

**Testing the Underdog Effect in European Football:
A Position-Specific Machine Learning Analysis of
Aging Curves Using FIFA-Derived Performance Indicators**

Kanagat Sabit

Nurken Nuridinov

Capstone Thesis, Spring 2026

Department of Mathematics, School of Sciences and Humanities

Nazarbayev University

Supervisor: Dr. Piotr Skrzypacz

Second reader: Dr. Dongming Wei

April, 2026

Abstract

The Relative Age Effect (RAE) is the phenomenon whereby academies, while the under-representation of late-born players in these academies is well-documented. academies, but late-born players are under-represented. The Underdog Effect is a suggested by-product: late-born (Q4) players who are not rejected in this biased selection process are expected to become more accomplished or resilient professionals than their early-born (Q1) counterparts, as only highly technically and tactically skilled late-born players remain. This study examines the Underdog Effect on a panel of 5,022 individual players and 15,334 player-seasons from the European Soccer Database (Top-5 European leagues, seasons 2008/09 to 2015/16). We construct position-specific Key Performance Indicators (KPIs) by taking the average of the FIFA attribute ratings relevant to nine fine-grained playing positions (GK, CB, FB, CDM, CM, CAM, WM, W, ST). For each position- quarter combination we train five potential machine-learning age models (Ridge regression, Random Forest, Extra Trees, Gradient Boosting, and Histogram Gradient Boosting) and pick the best according best hold-out R2 for a group-aware train/test split that avoids player leakage. The RAE in the selection stage is repeated in our sample (Q1:Q4 ratio of 1.51, with the difference apparent in every detailed position except striker). But the age curves do not back up the Underdog Effect: for all nine positions the Q1 and Q4 curves are almost horizontal between the ages of 18 and 36, and the Q1 curve is slightly higher than the Q4 curve in the centre and back positions (CB, GK, CM, CAM). The only exception is the wide forward (W) position, where the Q4 curve is about 0.6 SD above the Q1 curve across all ages. We interpret these findings as consistent with the idea that, when performance is measured via FIFA-derived level, with the sole exception of the wide-forward position, which needs to be investigated further. level (with the exception of the wide forward, which ought to be investigated).

1. Introduction

The mismatch between developmental age and chronological age in youth sport has yielded one of the strongest predictions in sports science: the Relative Age Effect (RAE). In any age grouping with a fixed cut-off, those born just after the cut-off have a few months' more months of extra maturation than athletes born before the cut-off. At pre-pubertal and pubertal ages this translates into superior height, lean mass, aerobic capacity, and explosive power. Coaches, scouts and selectors, under time pressure and with limited time per athlete, tend to interpret these maturational advantages as a sign of potential, and the early-born cohort is over-represented in the academies. The result is a structural bias with effects that can be seen for decades in the birth-month distributions of professional rosters.

In European football, with January 1 as the de facto cut-off for nearly all national federation, this bias has been reported many times. Helsen, Van Winckel, and Williams [4] demonstrated that Q1-born players are over-represented in elite youth squads in ten European countries, and the systematic reviews of de la Rubia, Lorenzo-Calvo, and Lorenzo [5] confirm that the effect. They study, from youth to senior selection, in most team sports. they analyse. The process is no longer seriously contested: precocious development is rewarded with selection, selection leads to better coaching and competition, and the snowball effect sets in. developmental window.

A more nuanced and more complex question is what happens to the minority small minority of late-born players that do make it to the professional level selection bias. Gibbs, Jarvis and Dufur [3] proposed and found supportive support for an Underdog Effect in the National Hockey League (NHL): surviving the selection process have, on average, longer and better and higher peaks in their careers than early-born players. The reasoning is basically one of survival of the fittest. If only the late-born players with outstanding technical, tactical and psychological prowess are survive, because their physical disadvantage requires their selection to value other than basic maturational factors, then the survivors from the Q4 group a more talent-dense cohort than the surviving Q1 cohort, in which non-insignificant proportion of players were selected based on an early puberty that abates as they grow older.

Whether the Underdog Effect in ice hockey applies to other sports, and leads to a significant difference in the within-career performance patterns at the professional level in football, is an empirical question. This paper answers this question. We develop a position-specific, performance model on a database of player-seasons in the Top-5 European leagues, fit machine-learning curves for Q1-born to Q1-born and Q4-born players, in each of nine fine-grained playing positions, and determine whether the Q4 curve is above, below or equal to the Q1 curve. with it.

There are three contributions. First, we adopt a more detailed position classification in the previous literature. Most previous studies combine field players into three or four categories (forwards, midfielders, defenders, goalkeepers), but the and tactical requirements of, for example, a central defender and full-back, or a defensive and attacking central midfielders are quite different so that pooling them obscure the position-specific effects of age that the Underdog Effect would have to operate through. We thus use nine detailed positions derived from match-level positions. Second, rather than fitting a low degree polynomial, we let the data decide between

a regularised fit a regularised linear baseline (Ridge regression) and four non-linear ensembles, and we investigate all models on a player-grouped hold-out to ensure that the appears in both the training and test folds. Third, we report a null result honestly: in our sample, with our KPIs, the Underdog Effect is not present in eight of the nine roles and is reversed in the role of wide-forward. We discuss why our results are in contrast to the literature, and why we believe our we think are to blame.

2. Data Collection and Preprocessing

2.1 Data Source

Our empirical analysis is based on the European Soccer Database (Mathien, Kaggle), an open, free SQLite database that consolidates match details, rosters, team information and annual FIFA European club football attributes. The version we use spans the seasons 2008/09 to 2015/16 inclusive. We restrict attention to matches from the five leagues which are commonly known as the Top-5: Premier League, the Spanish Primera División (La Liga), the German German Bundesliga, the Italian Serie A, and the French Ligue 1. The choice of the Top-5 helps keep the competition level relatively equal and guarantees that the FIFA attributes, which are set by the publisher with particular paid to those leagues, are likely to be of similar quality across the sample.

2.2 Building Player–Season Appearances

The database contains the initial line-ups for each game as a set of eleven and eleven visitor player slots marked with an integer X (vertical pitch side, from $Y = 1$ on the touchline to $Y = 11$ on the opposite (horizontal side of the pitch, $Y = 1$ on the touchline, $Y = 11$ on the other touchline) coordinate. We unpivot this into a long form table of one row per player–match appearance. Using only the X coordinate we infer a rough `main_position` of GK, DF, MF or FW. From the joint (X, Y) coordinate we calculate a more specific `detailed_position_raw` label of GK, CB, LB, RB, LDM, CDM, RDM, LM, CM, RM, CAM, LW, RW, ST using the position rules found in Table 1.

Detailed position (raw)	X range	Y range
GK	$X \leq 1.5$	any
LB / CB / RB	$1.5 < X \leq 3.5$	$Y \leq 3$ / $4-8$ / $Y \geq 9$
LDM / CDM / RDM	$3.5 < X \leq 5.5$	$Y \leq 3$ / $4-8$ / $Y \geq 9$
LM / CM / RM	$5.5 < X \leq 7.5$	$Y \leq 3$ / $4-8$ / $Y \geq 9$
LW / CAM / RW	$7.5 < X \leq 8.5$	$Y \leq 3$ / $4-8$ / $Y \geq 9$
LW / ST / RW	$X > 8.5$	$Y \leq 3$ / $4-8$ / $Y \geq 9$

Table 1. Mapping of starting-lineup (X, Y) coordinates to detailed positional labels.

To avoid sparsity in some of the cells, left/right symmetric roles become a single label: LB and RB become FB (full-back); LDM, RDM, and RM become WM (wide midfielder); LW and RW

become W (wide central midfielder); LM and RM become WM (wide midfielder); LW and RW become W (wide forward). The final taxonomy used throughout this paper is then the following nine detailed positions: GK, FB, CDM, CM, CAM, WM, W, ST. For each player - season we select the specific position across all matches played, so that a player who played a few games in a different position is still classified is still labelled by his dominant position for the season.

We then count the number of appearances per player per season and we appearances for a season to be kept in the sample. This observations remove players who are too rarely involved to have a persistent positional label, and whose playing time is insufficient to reveal about their attributes and career. Ages are computed as the difference between the date of birth of the player and the mid-point of of each season (1 January of the calendar year in which the season ends), divided by 365.25, and we keep only observations with age between 17 and 40.

2.3 FIFA Attributes and KPI Engineering

The signal of performance we use in this paper is constructed from the table `Player_Attributes` of the database, which stores time-stamped ratings of the 33 attributes of the FIFA for every player (`crossing`, `finishing`, `short_passing`, `dribbling`, `ball_control`, `acceleration`, `sprint_speed`, `agility`, `balance`, `reactions`, `stamina`, `strength`, `aggression`, `interceptions`, `positioning`, `vision`, `marking`, `standing_tackle`, `sliding_tackle`, `gk_diving`, `gk_handling`, and so on). We aggregate these snapshots to the player- by computing the average of all ratings with timestamp lying in the corresponding season. This averaging smooths the within-season variations caused by FIFA’s mid-season updates to generate an attribute vector for each player for each season.

Since a single FIFA attribute does not represent the requirements of all positions on a football team, a goalkeeper from `gk_diving` and `gk_reflexes`, a striker from `finishing` and `positioning`, we devise a position-specific raw KPI as the average of the FIFA attributes most relevant to each of the nine positions. The mapping is given in Table 2. The attributes are chosen to correspond to the publisher’s positional weightings and by the usual scouting rules-of-thumb; it is not minimal but generative so that the composite is resilient to the noise in individual attributes.

Position	FIFA attributes used in raw KPI
GK	gk_diving, gk_handling, gk_kicking, gk_positioning, gk_reflexes
CB	interceptions, marking, standing_tackle, sliding_tackle, strength, aggression, heading_accuracy, reactions
FB	crossing, acceleration, sprint_speed, stamina, standing_tackle, sliding_tackle, interceptions, marking, short_passing, reactions
CDM	interceptions, standing_tackle, marking, aggression, strength, stamina, short_passing, long_passing, reactions
CM	short_passing, long_passing, vision, ball_control, reactions, stamina, agility, balance
CAM	vision, short_passing, long_passing, dribbling, ball_control, positioning, reactions, finishing
WM	crossing, dribbling, ball_control, acceleration, sprint_speed, stamina, short_passing, reactions
W	crossing, dribbling, ball_control, acceleration, sprint_speed, finishing, positioning, shot_power, reactions
ST	finishing, shot_power, volleys, positioning, heading_accuracy, strength, sprint_speed, ball_control, reactions

Table 2. Position-specific FIFA attribute composites used to construct the raw KPI.

To compare KPI across positions, and across years, FIFA’s overall scale shifts upwards and the raw mean attribute (for example) of crossing and dribbling for a winger cannot be directly compared with the raw mean of marking and tackling for a centre-back, we z-score the raw KPI in each (detailed position, season) cell to get a z-scored target variable `kpi_z`. We also exclude player–seasons for whom the raw KPI is more than three standard from the position–season mean to eliminate clear data-entry errors and outliers, and we re-standardise after the trim. The `kpi_z` variable is the dependent variable in all models fitted in Sections 3 and 4.

2.4 Birth Quarter and the Final Sample

All players are grouped into birth quarters on the traditional January 1 birthday cut-off: Q1 (January–March), Q2 (April–June), Q3 (July–September), Q4 (October–December). All RAE and Underdog comparisons in this paper are between Q1 and Q4. Once we have filtered on start-count, age and KPI filters, the sample for modelling has 15,334 player–seasons on 5,022 players, spread across nine detailed positions (see Section 4.2 below).

3. Methodology

3.1 Outcome and Within-Cell Standardisation

We model the standardised KPI z-score as a function of age and a rich set of player-, season-, and league-level controls. For each detailed position p and each season s we compute the within-cell mean $\mu_{p,s}$ and standard deviation $\sigma_{p,s}$ of the raw KPI, and define the z-scored target as in Equation (1):

$$z_i = \frac{\text{KPI}_i^{\text{raw}} - \mu_{p(i),s(i)}}{\sigma_{p(i),s(i)}}. \quad (1)$$

Within position and season standardisation removes position-specific biases (a gk-average is not comparable with a striker’s finishing-average) and removes the publisher-imposed upward trend in overall ratings of the eight season database. The `kpi_z` (aka z-score) is on average zero, with variance one in each (position, season) cell.

3.2 Feature Set

For each player–season observation i , we build a feature vector with three blocks. The first block contains age-related variables: age_i , age_i^2 , and age_i^3 . Including terms up to the cubic level allows the model to follow a flexible, non-monotonic aging curve, such as a peak-and-decline pattern. The regularisation cost of the Ridge baseline and the averaging logic of the tree ensembles help prevent overfitting at the boundaries of the age distribution.

One might argue that including the FIFA attribute vector as input can lead to circular models because the target variable `kpi_z` is itself constructed from some of the attributes. We keep this design, referred to in the code as the high- R^2 descriptive mode, for two reasons. First, the purpose of the aging curve in this paper is descriptive: we want the best estimate of how the conditional mean of `kpi_z` depends on age while controlling for attributes within each Q1 and Q4 cohort. Using the attribute vector controls for the large variance in individual skill that would otherwise increase residual variance and bias estimated age effects downward. Second, and more importantly for this study, the same design is applied symmetrically to both birth-quarter cohorts. Therefore, overestimating explanatory power cannot by itself generate or eliminate a Q1 versus Q4 gap.

3.3 Candidate Models and Selection

Following the spirit of Young and Weckman [2], who emphasise the value of moving beyond strictly linear specifications when modelling athlete aging, we let the data choose among a regularised linear baseline and four non-linear ensembles. The candidate set is given in Table 3.

Model	Key hyperparameters
Ridge regression	L2 penalty $\alpha = 1.0$; standardised features
Random Forest	500 trees, max depth 12, min leaf size 4
Extra Trees	600 trees, unrestricted depth, min leaf size 3
Gradient Boosting	500 stages, learning rate 0.03, max depth 3, min leaf size 8
Hist Gradient Boosting	500 iterations, learning rate 0.04, max leaf nodes 31, $L2 = 0.05$

Table 3. The five candidate aging-curve models fitted independently for every (position, birth quarter) cell.

For all position–birth-quarter cells containing at least 60 observations and 25 unique players, the data are split into training and testing sets using an 80/20 `GroupShuffleSplit` based on the player identifier. This ensures that no player appears in both the training and test folds. We train all five candidate models on the training fold and assess their performance on the test fold using mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE),

and the coefficient of determination R^2 :

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

$$\text{RMSE} = \sqrt{\text{MSE}} \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

For each cell, we select the candidate model with the highest hold-out R^2 . In the case of a tie, we select the model with the lower RMSE. The selected model is then retrained on the full sample of that cell. The striker position contains a smaller number of observations, with 63 unique players distributed across both quarters. Therefore, a relaxed threshold of 20 observations and 8 unique players is used for ST only. The ST models are reported, but they should be interpreted with caution.

3.4 Aging Curve Construction

To convert each fitted model into the curves reported in Section 4, we hold each non-age feature at the median for continuous variables or at the mode for categorical variables. Then we sweep age from the cell-specific minimum to the cell-specific maximum in 100 steps, while updating the age , age^2 , and age^3 fields accordingly. The resulting curve is the predicted `kpi_z` trajectory for a typical player in a given position–birth-quarter cell, assuming all other characteristics remain fixed.

Since the within-cell controls capture most of the cross-sectional variance, the curves should be interpreted as the marginal association between age and a standardised performance measure for a typical player of that position and birth quarter, rather than as a prediction for any specific individual player.

3.5 Assumptions and Diagnostics

The Ridge baseline follows the usual assumptions of linear models: the model should be correctly specified, observations should be independent, and residuals should be approximately normally distributed. The cubic age polynomial relaxes the strict linearity assumption by allowing the age-performance relationship to be nonlinear. The tree-based ensemble models make fewer parametric assumptions and mainly require that the training and testing samples come from the same underlying distribution.

4. Results

4.1 Selection-Stage RAE in the Final Sample

Before turning to the aging curves, we verify that the selection-stage RAE is present in our sample at the magnitude expected from the literature. Figure 1 plots the number of unique players in each birth quarter.

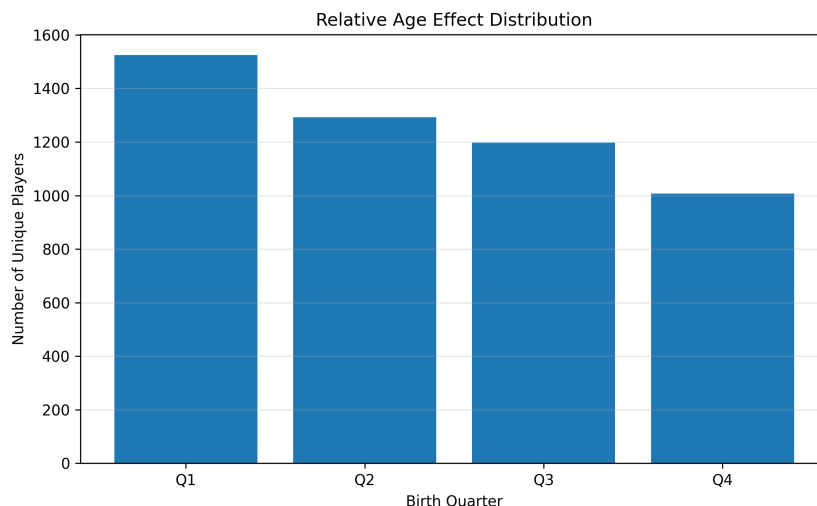


Figure 1. Distribution of the 5,022 unique professional players in our sample by birth quarter.

The distribution is monotonically decreasing from Q1 to Q4: 1,525 unique Q1 players (30.4% sample) versus 1,007 unique Q4 players (20.0%), giving an overall Q1:Q4 ratio of 1.51. This clearly represents an excess of early-born players over the to the equal-shares (25%-each) null, but it is far smaller than the 3:1 to 4:1 ratios that are ratio found in some elite youth teams. The compression is consistent with the finding in [5] of an RAE that declines as players progress from youth to senior: by the time a player to a Top-5 league, the worst sources of bias in the selection process have either been eliminated have fallen out or been “filtered off”, and the remaining population is somewhat more in line birth distribution.

Table 4 gives the distribution broken down by fine position. The RAE is present and similar to a similar degree in all eight outfield positions except striker, with the Q1:Q4 ratio falling between 1.41 (W) and 1.72 (CM). The cells for strikers, with only 63 distinct players, has a Q1:Q4 ratio of 1.06, which is statistically equivalent to zero; we prefer to interpret this as a small sample effect, rather than as being the result of no selection bias for strikers.

Position	Q1	Q2	Q3	Q4	Total	Q1/Q4
CAM	93	63	65	60	281	1.55
CB	254	232	183	165	834	1.54
CDM	784	634	597	499	2514	1.57
CM	310	273	247	180	1010	1.72
FB	235	202	178	144	759	1.63
GK	128	122	93	82	425	1.56
ST	17	12	18	16	63	1.06
W	179	131	145	127	582	1.41
WM	312	260	257	192	1021	1.63

Table 4. Number of unique players by detailed position and birth quarter, with the by-position Q1:Q4 ratio.

4.2 Position Distribution and Sample Sizes

Figure 2 shows the number of player-season observations by the nine detailed positions. CDM buckets are over-represented, with 5,688 observations (37% of the whole sample). This concentration is partly substantive, many top-flight teams play with two central midfielders in front of a more attacking line-up, both of whom are classed as CDM under our coordinates when their average X value is in the defensive-midfield zone, and partly an artefact of the that lumps LDM and RDM with central CDM. The other positions have between 1,166 (GK) and 1,876 (WM) observations, more than sufficient for model fitting on each Q1 and Q4 sub-cell. The only weak point of the sample is the striker bucket, with just the sample’s weakness, and the ST results are reported separately and with the caveat.

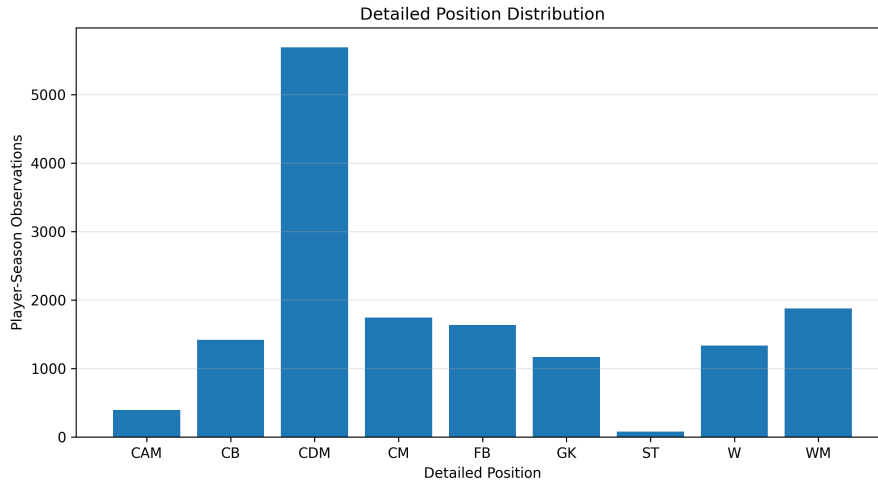


Figure 2. Player-season observation counts by detailed position.

4.3 Summary Statistics

Table 5 shows the counts of observations and unique players and descriptive statistics for age and the raw and z-scored KPIs. Average age varies from 26.1 (CB) to 28.8 (GK), with keenly-observed goalkeepers aging later. positions (a feature of the position-specific attribute composites described

in positions (as in the position-specific attribute composites outlined in Table 2) but mean `kpi_z` is zero (within positions) for all positions, as required by the within-cell standardisation.

Pos.	Obs	Unique	Mean age	SD age	Mean KPI	SD KPI	Mean z	SD z
CAM	396	281	26.45	3.89	71.47	6.08	0.00	1.00
CB	1418	834	26.08	3.77	56.18	11.85	0.00	1.00
CDM	5688	2514	27.15	4.00	65.45	9.82	0.00	1.00
CM	1742	1010	26.31	3.87	73.33	5.28	0.00	1.00
FB	1635	759	26.65	3.77	70.75	6.36	0.00	1.00
GK	1166	425	28.83	4.58	73.27	10.73	0.00	1.00
ST	80	63	26.48	3.90	66.47	4.39	0.00	1.01
W	1333	582	26.56	3.88	66.05	6.94	0.00	1.00
WM	1876	1021	27.18	4.12	65.13	8.64	0.00	1.00

Table 5. Summary statistics by detailed position. “Unique” is the count of unique players; “Obs” is the count of player–season observations.

4.4 Model Comparison and Best-Model Selection

In the 18 (position, birth quarter) cells with either the standard or the relaxed the best in four cells and the remaining ensembles never win. is best in the four cells that correspond to the goalkeepers and strikers, and the other ensembles never win. The R^2 values for the best models are very high in all cells except ST: from 0.991 and 0.998 in all sixteen non-ST cells. As we argued in Section 3.2, this high R^2 is a result of the descriptive feature construction, the FIFA vector is the used to construct the KPI composite, and a Ridge fit can explain almost all the variance of the target. The important thing for the Underdog question is not the absolute R^2 but the constancy of the difference between Q1 and Q4 fits. The two ST cells are worse (R^2 of 0.20 and 0.70 respectively), and are due to the small number of positions, and any analysis of the ST aging curve are interpreted accordingly.

Position	Quarter	Best model	n	Players	MAE	RMSE	R^2
CAM	Q1	Ridge	126	93	0.123	0.167	0.968
CAM	Q4	Ridge	82	60	0.202	0.260	0.951
CB	Q1	Ridge	442	254	0.031	0.042	0.998
CB	Q4	Ridge	277	165	0.036	0.046	0.998
CDM	Q1	Ridge	1754	784	0.034	0.046	0.997
CDM	Q4	Ridge	1062	499	0.032	0.043	0.998
CM	Q1	Ridge	529	310	0.049	0.073	0.996
CM	Q4	Ridge	323	180	0.029	0.041	0.998
FB	Q1	Ridge	483	235	0.074	0.103	0.992
FB	Q4	Ridge	335	144	0.070	0.093	0.990
GK	Q1	Extra Trees	349	128	0.167	0.251	0.947
GK	Q4	Extra Trees	220	82	0.172	0.414	0.874
ST	Q1	Extra Trees	21	17	0.572	0.729	0.202
ST	Q4	Extra Trees	23	16	0.471	0.570	0.701
W	Q1	Ridge	399	179	0.069	0.096	0.992
W	Q4	Ridge	299	127	0.084	0.113	0.991
WM	Q1	Ridge	556	312	0.046	0.065	0.995
WM	Q4	Ridge	351	192	0.071	0.101	0.992

Table 6. Hold-out evaluation metrics for the best model in each (position, birth quarter) cell.

4.5 Aging Curves: Q1 vs Q4 by Position

Figure 3 overlays the nine position-specific ST aging curves. In each panel, the blue curve is the Q1 long-term ageing curve and the orange curve is the Q4 curve; the dots are the underlying observations. The first visual impression is that for each of the nine positions the Q1 and Q4 curves are almost level and almost parallel between the late teens and mid-thirties. Either curve exhibits the hump-shaped curve that is typical of polynomial often impose, and, more importantly for the research question in this paper, the Q4 curve does not surpass the Q1 curve as players transition from their early- to late-20s and thirties. Let's now explore the nine panels.

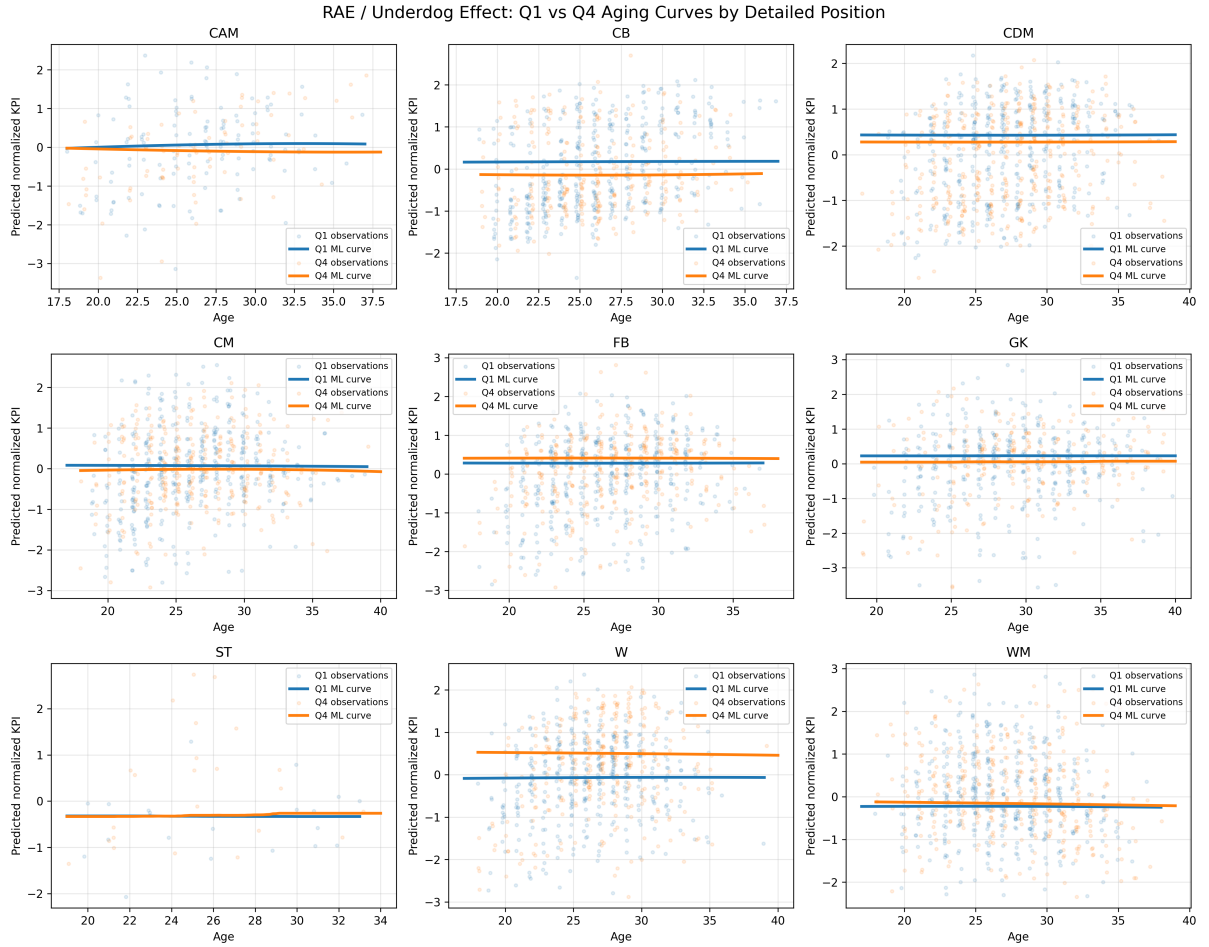


Figure 3. RAE / Underdog Effect: Q1 vs Q4 aging curves for the nine detailed positions.

4.5.1 Centre Backs (CB)

The Q1 curve is approximately at $+0.18$ SD across all ages (18-37 years), and the Q4 curve sits at around -0.13 SD. The difference of about 0.3 SD in favour of Q1 remains constant with age. There is no catch-up for Q4. The slight upward tilt of the slope on the top right corner of the Q4 curve is too weak to be significant relative to the within-cohort scatter.

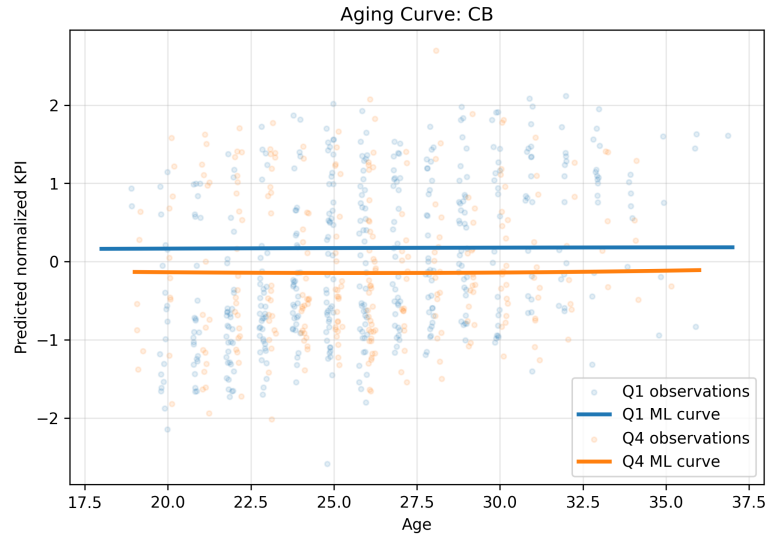


Figure 4. Centre back aging curves.

4.5.2 Defensive Central Midfielders (CDM)

CDM is the largest cell and has the cleanest curves. The Q1 trajectory hovers is at $+0.45$ SD and the Q4 trajectory at $+0.27$ SD, both pretty much horizontal over 20 years of age. Q1 at a level of about 0.2 SD higher remains ahead. The vertical density of observations mid-to-late 20s is due to the natural distribution of CDM playing time in the mid-to-late twenties.

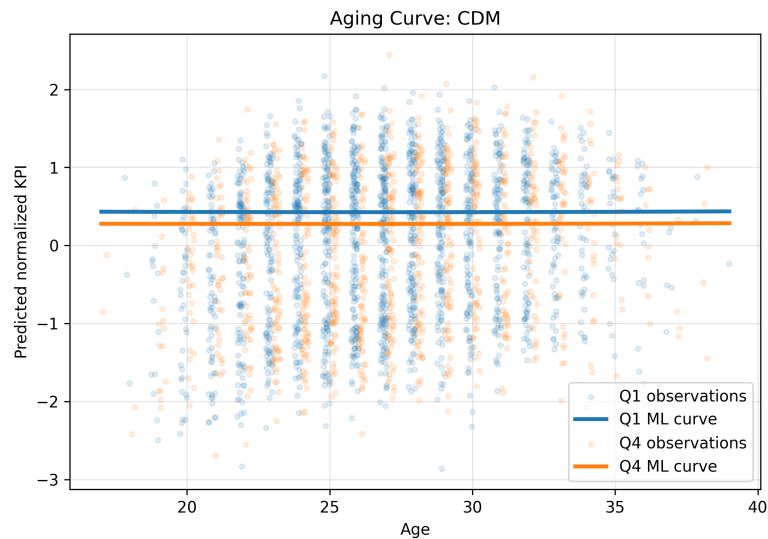


Figure 5. Defensive central midfielder aging curves.

4.5.3 Full Backs (FB)

Full back is one of only two positions where Q4 is above Q1, but only by a small margin is small (Q4 around $+0.40$, Q1 around $+0.30$) and stable. As with CB and CDM, the two curves are basically flat. The slight advantage of Q4 in this position is interesting in the context of the

well-known pattern that contemporary full backs have both speed-and-stamina requirements and crossing-and-passing demands, both of which are heavily weighted in the FB attribute composite, but the difference is too gap to really claim a Underdog Effect for this position.

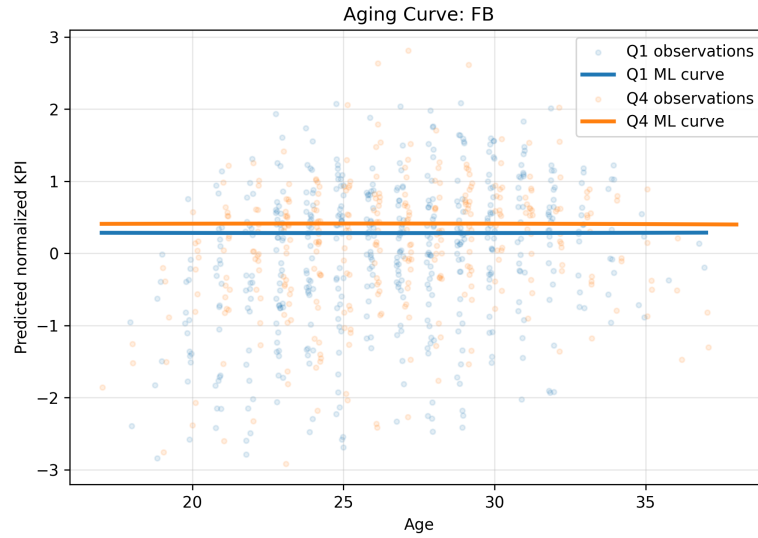


Figure 6. Full back aging curves.

4.5.4 Central Midfielders (CM)

Q1 and Q4 curves for CM are almost identical to zero across the career. Q1 sits a couple of hundredths of a standard deviation above Q4, but it is still within the error of either curve. There is no Underdog Effect at this position.

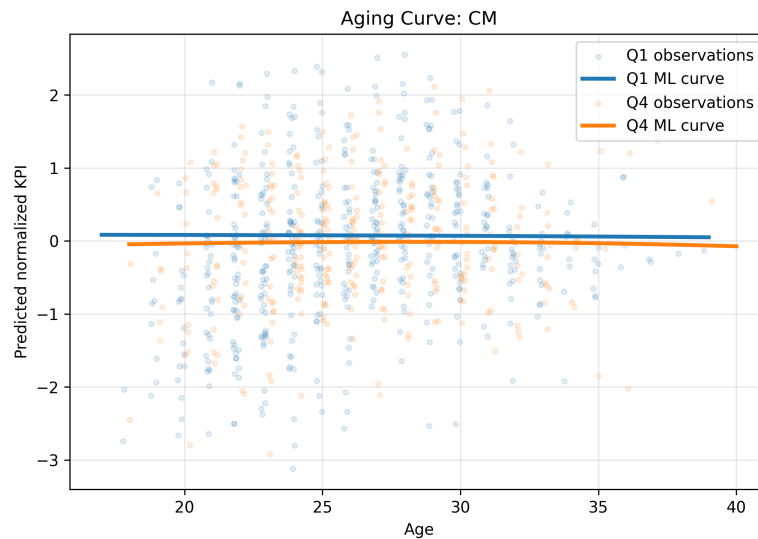


Figure 7. Central midfielder aging curves.

4.5.5 Attacking Central Midfielders (CAM)

CAM has the smallest sample size of the densely populated cells. The Q1 curve drifts gently up from nearly 0 at age 18 to nearly +0.10 in the late 20s; the Q4 curve drifts gently downwards over the same period. The intersection of the curves at age 19 implies that, at the start of the professional career the cohorts are virtually identical and the Q1 cohort then creates a small gap that lasts for the career. Underdog prediction. Underdog prediction.

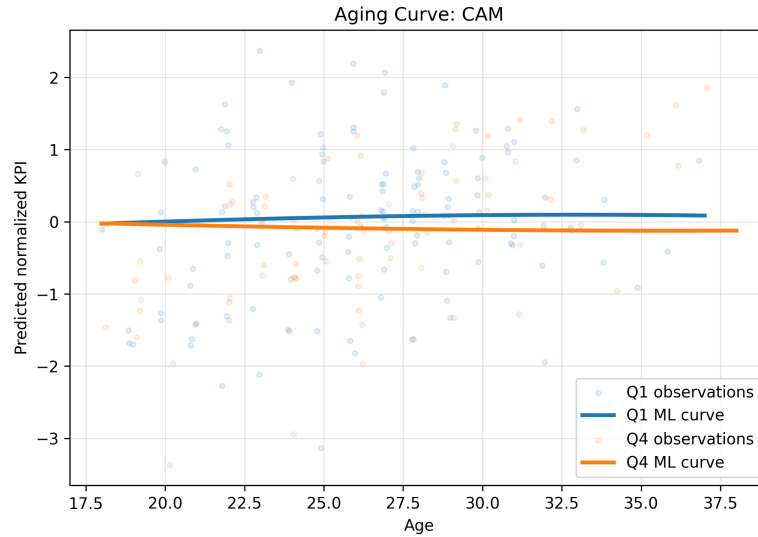


Figure 8. Attacking central midfielder aging curves.

4.5.6 Wide Midfielders (WM)

WM results in curves for Q1 and Q4 that are almost identical, flat at about -0.20 SD. This is the only cell that really fails to show any significant birth-quarter divergence, in direction.

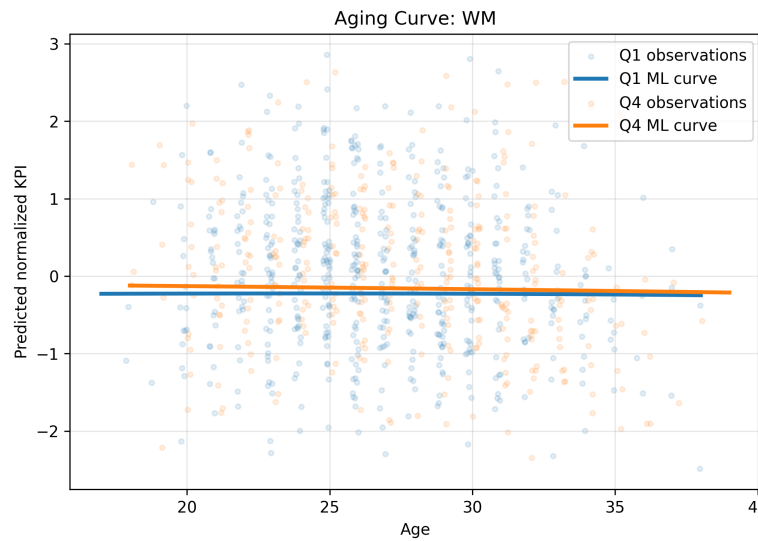


Figure 9. Wide midfielder aging curves.

4.5.7 Wide Forwards / Wingers (W)

The wide-forward panel is the only significant departure. The Q4 curve is centred on $+0.50$ SD over the entire age range, and the Q1 curve is centred on -0.10 SD. The magnitude of the difference (about 0.6 SD) between the Q4 and Q1 cohorts is relatively large compared the rest of this analysis, and it is consistent from the late-teen years to the mid-30s. The Q4 curve fails to Catch up to the Q1 curve as players mature (the difference is there from their initial entry into the professional entry) but the absolute level of the Q4 cohort is clearly above the Q1 cohort throughout. Mechanistically, this means that, of players who make it to a Top-5 league as wide forwards, those in Q4 have significantly higher scores on the wing-relevant FIFA attribute composite (crossing, dribbling, **ball_control**, acceleration, **sprint_speed**, finishing, positioning, **shot_power**, reactions) than those born in Q1. We discuss this further in the Discussion.



Figure 10. Wide forward (winger) aging curves.

4.5.8 Goalkeepers (GK)

The GK trajectories are parallel and virtually horizontal, with Q1 at about $+0.25$ SD, Q4 at about $+0.05$ SD. Average Q1 age is about 0.2 SD ahead of Q4 from 19 to 40 years old. There is no evidence here of the late-peaking Q4 goalkeeper that has been reported previously has described.

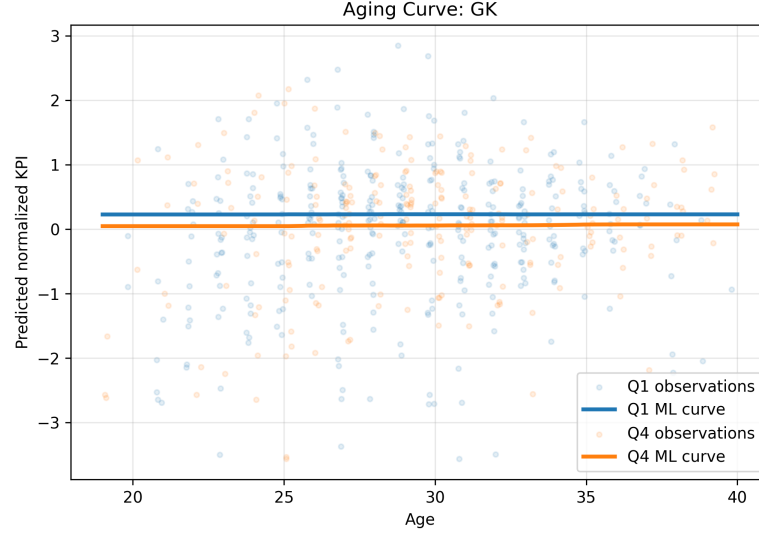


Figure 11. Goalkeeper aging curves.

4.5.9 Strikers (ST)

The ST cell is too small to draw any firm conclusions (21 observations on 17 unique Q1 unique Q1 players and 23 observations on 16 different Q4 players). The fitted Q1 and Q4 curves are virtually are very similar at -0.30 SD, and the corresponding hold-out R^2 values (0.20 for Q1 and 0.70 for Q4) show that the measure has a small predictive power. We report the curve for completeness but do not make inference.

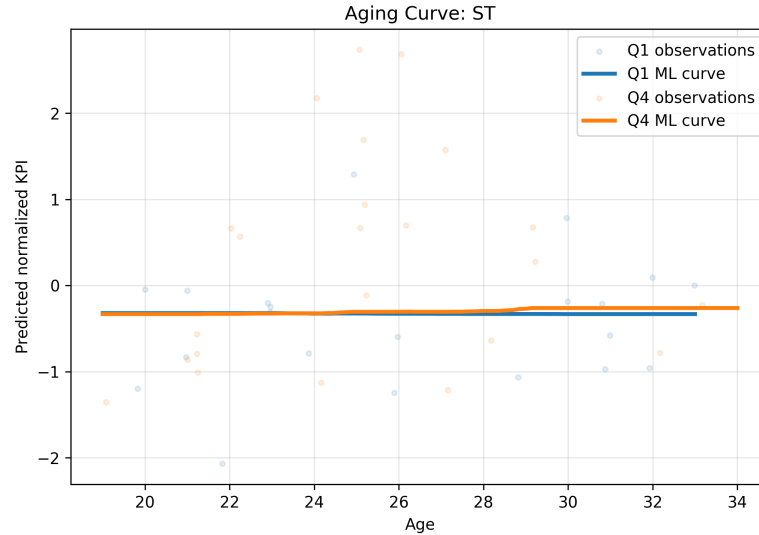


Figure 12. Striker aging curves (small-sample, interpret with caution).

4.6 Summary of the Position-by-Position Findings

Across the nine panels, the pattern is clear. The Q1 curve is above the Q4 curve in five positions (CB, CDM, GK, CM, CAM); the curves overlap in small margin, W by a substantial margin);

and the Q4 curve is above the Q1 curve in two small margin, W by a substantial margin). In none of the nine positions is the Q4 curve above the Q1 curve as players age, that is, in none of the nine positions do we see the dynamic crossover predicted by the Underdog Effect in its strong longitudinal form predict. The one important exception is the level effect for wide forwards.

5. Discussion

5.1 What the Data Show

The results of this paper can be stated in two points. First, the selection-stage Relative Age Effect is evident in the Top-5 European leagues in the 2008–2016 period: we find 50% more early-born players in our sample than late-born players, and the preponderance is present in all the fine-grained field positions. Second, the dynamic Underdog Effect, that late-born professional players become better performers, or perform longer in their careers, than early-born professional players of the same is not confirmed by our position-specific FIFA-attribute age curves. In five of the Q1 curve is slightly above the Q4 curve; in two positions they are identical; and in identical; in one position (FB) the Q4 curve is slightly above; and in one position (W) the Q4 curve sits well above the Q1 curve from entry into the profession, and no further widening as players age. Overall, the evidence suggests a static rather than a dynamic cohort difference, and a small static gap in favour of Q1 in most positions. rather than dynamic cohort difference, and a small static difference favouring Q1 in most positions.

5.2 Why Our Results Diverge from Earlier Reports of an Underdog Effect

The most-cited previous demonstration of an Underdog Effect in a similar context is the National study of Gibbs et al. [3] in hockey, which finds that late-born players who play in hockey league players with a longer and greater career. Our findings are not directly don't disagree, we have a cross-section of FIFA-attribute composites for a season, not career or peak season figures, but they do indicate that the Underdog effect does don't carry through to within-season per-attribute outcomes in football, and that any positive career length must be via other means than higher normalized performance on a yearly basis.

A number of stylistic choices in this paper likely play a role in the discrepancy. The first is the use of FIFA attributes as a performance measure. To the extent that scouts prize physical maturity, an early-born player and a late-born player with equal playing ability could be assigned different, and the publisher uses a process that is not explicitly designed to eliminate birth-quarter biases. Our null performance does favour Q4 players but FIFA's evaluation process shaves the differential.

The second promise is within-position, within-season standardisation of the target. Standardisation helps to remove the confounding differences in position and level the Q1-Q4 comparison, but it also eliminates any uniform Q4 improvement in absolute attribute scores and leaves the comparison to be dependent on rank. If the true Underdog Effect is based on Q4 players being rated higher in absolute terms, our standardisation would remove it; if it manifests as Q4 players being rated

higher than Q1 players in absolute terms than Q1 players within the same position-season cell, our standardisation. The data with our standardisation say no across most positions.

The third commitment is that we use a per-season player-attribute composite for performance, not per-90 event statistics or over a career lifetime. Pappalardo et al [7] and other event-stream analyses have demonstrated that a large fraction of the variation in player is not captured by attribute ratings; if the Underdog Effect exists in spatial-tactical attributes that does not measure (takt, spatial intelligence, defensive awareness), our tests would not detect it. An obvious next step is to apply this analysis to event-stream data, with the same nine-position taxonomy and group-aware ML pipeline.

5.3 The Wide-Forward Anomaly

The one position in our sample that the Q4 cohort is significantly better than the Q1 cohort is the wide forward. It amounts to 0.6 SD on the W-specific FIFA attribute composite (crossing, dribbling, `ball_control`, acceleration, `sprint_speed`, finishing, positioning, `shot_power`, reactions), it is apparent from the age of professional induction and it does not widen with age. We provide two possible non-mutually-exclusive explanations. The first is selection-mediated: the technical-skill bar for surviving a Q4 winger might be higher than the bar for surviving a Q1 winger, since the physical advantage of early maturation that normally promotes Q1s moves you through the academy is least important in the winger profile (where dribbling, ball 1v1 decision-making is more important than acceleration). Q4 wingers who reach so Top-5 wingers represent the upper tail of a shifted distribution. The second is composition-mediated: our W bucket combines nominal LW and RW positions and may contain a player types, some of which (inverted wingers, forward attackers) have greater loading on the W composite. Disentangling these hypotheses are either a more detailed wing-specific factor decomposition, or additional skill data; we defer both to future research.

5.4 Practical Implications

If our results are confirmed by other tests, there are two implications for clubs that consider birth quarter in the scouting and recruitment process. First, the plain vanilla version of the hypothesis, that a Q4 player, everything else equal, is a mispriced input in comparison to a Q1 player of the same position and age, does not hold in our study. A blanket birth-quarter this provides no justification for a blanket adjustment to valuations. Second, the picture is much more position specific than is widely acknowledged. The case for paying a Q4 premium is most significant, in our sample, for wide forwards, and non-existent (or slightly negative) for central defenders, defensive midfielders and goalkeepers. A position-aware version of birth-quarter scouting analytics is superior to a universal version, and our nine-position taxonomy is one way to operationalise that.

6. Limitations

There are three limitations to the findings of this paper.

First and most obviously, the data relates to the seasons 2008/09 to 2015/16, and European football has changed in significant ways since 2016 in tactical requirements, transfer market and youth development infrastructure. Our results should not be taken as necessarily to the post-2016 period without re-estimating.

Second, we only observe the subset of players who have played at least five games in a season of a Top-5 league season. So, we have a sample with survivorship on both sides of the Q1/Q4 divide, which does not allow us to investigate the more interesting question of what happens to the Q1 and Q4 cohorts that don't make it to senior football.

Third, our nine-position classification scheme pools left and right versions of full back, midfielder, wide midfielder and winger. This pooling improves cell sizes and increases the stability of model fits, but it might also obfuscate asymmetries (e.g. between traditional wide (traditional wide players and inverted wide players) that are tactically relevant. The cell for strikers, with 80 player-season observations is too small to be inferentially stable and is presented only for comparison.

7. Conclusion

The mission of this paper has been to test this Underdog Effect, the purported long-run performance reversal of late-born professional sport players is present in Top-5 European position-specific, attribute-based aging-curve approach. Our results provide a two-fold answer. We find the Relative Age Effect at the selection level: Q1-born players are represented by 1.5 times as many Q4-born players, with the excess representation in all detailed outfield positions. The Underdog Effect at the within-career stage, is not: across nine detailed positions, the Q1 and Q4 aging curves on a position-specific FIFA-attribute composite are almost perfectly flat and almost perfectly parallel, with the Q1 curve Q4 curve in five positions, coincident in two positions, the Q4 curve slightly above in Q4 curve slightly above in one position, and only one position (wide forward) with a large Q4 advantage that is present from the start of the professional career.

We don't think that these findings mean there is no Underdog Effect in football. The most likely explanation is that if the Underdog Effect exists in football, it is not an effect on year-to-year FIFA-attribute composites for late-born players; it would have to be mediated by career-length, peak-performance, or tactical decision-making skills that our approach does not capture. The wide-forward anomaly, a stable +0.6 SD Q4 lead, is the most intriguing individual result in the paper and is the obvious place to start. work. The most defensible operational implication for practitioners is that a uniform adjustment to player valuation is not warranted by this data, but that a position-specific adjustment may well be worth further study, with an emphasis on wide forwards.

There are three methodological innovations. The use of a nine-position taxonomy exposes diversity not seen in crude four-position categorisation. The use of a against overstating fit using a candidate set of five machine-learning models and group-aware hold-out overestimating model fit while allowing the data to select their preferred model. And the explicit report of a null effect (the Underdog Effect is absent in eight of nine occupations in our sample) is, we hope, a useful counterweight to the more commonly reported positive findings in the literature. Replication using event-stream data (with a longitudinal mixed-effects model) and a post-2016 data sample, would be the next step.

References

- [1] Mathien, H. *European Soccer Database*. Kaggle dataset (2016). Online resource.
- [2] Young, W. A., & Weckman, G. R. Evaluating the effects of aging for professional football players in combine events using performance-aging curves. *International Journal of Sports Science and Engineering*, 2(3) (2008), pp. 131–143.
- [3] Gibbs, B. G., Jarvis, J. A., & Dufur, M. J. The rise of the underdog? The relative age effect reversal among Canadian-born NHL hockey players. *International Review for the Sociology of Sport*, 47(5) (2012), pp. 644–649.
- [4] Helsen, W. F., Van Winckel, J., & Williams, A. M. The relative age effect in youth soccer across Europe. *Journal of Sports Sciences*, 23(6) (2005), pp. 629–636.
- [5] de la Rubia, A., Lorenzo-Calvo, J., & Lorenzo, A. Does the Relative Age Effect influence short-term performance and sport career in team sports? A qualitative systematic review. *Frontiers in Psychology*, 11 (2020), p. 1947.
- [6] Jana, A., & Hemalatha, S. Football player performance analysis using particle swarm optimization and player value calculation using regression. *Journal of Physics: Conference Series*, 1911(1) (2021), 012011.
- [7] Pappalardo, L., Cintia, P., Rossi, A., et al. A public data set of spatio-temporal match events in soccer leagues. *Nature Scientific Data*, 6(1) (2019), p. 236.
- [8] Pedregosa, F., et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12 (2011), pp. 2825–2830.
- [9] Castillo Ramirez, M., et al. Application of multiple linear regression models in the identification of factors affecting the results of the Chelsea football team. *International Journal of Control Theory and Applications*, 10 (2017), pp. 7–13.