

**Bachelor Thesis:
Favoritism Under Social Pressure: Evidence From English Premier League**

BY VOJTECH HERRMANN

Supervisor Report

September 3, 2017

This diploma work studies factors that determine the exact length of an overtime in a football match played in the English Premier League in the recent seasons. The main question is whether there is a systematic bias for some specific teams. An example of a systematic bias is that strong teams (such as members of Big 5: Arsenal, Chelsea, Liverpool, Manchester City, Manchester United) would play shorter overtimes when being ahead, but longer overtimes when being behind.

The author uses linear regression analysis to identify the major factors. He applies dummy variables that indicate the score differential from the perspective of a home team. This part has a small theoretical contribution that identifies whether the difference of the coefficients of the dummy variables is statistically significant or not. The conclusion of the data analysis is that the major factor explaining the length of the overtime is the actual score differential at the end of the regular time. The overtime is longer for games that can still change the outcome, i.e., the games with the score differential -1 , 0 , or $+1$. The games with a larger absolute score differential are at that point essentially decided and the overtime tends to be shorter. Interestingly, the score differential $+1$ (= home team is leading by a goal) extends the game rather than shortens it. This is different from the Spanish league where there is a systematic bias towards the home team. The analysis of the English Premier League suggests that the referees are not home biased, and in fact they are also not biased towards some specific teams.

The contribution of the thesis is the following. The author demonstrates that he can perform a linear regression analysis on a new data and understands the theoretical implications of such a model. The results are novel and interesting for both the specialists and the public. The thesis is written in solid English, which is a plus.

At the same time, the thesis is not perfect. The study of the team bias probably require a more complex approach of a Bayesian statistics in contrast to the presented approach that uses dummy variables. However, this is beyond what is expected in a bachelor thesis. I have two specific comments: the analysis of the normality of the overtime (or log of the overtime) on pages 16-17 is irrelevant as the regression model assumes only normality of the residual, not the normality of the explained variable. The second comment is that other types of significantly important stoppages, most notably substitutions and red cards, are not discussed in sufficient detail.

In conclusion, I recommend that the thesis to be accepted as a bachelor thesis. The

final mark should reflect both the above mentioned positives and negatives of the work.
My impression is that the positives overweight the negatives.

Jan Vecer,
KPMS, MFF UK,
Sokolovska 83
18675 Praha 8
Email: vecer@karlin.mff.cuni.cz