: Dummy equal to 1 if the post is *in*.
- effectpost: Effect of the post on the score if it were to get in.
- att,def,mid: Position of the player given by Transfermarkt.
- homeshooter: Dummy equal to 1 if the player is playing at home.
- starting: Dummy equal to 1 if the player is in the starting team.

[1] `https://www.optasports.com/`

- sub: Dummy equal to 1 if the player is a substitute.
- avteamnoteseason: Average rating of the team on the previous matches of the season.
- avteamnote2: Average rating of the team excluding the player which made the shot.
- avteammvstarting: Average market value of the team.
- avotherteammvstarting: Average market value of the opposite team.
- avteammvstarting2: Average market value of the team excluding the player which made the shot.
- sumgoalseason: Number of goals scored in the previous matches of the season.
- sumnoteseason: Average rating in the previous matches of the season.
- sumteamnoteseason: Average team rating in the previous matches of the season.
- sumppostin: Frequency of posts inside in the previous matches of the season.
- nextmatchindataset: Dummy equal to 1 if the next match of the season is in the dataset. It is not in the dataset if the match is the last one of the season.
- previousmatchindataset: Dummy equal to 1 if the previous match of the season is in the dataset. It is not in the dataset if the match is the first one of the season.
- nextnote: Rating received on the next match. Empty if no rating is received on the next match.
- previousnote: Rating received on the previous match. Empty if no rating is received on the previous match.
- nextnote2: Rating received on the next match the player received a rating.
- previousnote2: Rating received on the previous match the player received a rating.
- nextngoalmatch: Number of goal scored on the next match.
- previousngoalmatch: Number of goals scored on the previous match.
- nextngoalmatch2: Number of goals scored on the next match played.
- previousngoalmatch2: Number of goals scored on the previous match played.
- pplayerwin: Ex-ante probability to win the match given by the betting odds.
- diffplay: ΔPlay. -1 if played the previous match but not the next. 0 if did not played the previous and the next match or if played both previous and next match. 1 if did not played the previous match but played the next match.
- diffminplay: ΔMinute.

- diffstarting: ΔStart.
- avminplayseason, avstartingseason, avplayseason: Avg. playing time over the season
- absx3, absy3: x-y co-ordinates of the shots on a 0-100 scale (for Figure 1).

For more information please email romain.gauriot@gmail.com.