Opponent´s report on the dissertation by Jiří Lahvička

The dissertation of Jiří Lahvička consists of five separate articles whose subject matter refers to the economics of sports. Three of these articles already went through peer-review process and were published in academic journals with positive impact factor and the fourth one was accepted for publication. From this point of view, the author proved that he is highly capable of original scientific contribution in his chosen field.

In the first article called "Using Monte Carlo simulation to calculate match importance: The case of English Premier League" and published in Journal of Sports Economics, the author introduces and describes a new method of calculating match importance. The concept of match importance is an important one in economics of sports, because existing literature shows that there is a link between match importance and team performance, as well as spectators' attendance. First, given the more general audience of the dissertation compared to a specialized journal, I would appreciate a little more detailed description of sources of data (e.g. if Trefik is a usually used data source in academic papers, what is the main target audience of the database, etc.).

The article estimates probabilities of match results and then uses them to estimate probabilities of final ranks of individual teams. In the last step, a link between some chosen outcome (such as the final rank) and a particular match result is calculated. This link is then called the match importance. The second half of the article is dedicated to verification of the predictions. However, the author deals only with verification of the first two steps – match results predictions and final ranks predictions. The final match importance is only compared to importance obtained using other previously published methods. I understand that match importance is only calculated from the predictions. But I am not sure whether realistic match results prediction has to be a sufficient condition for correct match importance values, which is what the author claims on p. 18, if I understand it correctly. I suggest that the author addresses this issue either by providing some test of his match importance values (e.g. correlation with match attendance?) or by improving his explanation in subsection 1.6. The footnote on p. 22 seems to go in the right direction in this sense.

In the second article "The impact of playoffs on seasonal uncertainty in the Czech ice hockey Extraliga" published also in Journal of Sports Economics, the author estimates the impact of playoffs on probability of winning the whole Extraliga. This, again, should have an impact on spectators' attendance because higher seasonal uncertainty means higher number of important matches. The article is, overall, very nicely written and I have only a couple of remarks. First, on p. 30 the author states that "the major reason for using playoffs is unlikely to be to directly increase attendance … because the total attendance per week during the regular season is actually slightly higher than during the postseason." However, the author doesn't seem to take into account the fact that the mere existence of playoff stage means that a large number of teams is able to compete for the possibility to qualify and have a chance of winning the whole tournament. This has an impact on match importance (as the author shows in the first article) and, therefore, has a potential of increasing match attendance during the regular season. In other words, a simple comparison of total attendance during playoffs and during regular seasons may be misleading. Second, I would appreciate a more thorough explanation of the choice of values (probably the average number of expected goals?) on p. 35.

In the third article "Does match uncertainty increase attendance? A non-regression approach" Jiří Lahvička shows why empirical evidence regarding the question posed in the title of the article is mixed. His hypothesis is that the inconsistency of findings can be explained by wrong specification of regression equations. Then, using English Championship data, he introduces a non-regression approach of testing the uncertainty-attendance link. The majority of my suggestions have been already incorporated into the text. But I still wonder if it would be possible to add a couple of robustness checks that would help to rule out endogeneity problem? E.g. what if both quality (kind of an explanatory variable) and attendance (kind of an explained variable) depend on geographical home location of the teams? Also, I would suggest that the author sticks to the standard structure of a paper and add a Conclusion section.

The fourth article "What causes the favorite-longshot bias? Further evidence from tennis" uses a large dataset to analyze possible causes of the fact that bets on favorites tend to have a higher expected value than bets on longshots. The article was published in Applied Economics Letters. First, I suggest the author uses a more standard layout of tables with regression results (include no. of observations, R^2, etc.). Footnote 71 on p. 72 states that the author used heteroskedasticity-robust standard errors. However, I suspect that the observations are not fully independent, which would cause the standard errors to be biased downwards. To explain what I mean: The author uses betting odds to calculate an implied probability of a player winning. This implied probability is then used as an explanatory variable in regressions explaining match result. But during one tournament, the betting odds on individual matches of one specific player are hardly independent – they depend on the current form of the player, their physical and mental fitness, etc. A win in the first round of a tournament can be expected to have an impact on betting odds in the following rounds. Therefore, I suggest the author adds a robustness check where he would cluster variance on the tournament level to allow for intra-tournament correlation of observations, and check if the significance of results changes. Regarding the lack of statistical significance of *ImpliedProbability* variable in Table 19 on p. 73 – what is the number of observations with *LowerRank*, *LaterRound*, as well as *HighProfile* equal to zero? Is there enough observations (and therefore variance) to identify a relationship?

His last article "The Fibonacci strategy revisited: Can you really make money by betting on soccer draws?" was accepted for publication in the meantime between the so called small defense and the ordinary defense in the Journal of Gambling Business and Economics. In this article author uses simulated and real data and shows that the strategy leads to losing money. This article is very clear and concise and I would only suggest that the author tries to explain more the possible causes of previous positive results. Does he think that it is, indeed, solely the low number of trials, or could there be also a selectivity bias present?

I suppose that Mr. Jiří Lahvička has complied with all the necessary prerequisites required for the ordinary defense at the Faculty of Economics at the University of Economics, Prague. I recommend this dissertation for successful defense before the respective examination committee for dissertations defense.

Prague, October 6, 2014

Jiří Schwarz