Exploring the New Coach Effect with Data

Yiming Zang and Jiahui Fan

Research Question

- To evaluate whether the change of coach has a statistically and practically significant effect on team performance, using model-based inference and standardized metrics.
-

Data Preparation

Table 1: Example rows around coaching change

Team	Date	Coach	Change	Change ID	Phase	Points	Heterogeneity
Dortmund	2018-05-12	Peter Stöger	0	NA	before	0	0.7709338
Dortmund	2018-08-26	Lucien Favre	1	35	after	3	0.7717424
Dortmund	2021-05-22	Edin Terzic	0	NA	before	3	1.1793121
Dortmund	2021-08-14	Marco Rose	1	137	after	3	0.6322815
Dortmund	2022-05-14	Marco Rose	0	NA	before	3	1.1540355

- Scraped all Bundesliga match data (2017–2025)using the package`worldfootballR`
- 30 teams, total 4876 matches (only `ForAgainst = For` rows kept)

Coach Change Events and Hetergeneity

Hypothesis: When is a coaching change more effective? — When the team is homogeneous.

- Identified coach change points and assigned `event_id`
- Created windows before/after the change
- Constructed:
  - `relative_time` (time index relative to change)
  - `post` (binary: pre/post change)
  - `time_post` (time since change)

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

where $\text{SD}_{i,\,t-4:t}(k)$ denotes the standard deviation of variable k for team i over the five-match window from t-4 to t. The variables include: ▪ Sh: number of shots ▪ SoT: shots on target ▪ xG: expected goals –

Visualization

- Plotted average points over time by `relative_time`
- Compared pre- and post-change trends
- Goal: visually assess changes in team performance after coaching switches

Interrupted Time Series Model (ITS) and Poisson Regression Model

We model team performance using the following regression:

$$y_{it} = \alpha_i + \beta_1 t + \beta_2 \cdot \text{post}_{it} + \beta_3 \cdot \text{time\_post}_{it} + \beta_4 \cdot \text{home}_{it} + \beta_5 \cdot \text{elo}_{it} + \varepsilon_{it}$$

where α_i corresponds to C(Team) or heter— the fixed effect for each team.

We model count-type outcomes (e.g., goals or points) as:

$$y_{it} \sim \text{Poisson}(\lambda_{it})$$

with the log link function:

$$\log(\lambda_{it}) = \alpha_i + \beta_1 t + \beta_2 \cdot \text{post}_{it} + \beta_3 \cdot \text{time\_post}_{it} + \beta_4 \cdot \text{home}_{it} + \beta_5 \cdot \text{elo}_{it}$$

Further Perspective: Tactical Efficiency

- Define `eff = GF - xG_Expected`
- If post-change `eff > 0` and significant $\rightarrow$ better tactical efficiency
- If `xG_Expected` ↑ but `eff   0` $\rightarrow$ chances created but not converted