Evaluating Individual Football Player Performances

Robert Bajons

Institute for Statistics and Mathematics, Vienna University of Economics and Business

May 4, 2023

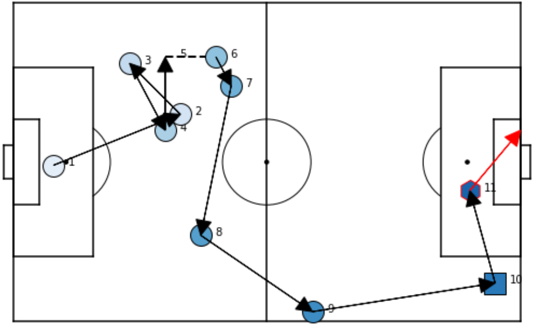
Evaluating Players in Football

- Classical approach: Use count based statistics such as goals, assist or number of shots to evaluate players (and even teams).

Advances in collecting data lead to more data driven approaches:

- Bottom-Up rating systems: assess player performances by assigning values to each action performed and then aggregate them for each player over the course of a relevant period (i.e. a match or a season).
  - Expected goals (xG, e.g. Robberechts and Davis (2020)).
  - Passing models (e.g Szczepański and McHale (2016) Håland et al. (2019)).
  - Action based models (VAEP Decroos et al. (2019)).
- Top-Down rating systems: evaluate players by breaking down the whole team performance and distributing credit onto players involved.
  - Plus-minus models (e.g. Hvattum (2019)).

Event Stream Data



	time	actiontype	player	team
1	26m2s	pass	M. ter Stegen	FC Barcelona
2	26m13s	pass	Sergio Busquets	FC Barcelona
3	26m16s	pass	Y. Mina	FC Barcelona
4	26m20s	pass	Sergio Busquets	FC Barcelona
5	26m23s	dribble	Y. Mina	FC Barcelona
6	26m26s	pass	Y. Mina	FC Barcelona
7	26m28s	pass	Iniesta	FC Barcelona
8	26m32s	pass	Piqué	FC Barcelona
9	26m34s	pass	Nélson Semedo	FC Barcelona
10	26m36s	cross	O. Dembélé	FC Barcelona
11	26m37s	shot	L. Suárez	FC Barcelona

Data Preprocessing

- Build a data set of "possessions":
  - Sequence of consecutive actions ended by either opponent team gaining possession or stopped by the referee.
- Extract important features from the data:
  - Possession relevant features: locational (start/end location, total distance, goal angle, etc.), temporal (game time, speed, etc.), discrete (freekick/corner, number of actions, number passes, score differential, etc.).
  - Player and team information: players involved, opponent team.
- Response variable:
  - Possession ends with goal.
  - Difference in possession value at the beginning and at the end (As measured by VAEP, xT, etc.).

Approaching the Problem of Evaluating Players

- Derive a rating from measuring a players contribution to (successful) possession.
- Naive approach: fit (generalized) linear model on possession data and interpret model coefficients $\beta \rightarrow$ problematic in high dimensional and imbalanced setup.
- Add regularization term to estimation of coefficients.
- Use a debiased machine learning (DML) approach (Chernozhukov et al. (2018)).

Debiased Machine Learning

Consider the following partially linear model:

$$\begin{aligned} Y &= D\theta_0 + g_0(X) + \epsilon, \qquad & \mathbb{E}[\epsilon|X, D] = 0 \\ D &= m_0(X) + \nu, \qquad & \mathbb{E}[\nu|X] = 0. \end{aligned} \tag{1}$$

θ_0 is parameter of interest, effect of the treatment variable D on the outcome Y. X are additional covariates influencing Y as well as D (confounders). g_0 and m_0 are nuisance functions.

Naive approach:

- Estimate $g_0(X)$ via a ML technique (in our high dimensional set-up e.g. via lasso, random forest, etc.) on a subsample of the data.
- Estimate θ_0 via OLS using the main sample.

$\Rightarrow$ introduces a (potentially severe) bias invalidating inference and interpretation of estimate of θ_0.

Debiased Machine Learning

DML idea: partialling out the effect of X, i.e. rewrite (1)

$$\begin{aligned} W &= \nu\theta_0 + \epsilon, & \mathbb{E}[\epsilon|D, X] &= 0, \\ W &= Y - \ell_0(X), & \ell_0(X) &= \mathbb{E}[Y|X] = m_0(X)\theta_0 + g_0(X), \\ \nu &= D - m_0(X), & m_0(X) &= \mathbb{E}[D|X]. \end{aligned} \tag{2}$$

DML steps:

1. Estimate m_0 and ℓ_0 using ML techniques, to obtain $\hat{W} = Y - \hat{\ell}_0(X)$ and $\hat{\nu} = D - \hat{m}_0(X)$.
2. Regress the residuals $\hat{W}$ on the residuals $\hat{\nu}$, to get an estimate of θ_0.

$\Rightarrow$ results in a debiased estimate for θ_0. Free of regularization bias (overfitting bias can be reduced by cross validation).

Setting up the model

In our setup:

- $Y \dots$ binary variable whether possession ends in a goal or not.
- $D \dots$ binary variable whether player is involved in possession or not.
- $X \dots$ sparse matrix of possession features, players involved and opponents.

Estimate of interest $\hat{\theta}_0 \Rightarrow$

- Use a DML approach as described above, to estimate θ_0.
- repeat procedure for every player, in order to estimate effect of each player.

Deriving a Player Rating

- $\hat{\theta}_{0,i}$ of player i can be interpreted as the level shift in probability of scoring, when player i is involved in the possession.
- Problem: Imbalance of data $\rightarrow$ overestimation (underestimation) of players with low (high) number of involvements.
- Derive a metric that accounts for number of involvement in possessions.

$$PCV_i = N_i(\hat{\theta}_{0,i} + \bar{p}).$$

- $N_i \ldots$ Number of involvement in possession of player i.
- $\bar{p} \ldots$ average probability of scoring from a possession.

Top 20 Players with PCV ranking

	Player	Role	Theta	Inv	PCV
1	L. Messi	Forward	0.014	1005	36.060
2	K. De Bruyne	Midfielder	0.007	1179	33.606
3	Malcom	Forward	0.013	783	27.775
4	Luis Alberto	Midfielder	0.011	836	27.236
5	L. Sané	Midfielder	0.021	630	27.185
6	F. Thauvin	Forward	0.011	816	27.105
7	T. Kroos	Midfielder	0.007	904	26.296
8	L. Suárez	Forward	0.019	630	25.988
9	R. Sterling	Forward	0.017	656	25.633
10	C. Eriksen	Midfielder	0.002	1051	25.565
11	Mohamed Salah	Forward	0.019	623	25.389
12	Suso	Forward	0.009	807	25.233
13	I. Perišić	Midfielder	0.010	761	24.543
14	K. Walker	Defender	0.011	751	24.420
15	Cristiano Ronaldo	Forward	0.025	519	24.317
16	Neymar	Forward	0.013	693	24.053
17	C. Immobile	Forward	0.030	460	23.927
18	L. Insigne	Forward	0.001	1037	23.798
19	Son Heung-Min	Forward	0.019	557	23.041
20	D. Payet	Midfielder	0.008	753	22.957

Validity of the Player Ratings

Hvattum and Gelade (2021)

"If ratings are accurate in assessing the capabilities of players, one should be able to accurately predict the outcome of a match based only on the ratings of the players involved."

- Use match outcome data of the 2017/18 season.
- Predict match results via two state of the art models:
  - Bivariate Poisson model (Karlis and Ntzoufras (2003)).
  - Ordered logistic regression (Arntzen and Hvattum (2021)).
- Use average player rating differences between teams as covariates.
- Evaluate predictive performances of the models with covariates using proper scoring rules (Brier score (BS), informational loss (IL)).

Validity Results

	Bivariate Poisson		Ordinal logistic	
Covariates	BS	IL	BS	IL
Intercept Only	0.638	1.526	0.638	1.526
PCV	0.575	1.400	0.575	1.397
VAEP	0.578	1.410	0.578	1.405
ELO	0.574	1.395	0.573	1.392

Conclusion

- Novel semi top-down player rating method using event stream data and a debiased machine learning approach.
- Applied to data of the 5 big European leagues for the 2017/18 season.
- Validity checks show that result are promising.
- Outlook and future work:
  - Perform reliability analysis: more data needed.
  - Validation of rating by experts and professionals.
  - Regularized regression approach.

The End

Thank You!

Contact: robert.bajons@wu.ac.at

References I

[1] Arntzen, H. and Hvattum, L. M. (2021). Predicting match outcomes in association football using team ratings and player ratings. *Statistical Modelling*, 21(5):449–470.

[2] Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68.

[3] Decroos, T., Bransen, L., Van Haaren, J., and Davis, J. (2019). Actions speak louder than goals: Valuing player actions in soccer. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, page 1851–1861, New York, NY, USA. Association for Computing Machinery.

[4] Hvattum, L. M. (2019). A comprehensive review of plus-minus ratings for evaluating individual players in team sports. *International Journal of Computer Science in Sport*, 18(1):1–23.

[5] Hvattum, L. M. and Gelade, G. A. (2021). Comparing bottom-up and top-down ratings for individual soccer players. *International Journal of Computer Science in Sport*, 20(1):23–42.

[6] Håland, E. M., Wiig, A. S., Stålhane, M., and Hvattum, L. M. (2019). Evaluating passing ability in association football. *IMA Journal of Management Mathematics*, 31(1):91–116.

References II

[7] Karlis, D. and Ntzoufras, I. (2003). Analysis of sports data by using bivariate poisson models. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(3):381–393.

[8] Robberechts, P. and Davis, J. (2020). How data availability affects the ability to learn good xg models. In Brefeld, U., Davis, J., Van Haaren, J., and Zimmermann, A., editors, *Machine Learning and Data Mining for Sports Analytics*, pages 17–27, Cham. Springer International Publishing.

[9] Szczepański, L. and McHale, I. (2016). Beyond completion rate: evaluating the passing ability of footballers. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 179(2):513–533.

Some more comparisons

	Bivariate Poisson		Ordinal logistic	
Covariates	BS	IL	BS	IL
PCV opp	0.584	1.417	0.583	1.415
PCV avg N opp	0.595	1.443	0.594	1.440
PCV avg N	0.586	1.427	0.586	1.423
Theta	0.606	1.465	0.606	1.464

Estimated Effect $\hat{\theta}_0$