ZÁPADOČESKÁ UNIVERZITA V PLZNI
FAKULTA APLIKOVANÝCH VĚD
KATEDRA MATEMATIKY

Diplomová práce

Modelování a odhadování výsledků hokejových utkání

Plzeň, 2018 Bc. Ivana Gabrišková

Prohlášení

Prohlašuji, že jsem diplomovou práci vypracovala samostatně a výhradně s použitím literatury a pramenů uvedených v seznamu.

V Plzni dne 18.5.2018

…………………

Ivana Gabrišková

Poděkování

Ráda bych poděkovala vedoucímu mé diplomové práce Ing. Patrice Markovi, Ph.D. za vedení, odborné rady a čas, který mi věnoval při zpracování této práce.

Dále bych chtěla poděkovat své mamince, příteli a rodině, která mě během mého studia neustále podporovala a bez nichž bych to nezvládla.

Abstrakt

Tato diplomová práce se zabývá modelováním a odhadováním výsledků hokejových utkání a poté využitím odhadů při sázení proti sázkovým kancelářím. Práce popisuje modely používané pro odhadování výhry domácího týmu, remízy nebo výhry hostujícího týmu. Dále se věnuje Dixon-Colesovým modelům upraveným podle článku [3]. Těmito modely jsou odhadovány výsledky zápasů české, polské a národní hokejové ligy (NHL) v sezóně 2015/2016. Modely jsou porovnané s triviálními schématy sázení.

Klíčová slova: Hokej, Poissonovo rozdělení, odhad hokejových výsledků, sázení

Abstract

This diploma thesis concerns with modeling and estimating results of hockey matches and use of estimation for betting against bookmakers. The thesis describes models used for estimating winning of home team, a draw or away team win. The thesis deals about Dixon-Coles models arranged according to the article [3]. Results of matches of Czech, Polish and National Hockey League in season 2015/2016 are estimated by using these models. Models are compared with trivial schemas of betting.

Key words: Hockey, Poisson distribution, estimate of hockey results, betting

OBSAH

1 Úvod 1

2 DATA 2

2.1 Extraliga (CZE) 2

2.2 Ekstraliga (POL) 3

2.3 NHL 4

3 Statistické pojmy a metody 5

3.1 Poissonovo rozdělení 5

3.2 P-hodnota 5

3.3 Chí-kvadrát test dobré shody 5

3.4 Bonferroniho korekce 6

3.5 Chí-kvadrát test nezávislosti v kontingenčních tabulkách 6

4 Testování předpokladů modelů 8

4.1 Předpoklad Poissonovo rozdělení 8

4.2 Test nezávislosti 11

5 Maherovy modely 13

5.1 Druhy modelů 13

5.1.1 Model 0 13

5.1.2 Model 1A, 1B 13

5.1.3 Model 2 14

5.1.4 Model 3C, 3D 14

5.1.5 Model 4 14

5.2 Zvolený model 14

5.3 Model pro Extraligu 16

5.3.1 Parametry 16

5.3.2 Výsledky 18

5.3.3 Chí-kvadrát test dobré shody 19

5.3.4 Zhodnocení 21

6 Dixon-Colesovy modely 21

6.1 Pravděpodobnost výhry, remízy a prohry 22

6.2 Dvojnásobný Poissonovo (DP) model 22

6.2.1 Sdružená pravděpodobnostní funkce 23

6.3 Dvourozměrný Poissonovo (BP) model 23

6.4 Diagonálně rozšířený model 23
6.5 Odhad parametrů 24
6.5.1 Věrohodnostní funkce 24
6.5.2 Logaritmická věrohodnostní funkce 25
6.5.3 Odhad parametru ξ 25
6.6 Model pro Extraligu (CZE) 27
6.6.1 Parametr ξ pro sezónu 2014/2015 27
6.6.2 Odhad parametrů BP model 2015/2016 29
6.6.3 Odhad výsledků zápasů 2015/2016 32
6.6.4 Porovnání modelů pro Extraligu 33
6.7 Model pro Ekstraligu (POL) 34
6.7.1 Parametr ξ pro sezónu 2014/2015 34
6.7.2 Odhad parametrů BP model 2015/2016 36
6.8 Model pro NHL 39
6.8.1 Parametr ξ pro sezónu 2014/2015 39
6.8.2 Odhad parametrů BP-DI model 2015/2016 41
7 Návrh vlastního modelu 45
7.1 Model pro Extraligu (CZE) 45
7.1.1 Parametr ξ pro sezónu 2014/2015 46
7.1.2 Odhad parametrů DP model 2015/2016 47
7.1.3 Porovnání vlastního modelu s původním modelem 48
7.2 Model pro Ekstraligu (POL) 49
7.2.1 Parametr ξ pro sezónu 2014/2015 49
7.2.2 Odhad parametrů DP-DI model 2015/2016 50
7.2.3 Porovnání vlastního modelu s původním modelem 52
7.3 Model pro NHL 52
7.3.1 Parametr ξ pro sezónu 2014/2015 53
7.3.2 Odhad parametrů BP-DI model 2015/2016 54
7.3.3 Porovnání vlastního modelu s původním modelem 57
8 Předvídací schopnost a sázecí strategie 58
8.1 Triviální strategie 58
8.2 Sázení dle modelu 61
8.2.1 Extraliga (CZE) 61

8.2.2 Ekstraliga (POL) ... 64
8.2.3 NHL ... 66
8.3 Sázení dle vlastního modelu ... 68
8.3.1 Extraliga (CZE) ... 68
8.3.2 Ekstraliga (POL) ... 69
8.3.3 NHL ... 71
8.4 Porovnání modelu s triviálními strategiemi ... 72
8.4.1 Extraliga (CZE) ... 72
8.4.2 Ekstraliga (POL) ... 73
8.4.3 NHL ... 74
8.5 Porovnání vlastního modelu s triviálními strategiemi ... 75
8.5.1 Extraliga (CZE) ... 75
8.5.2 Ekstraliga (POL) ... 76
8.5.3 NHL ... 77
9 Závěr ... 78

1 Úvod

Mnoho lidí považuje sázení za hazard a mohou se při něm dostat do osobního bankrotu. Sázení naslepo bez jakýchkoliv zkušeností je velmi rizikové, a proto by bylo vhodné vytvořit si strategii sázení. K tomu lze využít pravděpodobnostní modely a statistické metody. Cílem této práce bude prozkoumání základních modelů, které lze k odhadování výsledků zápasů využít, a porovnat proti sázkovým kancelářím.

V této práci se budeme zabývat zpracováním hokejových zápasů, respektive tří vybraných hokejových lig. K takovému zpracování je zapotřebí získat dostatečně dlouhou historii odehraných zápasů, na základě kterých se budeme snažit odhadovat výsledky budoucích zápasů, ve smyslu výhry domácích, remízy nebo výhry hostů.

Jako základ statistického zpracování a odhadování výsledků sportovních utkání lze použít Maherův model, který je ovšem určený pro data fotbalových utkání. V této práci se tedy pokusíme ověřit, zda je tento model stejně vhodný pro data hokejových utkání. Dalšími uvažovanými modely budou dvojnásobný Poissonovo model (DP) a dvourozměrný Poissonovo model (BP). Tyto modely však podhodnocují pravděpodobnosti remízy, proto využijeme ještě dalšího modelu, který je rozšířením uvedených DP a BP modelů.

Uvedené modely mezi sebou budeme porovnávat a určíme, který z nich je nejvhodnější pro odhad výsledků hokejových zápasů v nadcházející sezóně. Na základě tohoto porovnání najdeme vhodnou strategii, resp. model, podle kterého budeme určovat, zda vsadit na výhru domácích, remízu či výhru hostujících týmů.

Pomocí takových modelů zjistíme, na které zápasy je vhodné vsadit, a jak bude strategie výnosná.

Na závěr porovnáme vybranou strategii sázení pro každou z lig s triviálními strategiemi sázení, a určíme která ze strategií je nejvýnosnější.

2 DATA

Základem této práce jsou správná a kompletní data reálných zápasů z vybraných hokejových lig. Ke zpracování byly vybrány tři hokejové ligy, a to Extraliga hraná v ČR, Ekstraliga hraná v Polsku a národní hokejová liga NHL hraná v USA a Kanadě.

Veškerá historická data k těmto ligám jsou získána z webových stránek *Sfstats.net* [A], *Oddsportal.com* [B] a *Betexplorer.com* [C]. Data obsahují datum, jména a počty gólů týmů v odehraném zápasu, včetně kurzu na výhru domácích, remízu, výhru hostujících. Na stránkách zdroje [A] lze dohledat, že uvedené kurzy jsou průměrem z vybraných pěti sázkových kanceláří (sportingbet, gamebookers, bwin, expekt a bet365). Zdroj [B] naopak obsahuje detail kurzů jednotlivých sázkových kanceláří pro každý zápas. Zdroj [C] byl využit jen ve výjimečných případech k ověření dat nebo doplnění chybějících údajů.

V této práci je primárně využit zdroj [A], ale po důkladném zkoumání dat byly nalezeny chybějící zápasy a kurzy, nebo byl tým uveden v zápase jako domácí, ale ve skutečnosti hrál jako hostující a naopak. Tyto nedostatky v datech bylo nutné dohledat ze zdroje [B], ze kterého byly vypsány chybějící zápasy a dostupné kurzy z vybraných sázkových kanceláří a uloženy do souboru chybějící_kurzy.xlsx, kde se z nich ve stejné logice počítá průměr, který je zaznamenaný v souboru *data.xlsx* na listu dané ligy. Pro českou ligu chybělo celkem 26 kurzů, pro polskou ligu 69 a pro NHL 62.

Kurzy na výsledky zápasů jsou stanovené za základní hrací dobu 60 min, proto používáme i skóre za stejnou hrací dobu, tj. hra může skončit za nerozhodného stavu a nás tedy nezajímá, kdo nakonec zápas vyhrál v případném prodloužení.

V rámci této práce se použijí pouze pravidelné sezónní zápasy kromě zápasů play-off a play-out. Hlavním důvodem tohoto omezení je, že tyto zápasy jsou hrané v rámci vyřazovacího turnaje, který se hraje na čtyři vítězná utkání, na rozdíl od zápasů v základní části. Dalším důvodem je, že zápasy v play-off se mohou hrát různými taktikami oproti obvyklým zápasům odehraným v sezóně.

Data jsou rozdělena do dvou částí. První část zahrnuje všechny zápasy od sezóny 2009/2010 až do sezóny 2014/2015, druhá část obsahuje všechny zápasy ze sezóny 2015/2016 a používá se pouze pro posouzení předvídací schopnosti modelu.

Jeden hrací den může mít různý počet zápasů, a tak (jeden) hrací den nazveme jako jedno „kolo“. Tento standard využijeme pro celé zpracování.

2.1 Extraliga (CZE)

Struktura sezóny

V České republice se hraje nejvyšší hokejová soutěž, která se nazývá Extraliga a má následující pravidla [12]:

- počet týmů, které hrají v jedné sezóně je 14,
- základní část se hraje od září do února,
- v základní části se utká každý tým s každým dvakrát doma a dvakrát venku,

- po základní části se hraje rozšířené play-off, které začíná nejprve předkolem, ve kterém se utká sedmý s desátým a osmý s devátým týmem umístěným v tabulce po základní části,
- předkolo play-off se hraje na 3 vítězné zápasy,
- po této části následuje play-off, ve kterém hraje první tým s hůře umístěným vítězem předkola, druhý s lépe umístěným vítězem předkola, třetí s šestým a čtvrtý s pátým týmem,
- play-off se hraje na 4 vítězné zápasy,
- z play-off postoupí 4 týmy do semifinále a vítězové semifinále postupují do finále,
- vítěz finále je Mistrem Extraligy, druhé až desáté místo je určeno úspěšností týmů v play-off,
- zbývající 4 týmy tabulky po základní části hrají play-out skupinu o udržení v dané lize,
- v play-out (části) se každý tým s každým utká dvakrát, a započítávají se zde body ze základní části,
- týmy umístěné v play-out na 13. a 14. místě hrají se dvěma vítězi semifinále play-off (každý s každým) baráž o udržení v Extralize,
- týmy na prvním a druhém místě na konci baráže postoupí v další sezóně do Extraligy.

Data Extraligy

Soubor dat obsahuje 2 184 výsledků od sezóny 2009/2010 do 2014/2015 za celou hrací dobu v Extralize, a to včetně všech sezónních utkání kromě zápasů play-off a play-out. Data zahrnují i výsledky ze sezóny 2015/2016, které však slouží pouze k ověření modelu oproti sázkovým kancelářím, kde je proveden odhad konečného stavu utkání za základní hrací dobu (výhra domácích, remíza a výhra hostů). Stejně tak tomu je i u dalších lig Ekstraliga (POL) a NHL.

Chybějících kurzů bylo v této lize celkem 26 a jsou doplněné v datovém souboru *data.xlsx* (na listu *CZE*). Dohledávané kurzy jednotlivých sázkových kanceláří ze zdroje [B] jsou vypsané v souboru *chybějící_kurzy.xlsx* na listu *CZE*.

2.2 Ekstraliga (POL)

Struktura sezóny

V Polsku se hraje nejvyšší profesionální hokejová soutěž, která se nazývá Ekstraliga a má následující pravidla [13]:

- počet týmů, které hrají v jedné sezóně je 10 (v sezónách 2011/2012, 2012/2013 a 2013/2014 již není dodržováno),
- nejhorší tým sestupuje do 1. ligy,
- v první části ligy hraje každý tým s každým čtyři zápasy (například pro sezónu 2014/2015, kterou hrálo 10 týmů, to bylo 36 zápasů),
- liga se poté rozdělí do dvou skupin, kde prvních 6 týmů hraje tzv. silnější skupinu o umístění v tabulce před play-off, ve které každý tým s každým hraje jednou doma a jednou venku,

- zbývající 4 týmy hrají tzv. slabší skupinu, ve které hraje každý tým s každým dvakrát doma a dvakrát venku,
- dva nejlepší týmy postupují do play-off a poslední tým sestupuje do 1. ligy,
- následuje play-off, které se hraje na 3 vítězné zápasy v prvním kole,
- následně se hraje čtvrtfinále, semifinále a finále, které se hraje na 4 vítězné zápasy,
- dále se hraje na 2 vítězné zápasy i o konečné třetí místo.

Data Ekstraligy

Soubor dat obsahuje celkem 1 169 zápasů od sezóny 2009/2010 až do sezóny 2014/2015. Co se týče kvality dat z pohledu kurzů, tak v těchto datech celkově chybělo 69 záznamů, které bylo potřeba doplnit z alternativního zdroje [B]. V tomto zdroji však některé kurzy také chyběly a byly dohledány ze zdroje [C]. Data jsou opět doplněná v datovém souboru *data.xlsx* (na listu *POL*) a více informací o chybějících kurzech včetně zápasů lze nalézt v souboru *chybějící_kurzy.xlsx* na listu *POL*.

2.3 NHL

Struktura sezóny

Nejprestižnější hokejovou ligu světa je národní hokejová liga NHL, kterou hrají týmy pouze z USA a Kanady. Týmy jsou rozdělené do východní a západní konference, které jsou dále rozděleny na dvě divize z každé konference. Ve východní konferenci je v každé divizi osm týmů a v západní pouze sedm. Sezóna se skládá ze dvou částí, a to z tzv. základní části a play-off.

NHL má následující pravidla (dle [14] a [15]):

- základní část začíná první středu v říjnu a končí v polovině dubna, kdy hraje každý tým 82 zápasů, z toho polovinu zápasů doma a polovinu venku,
- ve východní konferenci hraje každý tým 30 zápasů proti týmům z jejich divize, 24 zápasů proti týmům ze své konference a 28 zápasů se čtrnácti zbývajícími týmy,
- v západní konferenci hraje každý tým 29 zápasů proti týmům z jejich divize, 21 zápasů proti týmům ze své konference a 32 zápasů s šestnácti zbývajícími týmy,
- tento popis sezóny byl uveden v platnost od sezóny 2013/2014,
- další část sezóny - play-off se hraje od dubna nejpozději do poloviny června, a to na 4 vítězné zápasy,
- nejlepší tři týmy z každé divize a dva týmy z každé konference s nejvyšším počtem bodů bez ohledu na divize postupují do play-off,
- President´s Trophy získává tým, který má největší počet bodů v základní části z obou konferencí,
- vítěz finále play-off získává Stanley Cup.

Data NHL

Soubor dat obsahuje 6 869 zápasů od sezóny 2009/2010 do 2014/2015. Chybějících kurzů bylo v této lize celkem 62 a jsou doplněné v datovém souboru *data.xlsx* (na listu *NHL*), kurzy jednotlivých sázkových kanceláří jsou opět v souboru *chybějící_kurzy.xlsx* na listu *NHL*.

3 Statistické pojmy a metody

Pro využití modelů určených k odhadování výsledků hokejových zápasů je třeba definovat pojmy z pravděpodobnosti a statistiky, které jsou použité v dalších kapitolách.

Všechny zkoumané modely jsou založeny na Poissonovo rozdělení pravděpodobnosti.

3.1 Poissonovo rozdělení

Poissonovo rozdělení je diskrétním rozdělením pravděpodobnosti náhodné veličiny a je označováno $Po(\lambda)$ (viz [6] str. 114-116).

Pravděpodobnostní funkce Poissonova rozdělení má tvar

$$P(X = x) = \frac{\lambda^x}{x!} \cdot e^{-\lambda}, \quad x=0,1,2,... \tag{3.1}$$

Střední hodnota a rozptyl Poissonova rozdělení mají tvar

$$E(X) = \lambda, \tag{3.2}$$

$$D(X) = \lambda. \tag{3.3}$$

3.2 P-hodnota

Definice *p*-hodnoty dle [8] je „pravděpodobnost, s jakou testovací statistika nabývá hodnot „horších“ (více svědčících proti testované hypotéze), než je pozorovaná hodnota statistiky.“

Hypotézu H_0 zamítáme na hladině významnosti α, právě když $p\text{-hodnota} < \alpha$.

3.3 Chí-kvadrát test dobré shody

Máme náhodný výběr rozsahu n z náhodné veličiny X. Testujeme hypotézu na hladině významnosti α, že rozdělení veličiny X má nějaké rozdělení, které známe až na hodnotu m neznámých parametrů (jestliže známe všechny parametry, pak $m = 0$) [7].

Při testování postupujeme následovně:

Rozdělíme obor hodnot na k disjunktních tříd ($k \geq 2$), a zjistíme kolik hodnot náhodného výběru se nachází v jednotlivých třídách, a tyto počty označíme n_i ($i = 1, 2, ..., k$). Poté odhadneme neznámé parametry m předpokládaného modelu. Pro každou třídu spočteme očekávaný počet hodnot o_i v této třídě podle následujícího vzorce

$$o_i = n \cdot p_i, \quad i = 1, 2, ..., k, \tag{3.4}$$

kde n je rozsah náhodného výběru a p_i je pravděpodobnost, že veličina X s předpokládaným rozdělením pravděpodobnosti nabude hodnoty patřící do i-té třídy. Součet očekávaných hodnot o_i je roven n.

Pokud je některý očekávaný počet $o_i < 5$ (někdy se nedodržuje důsledně pro všechny třídy, ale musí být vždy očekávané hodnoty $o_i > 1$), pak sdružíme danou třídu s některou z vedlejších tříd. Sdružená třída má pak očekávaný počet o_i roven součtu očekávaných počtů ze tříd, jejichž sdružením vznikla. Tento postup je nutné opakovat, dokud není splněna podmínka $o_i \geq 5$ pro každou třídu nebo alespoň, že všechny očekávané hodnoty $o_i > 1$. Nový počet tříd opět označíme k.

Hypotézu, že veličina se řídí předpokládaným rozdělením, zamítáme na hladině významnosti α, je-li

$$\sum_{i=1}^{k} \frac{(n_i - o_i)^2}{o_i} > \chi^2_{1-\alpha}(v), \tag{3.5}$$

kde $\chi^2_{1-\alpha}(v)$ je kvantil χ^2 rozdělení a v je počet stupňů volnosti $v = k - 1 - m$ $(v > 0)$.

3.4 Bonferroniho korekce

Bonferroniho korekce je jednou z nejznámějších korekčních procedur pro násobné testování hypotéz. Při testování složených hypotéz je třeba upravit hladinu významnosti α, k čemuž se používá Bonferroniho korekce (popis korekce lze nalézt v [17] str. 5-6):

$$\alpha^* = \frac{\alpha}{m}, \tag{3.6}$$

kde

α^* je upravená hladina významnosti pro jednotlivé hypotézy,
α je původní hladina významnosti pro složenou hypotézu,
m je počet provedených testů.

Bonferroniho korekce porovnává původní hladinu významnosti složené hypotézy s celkovým počtem provedených testů.

3.5 Chí-kvadrát test nezávislosti v kontingenčních tabulkách

Pro testování nezávislosti používáme χ^2-test nezávislosti uvedený v [9]. Pozorované četnosti zapisujeme do následující tabulky (tzv. kontingenční).

Kontingenční tabulka

	1	...	c	Σ
1	n_{11}	...	n_{1c}	$n_{1\bullet}$
2	n_{21}	...	n_{2c}	$n_{2\bullet}$
...	...	...	...	...
r	n_{r1}	...	n_{rc}	$n_{r\bullet}$
Σ	$n_{\bullet 1}$	$n_{\bullet 2}$	$n_{\bullet c}$	n

Obr. 1: Ukázka kontingenční tabulky

Pro proměnné na obrázku platí:

- n_{ij} jsou pozorované četnosti,
- n je celkový počet pozorovaných četností,
- pro řádkové a sloupcové součty platí

$$n_{i\cdot} = \sum_{j=1}^{c} n_{ij}, \quad n_{\cdot j} = \sum_{i=1}^{r} n_{ij}, \tag{3.7}$$

kde *i = 1, …, r; j = 1, …, c.*

Testujeme hypotézy:

H_0: počet gólů vstřelených domácím týmem a počet gólů vstřelených hostujícím týmem jsou nezávislé náhodné veličiny

H_1: počet gólů vstřelených domácím týmem a počet gólů vstřelených hostujícím týmem nejsou nezávislé náhodné veličiny

Testovací kritérium má tvar

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{ij} - o_{ij})^2}{o_{ij}} = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{ij} - \frac{n_{i\cdot}\cdot n_{\cdot j}}{n})^2}{\frac{n_{i\cdot}\cdot n_{\cdot j}}{n}}. \tag{3.8}$$

Při platnosti nulové hypotézy má testové kritérium asymptoticky χ^2 rozdělení, jehož počet stupňů volnosti je roven $v = r \cdot c - (r + c - 2) = (r - 1) \cdot (c - 1)$. Pokud pro hodnotu testovacího kritéria platí, že $\chi^2 > \chi^2_{(r-1)(c-1)}(\alpha)$, zamítáme hypotézu o nezávislosti těchto veličin. Ke shodě s limitním rozdělením se požaduje podmínka očekávaných četností $o_i \geq 5$. Není-li tato podmínka splněna (někdy se nedodržuje striktně pro všechny třídy, ale musí být vždy očekávané četnosti $o_i > 1$), pak sloučíme příslušné řádky nebo sloupce se sousedními v kontingenční tabulce. Počet řádků i sloupců se však nesmí zredukovat na jeden.

4 Testování předpokladů modelů

Maher v článku [1] ukázal na výsledcích fotbalových utkání anglických lig, že počet gólů vstřelených týmem v zápase se řídí Poissonovo rozdělením: držení míče (v našem případě puku) je důležitý aspekt, a pokaždé když má tým míč, má příležitost k útoku a skórování. Pravděpodobnost p, že útok bude mít za následek gól je malá, ale kolikrát má tým míč (resp. puk) v držení během zápasu, je velmi vysoká. Jestliže p je konstantní a útoky jsou nezávislé, počet gólů bude mít binomické rozdělení a za těchto okolností bude velmi dobře platit aproximace Poissonovo rozdělením.

Dle kapitoly 3.3 budeme pomocí χ^2-testu dobré shody ověřovat, zda se počty gólů vstřelených jednotlivými týmy v Extralize (resp. polské lize a NHL) řídí Poissonovo rozdělením.

Dalším testovaným předpokladem bude, že počet gólů domácích a počet gólů hostů jsou dvě nezávislé náhodné veličiny.

4.1 Předpoklad Poissonovo rozdělení

Tento předpoklad bude zkoumán pro počet gólů vstřelených doma a venku, protože týmy hrající zápas doma se zpravidla snaží více útočit a tím pádem i střílet více gólů než při hostujících zápasech.

Pomocí χ^2-testu dobré shody (kapitola 3.3) budeme testovat následující složenou hypotézu, která se skládá z 34 jednotlivých hypotéz (pro každý tým), na hladině významnosti $\alpha = 5\ \%$.

H_0: Počet gólů vstřelených týmy v domácím/hostujícím zápase se řídí Poissonovo rozdělením.

H_1: Počet gólů vstřelených týmy v domácím/hostujícím zápase se neřídí Poissonovo rozdělením.

Testy pro jednotlivé české (resp. polské a NHL) týmy jsou provedeny v souboru *CZE_testování_předpokladů.xlsx* (obdobně pro ostatní ligy). Jako příklad nyní uvedeme jeden z 34 testů pro českou ligu.

$H_{0\_Zlín}$: Počet gólů vstřelených týmem Zlín v domácích zápasech se řídí Poissonovo rozdělením.

$H_{1\_Zlín}$: Počet gólů vstřelených týmem Zlín v domácích zápasech se neřídí Poissonovo rozdělením.

Abychom zajistili, že chyba 1. druhu (α) bude korektní, je třeba použít jednu z metod pro složené testování, v našem případě Bonferroniho korekci (více v kapitole 3.4). Složená hypotéza se skládá z 34 jednotlivých hypotéz u testování české ligy (resp. z 60 jednotlivých hypotéz u NHL a z 28 u polské ligy), proto upravená hladina významnosti pro českou ligu je $\alpha^* = \frac{0{,}05}{34} = 0{,}147\ \%$.

V následujících tabulkách je uveden pozorovaný a očekávaný počet gólů týmu Zlína v domácích zápasech, kritická hodnota a *p-hodnota* pro tento test.

Počet gólů [x_i]	Pozorované četnosti [n_i]	Pravděpodobnosti [p_i]	Očekávané četnosti [o_i]
0	7	0,05	8,33
1	21	0,16	24,41
2	41	0,23	35,76
3	37	0,22	34,92
4	26	0,16	25,58
5	14	0,10	14,98
6 a více	10	0,08	12,01
Celkem	**156**	**1,00**	**156,00**

Tab. 1: Pozorovaný a očekávaný počet gólů pro domácí tým Zlín

testová statistika	1,98
stupeň volnosti	5
hladina významnosti	0,001
kritická hodnota	19,62
p-hodnota	0,85

Tab. 2: Výsledky χ^2-testu pro Zlín

Získané výsledky jednotlivých hypotéz pro všechny týmy v domácích i venkovních zápasech jsou přehledně zobrazeny v Tab. 3, více lze nalézt v souboru *CZE_testování_předpokladů.xlsx* na listu *Poisson_chí-kvadrát test*.

Tým	Výsledek testu doma	p-hodnota doma	Výsledek testu venku	p-hodnota venku
České Budějovice	nezamítáme H0	0,963	nezamítáme H0	0,860
Karlovy Vary	nezamítáme H0	0,040	nezamítáme H0	0,874
Kladno	nezamítáme H0	0,236	nezamítáme H0	0,314
Kometa Brno	nezamítáme H0	0,124	nezamítáme H0	0,132
Liberec	nezamítáme H0	0,131	nezamítáme H0	0,204
Litvínov	nezamítáme H0	0,231	nezamítáme H0	0,019
Mladá Boleslav	nezamítáme H0	0,998	nezamítáme H0	0,163
Pardubice	nezamítáme H0	0,941	nezamítáme H0	0,225
Plzeň	nezamítáme H0	0,127	nezamítáme H0	0,552
Slavia Praha	nezamítáme H0	0,132	nezamítáme H0	0,724
Sparta Praha	nezamítáme H0	0,131	nezamítáme H0	0,002
Třinec	nezamítáme H0	0,087	nezamítáme H0	0,992
Vítkovice	nezamítáme H0	0,968	nezamítáme H0	0,121
Zlín	nezamítáme H0	0,852	nezamítáme H0	0,497
Hradec Králové	nezamítáme H0	0,904	nezamítáme H0	0,138
Chomutov	nezamítáme H0	0,885	nezamítáme H0	0,327
Olomouc	nezamítáme H0	0,141	nezamítáme H0	0,040

Tab. 3: Přehled výsledků jednotlivých hypotéz – Extraliga (CZE)

Dle výsledků jednotlivých hypotéz u všech týmů Extraligy doma i venku není zamítnuta nulová hypotéza, a tak není ani zamítnuta složená hypotéza H_0: počet gólů vstřelených týmem v domácím/hostujícím zápase se řídí Poissonovo rozdělením.

Stejným způsobem se testují jednotlivé hypotézy pro všechny týmy v domácích i venkovních zápasech ostatních lig. Výsledky těchto jednotlivých hypotéz jsou uvedené v Tab. 4 a Tab. 5.

Tým	Výsledek testu doma	p-hodnota doma	Výsledek testu venku	p-hodnota venku
Anaheim	nezamítáme H0	0,863	nezamítáme H0	0,706
Atlanta	nezamítáme H0	0,751	nezamítáme H0	0,192
Boston	nezamítáme H0	0,232	nezamítáme H0	0,386
Buffalo	nezamítáme H0	0,195	nezamítáme H0	0,004
Calgary	nezamítáme H0	0,266	nezamítáme H0	0,050
Carolina	nezamítáme H0	0,794	nezamítáme H0	0,777
Colorado	nezamítáme H0	0,540	nezamítáme H0	0,075
Columbus	nezamítáme H0	0,582	nezamítáme H0	0,299
Dallas	nezamítáme H0	0,861	nezamítáme H0	0,125
Detroit	nezamítáme H0	0,636	nezamítáme H0	0,971
Edmonton	nezamítáme H0	0,472	nezamítáme H0	0,250
Florida	nezamítáme H0	0,785	nezamítáme H0	0,721
Chicago	nezamítáme H0	0,230	nezamítáme H0	0,125
Los Angeles	nezamítáme H0	0,491	nezamítáme H0	0,370
Minnesota	nezamítáme H0	0,263	nezamítáme H0	0,009
Montreal	nezamítáme H0	0,085	nezamítáme H0	0,746
Nashville	nezamítáme H0	0,051	nezamítáme H0	0,793
New Jersey	nezamítáme H0	0,732	nezamítáme H0	0,622
NY Islanders	nezamítáme H0	0,682	nezamítáme H0	0,305
NY Rangers	nezamítáme H0	0,473	nezamítáme H0	0,937
Ottawa	nezamítáme H0	0,234	nezamítáme H0	0,051
Philadelphia	nezamítáme H0	0,001	nezamítáme H0	0,487
Phoenix	nezamítáme H0	0,963	nezamítáme H0	0,398
Pittsburgh	nezamítáme H0	0,407	nezamítáme H0	0,078
San Jose	nezamítáme H0	0,370	nezamítáme H0	0,523
St. Louis	nezamítáme H0	0,841	nezamítáme H0	0,659
Tampa Bay	nezamítáme H0	0,001	nezamítáme H0	0,495
Toronto	nezamítáme H0	0,252	nezamítáme H0	0,131
Vancouver	nezamítáme H0	0,040	nezamítáme H0	0,246
Washington	nezamítáme H0	0,007	nezamítáme H0	0,149

Tab. 4: Přehled výsledků jednotlivých hypotéz – NHL

U NHL ligy jsou výsledky testování jednotlivých hypotéz podobné jako v předchozí lize, tj. u všech týmů doma i venku nezamítáme nulovou hypotézu, a zároveň není zamítnuta ani složená hypotéza H_0: počet gólů vstřelených týmy v domácím/hostujícím zápase se řídí Poissonovo rozdělením.

Tým	Výsledek testu doma	p-hodnota doma	Výsledek testu venku	p-hodnota venku
Bytom	nezamítáme H0	0,050	nezamítáme H0	0,131
Janów	nezamítáme H0	0,646	nezamítáme H0	0,527
Jastrzebie JKH GKS	zamítáme H0	< 0,001	zamítáme H0	< 0,001
Katowice	nezamítáme H0	0,209	nezamítáme H0	0,003
Kraków	zamítáme H0	< 0,001	nezamítáme H0	0,497
Krynica KTH	zamítáme H0	< 0,001	nezamítáme H0	0,188
Orlik Opole	nezamítáme H0	0,007	nezamítáme H0	0,033
Podhale Nowy Targ	nezamítáme H0	0,015	nezamítáme H0	0,107
Sanok	nezamítáme H0	0,015	zamítáme H0	0,001
Stoczniowiec Gdańsk	nezamítáme H0	0,328	nezamítáme H0	0,874
Toruň	nezamítáme H0	0,637	nezamítáme H0	0,609
Tychy	nezamítáme H0	0,270	nezamítáme H0	0,014
Unia Oświęcim	nezamítáme H0	0,003	zamítáme H0	0,001
Zagłębie Sosnowiec	nezamítáme H0	0,010	nezamítáme H0	0,370

Tab. 5: Přehled výsledků jednotlivých hypotéz – Ekstraliga (POL)

Při testování předpokladu Poissonova rozdělení pro polskou ligu jsme zamítli hypotézu H_0 u šesti individuálních testů, a zároveň tak zamítáme i složenou hypotézu H_0: počet gólů vstřelených týmy v domácím/hostujícím zápase se řídí Poissonovo rozdělením.

Na základě nesplnění tohoto předpokladu bychom neměli využívat modely s Poissonovo rozdělením pro polskou ligu. Mezi týmy v polské lize můžeme také sledovat obrovské rozdíly z pohledu skóre, a to také může ovlivňovat výsledek tohoto testu.

V dalších kapitolách přesto využijeme pro odhad výsledků zápasů polské ligy modely, které využívají Poissonovo rozdělení, a ověříme, zda model bude i přesto vhodný k sázení a sázející se v průběhu sezóny nedostane do bankrotu.

4.2 Test nezávislosti

Některé modely předpokládají nezávislost mezi počtem gólů domácích a hostujících, proto dalším testovaným předpokladem bude nezávislost, kterou budeme testovat pomocí χ^2-testu nezávislosti v kontingenčních tabulkách (kapitola 3.5) pro každou sezónu.

Hypotézy formulujeme na hladině významnosti 5 % takto:

H_0: počet gólů vstřelených domácími týmy a počet gólů vstřelených hostujícími týmy jsou nezávislé náhodné veličiny v sezóně 2009/2010 (resp. v dalších sezónách),

H_1: počet gólů vstřelených domácími týmy a počet gólů vstřelených hostujícími týmy nejsou nezávislé náhodné veličiny v sezóně 2009/2010 (resp. i dalších sezónách).

Testy nezávislosti pro takto formulované hypotézy jsou zpracované pro českou ligu v souboru *CZE_testování_předpokladů.xlsx* na listu *test nezávislosti* (stejně tak i pro ostatní ligy).

Třídy jednotlivých testů byly určeny dle počtu vstřelených gólů, např. pro českou ligu v sezóně 2014/2015 jsou třídy 0, 1, …, 9. Ovšem při testování nezávislosti u všech lig vycházely některé očekávané četnosti menší než 5, proto bylo potřeba sloučit třídy, například u české ligy pro sezónu 2014/2015 jsme slučovali třídy s 0 góly a 1 gólem pro domácí i hostující týmy, dále pak 6 až 9 gólů pro domácí týmy a 5 až 9 gólů pro hostující týmy. Po sloučení těchto tříd jsme vyjádřili znovu teoretické četnosti a určili testovou statistiku.

Na základě testové statistiky a kritické hodnoty testu jsme u české ligy došli k závěrům, že pro sezónu 2014/2015 nezamítáme hypotézu H_0, tj. počet gólů vstřelených domácími týmy (X_{ij}) a počet gólů vstřelených hostujícími týmy (Y_{ij}) jsou považovány za nezávislé náhodné veličiny. Ke stejným výsledkům jsme došli i u sezóny 2009/2010 a 2012/2013. V ostatních sezónách – 2010/2011, 2011/2012 a 2013/2014, naopak zamítáme hypotézu H_0, tj. přijímáme hypotézu H_1, že se jedná o závislé náhodné veličiny. Detaily těchto testů jsou zpracovány v souboru *CZE_testování_předpokladů.xlsx* na listu *test nezávislosti*.

Pro sezónu 2014/2015 jsou výsledky χ^2-testu nezávislosti zaznamenané v následující tabulce.

testová statistika	21,65
stupeň volnosti	20
hladina významnosti	0,05
kritická hodnota	31,41
p-hodnota	0,36

Tab. 6: Výsledky χ^2-testu nezávislosti pro sezónu 2014/2015 - Extraliga (CZE)

Další testovanou ligou byla NHL, kde jsme pomocí tohoto testu zamítli hypotézu H_0 pro všechny zpracované sezóny. Testy nezávislosti jednotlivých lig jednoznačně určili závislost mezi počtem vstřelených gólů domácími týmy a počtem gólů vstřelených hostujícími týmy. Detail těchto testů lze nalézt v *NHL_testování_předpokladů.xlsx* na listu *test nezávislosti*.

Výsledky χ^2-testu nezávislosti NHL ligy pro sezónu 2014/2015 jsou uvedené v následující tabulce.

testová statistika	125,50
stupeň volnosti	30
hladina významnosti	0,05
kritická hodnota	43,77
p-hodnota	< 0,001

Tab. 7: Výsledky χ^2-testu nezávislosti pro sezónu 2014/2015 – NHL

U třetí zkoumané polské ligy jsme pro sezónu 2011/2012 a 2012/2013 nezamítali hypotézu H_0, avšak ve zbývajících sezónách tomu bylo naopak. Jak je vidět v následující tabulce, hodnota testové statistiky byla vyšší než kritická hodnota, proto jsme pro sezónu 2014/2015 zamítali nulovou hypotézu, tj. přijímali hypotézu H_1, že náhodné veličiny X_{ij} a Y_{ij} nejsou nezávislé.

testová statistika	38,51
stupeň volnosti	16
hladina významnosti	0,05
kritická hodnota	26,30
p-hodnota	< 0,001

Tab. 8: Výsledky χ^2-testu nezávislosti pro sezónu 2014/2015 - Ekstraliga (POL)

Výsledky všech testů jsou uvedené v souboru *POL_testování_předpokladů.xlsx* na listu *test nezávislosti*.

Testováním předpokladu nezávislosti jsme se pokusili určit, na jaké modely se v následujících kapitolách máme zaměřit. U české a polské ligy nevyšly výsledky testování pro všechny sezóny jednoznačně pro nezávislost či závislost a nelze přesně určit, který model bude vhodnější. Pro NHL ligu jsme získali jednoznačný výsledek, který se přiklání k hypotéze o závislosti počtu vstřelených gólů domácích a hostů. Nicméně jsme se rozhodli, že v dalších částech budeme pokračovat ve zkoumání obou druhů modelů pro všechny ligy, které předpokládají tuto nezávislost nebo závislost a rozhodneme, který z modelů je pro danou ligu vhodnější.

5 Maherovy modely

V této kapitole popíšeme, jaké modely používá M. J. Maher pro odhadování fotbalových zápasů a vybraný model použijeme pro hokejová data.

5.1 Druhy modelů

Na základě poznatků, které byly uvedeny v článku [1] a v kapitole 4, M. J. Maher popsal několik modelů pro odhad výsledků fotbalových zápasů za použití Poissonova rozdělení, které zde uvedeme. Jestliže tým i hraje doma proti týmu j a pozorované skóre je (x_{ij}, y_{ij}), Maher předpokládá, že počet gólů X_{ij} vstřelených domácím týmem se řídí Poissonovo rozdělením s parametrem λ_{ij} a počet gólů Y_{ij} vstřelených hostujícím týmem se řídí Poissonovo rozdělením s parametrem μ_{ij} (ověřeno v předchozí kapitole), a také že X_{ij} a Y_{ij} jsou nezávislé.

Parametr λ je vyjádřen následujícím vzorcem

$$\lambda_{ij} = \alpha_i \cdot \beta_j, \tag{5.1}$$

kde α_i představuje sílu útoku domácího týmu i a β_j udává slabost obrany hostujícího týmu j.

Parametr μ je dán následujícím vzorcem

$$\mu_{ij} = \gamma_i \cdot \delta_j, \tag{5.2}$$

kde γ_i představuje slabost obrany domácího týmu i a δ_j sílu útoku hostujícího týmu j.

Je otázka, zda jsou všechny tyto parametry nezbytné pro zajištění odpovídajícího popisu skóre. Maher se v článku zamýšlí, jestli jsou patrnější rozdíly v útocích nebo obranách, pokud existují rozdíly mezi týmy, a zda je opravdu nutné mít odlišné parametry pro kvalitu útoku týmu doma a venku. Zvážení těchto otázek vede k možné hierarchii modelů, které se liší výpočtem jednotlivých parametrů a jsou popsány v další části podle Mahera z článku [1].

5.1.1 Model 0

Základní model je postaven na úvaze, že všechny týmy jsou stejně silné v útoku i v obraně. V takovém případě platí $\alpha_i = \alpha$, $\beta_i = \beta$, $\gamma_i = \gamma$ a $\delta_i = \delta$ pro všechna i. Aby množina odhadovaných parametrů byla jednoznačná, tak platí omezení $\alpha = \beta$ a $\gamma = \delta$, z čehož vyplývá, že pro tento model stačí odhadnout pouze dva nezávislé parametry, a to je výhodou tohoto modelu. Naopak nevýhodou je, že model bere všechny týmy za stejně silné, což pravděpodobně není pravda.

5.1.2 Model 1A, 1B

Další variantou může být úvaha, že obrana všech týmů je stejně silná, ale v útoku je síla týmů odlišná. Takovou variantu nazývá Maher jako model 1A a platí $\alpha_i = \delta_i$, $\beta_i = \beta$, $\gamma_i = \gamma$ pro všechna i a $\sum_i \alpha_i = \sum_i \beta_i$. Zde je potřeba odhadnout $n + 1$ nezávislých parametrů, kde n je počet týmů v lize.

Model 1B je opačnou variantou modelu 1A, kde útok všech týmů je stejně silný a obrana každého týmu je odlišná, tj. $\alpha_i = \alpha$, $\beta_i = \gamma_i$ a $\delta_i = \delta$ pro všechna i a $\sum_i \alpha_i = \sum_i \beta_i$. Stejně jako u modelu 1A je potřeba odhadnout $n + 1$ nezávislých parametrů.

5.1.3 Model 2

Model 2 je již variantou uvažující rozdílnou sílu týmů v útoku i obraně a navíc je zde zaveden parametr k (resp. k^2), který vyjadřuje poměr síly týmu venku a doma, tedy platí

$\delta_i = k \cdot \alpha_i, \gamma_i = k \cdot \beta_i$ pro všechna i a $\sum_i \alpha_i = \sum_i \beta_i$. Je potřeba odhadnout *2n* nezávislých parametrů.

5.1.4 Model 3C, 3D

V modelu 3C se počítá síla týmu v útoku a v obraně pro každý tým zvlášť. Síla týmu v útoku je brána za stejnou doma i venku, a platí $\alpha_i = \delta_i$ pro všechna i a $\sum_i \alpha_i = \sum_i \beta_i$. Naopak síla týmu v obraně je doma a venku různá. Zde je potřeba odhadnout $3n - 1$ nezávislých parametrů.

Model 3D je opačnou variantou modelu 3C, kde platí, že síla týmu v obraně je doma a venku stejná a v útoku se počítá zvlášť doma a venku.

5.1.5 Model 4

Nejsložitější model dle Mahera uvažuje zvlášť sílu týmu doma a venku, v útoku i obraně. Platí zde $\sum_i \alpha_i = \sum_i \beta_i$ a $\sum_i \gamma_i = \sum_i \delta_i$. V tomto případě je třeba odhadnout $4n - 2$ nezávislých parametrů.

5.2 Zvolený model

Pro naše vybraná data hokejových zápasů zvolíme model 2 (kapitola 5.1.3), který Maher také použil ve svém článku [1] jako nejvhodnější pro fotbalová data. Výhodou tohoto modelu je, že se již počítá s rozdílnou silou jednotlivých týmů v útoku i v obraně, a je třeba odhadnout méně parametrů oproti složitějším modelům. V tomto modelu se odhaduje parametr síly útoku α_i pro všechna i, parametr síly obrany β_j pro všechna j a parametr k (resp. k^2), který vyjadřuje sílu při venkovních zápasech oproti síle při domácích zápasech.

Náhodná veličina X_{ij} označuje počet gólů, které vstřelí domácí tým i a náhodná veličina Y_{ij} značí počet gólů vstřelených hostujícím týmem j v zápase. Předpokládáme, že X_{ij} a Y_{ij} jsou nezávislé a řídí se Poissonovo rozdělením

$$X_{ij} \sim Po\left(\alpha_i \cdot \beta_j\right) \tag{5.3}$$

$$Y_{ij} \sim Po(k^2 \cdot \alpha_j \cdot \beta_i) \tag{5.4}$$

Pro odhad parametrů ve zvoleném modelu 2 jsou v článku [1] na str. 114 odvozené metodou maximální věrohodnosti následující vzorce

$$\hat{\alpha}_i = \frac{\sum_{j \neq i}(x_{ij} + y_{ji})}{\left(1 + \hat{k}^2\right) \sum_{j \neq i} \hat{\beta}_j} \tag{5.5}$$

$$\hat{\beta}_j = \frac{\sum_{i \neq j}(x_{ij} + y_{ji})}{\left(1 + \hat{k}^2\right) \sum_{i \neq j} \hat{\alpha}_i} \tag{5.6}$$

$$\hat{k}^2 = \frac{\sum_i \sum_{j \neq i} y_{ij}}{\sum_i \sum_{j \neq i} x_{ij}}, \tag{5.7}$$

kde x_{ij} je počet gólů vstřelených týmem i týmu j v domácím zápase, y_{ij} je počet gólů vstřelených týmem i týmu j ve venkovním zápase.

Z předchozího vyplývají následující podmínky

$$\sum_i \sum_{j \neq i} \hat{\alpha}_i \cdot \hat{\beta}_j = \sum_i \sum_{j \neq i} x_{ij}, \tag{5.8}$$

$$\sum_i \sum_{j \neq i} \hat{k}^2 \cdot \hat{\alpha}_j \cdot \hat{\beta}_i = \sum_i \sum_{j \neq i} y_{ij,} \tag{5.9}$$

což znamená, že součet středních hodnot vhodných Poissonových rozdělení se rovná pozorovanému počtu vstřelených (resp. obdržených) gólů.

Maherovy modely jsou zaměřené na data z fotbalových utkání, kde hraje každý tým proti každému pouze jednou doma a jednou venku. V Extralize hraje každý tým proti každému dvakrát doma a dvakrát venku, proto bylo potřeba upravit rovnice (5.5), (5.6), (5.8) a (5.9) na tvar

$$\hat{\alpha}_i = \frac{\sum_{j \neq i}(x_{ij} + y_{ji})}{\left(1 + \hat{k}^2\right) \cdot 2 \cdot \sum_{j \neq i} \hat{\beta}_j} \tag{5.10}$$

$$\hat{\beta}_j = \frac{\sum_{i \neq j}(x_{ij} + y_{ji})}{\left(1 + \hat{k}^2\right) \cdot 2 \cdot \sum_{i \neq j} \hat{\alpha}_i} \tag{5.11}$$

$$\sum_i \left(\sum_{j \neq i} \hat{\alpha}_i \cdot \hat{\beta}_j \right) \cdot 2 = \sum_i \sum_{j \neq i} x_{ij} \tag{5.12}$$

$$\sum_i \left(\sum_{j \neq i} \hat{k}^2 \cdot \hat{\alpha}_j \cdot \hat{\beta}_i \right) \cdot 2 = \sum_i \sum_{j \neq i} y_{ij}. \tag{5.13}$$

5.3 Model pro Extraligu

Model 2 podle Mahera popsaný v předchozí podkapitole je použit pro data české ligy od sezóny 2009/2010 do 2014/2015 v sešitu *CZE_model_Maher.xlsm*. Na jednotlivých listech pojmenovaných podle sezón vycházíme z dat, která jsou uvedena na listu *data_uspořádaná*. Při zpracování vybraného Maherova modelu 2 vycházíme z [11].

5.3.1 Parametry

V této podkapitole uvádíme odhady parametrů pouze pro sezónu 2014/2015. Odhady pro předchozí sezóny jsou zpracované v sešitu *CZE_model_Maher.xlsm*.

Nejprve je třeba určit poměr síly týmu venku k síle v domácím prostředí. Tento vztah vyjadřuje parametr k^2 dle rovnice (5.7). Pro sezónu 2014/2015 je $k^2 = 0{,}83$ (tj. 0,83 krát méně gólů venku oproti počtu gólů v domácím zápase).

Poté je nutné vyjádřit pro každý tým sílu v útoku a obraně. Odhadneme parametry α_i a β_i uvedené v rovnicích (5.5) a (5.6). Odhady parametrů se provádí pomocí řešitele, který je analytickým doplňkem MS Excelu. Popis funkce řešitele lze nalézt v [16].

Postup odhadu parametrů α_i a β_i v Microsoft Excel 2013:

1. Zvolíme výchozí hodnoty parametrů α_i a β_i pro všechny týmy viz Obr. 2.
2. Spočteme odhady parametrů α_i a β_i dle rovnic (5.5) a (5.6).
3. Spustíme řešitele s následujícím nastavením
   a. nastavíme Gradientní metodu řešení[1] s použitím automatického měřítka,
   b. součet hodnot parametrů α_i pro všechny týmy se musí rovnat součtu hodnot parametrů β_i dle podmínky definované v kapitole 5.1.3,
   c. nastavíme další podmínku dle rovnice (5.12),
   d. měněné buňky jsou výchozí hodnoty parametrů α_i a β_i.
4. Na základě nových výchozích hodnot zjištěných řešitelem se dopočtou odhady parametrů α_i a β_i opět dle rovnic (5.5) a (5.6), které se v další iteraci použijí jako startovací hodnoty.
5. Tento postup se opakuje několikrát, do té doby, než kvadrát odchylek odhadů mezi jednotlivými iteracemi pro každý tým není menší než 0,0001[2]. V případě, že je tato podmínka splněna, odhad končí a odhadnuté parametry α_i a β_i považujeme za finální.

Na 1 iteraci je vytvořeno makro ve Visual Basic viz Příloha 1, které se spouští tlačítkem „Řešitel – 1 iterace“.

Na Obr. 2 je ukázka počátečního nastavení před první iterací.

---

[1] V řešiteli byla použita Gradientní metoda pro řešení nelineárních problémů (v Microsoft Excel 2010 se tato metoda nazývá GRG Nonlinear), více o této metodě lze nalézt v [10].

[2] Nastavili jsme si tuto podmínku, kde považujeme hodnotu 0,0001 za postačující, protože rozdíly mezi odhady menší než tato hodnota, již významně neovlivňují výsledky a závěry.

Tým	alfa_výchozí	beta_výchozí	suma alfa	suma beta	alfa_odhad	beta_odhad	Nastavení podmínky	alfa_výsledná	beta_výsledná
Hradec Králové	1,70	1,70	22,10	22,10	1,65	1,61	74,95	1,62	1,58
Karlovy Vary	1,70	1,70	22,10	22,10	1,34	1,68	60,66	1,32	1,64
Kometa Brno	1,70	1,70	22,10	22,10	1,86	1,61	84,53	1,83	1,60
Liberec	1,70	1,70	22,10	22,10	1,47	1,58	67,13	1,45	1,55
Litvínov	1,70	1,70	22,10	22,10	1,99	1,45	91,37	1,95	1,45
Mladá Boleslav	1,70	1,70	22,10	22,10	1,93	1,92	86,71	1,93	1,91
Olomouc	1,70	1,70	22,10	22,10	1,50	2,01	67,00	1,50	1,96
Pardubice	1,70	1,70	22,10	22,10	1,76	1,86	79,15	1,75	1,84
Plzeň	1,70	1,70	22,10	22,10	1,68	1,62	76,60	1,66	1,60
Slavia Praha	1,70	1,70	22,10	22,10	1,28	2,22	56,50	1,29	2,15
Sparta Praha	1,70	1,70	22,10	22,10	2,33	1,72	105,42	2,30	1,75
Třinec	1,70	1,70	22,10	22,10	2,22	1,41	101,75	2,16	1,42
Vítkovice	1,70	1,70	22,10	22,10	1,63	1,84	73,62	1,62	1,82
Zlín	1,70	1,70	22,10	22,10	1,73	1,83	78,12	1,72	1,81
Suma	**23,80**	**23,80**	**309,40**	**309,40**	**24,37**	**24,37**	**1 103,51**	**24,08**	**24,08**

Musí být splněna podmínka, že součet alfa = součet beta.

Tato hodnota se musí rovnat počtu gólů domácích (tj. 1 077 - buňka H20).

Obr. 2: Ukázka nastavení výchozích hodnot pro odhad parametrů α_i a β_i

Testovali jsme, zda jsou odhady parametrů α_i a β_i stabilní, tj. při různých výchozích hodnotách dojdeme po několika iteracích ke stejným odhadům a zároveň čím více iterací bude provedeno, tím přesnější by měl být odhad (menší hodnota kvadrátu odchylek mezi jednotlivými iteracemi). Výchozí hodnoty pro sezónu 2009/2010 volíme v rozmezí 1 až 2 po hodnotě 0,1, a také vyšší výchozí hodnoty například 4 a 5, protože nejběžnější hodnoty těchto parametrů by měly logicky vycházet například mezi 0 a 3. To vyplývá z významu parametrů α a β (kapitola 5.2). Detail porovnání odhadů je zpracován na listu *2009-2010 iterace*.

Po několika iteracích jsme při různých výchozích hodnotách došli ke stejným odhadům. Na základě sezóny 2009/2010 jsme určili výchozí hodnoty pro tuto sezónu a následující. Princip byl, aby byl co nejmenší počet iterací, který stačí pro splnění podmínky definované v postupu odhadu parametrů α_i a β_i (5. bod), a zároveň součet rozdílů mezi poslední a předposlední iterací byl co nejmenší. Tato pravidla splnily výchozí hodnoty nastavené na hodnotu 1,7.

Odhady parametrů α_i a β_i po jednotlivých iteracích jsou zobrazeny v následující tabulce, kde výchozí hodnoty všech parametrů byly nastaveny na hodnotu 1,7. Odhady z 5. iterace považujeme za finální.

	1. iterace		2. iterace		3. iterace		4. iterace		5. iterace	
Tým	α	β	α	β	α	β	α	β	α	β
Hradec Králové	1,63	1,59	1,62	1,58	1,62	1,58	1,62	1,58	**1,62**	**1,58**
Karlovy Vary	1,32	1,66	1,32	1,64	1,32	1,64	1,32	1,64	**1,32**	**1,64**
Kometa Brno	1,83	1,59	1,83	1,60	1,83	1,60	1,83	1,60	**1,83**	**1,60**
Liberec	1,45	1,57	1,45	1,55	1,45	1,55	1,45	1,55	**1,45**	**1,55**
Litvínov	1,97	1,43	1,95	1,45	1,95	1,45	1,95	1,45	**1,95**	**1,45**
Mladá Boleslav	1,91	1,90	1,92	1,91	1,93	1,91	1,93	1,91	**1,93**	**1,91**
Olomouc	1,48	1,98	1,50	1,96	1,50	1,96	1,50	1,96	**1,50**	**1,96**
Pardubice	1,74	1,83	1,75	1,84	1,75	1,84	1,75	1,84	**1,75**	**1,84**
Plzeň	1,66	1,60	1,66	1,60	1,66	1,60	1,66	1,60	**1,66**	**1,60**
Slavia Praha	1,28	2,19	1,29	2,15	1,29	2,15	1,29	2,15	**1,29**	**2,15**
Sparta Praha	2,30	1,70	2,30	1,75	2,30	1,75	2,30	1,75	**2,30**	**1,75**
Třinec	2,19	1,39	2,16	1,43	2,16	1,42	2,16	1,42	**2,16**	**1,42**
Vítkovice	1,61	1,82	1,62	1,82	1,62	1,82	1,62	1,82	**1,62**	**1,82**
Zlín	1,71	1,81	1,72	1,81	1,72	1,81	1,72	1,81	**1,72**	**1,81**

Tab. 9: Odhadování parametrů α_i a β_i pro sezónu 2014/2015

5.3.2 Výsledky

V této části budou ukázány výsledky pro dva vybrané týmy. Výsledky ostatních týmů jsou v sešitu *CZE_model_Maher.xlsm* na listech jednotlivých sezón.

Z odhadnutých parametrů z minulé části lze vypočítat odhady parametrů λ_{ij} dle rovnice (5.1) a μ_{ij} dle rovnice (5.2). Pokud bychom uvažovali zápas mezi Plzní a Spartou Praha, potom by byl odhad střední hodnoty počtu vstřelených gólů následující:

Pro Plzeň
$$\hat{\lambda} = \hat{\alpha}_{PLZ} \cdot \hat{\beta}_{SP} = 1{,}66 \cdot 1{,}75 = 2{,}89 \tag{5.11}$$

Pro Spartu Praha
$$\hat{\mu} = \hat{k}^2 \cdot \hat{\alpha}_{SP} \cdot \hat{\beta}_{PLZ} = 0{,}83 \cdot 2{,}30 \cdot 1{,}60 = 3{,}05 \tag{5.12}$$

Nyní můžeme určit pravděpodobnosti, kolik dá tým gólů v zápase.

Pro Plzeň, kde $\hat{\lambda}$ = 2,89 lze podle (3.1) dopočítat pravděpodobnost, že Plzeň dá Spartě 1 gól:

$$P(X = 1) = e^{-2{,}89} \cdot \frac{2{,}89^1}{1!} = 0{,}16. \tag{5.13}$$

Nebo pravděpodobnost, že Plzeň dá Spartě 6 a více gólů je

$$P(X \geq 6) = 1 - P(X \leq 5) = 1 - F(5) = 0{,}07. \tag{5.14}$$

Pro Spartu Praha, kde $\hat{\mu}$ = 3,05 lze podle (3.1) dopočítat pravděpodobnost, že dá Plzni 1 gól:

$$P(Y = 1) = e^{-3{,}05} \cdot \frac{3{,}05^1}{1!} = 0{,}14. \tag{5.15}$$

Můžeme také dopočítat pravděpodobnost, že zápas skončí remízou, například za stavu 1:1

$$\begin{aligned} P(X = 1, Y = 1) &= P(X = 1) \cdot P(Y = 1) = 0{,}16 \cdot 0{,}14 \\ &= 0{,}02. \end{aligned} \tag{5.16}$$

V Tab. 10 jsou uvedené pravděpodobnosti, že domácí tým dá 1 gól hostujícímu týmu ve vzájemných zápasech mezi uvedenými týmy.

P(X = 1)		Hosté													
		H. Králové	K. Vary	Kometa Brno	Liberec	Litvínov	M. Boleslav	Olomouc	Pardubice	Plzeň	Slavia Praha	Sparta Praha	Třinec	Vítkovice	Zlín
Domácí	Hradec Králové	x	0,188	0,195	0,205	0,225	0,140	0,133	0,152	0,195	0,107	0,167	0,230	0,156	0,156
	Karlovy Vary	0,259	x	0,256	0,265	0,283	0,203	0,195	0,215	0,256	0,167	0,231	0,288	0,219	0,220
	Kometa Brno	0,160	0,151	x	0,167	0,188	0,106	0,099	0,117	0,158	0,078	0,131	0,193	0,120	0,121
	Liberec	0,232	0,222	0,229	x	0,258	0,174	0,166	0,186	0,229	0,139	0,202	0,263	0,190	0,191
	Litvínov	0,141	0,132	0,138	0,148	x	0,090	0,084	0,100	0,138	0,064	0,113	0,173	0,103	0,104
	Mladá Boleslav	0,144	0,135	0,142	0,151	0,172	x	0,086	0,103	0,142	0,066	0,116	0,177	0,106	0,106
	Olomouc	0,221	0,212	0,218	0,228	0,248	0,163	x	0,176	0,218	0,129	0,191	0,253	0,179	0,180
	Pardubice	0,174	0,164	0,171	0,181	0,202	0,118	0,111	x	0,171	0,088	0,144	0,207	0,133	0,134
	Plzeň	0,190	0,181	0,188	0,198	0,218	0,133	0,126	0,145	x	0,101	0,160	0,223	0,149	0,149
	Slavia Praha	0,266	0,257	0,263	0,272	0,290	0,210	0,202	0,222	0,263	x	0,238	0,294	0,226	0,227
	Sparta Praha	0,095	0,087	0,093	0,101	0,119	0,054	0,049	0,061	0,093	0,035	x	0,123	0,064	0,064
	Třinec	0,111	0,103	0,109	0,118	0,137	0,066	0,061	0,075	0,109	0,045	0,086	x	0,077	0,078
	Vítkovice	0,197	0,187	0,194	0,204	0,224	0,139	0,132	0,151	0,194	0,107	0,166	0,229	x	0,155
	Zlín	0,179	0,169	0,176	0,186	0,207	0,122	0,115	0,134	0,176	0,092	0,149	0,211	0,137	x

Tab. 10: Pravděpodobnosti, že domácí tým dá 1 gól hostujícímu týmu

Obdobné tabulky s pravděpodobnostmi, že hostující tým dá 0, 1, … gólů domácímu týmu v zápase jsou uvedené v sešitu *CZE_model_Maher.xlsm* na listu konkrétní sezóny.

5.3.3 Chí-kvadrát test dobré shody

χ^2-test dobré shody uvedený v kapitole 3.3 využijeme k porovnání vypočtených pravděpodobností dle vybraného Maherova modelu s reálnými výsledky, stejně jako uvádí Maher ve svém článku [1]. Testujeme zvlášť góly doma a venku v sešitu *CZE_model_Maher.xlsm* na listu *chí-kvadrát test*. Následující hypotézy testujeme na hladině významnosti $\alpha = 5\ \%$.

H_0: Skutečná četnost zápasů s daným počtem vstřelených gólů pro domácí týmy (resp. hostující) v sezóně 2009/2010 (následně pro další sezóny) se neliší od četností zápasů s daným počtem vstřelených gólů pro domácí týmy (resp. hostující) určených dle zvoleného modelu.

H_1: Skutečná četnost zápasů s daným počtem vstřelených gólů pro domácí týmy (resp. hostující) v sezóně 2009/2010 (následně pro další sezóny) se liší od četností zápasů s daným počtem vstřelených gólů pro domácí týmy (resp. hostující) určených dle zvoleného modelu.

Jako první vyjádříme pozorované četnosti zápasů n_i podle výsledků zápasů pro každou sezónu. Následně určíme očekávané četnosti pro jednotlivé góly. Níže v tabulce můžeme vidět, že například očekávaná četnost zápasů pro 1 gól vstřelený domácími je 59 oproti skutečnosti, kdy bylo těchto zápasů 52. Očekávané četnosti zápasů jsme v tomto případě určili součtem hodnot tabulky $P(X = 1)$ vynásobené dvěma (protože každý tým české ligy hraje dva zápasy doma a dva venku, jak již bylo uvedeno výše).

Počet gólů	Pozorované četnosti	Očekávané četnosti
0	24	22
1	52	59
2	87	81
3	78	77
4	52	57
5	36	35
6 a více	35	33
Počet gólů celkem	**364**	**364**

Tab. 11: Pozorované a očekávané četnosti pro domácí zápasy – sezóna 2014/2015

testová statistika	1,96
stupně volnosti	5
hladina významnosti	0,05
kritická hodnota	11,07
p-hodnota	0,85

Tab. 12: Výsledky χ^2-testu pro domácí zápasy – sezóna 2014/2015

Počet gólů	Pozorované četnosti	Očekávané četnosti
0	21	35
1	96	79
2	107	91
3	57	73
4	36	45
5	26	24
6 a více	21	17
Počet gólů celkem	**364**	**364**

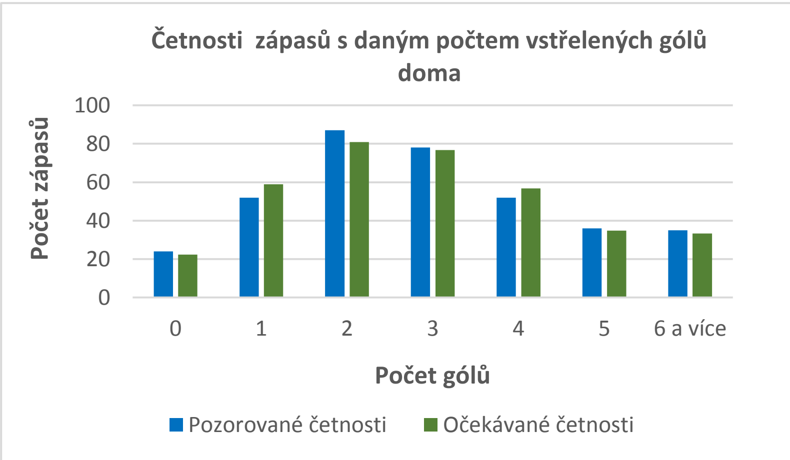
Tab. 13: Pozorované a očekávané četnosti pro venkovní zápasy – sezóna 2014/2015

testová statistika	19,00
stupně volnosti	5
hladina významnosti	0,05
kritická hodnota	11,07
p-hodnota	0,002

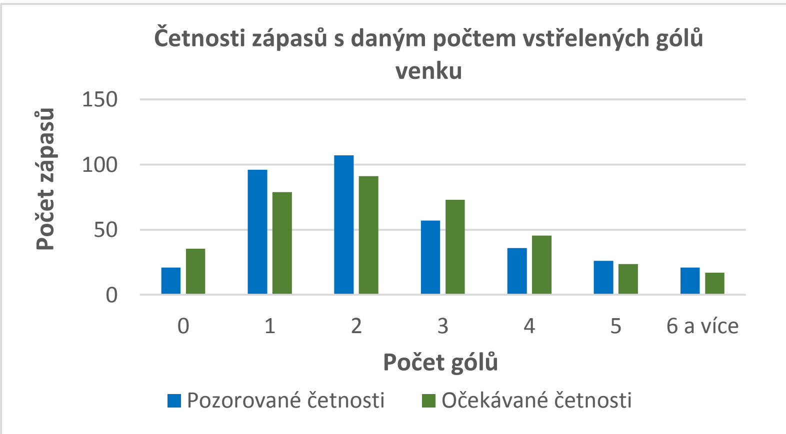
Tab. 14: Výsledky χ^2-testu pro venkovní zápasy – sezóna 2014/2015

V Tab. 11 a Tab. 13 je vidět, že očekávané četnosti u zápasů hraných doma téměř odpovídají skutečným četnostem. Naopak velký rozdíl mezi skutečnými a očekávanými četnostmi je u venkovních zápasů. Tyto závěry jen potvrzují p-hodnoty obou testů.

P-hodnota u testu domácích zápasů pro sezónu 2014/2015 vychází 0,855, což značí, že bychom hypotézu H_0 nezamítali. Naopak u testu venkovních zápasů vychází *p-hodnota* 0,002, proto bychom měli hypotézu H_0 zamítat, tj. přijímat hypotézu H_1, že počet gólů vstřelených všemi týmy venku se liší od počtu gólů určených ve zvoleném Maherově modelu. Ostatní sezóny jsou zpracované na listu *chí-kvadrát test*.



Obr. 3: Histogram četností – domácí zápasy



Obr. 4: Histogram četností – venkovní zápasy

5.3.4 Zhodnocení

Z výsledků předchozího testování vyplývá, že Maherův model není pro naše hokejová data zcela vhodný, dle zamítnutí hypotézy o shodě počtu vstřelených gólů venku s očekávaným počtem vstřelených gólů na základě vybraného modelu (Maherův model 2) pro českou ligu.

O Maherovo modelu 2 víme, že je třeba parametry odhadovat až na základě ukončené předchozí sezóny, kdy je počet zápasů každého týmu stejný. Všechny zápasy v sezóně mají stejnou váhu, a tím se neprojevuje v modelu aktuální forma z posledních zápasů.

Tento model jsme tedy použili alespoň pro českou ligu a v dalších částech již budou použity všechny zmíněné ligy, kde využijeme dalších modelů popsaných v kapitole 6, kde je zavedena váhová funkce pro jednotlivé zápasy.

6 Dixon-Colesovy modely

Mark J. Dixon a Stuart G. Coles vylepšili Maherův model, a tento model popisují ve svém článku [2]. Jejich model byl vytvořený pouze pro fotbalová data, proto používáme pro hokejová data upravené Dixon-Colesovy modely z článku [3]. Síla týmů se v čase může měnit, což je pro nás důležitý aspekt a tyto modely takovou dynamiku zohledňují.

U statistického modelu jsou požadovány následující vlastnosti [2]:

- měl by brát v úvahu rozdílné schopnosti obou týmů v utkání,
- měl by uvažovat skutečnost, že týmy hrající doma mají všeobecně nějakou výhodu tzv. „domácí výhodu“,
- nejrozumnější měřítko schopnosti týmu by mělo být založeno na tom, jakou výkonnost měl tým v minulosti, kdy největší váhu by měly mít nejnovější výsledky,
- měl by brát zvlášť schopnost týmu útočit (dát gól) a jeho schopnost bránit se (nedostat gól).

Všechny tyto předpoklady splňují jak Dixon-Colesovy modely z [2], tak i upravené Dixon-Colesovy z článku [3].

V následujících podkapitolách jsou popsány dva modely a diagonálně rozšířený model, který slouží k rozšíření těchto dvou modelů. Modely se mohou použít pro odhad sdružené pravděpodobnostní funkce (X_{ij}, Y_{ij}). Jak jsme uvedli na začátku kapitoly, vycházíme z modelů pro fotbalová data. Ovšem některé z těchto modelů uvažují, že počet gólů domácího a hostujícího týmu jsou nezávislé náhodné proměnné a jiné modely je naopak považují za závislé. Jak již bylo uvedeno v podkapitole 4.2, zkoumáme oba modely, protože nelze přesně definovat, která volba je vhodnější pro hokejová data.

V případě těchto modelů byla objevena skutečnost, že modely podhodnocují pravděpodobnosti remíz, což ukázali Karlis a Ntzoufras ve svém článku [4]. Na základě tohoto zjištění byla navržena modifikace obou modelů, uvedená v článku [3]. Takovou modifikaci v práci také používáme, jedná se o diagonálně rozšířené modely.

6.1 Pravděpodobnost výhry, remízy a prohry

Chceme určit, s jakou pravděpodobností dají týmy určitý počet gólů v utkání, a tím odhadnout, zda tým vyhraje, remízuje či prohraje. Náhodné proměnné X_{ij} a Y_{ij} označují výsledky zápasu mezi týmy i a j, kde náhodná proměnná X_{ij} představuje počet gólů vstřelených domácím mužstvem a počet gólů od hostujícího týmu je náhodná proměnná Y_{ij}, jak již bylo uvedeno v Maherově modelu (kapitola 5.1).

V případě, že známe sdruženou pravděpodobnostní funkci (X_{ij}, Y_{ij}), můžeme určit pravděpodobnosti výhry domácího týmu $\left(p_{ij}^{H}\right)$[3], remízy $\left(p_{ij}^{D}\right)$[4] a prohry, čili vítězství hostujícího týmu $\left(p_{ij}^{A}\right)$[5], dle následujících vzorců (viz [3])

$$p_{ij}^{H} = \sum_{x>y} P\left(X_{ij} = x, Y_{ij} = y\right), \tag{6.1}$$

$$p_{ij}^{D} = \sum_{x=y} P\left(X_{ij} = x, Y_{ij} = y\right), \tag{6.2}$$

$$p_{ij}^{A} = \sum_{x<y} P\left(X_{ij} = x, Y_{ij} = y\right). \tag{6.3}$$

6.2 Dvojnásobný Poissonovo (DP) model

DP model předpokládá, že náhodné proměnné X_{ij} a Y_{ij} jsou nezávislé (původně byl model použit Maherem v článku [1]) a dle článku [3] platí[6]:

$$X_{ij} \sim Po\left(\lambda_H = \mu \cdot \alpha_i \cdot \beta_j \cdot \gamma\right), \tag{6.4}$$

$$Y_{ij} \sim Po\left(\lambda_A = \mu \cdot \alpha_j \cdot \beta_i\right), \tag{6.5}$$

kde

α_i je síla útoku týmu (čím vyšší, tím lepší), $\alpha_i > 0$ pro všechna i,

β_i je naopak síla obrany týmu (čím menší, tím lepší), $\beta_i > 0$ pro všechna i,

μ je parametr, který vyjadřuje průměrný počet gólů hostujícího týmu, $\mu > 0$,

γ je míra výhody domácího utkání (více v článku [2] nebo [3]), $\gamma > 0$.

[3] H = home, používá se u parametrů týkajících se domácího týmu,
[4] D = draw, používá se u parametrů týkajících se remízy,
[5] A = away, používá se u parametrů týkajících se hostujícího týmu.
[6] Model použitý Maherem, ale upravený pro hokejová data, kde provádíme reparametrizaci přidáním parametru $\mu > 0$.

6.2.1 Sdružená pravděpodobnostní funkce

Sdružená pravděpodobnostní funkce pro výsledek zápasu mezi domácím týmem i a hostujícím týmem j má tvar [3]

$$P(x,y) = \frac{\lambda_H^x \cdot e^{-\lambda_H}}{x!} \cdot \frac{\lambda_A^y \cdot e^{-\lambda_A}}{y!}, \qquad x,y = 0,1,2,\dots \tag{6.6}$$

Je nezbytné použít omezující podmínky $\sum_i \alpha_i = N$ a $\sum_i \beta_i = N$, kde N je počet týmů, které vstupují do modelu. Tyto podmínky zajišťují, že průměrná hodnota všech útočných i obranných parametrů je rovna 1. V následujících modelech budeme uvažovat stejné omezující podmínky.

6.3 Dvourozměrný Poissonovo (BP) model

V tomto modelu z článku [3] autoři navrhují užít dvourozměrné Poissonovo rozdělení. Toto rozdělení však nelze použít vzhledem k záporné korelaci mezi počtem gólů domácích a hostů. Proto použijeme dvourozměrné Poissonovo rozdělení, které je definované takto[7] (detail tohoto rozdělení viz [5])

$$P(x,y) = \lambda_H^x \cdot \lambda_A^y \cdot e^{-\lambda_H-\lambda_A} \cdot \frac{1 + \lambda \cdot \left(e^{-x} - e^{-d\cdot\lambda_H}\right) \cdot \left(e^{-y} - e^{-d\cdot\lambda_A}\right)}{x! \cdot y!} \tag{6.7}$$

pro x, y = 0, 1, 2, …, kde $d = 1 - e^{-1}$, λ_H a λ_A jsou definované v rovnicích (6.4) a (6.5). Korelační koeficient mezi X_{ij} a Y_{ij} lze vyjádřit takto

$$\rho = \lambda \cdot \sqrt{\lambda_H \cdot \lambda_A} \cdot d^2 \cdot e^{-d\cdot(\lambda_H+\lambda_A)}. \tag{6.8}$$

Z předchozího vyplývá, že parametr $\lambda \in R$ musíme odhadovat. Parametr λ určuje závislost mezi góly vstřelenými domácími a hostujícími týmy. Tento parametr je globální pro všechny týmy, stejně jako parametry γ a μ.

6.4 Diagonálně rozšířený model

Na fotbalových datech bylo ukázáno v článku [2] a [4], že modely podhodnocují pravděpodobnost výsledku remízy. Proto použijeme diagonálně rozšířený model pro oba předchozí modely, který byl navržen v článku [4] v této formě

$$P_{DI}(x,y) = \begin{cases} (1-p)P(x,y) & x \neq y, \\ (1-p)P(x,y) + pD(x) & x = y, \end{cases} \tag{6.9}$$

kde $p \in [0,1]$ je pravděpodobnost remízy, $D(x)$ je diskrétní rozdělení, *P(x,y)* je sdružená pravděpodobnostní funkce definovaná pro DP a BP model (rovnice ((6.6) a ((6.7)).

[7] Tímto postupem nahrazujeme problematické řešení, které využívá propojení dvou Poissonovo rozdělení pomocí kopule [3].

Budeme uvažovat obecné rozdělení $D(x)$, pro které platí $P(X = k) = \theta_k$ pro k = 0, 1, …, kde $\theta_k \geq 0$ pro všechna k a $\sum_{k=0}^{K} \theta_k = 1$, kde K označuje celkový počet remízových gólů. V článku [4] autoři používali pro fotbalová data parametr $K = 3$ jako postačující. Ovšem tato hodnota nemusí být a ani není postačující pro hokejová data, kde je počet vstřelených gólů během zápasu vyšší. Například data z české ligy obsahují 463 remíz, z čehož nejméně zápasů skončilo remízou 6:6 nebo více gólů (konkrétně pouze čtyřikrát). Proto za dostatečný koeficient volíme $K = 5$. Stejná hodnota tohoto parametru se použije i pro ostatní zkoumané ligy.

Pokud ve dvojnásobném Poissonovo modelu a dvourozměrném Poissonovo modelu užijeme rovnici (6.9), získáme dva nové modely (DP-DI a BP-DI).

6.5 Odhad parametrů

Parametry z modelů definovaných v předchozích kapitolách 6.2 až 6.4 se odhadují metodou maximální věrohodnosti, tj. stejným způsobem jako v Maherových modelech.

6.5.1 Věrohodnostní funkce

Pro každý model uvedený v podkapitole 6.2 až 6.4 se definuje věrohodnostní funkce pro odhad parametrů zápasů, které označujeme $m = 1, 2, \ldots, M$ (kde $m = 1$ je nejstarší dostupné utkání), a skóre mezi týmy i a j v m-tém zápasu (x_m, y_m).

Pro dvojnásobný Poissonovo (DP) model má věrohodnostní funkce tvar [3]

$$L(\alpha_i, \beta_i, \gamma, \mu; i = 1, \ldots, N; k = 0, \ldots, K) = \prod_{m=1}^{M} P(x_m, y_m), \tag{6.10}$$

kde $P(x_m, y_m)$ je pravděpodobnostní funkce dvojnásobného Poissonovo rozdělení definovaná v rovnici (6.6).

Pro ostatní upravené Dixon-Colesovy modely se použije stejná definice věrohodnostní funkce, jako je uvedena výše, pouze s rozdílnou pravděpodobnostní funkcí $P(x_m, y_m)$, a to dle konkrétního modelu:

- pro dvourozměrný Poissonovo (BP) model definovaná v rovnici (6.7),
- pro diagonálně rozšířený dvojnásobný Poissonovo (DP-DI) model definovaná v rovnici (6.9), do které se dosadí pravděpodobnostní funkce dvojnásobného Poissonovo rozdělení definovaná v rovnici (6.6),
- pro diagonálně rozšířený dvourozměrný Poissonovo (BP-DI) model definovaná v rovnici (6.9), do které se dosadí pravděpodobnostní funkce dvourozměrného Poissonovo rozdělení definovaná v rovnici (6.7).

Je zřejmé, že výkonnosti týmů se mohou v průběhu sezóny měnit. Proto neodhadujeme konstantní parametry pro celou sezónu, jak tomu bylo u Maherových modelů, ale pro každý čas, kdy se hraje zápas. Stejně jako popisují autoři článku [2].

V této práci, stejně jako Dixon a Coles, volíme exponenciální funkci jako váhu pravděpodobnosti, kde váha předchozích zápasů exponenciálně klesá. Tito autoři odhadovali fotbalové zápasy v polo týdnech, my ale pro další odhad volíme vždy každé další kolo (resp. hrací den), tj. stejným způsobem, jako uvádí autoři v článku [3]. Tím bereme v úvahu čas mezi sezónami nebo některými krátkými přestávkami. Váhová funkce v čase odhadu T je definována dle [3] takto

$$\tau(t_m) = \begin{cases} e^{-\xi\cdot(T-t_m)/365,25}, & t_m < T, \\ 0, & t_m \geq T, \end{cases} \tag{6.11}$$

kde t_m je čas zápasu a $\xi \geq 0$; když mají všechny zápasy stejnou váhu, pak $\xi = 0$.

Z rovnice (6.10) s použitím váhové funkce sestavujeme věrohodnostní funkci, kde je pravděpodobnostní funkce umocněna na $\tau(t_m)$

$$L(i = 1, \dots, N; k = 0, \dots, K) = \prod_{m=1}^{M} \{P(x_m, y_m)\}^{\tau(t_m)}, \tag{6.12}$$

kde

$m = 1, 2, \dots, M$ ($m = 1$ je nejstarší zápas),

(x_m, y_m) je skóre mezi týmem i a j v m-tém zápase.

6.5.2 Logaritmická věrohodnostní funkce

Pro odhad parametrů jsou důležité pouze polohy bodů maxima. Využijeme proto zlogaritmování věrohodnostní funkce, čímž se odhady parametrů nezmění.

Logaritmická věrohodnostní funkce pro DP a BP model má následující tvar

$$ln\, L = ln\left(\prod_{m=1}^{M} \{P(x_m, y_m)\}^{\tau(t_m)}\right) = \sum_{m=1}^{M} \tau(t_m) \cdot \ln(P(x_m, y_m)). \tag{6.13}$$

Pro diagonálně rozšířené modely (DP-DI a BP-DI) má logaritmická věrohodnostní funkce tvar

$$ln\, L = ln\left(\prod_{m=1}^{M} \{P_{DI}(x_m, y_m)\}^{\tau(t_m)}\right) = \sum_{m=1}^{M} \tau(t_m) \cdot ln(P_{DI}(x_m, y_m)), \tag{6.14}$$

kde $P_{DI}(x_m, y_m)$ je definovaná v rovnici (6.9).

6.5.3 Odhad parametru ξ

Optimalizace výběru parametru ξ je problematická, protože rovnice (6.12) definuje posloupnost nezávislých pravděpodobností, ale požadujeme ξ takové, aby byla maximalizována celková prediktivní schopnost modelu [2]. Dixon a Coles zvolili ve svém článku parametr ξ tak, aby optimalizovali předvídání výsledků než skóre zápasů, což použijeme i v této práci. K nalezení nejlepší hodnoty parametru ξ definovali

$$S(\xi) = \sum_{m=1}^{M} (\delta_m^H \cdot \ln p_m^H + \delta_m^D \cdot \ln p_m^D + \delta_m^A \cdot \ln p_m^A), \tag{6.15}$$

kde δ_m je indikační funkce, pro kterou platí například:

$\delta_m^H = 1$, jestliže zápas m končí vítězstvím domácího týmu,

$\delta_m^H = 0$ v jiných případech,

p_m^H, p_m^D a p_m^A jsou pravděpodobnosti možných výsledků zápasu spočítané dle rovnic (6.1), (6.2) a (6.3) za použití maximálně věrohodných odhadů získaných z rovnice (6.12) s nastaveným váhovým parametrem na ξ.

Jak uvádí autoři článku [3], funkci $S(\xi)$ lze považovat za měřítko předvídací chyby výsledků, protože pro výsledky s odhadnutou malou pravděpodobností dává velké záporné číslo, zatímco pro výsledky s odhadnutou vysokou pravděpodobností udává záporné číslo blížící se k nule, tudíž lze říci, že maximalizace $S(\xi)$ vede k nejlepšímu odhadu parametru ξ. Jak bude ukázáno dále, výsledek je robustní v celé řadě hodnot parametru ξ.

Pro každý model se užije následující ověřovací postup převzatý z [3]:

1) Soubor dat historických výsledků je rozdělen do dvou souborů. První soubor obsahuje všechny zápasy, které byly před sezónou 2014/2015 a druhý soubor obsahuje zápasy, které se odehrály právě v této sezóně. Předpokládejme, že výsledky všech zápasů v druhém souboru jsou neznámé.
2) Nejprve vybereme nejbližší datum, kdy se hrály nějaké zápasy s neznámým výsledkem v sezóně 2014/2015 a stanovíme *T*, které je použito v rovnici (6.11), rovné tomuto datu.
3) Stanovíme *M* použité v rovnici (6.10), jako počet zápasů odehraných před tímto datem *T*.
4) Použijeme model definovaný v rovnici (6.12) nebo jeho logaritmickou verzi a odhadneme příslušné parametry pro daný model užitím všech informací ze souboru historických výsledků.
5) Užijeme odhadované parametry k výpočtu pravděpodobností definovaných v rovnicích (6.1), (6.2) a (6.3) pro každý zápas odehraný v čase *T*.
6) Přesuneme všechny zápasy a jejich reálné výsledky odehrané v čase *T* ze souboru s neznámými výsledky do souboru s historickými výsledky.
7) Budou-li v sezóně stále zápasy s neznámým výsledkem, vrátíme se na 2. bod tohoto postupu a použijeme aktualizované soubory z předchozího bodu.

Tento postup se opakuje pro různé hodnoty parametru ξ a odpovídající $S(\xi)$ je dopočítáno po uplynutí celé sezóny. Jak jsme zmiňovali již dříve, $S(\xi)$ může mít pouze záporné hodnoty, a čím více se hodnota $S(\xi)$ blíží k nule, tím lepší predikci získáme.

6.6 Model pro Extraligu (CZE)

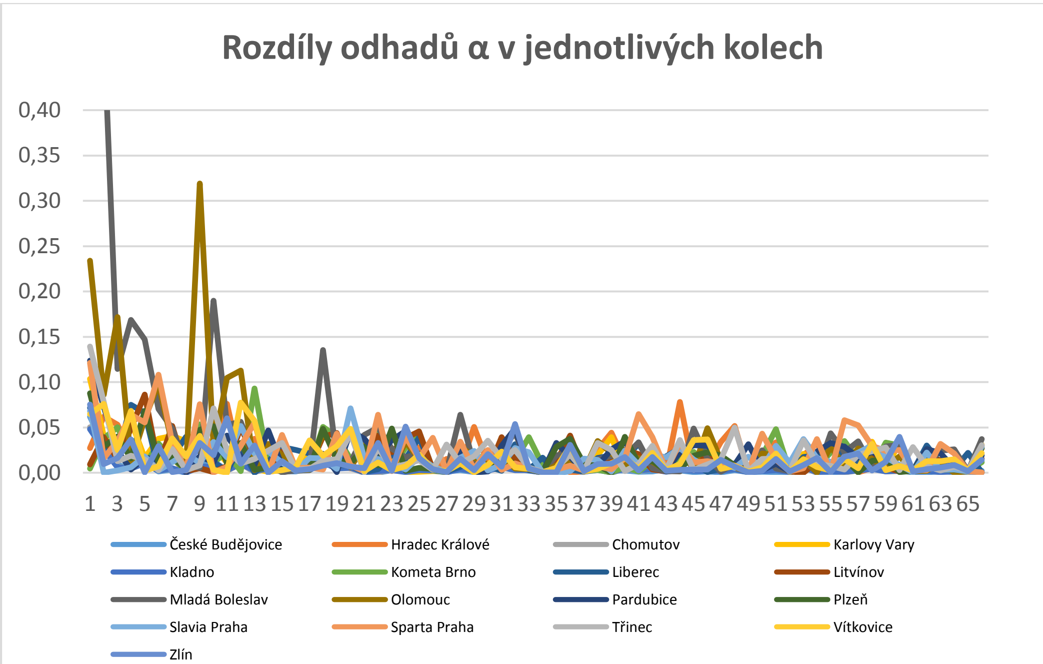
V této kapitole budeme odhadovat výsledky zápasů v sezóně 2015/2016. Zápasy ze starších sezón než 2009/2010 pro nás nejsou tolik důležité vzhledem k váhové funkci $\tau(t_m)$ a parametru ξ, protože váha těchto zápasů by byla velmi malá. Abychom mohli odhadovat výsledky utkání, nejprve je třeba určit váhu jednotlivých historických zápasů, k čemuž potřebujeme parametr ξ získaný maximalizací funkce $S(\xi)$ definované v rovnici (6.15). Poté se tento parametr použije pro odhadovanou sezónu 2015/2016.

6.6.1 Parametr ξ pro sezónu 2014/2015

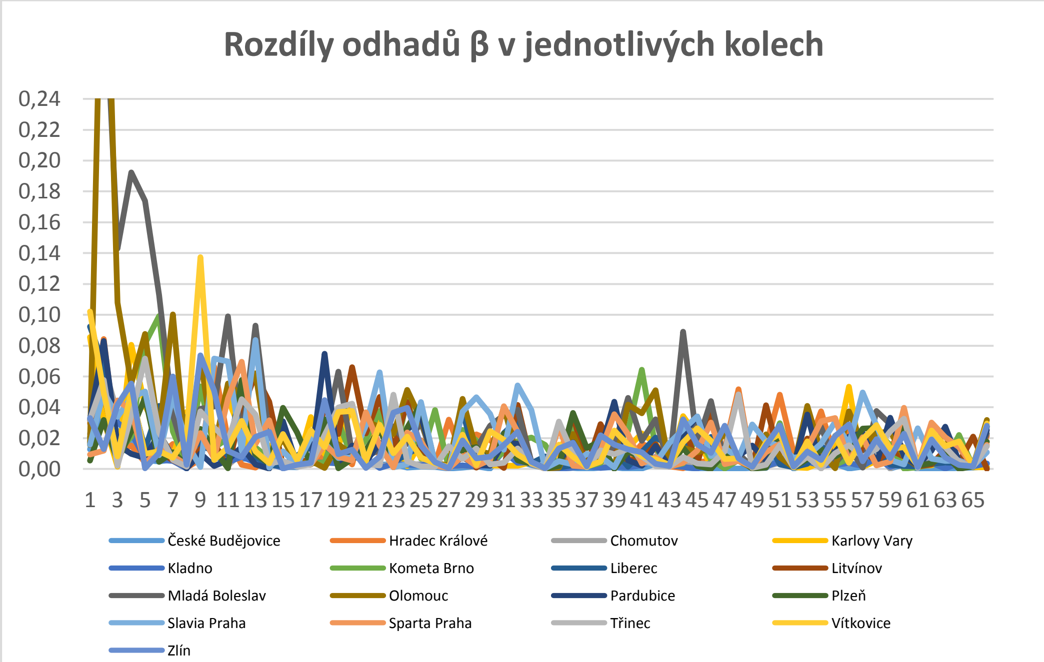
Ve všech modelech bylo potřeba nejprve odhadnout všechny parametry pro sezónu 2014/2015. Na této sezóně je možné maximalizací $S(\xi)$ (rovnice (6.15)) určit optimální hodnotu parametru ξ následně použitou pro novou odhadovanou sezónu 2015/2016.

Podle postupu z podkapitoly 6.5.3 se každý model z kapitoly 6.2 až 6.4 užil na odhadování pravděpodobností během sezóny a jejich síla je porovnána v této podkapitole. Výpočty pro všechny čtyři modely jsou v souboru *CZE_volba_ksi.xlsm*.

Při určování hodnoty parametru ξ odhadujeme parametry pro sezónu 2014/2015 až od 6. kola. Je to z toho důvodu, že odhady mohou být na začátku sezóny kolísavé, protože výkonnost mužstva před letní přestávkou a po ní se může významně změnit. Toto tvrzení je podloženo článkem [3] a zároveň vývojem rozdílů odhadů parametrů α a β v jednotlivých kolech pro každý tým zvlášť, který je zobrazen na Obr. 5 a Obr. 6. Odhady parametrů od 6. kola považujeme již za stabilní, kde rozdíly odhadovaných parametrů byly menší než 0,2. Při odhadování parametrů pro ostatní ligy také vynecháváme několik prvních kol, přesný počet bude uveden v podkapitole 6.7.1 a 6.8.1, kde je obdobné zpracování pro ostatní ligy.



Obr. 5: Rozdíly odhadů α v jednotlivých kolech – CZE

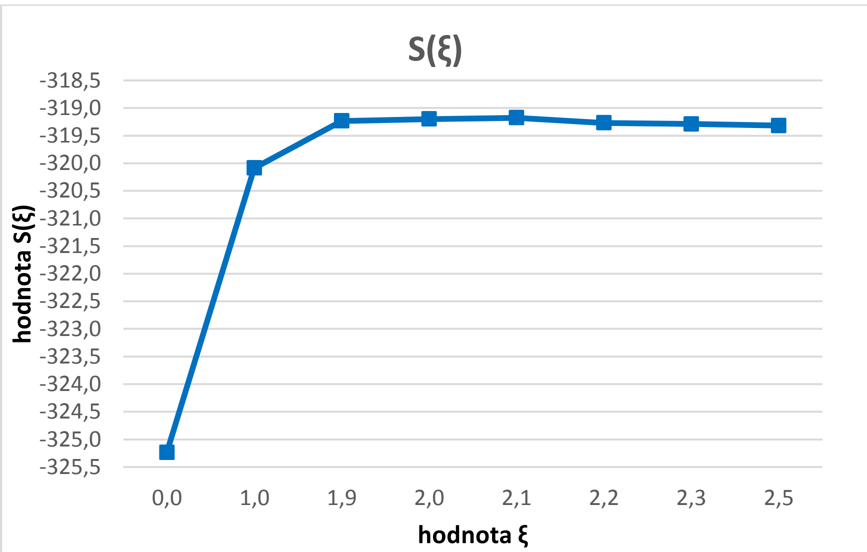


Obr. 6: Rozdíly odhadů β v jednotlivých kolech - CZE

Jako první musíme určit hodnotu $S(\xi)$, tj. míry chyby predikce výsledků v celé sezóně. Funkce $S(\xi)$ nám umožňuje porovnat modely a vybrat z nich ten nejlepší.

Na základě hodnoty této funkce vycházel nejlépe pro českou ligu model BP, který v porovnání s ostatními modely poskytoval maximální hodnotu $S(\xi)$ s optimální hodnotou parametru $\xi = 2{,}1$. Tato hodnota je optimální pro všechny modely. V okolí optimální hodnoty nejsou výsledky citlivé na malou změnu parametru ξ. Když je parametr $\xi = 2{,}1$, pak všechny modely vytvářejí nejlepší odhady z pohledu funkce $S(\xi)$. Pro výsledky zápasů starých jeden rok je váha 12,26 %, kterou jsme určili dosazením optimální hodnoty ξ do rovnice (6.11). Výsledky zápasů před dvěma lety mají váhu 1,50 %.

Hodnoty funkce $S(\xi)$ proti ξ pro celou sezónu jsou uvedeny na následujícím obrázku.



Obr. 7: Hodnoty $S(\xi)$ proti ξ pro BP model (CZE)

Přehled hodnot S(ξ) jednotlivých modelů pro různé hodnoty parametru ξ je uveden v Tab. 15. Mezi jednotlivými modely jsou nepatrné rozdíly v hodnotách S(ξ) v optimálním bodě ξ, a jsou vyznačeny žlutě. Model BP byl zvolen jako nejvhodnější, s hodnotou $S(\xi) = -319{,}177$, a je označena zeleně.

ξ	0,0	1,9	2,0	2,1	2,2	2,3	2,5
S(ξ) - DP	-325,240	-319,321	-319,311	-319,306	-319,309	-319,321	-319,368
S(ξ) - BP	-325,235	-319,236	-319,199	-319,177	-319,268	-319,289	-319,320
S(ξ) - DP-DI	-327,507	-319,980	-319,923	-319,466	-320,240	-320,252	-320,274
S(ξ) - BP-DI	-327,259	-319,865	-319,564	-319,500	-319,529	-319,541	-319,583

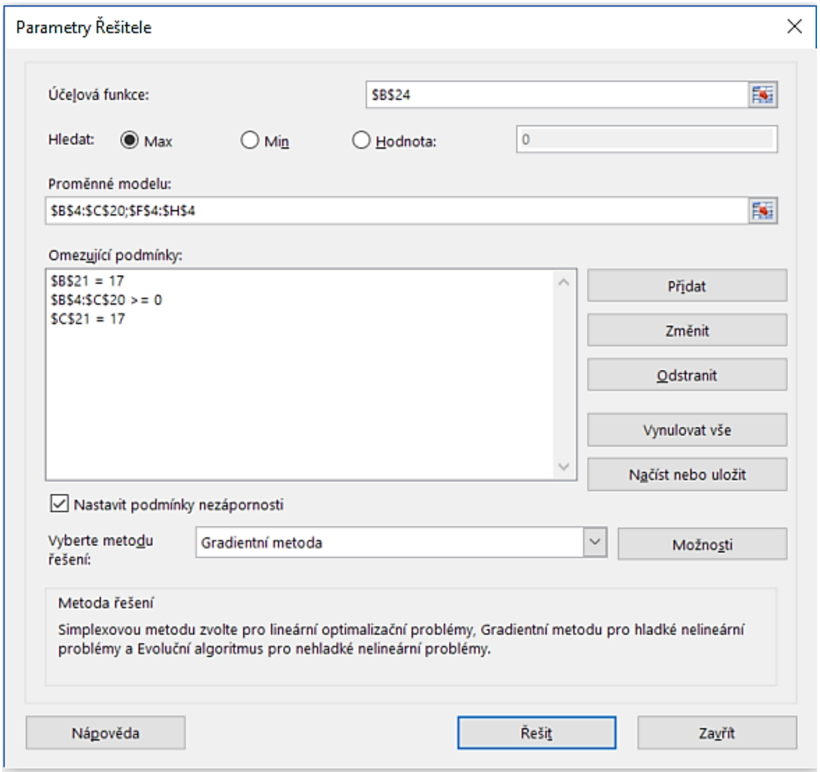
Tab. 15: Hodnoty $S(\xi)$ pro jednotlivé modely – Extraliga (CZE)

6.6.2 Odhad parametrů BP model 2015/2016

Všechny prezentované výsledky jsou založeny na nejlepším získaném modelu, což je podle předchozích výsledků BP model s parametrem ξ nastaveným na 2,1.

Jak již bylo uvedeno výše, odhady parametrů mohou být nestabilní v začátku sezóny. Z tohoto důvodu jsme vynechali odhad parametrů pro prvních 5 kol, tj. parametry odhadujeme od 6. kola. Parametry odhadnuté v tomto kole se následně použijí k odhadování pravděpodobností výsledků 7. kola (výhra domácích, remíza a výhra hostů), takto postupujeme i v dalších kolech.

Odhadování výsledků české ligy je vytvořeno v sešitu *CZE_BP_15-16.xlsm* na listu *BP model*. Odhad je proveden maximalizací věrohodnostní funkce (rovnice (6.12)). K maximalizaci se používá řešitel, ve kterém je nastavena přesnost omezující podmínky na 0,00001, tj. pokud se žádný z parametrů nezmění o více než 0,00001, výpočet skončí.



Obr. 8: Nastavení řešitele pro BP model v Microsoft Excel 2013

Při výpočtu se mění parametry síly útoku α_i a obrany β_i pro všechny týmy i, dále se mění parametr výhody domácího prostředí γ, parametr určující závislost mezi góly domácích a hostujících týmů λ a parametr μ.

Parametry α_i, β_i, γ a μ jsou pro všechny týmy i nezáporná reálná čísla, což vyplývá z logaritmické věrohodnostní funkce (6.18) a z významu parametrů α_i a β_i, které vyjadřují sílu útoku a obrany. Parametr λ je reálné číslo, které může být oproti ostatním uvedeným parametrům i záporné. Pro parametry α_i a β_i jsou nastavené podmínky z podkapitoly 6.2.1.

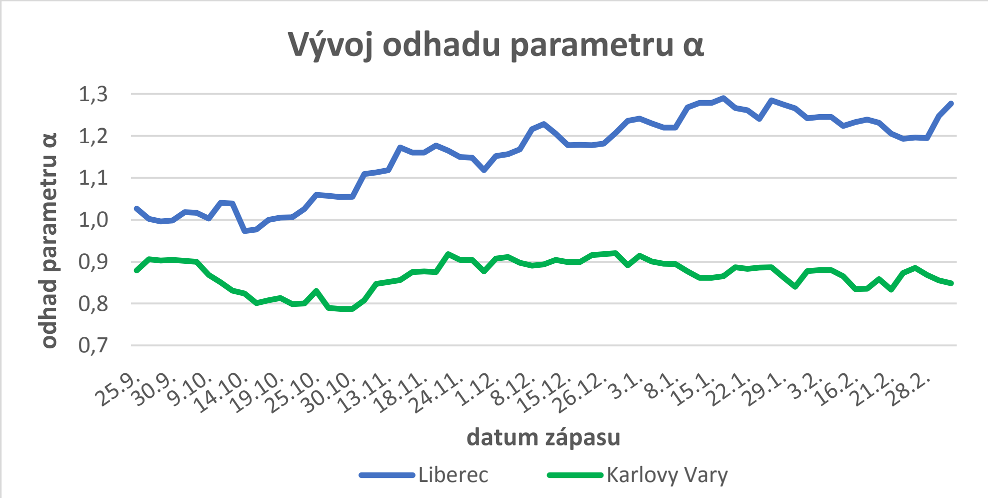
Před spuštěním řešitele, je třeba nastavit počáteční hodnoty parametrů. V tomto modelu byly nastaveny pro 6. kolo všechny parametry α_i, β_i, γ, λ a μ na 1. Pro další kola se vždy jako počáteční hodnoty používají hodnoty z předchozího kola.

Odhadnuté parametry 6. až 74. kola jsou uvedené v souboru *CZE_BP_model_15-16.xlsm* na listu *BP_model_parametry*. V následující tabulce jsou uvedené parametry α_i a β_i 73. kola použité k odhadu 74. kola.

Tým	α	β
České Budějovice	0,93	1,05
Hradec Králové	1,03	0,96
Chomutov	1,03	1,02
Karlovy Vary	0,86	1,25
Kladno	0,68	1,12
Kometa Brno	0,99	1,04
Liberec	1,25	0,79
Litvínov	0,84	0,92
Mladá Boleslav	1,15	1,00
Olomouc	0,84	0,83
Pardubice	0,89	1,07
Plzeň	1,18	0,97
Slavia Praha	0,82	1,28
Sparta Praha	1,50	0,92
Třinec	0,95	0,85
Vítkovice	1,08	1,07
Zlín	0,97	0,87
Celkem	**17,00**	**17,00**

Tab. 16: Odhad parametrů pro 74. kolo (tj. z výsledků do 73. kola včetně)

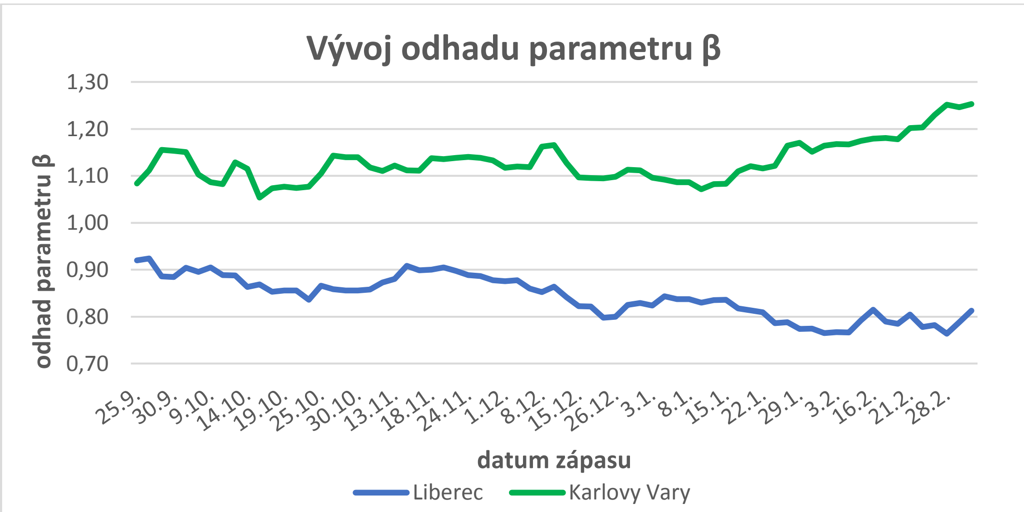
Na Obr. 9 je vidět, jak se mění odhad parametru útoku α během celé sezóny 2015/2016 pro nejlepší tým sezóny Liberec, a nejhorší mužstvo Karlovy Vary.



Obr. 9: Max. věrohodné odhady parametru útoku α pro nejlepší a nejhorší tým (CZE)

Jak je vidět, odhady jsou v souladu se skutečností, že Liberec byl lepší v útoku oproti Karlovým Varům, a to po celou sezónu. Odhad parametru síly útoku α pro Liberec měl rostoucí trend v celé sezóně.

Dobrá hodnota parametru pro útok ještě není zárukou toho, že mužstvo bude úspěšné, zároveň musí být podpořeno i kvalitní obranou - parametrem obrany β. Vývoj odhadu parametru obrany pro Liberec, nejlepší tým sezóny, a Karlovy Vary, nejhorší tým, je znázorněn na Obr. 10.



Obr. 10: Max. věrohodné odhady parametru obrany β pro nejlepší a nejhorší tým (CZE)

Jak již bylo uvedeno, čím nižší hodnota parametru β, tím je obrana mužstva lepší. Ve druhé části sezóny je zřejmé, že Liberec zlepšil svoji obranu a snížil tak hodnotu odhadu parametru pro obranu ještě více pod průměr, tj. pod hodnotu 1. Naopak Karlovy Vary zhoršily svoji obranu a zvýšila se tak hodnota odhadu parametru β na více než 1,2.

Další tři parametry jsou globální pro všechny týmy. Odhady parametru výhody domácího prostředí γ a parametru μ se téměř nemění v průběhu času. Větší změny těchto parametrů mohou být mezi sezónami, jelikož se mění sestavy jednotlivých týmů a hráči přechází mezi týmy, případně se dostávají do ligy noví hráči. Průměrná hodnota odhadu parametru γ je 1,24 a znamená to, že domácí tým skóruje přibližně 1,24 krát více gólů než hostující tým. Průměrná hodnota odhadu parametru μ je 2,28 a to lze chápat jako průměrný počet gólů hostujícího mužstva. Jestliže se tento parametr násobí odhadem parametru γ, může to být považováno za průměrný počet gólů domácího týmu, tj. 2,83. Průměrná hodnota odhadu parametru λ je 0,02.

6.6.3 Odhad výsledků zápasů 2015/2016

Pravděpodobnosti výsledků zápasů lze odhadnout pomocí sdružené pravděpodobnostní funkce pro BP model definované v rovnici (6.7). Pravděpodobnosti konkrétních výsledků zápasů se dopočítávají v souboru *CZE_BP_15-16.xlsm* na listu *BP sezóna 2015-2016* s použitím odhadnutých parametrů z listu *BP model_parametry*. Poté se vyjádří celkové pravděpodobnosti výher domácích, remíz a výher hostů.

Odhadnuté parametry pro zápas posledního hracího dne sezóny 4.3.2016 (tj. 74. „kola“) mezi Plzní a Chomutovem jsou zobrazeny v následující tabulce.

Parametr	Plzeň - domácí	Chomutov - hosté
α	1,183	1,032
β	0,974	1,024
γ	1,287	
λ	< 0,001	
μ	2,239	

Tab. 17: Odhadnuté parametry pro zápas Plzně s Chomutovem

Pro tento zápas je odhad průměrného počtu gólů domácích (λ_{PLZ}) a průměrného počtu gólů hostů (λ_{CHO}) uveden v následujících rovnicích

$$\hat{\lambda}_{PLZ} = \hat{\mu} \cdot \hat{\alpha}_{PLZ} \cdot \hat{\beta}_{CHO} \cdot \hat{\gamma} = 2{,}239 \cdot 1{,}183 \cdot 1{,}024 \cdot 1{,}287 = 3{,}492, \tag{6.16}$$

$$\hat{\lambda}_{CHO} = \hat{\mu} \cdot \hat{\alpha}_{CHO} \cdot \hat{\beta}_{PLZ} = 2{,}239 \cdot 1{,}032 \cdot 0{,}974 = 2{,}251. \tag{6.17}$$

Známe-li tyto parametry, můžeme je dosadit do pravděpodobnostní funkce pro BP model (rovnice (6.7)) a určit tak pravděpodobnost konkrétního výsledku.

Tento zápas skončil 3:1 výhrou pro domácí Plzeň, proto v následující rovnici uvádíme pravděpodobnost tohoto výsledku určeného před zápasem dle zvoleného BP modelu

$$P(X = 3, Y = 1) = (3{,}492^3 \cdot 2{,}251^1 \cdot e^{-3{,}492-2{,}251}) \cdot$$
$$\cdot \frac{1 - 0 \cdot \left(e^{-3} - e^{-(1-e^{-1})\cdot 3{,}492}\right) \cdot \left(e^{-1} - e^{-(1-e^{-1})\cdot 2{,}251}\right)}{3! \cdot 1!} = 0{,}051 \tag{6.18}$$

Stejným způsobem se určí pravděpodobnost pro všechny možné výsledky do stavu 15:15 (teoreticky však až do výsledku ∞:∞), což by měla být dostačující hodnota skóre vzhledem k tomu, že 15 gólů nedalo žádné mužstvo ve všech dostupných zápasech. Pravděpodobnost všech možností výsledku v jednotlivých zápasech (výhra domácí, remíza a výhra hosté) musí být rovna jedné, což nebylo ve všech případech splněno. Proto jsme spočítané pravděpodobnosti znormovali následujícím způsobem

$$\frac{p_v}{p},$$

kde p_v je pravděpodobnost jednotlivých možností výsledku zápasu $v = \{výhra\ domácí; remíza; výhra\ hostů\}$ a p je pravděpodobnost všech těchto možností.

V následující tabulce jsou pravděpodobnosti pro jednotlivé výsledky tohoto zápasu.

Počet gólů domácí/hosté		Chomutov							
		0	1	2	3	4	5	6	7 a více
Plzeň	0	0,003	0,007	0,008	0,006	0,003	0,002	0,001	0,000
	1	0,011	0,025	0,028	0,021	0,012	0,005	0,002	0,001
	2	0,020	0,044	0,050	0,037	0,021	0,009	0,004	0,002
	3	0,023	0,051	0,058	0,043	0,024	0,011	0,004	0,002
	4	0,020	0,045	0,050	0,038	0,021	0,010	0,004	0,002
	5	0,014	0,031	0,035	0,026	0,015	0,007	0,003	0,001
	6	0,008	0,018	0,020	0,015	0,009	0,004	0,001	0,001
	7 a více	0,007	0,015	0,017	0,013	0,007	0,003	0,001	0,000

Tab. 18: Pravděpodobnosti výsledků pro zápas Plzně s Chomutovem

Výsledek		
Výhra domácí [1]	Remíza [0]	Výhra hosté [2]
0,619	0,151	0,230

Tab. 19: Pravděpodobnost výhry domácích, remízy a výhry hostů

V předchozí tabulce jsou uvedené pravděpodobnosti výsledku tohoto zápasu, kde největší pravděpodobnost je, že vyhraje domácí tým Plzeň. Konkrétní výsledek zápasu nepotřebujeme znát a ani se ho nesnažíme odhadnout, chceme jen vědět pravděpodobnosti konečného stavu zápasu.

6.6.4 Porovnání modelů pro Extraligu

Tato podkapitola se zabývá porovnáním dvou zvolených modelů z předchozí části pro českou Extraligu, a to Maherův model 2 a BP model. Oba modely budeme nyní porovnávat na odhadované sezóně 2015/2016. Jak bylo uvedeno již výše, pro porovnání modelů nám poslouží funkce $S(\xi)$ (definovaná v rovnici (6.15), podle níž určíme, který z těchto dvou modelů je pro hokejová data z Extraligy vhodnější.

K porovnání těchto modelů musíme nejprve vyjádřit funkci $S(\xi)$ pro Maherův model pro sezónu 2015/2016, stejným způsobem jako jsme ji počítali pro BP model, tj. od 7. kola.

Vzhledem k tomu, že v Maherově modelu platí, že všechny zápasy v sezóně mají stejnou váhu, není třeba určovat optimální ξ pro váhu jednotlivých zápasů (definovanou v rovnici (6.11)), jako tomu bylo u BP modelu. V tomto případě, je tedy ξ rovno 0.

Odhad zápasů hraných v sezóně 2015/2016 včetně vypočtené hodnoty $S(\xi)$ je proveden v souboru *CZE_model_Maher* na listu *odhad sezóny 2015-2016*. Pro odhad těchto zápasů jsme použili odhadnuté parametry z předchozí sezóny (2014/2015).

Výsledné hodnoty funkce $S(\xi)$ pro oba zkoumané modely jsou uvedené v následující tabulce.

	Maherův model 2	BP model
S(ξ)	-372,33	-335,97

Tab. 20: Porovnání hodnot $S(\xi)$ pro Maherův model 2 a BP model

Jak již bylo uvedeno v předchozí kapitole, pro hodnotu $S(\xi)$ platí, že čím více se hodnota blíží k nule, tím lepší predikci získáme. Proto můžeme říci, že model s vyšší hodnotou této funkce je vhodnější pro naše data. Při porovnání těchto modelů vychází vyšší hodnota $S(\xi)$ ve prospěch BP modelu, který je tedy na základě takového kritéria vhodnější pro data z české ligy.

Tímto jsme také potvrdili tvrzení z podkapitoly 5.3.4.

6.7 Model pro Ekstraligu (POL)

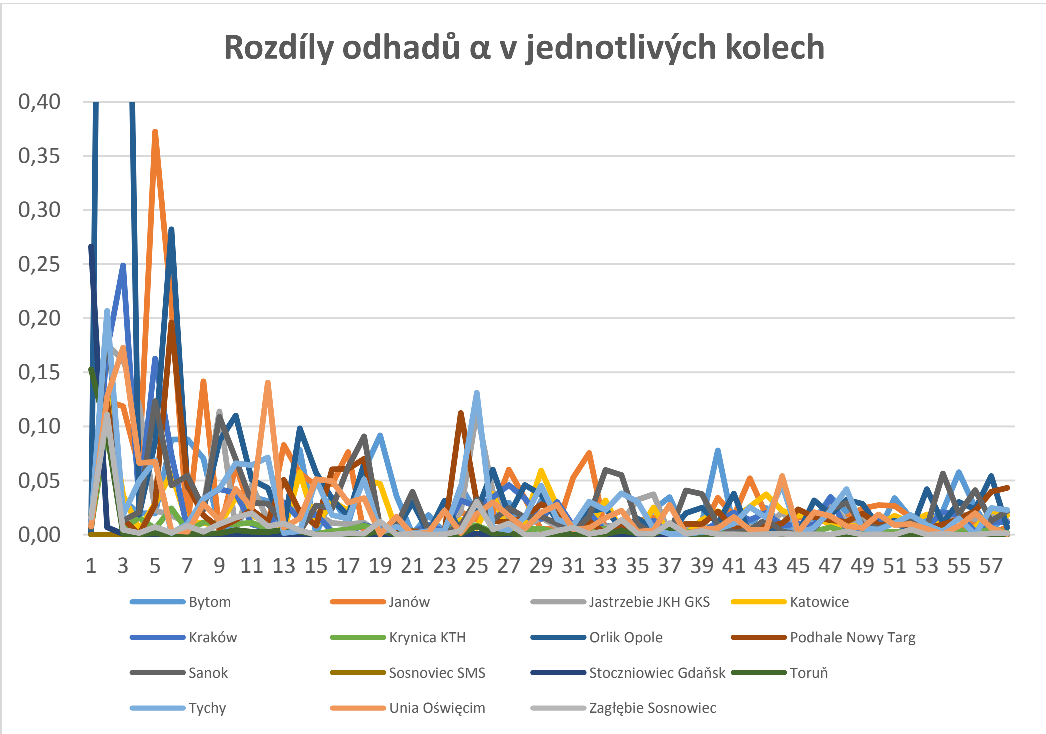
Stejně jako v předchozí podkapitole 6.6, i zde budeme odhadovat výsledky zápasů v sezóně 2015/2016, ovšem nyní pro polskou ligu. Odhadování výsledků je provedeno v sešitu *POL_BP 15-16.xlsm* na listu *BP model*.

Ale ještě před samotným odhadováním sezóny 2015/2016 je potřeba, stejně jako u české ligy, zvolit optimální hodnotu parametru ξ na sezóně 2014/2015 a poté tento parametr použít pro odhadování sezóny 2015/2016.

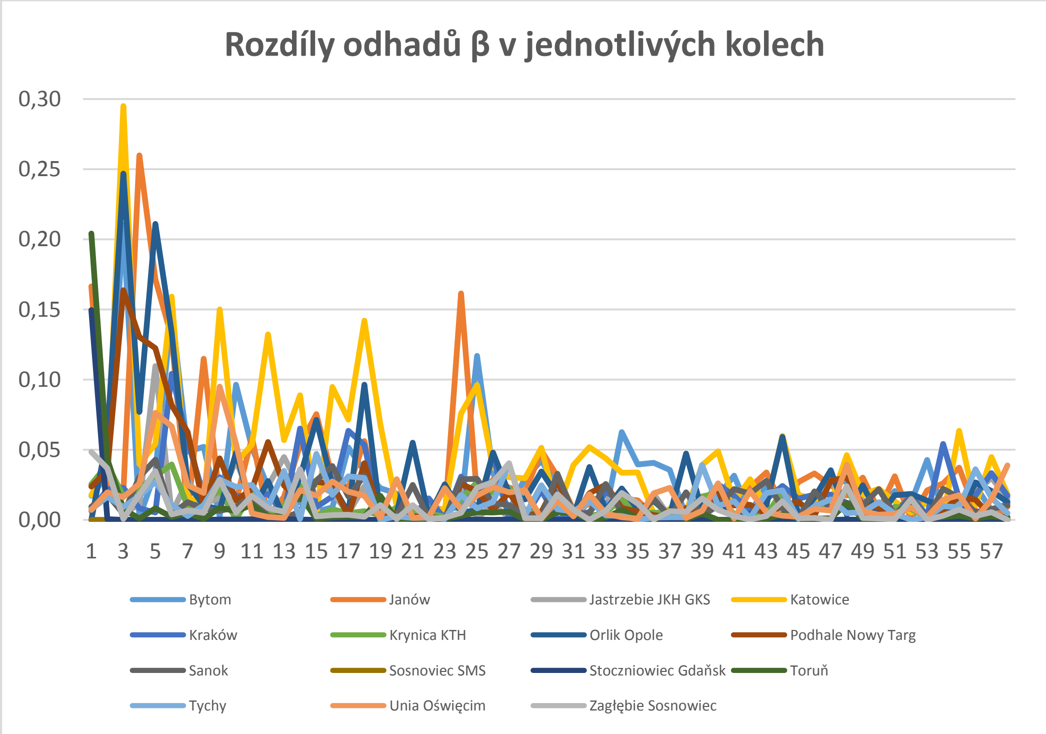
6.7.1 Parametr ξ pro sezónu 2014/2015

Při určování hodnoty parametru ξ odhadujeme parametry pro sezónu 2014/2015 od 7. kola. Výpočty pro všechny čtyři modely jsou v souboru *POL_volba_ksi.xlsm*.

V této lize odhadujeme všechny parametry až od 7. kola, protože v prvních šesti kolech byl odhad parametrů vychýlený, znázorněno na Obr. 11 a Obr. 12.



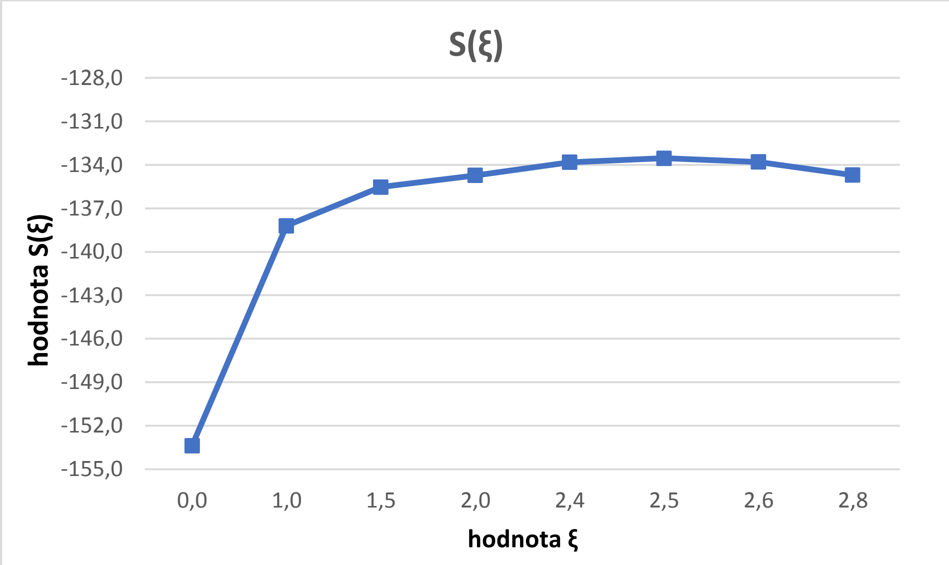
Obr. 11: Rozdíly odhadů α v jednotlivých kolech – POL



Obr. 12: Rozdíly odhadů β v jednotlivých kolech - POL

Nejprve musíme určit hodnotu $S(\xi)$ pro odhadované zápasy v sezóně 2014/2015 pro různé hodnoty parametru ξ, a určit podle toho optimální hodnotu ξ. Ukázka hodnot $S(\xi)$ pro různá ξ je zobrazena na Obr. 13. Ačkoliv BP-DI model poskytoval maximální hodnotu funkce $S(\xi)$ ze všech modelů, tak v porovnání s BP modelem byla tato hodnota pouze o 0,008 větší a odhady parametru p pro jednotlivá kola vycházely většinou nulové, tj. parametr p neměl téměř žádný vliv na zvýšení pravděpodobností remíz. Z tohoto důvodu byl zvolen jako nejvhodnější BP model, který je zároveň jednodušší. Tento závěr potvrzuje i skutečnost, že počet remíz v polské lize je velmi nízký, například v sezóně 2014/2015 jich bylo pouze 22.

U všech modelů závisí $S(\xi)$ na hodnotě parametru ξ. V modelu DP a DP-DI je optimální hodnota ξ rovna 2,6 a pro model BP i BP-DI vychází optimální hodnota tohoto parametru 2,5. Po dosazení $\xi = 2{,}5$ do rovnice (6.11) můžeme říci, že váha výsledků zápasů starých jeden rok je 8,22 %, zatímco stejné výsledky před dvěma lety mají váhu 0,68 %.



Obr. 13: Hodnoty $S(\xi)$ proti ξ pro BP model (POL)

V následující tabulce je uveden přehled hodnot $S(\xi)$ jednotlivých modelů pro různé hodnoty parametru ξ. Z tabulky je zřejmé, že mezi jednotlivými modely jsou velmi malé rozdíly hodnot $S(\xi)$ v optimálních bodech ξ, které jsou vyznačeny žlutě.

Pro zvolený BP model je optimální hodnota ξ označena v tabulce zelenou barvou.

ξ	0,0	1,0	2,0	2,4	2,5	2,6	2,7	3,0
S(ξ) - DP	-153,365	-138,206	-133,982	-133,584	-133,555	-133,550	-133,566	-133,724
S(ξ) - BP	-153,372	-138,204	-134,731	-133,818	-133,549	-133,802	-134,051	-133,709
S(ξ) - DP-DI	-153,484	-138,456	-133,987	-133,585	-133,557	-133,555	-133,575	-133,712
S(ξ) - BP-DI	-153,340	-139,035	-134,354	-133,573	-133,541	-133,546	-133,616	-133,700

Tab. 21: Hodnoty $S(\xi)$ pro jednotlivé modely – Ekstraliga (POL)

6.7.2 Odhad parametrů BP model 2015/2016

V této části jsou všechny výsledky založeny na nejvhodnějším modelu, což je podle předchozí kapitoly BP model s parametrem ξ nastaveným na hodnotu 2,5.

Odhad je proveden maximalizací věrohodnostní funkce (6.14). K maximalizaci se používá opět řešitel, který je nastaven obdobně jako u české ligy. Před spuštěním řešitele, je třeba zvolit počáteční hodnoty parametrů, které byly v tomto modelu pro 7. kolo nastavené stejně jako u české ligy.

Při výpočtu se opět mění parametry α_i, β_i pro všechny týmy i, dále se mění parametr γ, λ a parametr μ.

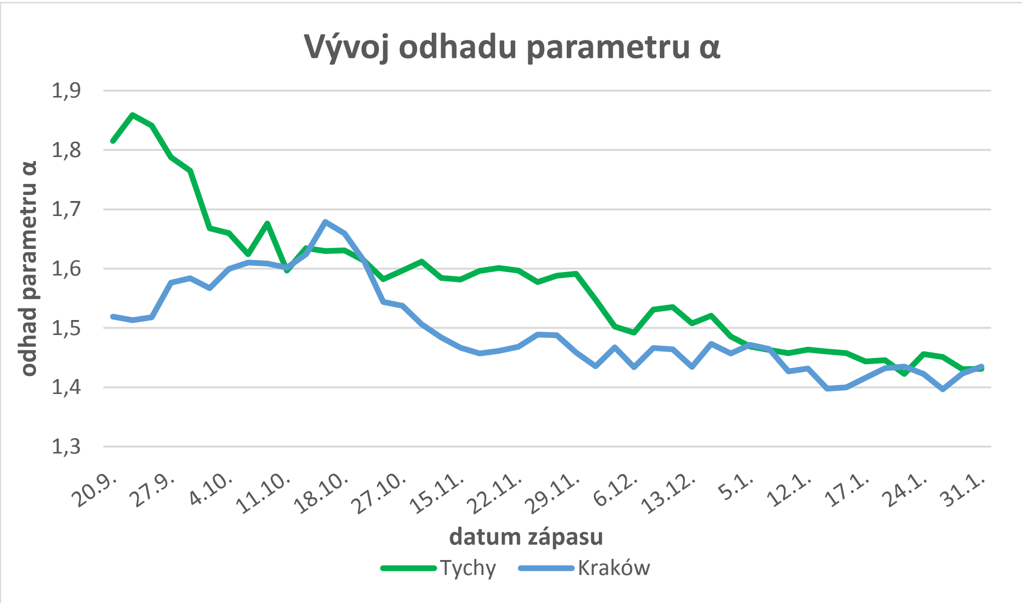
Tým	α	β
Bytom	1,03	0,92
Janów	0,76	1,12
Jastrzebie JKH GKS	1,02	0,53
Katowice	0,70	1,99
Kraków	1,42	0,49
Krynica KTH	0,74	1,29
Orlik Opole	1,04	0,86
Podhale Nowy Targ	1,31	0,60
Sanok	1,07	0,54
Sosnoviec SMS	0,47	1,90
Stoczniowiec Gdańsk	1,30	0,94
Toruň	0,86	1,38
Tychy	1,43	0,47
Unia Oświęcim	1,07	0,78
Zagłębie Sosnowiec	0,78	1,21
Celkem	**15,00**	**15,00**

γ	λ	μ
1,13	0,00	3,70

Obr. 14: Odhad parametrů pro 52. kolo (tj. z výsledků do 51. kola včetně)

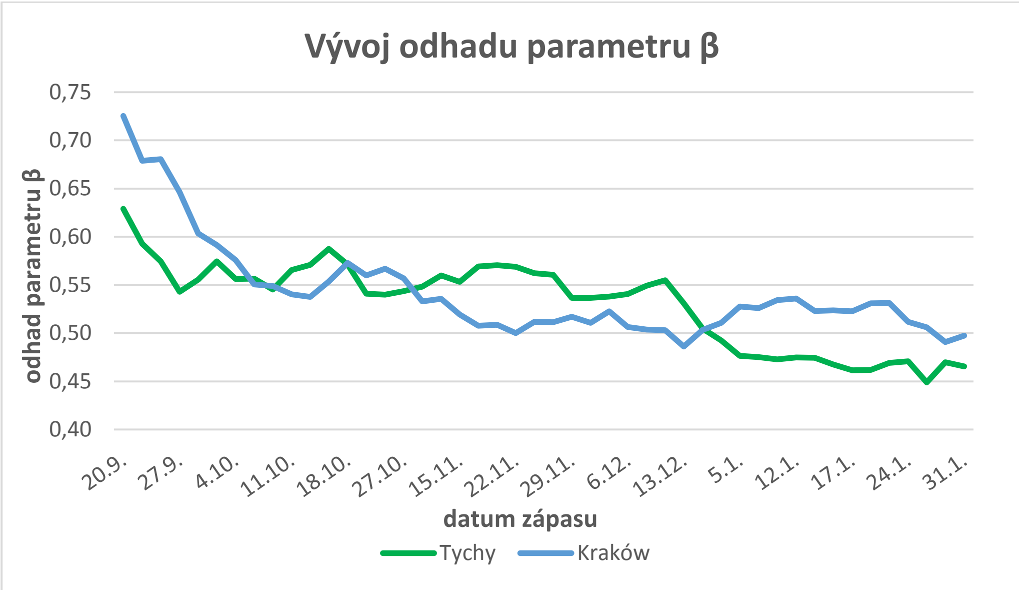
Parametry α_i, β_i, γ a μ jsou nezáporná reálná čísla pro všechny týmy i. Parametr λ je reálné číslo, které může být záporné. Pro odhady parametrů α_i a β_i jsou nastavené podmínky z podkapitoly 6.2.1.

Na Obr. 15 je zobrazen vývoj odhadu parametru útoku α během celé sezóny 2015/2016 pro dva nejlepší týmy sezóny – Kraków a Tychy. Parametry se odhadovaly pokaždé, když se hrál zápas, a získala se tak časová posloupnost 46 záznamů (pro 7. až 52. kolo). Oba týmy měly od 13. kola (tj. od 4.10.2015) až do konce sezóny velmi vyrovnanou sílu v útoku.



Obr. 15: Max. věrohodné odhady parametru útoku α pro dva nejlepší týmy (POL)

Vývoj odhadu parametru obrany pro Kraków, nejlepší tým sezóny, a Tychy, 2. nejlepší tým, je zobrazen na Obr. 16.



Obr. 16: Max. věrohodné odhady parametru obrany β pro dva nejlepší týmy (POL)

Od 39. kola (3.1.2016) Tychy zlepšil svoji obranu a snížila se tak hodnota odhadu parametru síly v obraně až na hodnoty pod 0,5. Naopak Kraków zhoršil svoji obranu, a proto se zvýšila hodnota odhadu parametru β.

Průměrná hodnota odhadu parametru γ je 1,13, tj. domácí tým skóruje přibližně 1,13 krát více gólů než hostující tým. Průměrná hodnota odhadu parametru μ je 3,55. Jestliže vynásobíme tento parametr odhadem parametru γ, může říci, že průměrný počet gólů domácího týmu je 4,01. Průměrná hodnota odhadu parametru λ je -0,16, a to znamená, že korelační koeficient ρ je záporný (dle rovnice (6.8)), tj. X_{ij} a Y_{ij} jsou dvě závislé náhodné veličiny, ale pouze dle hodnoty tohoto parametru.

Toto tvrzení bylo ještě ověřeno χ^2-testem nezávislosti (viz kapitola 3.5). Tento test je zpracován pro sezónu 2015/2016 v sešitu *POL_BP 15-16* na listu *chí-kvadrát test*. Výsledky pouze potvrzují závěr na základě korelačního koeficientu ρ, tj. X_{ij} a Y_{ij} nejsou nezávislé náhodné veličiny.

Nyní když jsou známé všechny parametry, můžeme odhadnout výsledky zápasů pomocí sdružené pravděpodobnostní funkce pro BP model. Odhad výsledků zápasů a pravděpodobnosti výher domácích, remíz a výher hostů jsou zaznamenány na listu *BP sezóna 2015-2016.*

6.8 Model pro NHL

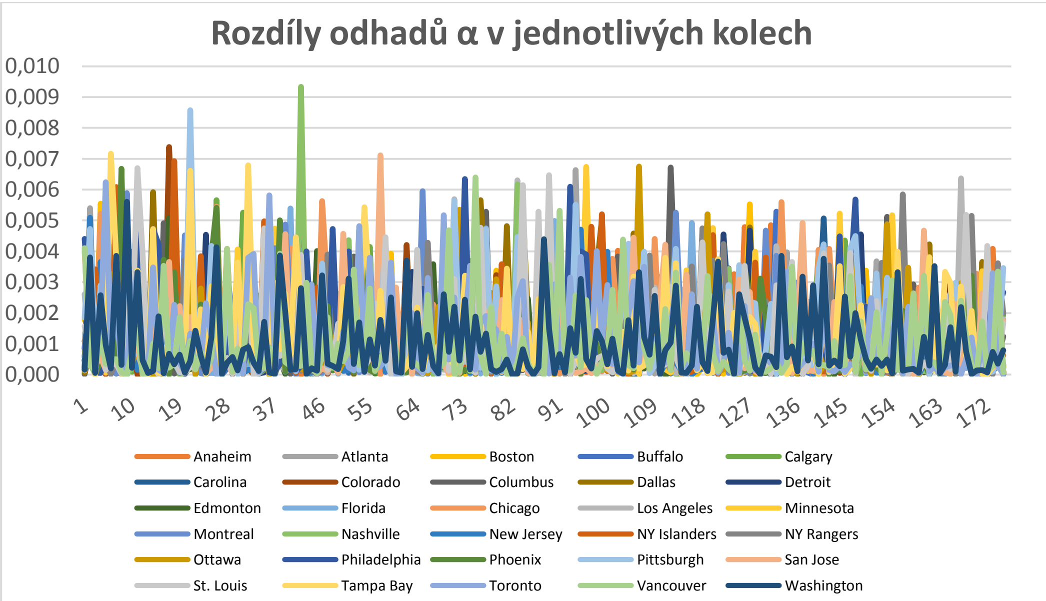
V této kapitole se budou odhadovat výsledky zápasů pro stejnou sezónu jako u ostatních lig. Odhadování výsledků zápasů NHL ligy pro sezónu 2015/2016 je provedeno v sešitu *NHL_BP-DI 15-16.xlsm* na listu *BP-DI model*.

Zde je stejný způsob odhadování jako u předchozích lig, tj. nejprve musíme určit hodnotu parametru ξ na předchozí sezóně.

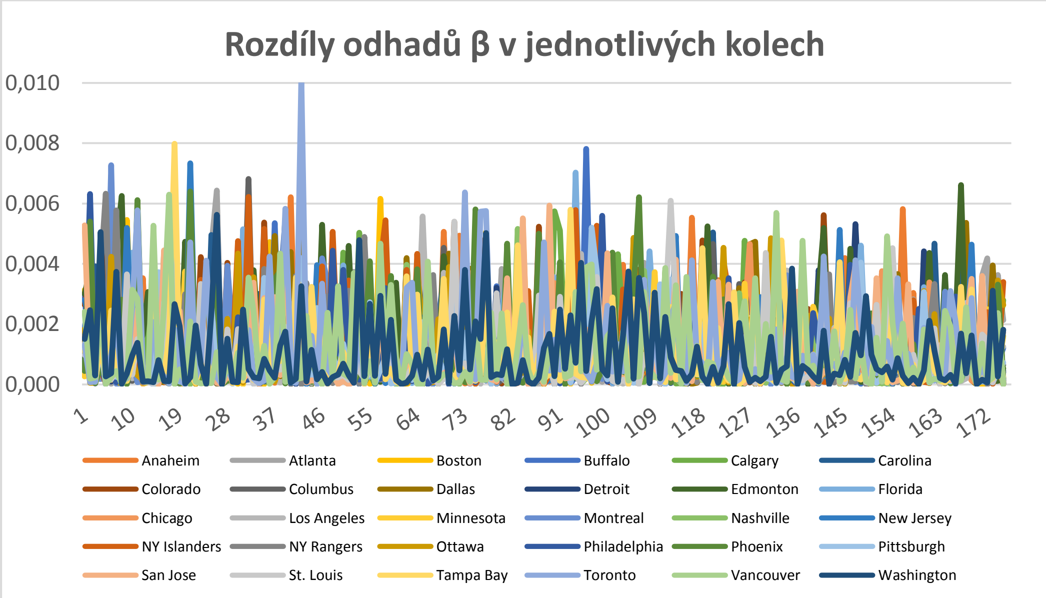
6.8.1 Parametr ξ pro sezónu 2014/2015

Při určování hodnoty parametru ξ odhadujeme parametry pro sezónu 2014/2015 od 15. kola. Výpočty jsou rozdělené do čtyř souborů nazvaných podle konkrétního modelu, například *NHL_DP_volba_ksi.xlsm*, atd.

V následujících grafech jsou znázorněny rozdíly mezi jednotlivými odhady, stejně jako u předchozích lig.



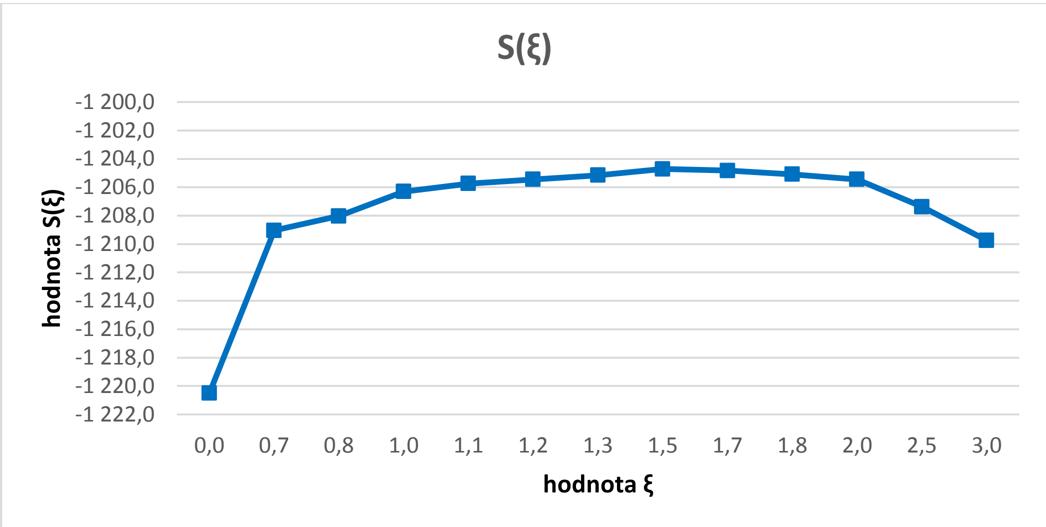
Obr. 17: Rozdíly odhadů α v jednotlivých kolech – NHL



Obr. 18: Rozdíly odhadů β v jednotlivých kolech - NHL

V NHL lize pozorujeme, že jednotlivé odhady jsou poměrně stabilní již od začátku sezóny. Nemůžeme však říci určitě, že toto platí vždy a pro každou sezónu a proto použijeme stejný postup jako u předchozích lig. Pravděpodobnosti výsledků zápasů budeme odhadovat od 16. kola, tj. od 16. hracího dne.

Nejdříve opět určíme hodnotu $S(\xi)$. Na Obr. 19 jsou hodnoty $S(\xi)$ pro různé ξ pro všechny zápasy dané sezóny. Porovnáme-li všechny čtyři modely na základě hodnoty $S(\xi)$, tak model BP-DI poskytoval maximální hodnotu $S(\xi)$. Optimální volba parametru ξ pro BP-DI a DP-DI model je 1,5. Pro model BP je optimální hodnota ξ rovna 1,9, a u modelu DP tato hodnota vychází 0,2. Optimální hodnota 1,5 pro BP-DI model v našem případě znamená, že váha výsledků zápasů starých jeden rok je 22,34 %, kdežto stejných výsledků před dvěma lety pouze 4,99 %.



Obr. 19: Hodnoty $S(\xi)$ proti ξ pro BP-DI model (NHL)

V následující tabulce je přehled hodnot $S(\xi)$ jednotlivých modelů pro různé hodnoty parametru ξ. Mezi modely jsou výrazné rozdíly v hodnotách $S(\xi)$ v optimálním bodě ξ, a jsou zvýrazněné žlutě. Na základě maximalizace $S(\xi)$ byl vybrán model BP-DI, který vyšel s hodnotou -1 204,719, tj. o téměř 600 vyšší hodnota oproti BP modelu, a v tabulce je vyznačen zeleně[8].

ξ	0,0	1,0	1,2	1,3	1,4	1,5	1,6	1,8	2,0	2,5	3,0
S(ξ) - DP-DI	-1 220,468	-1 206,284	-1 205,308	-1 204,972	-1 204,808	-1 204,721	-1 204,759	-1 205,050	-1 205,529	-1 207,470	-1 209,881
S(ξ) - BP-DI	-1 220,511	-1 206,302	-1 205,447	-1 205,156	-1 204,815	-1 204,719	-1 204,737	-1 205,087	-1 205,451	-1 207,379	-1 209,744

Tab. 22: Hodnoty $S(\xi)$ pro diagonálně rozšířené modely – NHL

ξ	0,0	0,1	0,2	0,3	0,4	0,5	1,0	2,0	2,1	2,2	2,6
S(ξ) - DP	-1 809,837	-1 809,213	-1 808,961	-1 809,082	-1 809,493	-1 810,149	-1 814,499	-1 821,466	-1 822,094	-1 822,719	-1 825,198

Tab. 23: Hodnoty S(ξ) pro DP model – NHL

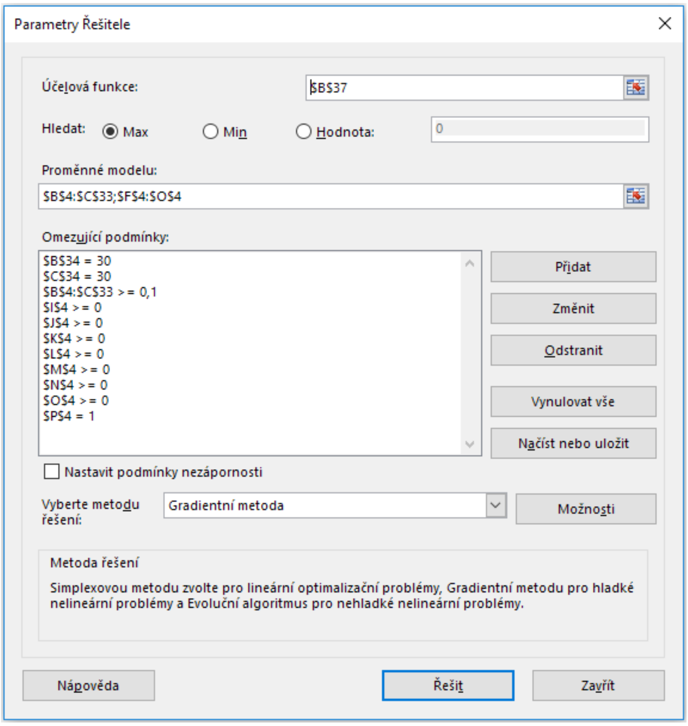
ξ	0,0	0,4	0,5	1,0	1,5	1,7	1,8	1,9	2,0	2,1	2,5	2,8
S(ξ) - BP	-1 809,331	-1 807,524	-1 807,151	-1 801,833	-1 798,949	-1 798,957	-1 798,168	-1 798,106	-1 798,310	-1 798,980	-1 799,419	-1 800,246

Tab. 24: Hodnoty S(ξ) pro BP model - NHL

6.8.2 Odhad parametrů BP-DI model 2015/2016

Všechny uvedené výsledky pro NHL ligu jsou založeny na nejlepším získaném modelu, což je dle předchozí podkapitoly BP-DI model s parametrem ξ nastaveným na 1,5.

Odhad je proveden maximalizací věrohodnostní funkce (6.14). K maximalizaci se používá opět řešitel, který je nastaven stejným způsobem jako u předchozích lig.



Obr. 20: Nastavení řešitele pro BP-DI model v Microsoft Excel 2013

[8] Jednotlivé tabulky mají odlišný obor hodnot z toho důvodu, že se nejprve snažíme zjistit okolí pro parametr ξ, kde jsou nejvyšší hodnoty funkce S(ξ). Podle toho jsme přizpůsobili obor hodnot pro lepší znázornění a určení optimální hodnoty parametru ξ. To samé platí i v kapitole 7.

Při výpočtu se mění parametry[9] α_i a β_i pro všechny týmy *i*, dále se mění parametr γ, λ, μ, p a parametry $\theta_{0,\dots,5}$.

Tým	α	β
Anaheim	1,02	0,89
Atlanta	0,98	1,04
Boston	1,06	1,03
Buffalo	0,87	1,02
Calgary	1,06	1,16
Carolina	0,91	1,01
Colorado	0,97	1,09
Columbus	1,01	1,12
Dallas	1,20	1,06
Detroit	0,97	1,02
Edmonton	0,91	1,16
Florida	1,04	0,95
Chicago	1,05	0,92
Los Angeles	0,98	0,88
Minnesota	0,99	0,90
Montreal	0,96	1,06
Nashville	1,04	0,93
New Jersey	0,79	0,94
NY Islanders	1,05	0,98
NY Rangers	1,09	0,95
Ottawa	1,05	1,08
Philadelphia	0,97	0,96
Phoenix	0,89	1,11
Pittsburgh	1,12	0,91
San Jose	1,07	0,97
St. Louis	1,03	0,88
Tampa Bay	1,05	0,92
Toronto	0,89	1,11
Vancouver	0,87	1,08
Washington	1,12	0,90
Celkem	**30,00**	**30,00**

γ	λ	μ	p	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	suma θ_k
1,09	-1,59	2,57	0,08	0,10	0,39	0,27	0,16	0,07	0,00	1,00

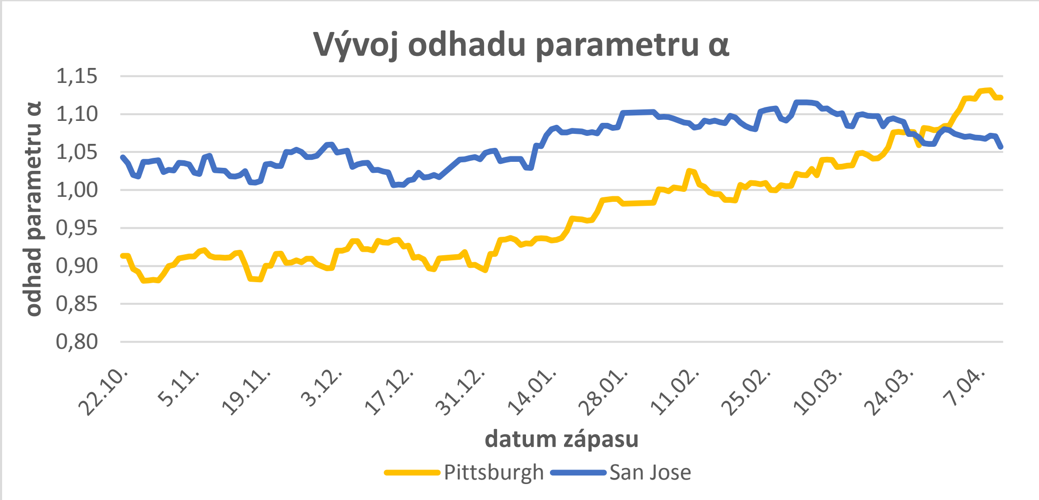
Obr. 21: Odhad parametrů pro 177. kolo (tj. z výsledků do 176. kola včetně)

Parametry α_i, β_i, γ a μ jsou nezáporná reálná čísla pro všechny týmy *i*. Parametr $p \in [0;1]$ a θ_k pro k = 0,…,5 platí $\theta_k \geq 0$ pro všechna k, $\sum_{k=0}^{5} \theta_k = 1$. Parametr λ je reálné číslo, které může být záporné.

Před spuštěním řešitele, je třeba nastavit počáteční hodnoty parametrů. V tomto modelu byly nastaveny pro 15. kolo parametry α_i, β_i, γ, λ a μ na 1, p na 0,5, $\theta_0 = 0{,}35$, $\theta_1 = 0{,}30$, $\theta_2 = 0{,}20$, $\theta_3 = 0{,}10$, $\theta_4 = 0{,}05$ *a* $\theta_5 = 0$. Pro další kola se vždy jako počáteční hodnoty používají hodnoty z předchozího kola.

Na Obr. 22 je vidět, jak se mění odhad parametru útoku α během celé sezóny 2015/2016 pro dva nejlepší týmy sezóny, Pittsburgh a San Jose. Jak již bylo uvedeno výše, parametry se odhadovaly pokaždé, když se hrál zápas, a tak se získala pro každý parametr časová posloupnost 163 záznamů (pro 15. až 177. kolo).

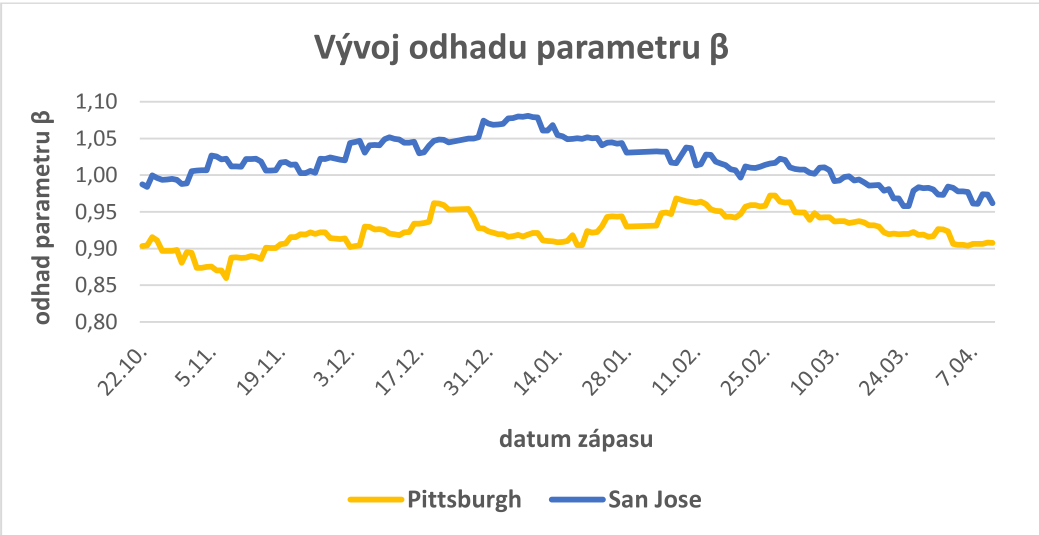
[9] Původní podmínka modelu byla, že parametr $\alpha_i > 0$ a $\beta_i > 0$, v tomto případě ale při nastavení této podmínky přestával řešitel fungovat a výpočet proběhl pouze pro první odhadované kolo a v ostatních kolech se odhad vůbec nezměnil. Z tohoto důvodu byla podmínka pozměněna a nastavena tak, že $\alpha_i \geq 0{,}1$ a $\beta_i \geq 0{,}1$. Nejhorší týmy (resp. nejlepší týmy pro parametr β) mají obvykle koeficienty mnohem vyšší.



Obr. 22: Max. věrohodné odhady parametru útoku α pro dva nejlepší týmy (NHL)

Odhad parametru síly útoku α Pittsburghu měl rostoucí trend po celou sezónu. Tým San Jose byl téměř celou sezónu lepší v útoku než Pittsburgh.

Vývoj odhadu parametru obrany pro Pittsburgh, nejlepší tým sezóny, a San Jose, 2. nejlepší tým, je znázorněn na Obr. 23.

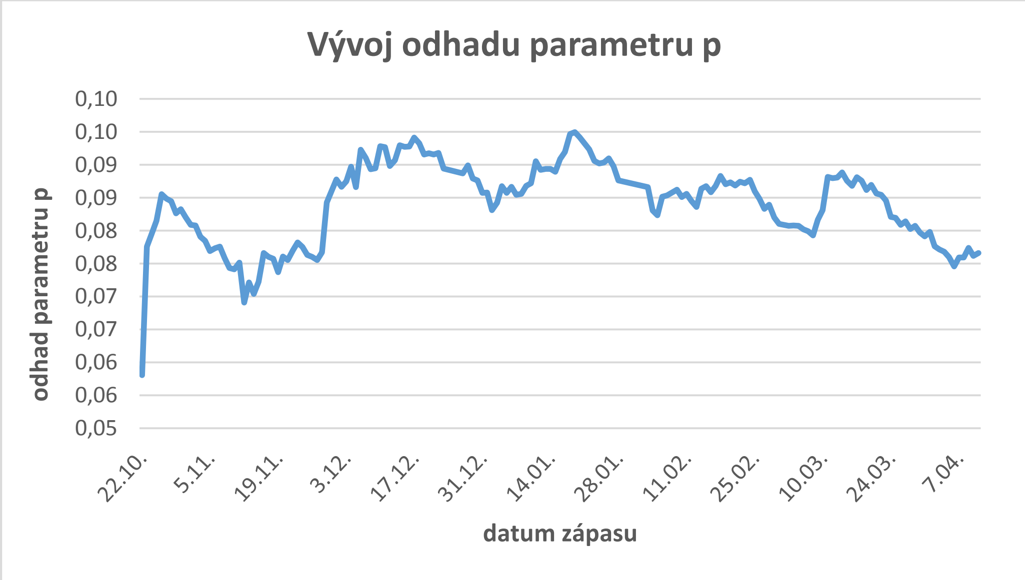


Obr. 23: Max. věrohodné odhady parametru obrany β pro dva nejlepší týmy (NHL)

Z předchozího obrázku je zřejmé, že Pittsburgh byl v obraně silnější než San Jose.

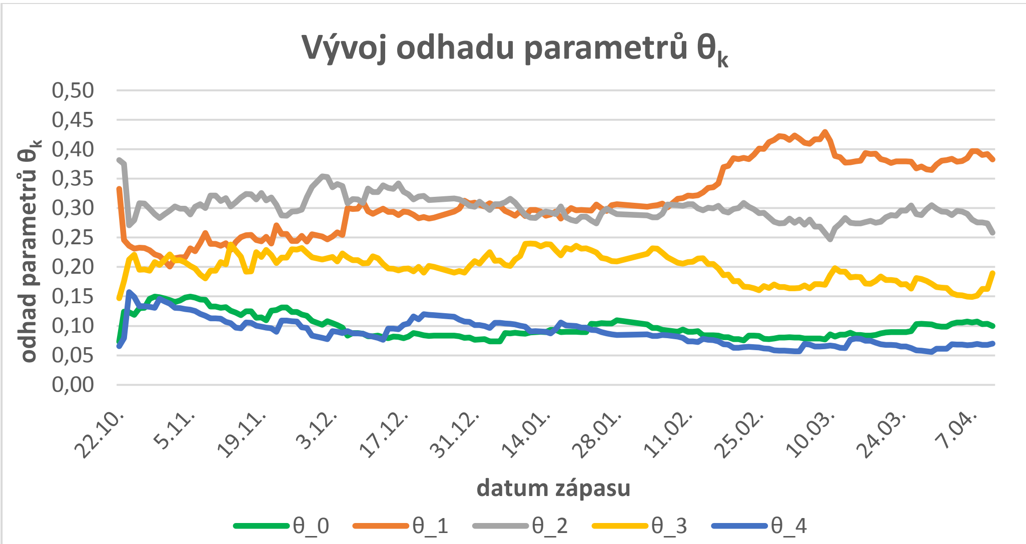
Odhady parametru výhody domácího prostředí γ a parametru μ se téměř nemění během jednotlivých kol. Průměrná hodnota odhadu parametru γ je 1,09. Průměrná hodnota odhadu parametru μ je 2,53, tj. průměrný počet gólů vstřelených hostujícím týmem. Průměrný počet gólů domácího týmu je 2,76. Průměrná hodnota odhadu parametru λ je -1,27.

Obr. 24 zobrazuje vývoj odhadu parametru p, tj. mixovacího parametru použitého v rovnici (6.9). Odhad tohoto parametru se mění během jednotlivých kol, ale ve všech případech je mezi 0,058 až 0,095. V tomto případě, vysoká hodnota parametru p značí vhodnost diagonálního rozšíření.



Obr. 24: Vývoj odhadu mixovacího parametru *p*

Poslední skupina parametrů je θ_k, pro k = 0,1,…,5, tyto parametry jsou použité pouze v diagonálně rozšířených modelech. Od prvního odhadovaného kola až po 91. kolo (tj. do 10.1.2016) měl odhad parametru θ_2 nejvyšší hodnotu ze všech odhadovaných parametrů θ_k, což značí, že parametr se použil ke zvýšení hodnoty pravděpodobnosti remízy se dvěma góly na každé straně. Téměř ve všech zbývajících kolech (92. až 177.) má nejvyšší hodnotu odhad parametru θ_1. Odhady parametrů $\theta_{0,\dots,4}$ jsou ukázány na Obr. 25. Při porovnání s dvourozměrným Poissonovo modelem (BP), odhady *p* a θ_k znamenají, že největší změny jsou v pravděpodobnostech remízových utkání s žádným, jedním, dvěma a třemi góly na každé straně. Menší změny jsou vytvořeny pro případ remízy se čtyřmi góly, naopak žádné změny nejsou vytvořeny pro remízy s pěti góly, proto odhad tohoto parametru není zahrnut v následujícím grafu.



Obr. 25: Vývoj odhadu parametrů $\theta_{0,\dots,4}$

Odhad výsledků zápasů je vytvořen na listu *BP-DI sezóna 2015-2016*.

7 Návrh vlastního modelu

Při zkoumání počtu gólů domácích a hostujících týmů v hokejových zápasech jsme zjistili, že domácí prostředí nemusí být pro všechny týmy výhodou, tzn., že všechny týmy nemusí mít stejnou hodnotu parametru γ vyjadřujícího výhodu domácího prostředí, což bude ukázáno v kapitole 7.1 pro českou ligu (resp. pro ostatní ligy v kapitole 7.2 a 7.3). Proto jsme navrhli úpravu modelů definovaných v kapitole 6, kde odhadujeme parametr γ pro každý tým i.

V návrhu vlastního modelu vycházíme z modelu původně užitého Maherem [1], kde jsou obě náhodné proměnné X_{ij} a Y_{ij} také nezávislé, ovšem pro λ_H platí upravený vzorec s přidaným parametrem γ_i, který je následující[10]:

$$X_{ij} \sim Po\left(\lambda_H = \mu \cdot \alpha_i \cdot \beta_j \cdot \gamma \cdot \gamma_i\right), \tag{7.1}$$

$$Y_{ij} \sim Po\left(\lambda_A = \mu \cdot \alpha_j \cdot \beta_i\right), \tag{7.2}$$

kde

α_i je síla útoku týmu (čím vyšší, tím lepší), $\alpha_i > 0$ pro všechna i,
β_i je naopak síla obrany týmu (čím menší, tím lepší), $\beta_i > 0$ pro všechna i,
μ je parametr, který vyjadřuje průměrný počet gólů hostujícího týmu, $\mu > 0$,
γ je míra výhody domácího utkání (globální parametr, dle článku [1] nebo [2]), $\gamma > 0$,
γ_i je míra výhody domácího utkání týmu, $\gamma_i > 0$ pro všechna i.

Takto upravený model budeme používat pro všechny zkoumané ligy.

7.1 Model pro Extraligu (CZE)

Tvrzení z předchozí části, že všechny týmy nemusí mít stejnou hodnotu parametru γ vyjadřujícího výhodu domácího prostředí potvrzuje následující tabulka s daty z české ligy.

Tým	Góly doma	Góly venku	Poměr gólů doma a venku
Hradec Králové	76	57	1,33
Karlovy Vary	66	42	1,57
Kometa Brno	76	74	1,03
Liberec	65	54	1,20
Litvínov	90	71	1,27
Mladá Boleslav	83	73	1,14
Olomouc	74	47	1,57
Pardubice	80	62	1,29
Plzeň	83	53	1,57
Slavia Praha	56	47	1,19
Sparta Praha	89	99	0,90
Třinec	93	86	1,08
Vítkovice	68	64	1,06
Zlín	78	62	1,26

Tab. 25: Počet gólů domácích/hostů pro týmy z Extraligy v sezóně 2014/2015

[10] Model použitý Maherem, ale upraveným pro hokejová data, kde provádíme také reparametrizaci přidáním parametru $\mu > 0$.

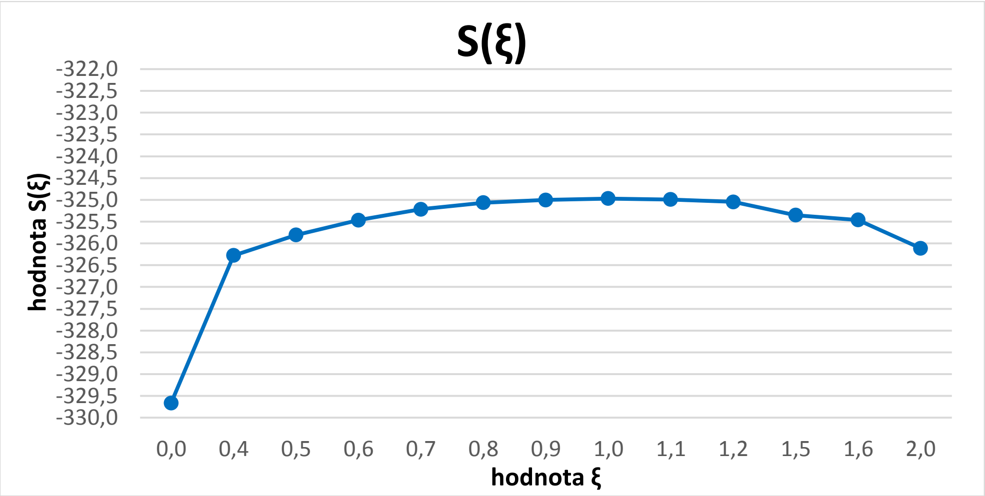
Poměr gólů doma a venku pro Spartu Praha (odpovídá parametru k^2 použitým v Maherově modelu a definovaným v rovnici (5.7)) vychází menší než 1, tzn., že v sezóně 2014/2015 dávala více gólů jako hostující tým, a proto pro tento tým nemusí být domácí prostředí výhodou.

Nyní, abychom určili odhady zápasů v sezóně 2015/2016, postupujeme stejným způsobem jako v předchozích kapitolách.

7.1.1 Parametr ξ pro sezónu 2014/2015

Stejně jako pro upravené Dixon-Colesovy modely z kapitoly 6, bylo potřeba nejprve vhodně zvolit parametr ξ pro sezónu 2015/2016 na základě odhadů z předchozí sezóny, detail lze nalézt v souboru *CZE_VM_volba_ksi.xlsm*. Odhady parametrů pro sezónu 2014/2015, které se použijí pro vyjádření funkce $S(\xi)$, jsou prováděny od 6. kola ze stejných důvodů, jako tomu bylo v původních modelech u české ligy.

Na Obr. 26 jsou zobrazeny hodnoty $S(\xi)$ různých hodnot ξ pro DP model.



Obr. 26: Hodnoty $S(\xi)$ proti ξ pro upravený DP model (CZE)

V následující tabulce je uveden přehled hodnot $S(\xi)$ jednotlivých modelů pro různé hodnoty parametru ξ. Mezi jednotlivými modely jsou malé rozdíly v hodnotách statistiky $S(\xi)$ v optimálním bodě ξ, a jsou označené žlutě. Pro nejvhodnější model DP je hodnota $S(\xi) = -324{,}97$, tj. vyšší než u ostatních modelů a v tabulce je zvýrazněna zeleně.

ξ	0,0	0,8	1,0	1,1	1,2	1,3	1,4	1,5	2,0
S(ξ) - DP	-329,663	-325,066	-324,970	-324,993	-325,046	-325,130	-325,216	-325,352	-326,112
S(ξ) - BP	-336,676	-330,060	-329,807	-329,741	-329,711	-329,682	-329,685	-329,694	-330,903
S(ξ) - DP-DI	-331,788	-326,171	-325,634	-326,076	-326,136	-326,188	-326,225	-326,507	-327,173
S(ξ) - BP-DI	-331,154	-325,828	-325,442	-325,614	-325,692	-325,763	-325,776	-326,059	-327,459

Tab. 26: Hodnoty $S(\xi)$ pro jednotlivé upravené modely (CZE)

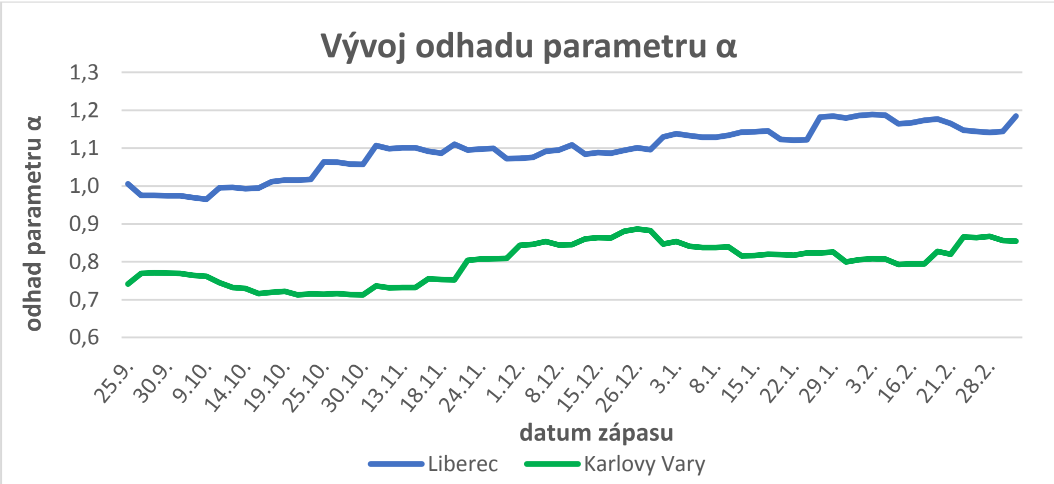
Pro model DP, DP-DI a BP-DI byla optimální hodnota ξ rovna 1, pouze u modelu BP tato hodnota vyšla rovna 1,3. Porovnáme-li všechny modely, tak zjistíme, že DP model poskytuje maximální hodnotu $S(\xi)$, proto jsme pro odhad výsledků sezóny 2015/2016 zvolili tento model s optimální hodnotou parametru ξ rovnou 1.

7.1.2 Odhad parametrů DP model 2015/2016

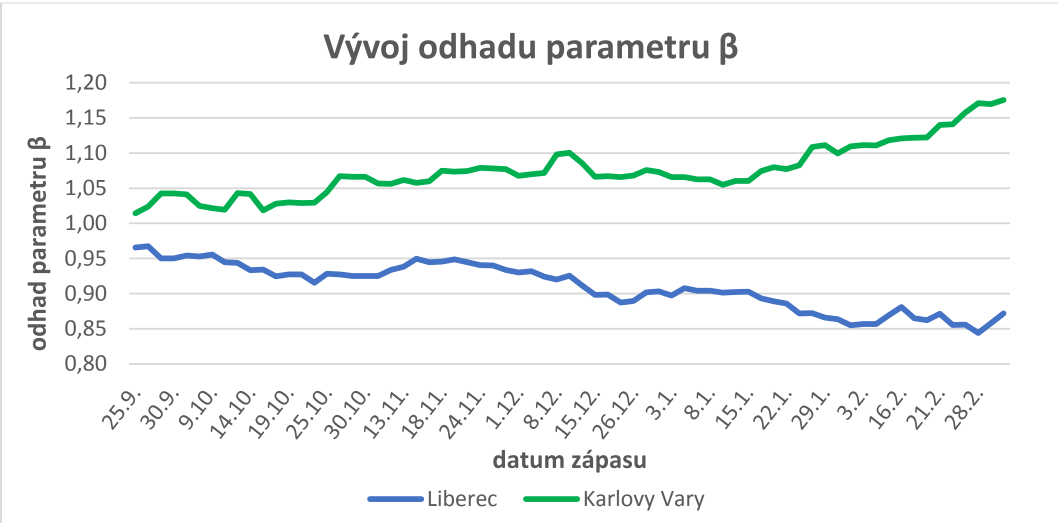
V této části budeme odhadovat pravděpodobnosti stavu zápasů (výhra domácích/remíza/výhra hostů) v sezóně 2015/2016 od 7. kola pomocí DP modelu zvoleného v předchozí části s parametrem $\xi = 1$. Zpracovány jsou v sešitu *CZE_VM_DP_15-16.xlsm* na listu *DP model*.

Odhad se provádí maximalizací věrohodnostní funkce definované v rovnici (6.12), stejně jako pro upravené Dixon-Colesovy modely. K maximalizaci se použije opět řešitel, který je v nastavení lehce pozměněn tak, aby pro odhad přidaného parametru γ_i platily stejné podmínky jako pro parametry α_i a β_i v původním modelu, tj. $\sum_i \gamma_i = N$, kde N je počet týmů, které vstupují do modelu.

Na následujících obrázcích je ukázán vývoj odhadu parametru útoku α_i, obrany β_i a výhody domácího prostředí γ_i během sezóny 2015/2016 pro nejlepší a nejhorší tým.



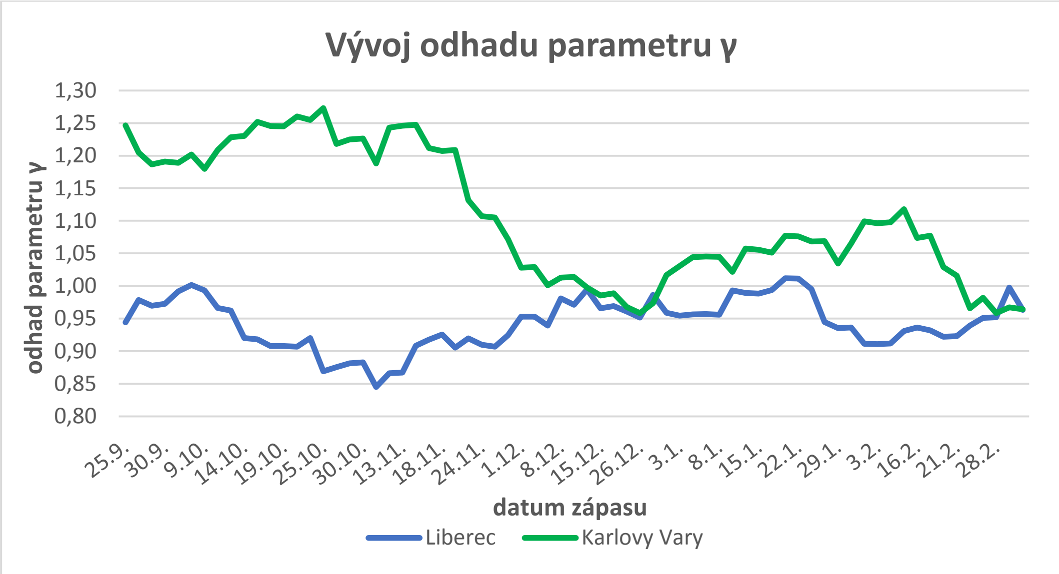
Obr. 27: Vývoj odhadu parametru útoku α – vlastní model (CZE)



Obr. 28: Vývoj odhadu parametru obrany β – vlastní model (CZE)

Z grafů jsou patrné rozdíly mezi oběma týmy, jak v útoku, tak i obraně. Odhady obou parametrů vycházejí podprůměrně pro Karlovy Vary a to odpovídá skutečnosti, že tento tým byl nejhorším týmem v sezóně.

V tomto upraveném modelu je parametr γ navíc odhadován pro každý tým zvlášť, jak již bylo uvedeno výše. Změny odhadu tohoto parametru v průběhu sezóny pro Liberec a Karlovy Vary jsou zobrazeny na Obr. 29.



Obr. 29: Vývoj odhadu parametru výhody domácího prostředí γ – vlastní model (CZE)

Z průběhu odhadu parametru $\gamma_{Karlovy\ Vary}$ můžeme vidět, že jeho hodnota klesla z úrovně kolem hodnoty 1,25 na úroveň blízkou hodnotě 0,95. To znamená, že výhoda domácího prostředí se v průběhu sezóny výrazně snížila. Naopak u týmu Liberce se úroveň odhadu parametru $\gamma_{Liberec}$ významně nezměnila, ale můžeme pozorovat, že v první polovině sezóny se výhoda domácího prostřední snížila, ale v druhé polovině sezóny se vrátila k původní úrovni jako na začátku odhadů tohoto parametru, a to přibližně k hodnotě 1. To znamená, že tento tým má vyrovnanější počet vstřelených gólů jak doma, tak venku.

7.1.3 Porovnání vlastního modelu s původním modelem

V této části využijeme upravený model pro českou ligu k porovnání s původním modelem. Určíme, který z těchto modelů je pro českou ligu vhodnější použít. Primárně se rozhodujeme podle hodnot funkce $S(\xi)$:

- Porovnáváme hodnoty funkce $S(\xi)$ pro oba modely v sezóně 2014/2015. Hodnota $S(\xi)$ pro vlastní model vychází nižší než v původním modelu, konkrétně o 5,8. To znamená, že původní model díky vyšší hodnotě $S(\xi)$ by měl lépe predikovat výsledky zápasů než navrhovaný model.

- Optimální hodnoty ξ se liší o 1,1. Tím, že je optimální hodnota ξ pro vlastní model nižší, přikládáme vyšší váhu starším zápasům, oproti původnímu modelu. To může mít vliv na odlišnosti mezi predikovanými a skutečnými výsledky zápasů, protože se výkon týmu může změnit mezi sezónami, ale i v průběhu sezóny.

- Rozdíly v odhadovaných výsledcích zápasů mohou být způsobené i volbou Dixon-Colesových modelů pro odhad parametrů sezóny 2015/2016. Na základě hodnoty $S(\xi)$ jsme zvolili DP model pro vlastní model, a u původního modelu jsme zvolili model BP.

Celkově vychází vhodnější použít původní model pro tuto ligu, i přesto, že jsme mohli vidět různý průběh parametru výhody domácího prostředí u jednotlivých týmů.

7.2 Model pro Ekstraligu (POL)

V této části odhadujeme výsledky zápasů sezóny 2015/2016 pro polskou ligu pomocí upraveného modelu definovaného na začátku kapitoly 7.

V tabulce níže je zobrazen poměr gólů doma a venku pro všechny týmy, které se zúčastnily Ekstraligy v sezóně 2014/2015. Je zde vidět, že více gólů vstřelily týmy v domácích zápasech, tj. pro všechny týmy z této ligy by mělo být domácí prostředí výhodou, avšak pro každý tým bude tato výhoda různě významná.

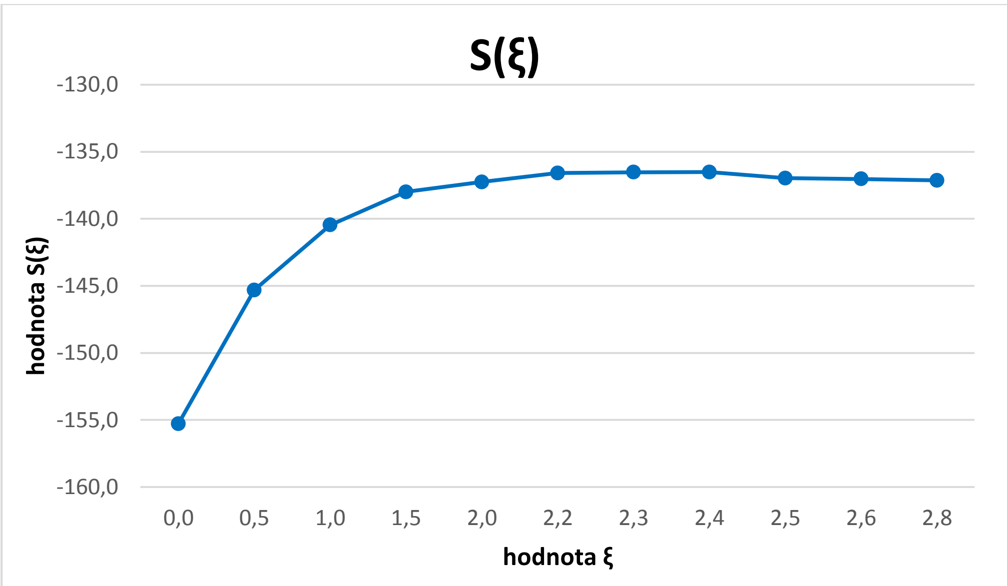
Tým	Góly doma	Góly venku	Poměr gólů doma a venku
Bytom	66	64	1,03
Janów	86	81	1,06
Jastrzebie JKH GKS	105	80	1,31
Katowice	51	39	1,31
Kraków	113	89	1,27
Orlik Opole	75	71	1,06
Podhale Nowy Targ	89	72	1,24
Sanok	120	100	1,20
Tychy	129	79	1,63
Unia Oświęcim	86	67	1,28

Tab. 27: Počet gólů domácích/hostů pro týmy z Ekstraligy pro sezónu 2014/2015

7.2.1 Parametr ξ pro sezónu 2014/2015

Znovu začneme tím, že určíme optimální hodnotu parametru ξ na sezóně 2014/2015 a tuto hodnotu použijeme pro odhadovanou následující sezónu.

Na následujícím obrázku je zobrazen graf $S(\xi)$ s různými hodnotami ξ pro vybraný DP-DI model.



Obr. 30: Hodnoty $S(\xi)$ proti ξ pro upravený DP-DI model (POL)

V následujících tabulkách je uveden přehled hodnot $S(\xi)$ jednotlivých modelů pro různé hodnoty parametru ξ. Mezi jednotlivými modely jsou malé rozdíly v hodnotách statistiky $S(\xi)$ v optimálním bodě ξ, a jsou označené žlutě. Pro model DP-DI je hodnota $S(\xi) = -136{,}52$, tj. vyšší než u ostatních modelů a v tabulce je zvýrazněna zeleně.

ξ	0,0	0,5	1,0	1,5	2,0	2,3	2,4	2,5	2,7
S(ξ) - BP	-155,251	-145,340	-140,430	-138,204	-137,068	-136,566	-136,528	-136,559	-136,634
S(ξ) - DP-DI	-155,287	-145,319	-140,465	-137,992	-137,256	-136,526	-136,517	-136,973	-137,089
S(ξ) - BP-DI	-155,088	-145,294	-140,493	-138,156	-137,303	-137,085	-136,873	-136,919	-137,090

Tab. 28: Hodnoty $S(\xi)$ pro jednotlivé upravené modely bez DP modelu (POL)

ξ	0,0	1,0	1,2	1,4	1,5	1,7	2,0	2,3	2,4	2,5
S(ξ) - DP	-156,052	-148,981	-148,307	-147,943	-148,121	-148,461	-149,535	-149,925	-150,605	-150,675

Tab. 29: Hodnoty $S(\xi)$ pro upravený DP model (POL)

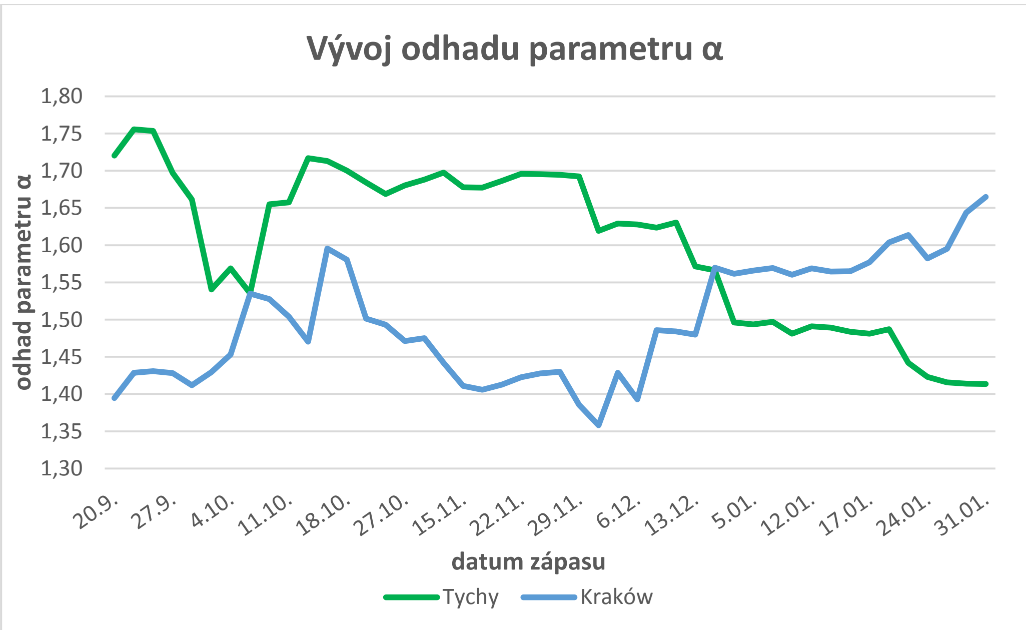
Na základě funkce $S(\xi)$ byly porovnány všechny čtyři upravené modely a maximální hodnotu této funkce poskytoval model DP-DI s optimální hodnotou $\xi = 2{,}4$. Pro model BP a BP-DI byla optimální hodnota ξ také rovna 2,4, pouze u modelu DP tato hodnota vyšla rovna 1,4. Proto model DP-DI s touto hodnotou parametru ξ zvolíme pro odhadovanou následující sezónu.

7.2.2 Odhad parametrů DP-DI model 2015/2016

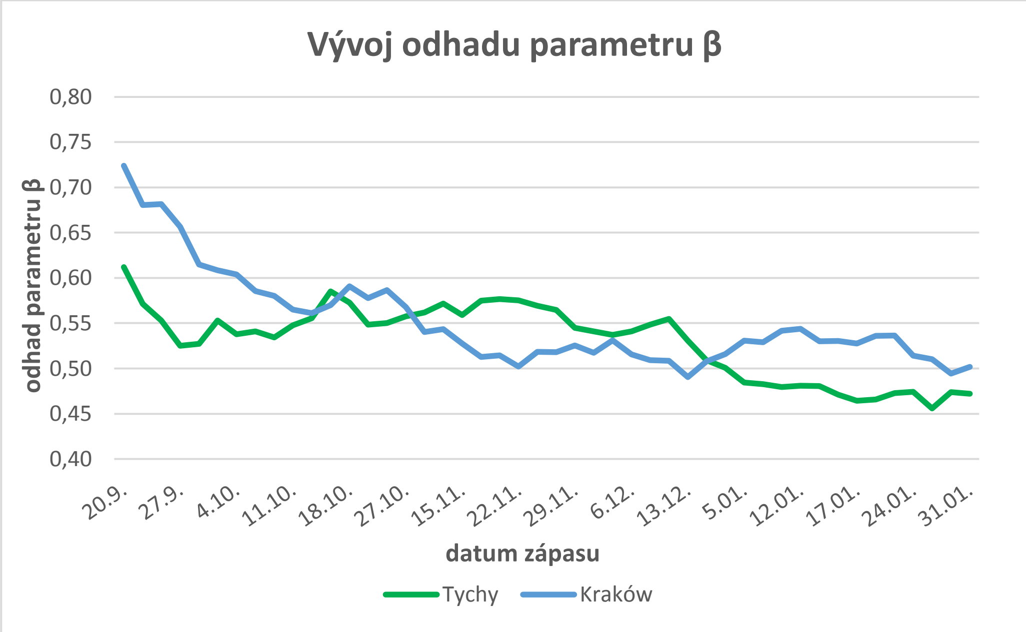
V této části budeme odhadovat pravděpodobnosti výsledků zápasů v sezóně 2015/2016 od 8. kola pomocí zvoleného DP-DI modelu v předchozí části. Zpracování je provedeno v sešitu *POL_VM_DP_15-16.xlsm* na listu *DP-DI model*.

Odhad se provádí maximalizací věrohodnostní funkce definované v rovnici (6.12) pomocí řešitele, stejně jako u české ligy.

Na následujících obrázcích je ukázán vývoj odhadu parametru útoku α_i, obrany β_i a výhody domácího prostředí γ_i během sezóny 2015/2016 pro dva nejlepší týmy.

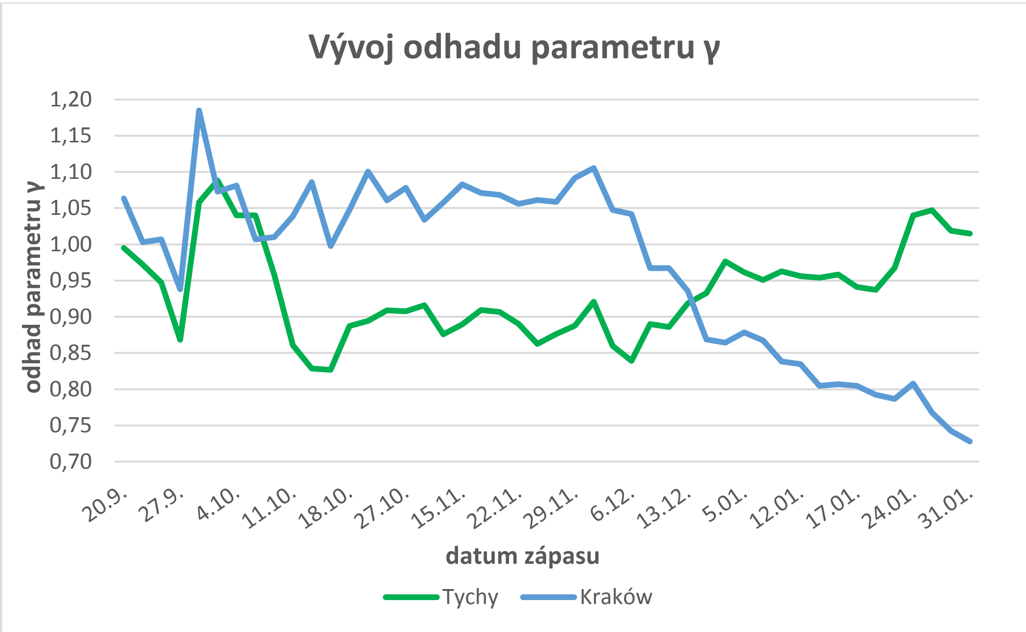


Obr. 31: Vývoj odhadu parametru útoku α – vlastní model (POL)



Obr. 32: Vývoj odhadu parametru obrany β – vlastní model (POL)

Z grafů jsou vidět nepatrné rozdíly mezi oběma týmy, jak v útoku, tak i obraně. Ve dvou třetinách sezóny byl Tychy silnější v útoku oproti týmu Kraków, ale v poslední části se vývoj odhadu parametru síly útoku u obou týmů změnil. Podobný průběh měl i parametr obrany pro oba týmy.



Obr. 33: Vývoj odhadu parametru výhody domácího prostředí γ – vlastní model (POL)

Úroveň hodnoty odhadu parametru γ_{Tychy} se v průběhu sezóny významně nezměnila, ale můžeme pozorovat, že v poslední části sezóny se výhoda domácího prostředí začala zvyšovat, až se vrátila k původní úrovni jako na začátku odhadů tohoto parametru, a to přibližně k hodnotě 1. To znamená, že tento tým má vyrovnanější počet vstřelených gólů jak doma, tak venku. Výhoda domácího prostředí pro tým Kraków se postupně v druhé polovině sezóny snižovala, až k hodnotě 0,7.

Můžeme vidět, že se pro oba týmy úroveň odhadu parametru domácího prostředí v druhé polovině sezóny mění, a tak je skutečně vhodné uvažovat odhad parametru γ pro každý tým zvlášť.

7.2.3 Porovnání vlastního modelu s původním modelem

V této části porovnáváme upravený model s původním modelem pro polskou ligu. Zvolíme, který z těchto modelů je pro tuto ligu vhodnější použít. Rozhodujeme se podle hodnot funkce $S(\xi)$, stejně jako u české ligy.

- Hodnota $S(\xi)$ pro vlastní model vychází nižší než v původním modelu, konkrétně o 2,97.
- Optimální hodnoty ξ se liší pouze o 0,1, tj. váhy jednotlivých zápasů se téměř neliší.
- Na základě hodnoty $S(\xi)$ jsme zvolili DP-DI model pro vlastní model, a u původního modelu jsme zvolili model BP. Rozdíly v odhadovaných výsledcích zápasů mohou být způsobené odlišným typem modelu.

Vhodnější je použít původní model pro tuto ligu, i přesto, že průběh odhadu parametru výhody domácího prostředí u jednotlivých týmů byl různý.

7.3 Model pro NHL

Stejně jako u předchozích lig odhadujeme pravděpodobnosti stavu zápasů pro sezónu 2015/2016, ale tentokrát od 16. kola, dle zpracování této ligy v předchozích kapitolách.

V tabulce vstřelených gólů doma a venku pro jednotlivé týmy můžeme vidět, že četnost týmů, které vstřelily více gólů venku, je oproti předchozím ligám významnější a tvoří zhruba 25 % z celkového počtu týmů v lize. Proto je určitě vhodné pro tuto ligu rozlišovat výhodu domácího prostředí pro jednotlivé týmy zvlášť.

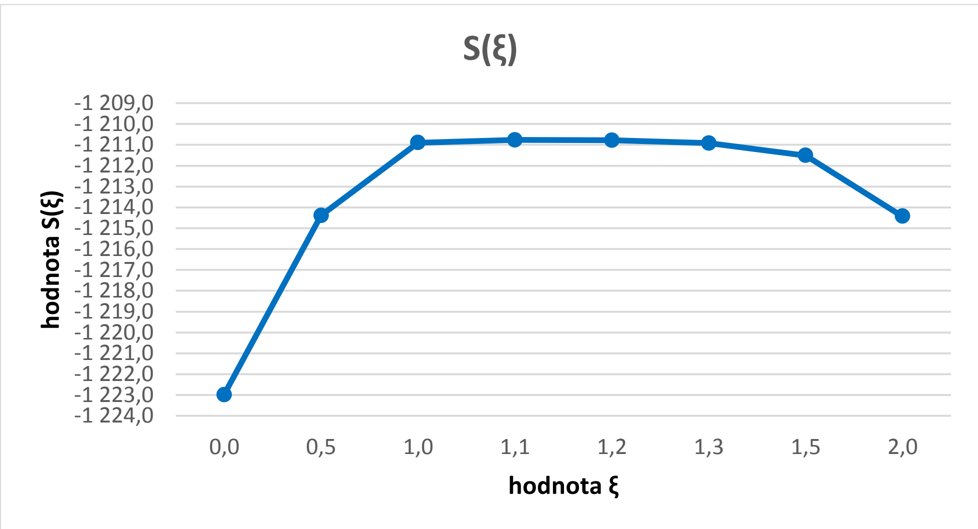
Tým	Góly doma	Góly venku	Poměr gólů doma a venku
Anaheim	107	113	0,95
Atlanta	122	97	1,26
Boston	104	96	1,08
Buffalo	78	74	1,05
Calgary	120	108	1,11
Carolina	89	93	0,96
Colorado	113	94	1,20
Columbus	109	113	0,96
Dallas	127	124	1,02
Detroit	117	107	1,09
Edmonton	93	98	0,95
Florida	105	92	1,14
Chicago	110	107	1,03
Los Angeles	113	104	1,09
Minnesota	124	99	1,25
Montreal	115	89	1,29
Nashville	112	106	1,06
New Jersey	86	89	0,97
NY Islanders	130	109	1,19
NY Rangers	119	123	0,97
Ottawa	130	95	1,37
Philadelphia	114	92	1,24
Phoenix	73	87	0,84
Pittsburgh	114	97	1,18
San Jose	112	110	1,02
St. Louis	120	114	1,05
Tampa Bay	137	118	1,16
Toronto	123	77	1,60
Vancouver	111	119	0,93
Washington	118	114	1,04

Tab. 30: Počet gólů domácích/hostů pro týmy z NHL v sezóně 2014/2015

7.3.1 Parametr ξ pro sezónu 2014/2015

Znovu začneme určením optimální hodnoty ξ na sezóně 2014/2015 a hodnotu potom použijeme pro odhadovanou následující sezónu.

Na obrázku níže jsou zobrazeny hodnoty $S(\xi)$ různých hodnot ξ pro BP-DI model.



Obr. 34: Hodnoty $S(\xi)$ proti ξ pro upravený BP-DI model (NHL)

V následujících tabulkách je uveden přehled hodnot $S(\xi)$ jednotlivých modelů pro různé hodnoty parametru ξ. Optimální hodnota ξ pro jednotlivé modely je zvýrazněná barevně.

ξ	0,0	0,2	0,4	0,5	0,8	1,0	1,3	1,5	2,0
S(ξ) - DP	-1 239,464	-1 235,002	-1 231,697	-1 230,595	-1 228,875	-1 228,695	-1 229,237	-1 229,959	-1 232,571

Tab. 31: Hodnoty $S(\xi)$ pro upravený DP model (NHL)

ξ	0,0	0,1	0,2	0,3	0,4	0,6	1,0	1,5	2,0
S(ξ) - BP	-1 279,292	-1 281,349	-1 280,707	-1 283,036	-1 283,504	-1 292,461	-1 316,136	-1 339,625	-1 357,368

Tab. 32: Hodnoty $S(\xi)$ pro upravený BP model (NHL)

ξ	0,0	0,1	0,5	1,0	1,1	1,2	1,3	1,5	2,0
S(ξ) - DP-DI	-1 253,547	-1 253,362	-1 251,975	-1 250,504	-1 250,794	-1 251,142	-1 251,706	-1 253,406	-1 259,714

Tab. 33: Hodnoty $S(\xi)$ pro upravený DP-DI model (NHL)

ξ	0,0	0,5	1,0	1,1	1,2	1,3	1,5	2,0
S(ξ) - BP-DI	-1 222,986	-1 214,394	-1 210,900	-1 210,768	-1 210,783	-1 210,921	-1 211,514	-1 214,430

Tab. 34: Hodnoty $S(\xi)$ pro upravený BP-DI model (NHL)

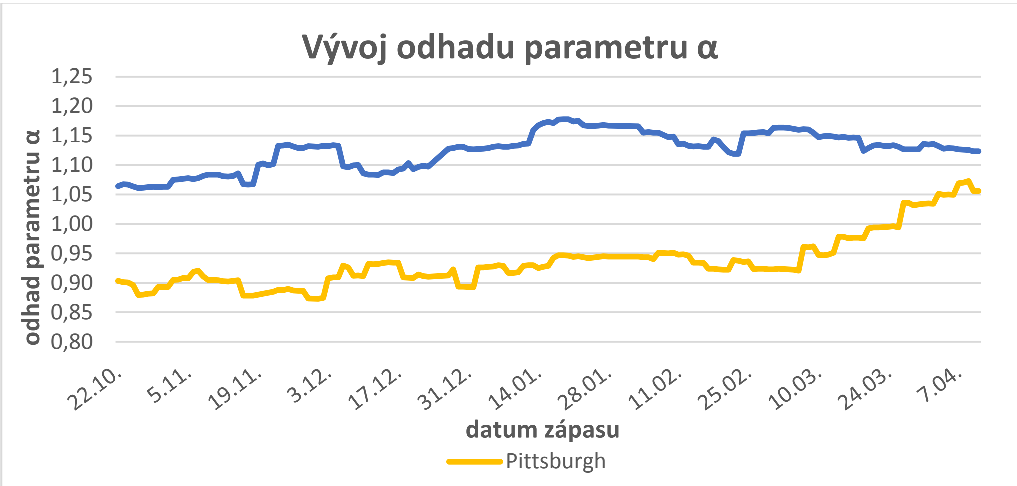
Při porovnání všech modelů jsme zjistili, že BP-DI model poskytuje maximální hodnotu $S(\xi) = -1\,210{,}77$, která je vyšší než u ostatních modelů a proto je zvýrazněna zeleně. Pro odhadovanou sezónu 2015/2016 jsme zvolili tento model s optimální hodnotou parametru ξ rovnou 1,1.

7.3.2 Odhad parametrů BP-DI model 2015/2016

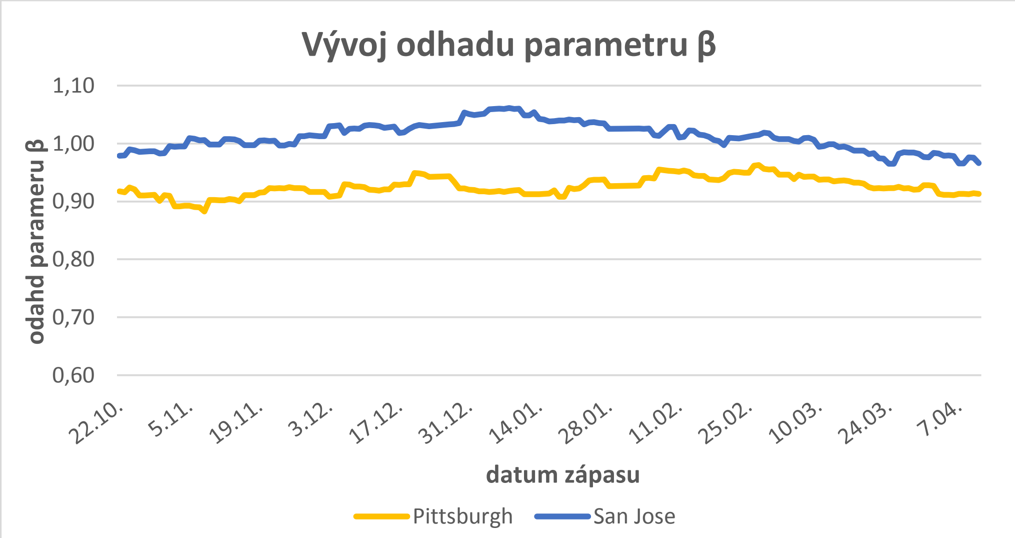
V této části budeme odhadovat pravděpodobnosti výsledků zápasů v sezóně 2015/2016 od 16. kola pomocí BP-DI modelu s parametrem $\xi = 1{,}1$ určeného v předchozí části. Zpracování lze nalézt v sešitu *NHL_VM_BP-DI_15-16.xlsm*.

Odhad se provádí opět maximalizací věrohodnostní funkce definované v rovnici (6.12) pomocí řešitele, stejně jako u ostatních lig.

Na následujících obrázcích je ukázán vývoj odhadu parametru útoku α_i, obrany β_i, výhody domácího prostředí γ_i, mixovacího parametru p a parametrů θ_k během sezóny 2015/2016 pro dva nejlepší týmy.



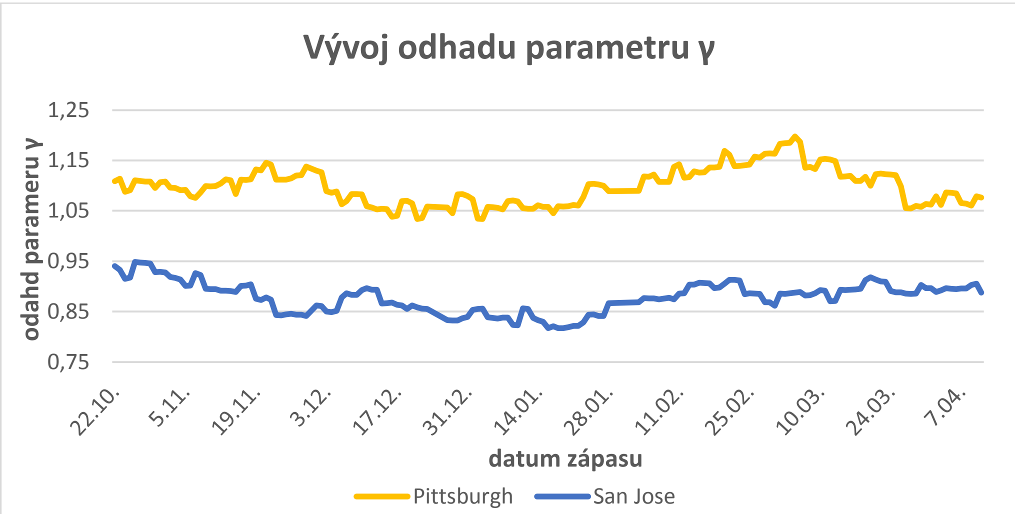
Obr. 35: Vývoj odhadu parametru útoku α – vlastní model (NHL)



Obr. 36: Vývoj odhadu parametru obrany β – vlastní model (NHL)

Vývoj odhadu parametru síly útoku a obrany pro San Jose je zhruba do poloviny sezóny rostoucí, nicméně od druhé poloviny sezóny je odhad parametru síly útoku stabilní až mírně klesající a zároveň odhad parametru síly obrany výrazně klesající, a vrací se na původní úroveň v začátku sezóny. Vývoj odhadu parametru síly útoku u týmu Pittsburgh je rostoucí, a to silněji ke konci sezóny, nicméně úroveň síly útoku je výrazně nižší než u týmu San Jose, ale ke konci sezóny se srovnává, téměř na úroveň týmu San Jose. Odhad parametru síly obrany je v tomto případě poměrně stabilní, ke konci sezóny mírně klesá, ale úrovně síly obrany obou týmů ke konci sezóny se také lehce srovnávají.

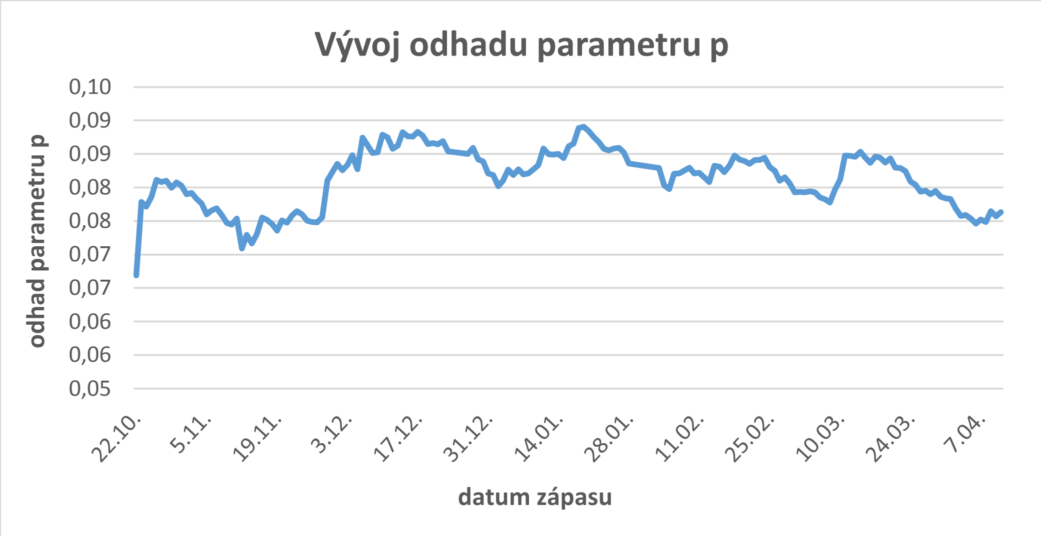
Změny odhadu parametru výhody domácího prostředí v průběhu sezóny pro Pittsburgh a San Jose jsou na Obr. 37.



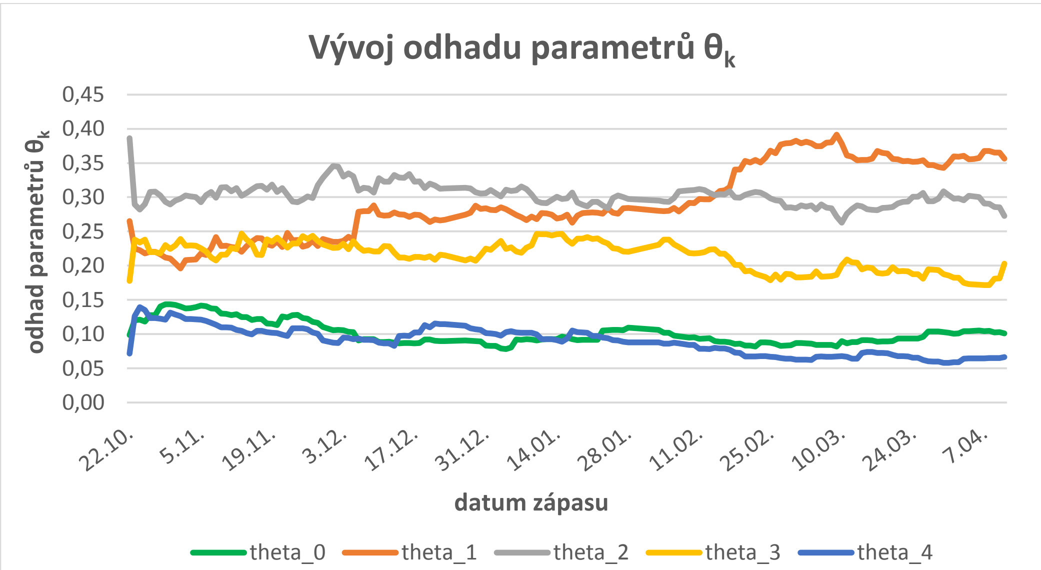
Obr. 37: Vývoj odhadu parametru výhody domácího prostředí γ – vlastní model (NHL)

Vývoj odhadu parametru výhody domácího prostředí pro tým San Jose ukazuje nižší úroveň než u týmu Pittsburgh a je poměrně stabilní po celou sezónu. Naopak tým Pittsburgh má odhad tohoto parametru více kolísající a v druhé polovině sezóny se zvyšoval.

Obr. 38 zobrazuje vývoj odhadu mixovacího parametru *p*. Odhad tohoto parametru se mění během sezóny a v průběhu sezóny je mezi 0,067 až 0,089. Vysoká hodnota parametru *p* značí vhodnost diagonálního rozšíření.



Obr. 38: Vývoj odhadu mixovacího parametru *p* – vlastní model (NHL)



Obr. 39: Vývoj odhadu parametrů $\theta_{0,\ldots,4}$ – vlastní model (NHL)

V předchozím grafu můžeme vidět, že od začátku sezóny až do 14.2.2016 měl odhad parametru θ_2 nejvyšší hodnotu ze všech odhadovaných parametrů θ_k, což značí, že parametr se použil ke zvýšení hodnoty pravděpodobnosti remízy se dvěma góly na každé straně. Ve zbývající části sezóny má nejvyšší hodnotu odhad parametru θ_1.

7.3.3 Porovnání vlastního modelu s původním modelem

V této části porovnáme upravený model pro NHL ligu s původním modelem. Podle hodnot funkce $S(\xi)$ rozhodneme, který z těchto modelů je pro tuto ligu vhodnější.

- Hodnota $S(\xi)$ pro vlastní model vychází nižší než v původním modelu, konkrétně o 6,05.
- Optimální hodnoty ξ se liší pouze o 0,4, kde vyšší hodnota tohoto parametru je pro původní model. Tím, že optimální hodnota ξ je pro původní model vyšší, přikládáme nižší váhu starším zápasům, oproti vlastnímu modelu. To může mít také vliv na odlišnosti mezi predikovanými a skutečnými výsledky zápasů.

Původní model vychází jako vhodnější varianta pro tuto ligu, i přesto, že odhad parametru γ byl odlišný u jednotlivých týmů.

8 Předvídací schopnost a sázecí strategie

V této kapitole použijeme odhady výsledků zápasů v sezóně 2015/2016 ze všech zkoumaných lig proti sázkovým kancelářím, aby bylo možné zhodnotit předvídací schopnost vybraného modelu, podobně jako je uvedeno v článku [3].

Strategií jakým způsobem sázet je celá řada, nicméně zde bude použita strategie zvaná Flat betting. Jedná se o velice jednoduchý systém, v němž používáme stále stejnou výši vkladu. V našem případě se jedná o 10 Kč pro každou ligu. Výhoda tohoto modelu spočívá v tom, že sázející nemusí řešit, kolik má vsadit na zápas, protože sází stále stejnou částku.

Pravděpodobnosti zápasů jsou odhadnuty užitím vybraného modelu, a podle očekávané hodnoty návratnosti (určené jako součin odhadů pravděpodobností a daných kurzů) zvolíme, zda na zápas sázet či nikoliv[11]. Nutným a logickým požadavkem je, aby se sázelo jen v případech, kdy očekávaná hodnota návratnosti převyšuje hodnotu 1. Proto zkoumáme hodnoty vyšší než 1, tj. sázíme pouze, když při zápase *m* očekávaná hodnota návratnosti je minimálně rovna L, kde $L > 1$. Pravidlo je možné zapsat takto [3]:

$$p_m^R \cdot o_m^R \geq L, R \in \{H, D, A\}, \tag{8.1}$$

kde p_m^R je odhad pravděpodobnosti předpokládaný modelem pro jednotlivé varianty R (R nabývá hodnoty H pro vítězství domácího týmu, D pro remízu a A pro vítězství hostujícího týmu) a o_m^R jsou kurzy pro danou variantu výsledku.

Pro testování se použily kurzy z [A], které jsou získané jako průměr z pěti vybraných sázkových kanceláří, jak bylo uvedeno v kapitole 2.

Podmínka definovaná v rovnici (8.1) může být splněna pro dva ze tří možných výsledků utkání. V případě, že by tato situace nastala, tak vsadíme pouze na variantu s nejvyšší očekávanou ziskovostí.

8.1 Triviální strategie

Pro každou ligu budeme modelovat jednoduché strategie sázení, které nazýváme triviální, z toho důvodu, že k nim nepotřebujeme žádné složité matematické výpočty.

Tyto strategie využijeme k tomu, abychom ukázali jejich ziskovost a především zhodnotili definované modely použité pro sázení.

Pracujeme s následujícími strategiemi sázení[12]:

- sázení na 100 % zápasů,
- náhodné sázení na 80 % zápasů,
- náhodné sázení na 60 % zápasů,
- sázení na výhru domácího týmu,
- sázení na remízu,

[11] Tato podmínka je převzata z článku [3].

[12] U náhodného sázení uvažujeme, že pravděpodobnost jednotlivých možností výsledku zápasu (výhra domácí, remíza a výhra hostů) je stejná.

- sázení na výhru hostujícího týmu,
- sázení na min. kurz,
- sázení na max. kurz.

Všechny uvedené triviální strategie budou v další části porovnány s vybraným modelem pro danou ligu.

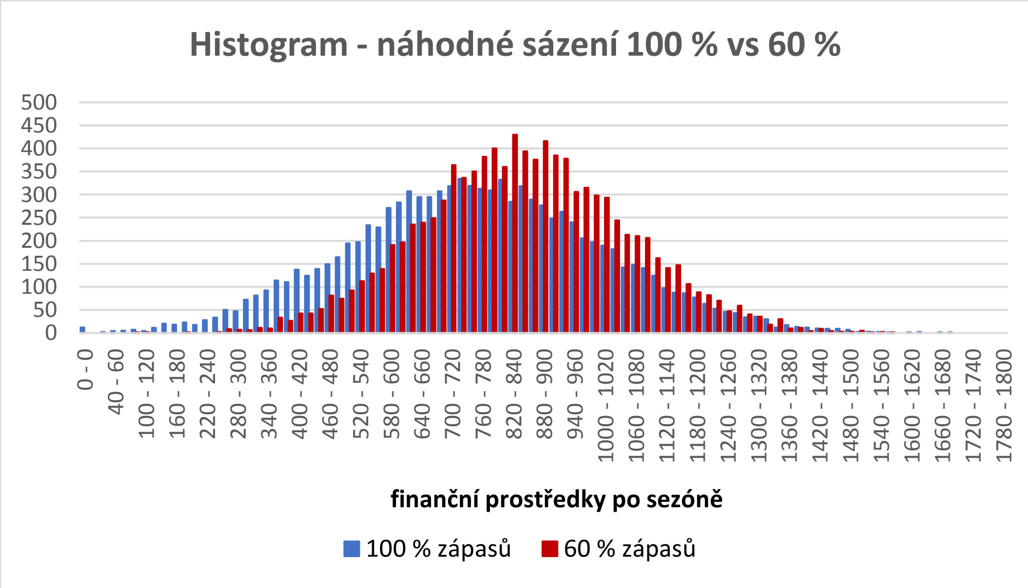
Při zpracování náhodného sázení jsme postupovali v následujících krocích (ukázka sázení na 80 %, resp. 60 % zápasů):

- vygenerovali jsme si v Microsoft Excel funkcí *randbetween(0;2)* náhodná celá čísla od 0 do 2 vyjadřující, na kterou ze tří možností výsledku vsadíme (1 = sázka na výhru domácích, 0 = sázka na remízu, 2 = sázka na výhru hostů),
- abychom nesázeli na všechny zápasy, přidali jsme další podmínku. Pomocí dalšího generování náhodných čísel (tentokrát funkcí *náhčíslo()*, která generuje čísla v intervalu [0;1] z oboru hodnot $\mathbb{R}$) určujeme, že v případě vygenerování čísla menšího než 0,2 (resp. 0,4), nevsadíme na tuto variantu, v opačném případě vsadíme,
- ke stavu dostupných finančních prostředků z předchozího zápasu přičteme hodnoty čistého zisku nebo ztráty po každém zápasu, abychom získali stav dostupných finančních prostředků na konci sezóny,

Náhodné generování, popsané v postupu výše, provádíme přes makro ve Visual Basicu, které pouze aktualizuje hodnoty generovaných náhodných čísel (klávesa F9) a uchovává dostupné finanční prostředky po celé sezóně. Tento postup se opakuje v cyklu 10 000 krát, abychom měli dostatečný počet simulací. Makro se spouští přes vytvořené tlačítko „Aktualizace“, ukázka viz *Příloha 4*.

Na závěr z takto uložených dostupných finančních prostředků po sezóně jsme si vytvořili histogram.

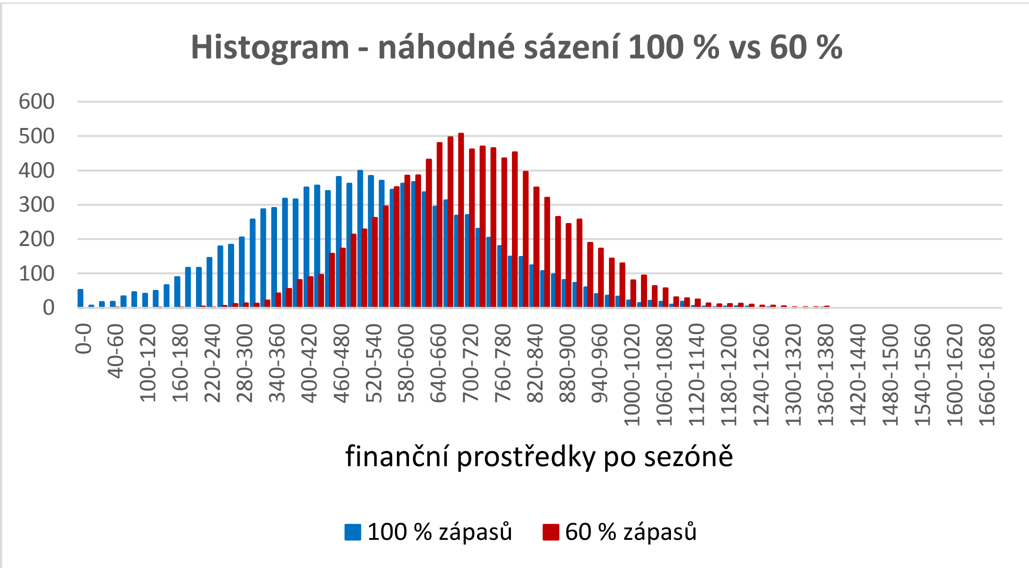
Porovnání náhodného sázení na 100 % a 60 % zápasů hraných v Extralize je zobrazeno na následujícím histogramu.



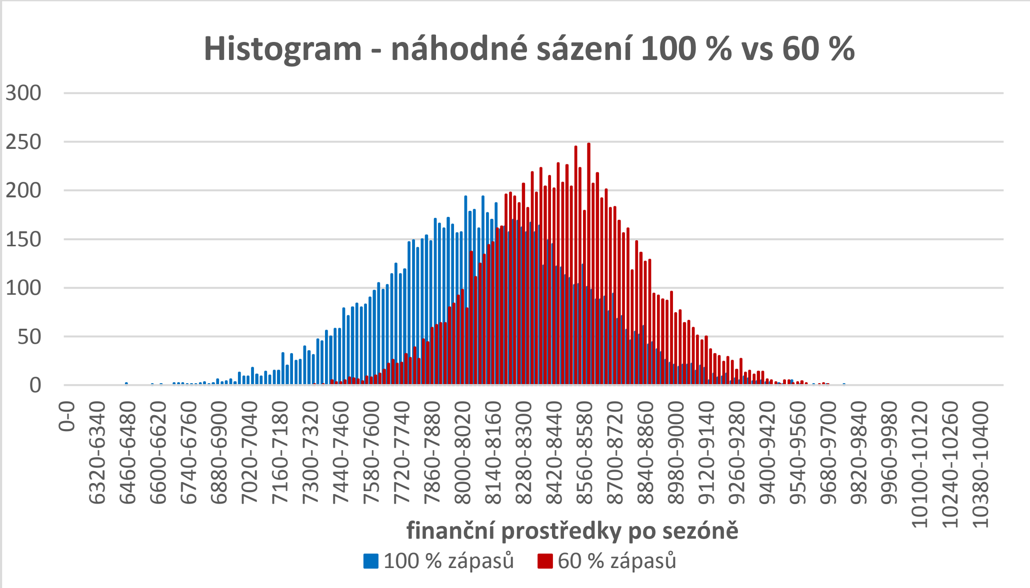
Obr. 40: Porovnání náhodného sázení na 100 % a 60 % zápasů – Extraliga (CZE)

V histogramu je vidět, že při sázení na 100 % zápasů nastala situace, kdy sázející zbankrotoval, a to konkrétně dvanáctkrát z 10 000 náhodných generování. Při sázení na 60 % zápasů již tato situace ani jednou nenastala. U varianty sázení na 80 % zápasů došlo k bankrotu ve 4 případech z 10 000. Je tedy vidět, že při vyšším procentu zápasů, na které sázíme, se snižuje střední hodnota finančních prostředků po sezóně. Pro tuto ligu jsme uvažovali počáteční finanční prostředky ve výši 1 000 Kč, stejně tak pro polskou ligu, a pro NHL 9 000 Kč z důvodu vyššího počtu zápasů.

Níže uvádíme obdobně histogramy pro polskou ligu a NHL.



Obr. 41: Porovnání náhodného sázení na 100 % a 60 % zápasů – Ekstraliga (POL)



Obr. 42: Porovnání náhodného sázení na 100 % a 60 % zápasů – NHL

Ostatní strategie není třeba popisovat, jelikož z názvu již vyplývá, jakým způsobem budeme sázet. Zpracování těchto všech strategií pro českou ligu je provedené v sešitě *CZE_BP 15-16.xlsm* na listech *sázení* (obdobně i pro polskou ligu a NHL).

8.2 Sázení dle modelu

V kapitole 6.6 (resp. 6.7 a 6.8) byl na základě sezóny 2014/2015 vybrán nejvhodnější model pro data z české ligy (resp. polské ligy a NHL), a použit pro výpočet pravděpodobností výher domácího týmu, remízy a vítězství hostujícího mužstva v následující sezóně. V této kapitole se budeme snažit využít těchto modelů k sázení počátečních finančních prostředků na vybrané zápasy, a zjistit zda je zvolený model pro danou ligu ziskový (resp. ztrátový). Kompletní zpracování této kapitoly je v souboru *CZE_BP_15-16.xlsm* na listu *sázení dle modelu* (obdobně pro polskou ligu a NHL).

8.2.1 Extraliga (CZE)

Pro českou ligu byl zvolen BP model jako nejvhodnější a pomocí něj odhadujeme pravděpodobnosti výhry domácích, remízy nebo hostujících od 7. kola sezóny 2015/2016, jak bylo uvedeno v předchozích kapitolách.

Předpokládali jsme dostupné finanční prostředky na začátku sázení pouze ve výši 1 000 Kč, aby mohla nastat situace, že sázející během sezóny zbankrotuje. Bankrotem je myšlena situace, kdy sázkař přijde o všechny své počáteční finanční prostředky, tj. v průběhu sázení by se dostal na hodnotu dostupných finančních prostředků nižší než je 0 Kč. Pokud by tato situace nastala, pak je ve výpočtech nastavena ošetřující podmínka, aby sázkař už nemohl dál sázet. Zajímá nás nejen, zda sázkař v průběhu sezóny zbankrotuje, ale hlavně stav finančních prostředků na konci sezóny pokud v průběhu nenastane situace bankrotu.

Nejprve jsme zvolili parametr L blízký hodnotě 1 tak, aby splňoval podmínku definovanou v rovnici (8.1), např. $L = 1{,}01$, tzn., že bychom vsadili dle BP modelu na 82 % zápasů (tj. 271 zápasů) a potřebovali bychom 2 710 Kč, abychom v nejhorším případě skončili po sezóně s 0 Kč.

Při nastavení různých hodnot parametru $L > 1$ se mění počet zápasů, na které bychom vsadili a samozřejmě také čistý zisk. Proto jsme zkoumali dopad různých hodnot tohoto parametru na ziskovost (resp. ztrátovost) modelu a na počet vsazených a vítězných zápasů, což lze vidět v tabulce níže. Vidíme, že při vyšších hodnotách parametru L bychom sázeli na méně procent zápasů.

	L												
	1,01	**1,03**	**1,06**	**1,07**	**1,08**	**1,09**	**1,12**	**1,15**	**1,18**	**1,21**	**1,24**	**1,27**	**1,30**
počet sázek	271	221	165	148	126	109	74	54	40	31	21	14	9
celkem vsazeno	2 710	2 210	1 650	1 480	1 260	1 090	740	540	400	310	210	140	90
počet vítězných sázek	110	92	70	67	57	48	31	24	16	11	8	6	3
celkem čistý zisk/ztráta	-240,00	-126,70	-4,00	117,30	78,70	36,30	26,60	44,90	15,20	-8,00	18,00	35,40	14,20
relativní hodnota majetku	0,91	0,94	1,00	1,08	1,06	1,03	1,04	1,08	1,04	0,97	1,09	1,25	1,16
na kolik % zápasů by se vsadilo	82%	67%	50%	45%	38%	33%	22%	16%	12%	9%	6%	4%	3%

Tab. 35: Sázení dle BP modelu pro různé hodnoty L – Extraliga (CZE)

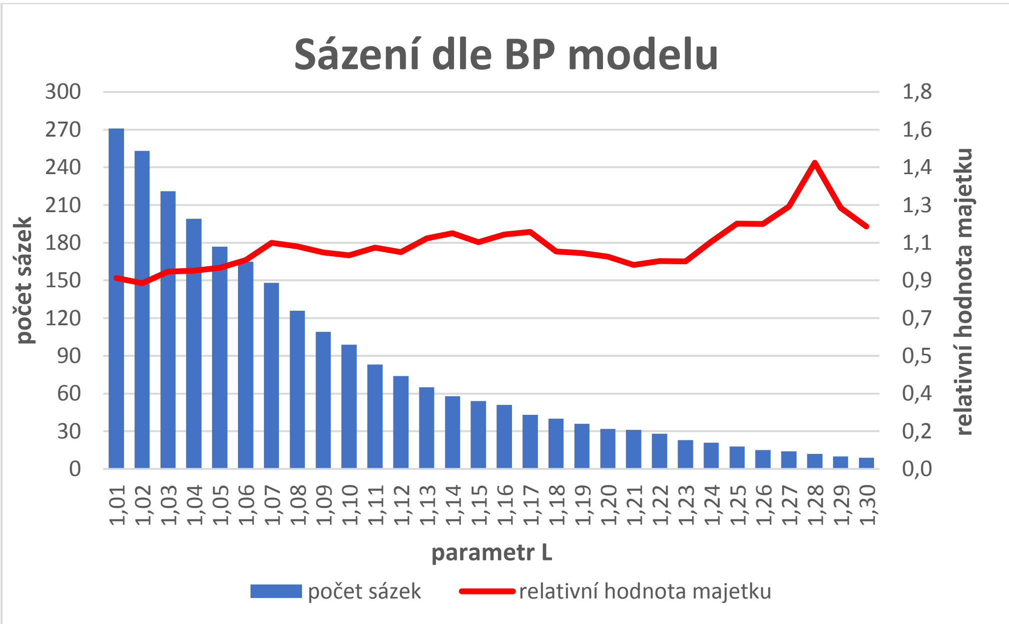
Nejvyššího celkového zisku model dosahuje při zvoleném parametru $L = 1{,}07$. To znamená, že na konci sezóny budeme mít k dispozici 1 117 Kč, tj. o 117 Kč více než je hodnota finančních prostředků na začátku sezóny. V tabulce je také uvedena relativní hodnota majetku, která vyjadřuje, kolikanásobek vsazených prostředků byl po konci sezóny, tj. 1,08 násobek pro $L = 1{,}07$. V průběhu sezóny nenastala situace bankrotu a tak můžeme říci, že při této hodnotě parametru L je model ziskový.

Pro znázornění je v následující tabulce uvedena část zápasů této ligy. Dle sloupců „Vsazeno/nevsazeno“ je vidět, na které zápasy dle modelu vsadíme, a dle sloupců „Výhra/Prohra“ zároveň i výši čistého zisku (případně ztráty) u vsazených zápasů, při zvoleném $L = 1{,}07$.

Datum	Kolo	Domácí	Hosté	Skóre domácí	Skóre hosté	Výsledek	Vsazeno/nevsazeno			Výhra/Prohra		
							1	X	2	1	X	2
27.9.2015	7	Olomouc	Chomutov	5	2	1	0	0	0	0,00	0,00	0,00
27.9.2015	7	Hradec Králové	Kometa Brno	3	1	1	0	0	0	0,00	0,00	0,00
27.9.2015	7	Liberec	Třinec	2	1	1	0	0	0	0,00	0,00	0,00
27.9.2015	7	Karlovy Vary	Mladá Boleslav	2	5	2	0	0	1	0,00	0,00	16,10
27.9.2015	7	Litvínov	Sparta Praha	2	2	0	1	0	0	-10,00	0,00	0,00
29.9.2015	8	Vítkovice	Plzeň	1	4	2	0	0	0	0,00	0,00	0,00
30.9.2015	9	Liberec	Sparta Praha	4	4	0	0	0	0	0,00	0,00	0,00
2.10.2015	10	Hradec Králové	Třinec	5	3	1	0	0	0	0,00	0,00	0,00
2.10.2015	10	Pardubice	Vítkovice	1	6	2	0	0	0	0,00	0,00	0,00
2.10.2015	10	Chomutov	Karlovy Vary	0	2	2	0	0	1	0,00	0,00	38,90
2.10.2015	10	Plzeň	Kometa Brno	4	3	1	1	0	0	12,10	0,00	0,00
2.10.2015	10	Litvínov	Mladá Boleslav	2	2	0	0	0	1	0,00	0,00	-10,00
4.10.2015	11	Kometa Brno	Sparta Praha	5	3	1	0	0	0	0,00	0,00	0,00
4.10.2015	11	Litvínov	Třinec	1	4	2	0	0	0	0,00	0,00	0,00
4.10.2015	11	Olomouc	Pardubice	2	1	1	0	0	0	0,00	0,00	0,00
4.10.2015	11	Hradec Králové	Plzeň	2	6	2	0	0	1	0,00	0,00	23,10
4.10.2015	11	Karlovy Vary	Zlín	1	2	2	0	0	1	0,00	0,00	18,30
4.10.2015	11	Liberec	Chomutov	2	2	0	0	0	0	0,00	0,00	0,00
9.10.2015	12	Sparta Praha	Plzeň	4	4	0	0	0	0	0,00	0,00	0,00
9.10.2015	12	Olomouc	Karlovy Vary	2	1	1	0	0	0	0,00	0,00	0,00
9.10.2015	12	Pardubice	Liberec	1	4	2	0	0	1	0,00	0,00	14,20
9.10.2015	12	Zlín	Vítkovice	3	5	2	0	0	0	0,00	0,00	0,00
9.10.2015	12	Mladá Boleslav	Hradec Králové	4	1	1	1	0	0	12,20	0,00	0,00
9.10.2015	12	Chomutov	Litvínov	1	0	1	0	0	1	0,00	0,00	-10,00
9.10.2015	12	Třinec	Kometa Brno	1	2	2	0	0	0	0,00	0,00	0,00
10.10.2015	13	Vítkovice	Karlovy Vary	6	1	1	0	0	0	0,00	0,00	0,00
11.10.2015	14	Kometa Brno	Chomutov	6	4	1	0	0	0	0,00	0,00	0,00
11.10.2015	14	Litvínov	Pardubice	1	2	2	0	0	0	0,00	0,00	0,00
11.10.2015	14	Hradec Králové	Zlín	2	1	1	0	0	1	0,00	0,00	-10,00
11.10.2015	14	Plzeň	Mladá Boleslav	3	1	1	0	0	1	0,00	0,00	-10,00
11.10.2015	14	Liberec	Olomouc	0	1	2	0	0	1	0,00	0,00	36,00
11.10.2015	14	Sparta Praha	Třinec	7	5	1	0	0	1	0,00	0,00	-10,00
14.10.2015	15	Sparta Praha	Karlovy Vary	1	1	0	0	0	0	0,00	0,00	0,00
14.10.2015	15	Třinec	Plzeň	5	1	1	0	0	1	0,00	0,00	-10,00
16.10.2015	16	Karlovy Vary	Litvínov	2	3	2	0	0	1	0,00	0,00	15,00
16.10.2015	16	Pardubice	Sparta Praha	2	4	2	0	0	1	0,00	0,00	11,60
16.10.2015	16	Chomutov	Hradec Králové	1	1	0	0	0	1	0,00	0,00	-10,00
16.10.2015	16	Zlín	Liberec	1	3	2	1	0	0	-10,00	0,00	0,00
16.10.2015	16	Mladá Boleslav	Kometa Brno	4	0	1	1	0	0	14,10	0,00	0,00
16.10.2015	16	Olomouc	Vítkovice	2	2	0	0	0	1	0,00	0,00	-10,00

Tab. 36: Výsledky sázení (zisk/ztráta) dle BP modelu ($L = 1{,}07$) - CZE

Relativní hodnota majetku pro kurzy ze zvoleného zdroje je znázorněna na Obr. 43. Rostoucí parametr L snižuje počet zápasů, na které bychom vsadili. Zde jsou znázorněné pouze výsledky, kdy bylo vsazeno alespoň třikrát za sezónu. Model je ziskový pro hodnoty parametru $L \in [1{,}07; 1{,}20] \cup [1{,}24; 1{,}30]$, což znamená, že pro hodnotu parametru $L = 1{,}07$ by bylo 67 vítězných sázek ze 148 vsazených zápasů (z 329 odhadovaných zápasů).



Obr. 43: Relativní hodnota majetku v porovnání s parametrem L – Extraliga (CZE)

Na průběhu křivky relativní hodnoty majetku můžeme sledovat mírně rostoucí vývoj, který je závislý na počtu zápasů, které skutečně dopadly tak, jak bylo odhadováno, v daném počtu zápasů, na které sázíme. Dle modelu se nám nejvíce vyplácí sázet na nižší počet zápasů (přibližně $L = 1{,}28$), při kterém máme nejvyšší zhodnocení vsazených finančních prostředků.

8.2.2 Ekstraliga (POL)

Pro polskou ligu jsme zvolili jako nejvhodnější model BP, pomocí kterého jsme vypočítali pravděpodobnosti jednotlivých variant výsledku zápasů od 8. kola v sezóně 2015/2016, které se použijí při sázení na tuto sezónu. Předpokládáme dostupné finanční prostředky na začátku sázení ve výši 1 000 Kč, aby mohla nastat situace, že sázející během sezóny zbankrotuje, stejně jako pro českou ligu.

Znovu jsme zkoumali dopad různých hodnot parametru $L > 1$ na ziskovost (resp. ztrátovost) modelu a na počet vsazených a vítězných zápasů, což lze vidět v Tab. 37. Při vyšších hodnotách parametru L bychom sázeli na méně procent zápasů.

	L												
	1,01	**1,03**	**1,04**	**1,06**	**1,07**	**1,08**	**1,09**	**1,13**	**1,15**	**1,18**	**1,21**	**1,27**	**1,33**
počet sázek	149	132	129	104	98	96	92	75	67	55	48	35	26
celkem vsazeno	1 490	1 320	1 290	1 040	980	960	920	750	670	550	480	350	260
počet vítězných sázek	71	60	59	45	41	39	38	31	26	20	17	11	8
celkem čistý zisk/ztráta	-0,40	28,70	42,70	-3,60	7,30	4,40	23,00	52,40	1,60	-58,10	-32,80	-18,80	14,50
relativní hodnota majetku	1,00	1,02	1,03	1,00	1,01	1,00	1,03	1,07	1,00	0,89	0,93	0,95	1,06
na kolik % zápasů by se vsadilo	69%	61%	60%	48%	46%	45%	43%	35%	31%	26%	22%	16%	12%

Tab. 37: Sázení dle BP modelu pro různé hodnoty L – Ekstraliga (POL)

Model je ziskový již pro hodnotu parametru $L = 1{,}03$. Pro danou hodnotu by v tomto případě bylo 60 vítězných sázek z celkově vsazených 132 zápasů (z 215 odhadovaných zápasů).

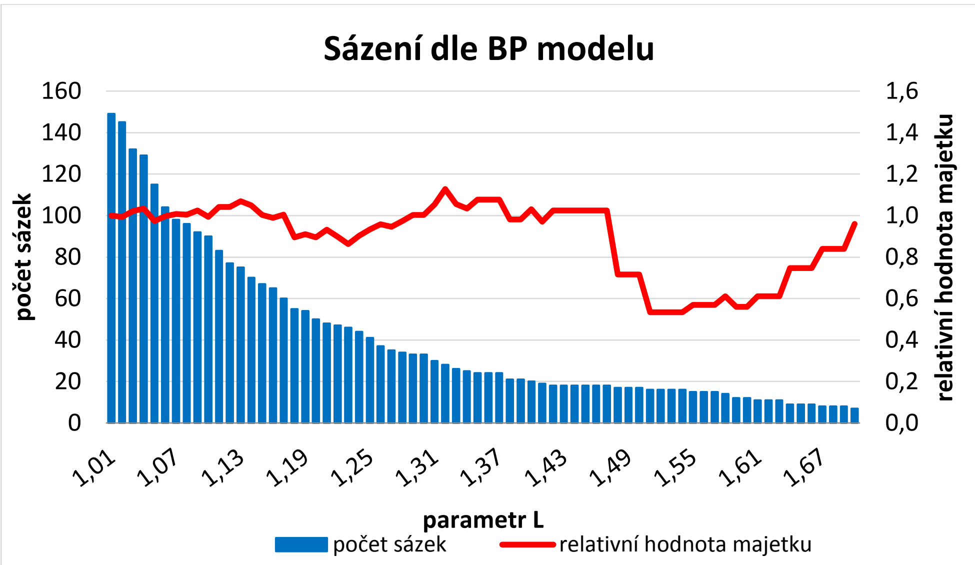
Nejvyššího celkového zisku model dosahuje při zvoleném parametru $L = 1{,}13$. To znamená, že na konci sezóny budeme mít k dispozici 1 052 Kč, tj. o 52 Kč více než je hodnota finančních prostředků na začátku sezóny. V průběhu sezóny nenastala situace, že by sázející zbankrotoval.

V tabulce níže je pro znázornění uvedena část zápasů této ligy, podobně jako v předchozí podkapitole pro českou Extraligu.

Datum	Kolo	Domácí	Hosté	Skóre domácí	Skóre hosté	Výsledek	Vsazeno/nevsazeno			Výhra/Prohra		
							1	X	2	1	X	2
22.9.2015	8	Tychy	Orlik Opole	4	1	1	0	0	0	0,00	0,00	0,00
25.9.2015	9	Sosnoviec SMS	Sanok	1	13	2	0	0	0	0,00	0,00	0,00
25.9.2015	9	Janów	Toruň	3	3	0	0	0	0	0,00	0,00	0,00
25.9.2015	9	Kraków	Podhale Nowy Targ	5	1	1	0	0	1	0,00	0,00	-10,00
25.9.2015	9	Orlik Opole	Tychy	1	3	2	0	0	0	0,00	0,00	0,00
25.9.2015	9	Unia Oświęcim	Jastrzebie JKH GKS	2	4	2	0	0	1	0,00	0,00	13,00
25.9.2015	9	Bytom	Zagłębie Sosnowiec	4	0	1	0	0	0	0,00	0,00	0,00
27.9.2015	10	Podhale Nowy Targ	Unia Oświęcim	5	0	1	0	0	0	0,00	0,00	0,00
27.9.2015	10	Jastrzebie JKH GKS	Orlik Opole	2	2	0	0	0	0	0,00	0,00	0,00
27.9.2015	10	Bytom	Toruň	8	3	1	0	0	1	0,00	0,00	-10,00
27.9.2015	10	Tychy	Janów	7	2	1	0	0	0	0,00	0,00	0,00
27.9.2015	10	Sanok	Kraków	1	4	2	1	0	0	-10,00	0,00	0,00
27.9.2015	10	Sosnoviec SMS	Zagłębie Sosnowiec	0	5	2	1	0	0	-10,00	0,00	0,00
29.9.2015	11	Zagłębie Sosnowiec	Toruň	5	2	1	0	0	1	0,00	0,00	-10,00
29.9.2015	11	Janów	Jastrzebie JKH GKS	3	6	2	0	0	0	0,00	0,00	0,00
29.9.2015	11	Orlik Opole	Podhale Nowy Targ	3	3	0	0	0	1	0,00	0,00	-10,00
29.9.2015	11	Kraków	Sosnoviec SMS	13	0	1	0	0	0	0,00	0,00	0,00
29.9.2015	11	Unia Oświęcim	Sanok	3	2	1	0	0	1	0,00	0,00	-10,00
29.9.2015	11	Bytom	Tychy	4	2	1	0	0	1	0,00	0,00	-10,00
2.10.2015	12	Sosnoviec SMS	Toruň	2	6	2	0	0	0	0,00	0,00	0,00
2.10.2015	12	Kraków	Unia Oświęcim	7	1	1	0	0	0	0,00	0,00	0,00
2.10.2015	12	Podhale Nowy Targ	Janów	12	1	1	0	0	0	0,00	0,00	0,00
2.10.2015	12	Jastrzebie JKH GKS	Bytom	1	0	1	1	0	0	8,30	0,00	0,00
2.10.2015	12	Sanok	Orlik Opole	2	4	2	1	0	0	-10,00	0,00	0,00
2.10.2015	12	Tychy	Zagłębie Sosnowiec	7	0	1	0	0	0	0,00	0,00	0,00
4.10.2015	13	Toruň	Tychy	2	5	2	0	0	0	0,00	0,00	0,00
4.10.2015	13	Orlik Opole	Kraków	1	5	2	1	0	0	-10,00	0,00	0,00
4.10.2015	13	Zagłębie Sosnowiec	Jastrzebie JKH GKS	1	4	2	0	0	1	0,00	0,00	4,20
4.10.2015	13	Unia Oświęcim	Sosnoviec SMS	10	1	1	0	0	0	0,00	0,00	0,00
4.10.2015	13	Bytom	Podhale Nowy Targ	4	2	1	0	0	1	0,00	0,00	-10,00
4.10.2015	13	Janów	Sanok	4	9	2	0	0	0	0,00	0,00	0,00
6.10.2015	14	Janów	Tychy	1	11	2	0	0	0	0,00	0,00	0,00
9.10.2015	15	Janów	Kraków	1	7	2	0	0	0	0,00	0,00	0,00
9.10.2015	15	Zagłębie Sosnowiec	Podhale Nowy Targ	2	3	2	0	0	1	0,00	0,00	3,20
9.10.2015	15	Orlik Opole	Unia Oświęcim	1	3	2	0	0	0	0,00	0,00	0,00
9.10.2015	15	Toruň	Jastrzebie JKH GKS	3	5	2	0	0	0	0,00	0,00	0,00
9.10.2015	15	Tychy	Sosnoviec SMS	8	2	1	0	0	0	0,00	0,00	0,00
9.10.2015	15	Bytom	Sanok	5	2	1	0	0	1	0,00	0,00	-10,00
11.10.2015	16	Kraków	Bytom	8	2	1	1	0	0	2,90	0,00	0,00
11.10.2015	16	Podhale Nowy Targ	Toruň	10	3	1	0	0	0	0,00	0,00	0,00
11.10.2015	16	Jastrzebie JKH GKS	Tychy	3	6	2	0	0	0	0,00	0,00	0,00
11.10.2015	16	Orlik Opole	Sosnoviec SMS	7	3	1	0	0	0	0,00	0,00	0,00
11.10.2015	16	Sanok	Zagłębie Sosnowiec	3	0	1	1	0	0	3,30	0,00	0,00
11.10.2015	16	Unia Oświęcim	Janów	3	1	1	0	0	1	0,00	0,00	-10,00
13.10.2015	17	Janów	Orlik Opole	2	6	2	0	0	0	0,00	0,00	0,00
13.10.2015	17	Zagłębie Sosnowiec	Kraków	2	9	2	0	0	0	0,00	0,00	0,00
13.10.2015	17	Toruň	Sanok	1	3	2	0	0	0	0,00	0,00	0,00
13.10.2015	17	Tychy	Podhale Nowy Targ	3	4	2	0	0	1	0,00	0,00	57,20
13.10.2015	17	Jastrzebie JKH GKS	Sosnoviec SMS	7	0	1	0	0	0	0,00	0,00	0,00
13.10.2015	17	Bytom	Unia Oświęcim	2	5	2	0	0	1	0,00	0,00	26,30

Tab. 38: Výsledky sázení (zisk/ztráta) dle BP modelu ($L = 1{,}13$) - POL

Na Obr. 44 je znázorněna relativní hodnota majetku. Pro nejvíce ziskovou variantu, při $L = 1{,}13$, by bylo 31 vítězných sázek ze 75 vsazených zápasů. V grafu jsou znázorněny pouze výsledky, když bychom vsadili alespoň na sedm zápasů za sezónu.



Obr. 44: Relativní hodnota majetku v porovnání s parametrem L – Ekstraliga (POL)

V této lize můžeme vidět průběh relativní hodnoty při nižších hodnotách L vyšší. Z toho můžeme usuzovat, že při sázení na nižší počet zápasů se zvyšuje riziko zápasů, při kterých bude výsledek zápasů jiný než odhadovaný výsledek.

8.2.3 NHL

Pro NHL ligu jsme zvolili BP-DI model jako nejvhodnější, pomocí kterého jsme vypočítali pravděpodobnosti jednotlivých variant výsledku zápasů od 16. kola v sezóně 2015/2016, jak bylo uvedeno v předchozích kapitolách. Předpokládáme dostupné finanční prostředky na začátku sázení ve výši 9 000 Kč, kvůli vyššímu počtu zápasů oproti ostatním ligám, aby nedocházelo k častým bankrotům.

Model je opět ziskový, ale u této ligy pouze pro vyšší hodnoty parametru L, tj. 1,18 a vyšší (kromě $L = 1{,}20$). Při hodnotě $L = 1{,}18$ by to znamenalo vsadit pouze na 74 zápasů z 1 139 možných (z toho by bylo vítězných zápasů pouze 25).

	L												
	1,01	**1,03**	**1,06**	**1,07**	**1,09**	**1,12**	**1,15**	**1,18**	**1,21**	**1,24**	**1,27**	**1,30**	**1,33**
počet sázek	686	557	401	358	288	199	124	74	43	32	20	11	5
celkem vsazeno	6 860	5 570	4 010	3 580	2 880	1 990	1 240	740	430	320	200	110	50
počet vítězných sázek	218	182	130	117	96	61	36	25	14	12	8	5	4
celkem čistý zisk/ztráta	-923,80	-504,60	-270,90	-164,40	-97,70	-189,90	-160,30	41,60	16,10	64,90	62,10	36,70	71,50
relativní hodnota majetku	0,87	0,91	0,93	0,95	0,97	0,90	0,87	1,06	1,04	1,20	1,31	1,33	2,43
na kolik % zápasů by se vsadilo	60%	49%	35%	32%	25%	18%	11%	7%	4%	3%	2%	1%	0%

Tab. 39: Souhrn pro sázení dle BP-DI modelu – NHL

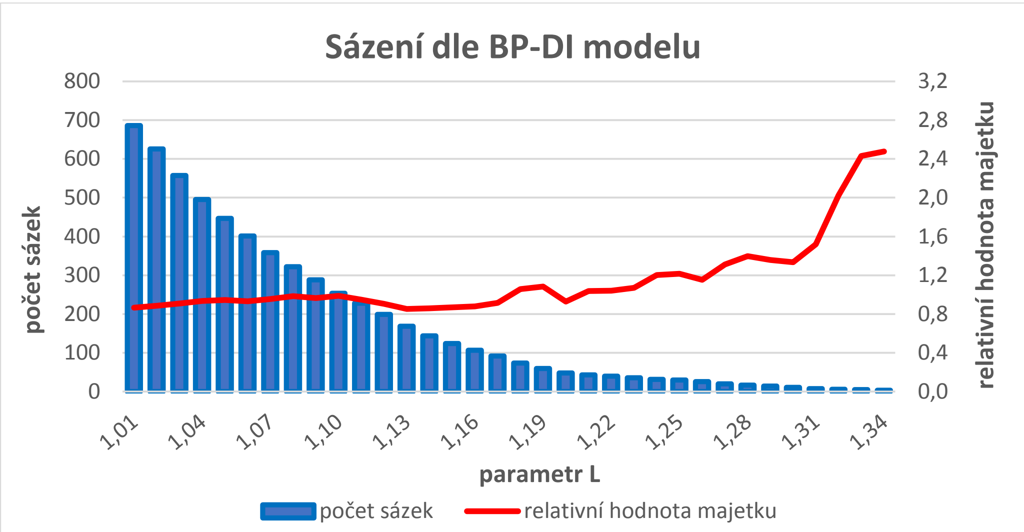
Nejvyššího celkového zisku model dosahuje až při $L = 1{,}33$. V tom případě budeme mít na konci sezóny k dispozici 9 072 Kč, tj. o 72 Kč více než byla hodnota finančních prostředků na začátku sezóny. V průběhu sezóny nenastala situace, že by sázející zbankrotoval.

V tabulce níže je pro znázornění uvedena část zápasů této ligy, na které dle modelu vsadíme, včetně výše čistého zisku (případně ztráty) u jednotlivých zápasů, při zvoleném $L = 1{,}18$.

Datum	Kolo	Domácí	Hosté	Skóre domácí	Skóre hosté	Výsledek	Vsazeno/Nevsazeno			Výhra/Prohra		
							1	X	2	1	X	2
31.10.2015	24	Los Angeles	Nashville	3	3	0	0	0	0	0,00	0,00	0,00
31.10.2015	24	Dallas	San Jose	5	3	1	0	0	0	0,00	0,00	0,00
31.10.2015	24	New Jersey	NY Islanders	2	2	0	0	0	0	0,00	0,00	0,00
31.10.2015	24	Phoenix	Vancouver	3	4	2	0	0	1	0,00	0,00	15,20
31.10.2015	24	Calgary	Montreal	2	6	2	0	0	1	0,00	0,00	14,70
31.10.2015	24	Minnesota	Chicago	5	4	1	0	0	0	0,00	0,00	0,00
31.10.2015	24	Florida	Boston	1	3	2	0	0	0	0,00	0,00	0,00
31.10.2015	24	Detroit	Ottawa	1	3	2	0	0	0	0,00	0,00	0,00
31.10.2015	24	Washington	Columbus	2	1	1	0	0	0	0,00	0,00	0,00
31.10.2015	24	NY Rangers	Toronto	3	1	1	0	0	0	0,00	0,00	0,00
31.10.2015	24	Carolina	Colorado	3	2	1	0	0	0	0,00	0,00	0,00
31.10.2015	24	Buffalo	Philadelphia	3	1	1	0	0	1	0,00	0,00	-10,00
1.11.2015	25	Carolina	Tampa Bay	3	4	2	0	0	0	0,00	0,00	0,00
1.11.2015	25	Colorado	San Jose	3	4	2	0	0	0	0,00	0,00	0,00
1.11.2015	25	Edmonton	Calgary	4	5	2	0	0	1	0,00	0,00	18,80
1.11.2015	25	St. Louis	Minnesota	2	2	0	0	0	0	0,00	0,00	0,00
1.11.2015	25	Toronto	Pittsburgh	0	4	2	0	0	0	0,00	0,00	0,00
1.11.2015	25	Tampa Bay	Boston	1	3	2	0	0	1	0,00	0,00	23,70
1.11.2015	25	Ottawa	Detroit	3	5	2	0	0	0	0,00	0,00	0,00
1.11.2015	25	Florida	Washington	1	1	0	0	0	0	0,00	0,00	0,00
1.11.2015	25	Columbus	Atlanta	2	3	2	0	0	0	0,00	0,00	0,00
2.11.2015	26	Anaheim	Nashville	4	2	1	0	0	0	0,00	0,00	0,00
2.11.2015	26	NY Islanders	Buffalo	1	2	2	0	0	0	0,00	0,00	0,00
2.11.2015	26	Montreal	Atlanta	5	1	1	0	0	0	0,00	0,00	0,00
3.11.2015	27	Vancouver	Philadelphia	4	1	1	0	0	0	0,00	0,00	0,00
3.11.2015	27	Chicago	Los Angeles	4	2	1	0	0	0	0,00	0,00	0,00
3.11.2015	27	Toronto	Dallas	4	1	1	0	0	0	0,00	0,00	0,00
4.11.2015	28	San Jose	Columbus	2	5	2	0	0	0	0,00	0,00	0,00
4.11.2015	28	Edmonton	Philadelphia	4	2	1	0	0	1	0,00	0,00	-10,00
4.11.2015	28	Colorado	Calgary	6	3	1	0	0	0	0,00	0,00	0,00
4.11.2015	28	St. Louis	Los Angeles	0	3	2	0	0	0	0,00	0,00	0,00
4.11.2015	28	Montreal	Ottawa	1	1	0	0	0	0	0,00	0,00	0,00
4.11.2015	28	Detroit	Tampa Bay	2	1	1	0	0	0	0,00	0,00	0,00

Tab. 40: Výsledky sázení (zisk/ztráta) dle BP-DI modelu ($L = 1{,}18$) – NHL

Na následujícím obrázku je ukázán graf počtu vsazených zápasů při různých hodnotách parametru L pro NHL ligu. Jsou zde znázorněné pouze výsledky, kdy bylo vsazeno alespoň čtyřikrát za sezónu.



Obr. 45: Relativní hodnota majetku v porovnání s parametrem L – NHL

Vidíme, že průběh relativní hodnoty majetku je lehce rostoucí, kromě posledních hodnot parametru L, kde se relativní hodnota majetku významně zvedá – zde bychom tedy sázeli na zápasy s velmi vysokou pravděpodobností výhry, jejichž počet je jen velmi malý.

8.3 Sázení dle vlastního modelu

V této části využijeme upravené modely definované v kapitole 7 k sázení proti sázkovým kancelářím, stejně jako v předchozí kapitole 8.2. Detail zpracování lze nalézt v souboru *CZE_VM_DP_15-16.xlsm* na listu *sázení dle modelu* (obdobně pro polskou ligu a NHL).

8.3.1 Extraliga (CZE)

Pro data z české ligy jsme zvolili jako nejvhodnější upravený DP model. Předpokládáme opět dostupné finanční prostředky na začátku sázení ve výši 1 000 Kč.

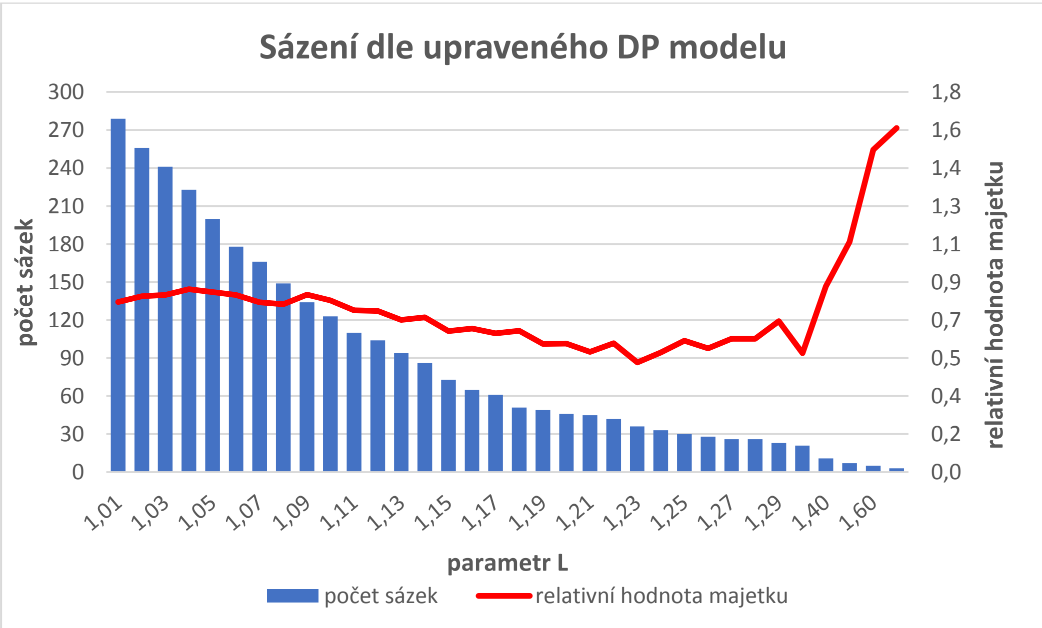
Zkoumali jsme vliv různých hodnot parametru L na ziskovost (resp. ztrátovost) modelu a na počet vsazených a vítězných zápasů, což lze vidět v tabulce níže. V průběhu sezóny nedošlo při sázení dle tohoto modelu k bankrotu.

	L												
	1,01	**1,03**	**1,06**	**1,09**	**1,12**	**1,15**	**1,18**	**1,21**	**1,24**	**1,27**	**1,30**	**1,50**	**1,60**
počet sázek	279	241	178	134	104	73	51	45	33	26	21	7	5
celkem vsazeno	2 790	2 410	1 780	1 340	1 040	730	510	450	330	260	210	70	50
počet vítězných sázek	100	89	64	48	33	20	14	10	7	6	4	2	2
celkem čistý zisk/ztráta	-543,50	-386,20	-286,80	-214,20	-246,80	-241,90	-168,50	-193,90	-143,10	-95,70	-91,60	6,30	26,30
relativní hodnota majetku	0,81	0,84	0,84	0,84	0,76	0,67	0,67	0,57	0,57	0,63	0,56	1,09	1,53
na kolik % zápasů by se vsadilo	85%	73%	54%	41%	32%	22%	16%	14%	10%	8%	6%	2%	2%

Tab. 41: Souhrn pro sázení dle upraveného DP modelu – vlastní model (CZE)

Nejvyššího celkového zisku model dosahuje při zvoleném parametru $L \in [1{,}56; 1{,}61]$. Při jakékoliv hodnotě z tohoto intervalu bychom měli mít na konci sezóny k dispozici 1 026 Kč, tj. pouze o 26 Kč více než je hodnota finančních prostředků na začátku sezóny. Pro ostatní zkoumané hodnoty tohoto parametru vycházel model jako ztrátový.

Relativní hodnota majetku je znázorněna na Obr. 46, kde jsou zobrazeny pouze výsledky při alespoň třech vsazených zápasech za sezónu.



Obr. 46: Relativní hodnota majetku v porovnání s parametrem *L* – vlastní model (CZE)

Průběh křivky relativní hodnoty majetku naznačuje, že se nám nejvíce vyplácí sázet na nízký počet zápasů, kde jsou ve větší míře správné odhady výsledků zápasů. Graf ovlivňují poslední vysoké hodnoty tohoto parametru, a pokud se podíváme na detail tohoto ukazatele pro vyšší počet zápasů, na které sázíme, vidíme, že se jeho hodnota zvedá s narůstajícím počtem zápasů a může se zdát, že se tak snižuje riziko ztráty. Nicméně i tak je relativní hodnota majetku pod hranicí výnosnosti.

Upravený DP model vychází v porovnání s původním modelem pro českou ligu méně výhodný z pohledu zisku. Navržený model by byl sice výhodnější než ostatní triviální strategie sázení, ale až při vyšších hodnotách parametru L, kde bychom sázeli na minimální počet zápasů. Oproti původnímu modelu použitého pro data z české ligy, je tento navržený model téměř pro všechny hodnoty L neziskový.

8.3.2 Ekstraliga (POL)

Pro polskou ligu jsme zvolili jako nejvhodnější upravený DP-DI model.

Předpokládáme dostupné finanční prostředky na začátku sázení ve výši 1 000 Kč. V průběhu sázení nedošlo k bankrotu sázejícího.

Vliv různých hodnot parametru $L > 1$ na ziskovost (resp. ztrátovost) modelu a na počet vsazených a vítězných zápasů je zobrazen v Tab. 42.

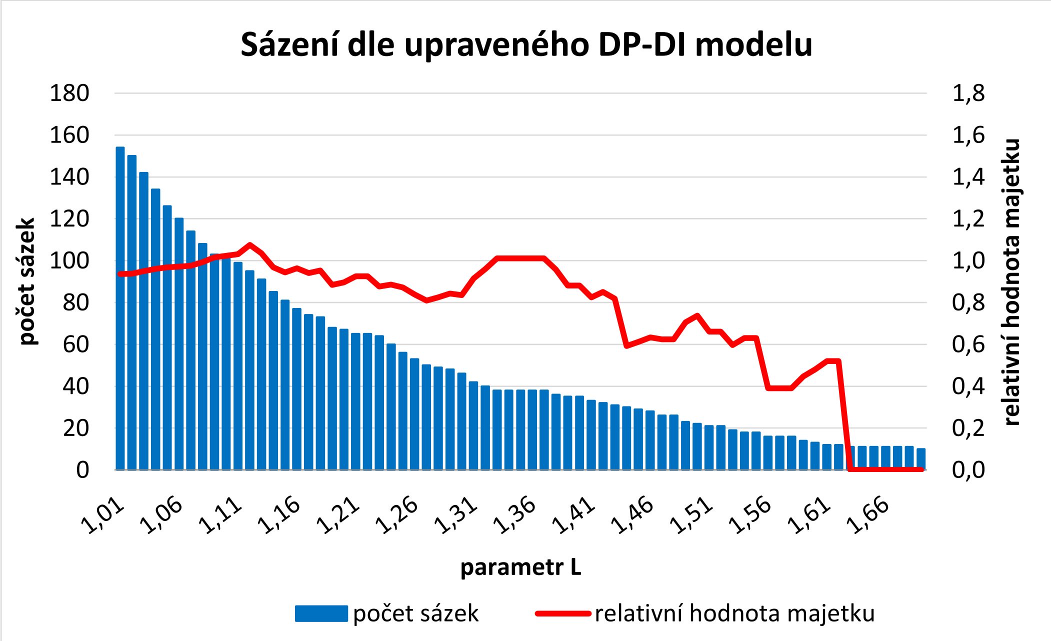
	L												
	1,01	1,03	1,06	1,09	1,12	1,13	1,15	1,18	1,21	1,24	1,27	1,30	1,33
počet sázek	154	142	120	103	95	91	81	73	65	60	50	46	38
celkem vsazeno	1 540	1 420	1 200	1 030	950	910	810	730	650	600	500	460	380
počet vítězných sázek	70	62	49	40	38	36	29	26	21	18	13	12	12
celkem čistý zisk/ztráta	-99,20	-70,60	-34,90	16,70	71,40	31,80	-45,30	-34,30	-48,90	-68,20	-95,70	-75,50	4,50
relativní hodnota majetku	0,94	0,95	0,97	1,02	1,08	1,03	0,94	0,95	0,92	0,89	0,81	0,84	1,01
na kolik % zápasů by se vsadilo	72%	66%	56%	48%	44%	42%	38%	34%	30%	28%	23%	21%	18%

Tab. 42: Sázení dle upraveného DP-DI modelu pro různé hodnoty L – Ekstraliga (POL)

Model je ziskový pro hodnotu parametru $L = 1{,}09$. V tomto případě by bylo 40 vítězných sázek z celkově vsazených 103 zápasů.

Nejvyššího celkového zisku model dosahuje při zvoleném parametru $L = 1{,}12$. To znamená, že na konci sezóny budeme mít k dispozici 1 071 Kč, tj. o 71 Kč více než je hodnota finančních prostředků na začátku sezóny.

Na obrázku níže je znázorněna relativní hodnota majetku. V grafu jsou znázorněny pouze výsledky, když bychom vsadili alespoň na deset zápasů za sezónu.



Obr. 47: Relativní hodnota majetku v porovnání s parametrem L – vlastní model (POL)

Průběh této křivky je opačný, než u ostatních modelů. Při sázení dle tohoto modelu s parametrem $L = 1{,}56$ a vyšších hodnot vychází, že by vítězná sázka byla pouze jedna nebo žádná. To je dáno vysokým kurzem na zápasy, kde model odhaduje podobnou pravděpodobnost výhry domácího a hostujícího týmu. Tím, že je kurz jednoho z týmů násobně vyšší (například nad hodnotou 6), tak dle pravidla sázení vsadíme na tento kurz, protože můžeme vydělat potenciálně mnohem více. Ve skutečnosti výsledky zápasů dopadly opačně. Tento fakt ovlivňuje průběh křivky relativní hodnoty majetku.

Upravený DP-DI model vychází v porovnání s původním modelem pro polskou ligu výhodnější z pohledu zisku, ale pouze při zvolení hodnoty parametru $L = 1{,}12$. Při této hodnotě poskytoval vlastní model čistý zisk ve výši 71 Kč, kdežto původní model pouze 32 Kč. Porovnáme-li ztrátovost při různých hodnotách parametru L, dosahoval původní model mnohem nižších ztrát než upravený model (i pro nižší hodnoty parametru L). Naopak při vyšších hodnotách parametru L dosahoval vlastní model ztráty, oproti původnímu modelu, který byl ziskový. Tudíž lze říci, že lepší variantou pro data z polské ligy je původní model, kde bylo větší množství hodnot parametru L, při kterém bychom dosahovali zisku, a tím i menší riziko ztráty vsazených prostředků.

Oproti původnímu modelu použitého pro data z polské ligy, je tento navržený model téměř pro všechny hodnoty L neziskový.

8.3.3 NHL

Pro NHL ligu byl vybrán BP-DI model jako nejvhodnější.

Předpokládáme dostupné finanční prostředky na začátku sázení ve výši 9 000 Kč.

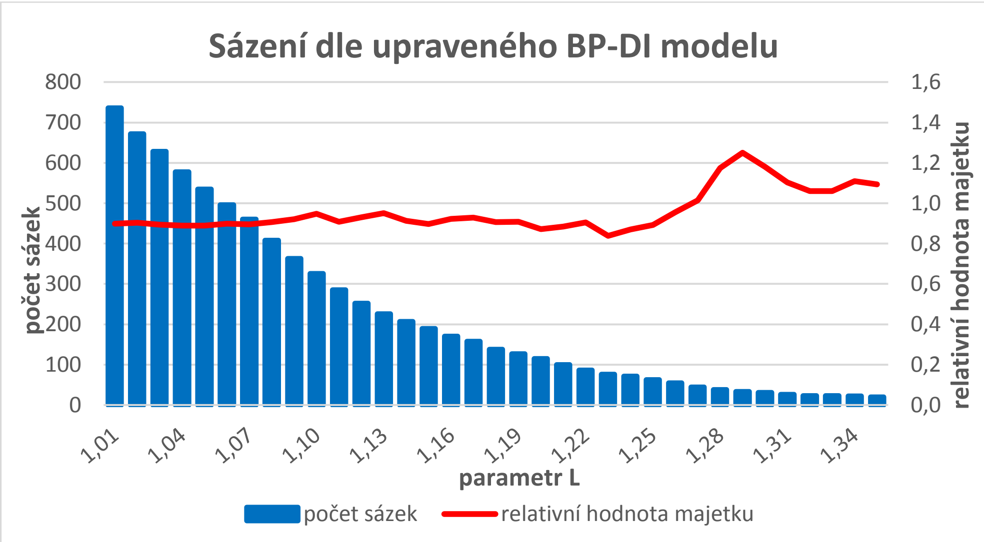
Model je opět ziskový, ale u tohoto upraveného modelu pouze pro vyšší hodnoty parametru L, tj. 1,27 a vyšší. Při hodnotě $L = 1{,}27$ bychom vsadili dle sázecího kritéria pouze na 44 zápasů z 1 139 možných (z toho by bylo vítězných zápasů pouze 13).

	L												
	1,01	**1,03**	**1,06**	**1,09**	**1,12**	**1,15**	**1,18**	**1,21**	**1,27**	**1,29**	**1,30**	**1,33**	**1,35**
počet sázek	736	628	496	363	252	189	138	100	44	33	31	23	20
celkem vsazeno	7 360	6 280	4 960	3 630	2 520	1 890	1 380	1 000	440	330	310	230	200
počet vítězných sázek	249	207	159	118	81	56	40	28	13	12	10	7	6
celkem čistý zisk/ztráta	-742,50	-665,40	-505,80	-289,70	-175,10	-195,30	-129,80	-115,50	6,30	82,60	56,10	13,90	18,70
relativní hodnota majetku	0,90	0,89	0,90	0,92	0,93	0,90	0,91	0,88	1,01	1,25	1,18	1,06	1,09
na kolik % zápasů by se vsadilo	65%	56%	44%	32%	22%	17%	12%	9%	4%	3%	3%	2%	2%

Tab. 43: Souhrn pro sázení dle upraveného BP-DI modelu – NHL

Nejvyššího celkového zisku model dosahuje až při $L = 1{,}29$. V tom případě budeme mít na konci sezóny k dispozici 9 083 Kč, tj. o 83 Kč více než byla hodnota finančních prostředků na začátku sezóny. V průběhu sezóny nenastala situace, že by sázející zbankrotoval.

Na následujícím obrázku je ukázán graf počtu vsazených zápasů při různých hodnotách parametru L pro NHL ligu. Jsou zde znázorněné pouze výsledky, kdy bylo vsazeno alespoň dvacetkrát za sezónu.



Obr. 48: Relativní hodnota majetku v porovnání s parametrem L – vlastní model (NHL)

Pro vyšší počet vsazených zápasů je relativní hodnota majetku na podobné úrovni, kolem hodnoty 0,9, tj. pod hranicí ziskovosti. Pro vysoké hodnoty parametru L se relativní hodnota majetku zvyšuje až nad hranici, kdy máme finanční prostředky na konci sezóny vyšší než na začátku.

Finanční prostředky na konci sezóny by byly pro vlastní model o 11 Kč vyšší než u maximální výše finančních prostředků v původním modelu. Dle upraveného modelu BP-DI bychom sázeli na podstatně více zápasů než v původním modelu.

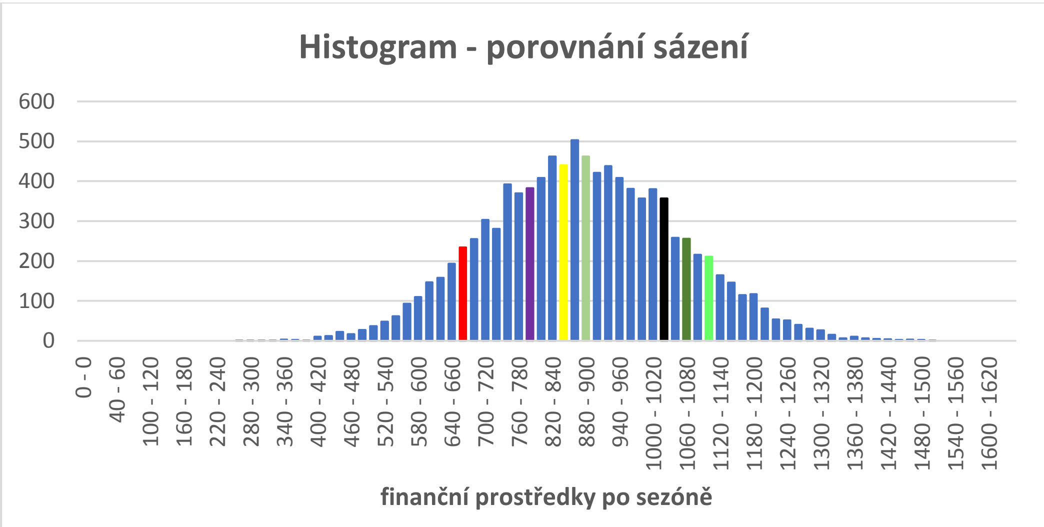
8.4 Porovnání modelu s triviálními strategiemi

V této části porovnáváme vybrané modely z kapitoly 8.2 s triviálními strategiemi popsanými v kapitole 8.1. Abychom ukázali, že definované modely jsou výhodnější než triviální strategie sázení.

8.4.1 Extraliga (CZE)

Pro Extraligu porovnáváme vybraný BP model s nastaveným parametrem $L = 1{,}07$, který přinášel nejvyšší čistý zisk, s jednotlivými triviálními strategiemi. U tohoto modelu vycházelo, že bychom vsadili pouze na 45 % zápasů. Z toho důvodu je zpracovaná ještě jedna varianta náhodného sázení na 45 % zápasů, kterou využíváme k porovnání s vybraným modelem a s triviálními schématy sázení (na listu *porovnání sázení*).

Do histogramu náhodného sázení na 45 % zápasů jsme přidali triviální strategie sázení, a to konkrétně sázení na výhru domácích, remízu, výhru hostů, na minimální a maximální kurzy. Přidali jsme také sázení dle BP modelu pro tři různé hodnoty parametru L ($L = 1{,}04$; $1{,}07$ a $1{,}14$), abychom viděli porovnání daného modelu i při jiných hodnotách parametru L. Tyto možnosti jsou barevně vyznačené v následujícím histogramu, který zobrazuje četnosti finančních prostředků na konci sezóny v daných intervalech.



Obr. 49: Porovnání různých typů sázení (CZE)

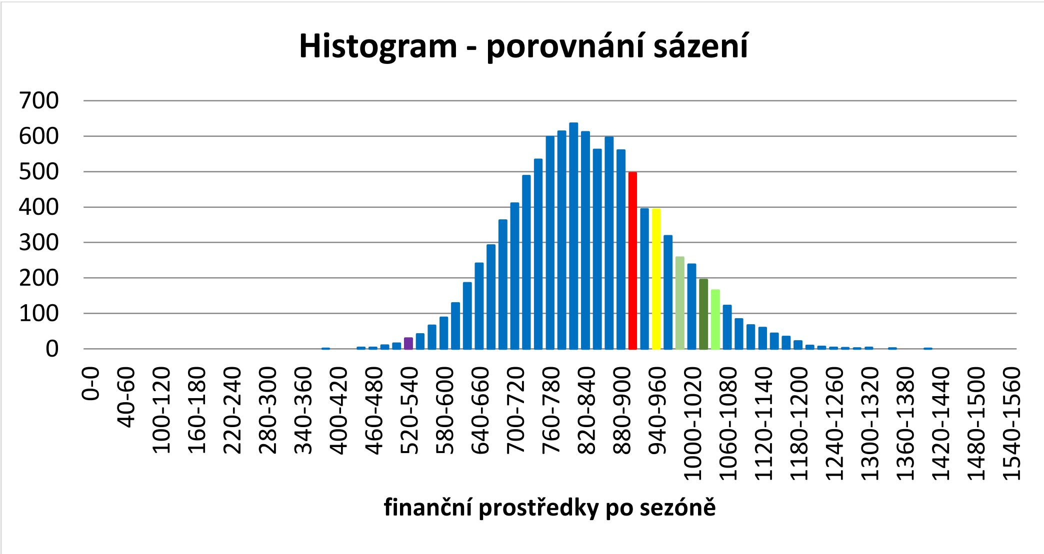
Odstíny zelené barvy v histogramu jsou použité pro variantu sázení dle BP modelu s vybranými třemi hodnotami parametru L. Zelený sloupec zprava značí model s parametrem $L = 1{,}07$, kdy tento model vycházel nejvíce ziskový, sloupec vlevo od této hodnoty je pro model s $L = 1{,}14$ a světle zelený pro $L = 1{,}04$. Černý sloupec je pro variantu sázení na max. kurzy, žlutý pro min. kurzy, fialový pro výhru domácích a remízu a červený pro výhru hostů.

Hodnoty znázorněné v předchozím histogramu v pravé části zobrazují vyšší zisk. Z tohoto srovnání je patrné, že model je nejvhodnější variantou oproti triviálním strategiím sázení. Model není ziskový pouze při menších hodnotách parametru L.

8.4.2 Ekstraliga (POL)

Vybraný BP model pro polskou ligu jsme také porovnali s triviálními strategiemi sázení, abychom ukázali, že je pro naše data výhodnější z pohledu sázení.

Nejvhodnější a zároveň nejvíce ziskovou variantou je pro naše data BP model s hodnotou parametru $L = 1{,}13$, kde bychom vsadili pouze na 35 % zápasů, proto porovnáváme tento model s triviálními strategiemi sázení na histogramu náhodného sázení pro stejné procento zápasů, stejným způsobem jako u české ligy.



Obr. 50: Porovnání různých typů sázení (POL)

Odstíny zelené barvy v histogramu jsou použité pro variantu sázení dle BP modelu s vybranými třemi hodnotami parametru L. Zelený sloupec zprava značí model s parametrem $L = 1{,}13$, kdy tento model vycházel nejvíce ziskový, sloupec vlevo od této hodnoty je pro model s $L = 1{,}09$ a světle zelený pro $L = 1{,}02$. Žlutý sloupec zobrazuje sázení na min. kurzy, červený výhru hostů a fialový výhru domácích. Strategie sázení na maximální kurzy skončila v mínusu, tzn., že bychom měli finanční prostředky po sezóně ve výši 281 Kč, stejně tak varianta sázení na remízy, kde výše finančních prostředků po sezóně byla 142 Kč.

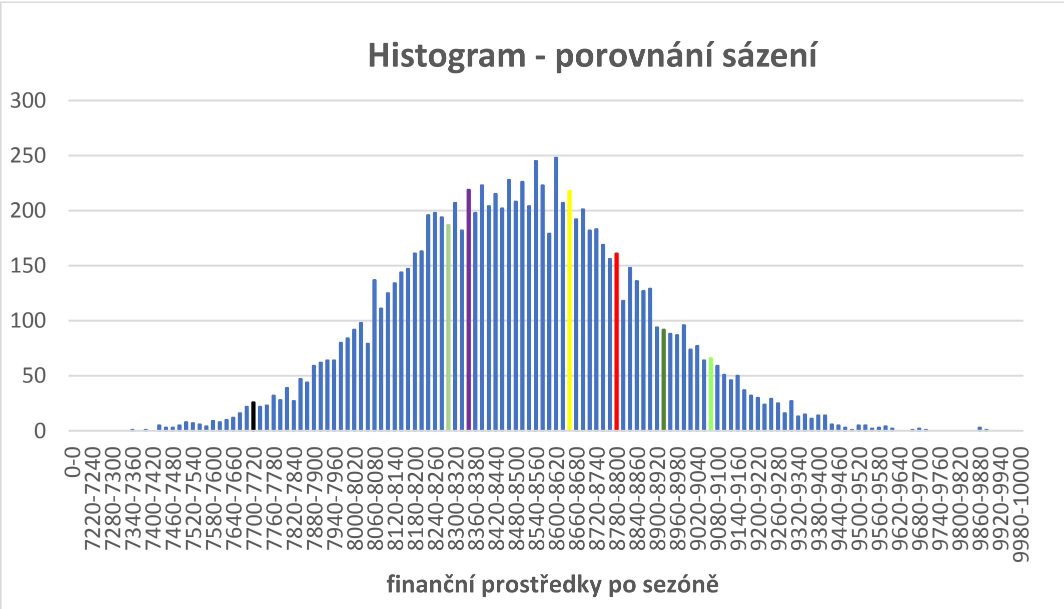
Lze tedy říci, že nejvýhodnější variantou je sázení dle modelu pro všechny tři zkoumané hodnoty parametru L. Nejvýhodnější variantou z triviálních strategií se ukázalo sázení na minimální kurzy, kde byla výše finančních prostředků 955 Kč, tj. nižší než počáteční finanční prostředky.

8.4.3 NHL

Vybraný BP-DI model jsme porovnali s triviálními strategiemi sázení, abychom ukázali, že je pro naše data vhodnější použít tento model.

Vybraný model přinášel nejvyšší čistý zisk s nastaveným parametrem $L = 1{,}33$, kde vycházelo, že bychom vsadili pouze na 0,4 % zápasů, což je příliš malé procento na srovnání s náhodným sázením, a proto model porovnáváme s triviálními strategiemi sázení na histogramu náhodného sázení s 60 % zápasů (při minimální zkoumané hodnotě parametru L, tj. 1,01).

V histogramu níže porovnáváme BP-DI model pro tři různé hodnoty parametru L (L = *1,02*; *1,09* a *1,18*) s triviálními strategiemi sázení. Tyto možnosti jsou barevně vyznačené v následujícím histogramu.



Obr. 51: Porovnání různých typů sázení (NHL)

Odstíny zelené barvy v histogramu jsou použité pro variantu sázení dle BP-DI modelu s vybranými třemi hodnotami parametru L. Zelený sloupec zprava značí model s parametrem $L = 1{,}18$, sloupec vlevo od této hodnoty je pro model s $L = 1{,}09$ a světle zelený pro $L = 1{,}02$. Červený sloupec je pro výhru hostů a žlutý sloupec pro min. kurzy. Fialový sloupec zobrazuje variantu sázení na výhru domácích a černý sloupec sázení na max. kurzy. Varianta sázení na remízu se ukázala jako nejméně výhodná varianta z pohledu zisku, kde bychom měli po celé sezóně finanční prostředky ve výši pouze 7 223 Kč. Nejvýhodnější varianta sázení je tedy zvolený model s parametrem $L = 1{,}18$.

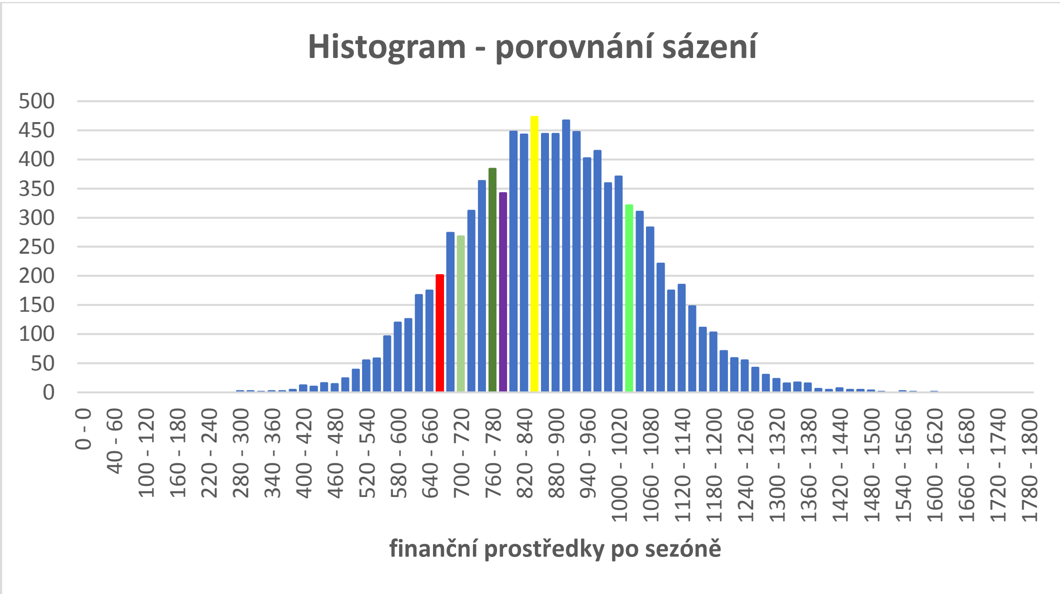
8.5 Porovnání vlastního modelu s triviálními strategiemi

Postup v této části je shodný s kapitolou 8.4, tedy porovnáváme upravený model s triviálními strategiemi.

8.5.1 Extraliga (CZE)

Na základě upraveného DP modelu a výběru parametru $L = 1{,}60$ byl model nejvíce ziskový, a vycházelo, že bychom vsadili pouze na 2 % zápasů. To je ovšem velmi malé procento na srovnávání s náhodným sázením, a proto model porovnáváme na histogramu náhodného sázení na 45 % zápasů (pro názornost stejně jako pro českou ligu v původním modelu) s triviálními strategiemi sázení.

V histogramu níže porovnáváme upravený DP model pro tři různé hodnoty parametru L ($L = 1{,}04$; $1{,}14$ a $1{,}60$) s triviálními strategiemi sázení. Tyto možnosti jsou barevně vyznačené v následujícím histogramu.



Obr. 52: Porovnání různých typů sázení – vlastní model (CZE)

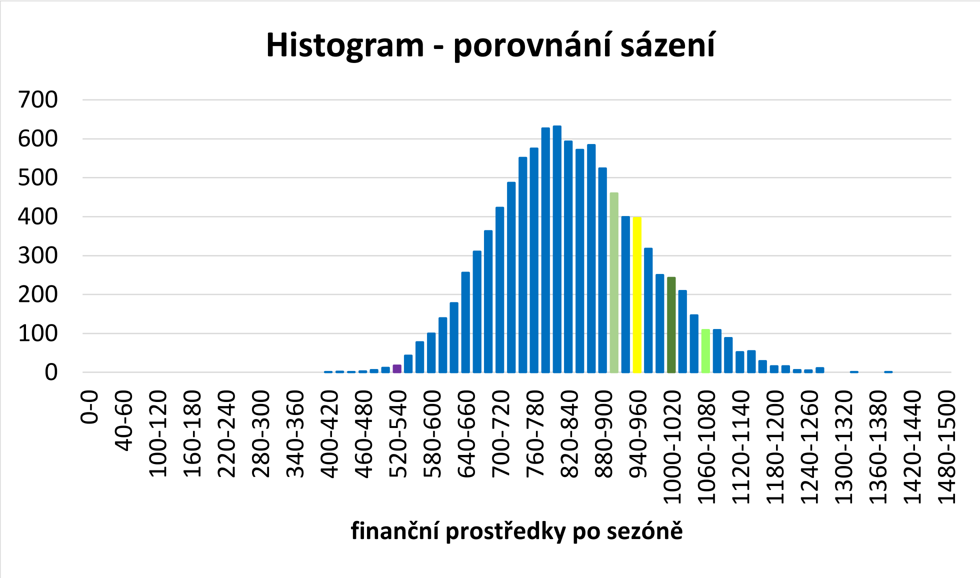
Odstíny zelené barvy v histogramu jsou použité pro variantu sázení dle upraveného DP modelu s vybranými třemi hodnotami parametru L. Zelený sloupec zprava značí model s parametrem $L = 1{,}60$ a sázení na max. kurzy, kdy tato strategie vycházela nejvíce zisková, světle zelený vlevo označuje výši finančních prostředků po sezóně pro model s $L = 1{,}04$ a tmavě zelený vedle je pro $L = 1{,}14$. Žlutý sloupec je pro min. kurzy, a fialový sloupec pro výhru domácích i remízu, a poslední červený sloupec pro výhru hostů.

Je tedy zřejmé, že v tomto případě je nejvhodnější variantou sázení na maximální kurzy, což značí, že takto upravený DP model nebude nejvhodnější modelem pro českou ligu.

V této lize se objevují vysoké kurzy na výhru velmi slabého týmu, kde zápas skončí právě výhrou slabého týmu, proto je strategie sázení na maximální kurzy tak výhodná. Toto platí pouze pro českou ligu, u ostatní lig byla tato strategie naopak nevýhodná.

8.5.2 Ekstraliga (POL)

Upravený DP-DI model jsme porovnali s triviálními schématy sázení na následujícím histogramu pro sázení na 35 % zápasů, stejně jako u původního modelu. Nejvhodnější a zároveň nejvíce ziskovou variantou je pro naše data upravený DP-DI model s hodnotou parametru $L = 1{,}12$.



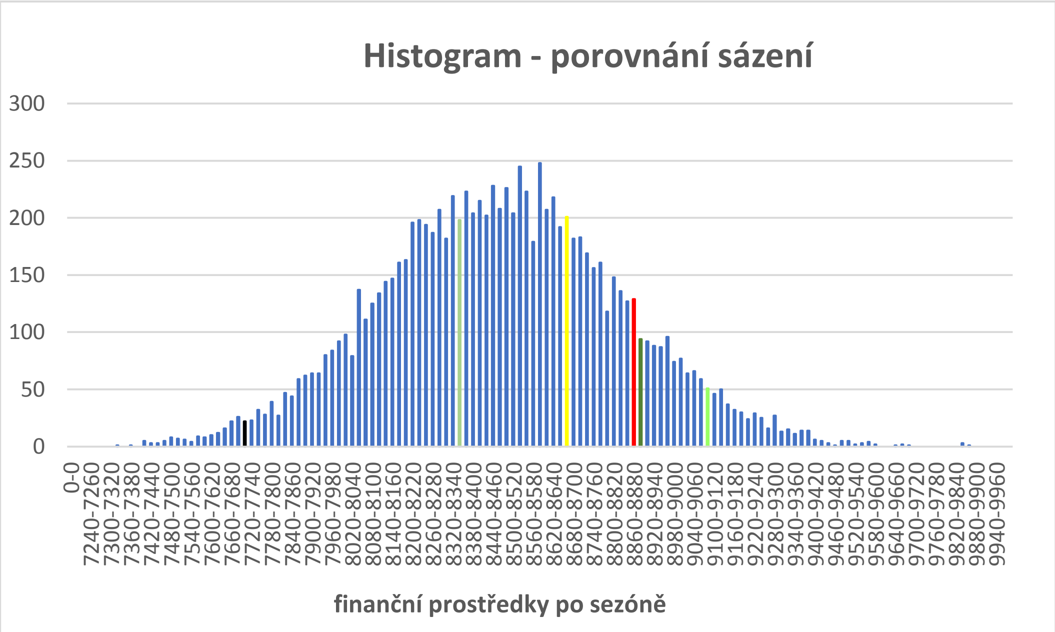
Obr. 53: Porovnání různých typů sázení – vlastní model (POL)

Odstíny zelené barvy v histogramu jsou použité pro variantu sázení dle upraveného DP-DI modelu s vybranými třemi hodnotami parametru L. Zelený sloupec zprava značí model s parametrem $L = 1{,}12$, kdy tento model vycházel nejvíce ziskový, sloupec vlevo od této hodnoty je pro model s $L = 1{,}09$ a světle zelený pro $L = 1{,}02$ a také pro variantu sázení na výhru hostů, kde výše finančních prostředků po sezóně leží ve stejném intervalu. Žlutý sloupec je pro variantu sázení na min. kurzy a fialový pro výhru domácích. Strategie sázení na maximální kurzy skončila v mínusu, tzn., že bychom měli finanční prostředky po sezóně ve výši 281 Kč, stejně tak varianta sázení na remízy, kde výše finančních prostředků po sezóně byla 142 Kč.

Z histogramu je vidět, že nejvíce ziskovou variantou je sázení dle vybraného modelu při hodnotě parametru $L = 1{,}12$. Při porovnání všech triviálních strategií je vidět, že dopadla nejlépe strategie sázení na min. kurzy, i když by byla pro sázejícího ztrátová, tj. měl by o 45 Kč méně finančních prostředků po sezóně než na začátku sázení.

8.5.3 NHL

K porovnání upraveného BP-DI modelu pro vybrané hodnoty parametru *L* (*L* = *1,02*; *1,18* a *1,29*) s triviálními schématy sázení využíváme histogram náhodného sázení na 60 % zápasů, stejně jako u původního modelu.



Obr. 54: Porovnání různých typů sázení – vlastní model (NHL)

Odstíny zelené barvy v histogramu jsou použité pro variantu sázení dle BP-DI modelu s vybranými třemi hodnotami parametru *L*. Zelený sloupec zprava značí model s parametrem $L = 1{,}29$, sloupec vlevo od této hodnoty je pro model s $L = 1{,}18$ a světle zelená pro $L = 1{,}02$, a také pro variantu sázení na výhru domácích, kde výše finančních prostředků po sezóně leží ve stejném intervalu. Červený sloupec je pro výhru hostů a žlutý sloupec je pro min. kurzy. Černý sloupec zobrazuje variantu sázení na max. kurzy, a hnědá pro remízu, která se ukázala jako nejméně výhodná varianta z pohledu zisku, kde bychom měli po celé sezóně finanční prostředky ve výši pouze 7 223 Kč.

9 Závěr

Na začátku této práce byly popsány modely pro odhadování výsledků fotbalových utkání, které používal M. J. Maher ve svém článku [1]. Z těchto modelů jsme vybrali model 2, který jsme využili ke zpracování výsledků hokejových zápasů české ligy, na kterém jsme ověřili, že není příliš vhodný, a proto jsme se zaměřili na jiné modely.

Další část práce byla zaměřena na upravené Dixon-Colesovy modely použité na české lize, polské lize a NHL, které jsou již přizpůsobeny hokejovým datům. Tyto modely zahrnují informace z historických výsledků zápasů a za použití váhové funkce přikládají vyšší váhu výsledkům nejnovějších zápasů, tj. zohledňují při odhadech parametrů, jak týmy hrají v průběhu sezóny.

Definovali jsme funkci $S(\xi)$, kterou lze považovat za měřítko předvídací chyby výsledků, a snažíme se jí maximalizovat. Všechny upravené Dixon-Colesovy modely vycházely pro českou ligu téměř shodně z pohledu této funkce, nicméně BP model vycházel nejlépe. Tento model jsme použili pro odhad výsledku zápasu (výhru domácích, remízu nebo výhru hostů) v sezóně 2015/2016 a poté k porovnání s jednoduššími strategiemi sázení (například sázení pouze na výhru domácích, nebo na min. kurz či max. kurz, atd.). Model se ukázal jako nejziskovější varianta při sázení. Při aplikování tohoto modelu pro sezónu 2015/2016 jsme uvažovali počáteční finanční prostředky ve výši 1 000 Kč se sázkou 10 Kč na každý zápas. Ze 148 vybraných zápasů, z nichž 67 bylo pro nás vítězných, jsme docílili zisku na konci sezóny ve výši 117 Kč.

Pro polskou ligu byly modely dle hodnoty funkce $S(\xi)$ také téměř shodné, a nejvhodněji vycházel BP model, který se použil pro odhad výsledků zápasů v sezóně 2015/2016 a následně k porovnání s jednoduchými strategiemi sázení. Model byl také výhodnější oproti triviálním strategiím sázení. Při použití stejných počátečních finančních prostředků jako v české lize a shodné výši sázky, nám při 75 vybraných zápasech (z nichž vítězných bylo 31) vycházel zisk na konci sezóny ve výši 52 Kč.

Pro NHL ligu byl zvolen model BP-DI, pro který na rozdíl od ostatních lig vycházela hodnota $S(\xi)$ významně vyšší v porovnání s ostatními modely. K dosažení zisku při sázení dle tohoto modelu, bychom vsadili pouze na 74 zápasů (kdy 25 bylo vítězných) z celkem 1 139 možných. V tomto případě jsme dosáhli zisku 42 Kč.

Další část práce obsahuje zpracování vlastního modelu (resp. upraveného Dixon-Colesova modelu), kde jsme se snažili odhadnout parametr γ pro každý tým zvlášť, který vyjadřuje výhodu domácího prostředí. U těchto modelů se nakonec ukázalo, že výhoda domácího prostředí se mezi jednotlivými týmy skutečně liší, ale tato úprava nepřinesla vyšší zisk v porovnání s původními modely pro všechny zkoumané ligy.

Další možné rozšíření práce by bylo zkoumání různých úprav modelů, například vytvoření nové váhové funkce nebo její úprava. Dále by bylo vhodné prozkoumat, proč výhoda domácího prostředí pro každý tým zvlášť neměla pozitivní vliv na lepší odhady výsledků zápasů.

Seznam obrázků

Obr. 1: Ukázka kontingenční tabulky 6
Obr. 2: Ukázka nastavení výchozích hodnot pro odhad parametrů α_i a β_i 17
Obr. 3: Histogram četností – domácí zápasy 20
Obr. 4: Histogram četností – venkovní zápasy 20
Obr. 5: Rozdíly odhadů α v jednotlivých kolech – CZE 27
Obr. 6: Rozdíly odhadů β v jednotlivých kolech - CZE 28
Obr. 7: Hodnoty $S(\xi)$ proti ξ pro BP model (CZE) 28
Obr. 8: Nastavení řešitele pro BP model v Microsoft Excel 2013 29
Obr. 9: Max. věrohodné odhady parametru útoku α pro nejlepší a nejhorší tým (CZE) 31
Obr. 10: Max. věrohodné odhady parametru obrany β pro nejlepší a nejhorší tým (CZE) 31
Obr. 11: Rozdíly odhadů α v jednotlivých kolech – POL 34
Obr. 12: Rozdíly odhadů β v jednotlivých kolech - POL 35
Obr. 13: Hodnoty $S(\xi)$ proti ξ pro BP model (POL) 36
Obr. 14: Odhad parametrů pro 52. kolo (tj. z výsledků do 51. kola včetně) 37
Obr. 15: Max. věrohodné odhady parametru útoku α pro dva nejlepší týmy (POL) 37
Obr. 16: Max. věrohodné odhady parametru obrany β pro dva nejlepší týmy (POL) 38
Obr. 17: Rozdíly odhadů α v jednotlivých kolech – NHL 39
Obr. 18: Rozdíly odhadů β v jednotlivých kolech - NHL 40
Obr. 19: Hodnoty $S(\xi)$ proti ξ pro BP-DI model (NHL) 40
Obr. 20: Nastavení řešitele pro BP-DI model v Microsoft Excel 2013 41
Obr. 21: Odhad parametrů pro 177. kolo (tj. z výsledků do 176. kola včetně) 42
Obr. 22: Max. věrohodné odhady parametru útoku α pro dva nejlepší týmy (NHL) 43
Obr. 23: Max. věrohodné odhady parametru obrany β pro dva nejlepší týmy (NHL) 43
Obr. 24: Vývoj odhadu mixovacího parametru p 44
Obr. 25: Vývoj odhadu parametrů $\theta_{0,\ldots,4}$ 44
Obr. 26: Hodnoty $S(\xi)$ proti ξ pro upravený DP model (CZE) 46
Obr. 27: Vývoj odhadu parametru útoku α – vlastní model (CZE) 47
Obr. 28: Vývoj odhadu parametru obrany β – vlastní model (CZE) 47
Obr. 29: Vývoj odhadu parametru výhody domácího prostředí γ – vlastní model (CZE) 48
Obr. 30: Hodnoty $S(\xi)$ proti ξ pro upravený DP-DI model (POL) 49
Obr. 31: Vývoj odhadu parametru útoku α – vlastní model (POL) 50
Obr. 32: Vývoj odhadu parametru obrany β – vlastní model (POL) 51
Obr. 33: Vývoj odhadu parametru výhody domácího prostředí γ – vlastní model (POL) 51
Obr. 34: Hodnoty $S(\xi)$ proti ξ pro upravený BP-DI model (NHL) 53
Obr. 35: Vývoj odhadu parametru útoku α – vlastní model (NHL) 54
Obr. 36: Vývoj odhadu parametru obrany β – vlastní model (NHL) 55
Obr. 37: Vývoj odhadu parametru výhody domácího prostředí γ – vlastní model (NHL) 55
Obr. 38: Vývoj odhadu mixovacího parametru p – vlastní model (NHL) 56
Obr. 39: Vývoj odhadu parametrů $\theta_{0,\ldots,4}$ – vlastní model (NHL) 56
Obr. 40: Porovnání náhodného sázení na 100 % a 60 % zápasů – Extraliga (CZE) 60
Obr. 41: Porovnání náhodného sázení na 100 % a 60 % zápasů – Ekstraliga (POL) 60
Obr. 42: Porovnání náhodného sázení na 100 % a 60 % zápasů – NHL 61
Obr. 43: Relativní hodnota majetku v porovnání s parametrem L – Extraliga (CZE) 63
Obr. 44: Relativní hodnota majetku v porovnání s parametrem L – Ekstraliga (POL) 65

Obr. 45: Relativní hodnota majetku v porovnání s parametrem *L* – NHL 67
Obr. 46: Relativní hodnota majetku v porovnání s parametrem *L* – vlastní model (CZE) .. 68
Obr. 47: Relativní hodnota majetku v porovnání s parametrem *L* – vlastní model (POL) .. 70
Obr. 48: Relativní hodnota majetku v porovnání s parametrem *L* – vlastní model (NHL) . 71
Obr. 49: Porovnání různých typů sázení (CZE) 72
Obr. 50: Porovnání různých typů sázení (POL) 73
Obr. 51: Porovnání různých typů sázení (NHL) 74
Obr. 52: Porovnání různých typů sázení – vlastní model (CZE) 75
Obr. 53: Porovnání různých typů sázení – vlastní model (POL) 76
Obr. 54: Porovnání různých typů sázení – vlastní model (NHL) 77

Seznam tabulek

Tab. 1: Pozorovaný a očekávaný počet gólů pro domácí tým Zlín 9
Tab. 2: Výsledky χ^2-testu pro Zlín 9
Tab. 3: Přehled výsledků jednotlivých hypotéz – Extraliga (CZE) 9
Tab. 4: Přehled výsledků jednotlivých hypotéz – NHL 10
Tab. 5: Přehled výsledků jednotlivých hypotéz – Ekstraliga (POL) 10
Tab. 6: Výsledky χ^2-testu nezávislosti pro sezónu 2014/2015 - Extraliga (CZE) 11
Tab. 7: Výsledky χ^2-testu nezávislosti pro sezónu 2014/2015 – NHL 12
Tab. 8: Výsledky χ^2-testu nezávislosti pro sezónu 2014/2015 - Ekstraliga (POL) 12
Tab. 9: Odhadování parametrů α_i a β_i pro sezónu 2014/2015 17
Tab. 10: Pravděpodobnosti, že domácí tým dá 1 gól hostujícímu týmu 18
Tab. 11: Pozorované a očekávané četnosti pro domácí zápasy – sezóna 2014/2015 19
Tab. 12: Výsledky χ^2-testu pro domácí zápasy – sezóna 2014/2015 19
Tab. 13: Pozorované a očekávané četnosti pro venkovní zápasy – sezóna 2014/2015 19
Tab. 14: Výsledky χ^2-testu pro venkovní zápasy – sezóna 2014/2015 20
Tab. 15: Hodnoty $S(\xi)$ pro jednotlivé modely – Extraliga (CZE) 29
Tab. 16: Odhad parametrů pro 74. kolo (tj. z výsledků do 73. kola včetně) 30
Tab. 17: Odhadnuté parametry pro zápas Plzně s Chomutovem 32
Tab. 18: Pravděpodobnosti výsledků pro zápas Plzně s Chomutovem 33
Tab. 19: Pravděpodobnost výhry domácích, remízy a výhry hostů 33
Tab. 20: Porovnání hodnot $S(\xi)$ pro Maherův model 2 a BP model 34
Tab. 21: Hodnoty $S(\xi)$ pro jednotlivé modely – Ekstraliga (POL) 36
Tab. 22: Hodnoty $S(\xi)$ pro diagonálně rozšířené modely – NHL 41
Tab. 23: Hodnoty S(ξ) pro DP model – NHL 41
Tab. 24: Hodnoty S(ξ) pro BP model - NHL 41
Tab. 25: Počet gólů domácích/hostů pro týmy z Extraligy v sezóně 2014/2015 45
Tab. 26: Hodnoty $S(\xi)$ pro jednotlivé upravené modely (CZE) 46
Tab. 27: Počet gólů domácích/hostů pro týmy z Ekstraligy pro sezónu 2014/2015 49
Tab. 28: Hodnoty $S(\xi)$ pro jednotlivé upravené modely bez DP modelu (POL) 50
Tab. 29: Hodnoty $S(\xi)$ pro upravený DP model (POL) 50
Tab. 30: Počet gólů domácích/hostů pro týmy z NHL v sezóně 2014/2015 53
Tab. 31: Hodnoty $S(\xi)$ pro upravený DP model (NHL) 54
Tab. 32: Hodnoty $S(\xi)$ pro upravený BP model (NHL) 54
Tab. 33: Hodnoty $S(\xi)$ pro upravený DP-DI model (NHL) 54
Tab. 34: Hodnoty $S(\xi)$ pro upravený BP-DI model (NHL) 54
Tab. 35: Sázení dle BP modelu pro různé hodnoty L – Extraliga (CZE) 62
Tab. 36: Výsledky sázení (zisk/ztráta) dle BP modelu ($L = 1{,}07$) - CZE 63
Tab. 37: Sázení dle BP modelu pro různé hodnoty L – Ekstraliga (POL) 64
Tab. 38: Výsledky sázení (zisk/ztráta) dle BP modelu ($L = 1{,}13$) - POL 65
Tab. 39: Souhrn pro sázení dle BP-DI modelu – NHL 66
Tab. 40: Výsledky sázení (zisk/ztráta) dle BP-DI modelu ($L = 1{,}18$) – NHL 67
Tab. 41: Souhrn pro sázení dle upraveného DP modelu – vlastní model (CZE) 68
Tab. 42: Sázení dle upraveného DP-DI modelu pro různé hodnoty L – Ekstraliga (POL) . 69
Tab. 43: Souhrn pro sázení dle upraveného BP-DI modelu – NHL 71

Literatura a zdroje

[1] MAHER, M. J. (1982). Modelling association fotball scores. Statistica Neerlandica, 36, pp. 109 - 118

[2] DIXON, M. J., & COLES, S. G. (1997). Modelling Association Fotball Scores and Inefficiencies in the Fotball Betting Market. Journal of the Royal Statistical Society. Series C (Applied Statistics), 46(2), pp. 265 - 280.

[3] MAREK, P., ŠEDIVÁ, B., ŤOUPAL, T. (2014). Modeling and prediction of ice hockey match results. *Journal of Quantitative Analysis in Sports*, 10(3), pp. 357 – 365.

[4] KARLIS, D., NTZOUFRAS, I. (2003). Analysis of sports data by using bivariate Poisson models. The Statistician, 52(3), pp. 381 - 393

[5] FAMOYE, F. (2010). A new bivariate generalized Poisson distribution. *Statistica Neerlandica*, 64(1), pp. 112 - 124

[6] HÁTLE, J., & LIKEŠ, J. (1974). Základy počtu pravděpodobnosti a matematické statistiky. Praha: SNTL

[7] REIF, J. *Metody matematické statistiky*. Plzeň: Západočeská univerzita, 2004, s. 61 – 63. ISBN 80-7043-302-7

[8] Pravděpodobnost a statistika hypertextově. P-hodnota. [online]. 2014 [cit. 2017-08-03]. Dostupné z: http://home.zcu.cz/~friesl/hpsb/phodn.html

[9] ŠEDIVÁ, B. Přednášky MSM. [online]. 2007-2017 [cit. 2017-08-03]. Dostupné z: https://courseware.zcu.cz/portal/studium/courseware/kma/msm/prednasky.html

[10] Microsoft. *GRG Algorithm.* [online]. © 2017 [cit. 2016-05-20]. Dostupné z: https://support.microsoft.com/en-us/kb/82890

[11] ŠPAČEK, J. Modelování a odhadování výsledků sportovních utkání: bakalářská práce. Plzeň: ZČU Plzeň, Fakulta aplikovaných věd, 2015.

[12] Extraliga ledního hokeje – Wikipedie. [online]. Dostupné z: https://cs.wikipedia.org/wiki/Extraliga_ledn%C3%ADho_hokeje

[13] Ekstraliga w hokeju na lodzie – Wikipedie. [online]. 2018 [cit. 2018-03-20]. Dostupné z: https://cs.wikipedia.org/wiki/Ekstraliga_w_hokeju_na_lodzie

[14] Pravidla hry - nhlnews.cz. *nhlnews.cz - NHL, novinky, video, výsledky NHL, Jágr v NHL* [online]. Copyright © Copyright nhlnews.cz. [cit. 2017-08-09]. Dostupné z: http://nhlnews.cz/historie/pravidla-hry/

[15] National Hockey League – Wikipedie. [online]. 2018 [cit. 2018-03-20]. Dostupné z: https://cs.wikipedia.org/wiki/National_Hockey_League

[16] Definice řešitele – support.office.com – nápověda k MS Office [online]. Copyright © Microsoft 2018. [cit. 2018-08-09]. Dostupné z: https://support.office.com/cs-cz/article/definov%C3%A1n%C3%AD-a-vy%C5%99e%C5%A1en%C3%AD-probl%C3%A9mu-pomoc%C3%AD-%C5%98e%C5%A1itele-5d1a388f-079d-43ac-a7eb-f63e45925040

[17] ABDI, Herve. The University of Texas at Dallas. The Bonferonni and Šidák Corrections for Multiple Comparisons. [online]. 2007 [cit. 2018-03-23]. Dostupné z: http://www.utdallas.edu/~herve/Abdi-Bonferroni2007-pretty.pdf

Zdroje dat

[A] Sfstats. [online]. © 2006 - 2012 [cit. 2016-20-05]. Dostupné z: http://www.sfstats.net/

[B] Odds portal. [online]. © 2008 - 2016 [cit. 2016-20-05]. Dostupné z: http://www.oddsportal.com/

[C] BetExplorer. [online] © 2003 - 2016 [cit. 2016-20-05]. Dostupné z: http://www.betexplorer.com/hockey/

Přílohy

Přílohy obsažené na CD:

DP_Gabrišková.pdf - soubor s kompletním textem diplomové práce v elektronické podobě,

data.xlsx – soubor obsahující data pro všechny tři ligy,

chybějící_kurzy.xlsx – soubor s chybějícími kurzy a zápasy, které bylo nutné dohledávat,

CZE – složka obsahující veškeré soubory s testy, odhady a sázení pro českou ligu

POL - složka obsahující veškeré soubory s testy, odhady a sázení pro polskou ligu

NHL – složka obsahující veškeré soubory s testy, odhady a sázení pro NHL ligu

vlastní model - složka obsahující veškeré soubory s testy, odhady a sázení pro upravený model

Přílohy tištěné

Příloha 1: kód vytvořeného makra pro odhad parametrů Maherova modelu (1 iterace) v sezóně 2014/2015

```
Sub Odhad_2014_2015()

' Hledání optimálního odhadu parametrů alfa a beta pomocí řešitele

Application.ScreenUpdating = False

' Spuštění řešitele pro sezónu 2014/2015
 SolverReset

 SolverOk SetCell:="$R$18", MaxMinVal:=3, ValueOf:=1077, ByChange:="$L$4:$M$17" _
        , Engine:=1, EngineDesc:="GRG Nonlinear"

 SolverAdd CellRef:=Range("$P$18"), Relation:=2, FormulaText:="$Q$18"

 SolverSolve UserFinish:=True

' nastavení odhadnutých hodnot jako výchozí pro další iteraci

    Range("P4:Q17").Select
    Selection.Copy
    Range("L4").Select
    Selection.PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks _
        :=False, Transpose:=False


End Sub
```

Příloha 2: ukázka kódu vytvořeného makra pro odhad parametrů BP modelu pro českou ligu v sezóně 2015/2016

```
Sub Odhad_BP_model()

Application.ScreenUpdating = False
Application.Calculation = xlAutomatic

Sheets("BP model").Select
Start = ActiveSheet.Range("B26").Value
Max = Start + ActiveSheet.Range("B27").Value
If Max > ActiveSheet.Range("C27").Value Then
Max = ActiveSheet.Range("C27").Value
End If

For i = Start To Max
Application.Calculation = xlAutomatic
Sheets("BP model").Select
ActiveSheet.Range("B28").Offset(i, 0).Select
Application.CutCopyMode = False
Selection.Copy
ActiveSheet.Range("B25").Select
Selection.PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks _
        :=False, Transpose:=False


' start řešitel maximalizace

SolverReset

  SolverOk SetCell:="$B$24", MaxMinVal:=1, ValueOf:=0, ByChange:= _
        "$B$4:$C$20,$F$4:$H$4"
  SolverOptions MaxTime:=99999, Iterations:=32000, Scaling:=True, Precision:=0.00001
  SolverAdd CellRef:=Range("B21"), Relation:=2, FormulaText:="17"
  SolverAdd CellRef:=Range("C21"), Relation:=2, FormulaText:="17"
  SolverAdd CellRef:=Range("B4:C20"), Relation:=3, FormulaText:="0"

  SolverSolve UserFinish:=True


' uložení spočtených parametrů

    Sheets("BP model").Select
    Range("B4:B20").Select
    Application.CutCopyMode = False
    Selection.Copy
    Sheets("BP model_parametry").Select
    Range("XFD6").Select
    Selection.End(xlToLeft).Offset(0, 1).Select
    Selection.PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks _
        :=False, Transpose:=False

    Sheets("BP model").Select
    Range("C4:C20").Select
    Application.CutCopyMode = False
    Selection.Copy
    Sheets("BP model_parametry").Select
    Range("XFD28").Select
    Selection.End(xlToLeft).Offset(0, 1).Select
    Selection.PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks _
```

Příloha 3: ukázka kódu vytvořeného makra pro odhad parametrů BP-DI modelu pro NHL ligu v sezóně 2015/2016

```
Sub Odhad_BP_DI_model()

Application.ScreenUpdating = False
Application.Calculation = xlAutomatic

Sheets("BP-DI model").Select
Start = ActiveSheet.Range("B39").Value
Max = Start + ActiveSheet.Range("B40").Value
If Max > ActiveSheet.Range("C40").Value Then
Max = ActiveSheet.Range("C40").Value
End If

For i = Start To Max
Application.Calculation = xlAutomatic
Sheets("BP-DI model").Select
ActiveSheet.Range("B41").Offset(i, 0).Select
Application.CutCopyMode = False
Selection.Copy
ActiveSheet.Range("B38").Select
Selection.PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks _
        :=False, Transpose:=False


' start řešitel maximalizace

  SolverReset

  SolverOk SetCell:="$B$37", MaxMinVal:=1, ValueOf:=0, ByChange:= _
        "$B$4:$C$33,$F$4:$N$4"
  SolverOptions MaxTime:=9999, Iterations:=32000, Scaling:=True, Precision:=0.00001
  SolverAdd CellRef:=Range("B34"), Relation:=2, FormulaText:="30"
  SolverAdd CellRef:=Range("C34"), Relation:=2, FormulaText:="30"
  SolverAdd CellRef:=Range("B4:C33"), Relation:=3, FormulaText:="0,1"
  SolverAdd CellRef:=Range("H4"), Relation:=3, FormulaText:="0"
  SolverAdd CellRef:=Range("P4"), Relation:=2, FormulaText:="1,000"
  SolverAdd CellRef:=Range("I4"), Relation:=3, FormulaText:="0"
  SolverAdd CellRef:=Range("J4"), Relation:=3, FormulaText:="0"
  SolverAdd CellRef:=Range("K4"), Relation:=3, FormulaText:="0"
  SolverAdd CellRef:=Range("L4"), Relation:=3, FormulaText:="0"
  SolverAdd CellRef:=Range("M4"), Relation:=3, FormulaText:="0"
  SolverAdd CellRef:=Range("N4"), Relation:=3, FormulaText:="0"

  SolverSolve UserFinish:=True

' uložení spočtených parametrů

    Sheets("BP-DI model").Select
    Range("B4:B33").Select
    Application.CutCopyMode = False
    Selection.Copy
    Sheets("BP-DI model_parametry").Select
    Range("XFD5").Select
    Selection.End(xlToLeft).Offset(0, 1).Select
    Selection.PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks _
        :=False, Transpose:=False
```

Příloha 4: ukázka kódu makra použité při náhodném sázení na 100 % zápasů

```
Sub Aktualizace_100()

' Aktualizace (F9) a uchování dostupných finančních prostředků po sezóně

Application.ScreenUpdating = False

For i = 1 To 10000

    Calculate
    Range("N366").Select
    Selection.Copy
    ActiveSheet.Range("R41").Offset(i, 0).Select
    Selection.PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks _
        :=False, Transpose:=False
    Application.CutCopyMode = False

Next

End Sub
```