European Football Player Valuation: Integrating Financial Models and Network Theory

AMS 2024 Fall Western Sectional Meeting

Dr. Jimmy Risk
Cal Poly Pomona

10/27/24

Motivation (Part 1)

- It is difficult to quantify the importance of any single player
- However, this is required (e.g. salary)
- Common numerical methodologies include leveraging marginal post-game stats
- **Opta Points** is a well known system what quantifies a player's game performance as a linear combination of

$$\text{Goals, Shots on/off target, Blocked Shots, Assists, Passes}, \cdots$$

Motivation (Part 2)

https://www.youtube.com/watch?v=kiU0IAkw0vI

- As the video shows, marginal post-game stats are insufficient in telling the whole story of plays in soccer

Q: So how do we attempt to summarize player performance?

Motivation (Part 2)

https://www.youtube.com/watch?v=kiUOIAkw0vI

- As the video shows, marginal post-game stats are insufficient in telling the whole story of plays in soccer

Q: So how do we attempt to summarize player performance?

(Augmented) Passing Matrices

Consider the network formed by any given team possession. If player i has the ball, we can exhaust all possible ways the ball exchanges control:

- pass to player j
- shot on goal
- turnover (or unsuccessful pass)

Let

$$(X_n)_{n=0}^{\infty}$$

be the process governing which player has control of the ball in any given team possession.

Ball Possession Example

Example

- ($X_0 = A$) Alan begins with the ball.
- ($X_1 = C$) Alan passes to Carl.
- ($X_2 = A$) Carl passes to Alan.
- ($X_3 = D$) Alan passes to David.
- ($X_4 = S$) David shoots at the goal.

- $X_k = S$ means the ball is shot at the goal from the player that had it at time $k - 1$[1]
- $X_k = U$ means "unsuccessful pass" (or turnover) by the player at time $k - 1$

This can be modeled as a Markov Chain with absorbing states[2] S and U!

[1] shot at goal, an actual goal, or a weighted combination of both; more discussion later

[2] since the team possession ends after a shot or turnover

Formal Markov Chain Definitions

Consider a team with M players.

- P be the augmented passing transition matrix.
  - $p_{i,j}$ is the probability player i passes to player j, $1 \leq i, j, \leq M$
  - $p_{i,S}$ is the probability player i has a shot on goal
  - $p_{i,U}$ is the probability of player i committing a turnover

$$p_{S,S} = p_{U,U} = 1 \qquad \text{(absorbing states)}$$

- $(\alpha_i)_{i=1}^{M}$ is the initial distribution, i.e. $\alpha_i = \mathbb{P}(X_0 = i)$

Assigning Player Value

A measure of player value can be determined from P and α.

- First attempt: **PageRank** *(Google search algorithm)*
  - Used in Pena et al. (2012) (PT12) (on passing transition matrix)
  - "A player is popular if he gets passes from other players."
  - Roughly assigns to each player the probability of having the ball after a reasonable number of transitions has been made *(with some regularization)*
  - Did not do well in our analysis
- How to incorporate
  - the importance of scoring;
  - the harmonious dance of the ball among players in a scoring attempt;
  - disincentives for turnovers?

Assigning Player Value

A measure of player value can be determined from P and α.

- First attempt: **PageRank** *(Google search algorithm)*
  - Used in Pena et al. (2012) (PT12) (on passing transition matrix)
  - "A player is popular if he gets passes from other players."
  - Roughly assigns to each player the probability of having the ball after a reasonable number of transitions has been made *(with some regularization)*
  - Did not do well in our analysis
- How to incorporate
  - the importance of scoring;
  - the harmonious dance of the ball among players in a scoring attempt;
  - disincentives for turnovers?

Markov Chain Approach

- At a basic level, a player brings value to a team if they are **involved** in a play that result in a scoring attempt.

Let

$$A_i = \{X_\ell = i \text{ for some } \ell = 0, 1, 2, \dots\}$$
$$q_i = \mathbb{P}(A_i | X_\infty = S)$$

- A_i is the event that player i has possession of the ball during a given team possession[3]
- q_i is the probability that player i is involved in a play that ended in a scoring attempt *(before a turnover)*.
- One can think: how crucial is player i to possessions that score?
- Thinking back to masters/PhD level classes on Markov chains, q_i is straightforward to calculate!

[3] assuming each team possession is iid

Remark on Literature

- A similar (same?) metric was developed by Duch et al. (2010) (DWA10), dubbed *flow centrality*.
- The paper did not clearly lay out the mathematics.[4]
- It was difficult to find any other works that focused specifically on a Markov chain approach.

[4] so I'm not actually sure if it's the same as what we're doing or not!

Markov Chain Calculations

For simplicity assume that all states communicate and note that the Markov chain is finite. For brevity, denote $S := \{X_\infty = S\}$

$$q_i = \mathbb{P}(A_i|S) = \frac{\mathbb{P}(A_i \cap S)}{\mathbb{P}(S)}.$$

- Compute numerator and denominator separately using
  - law of total probability, and
  - knowledge of absorption probabilities.

Markov Chain Calculations (Numerator)

Numerator (recall $\alpha_i = \mathbb{P}(X_0 = i)$):

$$\mathbb{P}(A_i \cap S) = \sum_j \alpha_j \mathbb{P}(A_i \cap S | X_0 = j)$$
$$= \sum_j \alpha_j \left[\mathbb{P}(S|X_0 = j) - \mathbb{P}(S \backslash A_i | X_0 = j)\right]$$

- $\mathbb{P}(S|X_0 = j)$ is an absorption probability and is classical to calculate;
- $\mathbb{P}(S \backslash A_i | X_0 = j)$ can be computed by considering a new Markov chain in which state i is treated as absorbing, and by determining the probability of reaching state S before reaching state i

Denominator:

$$\mathbb{P}(S) = \sum_j \alpha_j \mathbb{P}(S|X_0 = j)$$

Markov Chain Result

In summary: the **probability that player i is involved in a team possession that ended in a score** is

$$q_i = \mathbb{P}(A_i | S)$$
$$= \boxed{\frac{\sum_j \alpha_j \left[\mathbb{P}(S|X_0 = j) - \mathbb{P}(S \backslash A_i | X_0 = j)\right]}{\sum_j \alpha_j \mathbb{P}(S|X_0 = j)}}$$

Note: the initial distribution $\alpha_j = \mathbb{P}(X_0 = j)$ appears in both the numerator and denominator

- Provides weight according to the probability that player j begins a team possession.
- Important for defensive players!

Illustrative Example: Futsal Passing Matrix

Consider a simplified example of **Futsal**, a five-player team game similar to European football.

- Players: **A, B, C, D, E**
- Augmented passing matrix P:

$$P = \left[\begin{array}{ccccccc} \multicolumn{5}{c}{\left(\begin{array}{ccccc} 0 & 0.40 & 0.25 & 0.20 & 0.10 \\ 0.15 & 0 & 0.34 & 0.25 & 0.10 \\ 0.05 & 0.15 & 0 & 0.20 & 0.30 \\ 0.05 & 0.15 & 0.20 & 0 & 0.20 \\ 0 & 0.05 & 0.25 & 0.25 & 0 \end{array}\right)} & \begin{array}{c} 0 \\ 0.01 \\ 0.05 \\ 0.10 \\ 0.15 \end{array} & \begin{array}{c} 0.05 \\ 0.15 \\ 0.25 \\ 0.20 \\ 0.30 \end{array} \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

Initial possession distribution:

$$\alpha = (0.35, 0.26, 0.17, 0.17, 0.05).$$

Metric Calculation

$$P = \left[\begin{array}{ccccc|cc} 0 & 0.40 & 0.25 & 0.20 & 0.10 & 0 & 0.05 \\ 0.15 & 0 & 0.34 & 0.25 & 0.10 & 0.01 & 0.15 \\ 0.05 & 0.15 & 0 & 0.20 & 0.30 & 0.05 & 0.25 \\ 0.05 & 0.15 & 0.20 & 0 & 0.20 & 0.10 & 0.20 \\ 0 & 0.05 & 0.25 & 0.25 & 0 & 0.15 & 0.30 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

$$\alpha = (0.35, 0.26, 0.17, 0.17, 0.05).$$

- **Results**:

Metric	q_j	$\pi_j = q_j / \sum_i q_i$
Player A	0.444	0.159
Player B	0.518	0.185
Player C	0.613	0.219
Player D	**0.639**	**0.229**
Player E	0.579	0.207

Alternative Scenarios

Consider two additional scenarios:

- **Scenario (ii)**: Increase $p_{5,S}$ from 0.15 to 0.25; decrease $p_{5,U}$ from 0.30 to 0.20.
  - Player **E** becomes more successful at scoring.
- **Scenario (iii)**: Change α to $(0.30, 0.45, 0.10, 0.10, 0.05)$.
  - Player **B** begins possessions more often.

Metric	q_j			$\pi_j = q_j / \sum_i q_i$		
Scenario	(i)	(ii)	(iii)	(i)	(ii)	(iii)
Player A	0.444	0.447	0.430	0.159	0.157	0.149
Player B	0.518	0.515	**0.644**	0.185	0.181	**0.224**
Player C	0.613	0.611	0.604	0.219	0.215	0.209
Player D	**0.639**	0.604	0.618	**0.229**	0.212	0.214
Player E	0.579	**0.669**	0.587	0.207	**0.235**	0.204

Toward Valuation Models (I)

- For a game at time t, populate the i, j entry of matrix with $n_{i,j,t}$, the number of passes from player i to j.
- $n_{i,S,t}$ is a convex combination of shots made and missed:

$$n_{i,S,t} = \frac{5}{6} n_{i,\text{score},t} + \frac{1}{6} n_{i,\text{miss},t}, \tag{1}$$

  where $n_{i,\text{score},t}$ and $n_{i,\text{miss}}$ are the number of times player i scored or missed at time t.
- $n_{i,U,t}$ the number of missed passes by player i.
- Let $\alpha_t \in \mathbb{R}^M$ where $\alpha_{i,t}$ is the proportion of times player i began possession at time t.

Toward Valuation Models (II)

- The resulting frequency data matrix is row-normalized to produce a probability transition matrix giving P_t and $q_{i,t}$ for each i and t
- To apply our valuation methods, $0 \leq \pi \leq 1$ needs to carry performance across games by using a 6-game moving average[5]:

$$\pi_{j,t} = \frac{1}{6}\left(\frac{q_{j,t}}{\sum_i q_{i,t}} + \frac{q_{j,t-1}}{\sum_i q_{i,t-1}} + \cdots + \frac{q_{j,t-5}}{\sum_i q_{i,t-5}}\right).$$

[5] following Opta

Valuation Model

For flexibility and consistency with financial mathematics methods, we model[6] $\pi_t := \pi_{j,t}$[7] as a continuous time process

$$\boxed{d\pi_t = -\theta(\pi_t - \pi^*)dt + \sigma_\pi \sqrt{\pi_t(1-\pi_t)}dW_t^\pi} \quad (2)$$

- W^π is a standard Brownian motion.
- $\sigma_\pi, \theta > 0$, $0 \leq \pi^* \leq 1$, and $\min(\pi^*, 1-\pi^*) \geq \sigma_\pi^2/(2\theta)$;
- π^* is the stationary mean;
- θ controls the rate of reversion to π^*;
- σ_π is a volatility parameter that controls the strength of stochastic fluctuations.

[6]Please see the paper for the full valuation model Cohen and Risk (2024) (CR23)

[7]for brevity, we drop the implied j subscript in the consequent text

More on the π_t model

$$\boxed{d\pi_t = -\theta(\pi_t - \pi^*)dt + \sigma_\pi \sqrt{\pi_t(1-\pi_t)}dW_t^\pi}$$

- π_t is a Fisher-Wright diffusion and lies in the family of Pearson diffusion processes (FS08). Resultingly:
  - π_t is guaranteed to be bounded on $(0, 1)$;
  - π_t a stationary (long-term) Beta distribution

$$\mathbb{E}[\pi_t] = \pi^*, \quad \text{var}(\pi_t) = \pi^*(1-\pi^*)\frac{\sigma_\pi^2}{2\theta + \sigma_\pi^2}, \quad \varrho_\pi(h) = \exp(-\theta h),$$

  - $\mathbb{E}[\pi_t]$ and $\text{var}(\pi_t)$ are the *stationary (long-run) mean* and *variance* of π_t
  - $\varrho_\pi(h) = \text{corr}(\pi_t, \pi_{t+h})$ is the *autocorrelation* of π across h time units (BR10; FS08).

Case Study: European Football (EPL)

- Data from English Premier League (EPL), seasons 2018–2023.
- Each season has 38 games.
- **Objective**: Examine dynamics of π_t across players and positions.
  - Obtain historical data[8] ($q_{i,t} \mapsto \pi_{i,t}$)
  - Estimate parameters for each player
  - Note: $\pi_t = 1/11 \approx 0.0909$ is a benchmark for a high performing starter as 11 players are on the field at any given time[9]
- Time measured in years ($t = 0$ at start of 2018 season).

[8] Data collected per game (`https://www.whoscored.com`).

[9] Of course, a team rotates players in a game, so attaining 0.0909 is a remarkable feat

Selected Players and Teams

Liverpool:

- Mohamed Salah (Right Winger): Primary offensive threat.
- Trent Alexander-Arnold (Right-Back): Creative playmaker.
- Virgil van Dijk (Center-Back): Defensive leader.

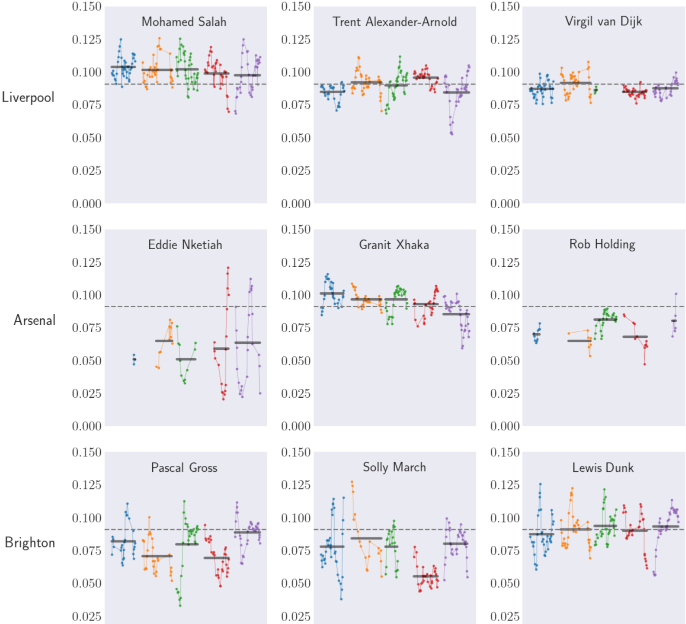
Arsenal:

- Eddie Nketiah (Striker): Young goal poacher.
- Granit Xhaka (Midfielder): Controls game's tempo.
- Rob Holding (Center-Back): Defensive player.

Brighton:

- Pascal Gross (Attacking Midfielder): Creative engine.
- Solly March (Winger/Full-Back): Versatile workhorse.
- Lewis Dunk (Center-Back): Defensive cornerstone.

Historical Data



Estimation Results

Using maximum likelihood estimation[10]

Team Name	Player Name	$\hat{\pi}^*$	$\hat{\theta}$	$\hat{\sigma}_\pi$
Liverpool	Salah	0.101	9.545	0.206
Liverpool	Alexander-Arnold	0.090	3.595	0.140
Liverpool	van Dijk	0.087	11.31	0.121
Arsenal	Nketiah	0.046	3.992	0.392
Arsenal	Xhaka	0.093	3.992	0.132
Arsenal	Holding	0.077	7.271	0.176
Brighton	Gross	0.077	4.914	0.211
Brighton	March	0.072	4.704	0.270
Brighton	Dunk	0.090	5.780	0.208

[10] using Kessler density estimate

Results and Observations

- **Long-term Performance Share** ($\hat{\pi}^*$):
  - Salah: Consistently above benchmark ($\hat{\pi}^* = 0.101$).
  - Xhaka: Slightly above benchmark ($\hat{\pi}^* = 0.093$).
  - Gross and March: Below benchmark ($\hat{\pi}^* = 0.077, 0.072$).
- **Non-starters**:
  - Nketiah ($\hat{\pi}^* = 0.046$), Holding ($\hat{\pi}^* = 0.077$).
  - Lower π^* as expected; Nketiah shows potential as a starter.
- **Defensive Value**:
  - van Dijk ($\hat{\pi}^* = 0.087$), Dunk ($\hat{\pi}^* = 0.09$): Averages near benchmark, reflecting strong defensive impact.

Variability and Consistency

- $(\hat{\sigma}_\pi)$ *(**Variability in** π_t)*:
  - Salah, Nketiah, Gross: High variability ($\hat{\sigma}_\pi = 0.206$, 0.392, 0.176).
  - Reflects offensive roles with fluctuating involvement.
- **Consistency**:
  - van Dijk: Low variability ($\hat{\sigma}_\pi = 0.121$).
  - Xhaka, Alexander-Arnold: Next most consistent $(0.132, 0.140)$.
- **Brighton Players**:
  - Higher variability, possibly due to more polarized match-ups.

Mean Reversion and Autocorrelation

- **Mean Reversion** ($\hat{\theta}$):
  - van Dijk: Higher $\hat{\theta} = 11.31$, quickly returns to mean.
  - Xhaka, Alexander-Arnold: Lower $\hat{\theta}$ (3.992, 3.595), more "streaky".
- **Autocorrelation**: $\widehat{\rho}_\pi(\frac{h}{52}) = \exp(-\hat{\theta}\frac{h}{52})$
- Weekly autocorrelation ($h = 1$ week):
  - van Dijk: 0.805
  - Xhaka: 0.926
  - Alexander-Arnold: 0.933
- At 10 weeks ($h = 10$):
  - van Dijk: 0.114
  - Xhaka: 0.464
  - Alexander-Arnold: 0.500
- **Interpretation**:
  - van Dijk "forgets" highs/lows quickly.
  - Xhaka, Alexander-Arnold exhibit persistent performance trends.

Takeaways

Key Takeaways:

- Introduced a novel framework using Markov chains and augmented passing matrices to quantify player performance.
- Improves over existing measures that focus on marginal post-game statistics
- Computed player involvement probabilities providing a dynamic and individualized measure of influence.
- Captured both offensive and defensive contributions effectively.
- Empirical analysis with Liverpool, Arsenal, and Brighton demonstrated practical applicability.
- Methodology can extend to other leagues and sports where player impact is quantifiable through data matrices.

Future Work

Future Work:

- Adjust model for players with inconsistent playing time.
- Refine approach to account for different player roles, especially goalkeepers.
- Extend model to include metrics specific to specialized positions.
- Model interactions through multiple correlated stochastic processes to capture collective team impact.
- Include opposing teams into the network for more comprehensive game analysis
- Extend to other leagues and sports where player impact is hard to quantify through post game stats and game dynamics can be measured through transition networks.

References

[BR10] J Bakosi and JR Ristorcelli, *Exploring the beta distribution in variable-density turbulent mixing*, Journal of Turbulence (2010), no. 11, N37.

[CR23] Albert Cohen and Jimmy Risk, *European football player valuation: Integrating financial models and network theory*, arXiv preprint arXiv:2312.16179 (2023).

[DWA10] Jordi Duch, Joshua S Waitzman, and Luís A Nunes Amaral, *Quantifying the performance of individual players in a team activity*, PloS one **5** (2010), no. 6, e10937.

[FS08] Julie Lyng Forman and Michael Sørensen, *The pearson diffusions: A class of statistically tractable diffusion processes*, Scandinavian Journal of Statistics **35** (2008), no. 3, 438–465.

[PT12] Javier López Pena and Hugo Touchette, *A network theory analysis of football strategies*, arXiv preprint arXiv:1206.6904 (2012).