

# Выявление и оценка факторов, влияющих на результат футбольного матча

Манукян А. А.

Национальный исследовательский университет «Высшая школа экономики»

(Россия, Москва)

sargis-97@mail.ru

Катышев П. К.

Департамент прикладной экономики факультета экономических наук НИУ ВШЭ

(Россия, Москва)

pkatyshev@hse.ru

**Ключевые слова:** пробит–модель упорядоченного, вероятность исхода футбольного матча, моделирование выигрышной маржи, оценку прогноза

## Ֆուտբոլային խաղի ելքի վրա ազդող գործոնների բացահայտում և գնահատում

Մանուկյան Ա. Ա.

Ազգային Հետազոտական համալսարան Բարձրագույն Տնտեսագիտության Դպրոց

(Ռուսաստան, Մոսկվա)

sargis-97@mail.ru

Պ. Կ. Կատիշեվ

Ազգային Հետազոտական համալսարան Բարձրագույն Տնտեսագիտության Դպրոց

(Ռուսաստան, Մոսկվա)

pkatyshev@hse.ru

**Անփոփում:** Այս շխատանքի ակնկալվող արդյունքն է՝ ստեղծել մի մոդել, որը թույլ կտա ստանալ ֆուտբոլային խաղի արդյունքները կանխատեսելու լավագույն հավանականության գնահատականներ: Մենք կուսումնասիրենք որոշակի ժամանակահատվածի համար Անգլիական Պրեմիեր Լիգայի խաղերին վերաբերող պաշտոնական կայքերում առկա տեղեկատվությունը: Այս մոդելը կկառուցվի պատմական տվյալների հիման վրա: Ապա կկատարվի մոդելի փորձարկում որոշակի ժամանակահատվածի համար, և եթե այն հաջող լինի, ապա կօգտագործվի հետագա խաղերի համար: Մենք կփորձենք հասկանալ, թե արդյոք հնարավոր է խաղալ բուքմեյքերի դեմ:

**Վճռորոշ բառեր՝** կարգավորված պրոբիտ-մոդել, ֆուտբոլային խաղի ելքի հավանականություն, հաղթող մարժայիմոդելավորում, կանխատեսման գնահատական

## Identification and Estimation of Factors Affecting the Football Match Outcome

Manukyan A. A.

National Research University Higher School of Economics (Russia, Moscow)

sargis-97@mail.ru

Katyshev P. K.

Faculty of Economic Sciences / Department of Applied Economics HSE (Russia, Moscow)

pkatyshev@hse.ru

**Abstract:** The expected result of this work will be creating a model which will allow getting good probability estimates to forecast football match results. We will study the information available on the official websites which relate to the English Premier League matches for a specific period of time. The model will be based on the historic data. This will follow with a check up for a specific time period, and if it proves to be successful then it will be used for future matches.

We will try to make out whether it is possible to play against the book maker.

**Keywords:** ordered probit-model, the probability of the football match outcome, modelling the winning margin, forecast estimate

### Данные

В данной работе строятся модели для предсказания результатов футбольного матча. Выб-

ранные данные определяют качество модели, поэтому им уделялось особое внимание. Мы рассматриваем сезон 2015-2016 годов Английской

Премьер Лиги. Один матч — это одно наблюдение. Количество наблюдений  $n = 342$ .

Объясняющие переменные:

HS – Home team Shots; AS – Away team Shots; HST – Home team Shots on target; AST – Away team Shots on target; Home\_A – Home team Attack parameter; Home\_M – Home team Midfield parameter; Home\_D – Home team Defense parameter; Away\_A – Away team Attack parameter; Away\_M – Away team Midfield parameter; Away\_D – Away team Defense parameter; H – Home advantage; TW\_Home – Home team Transfer weight – общая стоимость игроков домашней команды; TW\_Away – Away team Transfer weight – общая стоимость игроков гостевой команды

Так же иногда в место Home\_A, Home\_M, Home\_D, Away\_A, Away\_M, Away\_D иногда используем D\_A, D\_M, D\_D, где

D\_A – Difference Attack parameter = Home\_A - Away\_A

D\_M – Difference Midfield parameter = Home\_M - Away\_M,

D\_D – Difference Defense parameter = Home\_D - Away\_D

Данные об ударах и ударах в створ брались из сайта <http://www.football-data.co.uk/>. Параметры атаки, полузащиты и защиты брались из официального сайта FIFA. Естественно, эти параметры для определенного сезона, потому что для сезона  $n+1$  они будут существенно различаться из-за новых трансферов или из-за вариации характеристики игроков.

Данные о трансферных стоимостях были заимствованы с сайта <https://www.transfermarkt.ru>. Мы включили в модель трансферные стоимости всех игроков для всех команд, потому что стоимость сама по себе является экспертной оценкой всех характеристик игрока. Когда мы рассматриваем параметры атаки или защиты, речь идет об общей команде. Здесь больше оценивается взаимопонимание игроков, выбранная тактика команды, то есть командная сила. С другой стороны, необходимо учесть индивидуальные качества и умение игроков, потому что в трудные моменты, когда не идет игра команды, эти личности берут ответственность за игру на себя. И очень много случаев, когда с помощью таких игроков команда смогла победить.

Так же модель будет включать параметр  $h$ , который показывает преимущество своего поля.

Чтобы правильно оценить преимущество своего поля, нам нужно моделировать способности команд. Мы используем модель аналогичную той, которую использовали Stefani and Clarke (1992).

Выигрышная маржа между командами  $i$  и  $j$ , которые играли на поле команды  $i$  имеет вид:

$y_{ij} = a_i - a_j + h_i + \varepsilon_{ij}$ , где  $a_i$  и  $a_j$  – параметры силы команды,  $i$  и  $j$  соответственно,  $h_i$  – преимущество команды  $i$  на своем поле,  $\varepsilon_{ij}$  – ошибка.

### Описание модели

Мы используем пробит–модель упорядоченного, чтобы предсказать исход футбольного матча. Модель позволяет находить оценки вероятности для домашней победы, ничьи и победы гостей в каждом матче. Она включает скрытую переменную, удовлетворяющую стандартному уравнению линейной регрессии

$$y_i^* = x_i' \beta + \varepsilon_i, \varepsilon_i \sim N(0, 1)$$

где ошибки  $\varepsilon_i$  не зависят от всех векторов  $x_i$ , поскольку  $\varepsilon_i$  стандартно нормально распределено и  $y_i^*$  линейно зависит от  $\varepsilon_i$ , то  $y_i^*$  также нормально распределено. У каждой игры  $i$  есть три исхода: победа домашней команды, победа гостей и ничья. Будем обозначать эти исходы  $y_i = 1, y_i = -1, y_i = 0$ , соответственно. Связь между скрытой переменной  $y_i^*$  и наблюдаемыми значениями  $y_i$  определяется следующими соотношениями:

- (1)  $y_i = 1$ , если  $y_i^* > \mu_2$
- (2)  $y_i = 0$ , если  $y_i^* < \mu_2$  и  $y_i^* > \mu_1$
- (3)  $y_i = -1$ , если  $y_i^* > \mu_1$

где  $\mu_1 < \mu_2$  — некоторые границы (которые, как правило, не заданы априори и подлежат оцениванию). Имея значения коэффициентов  $\mu_1, \mu_2$  и  $\beta$ , мы можем оценить вероятность выигрыша домашней команды, ничьи и победа гостевой команды. Вероятность выигрыша домашней команды может быть рассчитана с помощью уравнения (1), для расчета вероятности выигрыша гостевой команды мы используем уравнения (3), а вероятность ничьи с использованием уравнения (2) и используя вероятность выигрыша домашней и гостевой команды.

$$P[y_i = 1] = P(\varepsilon_i > \mu_2 - x_i' \beta) = 1 - \Phi(\mu_2 - x_i' \beta)$$

$$P[y_i = -1] = P(\varepsilon_i < \mu_1 - y_i^*) = \Phi(\mu_1 - x_i' \beta)$$

$$P[y_i = 0] = \Phi(\mu_2 - x_i' \beta) - \Phi(\mu_1 - x_i' \beta) = 1 - P[y_i = 1] - P[y_i = -1]$$

Анализ данных для нахождения модели

Для того, чтобы анализировать данные и найти подходящую модель, мы используем программу EViews. Так как все данные нам уже известны, то мы просто подставляем их в программу EViews.

Сначала в модель включаем все объясняющие переменные (которые мы показали в части «Данные»), за исключением переменных Home\_A, Home\_M, Home\_D, Away\_A, Away\_M, Away\_D, потому что мы используем разницу между ними. То есть, модель будет включать

переменные Y, AS, AST, HS, HST, D\_A, D\_M, D\_D, H, TW\_Away, TW\_Home.

Табл.1

Dependent Variable: Y  
 Method: ML - Ordered Probit (Newton-Raphson / Marquardt steps)  
 Date: 05/09/18 Time: 20:23  
 Sample: 1 342  
 Included observations: 342  
 Number of ordered indicator values: 3  
 Convergence achieved after 5 iterations  
 Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
AS	0.002164	0.020345	0.106343	0.9153
AST	-0.210503	0.041323	-5.094055	0.0000
HS	-0.078444	0.017241	-4.549979	0.0000
HST	0.268927	0.036120	7.445464	0.0000
D_A	0.049777	0.025976	1.916292	0.0553
D_M	0.075058	0.040772	1.840915	0.0656
D_D	-0.045568	0.037008	-1.231322	0.2182
H	0.180064	0.130363	1.381257	0.1672
TW_AWAY	0.000908	0.001079	0.842006	0.3998
TW_HOME	-0.000430	0.001073	-0.400924	0.6885

  

Limit Points				
LIMIT_0:C(11)	-1.105039	0.353190	-3.128737	0.0018
LIMIT_1:C(12)	-0.179431	0.349595	-0.513255	0.6078

  

Pseudo R-squared	0.168850	Akaike info criterion	1.869463
Schwarz criterion	2.004017	Log likelihood	-307.6781
Hannan-Quinn criter.	1.923065	Restr. log likelihood	-370.1835
LR statistic	125.0109	Avg. log likelihood	-0.899644
Prob(LR statistic)	0.000000		

Анализ показывает, что на уровне 5% значимы только переменные AST, HS, HST. Это означает, что результат футбольного матча можно объяснить, имея только данные об ударах и ударах в створ домашней команды и об ударах в створ гостевой команды. Это для нас неприемлемо, потому что наша модель должна полагаться на то, что именно голы в большой степени влияют на исход матча, а не удары.

Это означает, что мы будем рассматривать несколько моделей, где каждый раз будет включена определенная группа переменных. Высокое значения P в большинстве данных объясняется тем, что некоторые данные сильно коррелированы между собой.

Сортируем несколько переменных по важности влияния на исход матча. С точки зрения голов, которые сильно влияют на исход матча, мы отберем переменную H, так как H вычисляется именно на основе забитых и пропущенных голов. Можно сказать, что H в большой степени несет информацию именно о команде. Тогда с другой стороны, нам надо выбрать переменную которое демонстрирует информацию об игроках команды. Такими переменными у нас являются TW\_Away, TW\_Home.

Получается, что первая отобранная модель будет включать в себя переменные TW\_Away, TW\_Home, H.

Табл.2

Dependent Variable: Y  
 Method: ML - Ordered Probit (Newton-Raphson / Marquardt steps)  
 Date: 05/09/18 Time: 19:32  
 Sample: 1 342  
 Included observations: 342  
 Number of ordered indicator values: 3  
 Convergence achieved after 4 iterations  
 Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
H	0.241548	0.118710	2.034776	0.0419
TW_AWAY_	-0.001377	0.000451	-3.053405	0.0023
TW_HOME_	0.001385	0.000452	3.065580	0.0022

  

Limit Points				
LIMIT_0:C(4)	-0.382437	0.174097	-2.196686	0.0280
LIMIT_1:C(5)	0.368996	0.174016	2.120475	0.0340

  

Pseudo R-squared	0.031504	Akaike info criterion	2.125856
Schwarz criterion	2.181920	Log likelihood	-358.5213
Hannan-Quinn criter.	2.148190	Restr. log likelihood	-370.1835
LR statistic	23.32441	Avg. log likelihood	-1.048308
Prob(LR statistic)	0.000035		

Можно увидеть, что все переменные значимы на уровне 5%. Коэффициент H-а ровняется на 0,241548, что означает, что модель является логически правильной. Коэффициент TW\_Away - а ровняется на - 0,001377. Так как TW\_Away - это трансферная стоимость игроков гостевой команды, то логично, что коэффициент примет отрицательное значение. Аналогично также как для H-а, так и для TW\_Home, коэффициент принимает положительное значение.

Модель принимает вид  
 $P[y_i = 1] = P(\epsilon_i > \mu_2 - x' \beta) = 1 - \Phi(0.368996 - x' \beta_i^*)$

$P[y_i = -1] = P(\epsilon_i < \mu_1 - y_i^*) = \Phi(-0.382437 - x' \beta_i^*)$

$P[y_i = 0] = 1 - P[y_i = 1] - P[y_i = -1]$

Для данной модели, можно получить оценку прогноза.

Табл.3

Dep. Value	Obs.	Correct	Incorrect	% Correct	% Incorrect
-1	104	43	61	41.346	58.654
0	95	0	95	0.000	100.000
1	143	114	29	79.720	20.280
Total	342	157	185	45.906	54.094

Посмотрим на рисунок. Первый столбец (Dep. Value), это наши  $y_i = \{1, 0, -1\}$ . Второй столбец (Obs.) показывает количество наблюдений. То есть, 104 раза матч завершился победой гостевой команды, 95 раз результатом вничью и 143 раза победой хозяев. Третий столбец (Correct) показывает сколько раз модель правильно предсказала исход матча имея определенное количество наблюдений. Аналогично столбец (Incorrect) показывает сколько раз модель ошибалась. Можно увидеть, что модель на 41.346% успешно предсказала, что матч закончится с победой

гостевой команды. На 79,72% успешно предсказала, что результатом матча будет победа домашней команды. Но модель полностью неправильно предсказала результат вничью. Проблема в том, что модель не отображает вероятности для исхода. Внутри программы она создает оценочные вероятности исхода, после чего выбирает наиболее вероятный исход и только отображается на таблице в виде единицы.

Посмотрим данную проблему на примере.

Табл.4

Y	Date	Home Team	Away Team	1/P(Y=-1)	1/P(Y=0)	1/P(Y=1)
-1	08.08.2015	Bournemouth	Aston Villa	2,2	3,7	4
0	08.08.2015	Chelsea	Swansea	1,57	4,94	7,8
0	08.08.2015	Everton	Watford	2,78	3,32	2,88
1	08.08.2015	Man. United	Tottenham	1,61	3,72	5,92
-1	08.08.2015	Norwich	Crystal Palace	3,77	2,95	2,35
0	09.08.2015	Newcastle	Southampton	2,63	3,31	2,74

$P(Y_i)$  – это вероятность каждого исхода. Букмекеры предлагают коэффициенты в виде  $1/P(Y_i)$ . То есть, чем больше коэффициент, тем меньше вероятность. Смотрим на матч между командами Bournemouth и Aston Villa. Модель выбирает значение  $1/P(Y_i) = 2,2$ , так как он предполагает наиболее вероятный исход. Матч закончился победой команды Aston Villa. Это означает, что предсказание модели неправильное. Собственно, и прогнозы букмекеров были неверными. Для таких моделей неправильное предсказание – нормальное явление, так как они не дают 100% вероятность. В таблице 4 существует 6 наблюдений, и модель предсказала верно только 2 из них. Но также у нас 3 ничейных результата, и если просто вычитать их оттуда, то в конечном счете, модель допускала ошибку только 1 раз. Если посмотреть на матч между Everton и Watford, то увидим, что все 3 вероятности очень близки друг к другу. Очень мало случаев, когда вероятность ничьи превышает вероятности побед.

Таким образом, мы показываем, что все 95 неправильных предсказаний касательно ничьей являются не такими уж ошибочными. Мы не можем просто вычитать линию ничейных наблюдений и рассмотреть таблицу в виде

Табл.5

Dep. Value	Obs.	Correct	Incorrect	% Correct	% Incorrect
-1	104	43	61	41.346	58.654
1	143	114	29	79.720	20.280

Проблема в том, что за каждое неправильное предсказание результата вничью, модель выбрала либо  $y_i = 1$ , либо  $y_i = -1$ . То есть, в данной таблице существуют дополнительные невидимые предсказания, которые влияют на процент корректных и не корректных предсказаний.

Вторая модель будет включать переменные Y, AST, Home\_A, Away\_A, H. То есть предполагается, что исход матча может быть объяснен с помощью параметров атаки домашней и гостевой команды, преимуществом, которым владеет домашняя команда, играя на своем поле, а также с помощью ударов в створ гостевой команды.

Табл.6

Dependent Variable: Y  
 Method: ML - Ordered Probit (Newton-Raphson / Marquardt steps)  
 Date: 05/11/18 Time: 16:58  
 Sample: 1 342  
 Included observations: 342  
 Number of ordered indicator values: 3  
 Convergence achieved after 6 iterations  
 Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
HOME_A	0.057317	0.020389	2.811153	0.0049
AWAY_A	-0.041114	0.020068	-2.048778	0.0405
H	0.246384	0.122031	2.019035	0.0435
AST	-0.190776	0.031260	-6.102886	0.0000

  

Limit Points				
LIMIT_0:C(5)	0.109978	2.256583	0.048737	0.9611
LIMIT_1:C(6)	0.921318	2.256846	0.408233	0.6831

  

Pseudo R-squared	0.084404	Akaike info criterion	2.017185
Schwarz criterion	2.084462	Log likelihood	-338.9386
Hannan-Quinn criter.	2.043986	Restr. log likelihood	-370.1835
LR statistic	62.48995	Avg. log likelihood	-0.991048
Prob(LR statistic)	0.000000		

Можно увидеть, что на уровне 5% все переменные значимы. Знаки коэффициентов логически правильны. Для домашней команды коэффициент  $\beta$  принимает знак «+», а для гостевой команды «-». Значимость AST объясняется тем, что гостевая команда обладает отрицательным преимуществом из-за существования H. Таким образом, контроль мяча, в основном, принадлежит домашней команде. Гостевые команды побольше играют на контратаках, и поэтому, в среднем, каждый выход у ворот домашней команды является более опасным, чем у ворот гостевой команды. Отсюда и каждый удар в створ может стать голевым для гостевой команды.

В таблице 7 приведены данные о матче Арсенал против Вест Хем. В матче домашняя команда ударила по воротам 22 раза, из них 6 раз попала в створ. Гостевая команда ударила по воротам 8 раз, попав в створ 4 раза. Преимущество по общему количеству ударов у Арсенала значительно большое, но матч закончился с победой Вест Хем-а со счетом 0:2. Статистика Вест Хем-а (удары 8, в створ 4, голы 2) против Арсенала (удары 22, в створ 6, голы 0) говорит о том,

удары в створ Вест Хем-а более опасны, чем удары в створ Арсенала. Здесь не говорится о том, что удары в створ домашней команды не имеют значение (они даже очень значимы), а мы показываем, что AST более «голевые», чем HST.

Табл. 7

Y	Date	HomeTeam	AwayTeam	HS	AS	HST	AST
-1	09.08.2015	Arsenal	West Ham	22	8	6	4

Табл. 8

Dep. Value	Obs.	Correct	Incorrect	% Correct	% Incorrect
-1	104	58	46	55.769	44.231
0	95	0	95	0.000	100.000
1	143	116	27	81.119	18.881
Total	342	174	168	50.877	49.123

Модель на 55,769% правильно предсказала исход гостевой команды для 104 наблюдений, всегда ошибалась на счет ничьи и на 81,119% правильно предсказала победы домашней команды. Общая корректность прогноза составляет 50,877%.

Теперь добавим в ту же модель переменные HS, AST, AS.

Табл. 9

Variable	Coefficient	Std. Error	z-Statistic	Prob.
HOME_A	0.064480	0.022045	2.924884	0.0034
AWAY_A	-0.047144	0.021252	-2.218340	0.0265
H	0.238126	0.126627	1.880525	0.0600
AST	-0.210944	0.040958	-5.150291	0.0000
HS	-0.077520	0.017161	-4.517317	0.0000
HST	0.262640	0.035665	7.364028	0.0000
AS	0.004327	0.020179	0.214453	0.8302

Видно, что из модели надо исключить переменную AS.

С чем связан отрицательный знак перед HS? В целом, количество ударов домашней команды в большинстве случаев превосходит количество ударов гостевой команды. Тут имеет место 2 случая. Во-первых, когда домашняя команда очень сильно превосходит противника по качеству, доминирование продолжается тогда, когда даже счет матча была бы 3:0. Во-вторых, когда домашняя команда проигрывает. Тогда, как правило, гостевая команда всеми силами обороняется, а домашняя большими силами атакует. Концентрация на половине гостей такая большая, что домашняя команда во многом, не находя пространства, начинает просто издали ударить по воротам. Получается, что каждый удар с низкой вероятностью может попасть в воротах. Поэтому, количество HS становится в течение

матча таким большим. Модель показывает, что в сезоне 2015-2016 по большому счету были именно такие матчи. И знак «-» говорит о том, что чем больше HS, тем вероятнее, что победит гостевая команда.

Табл. 10

Dependent Variable: Y  
Method: ML - Ordered Probit (Newton-Raphson / Marquardt steps)  
Date: 05/11/18 Time: 18:26  
Sample: 1 342  
Included observations: 342  
Number of ordered indicator values: 3  
Convergence achieved after 4 iterations  
Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
HOME_A	0.064380	0.022039	2.921250	0.0035
AWAY_A	-0.047279	0.021241	-2.225860	0.0260
H	0.237286	0.126556	1.874949	0.0608
AST	-0.205707	0.032850	-6.262101	0.0000
HS	-0.078196	0.016868	-4.635652	0.0000
HST	0.262453	0.035642	7.363614	0.0000

Limit Points

Variable	Coefficient	Std. Error	z-Statistic	Prob.
LIMIT_0:C(7)	0.135249	2.331886	0.058000	0.9537
LIMIT_1:C(8)	1.053347	2.331790	0.451733	0.6515

Pseudo R-squared 0.163550 Akaike info criterion 1.857545  
Schwarz criterion 1.947248 Log likelihood -309.6401  
Hannan-Quinn criter. 1.893280 Restr. log likelihood -370.1835  
LR statistic 121.0868 Avg. log likelihood -0.905381  
Prob(LR statistic) 0.000000

Табл. 11

Dep. Value	Obs.	Correct	Incorrect	% Correct	% Incorrect
-1	104	71	33	68.269	31.731
0	95	6	89	6.316	93.684
1	143	115	28	80.420	19.580
Total	342	192	150	56.140	43.860

Посмотрим на оценку прогноза. Модель на 68,269% правильно предсказала победы гостевой команды. На 80.420% правильно предсказала победы домашней команды. Можно так же увидеть, что 6 раз модель предсказала вероятность ничьи правильно. Это очень хороший результат, так как мы уже говорили, что вероятность ничьи в большинстве случаев выше вероятности побед. На этот раз общая корректность модели составляет 56,140%. По всем показателям модель\_2 лучше, чем модель\_1.

Третья модель будет включать переменные HOME\_M, AWAY\_M, AST, HST, HS.

Видно, что на уровне 5% переменная AWAY\_M незначима. Поэтому вместо переменных HOME\_M, AWAY\_M будем включать в модель переменную D\_M.

На уровне 5% все переменные значимы. Знак коэффициента D\_M положительный. Это логично, потому что изначально обозначили  $D_M = HOME_M - AWAY_M$ . То есть, если  $D_M > 0 \Rightarrow HOME_M > AWAY_M \Rightarrow$  вероятность домашней победы становится больше.

Табл.12

Variable	Coefficient	Std. Error	z-Statistic	Prob.
HOME_M	0.064470	0.019389	3.325082	0.0009
AWAY_M	-0.029758	0.019216	-1.548653	0.1215
AST	-0.209361	0.032690	-6.404408	0.0000
HS	-0.075584	0.016465	-4.590453	0.0000
HST	0.259183	0.035490	7.303045	0.0000
Limit Points				
LIMIT_0:C(6)	1.317131	2.035184	0.647180	0.5175
LIMIT_1:C(7)	2.229824	2.036765	1.094787	0.2736

Табл.13

Dependent Variable: Y  
 Method: ML - Ordered Probit (Newton-Raphson / Marquardt steps)  
 Date: 05/12/18 Time: 01:01  
 Sample: 1 342  
 Included observations: 342  
 Number of ordered indicator values: 3  
 Convergence achieved after 5 iterations  
 Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
D_M	0.046976	0.013881	3.384238	0.0007
AST	-0.204102	0.032395	-6.300318	0.0000
HS	-0.074555	0.016443	-4.534031	0.0000
HST	0.258899	0.035465	7.300202	0.0000
Limit Points				
LIMIT_0:C(5)	-1.302699	0.248225	-5.248057	0.0000
LIMIT_1:C(6)	-0.393147	0.241702	-1.626577	0.1038
Pseudo R-squared	0.158058	Akaike info criterion	1.857737	
Schwarz criterion	1.925014	Log likelihood	-311.6730	
Hannan-Quinn criter.	1.884539	Restr. log likelihood	-370.1835	
LR statistic	117.0210	Avg. log likelihood	-0.911325	
Prob(LR statistic)	0.000000			

Табл.14

Dep. Value	Obs.	Correct	Incorrect	% Correct	% Incorrect
-1	104	69	35	66.346	33.654
0	95	14	81	14.737	85.263
1	143	115	28	80.420	19.580
Total	342	198	144	57.895	42.105

Модель на 66,346% правильно предсказала победы гостевой команды. На 80,42% правильно предсказала победы домашней команды и на 14,737% ничью. Данная модель показывает самый хороший результат с точки зрения предсказания результата вничью. По сравнению с моделью\_2, модель\_3 на 2 наблюдения больше ошибалась в предсказаниях победы гостевой команды, на 8 наблюдений больше корректно спрогнозировала результаты вничью и предсказала победы домашней команды ровно столько, сколько предсказала модель\_2. Общая корректность модели составляет 57,895%, что является лучшим результатом среди всех моделей.

### Заключение

Данная работа дает понять, что для определения исхода матча не нужно только ориентироваться на силы команд или на место, которое они занимают в таблице. Существует ряд факторов, которые сильно влияют на исход матча, как например, преимущество своего поля. Ведь существуют команды, которые за сезон в домашних матчах проигрывают около одного или двух раз, но при этом, играя в гостях выигрывают лишь 3 раза. Для определения качества игроков всегда можно посмотреть на их трансферную стоимость, ведь они являются оценками характеристики игроков. Не количество ударов, а количество ударов в створ показывает большое влияние на исход матча. Данная работа доказывает теорию относительно того, что не всегда коэффициенты, предлагаемые БК, отображают истинную вероятность исхода событий, потому что максимизация прибыли приводит к изменениям коэффициентов.

Мы получили три разные модели для прогнозирования. Несмотря на то, что модель\_3 показывает наиболее хорошие результаты, всё же не стоит прогнозировать матч только на основе этой модели. Лучше прогнозировать по трем моделям и взять среднее значение, потому что эти 3 модели включают разные переменные, и мы не можем предсказать какая из них для конкретного события будет более точной.

### Список использованной литературы

1. Maher M. J. (1982). Modelling association football scores. *Statistica Neerlandica*, Volume36, Issue3, Pages 109-118.
2. Stephen R. Clarke and John M. Norman (1995). Home Ground Advantage of Individual Clubs in English Soccer. *Journal of the Royal Statistical Society. Series D (The Statistician)*, Vol. 44, No. 4, pp. 509-521.
3. Van der Zaan (2017). Predicting outcome of soccer matches in order to make money with betting.
4. Richard Pollard (1986). Home advantage in soccer: A retrospective analysis. *Journal of Sports Sciences*, 4(3):237-48.
5. Snyder and Purdy (1985). The Home Advantage in Collegiate Basketball. *Sociology of Sport Journal*, Volume 2, Issue 4, Pages 352-356.
6. Mark J. Dixon, Stuart G. Coles (1997). Modelling Association Football Scores and Inefficiencies in the Football Betting Market.