Probabilistic Modelling of Football Results

Background - Football

- Goals are very rare events (often times the final score is 0-0), but still the main measure of success in a match.
- Goal-scoring models include Expected Goals (xG), and Poisson.
- Poisson focuses more on match/outcome results, e.g. probability of wins, draws, losses.
    - Limited in-depth analysis of a team’s performance.

Research Question

- **Can we model football results using a Poisson distribution with informative, team-specific parameters?**

- Do win-streaks influence a team's results?

Methodology

Assumptions

- Number of goals scored by a team in a match is a Poisson variable.
  - Each goal is independent of another.
  - Match duration is constant.
- Number of goals scored by a team depends only on their offensive strength and the opponent's defensive strength.
- Given the model's parameters, matches are independent of each other.
- A team's strengths/weaknesses do not change throughout the season.
- We focused on a single season: 2014-15 English Premier League.

Model

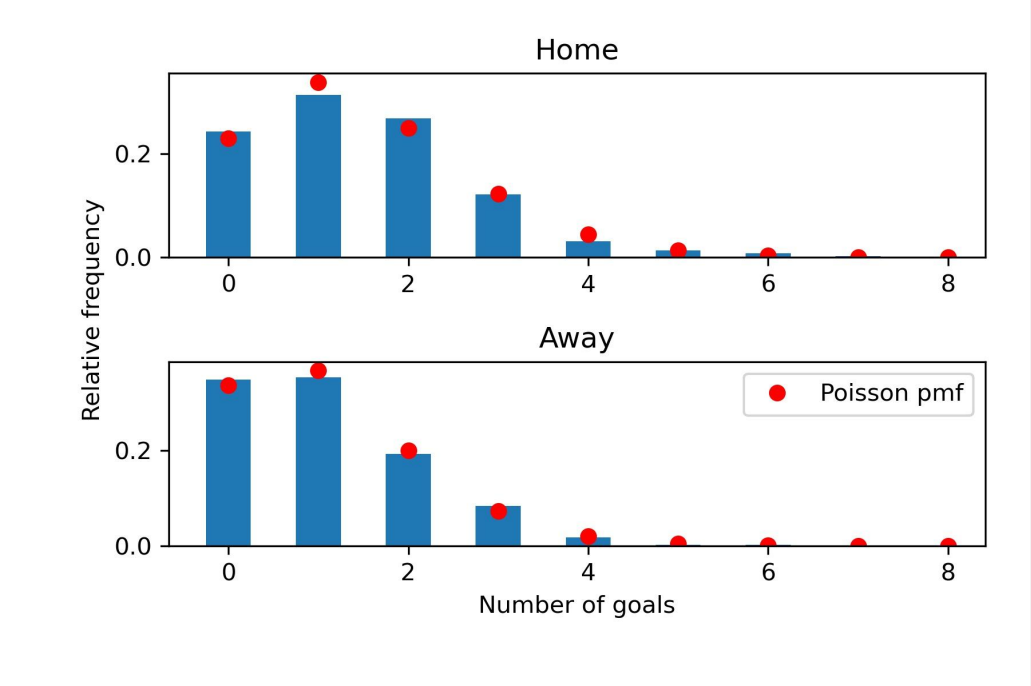
Match between teams i and j, with i playing at home, and a final score of ***n-m***

$$X \sim \text{Pois}(O_{H,i} D_{A,j})$$
$$Y \sim \text{Pois}(O_{A,j} D_{H,i})$$

$$P(X = n) = \frac{(O_{H,i} D_{A,j})^n \exp(-O_{H,i} D_{A,j})}{n!}$$
$$P(Y = m) = \frac{(O_{A,j} D_{H,i})^m \cdot \exp(-O_{A,j} D_{H,i})}{m!}$$

Statistical Significance

- Chi-Squared goodness of fit test
- P-values: 0.603 and 0.75 for home and away respectively.



Model Implementation

- Our main goal is to estimate the model parameters numerically
  - Use gradient descent to maximise log-likelihood of the data given our model parameters
  - Independence assumptions allow us to look at each match separately

$$P(X = n) = \frac{(O_{H,i}D_{A,j})^n \exp(-O_{H,i}D_{A,j})}{n!}$$

$$P(Y = m) = \frac{(O_{A,j}D_{H,i})^m \cdot \exp(-O_{A,j}D_{H,i})}{m!}$$

Model Implementation

$$O_{H,i} \leftarrow O_{H,i} + \gamma * \left(\frac{n}{O_{H,i}} - D_{A,j} \right)$$

$$D_{H,i} \leftarrow D_{H,i} + \gamma * \left(\frac{m}{D_{H,i}} - O_{A,j} \right)$$

$$O_{A,j} \leftarrow O_{A,j} + \gamma * \left(\frac{m}{O_{A,j}} - D_{H,i} \right)$$

$$D_{A,j} \leftarrow D_{A,j} + \gamma * \left(\frac{n}{D_{A,j}} - O_{H,i} \right)$$

Momentum

$$X \sim \text{Poisson}(O_{H,i} D_{A,j} \mathbf{s^{d_i}})$$

$$s \leftarrow s + \gamma * \left(\frac{n d_i}{s} - d_i O_{H,i} D_{A,j} s^{d_i - 1} \right)$$

$$s \leftarrow s + \gamma * \left(\frac{m d_j}{s} - d_j O_{A,j} D_{H,i} s^{d_j - 1} \right)$$

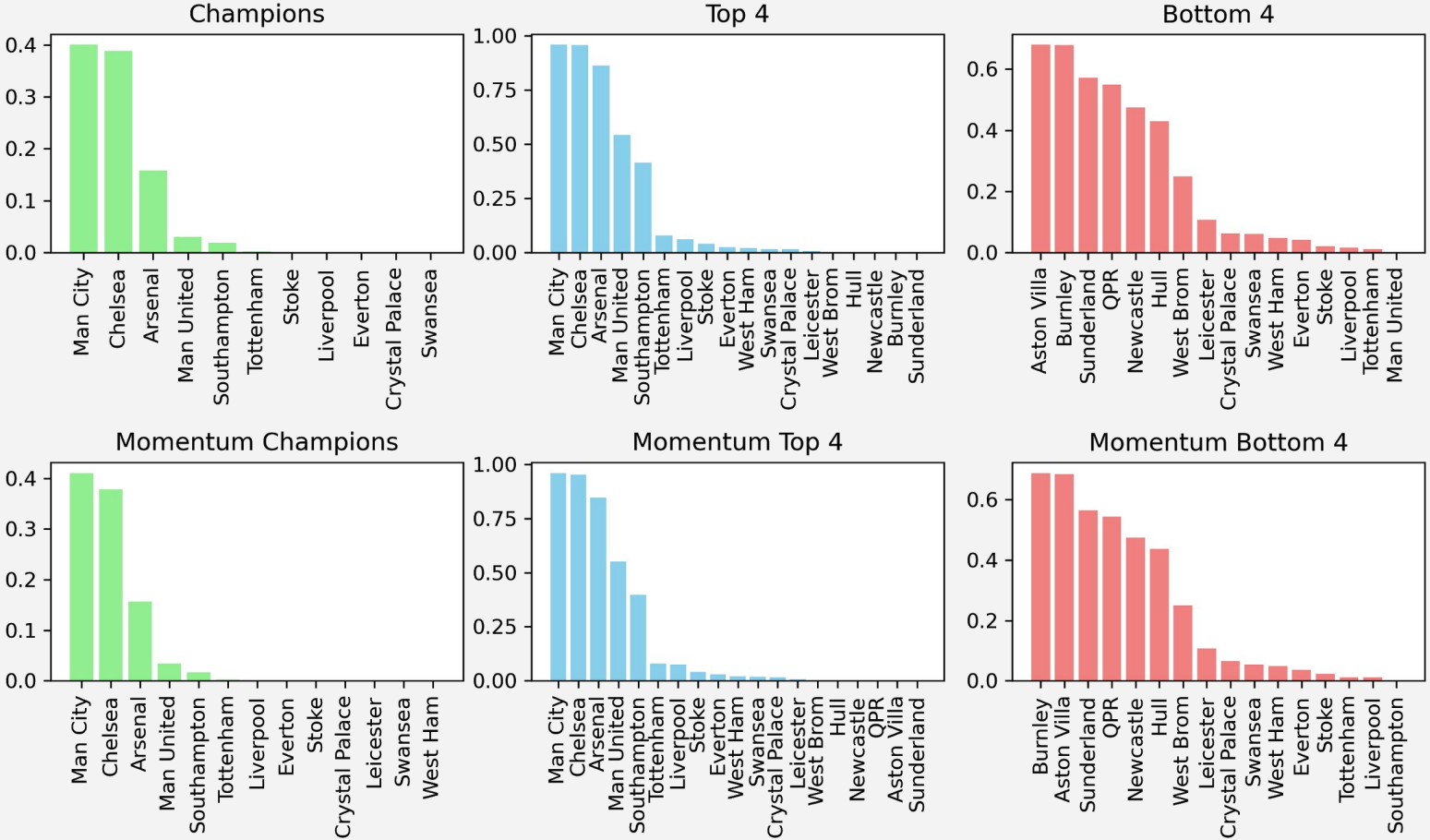
Results

Estimated parameters:

- Momentum parameters were very similar

$$s = 1.0033$$

	Team	Points	W	D	L	GD	GF	GA	O_H	D_H	O_A	D_A
1	Chelsea	87	26	9	3	41	73	32	1.53	0.48	1.78	1.00
2	Man City	79	24	7	7	45	83	38	1.87	0.76	1.91	1.08
3	Arsenal	75	22	9	7	35	71	36	1.74	0.74	1.47	0.98
4	Man United	70	20	10	8	25	62	37	1.74	0.77	1.03	0.99
5	Tottenham	64	19	7	12	5	58	53	1.33	1.24	1.35	1.28
6	Liverpool	62	18	8	12	4	52	48	1.30	1.03	1.09	1.24
7	Southampton	60	18	6	14	21	54	33	1.58	0.66	0.83	0.89
8	Swansea	56	16	8	14	-3	46	49	1.15	1.12	0.95	1.19
9	Stoke	54	15	9	14	3	48	45	1.36	1.11	0.80	1.02
10	Crystal Palace	48	13	9	16	-4	47	51	0.90	1.40	1.31	1.04
11	Everton	47	12	11	15	-2	48	50	1.17	1.07	1.03	1.26
12	West Ham	47	12	11	15	-3	44	47	1.09	0.92	0.94	1.26
13	West Brom	44	11	11	16	-13	38	51	1.03	1.31	0.71	1.09
14	Leicester	41	11	8	19	-9	46	55	1.22	1.12	0.91	1.45
15	Newcastle	39	10	9	19	-23	40	63	1.13	1.36	0.70	1.58
16	Aston Villa	38	10	8	20	-26	31	57	0.79	1.26	0.66	1.38
17	Sunderland	38	7	17	14	-22	31	53	0.69	1.35	0.75	1.10
18	Hull	35	8	11	19	-18	33	51	0.82	1.20	0.70	1.17
19	Burnley	33	7	12	19	-25	28	53	0.61	1.05	0.70	1.37
20	QPR	30	8	6	24	-31	42	73	1.03	1.22	0.95	2.12


Champions
Man City
Chelsea
Arsenal
Man United
Southampton
Tottenham
Stoke
Liverpool
Everton
Crystal Palace
Swansea
Top 4
Man City
Chelsea
Arsenal
Man United
Southampton
Tottenham
Liverpool
Stoke
Everton
West Ham
Swansea
Crystal Palace
Leicester
West Brom
Hull
Newcastle
Burnley
Sunderland
Bottom 4
Aston Villa
Burnley
Sunderland
QPR
Newcastle
Hull
West Brom
Leicester
Crystal Palace
Swansea
West Ham
Everton
Stoke
Liverpool
Tottenham
Man United
Momentum Champions
Man City
Chelsea
Arsenal
Man United
Southampton
Tottenham
Liverpool
Everton
Stoke
Crystal Palace
Leicester
Swansea
West Ham
Momentum Top 4
Man City
Chelsea
Arsenal
Man United
Southampton
Tottenham
Liverpool
Stoke
Everton
West Ham
Swansea
Crystal Palace
Leicester
West Brom
Hull
Newcastle
QPR
Aston Villa
Sunderland
Momentum Bottom 4
Burnley
Aston Villa
Sunderland
QPR
Newcastle
Hull
West Brom
Leicester
Crystal Palace
Swansea
West Ham
Everton
Stoke
Tottenham
Liverpool
Southampton

Observations

- Momentum had very little impact on overall performance
- As expected, goal-scoring proficiency has a great impact on a team's performance, according to our model
    - Manchester City were favoured
    - Southampton performed much better than their final standing suggests
    - Aston Villa over-performed - should have been relegated

Conclusion

Final Remarks

- Model relies heavily on goal-scoring proficiency, and less so on results and playstyle (because of assumptions).
- The model is able to tell whether a team is performing or not, but does not give the underlying reasons.
- Momentum has a very small impact on a team's performance.
- Somewhat expensive parameter calculation.
- Possible application to betting.

Possible Extensions

- Look at more data, and train for a longer time period to account for computational costs.
- Team dynamics such as player choices, injuries,
- Compare with more complex models that use in-game statistics.
- Compare the Premier League with other competitions.
    - The 2014-2015 English Premier League was somewhat one-sided.

Questions?