Finding the imaginary winner of the Euroleague 2019/20

Sergio Olmos

DSPT

Wed, March 31, 2021

About me

Basketball analytics in Europe

Overview

1. Get data from Euroleague's website
2. Predict in-game win probability
3. How to rank each team?
4. Elo rating system
5. Point-spread predictive model
6. Simulation of future games

Data

- Game results
- Boxscore
- Shot location
- Play-by-play
- Player-tracking

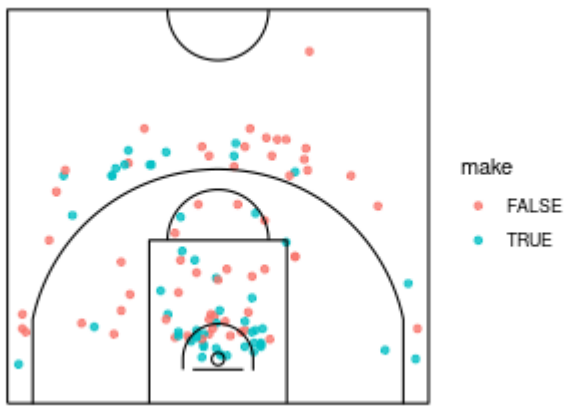

```
{
  "Live": false,
  "TeamA": "Zenit St Petersburg",
  "TeamB": "Zalgiris Kaunas",
  "CodeTeamA": "DYR       ",
  "CodeTeamB": "ZAL       ",
  "ActualQuarter": 5,
  "FirstQuarter": [
    {
      "TYPE": 0,
      "NUMBEROFPLAY": 2,
      "CODETEAM": "          ",
      "PLAYER_ID": "          ",
      "PLAYTYPE": "BP",
      "PLAYER": null,
      "TEAM": null,
      "DORSAL": null,
      "MINUTE": 1,
      "MARKERTIME": "",
      "POINTS_A": null,
      "POINTS_B": null,
      "COMMENT": "",
      "PLAYINFO": "Begin Period"
    },
    {
      "TYPE": 0,
      "NUMBEROFPLAY": 3,
      "CODETEAM": "DYR       ",
      "PLAYER_ID": "P005927   ",
      "PLAYTYPE": "TPOFF",
      "PLAYER": "AYON, GUSTAVO",
      "TEAM": "Zenit St Petersburg",
      "DORSAL": "34",
      "MINUTE": 1,
      "MARKERTIME": "09:59",
```

{eurolig}

- API wrapper around the official website of the Euroleague.
- Results, play-by-play and shot location data in a tidy format within R.
- Functions to analyze and visualize the data.

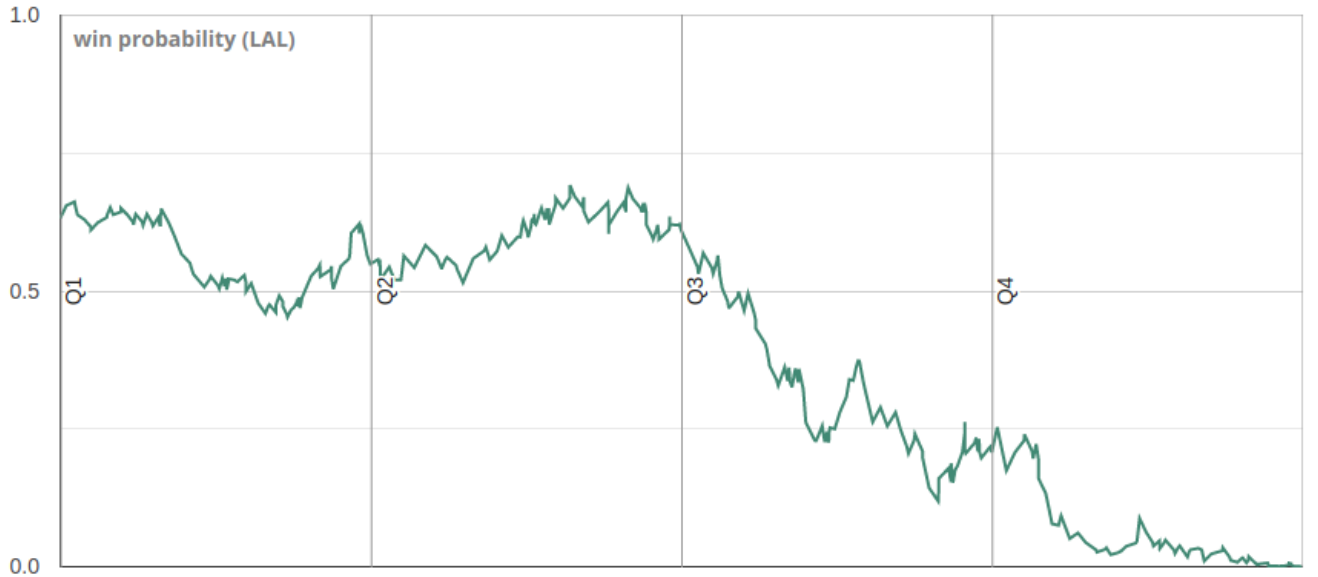
```
library(eurolig)

shots <- extractShots(260, 2017)
plotShotchart(shots)
```



In-Game win probabilities

- Play-by-play data
- Win probability at each event
- Logistic regression
- Needs baseline rating of each team in each game



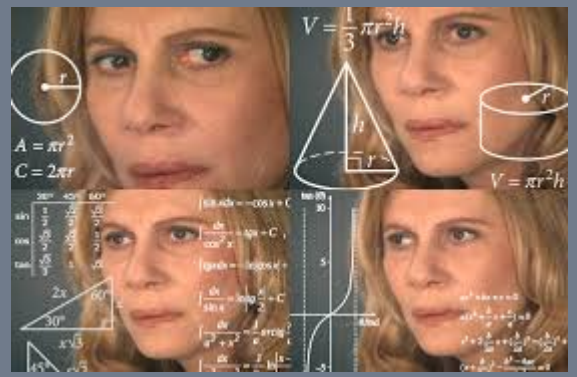
Source: `inpredictable.com`

Elo rating system

Elo rating system

- Everyone starts with the same initial rating
- A teams's rating is updated after every game
- Win + | Loss -
- Zero-sum
- Varying amount of points won/lost

Stay with me!



Blog post

Updating

- $R_A =$ Elo rating of Team A before the game
- $R'_A =$ Elo rating of Team A after the game

Algorithm

$$R'_A = R_A + k(S_A - P_A)$$

where

- $S_A = \begin{cases} 1 \text{ if A wins} \\ 0 \text{ if A looses} \end{cases}$
- P_A is A's expected probability of winning the game
- k is a tuning parameter

Example

Real Madrid vs. FC Barcelona: 85 - 77

Previous Elo rating

- Real Madrid: 1,600
- FC Barcelona: 1,500

Update

$$\text{New rating} = \text{Previous rating} + \text{How surprised}$$

Real Madrid

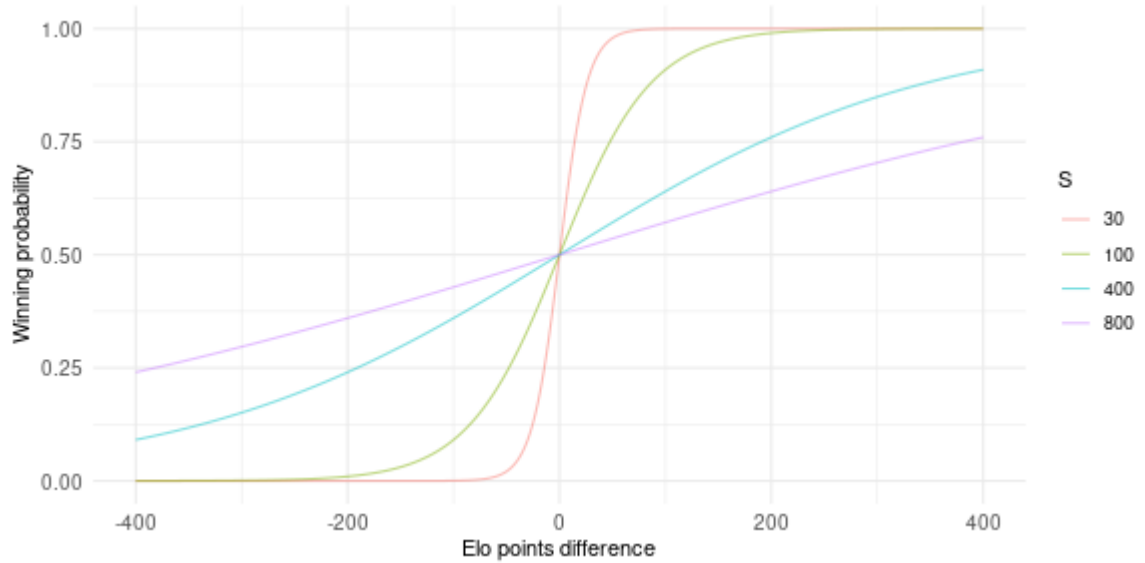
$$R'_{Madrid} = 1600 + k(1 - P_{Madrid})$$

Expected winning probability

$$P_A = \frac{1}{1 + 10^{-\Delta_{AB}/s}}$$

where

- Δ_{AB} is the difference in Elo points between team A and team B
- s is a scale tuning parameter.



Adding more context

Home court advantage

$$\Delta_{AB} = (R_A - R_B) + H_A$$

Margin of victory

More points to wider wins

Final updating rule

$$R'_A = R_A + k \times m_A(S_A - P_A)$$

where m_A is the margin of victory multiplier.

- Formula[1]:

$$m_A = \frac{(d+3)^{0.8}}{7.5 + 0.006\Delta_{AB}}$$

[1] From FiveThirtyEight's Elo model description.

Year-to-year carry over

- System starts in 2000/2001 season
- All teams start with an Elo rating of 1300

Should we keep previous season's ratings?

- Consistency in ratings between seasons
- Regression to the mean
- Weight parameter w

Model parameters

Grid search

- $k = 10, 11, \ldots, 50$
- $s = 200, 300, \ldots, 500$
- $h = 0, 1, \ldots, 150$
- $w = 0.5, 0.6, \ldots, 1$

Loss function

- Highest Elo rating ($\pm h$) wins
- Proportion of correct win predictions

Final parameter values

$k = 25$, $s = 400$, $h = 100$, $w = 0.8$

Example

Real Madrid vs. FC Barcelona: 85 - 75

Previous Elo rating

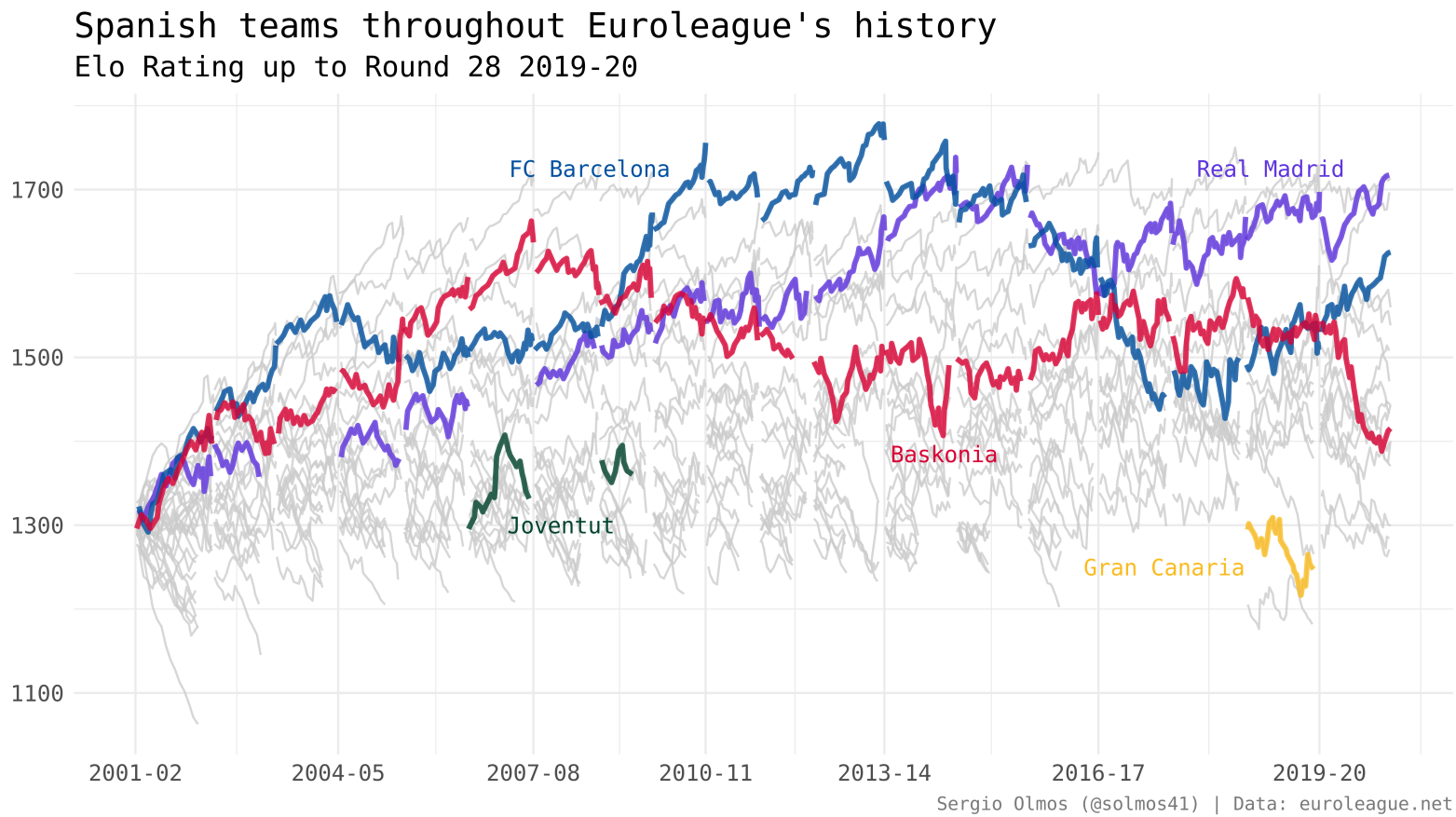
- Real Madrid: 1,600
- FC Barcelona: 1,500

Real Madrid

$$R'_{Madrid} = 1600 + 25 \times (1 - 0.6) = 1610$$


You made it!

Results


Spanish teams throughout Euroleague's history
Elo Rating up to Round 28 2019-20
FC Barcelona
Real Madrid
Baskonia
Joventut
Gran Canaria
1700
1500
1300
1100
2001-02
2004-05
2007-08
2010-11
2013-14
2016-17
2019-20
Sergio Olmos (@solmos41) | Data: euroleague.net

Simulation-based predictions

Initial Elo ratings

Pos	Equipo	Elo
1	Real Madrid	1717
2	Anadolu Efes Istanbul	1713
3	CSKA Moscow	1697
4	FC Barcelona	1625
5	Maccabi FOX Tel Aviv	1560

Point-spread predictive model

The margin-of-victory multiplier m_A needs the resulting point difference

Linear regression

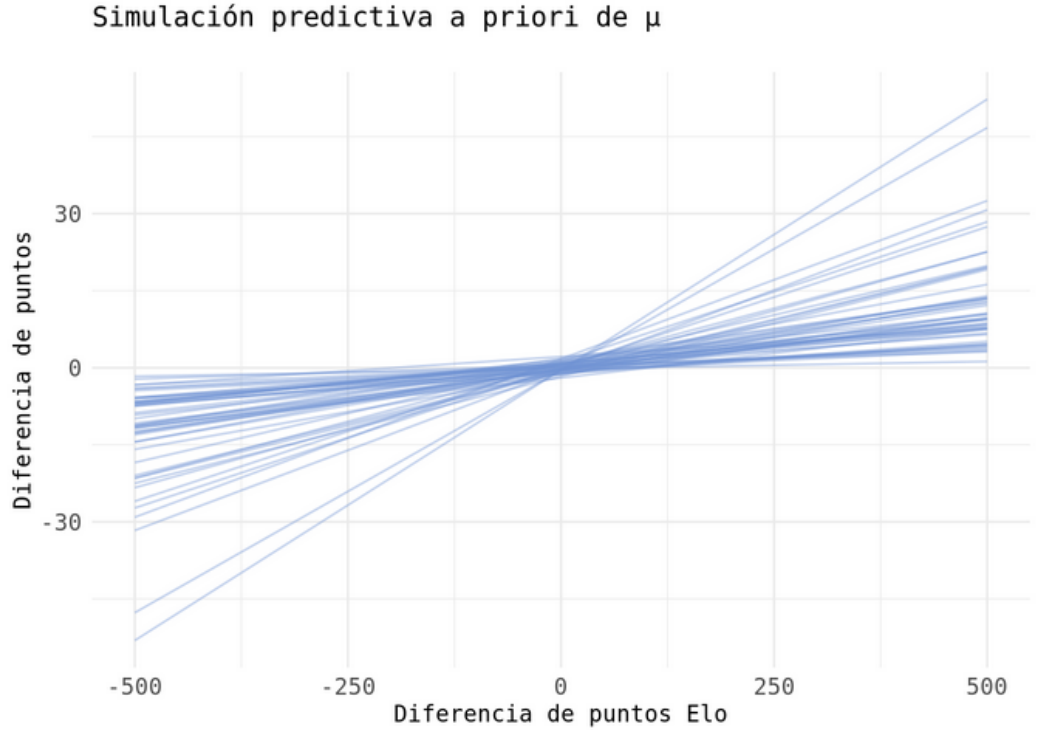
$$\text{Point difference} = \alpha + \beta \times \text{Elo difference}$$

Full model

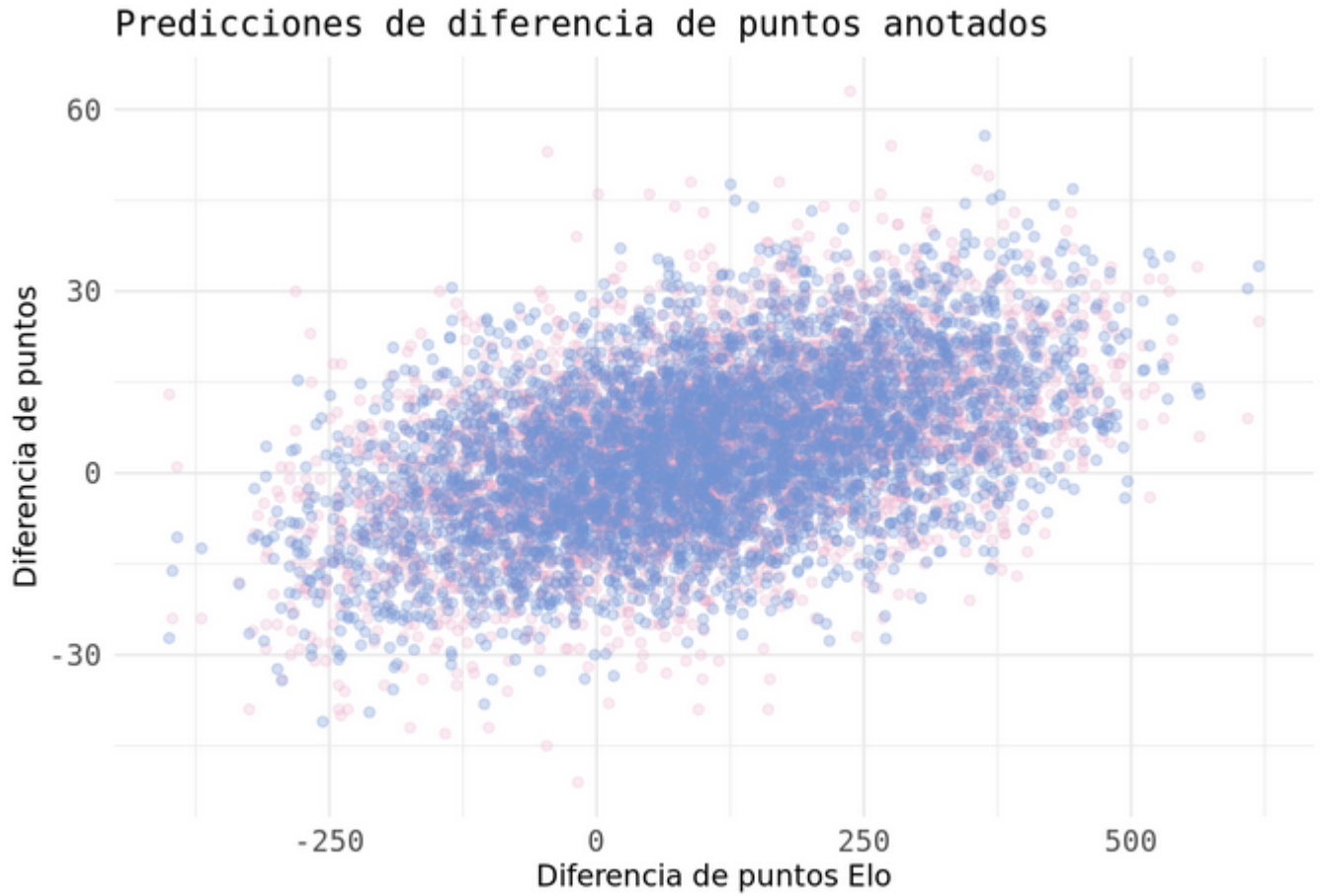
$$\begin{aligned} D_i &\sim \text{Normal}(\mu_i, \sigma^2) \\ \mu_i &= \alpha + \beta E_i \\ \alpha &\sim \text{Normal}(0, 1) \\ \beta &\sim \text{Log-Normal}(-4, 0.8) \\ \sigma &\sim \text{Uniform}(0, 50) \end{aligned}$$

Why Bayesian?

- Assign priors
- Samples of parameters to simulate from



Posterior predictive check



Simulations

1. Predict point-spread for each game sequentially
2. Update Elo Rating based on this scored-point difference
3. Repeat above steps 10,000 times

Challenges

Playoffs

- Choose best 8 teams
- Simulate best-of-5 playoff series

Final Four

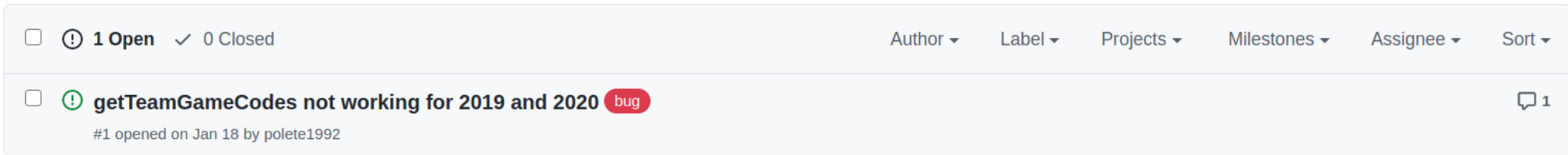
- Assumed all 4 teams played away

Predicted probabilities

Equipo	Elo	Playoffs	Final 4	Final	Campeón
Anadolu Efes Istanbul	1713	100.00%	91.84%	51.64%	28.74%
Real Madrid	1717	100.00%	68.44%	40.51%	21.33%
CSKA Moscow	1697	100.00%	73.63%	36.35%	20.18%
FC Barcelona	1625	100.00%	72.46%	33.57%	14.28%
Maccabi FOX Tel Aviv	1560	100.00%	28.19%	10.88%	4.66%

Future development

- Weekly updated probabilities
- In-game win probabilities
- {eurolig} needs your help!



Thanks!