

Report on Bachelor / Master Thesis

Institute of Economic Studies, Faculty of Social Sciences, Charles University

Student:	Sebastian Černý
Advisor:	Luboš Hanus
Title of the thesis:	xG Statistics in Football Matches: Predictions and Betting

OVERALL ASSESSMENT (provided in English, Czech, or Slovak):

Please provide a short summary of the thesis, your assessment of each of the four key categories, and an overall evaluation and suggested questions for the discussion. The minimum length of the report is 300 words.

Short summary

The thesis takes five years of football matches from five European leagues and makes predictions on the results of the last two seasons, which are then used in various betting systems to produce an economic outcome. In this exercise, the thesis uses a relatively new data and statistical measure of sporting performance in predicting matches, called expected goals (xG). The results are compared first in terms of prediction accuracy and then in terms of return on betting. Using a logistic regression model, the thesis shows positive returns from this data-driven approach when betting on matches where there are discrepancies between the predicted outcome and the odds offered by betting companies.

Contribution

The contribution of the paper is appropriate and fits well in the literature. A relatively new statistic called expected goals (xG) has been introduced in the sports data analysis literature, but there are not many studies reporting the predictive power of xG. The author addresses this issue using publicly available data. Two sources of xG data are used. The paper also provides a comparison and economic evaluation of models using different input variables. The thesis shows economic gains in return on investment (ROI) when the prediction and betting strategy is correctly chosen.

Methods

Since predicting the outcome of a game is a classification problem and the outcome is a binary variable, the author correctly chooses logistic regression to perform the task. He describes the method well and adequately, reports and tests the assumptions of the model and describes them correctly. To make the out-of-sample predictions valid, the estimation is done using an expanding window for each new week of the test season period. The formulation of the xG statistics and related variables is also shown. In the economic analysis, the various betting strategies and criteria are covered and explained.

Literature

The author uses the literature correctly. The author is aware of the approaches used in the literature and uses the sources correctly. The bibliography lists all the sources used in the text, and the number is sufficient.

Manuscript form

The form of the manuscript is correct and conforms to IES standards. The text is well written and structured, although the author sometimes "over-explains" himself, but this is not a serious problem. All tables and figures are clearly presented. Figures are coherent and well depicted.

Report on Bachelor / Master Thesis

Institute of Economic Studies, Faculty of Social Sciences, Charles University

Student:	Sebastian Černý
Advisor:	Luboš Hanus
Title of the thesis:	xG Statistics in Football Matches: Predictions and Betting

Overall evaluation and suggested questions for the discussion during the defense

To sum up, I am satisfied with Mr Černý's bachelor thesis, which he approached the task with enthusiasm and a great deal of interest. The thesis fulfils the requirements for a bachelor thesis at the IES, Faculty of Social Sciences, Charles University. In my opinion, it is a great work in the sense of IES bachelor theses. I recommend it for defense and propose a grade A.

The results of the Turnitin analysis do not indicate significant text similarity with other available sources.

Q1: *What would improve the statistical and economic results of your analysis?*

Q2: *What are the differences between your betting criteria and why does the Kelly criterion perform best?*

SUMMARY OF POINTS AWARDED (for details, see below):

CATEGORY	POINTS
<i>Contribution (max. 30 points)</i>	27
<i>Methods (max. 30 points)</i>	30
<i>Literature (max. 20 points)</i>	20
<i>Manuscript Form (max. 20 points)</i>	18
TOTAL POINTS (max. 100 points)	95
GRADE (A – B – C – D – E – F)	A

NAME OF THE REFEREE: *Luboš Hanus*

DATE OF EVALUATION: *May 6, 2024*

Digitálně podepsáno (6. 5. 2024)
Luboš Hanus

Referee Signature

EXPLANATION OF CATEGORIES AND SCALE:

CONTRIBUTION: *The author presents original ideas on the topic demonstrating critical thinking and ability to draw conclusions based on the knowledge of relevant theory and empirics. There is a distinct value added of the thesis.*

METHODS: *The tools used are relevant to the research question being investigated, and adequate to the author's level of studies. The thesis topic is comprehensively analyzed.*

LITERATURE REVIEW: *The thesis demonstrates author's full understanding and command of recent literature. The author quotes relevant literature in a proper way.*

MANUSCRIPT FORM: *The thesis is well structured. The student uses appropriate language and style, including academic format for graphs and tables. The text effectively refers to graphs and tables and disposes with a complete bibliography.*

Overall grading:

TOTAL	GRADE
91 – 100	A
81 - 90	B
71 - 80	C
61 – 70	D
51 – 60	E
0 – 50	F