Replication archive for Peeters and van Ours (2021)

Estimation code

We conduct the entire analysis of the paper in one stata do-file "homeadvantage.do". In part 1, this file merges the relevant datasets and creates seasonal variables. In part 2, we conduct regression analyses, followed by drawing graphs and tables in part 3. The whole procedure is logged in homeadvantage.log.

Dataset

Construction

The dataset is built up from 3 source files.

- Englandgames.dta
- Englandmanagers.dta
- Englandfinance.dta

These are combined into Englandgamelevel.dta, a dataset at the game level. In the course of the code the data is further collapsed to the team-season level and stored as seasonalHAEngland.dta.

Variable definitions

The variable definitions in Englandgamelevel.dta are given below. We refer to a game i, played by team c, under manager m, in year t. The variables get the prefix op_ if they refer to the opponent team, no prefix for own team. Note that each game feature twice, once from each team's perspective.

- d — calendar day of game i
- m — calendar month of game i
- y — calendar year of game i
- team — name of team c
- id — identification number of team c
- op_team — name of opponent team in game i
- op_id — identification number of opponent team in game i
- year — year season t ends in which game i is played
- attendance — attendance at the ground for game i
- match_id — identifier of game i
- home — indicator whether team c is home team
- sc — number of goals made by team c in game i
- op_sc — number of goals conceded by team c in game i
- division — division in which game i is played, for EL and CL, round of play
- res — result of game from perspective of team c, 2= win, 1= draw, 0= loss
- win — indicator for a win by team c in game i
- draw — indicator for a draw by team c in game i
- loss — indicator for a loss by team c in game i

- manager — name of manager m
- manid — identifier for manager m can differ by spell
- manidfe — identifier for manager m consistent over spells
- club — team c name
- div — division team c plays in season t
- played — number of games played in season t by team c
- wins — number of games won in season t by team c
- draws — number of games drawn in season t by team c
- losses — number of games lost in season t by team c
- gopro — number of goals scored in season t by team c
- goag — number of goals conceded in season t by team c
- godif — goal difference in season t for team c
- seasonpoints — earned points in season t for team c
- rank — end rank in season t for team c across divisions, 1=champion
- rev — total revenues of team c in season t inclusive transfer income
- wage — total wage cost team c (incl social security and tax) in season t
- stadname — name of home ground team c in season t
- stadcap — capacity of home ground team c in season t
- avatt — average attendance at home ground team c in season t
- newstaddum — indicator if team c has new home ground in season t
- newstadyear — season in which team c first played in current home ground

Output

The code creates the output tables and graphs, which correspond to the figures and tables in the published paper:

- Attendance.gph/.png — Figure 1
- HA-div1.gph/.png — Panel 1 figure 2
- HA-div2.gph/.png — Panel 2 figure 2
- HA-div3.gph/.png — Panel 3 figure 2
- HA-div4.gph/.png — Panel 4 figure 2
- HAyearfe.gph/.png — Figure 4
- clubtable1.xls + clubtable2.xls — Table 1
- mantable1.xls + mantable2.xls — Table 2
- HAreg.xml/.txt — Table 3
- Qualityreg.xml/.txt — Table 4

Figure 3 is created outside of the code file.