

Analysis and prediction of league game results

by Filip Šimsa

Referee Report

May 27th, 2015

This referee report evaluates a diploma thesis of Mr. Filip Šimsa submitted for the thesis defense at the Department of Probability and Statistics, Charles University. It focuses on analysis and prediction of the Czech ice hockey league results. The thesis has two major parts, the first three chapters summarize theoretical statistical techniques used in the thesis (namely 1. Theoretical background, 2. Dynamic linear model, and 3. Generalizations of linear model), followed by two chapters where these techniques are applied on the Czech ice hockey league data (chapters 4. Analysis of Czech extraliga 1999 – 2014, and 5. Search for profitable strategy).

The theoretical part represents a solid overview of the statistical techniques used in the thesis. This part is not expected to bring anything substantially new. The author proposes to use Gini coefficient as a goodness of fit criterion for generalized linear models which seems to be a novelty as the author cites mostly his own previous work in that direction. This would probably deserve a more detailed discussion of the goodness of fit used in other literature. It is not clear from the thesis what is the advantage of this approach.

The core of the thesis is in applying the statistical techniques described in the first three chapters on predicting the outcomes of the Czech ice hockey extraliga games. Unfortunately, the author does not clearly summarize the main new results in the introduction of the thesis or in the introduction of the chapter 4, so it is less obvious for the reader what is the exact contribution of the work. The quality of English also somewhat decreases in chapter 4 in comparison to the previous text, suggesting that the author did not have enough time for proper language editing. The main findings from the text are:

- it is better to estimate the quality of the team for each season separately,
- goal differences are more informative than the win/draw/loss result,
- the only significant predictor is the home team advantage, while other factors such as the result or tiredness from the last match, or the outcome from the last head to head match proved to be insignificant,
- it turns out that the quality of the team is better measured by the current league standing in contrast to the outcomes of the last few games.

I find most of these results as expected, there are no counterintuitive conclusions. The last chapter focuses on the possibility of finding a profitable betting strategy to see whether the odds quoted by the betting companies are systematically mispriced in comparison to the proposed model. It turns out that there could be a small bias based on the suggested evidence that the betting companies may underestimate the home team advantage. However, the author shows only the profit and loss resulting from the proposed betting strategy (observation of a single random variable), but he does not include the possible variability of these outcomes, which would give a confidence interval on the likely range of the profit-loss. One can conclude the bias only if the substantial percentage of the profit-loss distribution ends up in profit, which is not discussed in the thesis. The scale of the theoretical win from the proposed betting does not seem to be convincingly above zero, so it can easily happen by a pure coincidence. The author mentions that the theoretical edge decreased in the recent years and thus the quoted odds are reasonably reflecting the reality, making the possibility of a profitable strategy virtually impossible.

I find the overall quality of the thesis as quite decent, the author has shown his competence in both theoretical and practical aspects of this work. He also demonstrated a good knowledge of the existing literature. The author has obviously spent some nontrivial time into this project. At the same time, the work could have been improved by a more careful editing. Despite of any possible shortcomings of this work, I find the applications of the modern statistical techniques as quite original, supporting the fact that the author had to develop them from the scratch without any previous approach described in the literature.

Jan Vecer
Professor of Quantitative Finance
Frankfurt School of Finance and Management
Sonnemannstrasse 9–11
Frankfurt am Main, 60314
Germany