

New coach effect

xx and king

2025-05-16

模型设计

模型考虑的关键因素

- (待定) 同质性 (Homogeneity): 队员之间的相似性, 如技能、年龄、经验、薪资、位置等
- 激励不对称 (Asymmetry): 参与者在机会或动机上的不平等
- 赛季性因素 (Seasonality): 周期性规律影响, 如主客场轮换、赛程密度等

模型结构与变量说明

本研究的核心目的是分析在不同球队异质性水平下, 主教练更换是否会影响球队表现。我们以比赛得分 (0/1/3 分) 为主要衡量指标, 构建线性回归模型 (OLS) 以识别换帅的平均效应及其与球队结构的交互作用。同时, 为了验证主模型的稳健性, 我们使用进球数为因变量, 构建 Poisson 回归模型, 对换帅前后球队进攻表现的变化进行补充分析。

我们重点关注以下几个问题: 换帅是否提升得分? 是否提升进球数? 短期与长期效果是否不同? 球队异质性是否调节了换帅效果? 需要指出的是, “是否提升得分”与“是否提升进球数”是两个不同但相关的问题。进球数增加

并不必然导致得分增加，原因在于若球队同时失球更多，比赛结果可能并未改善。因此，我们采用得分作为主模型（OLS）的因变量，用以衡量整体表现是否提升；进球数则用于稳健性检验（Poisson），辅助判断进攻能力是否确有改善。

控制变量总结表

变量名	类型	说明
HomeDummy	虚拟变量	是否主场 (主场 =1, 客场 =0)
OpponentElo	连续变量	对手 Elo 分数，衡量对手实力
TeamElo	连续变量	本队 Elo 分数，衡量球队基础实力
PreForm	连续变量	换帅前 5 场比赛的平均得分，用于控制换帅决策动机
CoachTenure	连续变量	教练执教的场次，用于控制经验效应
MatchDay	类别变量或连续变量	比赛轮次，控制赛季中的时间趋势

OLS

$$\text{Points}_{ijt} = \alpha + \beta_1 \cdot \text{NewCoach}_{it} + \beta_2 \cdot (\text{NewCoach}_{it} \times \text{Heterogeneity}_{it}) + \gamma' \mathbf{X}_{ijt} + \varepsilon_{ijt}$$

- Points_{ijt} : 球队 i 在对阵 j 的比赛中获得的积分
- NewCoach_{it} : 是否在比赛前换帅的虚拟变量
- $\text{Heterogeneity}_{it}$: 球队的异质性指标
- \mathbf{X}_{ijt} : 控制变量（如主客场、对手 Elo 等）
- ε_{ijt} : 误差项

Poisson

$$\log(\lambda_{it}) = \alpha + \delta \cdot \text{NewCoach}_{it} + \gamma' \mathbf{X}_{it} \quad \text{with} \quad y_{it} \sim \text{Poisson}(\lambda_{it})$$

- y_{it} : 球队在第 t 场比赛中的进球数； λ_{it} : 进球的期望； \mathbf{X}_{it} : 控制变量与上面类似。

数据来源与处理

（描述数据来源、样本区间、清洗方式等）

在本研究中，我们构造了一个比赛级的“异质性（heterogeneity）”指标，用于衡量球队在近期比赛中的表现波动程度，作为球队结构一致性或稳定性的代理变量。该变量的引入旨在探讨换帅效应是否依赖于球队自身的结构特征。

考虑到我们当前数据集中并不包含球员级别的年龄、位置或薪资等信息，我们采用了基于比赛表现变量的滚动标准差法作为替代方案。

具体地，我们选取了三个反映球队进攻质量的变量：• Sh_Standard（射门次数）• SoT_Standard（射正次数）• xG_Expected（预期进球）

我们按球队和时间排序，并基于每支球队过去五场比赛中的上述三个变量，计算它们的标准差并取均值，作为球队在该场比赛的异质性指标。数学表达如下：

$$\text{Heterogeneity}_{it} = \frac{1}{3} \sum_{k \in \{\text{Sh, SoT, xG}\}} \text{SD}_{i,t-4:t}(k)$$

其中， $\text{SD}_{i,t-4:t}(k)$ 表示球队 i 在 $t-4$ 到 t 五场比赛中变量 k 的标准差。

构造该变量的核心逻辑在于，若一支球队近期表现波动较大，其内部执行力、战术协调或人员配置可能更具“异质性”；反之则更稳定、集中，可能更容易响应教练更换所带来的变化。

异常值说明

实证建模中，异常值（outliers）是指显著偏离变量典型分布的观测值。这些极端值可能来源于记录错误、测量波动、真实但罕见的事件等。

在本研究中，我们使用了普通最小二乘法（OLS）和 Poisson 回归模型来评估换帅效应，这两种模型对异常值的存在都具有一定程度的敏感性。具体而言，OLS 模型最小化的是残差的平方和（Sum of Squared Residuals），这使得远离回归线的观测点具有更大的权重，显著影响估计系数的方向和显著性。极端得分（例如单场比赛中极高的进球数或积分）可能导致换帅效应被高估或低估。而在 Poisson 模型中，虽然对异常值的敏感性相较于 OLS 较低，但由于其对因变量（进球数）假定为计数数据，且默认条件均值和方差相等，极端大值（例如单场 7-0、8-1 的比赛）仍可能导致 λ （条件均值）的估计偏离常规范围，进而影响所有协变量的边际效应估计。

通过对本数据集中各主要变量进行 Z 分数标准化检查，我们发现如 `npG_Expected`、`G_per_Sh_Standard`、`GF` 等变量中存在大量异常值（Z 分数 > 3 ），例如 `GF` 中有超过 100 个异常观测值，这些异常值多集中在个别高得分场次。若这些极端场次恰好出现在换帅后的观察期中，则可能导致对换帅效果的过度估计。此外，在控制变量如 `xG_Expected` 和 `elo_pre` 中也存在少量异常值，若未加以处理，可能使模型在解释球队进攻能力或对手强度时产生偏误。

因此，在后续的稳健性分析中，我们将采用如下策略应对异常值的影响：一是对高于 99 分位或低于 1 分位的变量进行截尾（winsorization）处理，二是剔除明显的极端观测，并对比清洗前后的估计结果，从而验证模型对异常值的敏感程度并增强结论的稳健性。

建模公式与估计策略

（写出 OLS 和 Poisson 模型的数学公式）

实证结果与解释

（表格形式输出结果，并用文字解释）

稳健性分析

（如使用短期窗口分析、替代变量、子样本等）

结论与讨论

（总结发现、理论含义、政策建议等）

附录

数据集变量说明

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

filter, lag

The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

表 2: 数据集变量定义表

变量名	英文全称	中文解释
GF	Goals For	本队在比赛中的进球数
GA	Goals Against	本队在比赛中被对手攻入的球数
Gls_Standard	Goals (Standard)	常规进球数 (不含点球)
Sh_Standard	Shots	射门次数
SoT_Standard	Shots on Target	射正次数
SoT_percent_Standard	Shot Accuracy (%)	射正率 (射正/射门)
G_per_Sh_Standard	Goals per Shot	每次射门的进球率
G_per_SoT_Standard	Goals per Shot on Target	每次射正的进球率
Dist_Standard	Average Shot Distance	平均射门距离
PK_Standard	Penalty Goals	点球进球数
PKatt_Standard	Penalty Attempts	点球尝试次数
FK_Standard	Free Kick Attempts	任意球尝试次数
xG_Expected	Expected Goals (xG)	预期进球数
npxG_Expected	Non-Penalty Expected Goals	非点球预期进球
npxG_per_Sh_Expected	Non-Penalty xG per Shot	每次射门的非点球预期进球
G_minus_xG_Expected	Goals minus xG	实际进球与预期进球之差
np:G_minus_xG_Expected	Non-Penalty Goals minus xG	非点球进球与非点球 xG 之差
elo_pre	Pre-match Elo Rating	比赛前本队 Elo 等级评分
opp_elo_pre	Opponent Pre-match Elo	比赛前对手 Elo 等级评分