丘成桐中学科学奖（数学奖）研究报告

职业足球球员状态与相对溢价的定义、计算与应用

——基于能力、比赛表现与球员身价关系的样本外研究

作者：张桐宁

学校：北京一零一中学

省份：北京　　国家：中国

指导老师：北京大学 文再文

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

职业足球球员状态与相对溢价的定义、计算与应用

——基于能力、比赛表现与球员身价关系的样本外研究

作者：张桐宁　　北京一零一中学

摘要

职业足球中，“状态好”和“身价偏高”是评价球员时常见的判断，但二者都缺少统一的量化基准。本文基于 83,527 条球员—赛季比赛表现记录，并最终匹配得到 8,316 名职业球员样本，在条件期望框架下分析球员能力、比赛表现与身价的关系，并分别定义 State（球员状态）和 Premium（相对溢价）：State 衡量总体比赛评分相对于能力基线条件预期的偏离，Premium 衡量当前身价相对于能力和前一期身价条件预期的偏离。两个指标均采用十折样本外计算。结果显示，Ability 已能解释球员身价的大部分差异，但加入 Performance（比赛表现）后，样本外原尺度 R^2 由 0.867 提高到 0.921，MAE 下降约 26.08 万欧元，说明比赛表现仍包含 Ability（球员能力）之外的增量信息。State 与原始比赛评分保持较强相关，而与单项 Ability 的最大绝对相关仅为 0.022；将完整 Performance 纳入身价基线后，State 与身价残差的相关降至 0.004，与条件期望残差的正交性质一致。最后，将 State 与 Premium 置于同一二维坐标形成 State–Premium 四象限图，同时比较球员近期表现相对于自身能力的位置与当前身价相对于估值基线的位置，用于大规模候选球员的初步筛查与分层分析。

关键词：条件期望；正交投影；交叉拟合；足球数据分析；球员状态；相对溢价；球员身价

目录

1 研究问题
2 数据与方法
3 Ability、Performance 与球员身价
4 State 的定义与计算
5 Premium 的定义与计算
6 理论依据与稳健性
7 足球应用
8 局限性与结论
参考文献

1 研究问题

这项研究关注球员能力、比赛表现与数据库球员身价之间的关系。实际观察中，比赛数据亮眼的球员不一定身价更高，能力评价接近的球员也可能有明显不同的身价。本文通过样本外预测比较这些信息各自能够解释什么，并进一步把通常较为主观的“状态”和“相对溢价”转化为可计算、可检验的数学量。

本文将主要信息分为相对稳定的 Ability（球员能力，下文简称 Ability）和样本期 Performance（球员赛场表现，下文简称 Performance）。Ability 对应 EA Sports FC 25 数据库记录的球员能力指标，可作为球员能力的可观测代理 [9,10]；Performance 来自正式比赛统计，记录球员在样本期内的实际

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

发挥。考虑到年龄和场上位置与球员估值存在系统性关系 [7]，后文将 Ability 与连续年龄和宽泛位置类别共同组成能力基线 A，其中位置采用 one-hot 编码。

研究依次包括六个步骤：构建样本并采用十折样本外计算；比较不同信息组合的身价预测能力；以实际表现相对能力条件期望的偏差定义 State；以后一期身价相对基础条件期望的偏差定义 Premium；检验残差定义的投影性质与稳健性；最后构建用于候选球员筛选的二维坐标。

本文使用的球员身价来自 EA Sports FC 24 与 FC 25 数据库，并非实际转会成交价。因此，Premium 只是相对于模型基线的残差，不能直接等同于真实转会市场中的“高估”或“低估”。

1.1 三类核心信息

本文使用三类信息。第一类是能力基线，记为 A，包括 EA Sports FC 25 数据库记录的球员 Ability 指标，如总体评分以及技术、身体和意识等能力数据，同时加入年龄和场上位置。第二类是样本期比赛表现，记为 X，包括比赛评分、出场、分钟、进球、助攻等统计及相应的每 90 分钟指标。其中，总体比赛评分（rating）记为 B，用于后文计算 State。第三类是球员身价，V_0 表示 EA Sports FC 24 数据库中的球员身价，V_1 表示 FC 25 数据库中的球员身价；身价模型主要以 $Y=\log(V_1)$ 为结果变量。

Ability 和 Performance 关注的是球员的两个不同方面。Ability 主要反映球员相对稳定的能力水平，Performance 则记录球员在样本期内实际踢出的表现。两者通常相关，但包含的信息并不完全相同。本文先比较它们各自与球员身价的关系，再以能力基线 A 计算 State；计算 Premium 时，则进一步加入前一期身价 V_0，用来提供球员已有的估值水平信息。为简化表述，后文模型比较中使用的“Ability”均指这一能力基线 A。

1.2 预测模型：信息比较工具

球员能力值和比赛统计维度较高，变量之间可能存在非线性关系。因此，本文使用 XGBoost 模型拟合这些特征与球员身价之间可能存在的非线性关系 [5]。

本文以样本外预测性能比较不同信息集合所提供的预测信息，分别估计 Ability、Performance 及其组合的样本外身价预测结果，并观察加入变量后解释度（R^2）的变化。

1.3 相关研究与本文定位

已有研究从多个角度讨论了职业足球中的球员评价与价值预测。年龄、合同、球员能力和比赛表现等信息已被用于解释或预测球员身价 [7,8]；FIFA 类评分也曾被引入球员表现评价模型并与其他指标结合使用 [9]。在比赛表现评价方面，expected goals 通过比较实际结果与给定条件下的预期结果来刻画表现 [10]；其他研究则进一步指出，球员评价和检索需要考虑场上角色、比赛情境以及球员个体差异 [11,12,14]。这些研究说明，无论评价比赛表现还是球员价值，单纯比较绝对数值往往是不够的，还需要把球员放在适当的条件基线下进行比较。

在这些研究基础上，本文进一步把球员能力、样本期比赛表现和球员身价放在同一框架下分析。具体地说，本文分别以能力基线和估值基线为参照，用样本外条件残差定义 State 和 Premium，并考察二者与比赛表现和球员身价之间的关系。这样可以把足球中常说的“状态”和“相对溢价”转化为具有明确基线、可以计算和比较的相对指标。

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

2 数据与方法

2.1 数据来源与样本构建

比赛表现数据来自 WhoScored 的历史记录，原始工作簿共有 83,527 条球员—赛季记录。重构数据时只保留赛季标签为 2024 或 2024/2025 的记录，所有标记为 2025/2026 的记录都被排除，避免出现未来信息泄露。能力变量来自 EA Sports FC 25 数据库；前一期和后一期球员身价分别来自 EA Sports FC 24 与 FC 25 数据库。

比赛统计按变量类型进行聚合：进球、助攻、出场、黄牌和红牌等事件累计求和；评分、传球成功率等均值或比率指标按分钟或出场次数加权；每 90 分钟数据在累计后计算。Performance 与历史能力/球员身价主表通过标准化姓名对照进行拼接，得到 8,316 名球员。主表内部跨版本拼接包含 7,104 个精确匹配、1,051 个高度可信模糊匹配和 161 个中度可信模糊匹配，因此本文还进行了仅采用精确匹配数据的完整重训。建模时年龄保持连续数值，位置由三个 one-hot 编码值体现，Defender 作为参照变量。

2.2 十折样本外计算

无论是球员身价预测，还是 State 和 Premium 的计算，本文都采用样本外预测。原因在于，如果一个球员参与了模型拟合，再用同一个模型计算他的预测残差，训练内拟合可能导致过拟合，使残差人为缩小。对于本文而言，这会直接影响 State 和 Premium 的数值，使它们不能真实反映模型对未参与训练球员的预测偏差。

因此，本文将全部球员随机划分为 K=10 折。计算第 k 折球员的预测值时，只使用其余九折数据训练模型，再将训练好的模型应用于第 k 折。十折依次完成后，每名球员都得到一个由未使用该球员样本进行拟合的模型产生的预测值。本文依据交叉验证的样本外评价思想 [3]，将这一方法同时用于模型评价和残差计算，从而得到用于计算 State 和 Premium 的样本外残差。本文的十折划分按球员随机进行，因此文中的“样本外”均指同一数据环境中对未参与训练球员的预测结果。下文将这种逐折得到的样本外预测简称为 OOF（out-of-fold）预测。XGBoost 回归采用平方误差目标，与条件期望的均方误差定义一致；实际计算得到的是 cross-fitted 样本外残差，后文在不混淆时仍简称 State 和 Premium。

2.3 评价尺度与对数还原

球员身价分布高度右偏，因此核心模型在 $Y=\log(V_t)$ 上训练，同时报告对数尺度与原尺度（欧元）的结果。评价指标主要使用 R^2、MAE 和 RMSE。但简单地对预测结果取指数会产生系统性回转偏差，因此对原尺度球员身价使用每个训练折内部估计的 smearing 因子进行修正 [4]：

$$\hat{V}_i = \exp(\hat{Y}_i) \cdot [1/n_{-k}] \Sigma_j \notin I_k \exp(Y_j - \hat{Y}_j).$$

同一个模型在对数 R^2 和原尺度 R^2 上的大小关系可能并不一致。因此后文都会注明评价尺度，不把某一尺度上的优势直接解释为模型整体更好。

3 Ability、Performance 与球员身价

为了比较 Ability 和 Performance 各自提供的信息，本文先用不同的信息组合预测后一期球员身价，再比较样本外结果。

2026 S.T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

预测信息	Log R^2	原尺度 R^2	原尺度 MAE（€）
Performance + 年龄 + 位置	0.482664	0.502013	3,086,359
Ability only	0.845945	0.867229	1,420,602
前一期身价 only	0.838088	0.802448	1,787,683
Ability + Performance	0.883927	0.921314	1,159,818
Ability + Performance + 前一期身价	0.908113	0.906537	1,211,224

只使用 Ability 时，原尺度样本外 R^2 为 0.867229；加入 Performance 后提高到 0.921314，MAE 由约 142.06 万欧元降至约 115.98 万欧元。Ability 已经解释了球员身价的大部分差异，但在其基础上加入 Performance 后，样本外预测仍有明显改善。

只使用 Performance 并加入年龄和位置时，原尺度 R^2 为 0.502013，说明比赛数据能够解释部分身价差异，但解释力明显弱于 Ability。

样本期表现会受到上场时间、队友、对手、战术、角色和随机事件等多种因素影响，因此单独预测相对稳定的球员身价时不如 Ability。Ability 更接近球员所处的能力层级，而 Performance 则补充了能力相近球员在样本期实际发挥上的差异。

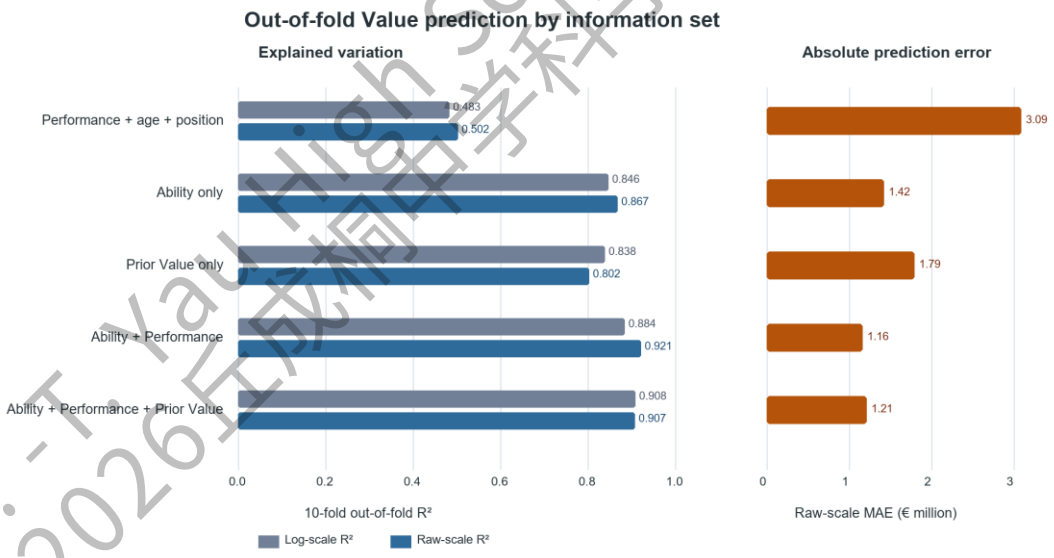


图 1　不同信息集合的十折样本外球员身价预测结果比较。左图同时报告对数尺度与原尺度 R^2；右图报告原尺度 MAE。

3.1 嵌套信息下的经验预测增量

为考察 Performance 在已有 Ability 信息之外还能增加多少预测信息，本文比较只使用 Ability 的模型与同时使用 Ability 和 Performance 的模型。前者的样本外 log-R^2 为 0.845945，后者提高到 0.883927。

这一比较可以写成嵌套信息集合的形式。设结果变量 $Y \in L^2$，较小信息集合为 H_0，较大信息集合为 H_1，且 $H_0 \subseteq H_1$。定义

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

$$\mu_j = E[Y \mid H_j],\ R_j = Y - \mu_j,\ j \in \{0,1\},$$

并记新增信息带来的投影增量为 $G=\mu_1-\mu_0$。由于 $R_0=R_1+G$，且 R_1 与 H_1 中的可测函数正交，有

$$E[R_1 G] = E\{G \cdot E[R_1 \mid H_1]\} = 0,$$

$$E[R_0^2] - E[R_1^2] = E[G^2] \geq 0.$$

这个分解给出了一个直接的解释：当信息集合扩大时，最优平方误差减少的量，正好对应新增投影部分的平方范数。基于这一点，进一步定义归一化信息增益

$$I(H_1 : H_0) = E[G^2] / \mathrm{Var}(Y).$$

在本文中，H_0 对应能力基线 A，H_1 对应 A + Performance。用十折样本外 log-R^2 差作为经验对应量，得到

$$\hat{I}_x|_a = R^2(A,X) - R^2(A) = 0.883927 - 0.845945 = 0.037982.$$

在本文的样本和模型下，加入 Performance 后，样本外对数球员身价的经验预测增量为 0.037982，表明在已有能力基线 A 的基础上，Performance 仍提供了额外预测信息。

4 State 的定义与计算

即使能力水平接近，不同球员在同一段时间里的实际发挥也可能很不一样。State 用于刻画这种“相对自身能力基线预期的偏离”。

例如，同样是 7.2 分的比赛评分，对能力基线不同的球员含义并不相同：对高能力球员而言，它可能只是接近正常水平；而对较低能力球员来说，则可能明显高于预期。因此，本文把 State 定义为实际比赛评分相对于能力基线条件预期的偏离。

4.1 数学定义

为了刻画球员实际表现相对于其能力基线的偏离，本文先估计在给定 Ability、年龄和位置条件下，球员通常能够达到怎样的比赛评分。记球员的总体比赛评分为 B，由 Ability、连续年龄和位置 one-hot 变量组成的能力基线向量为 A。State 需要一个能够用于不同位置球员的统一标量表现变量，因此本文选用总体比赛评分 B，而不是进球、助攻等更依赖场上角色的单项统计。补充样本外检验也支持这一选择：分别在出场时间不少于 450 分钟和 900 分钟的样本中，去掉 B 后的其余 Performance 指标仍可分别以 OOF R^2 =0.698 和 0.709 重建 B；在能力基线 A 上，单独加入 B 所获得的 log-R^2 增量约占加入完整 Performance 后增量的 85.8%。在所有只依赖 A 的预测函数中，条件期望

$$m(A) = E[B \mid A]$$

是在均方误差意义下对 B 的最优预测。于是本文把 State 定义为

$$S = B - m(A).$$

其中，$S>0$ 表示球员实际评分高于其能力基线下的预期，$S<0$ 则表示低于预期。State 由此反映球员样本期表现相对能力基线的偏离。

4.2 投影依据

该定义的理论依据是条件期望的投影性质 [1,2]。XGBoost 用于数值逼近未知的条件期望。对任意平方可积的 g(A)，都有

2026 Qiu Ti Yuan High School Science Award
仅用于2026丘成桐中学科学奖论文公示

$$\mathrm{E}[(\mathrm{B}-\mathrm{g}(\mathrm{A}))^2] \;=\; \mathrm{E}[(\mathrm{B}-\mathrm{m}(\mathrm{A}))^2] \;+\; \mathrm{E}[(\mathrm{m}(\mathrm{A})-\mathrm{g}(\mathrm{A}))^2].$$

右侧第二项恒不小于 0，所以 m(A)=E[B|A] 确实使均方误差最小。进一步，因为 E[B−m(A)|A]=0，可得

$$\mathrm{S} \;=\; \mathrm{B} \;-\; \mathrm{m}(\mathrm{A}),\;\; \mathrm{E}[\mathrm{S}\cdot\mathrm{h}(\mathrm{A})] \;=\; 0$$

这一结论对任意平方可积的 h(A) 都成立。因此，在总体层面扣除能力基线 A 所能解释的部分后，State 不再沿 A 的函数方向呈现系统性变化。本文利用这一投影性质来构造并检验球员的相对状态指标。

4.3 样本外计算

实际计算中，条件期望 m(A) 并不能直接得到，因此需要通过模型进行估计。本文使用十折交叉验证：若球员 i 位于第 k 折，就用其余九折拟合 $\hat{m}^{(-k)}$，再计算

$$\hat{\mathrm{S}}_\mathrm{i} \;=\; \mathrm{B}_\mathrm{i} \;-\; \hat{\mathrm{m}}^{(-k)}(\mathrm{A}_\mathrm{i}).$$

本文利用能力基线向量 A 预测比赛评分，包括记录的 Ability、连续年龄和位置 one-hot 变量。该模型的 pooled OOF R^2 为 0.264891，表明这些信息能够解释比赛评分中的一部分差异。由此得到的 State 与原始总体比赛评分 B 的相关系数为 0.884502，仍保留了较强的原始表现信息；与此同时，State 与各单项 Ability 变量的最大绝对 Pearson 相关仅为 0.021586。进一步使用完整能力基线 A 重新预测 State 时，Ridge 与 XGBoost 的样本外 R^2 均为 −0.003，说明残差化后 A 已不能在新球员上继续预测 State。

4.4 与原始评分的比较

为了公平比较原始总体比赛评分 B 与 State，两组模型使用完全相同的能力基线 A + 前一期身价、缺失值处理、标准化和 RidgeCV 流程，唯一变化是新增变量分别换成总体比赛评分 B 或 State。这样可以尽量排除模型和预处理差异造成的影响。

在 Ability + 前一期身价上新增的变量	Log R^2	Log MAE
总比赛评分 B	0.874628	0.343846
Cross-fitted State	0.874730	0.343828

两组结果几乎没有差别。State−rating 的 log-R^2 差为 0.000102，5,000 次配对 bootstrap 的 95% 区间为 [−0.000226, 0.000425]；log-MAE 差为 −0.000019，区间为 [−0.000569, 0.000531]。这些区间都跨过 0，因此没有证据支持 State 在预测能力上显著优于原始评分。

两组预测性能几乎相同。因此，State 的主要价值不在于提高预测精度，而在于把绝对评分转换为相对于能力基线 A 的表现偏差，同时基本保留其中与球员身价有关的信息。

5 Premium 的定义与计算

球员身价会受到既有能力、历史估值以及其他未观测因素的共同影响。为刻画当前身价相对于自身基础条件的位置，本文用能力基线 A 和前一期身价 V_0 共同构造估值基线。球员身价具有较强的时

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

间延续性，保留 V_0 可以反映球员既有的估值水平。后一期身价高于条件预期时残差为正，低于条件预期时残差为负；本文将这一样本外残差定义为相对溢价 Premium。

5.1 相对溢价的正式定义

令 $Y=\log(V_1)$。定义两个嵌套的球员身价基线：

$$q_0(A,V_0) = E[Y \mid A,V_0],\ q_1(A,V_0,X) = E[Y \mid A,V_0,X].$$

对应的十折样本外残差为

$$Q_0 = Y - \hat{q}_0^{(-k)}(A,V_0),\ Q_1 = Y - \hat{q}_1^{(-k)}(A,V_0,X).$$

本文把 Q_0 作为相对溢价 Premium 的核心坐标。$Q_0>0$ 表示后一期球员身价高于能力基线 A + 前一期身价所对应的条件预期，$Q_0<0$ 表示低于这一基线。$\exp(Q_0)$ 可以理解为观测身价与 exp(预测对数身价) 的比值；若使用 smearing 修正后的条件均值，这一比值还需要相应调整。

5.2 State 与 Premium 的关系

如果身价模型没有纳入样本期表现，State 中的表现偏离可能仍反映在身价残差中。因此，在只使用能力基线 A 和前一期身价构造基线时，State 与身价残差 Q_0 可能存在一定正相关。样本中 $\mathrm{corr}(State,Q_0)=0.175587$。

将完整的 Performance 向量 X 加入身价模型，并将新的样本外残差记为 Q_1。此时，State 所包含的比赛表现信息已经进入条件信息集合。按照条件期望残差的正交性质，若模型能够较好地逼近相应的条件期望，Q_1 不应再与 State 保持系统性的线性关系。因此有

$$E[SQ_1] = 0.$$

记 $F_1=\sigma(A,V_0,X)$。由于 B 是 X 的一个分量，$S=B-E[B|A]$ 对 F_1 可测；$Q_1=Y-E[Y|F_1]$ 满足 $E[Q_1|F_1]=0$，因此

$$E[SQ_1] = E\{S \cdot E[Q_1 \mid F_1]\} = 0.$$

实际样本中 $\mathrm{corr}(State,Q_1)=0.004378$，接近 0；而 $\mathrm{corr}(State,Q_0)=0.175587$。二者之差

$$\Delta_r = \mathrm{corr}(S,Q_0) - \mathrm{corr}(S,Q_1) = 0.171209.$$

加入完整 Performance 后，$\mathrm{corr}(State,Q_1)=0.004378$，较基线模型中的 $\mathrm{corr}(State,Q_0)=0.175587$ 明显下降并接近 0，与前述正交性推导一致。

6 理论依据与稳健性

6.1 条件投影基线

State 的定义建立在条件期望的投影性质上：给定能力基线 A，E(B|A) 是均方误差意义下对总体比赛评分 B 的最优预测，因此 B−E[B|A] 表示不能由 A 解释的相对偏离。实际计算中，本文用 XGBoost 估计未知的条件期望 m(A)=E[B|A]。

6.2 嵌套信息下的经验预测增量

将能力基线 A 与 A + Performance 看作两个嵌套的信息集合时，两组模型样本外 log-R^2 的差值为 0.037982，作为 Performance 新增预测信息的经验量。

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学论文公示

6.3 不确定性区间

为检验这一增量结果的稳定性，本文基于已有的样本外预测结果，在球员层面进行了 5,000 次配对 bootstrap [6]。Ability + Performance 相比 Ability only 的原尺度 R^2 增量为 0.054085，95% 区间为 [0.041127, 0.067310]；MAE 变化为 −€260,784，区间为 [−€305,298, −€216,328]。两个区间都远离 0，说明 Performance 的增量明显大于这一层面的重采样波动。上述 bootstrap 直接对已保存的样本外预测进行球员层面的配对重采样，每次重采样不重新训练模型，因此区间反映的是当前样本外预测结果的重采样稳定性。

6.4 稳健性检验

稳健性变体	n	Ability 原尺度 R^2	Ability + Performance 原尺度 R^2	增量
独立重训基准：当前能力定义	8,316	0.871333	0.921130	0.049797
Ability + Potential（潜力评分）	8,316	0.865282	0.913187	0.047905
仅保留跨版本 exact 匹配	7,104	0.865819	0.912994	0.047175
排除标准化姓名同名冲突	8,086	0.875925	0.922177	0.046252

四组结果中，加入 Performance 后，原尺度 R^2 的增量均保持在约 0.046—0.050：将 Potential 纳入 Ability 后为 0.047905，仅保留 exact 匹配时为 0.047175，排除同名冲突后为 0.046252。补充潜力信息或采用更严格的身份连接样本后，Performance 的增量仍基本保持。

表中的“独立重训基准”与三种稳健性变体来自同一套独立重训流程，因此其数值与前文最终主模型的 0.867229 和 0.921314 略有差异。稳健性检验关注的是同一行内加入 Performance 前后的配对变化，不宜直接比较不同重训管线之间的绝对数值。此外，在完全不直接使用原始 B 的情况下，先由其余 Performance 样本外重建 B，再按能力基线 A 残差化得到替代 State；该指标与主 State 的 Pearson 相关为 0.987，正负号一致率为 96.7%，进一步支持了以 B 构造 State 的稳定性。

6.5 分层可靠性与模型比较

前一期身价在原尺度上明显右偏，样本偏度为 6.590，取对数后降至 0.581。若直接按前一期身价分层，可靠性分析结果很容易被这一偏斜影响，因此本文改用 Ability + Performance + 前一期身价的完整模型的十折样本外预测对数身价四分位分层。完整模型预测值偏度为 0.585，每层 2,079 名球员。总体校准斜率为 1.005、截距为 −0.071，平均对数残差为 0.0012，说明样本外预测在整体尺度上接近校准。

表 4　按完整模型十折样本外预测对数身价分层的可靠性诊断。组内原尺度 R^2 受分层后的范围收缩影响，主要结合 MAE 和中位数进行联合分析。

完整模型预测层	n	观测中位数（欧元）	预测中位数（欧元）	log-MAE	原尺度 R^2
VQ1 低	2,079	475,000	520,772	0.3167	0.2293
VQ2	2,079	1,100,000	1,137,244	0.2772	0.1140
VQ3	2,079	2,000,000	2,099,393	0.2742	0.1460
VQ4 高	2,079	8,000,000	8,124,608	0.3052	0.8684

2026 STS Taylor High School Science Award
仅用于2026丘成桐中学科学奖论文公示

原尺度 R^2 在不同身价层之间差异较大，而对数尺度下的误差相对稳定。这与球员身价高度右偏的分布特征一致；采用对数身价建模有助于减弱极端高身价样本的影响，使不同身价区间的比较处在更接近的尺度上。

按位置复核时，Defender、Forward、Goalkeeper 和 Midfielder 的原尺度样本外 R^2 分别为 0.897、0.904、0.888 和 0.912，平均对数残差介于 −0.0179 和 0.0142 之间。按年龄分为 ≤21、22—25、26—29 和 ≥30 岁后，原尺度样本外 R^2 分别为 0.695、0.913、0.915 和 0.908。整体上，各位置组表现较一致，而 21 岁及以下球员仍然是明显更难预测的一组。

为比较同一分层中的模型，所有球员身价层都固定用完整模型的十折样本外预测对数身价划分，比较保留 Ability only、Ability + Performance 和 Ability + Performance + 前一期身价三个信息集合，并报告完整模型的 log-MAE。

表 5　固定完整模型预测球员身价层与年龄组的样本外模型比较。前三个模型列为原尺度 R^2；最后一列为完整模型 log-MAE。组内 R^2 可以为负，表示模型在该组内的预测误差大于直接使用该组均值作为预测值的误差。

分组	n	Ability	Ability + Performance	完整模型	完整模型 log-MAE
预测球员身价 VQ1	2,079	−0.120	0.184	0.229	0.3167
预测球员身价 VQ2	2,079	−0.325	0.016	0.114	0.2772
预测球员身价 VQ3	2,079	−0.223	0.144	0.146	0.2742
预测球员身价 VQ4	2,079	0.815	0.888	0.868	0.3052
年龄 ≤21	783	0.607	0.759	0.695	0.3461
年龄 22—25	2,541	0.878	0.932	0.913	0.3069
年龄 26—29	2,732	0.878	0.924	0.915	0.2797
年龄 ≥30	2,260	0.866	0.907	0.908	0.2763

按预测身价分层或按年龄分组后，加入 Performance 的原尺度 R^2 均高于只使用 Ability 的模型，说明这一增量并非由单一球员群体驱动。前一期身价的作用则存在组间差异：完整模型在部分低、中身价层表现更好，但在最高身价层及部分年龄组的原尺度 R^2 出现回落。原尺度 R^2 的这些波动也反映了组内身价范围和高身价样本的影响。

7 足球应用

7.1 State

Performance 由多项比赛统计构成，其中总体比赛评分 B 提供统一的标量表现尺度。State 考察这一评分相对于球员自身能力基线预期的位置。由于 Ability、年龄和位置不同，不同球员对应的预期评分也不同：一名总体评分较高的球员如果能力基线同样很高，State 可能接近 0；相同的比赛评分对较低能力基线的球员而言，则可能得到明显为正的 State。

已有研究表明，比赛情境、位置角色和球员个体差异都会影响实际比赛表现 [12,14]。因此，State 更适合作为经过 Ability、年龄和位置基线调整后的相对表现指标；实际筛选仍需结合完整 Performance、位置需求和比赛录像等信息。

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

7.2 Premium

从数学定义上，Premium 是后一期对数身价相对于 Ability 和前一期身价条件预期的残差；在应用解释中，它描述当前身价相对于这一估值基线的位置。正 Premium 表示高于基础预期，负 Premium 表示低于基础预期，因此它区分的是绝对身价与“相对自身基线的身价位置”。

Premium 可用于比较球员当前身价相对于 Ability 和前一期身价基线的偏离程度。在基础条件接近的候选球员之间，它提供一个相对价格位置的比较维度；实际转会判断仍需结合合同年限、工资、伤病、战术适配、俱乐部议价和真实报价等信息 [7,8]。

7.3 State–Premium 四象限

State 和 Premium 分别表示球员表现相对于能力预期的偏离，以及当前身价相对于基础估值的偏离。将两者作为坐标轴画成四象限图，可以同时比较球员在表现和身价两个方面的相对位置。

第一象限（State>0，Premium>0）表示样本期发挥高于能力预期，同时球员身价也高于基础基线，即“相对表现较高、相对身价也较高”。在预算允许且更看重近期表现的筛选条件下，这类球员可进入进一步考察范围。

第二象限（State>0，Premium<0）表示球员发挥高于能力预期，但当前身价低于能力基线 A + 前一期身价所给出的相对位置，形成“状态偏高、相对价格位置偏低”的组合。在本文的筛选框架下，可优先结合年龄、位置需求、合同和比赛录像进一步复核。

第三象限（State<0，Premium>0）表示样本期发挥低于能力基线，而当前身价高于基础预期。在只考虑本文两个指标时，这一组合需要更谨慎地复核近期表现、伤病和战术角色。

第四象限（State<0，Premium<0）表示样本期表现和相对身价位置均低于各自基线。这类球员需要结合基础能力、年龄、伤病和更长时间的表现记录判断；若基础能力达到球队要求且价格条件合适，仍可继续考察。

四象限图主要用于球员初筛。绝对 Performance 难以说明同样的比赛表现对于不同能力水平球员意味着什么，绝对身价也不能反映当前价格相对于自身基础条件的位置。已有研究表明，数据模型可以用于球员检索和球探分析 [11,12]；本文用 State 和 Premium 同时引入能力基线和身价基线，对较大的候选池进行初步分层。

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

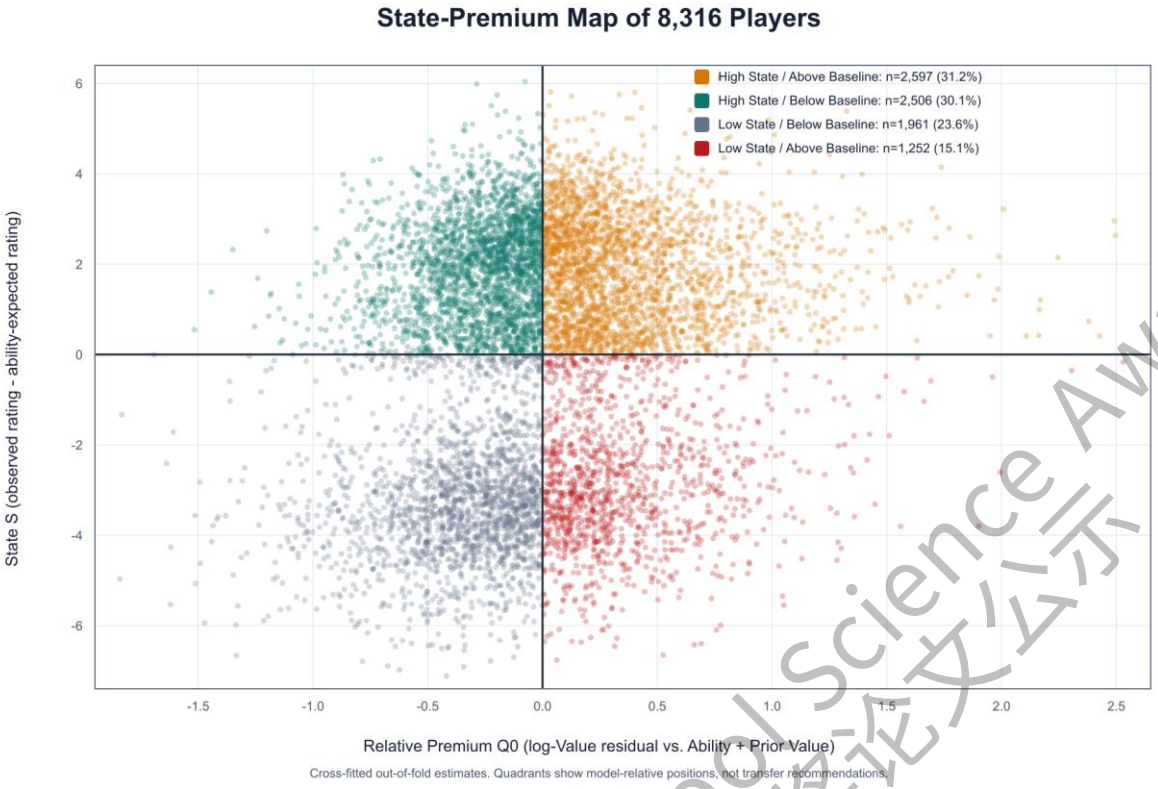


图 2　8,316 名球员的 State–Premium 四象限分布。横轴为相对溢价 Q_0，纵轴为 State。

以理论零点划分后，High State / Above Baseline 有 2,597 人（31.2%），High State / Below Baseline 有 2,506 人（30.1%），Low State / Below Baseline 有 1,961 人（23.6%），Low State / Above Baseline 有 1,252 人（15.1%）。这些结果展示了样本中球员在 State – Premium 坐标上的分布情况。

7.4 应用流程

筛选方式可以根据球队的引援目标调整。若需要在当前转会窗口补强某个位置，可先根据位置、预算和绝对 Performance 划定候选范围，再利用 State 找出近期表现高于自身能力预期的球员，并结合 Premium 比较其相对价格位置。对于更偏长期的引援，则可从年龄、Ability 和位置需求出发，重点关注基础能力达到要求、Premium 较低且 State 较高的球员。

在较大的候选池中，首轮筛查的任务是确定哪些球员值得继续观察。本文最终连接得到 8,316 名球员，与已有大型跨国职业足球研究的样本规模相近。Bosker 和 Gürtler [15] 将 Wyscout 比赛数据与 Transfermarkt 身价数据结合，覆盖 99 个国家和 431 项赛事，共涉及 11,098 名球员，进一步筛选后形成 9,391 名球员的分析样本。从样本规模看，本文可以在较大候选范围内展示这一框架的首轮筛查和分层结果。

State–Premium 四象限更适合用于引援流程前端的候选分层，而不是替代后续的比赛录像、球探报告和合同评估。已有研究已将数据模型用于球员检索和球探分析 [11]，随着足球数据规模扩大，可扩展的数据处理与分析流程也越来越重要 [13]。

8 局限性与结论

8.1 局限性

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

第一，本文使用的球员身价来自 EA Sports FC 24 与 FC 25 数据库记录，并非实际转会成交价。因此，本文的身价分析属于数据库身价的跨版本比较，不能直接等同于真实转会市场的前瞻预测。

第二，本文以 EA Sports FC 25 数据库中的 Ability 指标作为球员能力的可观测代理，并结合年龄和宽泛位置构成 State 的能力基线。球队、联赛、对手、战术以及更细的位置角色等因素没有进入这一基线，因此 State 的含义仍取决于当前能力指标和模型所包含的信息，部分比赛环境的影响也可能保留在 State 中。

第三，Performance 数据与能力、身价数据需要通过球员身份进行连接，其中除 7,104 名精确匹配球员外，还包含部分高可信度的模糊匹配。本文已经通过仅保留精确匹配以及排除同名冲突重新训练模型，两种情况下主要结果均保持稳定，但数据连接仍可能存在少量识别误差。

第四，本文采用随机球员十折划分进行样本外计算，因此检验的是同一数据环境下模型对未参与训练球员的预测表现。跨赛季、跨联赛以及更长期的预测能力，还需要在具有明确时间顺序或外部联赛的数据上进一步检验。

第五，本文报告的 bootstrap 区间是在已经得到的样本外预测结果上进行球员层面的配对重采样，并没有在每一次重采样中重新训练全部模型。因此，这些区间主要反映当前预测结果的稳定性，没有包含模型选择、折划分和数据连接等全部不确定性。

第六，不同球员群体的预测难度并不完全相同。分组结果中，各位置组的预测表现较为接近，但 21 岁及以下球员的原尺度样本外 R^2 为 0.695，明显低于其他年龄组；不同身价层的原尺度误差表现也存在差异。因此，在使用 State–Premium 坐标比较具体球员时，年龄和身价区间仍是需要同时考虑的信息。

8.2 结论

本文首先比较 Ability、Performance 与球员身价之间的关系。Ability 是预测球员身价的主要信息来源；加入 Performance 后，原尺度 R^2 由 0.867229 提高到 0.921314，MAE 下降约 26.08 万欧元，说明样本期比赛表现仍提供 Ability 之外的预测信息。

本文进一步定义 State 和 Premium。State 是实际总体比赛评分与能力基线 A 条件下预期评分之差，用于衡量样本期发挥相对于自身能力基线的偏离；Premium 是后一期对数身价与能力基线 A、前一期身价条件下预期对数身价之差，用于表示当前身价相对于自身估值基线的位置。

两个指标都以条件期望为基线，并用实际值与样本外条件预期之间的残差表示相对偏离。十折 cross-fitting 保证每名球员的指标均由未使用该球员样本进行拟合的模型计算。

样本结果支持上述构造。State 与原始总体比赛评分 B 保持较高相关，而与各单项 Ability 变量的线性相关很弱；在相同的身价预测流程下，加入 State 或原始总体比赛评分 B 的结果也几乎一致。只使用 Ability 和前一期身价时，State 与身价残差存在一定正相关；加入完整 Performance 后，这一相关降至接近 0，与条件期望残差的正交性质相符。

将 State 和 Premium 放在同一坐标中形成 State–Premium 四象限图，可以同时比较球员近期表现相对于自身能力预期的位置，以及当前身价相对于基础估值的位置。基于 8,316 名球员的匹配样本，这一框架可用于较大候选范围的初步筛查和分层比较。

2026 Siu Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

参考文献

[1] Billingsley P. Probability and Measure: Anniversary Edition. Hoboken, NJ: John Wiley & Sons; 2012.

[2] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. New York: Springer; 2009.

[3] Stone M. Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society: Series B (Methodological). 1974;36(2):111-133. doi:10.1111/j.2517-6161.1974.tb00994.x.

[4] Duan N. Smearing estimate: A nonparametric retransformation method. Journal of the American Statistical Association. 1983;78(383):605-610. doi:10.1080/01621459.1983.10478017.

[5] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM; 2016. p. 785-794. doi:10.1145/2939672.2939785.

[6] Efron B. Bootstrap methods: Another look at the jackknife. The Annals of Statistics. 1979;7(1):1-26. doi:10.1214/aos/1176344552.

[7] Franceschi M, Brocard J-F, Follert F, Gouguet J-J. Determinants of football players' valuation: A systematic review. Journal of Economic Surveys. 2024;38(3):577-600. doi:10.1111/joes.12552.

[8] Müller O, Simons A, Weinmann M. Beyond crowd judgments: Data-driven estimation of market value in association football. European Journal of Operational Research. 2017;263(2):611-624. doi:10.1016/j.ejor.2017.05.005.

[9] Matano F, Richardson L, Pospisil T, Politsch CA, Qin J. Augmenting adjusted plus-minus in soccer with FIFA ratings. Journal of Quantitative Analysis in Sports. 2023;19(1):43-49. doi:10.1515/jqas-2021-0005.

[10] Mead J, O'Hare A, McMenemy P. Expected goals in football: Improving model performance and demonstrating value. PLOS ONE. 2023;18(4):e0282295. doi:10.1371/journal.pone.0282295.

[11] Pappalardo L, Cintia P, Ferragina P, Massucco E, Pedreschi D, Giannotti F. PlayeRank: Data-driven performance evaluation and player ranking in soccer via a machine learning approach. ACM Transactions on Intelligent Systems and Technology. 2019;10(5):Article 59, 27 pages. doi:10.1145/3343172.

[12] Decroos T, Bransen L, Van Haaren J, Davis J. Actions speak louder than goals: Valuing player actions in soccer. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM; 2019. p. 1851-1861. doi:10.1145/3292500.3330758.

[13] Rein R, Memmert D. Big data and tactical analysis in elite soccer: Future challenges and opportunities for sports science. SpringerPlus. 2016;5:1410. doi:10.1186/s40064-016-3108-2.

[14] Forcher L, Forcher L, Härtel S, Jekauc D, Wäsche H, Woll A, Gross T, Altmann S. Does technical match performance in professional soccer depend on the positional role or the individuality of the player? Frontiers in Psychology. 2022;13:813206. doi:10.3389/fpsyg.2022.813206.

[15] Bosker J, Gürtler M. The impact of cultural differences on the success of elite labor migration—Evidence from professional soccer. Annals of Operations Research. 2024;341:781-824. doi:10.1007/s10479-024-06200-4.

2026 C. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示