

UNIVERZA V LJUBLJANI  
EKONOMSKA FAKULTETA

MAGISTRSKO DELO

**UPORABA VERJETNOSTNIH MODELOV ZA NAPOVEDOVANJE  
PRIHODNIH NAKUPOV STRANK  
NA PRIMERU IZBRANEGA PODJETJA**

Ljubljana, avgust 2016

ANDREJ BLAHA

## IZJAVA O AVTORSTVU

Podpisani Andrej Blaha, študent Ekonomske fakultete Univerze v Ljubljani, avtor predloženega dela z naslovom Uporaba verjetnostnih modelov za napovedovanje prihodnjih nakupov strank, pripravljenega v sodelovanju s svetovalcem dr. Jurijem Jakličem in sosvetovalko dr. Tanjo Dmitrović

### IZJAVLJAM

1. da sem predloženo delo pripravil samostojno;
2. da je tiskana oblika predloženega dela istovetna njegovi elektronski obliki;
3. da je besedilo predloženega dela jezikovno korektno in tehnično pripravljeno v skladu z Navodili za izdelavo zaključnih nalog Ekonomske fakultete Univerze v Ljubljani, kar pomeni, da sem poskrbel, da so dela in mnenja drugih avtorjev oziroma avtoric, ki jih uporabljam oziroma navajam v besedilu, citirana oziroma povzeta v skladu z Navodili za izdelavo zaključnih nalog Ekonomske fakultete Univerze v Ljubljani;
4. da se zavedam, da je plagiatorstvo – predstavljanje tujih del (v pisni ali grafični obliki) kot mojih lastnih – kaznivo po Kazenskem zakoniku Republike Slovenije;
5. da se zavedam posledic, ki bi jih na osnovi predloženega dela dokazano plagiatorstvo lahko predstavljalo za moj status na Ekonomski fakulteti Univerze v Ljubljani v skladu z relevantnim pravilnikom;
6. da sem pridobil vsa potrebna dovoljenja za uporabo podatkov in avtorskih del v predloženem delu in jih v njem jasno označil;
7. da sem pri pripravi predloženega dela ravnal v skladu z etičnimi načeli in, kjer je to potrebno, za raziskavo pridobil soglasje etične komisije;
8. da soglašam, da se elektronska oblika predloženega dela uporabi za preverjanje podobnosti vsebine z drugimi deli s programsko opremo za preverjanje podobnosti vsebine, ki je povezana s študijskim informacijskim sistemom članice;
9. da na Univerzo v Ljubljani neodplačno, neizključno, prostorsko in časovno neomejeno prenašam pravico shranitve predloženega dela v elektronski obliki, pravico reproduciranja ter pravico dajanja predloženega dela na voljo javnosti na svetovnem spletu preko Repozitorija Univerze v Ljubljani;
10. da hkrati z objavo predloženega dela dovoljujem objavo svojih osebnih podatkov, ki so navedeni v njem in v tej izjavi.

V Ljubljani, dne 24. 08. 2016

Podpis študenta:

# KAZALO

<b>UVOD .....</b>	<b>1</b>
<b>1 ŽIVLJENJSKA VREDNOST STRANK.....</b>	<b>3</b>
1.1 Oprelitev življenjske vrednosti strank .....	5
1.2 Kontekst življenjske vrednosti strank.....	7
1.3 modeli življenjske vrednosti strank .....	9
1.3.1 Modeli za izračun življenjske vrednosti strank .....	10
1.3.2 Modeli za analizo baze strank .....	11
1.3.3 Normativni modeli življenjske vrednosti strank.....	12
1.4 Izzivi modeliranja življenjske vrednosti strank.....	13
<b>2 VERJETNOSTNI NAPOVEDNI MODELI ZA ANALIZO BAZE STRANK ...</b>	<b>14</b>
2.1 Model NBD .....	15
2.2 Model Pareto/NBD .....	16
2.3 Model BG/NBD.....	19
2.4 Ostale izpeljanke modelov »kupuj do smrti« .....	20
2.4.1 Model CBG/NBD .....	21
2.4.2 Model CBG/CNBD-k.....	22
2.4.3 Model HB Pareto/NBD .....	23
2.5 Primerjava modelov in programska oprema za implementacijo modelov .....	24
<b>3 BLACK DIAMOND EQUIPMENT .....</b>	<b>27</b>
3.1 Trženje v podjetju.....	27
3.2 Vedenje strank in nakupno odločanje v primeru nakupa športnih rekvizitov .....	28
3.2.1 Motivacijski proces .....	30
3.2.2 Nakupno odločanje.....	31
3.3 Raziskovalna analiza podatkov .....	33
3.4 Izbira modelov in kriteriji primerjave .....	38
3.5 Uporaba modelov za napovedovanje prihodnjih nakupov strank .....	40
3.5.1 Uporaba modela Pareto/NBD .....	41
3.5.2 Uporaba modela BG/NBD.....	47
3.5.3 Uporaba modela CBG/NBD .....	51
3.5.4 Uporaba modela CBG/NBD-k.....	54
3.6 Primerjava posameznih modelov.....	60
3.6.1 Skladnost .....	60
3.6.2 Točnost .....	62
3.6.3 Hevristično odločanje in linearna regresija .....	63
3.7 Ugotovitve in nadaljnji koraki .....	64
3.8 Pomanjkljivosti in omejitve življenjske vrednosti strank.....	65
<b>SKLEP .....</b>	<b>67</b>
<b>LITERATURA IN VIRI .....</b>	<b>69</b>

## PRILOGE

### KAZALO TABEL

Tabela 1: Osredotočenost na izdelke oziroma stranke .....	4
Tabela 2: Kontekst CLV z vidika dveh dejavnikov .....	8
Tabela 3: Kontekst CLV z vidika treh dejavnikov .....	9
Tabela 4: Predpostavke modela NBD .....	16
Tabela 5: Predpostavke modela Pareto/NBD .....	18
Tabela 6: Predpostavke modela BG/NBD .....	19
Tabela 7: Matrika poslovnih odnosov s primeri ustreznih modelov .....	21
Tabela 8: Predpostavke modela CBG/NBD .....	21
Tabela 9: Predpostavke modela CBG/CNBD-k .....	22
Tabela 10: Predpostavke modela HB Pareto/NBD .....	23
Tabela 11: Primerjava modelov z vidika predpostavk in ocenjevanja parametrov .....	24
Tabela 12: Pregled osnovnih značilnosti razpoložljivih podatkov .....	34
Tabela 13: Razpoložljivi podatki .....	40
Tabela 14: Pripravljeni podatki .....	41
Tabela 15: Optimizacija in konvergiranje parametrov .....	42
Tabela 16: Dejanski in pričakovani povprečni nakupi ter število strank po skupinah v modelu Pareto/NBD .....	46
Tabela 17: Dejanski in pričakovani povprečni nakupi ter število strank po skupinah v modelu BD/NBD .....	49
Tabela 18: Dejanski in pričakovani povprečni nakupi ter število strank po skupinah v modelu CBG/NBD .....	53
Tabela 19: Dejanski in pričakovani povprečni nakupi, ter število strank po skupinah v modelu CBG/CNBD-2 .....	58
Tabela 20: Ocenjene vrednosti parametrov .....	60
Tabela 21: Logaritmi verjetja .....	61
Tabela 22: Dejanski in pričakovani celotni nakupi v kalibracijskem obdobju .....	61
Tabela 23: Dejansko in ocenjeno število nakupov na koncu validacijskega obdobja .....	63
Tabela 24: Mere točnosti uporabljenih modelov .....	63
Tabela 25: Mere točnosti za enostavno hevristično odločanje in linearno regresijo .....	64

### KAZALO SLIK

Slika 1: Konceptualni okvir življenjske vrednosti strank .....	6
Slika 2: Dejavniki, ki vplivajo na življenjsko vrednost stranke .....	7
Slika 3: Tipično gibanje števila kumulativnih transakcij v scenariju »kupuj do smrti« .....	17
Slika 4: Negativna binomska porazdelitev z različnimi parametri .....	25
Slika 5: Poissonova porazdelitev z različnimi parametri .....	25
Slika 6: Gama porazdelitev z različnimi parametri .....	26

Slika 7: Beta porazdelitev z različnimi parametri .....	26
Slika 8: Motivacijski proces potrošnika .....	30
Slika 9: Nakupno odločanje.....	32
Slika 10: PCM model .....	33
Slika 11: Skupno število nakupov .....	34
Slika 12: Število dni od zadnjega nakupa.....	35
Slika 13: Povprečni čas med posameznimi nakupi .....	36
Slika 14: Vzorci nakupovanja v času desetih naključno izbranih strank .....	37
Slika 15: Deleži strank po številu nakupov .....	37
Slika 16: Število transakcij v letu 2014 .....	38
Slika 17: Heterogenost v nakupnem procesu .....	43
Slika 18: Heterogenost v procesu osipa.....	43
Slika 19: Pogostost ocenjenih in dejanskih ponovnih nakupov v kalibracijskem obdobju v modelu Pareto/NBD .....	44
Slika 20: Pričakovani nakupi v validacijskem obdobju glede na število nakupov v kalibracijskem obdobju v modelu Pareto/NBD.....	45
Slika 21: Gibanje ocenjenih in dejanskih nakupov po tednih v modelu Pareto/NBD.....	46
Slika 22: Gibanje ocenjenih in dejanskih kumulativnih nakupov po tednih v modelu Pareto/NBD .....	47
Slika 23: Porazdelitev stopnje nakupov $\alpha$ in stopnje osipa $p$ v modelu BG/NBD.....	48
Slika 24: Pogostost ocenjenih in dejanskih ponovnih nakupov v kalibracijskem obdobju.....	49
Slika 25: Pričakovani nakupi v validacijskem obdobju glede na število nakupov v kalibracijskem obdobju v modelu BD/NBD .....	50
Slika 26: Gibanje ocenjenih in dejanskih kumulativnih nakupov po tednih v modelu BG/NBD .....	50
Slika 27: Porazdelitev stopnje nakupov $\alpha$ in stopnje osipa $p$ v modelu CBG/NBD .....	52
Slika 28: Pogostost ocenjenih in dejanskih ponovnih nakupov v kalibracijskem obdobju v modelu CBG/NBD .....	52
Slika 29: Pričakovane transakcije v validacijskem obdobju glede na število transakcij v kalibracijskem obdobju v modelu CBG/NBD .....	53
Slika 30: Gibanje ocenjenih in dejanskih kumulativnih transakcij po tednih v modelu CBG/NBD .....	54
Slika 31: Porazdelitev ocenjenega parametra oblike za stranke, ki so opravile najmanj 5 nakupov .....	55
Slika 32: Zglajen histogram mere regularnosti za stranke, ki so opravile najmanj 3 nakupe .....	56
Slika 33: Porazdelitev stopnje nakupov $\alpha$ in stopnje osipa $p$ v modelu CBG/CNBD-2 .....	57
Slika 34: Pogostost ocenjenih in dejanskih ponovnih nakupov v kalibracijskem obdobju v modelu CBG/CNBD-2 .....	58
Slika 35: Pričakovani nakupi v validacijskem obdobju glede na število nakupov v kalibracijskem obdobju v modelu CBG/NBD.....	59

Slika 36: Gibanje ocenjenih in dejanskih kumulativnih nakupov po tednih v modelu CBG/CNBD-2 .....	59
Slika 37: Dejanski in pričakovani celotni nakupi v kalibracijskem obdobju.....	62
Slika 38: Pristopi k modeliranju verjetnostnih modelov .....	67

## UVOD

Izziv razumevanja in vrednotenja strank se je v zadnjih dveh desetletjih razvil v koncept, imenovan osrediščenost v stranki (angl. *customer centricity*)<sup>1</sup>. Koncept stranke obravnava kot neotipljivo premoženje podjetja, ki denarnega toka ne generira samo v preteklosti, temveč tudi v prihodnosti (Fader & Hardie, 2015).

Ključna metrika tega koncepta je življenjska vrednost stranke (angl. *customer lifetime value*, v nadaljevanju CLV), ki je definirana kot razlika med vsemi prihodki, ki jih posamezna stranka prinaša podjetju, in vsemi odhodki, povezanimi s pridobitvijo, prodajo in storitvami, povezanimi s to stranko, ob upoštevanju spremembe vrednosti denarja v času (Berger & Nasr, 1998).

Aktivnosti podjetja vplivajo na obnašanje strank, ki vpliva na vrednost posameznih strank. V strokovni literaturi je mogoče zaslediti vrsto modelov, ki se ukvarjajo z razumevanjem, kako te aktivnosti vplivajo na vrednost strank. In čeprav skušajo v osnovi vsi modeli odgovoriti na ista vprašanja (na primer: Kdo so naše najboljše stranke?), jih lahko razvrstimo v šest skupin (Gupta et al., 2006):

- modeli RFM,
- verjetnostni modeli (angl. *probability models*),
- ekonometrični modeli (angl. *econometric models*),
- vztrajnostni modeli (angl. *persistence models*),
- računalniški modeli (angl. *computer science models*),
- difuzijski modeli/modeli rasti (angl. *diffusion/growth models*).

Med vsemi naštetimi se verjetnostni modeli neposredno osredotočajo na izračun CLV in dve ključni vprašanji, na kateri moramo odgovoriti, če želimo napovedati obnašanje posamezne stranke v (več zaporednih) prihodnjih obdobjih, in sicer:

- ali bo stranka v opazovanem obdobju aktivna?
- Če bo v opazovanem obdobju aktivna, koliko transakcij bo v tem obdobju izvedla?

Verjetnostni modeli obnašanje strank interpretirajo kot izid stohastičnega procesa, ki je posledica latentnih vedenjskih nagnjenj, ki se med posameznimi strankami razlikujejo (Gupta et al., 2006). Takšni verjetnostni modeli obnašanja strank imajo v tem kontekstu dve komponenti. Prvič, verjetnost, da je stranka v prihodnjem obdobju še aktivna, sledi verjetnostni porazdelitvi, katere parametri odražajo vedenjska nagnjenja posamezne stranke (angl. *behavioral propensities*). In drugič, razlike v vedenjskih nagnjenjih posameznih

---

<sup>1</sup> Prevod Lieber, 2014.

strank so zajete v dodatni verjetnostni porazdelitvi. Ti komponenti skupaj tvorita mešano porazdelitev, ki karakterizira obnašanje naključno izbrane stranke (Fader & Hardie, 2015).

Različni verjetnostni modeli uporabljajo različne verjetnostne porazdelitve, s pomočjo katerih razlagajo obnašanje strank, kar pomembno vpliva na rezultate modela. Izbira ustreznega verjetnostnega modela je odvisna od značilnosti poslovnega odnosa med podjetjem in stranko. Predvsem sta relevantni dve dimenziji tega odnosa, in sicer tip odnosa in priložnosti za transakcije. Pri tipu odnosa razlikujemo med zavezujočim, pogodbenim odnosom, kjer je poslovni odnos točno določen, in nezavezujočim odnosom, kjer je poslovni odnos nedoločen. Z dimenzijo priložnosti za transakcije razlikujemo med zveznimi in diskretnimi priložnostmi za transakcije. Zvezna priložnost pomeni, da se lahko transakcija izvede kadarkoli, medtem ko se lahko v okviru diskretnih priložnosti transakcije izvedejo samo ob določenih trenutkih.

Posebna pozornost je v strokovni literaturi namenjena modelom, ki obravnavajo zvezne, nezavezujoče kontekste poslovnih odnosov med podjetji in strankami. Z vidika napovedovanja prihodnjih nakupov gre za najzahtevnejšo obliko poslovnih odnosov, saj podjetja nikdar zares ne vedo, kdaj izgubijo stranke (Fader & Hardie, 2009). Wharton Customer Analytics Initiative te modele označuje kot modeli »kupuj do smrti« (angl. *Buy 'Til You Die Models*), ki so jih izdali v R paketu, imenovanem BTYD. Platzer (2008) je paket dopolnil in razvija BTYDplus, ki vsebuje dodatne modele. Za uporabo teh modelov in napoved prihodnjih nakupov potrebujemo samo tri podatke o vsaki stranki:

- število nakupov, ki jih je stranka izvedla v opazovanem preteklem obdobju (angl. *frequency*),
- čas zadnjega nakupa v opazovanem preteklem obdobju (angl. *recency*),
- trajanje opazovanega obdobja.

**Namen** tega magistrskega dela je raziskati, kako uporabiti modele BTYD za napovedovanje prihodnjih nakupov strank pri izračunavanju CLV.

Osnovni **cilj** magistrskega dela je izmed modelov BTYD izbrati najprimernejšega za napovedovanje prihodnjih nakupov strank podjetja. Za doseg tega cilja bom najprej definiral poslovno okolje, v katerem deluje podjetje, in identificiral napovedne modele, ki so relevantni v tem kontekstu. Modele bom analiziral in ugotovil, kakšne podatke potrebujemo za njihovo uporabo in na kaj je treba biti pri tem pozoren.

Naslednji cilj je priprava zahtevanih podatkov, izvedba raziskovalne analize (angl. *exploratory analysis*) razpoložljivih podatkov in definiranje kriterija uspešnosti, s pomočjo katerega bom primerjal modele. S pomočjo izbranih modelov bom napovedal prihodnje nakupe, primerjal pridobljene rezultate in izbral najprimernejši napovedni model za podjetje.



Na koncu bom napovedne modele postavil še v kontekst izračuna CLV in opozoril na njihove pomanjkljivosti.

V prvem delu magistrske naloge bom podrobno pregledal strokovno literaturo, raziskave in članke z obravnavanega področja. Ta del magistrskega dela bo analiziran s pomočjo opisne metode in metode kompilacije, s katero bom združil spoznanja mnogih avtorjev.

V drugem, empiričnem delu magistrske naloge bom s pomočjo ustreznih računalniških programov in statističnih metod najprej opravil raziskovalno analizo razpoložljivih podatkov in opisno ter grafično predstavil njihove statistične značilnosti. V nadaljevanju pa bom relevantne podatke uporabil v izbranih modelih in empirične rezultate primerjal in analiziral. V sklepu bom z metodo sinteze magistrsko delo zaključil na osnovi spoznanj iz prejšnjih poglavij.

## **1 ŽIVLJENJSKA VREDNOST STRANK**

Dobičkonosna stranka je tista, katere celotni prihodki presegajo celotne odhodke, povezane s pridobitvijo, prodajo in strežbo te stranke v času. Maksimiziranje dolgoročne dobičkonosnosti strank je zajeto v konceptu življenjske vrednosti strank (Kotler, Keller, Manceau & Hemonnet-Goujot, 2015).

Jain in Singh (2002) ugotavljata, da velja prepričanje, da so lojalne stranke bolj dobičkonosne. Svoje razloge za takšno prepričanje utemeljujeta s trditvami, da so lojalne stranke pripravljene plačati določeno premijo za izdelke podjetja, pomagajo širiti dober glas o izdelkih in tako privabljajo nove stranke, podjetju prinašajo več prihodkov, management odnosov z njimi pa predstavlja manjši strošek. Trženjske aktivnosti se tako osredotočajo predvsem na pridobivanje novih strank in krepitev baze lojalnih strank v prepričanju, da so lojalne stranke bolj dobičkonosne.

Kumar in Rajan (2009) takšen pristop označujeta kot tradicionalni pristop k dobičkonosnosti. Inovativni in visoko kakovostni izdelki privabljajo stranke, pri čemer boljši izdelki zagotavljajo večjo bazo strank, ki so tudi bolj donosne. Nove stranke podjetje nagradi z boljšo izkušnjo, kar poveča njihovo zadovoljstvo. Takšne stranke so bolj zveste, širijo dober glas in privabljajo nove stranke ter s podjetjem opravijo več transakcij. Vse to pomeni višje prihodke in posledično višji dobiček. Ta se naloži v razvoj novih izdelkov, kar ponovno poveča bazo lojalnih strank. Tabela 1 prikazuje nekaj razlik med podjetji, osredotočenimi na izdelke, in tistimi, ki se osredotočajo na stranke.

Tabela 1: Osredotočenost na izdelke oziroma stranke

Področje	Podpodročje	Osredotočenost na izdelke	Osredotočenost na stranke
Strategija	Cilj	Najboljši izdelek za stranko	Najboljša rešitev za stranko
	Glavna ponudba	Novi izdelki	Poosebljena ponudba izdelkov, podpore za stranko
	Ustvarjanje dodane vrednosti	Tehnološko napredni izdelki, napredne funkcije in uporabe	Poosebitev pogojev za stranko
	Najpomembnejše stranke	Uporabniki najnaprednejših izdelkov	Najbolj dobičkonosne in lojalne stranke
	Prioriteta	Nabor produktov	Nabor (dobičkonosnih) strank
Struktura podjetja	Organizacija	Izdelki, kot so profitni centri, produktni vodje in produktni timi	Segmentacija strank, managerji strank in bilance strank
	Najpomembnejši proces	Razvoj novih izdelkov	Management odnosov s strankami
Uspešnost	Pomembni kazalniki uspeha	Število novih izdelkov in prejetih nagrad, tržni delež	Delež dobičkonosnih strank, zadovoljstvo strank, CLV
Ljudje	Pristop	Moč je pri tistih, ki snujejo izdelke	Moč je pri tistih, ki poznajo stranke
	Miselni proces	Različnost – koliko načinov uporabe izdelka	Usklajenost – kakšna kombinacija izdelkov je najboljša za stranko
	Kultura	Kultura novih izdelkov – odprtost za nove izdelke in eksperimentiranje	Kultura managementa odnosa s strankami – iskanje potreb strank, ki jih lahko zadovoljimo

Vir: J. Galbraith, *Designing the customer-centric organization: A guide to strategy, structure, and process*, 2011, str. 10, tabela 1.1.

V strokovni literaturi se pojavljajo pomisleki glede vpliva lojalnosti strank na dobičkonosnost. Reinartz in Kumar (2002) opozarjata, da ni dokazov, da je management lojalnih strank za podjetje nujno cenejši, da lojalne stranke opazno pripomorejo k pridobitvi novih strank ali da so dejansko pripravljene plačevati premijo na izdelke podjetja. Berger in Nasr (1998) svetujeta, naj se podjetja ne trudijo in poskušajo pridobiti in zadovoljiti vseh

strank, ampak naj se osredotočajo le na tiste stranke, ki so dobičkonosne. Prav tako Kumar, Pozza, Petersen in Shah (2009) trdijo, da management strank za lojalnost ni enako kot management strank za dobičkonosnost in da je treba lojalnost in dobičkonosnost strank managerati simultano.

## 1.1 Opredelitev življenjske vrednosti strank

Dwyer (1997) CLV definira kot neto sedanjo vrednost pričakovanih koristi, zmanjšano za stroške, povezane s stranko. Berger in Nasr (1998) življenjsko vrednost stranke definirata kot razliko med vsemi prihodki, ki jih posamezna stranka prinaša podjetju, in vsemi odhodki, povezanimi s pridobitvijo, prodajo in storitvami, povezanimi s to stranko, ob upoštevanju spremembe vrednosti denarja v času. Jain in Singh (2002) CLV razumeta kot neto dobiček ali izgubo, ki ga (oz. jo) podjetje dosega s transakcijami skozi celotno življenjsko dobo stranke. Gupta in Lehmann (2003) pravita, da je CLV sedanja vrednost vseh prihodnjih dobičkov, ki jih stranka ustvari. Kumar in Rajan (2009) CLV opredelita kot vsoto vseh denarnih tokov stranke, diskontiranih s tehtanimi povprečnimi stroški kapitala, skozi celotno življenjsko obdobje. Fader (2012) pa CLV definira kot sedanjo vrednost prihodnjih (neto) denarnih tokov posamezne stranke.

V najbolj osnovni obliki lahko CLV matematično zapišemo kot:

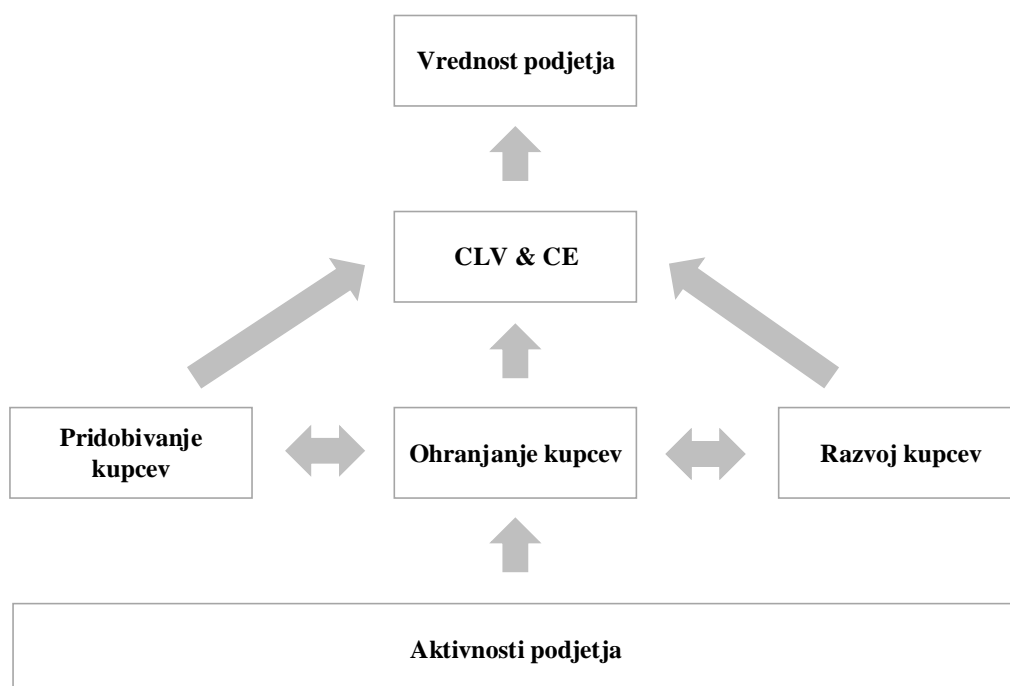
$$CLV = \sum_{i=1}^n \frac{(R_i - C_i)}{(1+d)^{i-0.5}}, \quad (1)$$

kjer  $i$  predstavlja obdobje transakcije;  $R_i$  prihodke v obdobju;  $C_i$  stroške v obdobju;  $d$  diskontno stopnjo;  $n$  število obdobj (Jain & Singh, 2002).

Konceptualni model CLV je prikazan na Sliki 1. Aktivnosti podjetja vplivajo na obnašanje strank, ki vpliva na vrednost posameznih strank. Vsota vrednosti celotne baze strank podjetja (angl. *customer equity*, v nadaljevanju CE) pa pomembno vpliva na vrednost podjetja na trgu.

Čeprav so si definicije CLV na prvi pogled zelo podobne, je treba biti pozoren na razlike. Dwyer (1997) med koristi šteje tako prihodke kot tudi neotipljive koristi (dober glas, priporočila). Gupta in Lehmann (2003) se osredotočata samo na prihodnje prihodke in odhodke in ne upoštevata preteklih transakcij. Berger in Nasr (1998) vključujeta stroške, medtem ko Gupta in Lehmann (2003) ter Kumar et al. (2009) stroškov ne upoštevajo.

Slika 1: Konceptualni okvir življenjske vrednosti strank



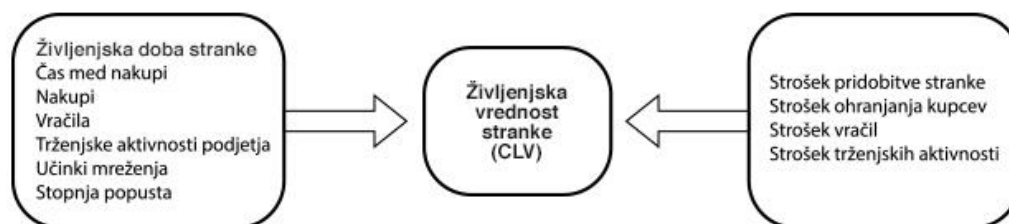
Vir: S. Gupta et al., *Modeling Customer Lifetime Value*, 2006, str. 140.

Pri razumevanju CLV je treba dodatno izpostaviti štiri ključne točke (Fader, 2012):

- CLV je koncept, usmerjen v prihodnost, ki ga ne smemo enačiti s preteklo dobičkonosnostjo stranke. Pretekla dobičkonosnost je lahko indikator prihodnje dobičkonosnosti in obnašanja stranke, ni pa CLV.
- Pri izračunu CLV je treba upoštevati samo relevantne podatke za posamezno stranko.
- Izračun CLV je napoved in kot pri vsaki napovedi obstaja določena negotovost.
- Metoda izračuna CLV je odvisna od poslovnega okolja, v katerem podjetje deluje, in od tipa razmerij, ki jih ima s svojimi strankami.

Da lahko CLV uspešno uporabimo, jo je treba najprej izračunati. To je zahtevna naloga, saj na CLV vpliva veliko dejavnikov, prikazanih na Sliki 2. Različni avtorji so poizkušali te dejavnike vključiti v svoje modele izračuna CLV, in sicer v okviru razpoložljivih podatkov in konteksta poslovnega odnosa med stranko in podjetjem. Pri izbiri ustreznega modela izračuna CLV je ključno prav razumevanje dejavnikov v okviru konteksta poslovnega odnosa, kar obravnavam v nadaljevanju.

Slika 2: Dejavniki, ki vplivajo na življenjsko vrednost stranke



Vir: D. Jain & S. S. Singh, *Measuring Customer Lifetime Value*, 2010, str. 5.

## 1.2 Kontekst življenjske vrednosti strank

Ko govorimo o kontekstu CLV, imamo v mislih kontekst poslovnega odnosa med stranko in podjetjem, v katerem so nastali podatki, ki jih bomo uporabili pri izračunu CLV. Ta kontekst igra ključno vlogo tako pri modeliranju kot pri interpretaciji in managementu CLV. Z vidika modeliranja kontekst določa podatke, ki jih uporabimo pri izračunu, z vidika managementa pa kontekst osvetljuje ključne dejavnike, ki vplivajo na management s CLV (Jain & Singh, 2010). Pomen konteksta pri izračunu CLV lahko ponazorimo z naslednjimi primeri: (a) nakup potrošnih dobrin v trgovini, (b) nakup storitev in (c) telefonska naročnina.

V primeru nakupa potrošnih dobrin (a) podjetje ne ve, kdaj je oziroma bo stranko izgubilo. Prav tako podjetje težko oceni, koliko nakupov bo stranka v prihodnosti opravila in kakšen denarni tok bo generirala. Ta negotovost z vidika managementa CLV pomeni, da se mora podjetje posebej posvečati zadržanju in prodaji strank, če želi povečati njihovo CLV. Z vidika modeliranja CLV pa takšna negotovost predstavlja dodatno kompleksnost. Po drugi strani v primeru nakupa storitev (b) in naročnin (c) podjetje načeloma predvidi, kdaj stranko izgubi. Prav tako lažje oceni prihodnje nakupe strank in generiran denarni tok. V primeru nekaterih storitev (c) tudi ve, da bo stranko izgubilo, ko se storitev opravi, ne glede na to, kako jo poskuša zadržati. Z vidika managementa CLV je torej boljše, da se podjetje osredotoča na navzkrižno prodajo (angl. *cross-sell*) ali na prodajo dražjih različic izdelka ali storitev (angl. *up-sell*). V primeru naročnin (c) pa naj se podjetje osredotoča na zadržanje strank, saj daljša naročnina pomeni višjo CLV. V obeh primerih, (b) in (c), je modeliranje CLV relativno enostavnejše, saj poznamo trajanje življenjske dobe stranke.

Kontekste CLV je mogoče klasificirati na tri načine (Jain & Singh, 2010):

- izgubljeni za vedno (angl. *lost-for-good*) in vedno na voljo (angl. *always-a-share*); v scenariju »izgubljeni za vedno« so stranke do določenega trenutka zveste podjetju, nato pa ga za vedno zapustijo oziroma se mora podjetje potruditi, da jih spet pridobi nazaj. Primer naročnin (c) iz prejšnjega odstavka je tipičen primer takšnega konteksta. V

scenariju »vedno na voljo« pa stranka kupuje pri različnih podjetjih in obdobje neaktivnosti še ne pomeni, da je za podjetje izgubljena. Nakup potrošnih dobrin (a) je tipičen primer takšnega konteksta.

- Članstvo in ne-članstvo; stranke morajo pred nakupom postati člani, v nasprotnem primeru ne morejo postati stranke podjetja. V okviru te klasifikacije je treba opozoriti na razliko med članstvom in pogodbenim odnosom med podjetjem in stranko. Članstvo je oblika pogodbenega odnosa, ni pa vsak pogodben odnos tudi članstvo.
- Pogodben in nezavezujoč; v pogodbenem kontekstu je razmerje med podjetjem in stranko urejeno s pogodbo, zato podjetje točno ve, kdaj stranko potencialno izgubi. Trajanje razmerja je direktno povezano z denarnim tokom, kar se odraža v višji CLV strank z daljšim pogodbenim razmerjem. Management strank se osredotoča predvsem na zadržanje strank. Primer naročnin (c) je tipičen primer takšnega konteksta. V nezavezujočem kontekstu je razmerje med podjetjem in stranko negotovo, zato podjetje ne ve, kdaj stranko izgubi. Zaradi te negotovosti je treba pri modeliranju CLV oceniti trajanje razmerja in generiran denarni tok. Primer nakupa potrošnih dobrin (a) je tipičen primer takšnega konteksta.

Pogodben in nezavezujoč kontekst je najbolj obravnavan kontekst v strokovni literaturi. Problem teh klasifikacij je, da so preveč široke in ne upoštevajo heterogenosti znotraj posameznih kategorij, ki lahko pomembno vplivajo na izračun CLV (Jain & Singh, 2010). Fader in Hardie (2009) opozarjata, da podjetja tem razlikam namenjajo veliko premalo pozornosti, in predlagata razširjeno klasifikacijo z dodano dimenzijo priložnosti za transakcije, kot prikazuje Tabela 2. Medtem ko je v kontekstu zveznih priložnosti za transakcije ključnega pomena, kdaj bo stranka opravila transakcijo, je v diskretnem kontekstu vprašanje, ali bo stranka v obdobju priložnosti za transakcije le-to opravila ali ne.

*Tabela 2: Kontekst CLV z vidika dveh dejavnikov*

Priložnost za transakcije	Tip odnosa s stranko	
	Nezavezujoč	Pogodben
Zvezna – transakcija se lahko izvede kadarkoli	Nakup potrošnega blaga Obisk hotelov	Kreditne kartice
Diskretna – transakcija se lahko izvede samo ob določenih trenutkih	Donacije Obisk dogodkov	Zavarovalna polica Naročnina na revije

*Vir: P. Fader & B. Hardie, Probability Models for Customer-Base Analysis, 2009, str. 63, slika 2.*

Takšna razširjena klasifikacija upošteva razlike v življenjski dobi in čas med nakupi, ne vključuje pa tretjega ključnega dejavnika – zneska nakupa. Znesek nakupa posamezne nakupne priložnosti je lahko fiksni ali variabilni. Fiksne zneske je enostavno napovedati in so med posameznimi nakupi enaki, medtem ko je lahko v kategoriji variabilnih zneskov

vrednost nakupa kakršnakoli. V kontekstu fiksnih zneskov se podjetje primarno osredotoča na čas med nakupi in življenjsko dobo strank, v kontekstu variabilnih zneskov pa se mora podjetje dodatno osredotočati tudi na znesek nakupa, če želi povečati CLV. Tabela 3 prikazuje takšno razširjeno klasifikacijo. Z vidika zahtevnosti modeliranja je ta največja v kontekstu variabilnih, zveznih transakcij v nezavezujočih odnosih s stranko. V tem kontekstu namreč ne poznamo niti trenutka, ko stranko izgubimo, ne frekvence nakupov, niti njihove vrednosti (Jain & Singh, 2010).

Kategorizacija kontekstov nam pomaga razumeti in izbrati ustrezne modele CLV za posamezne situacije. In čeprav na CLV vpliva veliko dejavnikov, sem skozi zgornje odstavke ugotovil, da so ključni predvsem trije:

- življenjska doba,
- čas med nakupi,
- znesek nakupa.

*Tabela 3: Kontekst CLV z vidika treh dejavnikov*

Priložnost za transakcije	Tip odnosa s stranko			
	Nezavezujoč		Pogodben	
	Znesek nakupa			
	Fiksen	Variabilen	Fiksen	Variabilen
Zvezna – transakcija se lahko izvede kadarkoli	Najem filmov na spletu	Nakup potrošnega blaga	Študentski boni	Kreditne kartice
Diskretna – transakcija se lahko izvede samo ob določenih trenutkih	Obisk dogodkov	Donacije	Zavarovalna polica Naročnina na revije	Telefonska naročnina

*Vir: D. Jain & S. S. Singh, Measuring Customer Lifetime Value, 2010, str. 16, tabela 2.*

### 1.3 Modeli življenjske vrednosti strank

Kot odgovor na vrsto kompleksnih izzivov, ki izhajajo iz različnih kontekstov poslovnega odnosa, v strokovni literaturi najdemo veliko predlogov različnih modelov za izračun CLV. Jain in Singh (2002) modele klasificirata v tri skupine:

- modeli za izračun CLV,
- modeli za analizo baze strank,
- normativni modeli CLV.

Gupta in Lehmann (2006) modele razdelita v šest skupin:

- modeli RFM,
- verjetnostni modeli (angl. *probability models*),
- ekonometrični modeli (angl. *econometric models*),
- vztrajnostni modeli (angl. *persistence models*),
- računalniški modeli (angl. *computer science models*),
- difuzijski modeli/modeli rasti (angl. *diffusion/growth models*).

V nadaljevanju na kratko obravnavam značilnosti posameznih modelov, kot jih vidijo omenjeni avtorji.

### 1.3.1 Modeli za izračun življenjske vrednosti strank

Ti modeli so namenjeni specifično izračunu CLV oziroma izbiri optimalne metode alokacije razpoložljivih sredstev za optimizacijo CLV. Primerni so kot podlaga za strateško in taktično odločanje. Obstajajo štiri različni tipi takih modelov, njihovih variacij pa je lahko mnogo več, odvisno od konteksta.

**Osnovni strukturni model** CLV je najenostavnejši tip modelov, ki sta jih obravnavala Berger in Nasr (1998) in jih lahko zapišemo z že predstavljeno enačbo:

$$CLV = \sum_{i=1}^n \frac{(R_i - C_i)}{(1 + d)^{i-0.5}}, \quad (2)$$

kjer  $i$  predstavlja obdobje transakcije;  $R_i$  prihodke v obdobju;  $C_i$  stroške v obdobju;  $d$  diskontno stopnjo;  $n$  število obdobj. Matematično gre za izračun neto sedanje vrednosti prihodnjih denarnih tokov strank. Model je lahko razumljiv in enostaven za uporabo, vendar ima precej pomanjkljivosti. Največja je zagotovo dejstvo, da ne upošteva stohastične narave obnašanja strank pri nakupih, ampak predpostavlja konstantno periodo prihodnjih denarnih tokov.

**Model migracije strank** (angl. *customer migration model*) je razvil Dwyer (1997), ki je predlagal razdelitev strank v dve skupini: izgubljeni za vedno (angl. *lost-for-good*) in vedno na voljo (angl. *always-a-share*). Za prvo skupino je uporabil kar nekoliko prilagojeni osnovni strukturni model CLV, za drugo pa je uvedel nov model migracije strank, ki temelji na času zadnjega nakupa (angl. *purchase recency*). S tem je razširil osnovni model in poskusil odgovoriti na njegovo ključno pomanjkljivost, modeliranje verjetnostne narave obnašanja strank pri nakupih. Model je še vedno dovolj enostaven in uporaben, vendar samo v določenih situacijah, kot so na primer naročniška razmerja, saj predpostavlja konstantno periodo prihodnjih denarnih tokov.



**Model optimalne alokacije sredstev** (angl. *optimal resource allocation model*) sta predlagala Blattberg in Deighton (1996) in se ukvarja z iskanjem optimalnega razmerja med izdatki za pridobitev strank na eni in izdatki za ohranjanje strank na drugi strani, in sicer s ciljem maksimizacije CLV. Model uporabi CLV kot osnovo za odločanje o optimalni razporeditvi sredstev za trženje in torej ni namenjen izključno izračunu CLV. Prav tako je model bolj kompleksen, saj za izračun CLV upošteva tudi izdatke za pridobitev in ohranjanje strank. Še vedno pa ima enake pomanjkljivosti kot prej predstavljena modela, ker ravno tako predpostavlja konstantno periodo prihodnjih denarnih tokov.

**Model odnosa s strankami** (angl. *customer relationship model*) sta razvila Pfeifer in Carraway (2000) in predlagala, da so za modeliranje odnosov s strankami primerni modeli Markovske verige (angl. *Markov chain models*). Ti modeli naj bi odpravili pomanjkljivosti, izpostavljene v zgornjih modelih. Ker gre v osnovi za verjetnostne modele, so primerni za modeliranje stohastičnosti v odnosih s strankami. Od vseh modelov v tej skupini so modeli Markov najbolj fleksibilni, vendar tudi najbolj zahtevni za izračun in v osnovi še vedno predpostavljajo konstantno periodo prihodnjih denarnih tokov.

### 1.3.2 Modeli za analizo baze strank

Rezultat teh modelov je verjetnost nakupov v prihodnjih obdobjih, kar je eden od parametrov izračuna CLV. Verjetnost je izračunana na osnovi vseh preteklih nakupov celotne baze strank ob upoštevanju stohastičnega vedenja posameznih strank. Tipičen primer takšnega modela je model Pareto/NBD.

**Model Pareto/NBD**, ki so ga predlagali Schmittlein, Morrison in Colombo (1987), izračuna verjetnost, da je stranka še vedno aktivna. S pomočjo tega modela lahko podjetje ugotovi, koliko aktivnih strank ima in koliko nakupov lahko od njih pričakuje v prihodnjih obdobjih, posamično in kumulativno. Model se lahko uporabi v kontekstu poslovnega odnosa, pri čemer podjetje ne ve, kdaj bo stranko izgubilo, stranka lahko nakupuje kadar koli in tudi postane neaktivna, kadar želi. Glavni problem teh modelov je, da so v praksi matematično izredno zahtevni za izračun. Prav tako njihova uporabnost zbledi, kadar imamo veliko število strank, ki opravljajo majhno število nakupov. Treba je biti izredno pazljiv, kaj štejemo kot prvi nakup in kolikšen del zgodovine upoštevamo. Vse te odločitve namreč pomembno vplivajo na vhodne podatke modela in lahko izredno izkrivijo rezultate.

**Modeli RFM** se v neposrednem trženju uporabljajo že več kot 30 let. Z njihovo pomočjo stranke razdelimo v skupine glede na 3 spremenljivke: čas od zadnjega nakupa (angl. *recency* – *R*), število nakupov (angl. *frequency* – *F*) in vrednost nakupa (angl. *monetary value* – *M*). Največkrat se stranke razvrsti v pet skupin znotraj vsake posamezne spremenljivke, s čimer dobimo 125 (5 x 5 x 5) skupin. Rezultate modela RFM uporabimo kot vhodni parameter pri izračunu CLV, saj poizkušajo napovedati prihodnje nakupe strank,

pri čemer že samo ime odraža tudi vpliv posameznih spremenljivk:  $R > F > M$ . Čeprav se v praksi pogosto uporabljajo, imajo precej pomanjkljivosti (Fader et al., 2005):

- Prihodnje nakupe so sposobni napovedati samo za naslednje obdobje, ne pa tudi za vsa ostala prihodnja obdobja, kar je ključno za izračun CLV.
- Spremenljivke RFM niso popoln pokazatelj dejanskega obnašanja strank.
- Ne upoštevajo možnosti, da je lahko preteklo obnašanje strank posledica specifičnih trženjskih aktivnosti podjetja v preteklosti.

**Verjetnostni modeli** obnašanje strank interpretirajo kot izid stohastičnega procesa, ki je posledica latentnih vedenjskih nagnjenj, ki se med posameznimi strankami razlikujejo. Fokus modeliranja je povedati strukturirano zgodbo, ki opisuje in napoveduje opazovano obnašanje, ne pa osredotočanje na razlaganje razlik v tem obnašanju. Te razlike sprejemamo in modeliramo kot katero od verjetnostnih porazdelitev. Eden prvih verjetnostnih modelov je bil že omenjeni model Pareto/NBD, kasneje pa njegove izpeljanke. Med verjetnostne modele lahko delno prištejemo tudi modele Markovskih verig.

**Ekonometrični modeli** temeljijo na filozofiji verjetnostnih modelov, a se osredotočajo predvsem na modeliranje pridobitve, ohranitve in razširitve (navzkrižna prodaja ali marže) strank. Te elemente na koncu združijo, da lahko izračunajo CLV.

Tudi **vztrajnostni modeli** se osredotočajo na iste komponente kot ekonometrični modeli. Glavna prednost vztrajnostnih modelov je, da omogočajo modeliranje teh komponent na dolgi rok, kot del enotnega dinamičnega sistema. Podjetju omogočajo analiziranje, kako spremembe v eni od spremenljivk (npr. nova kampanja za pridobitev strank) vplivajo na ostale in na celotni CLV skozi čas.

**Računalniški modeli** temeljijo na podatkovnem rudarjenju (angl. *data mining*), strojnem učenju (angl. *machine learning*) in podobnih modelih napovedne analitike. Sem štejemo na primer odločitvena drevesa (angl. *decision trees*), nevronske mreže (angl. *neural networks*) in metodo podpornih vektorjev (angl. *support vector machines*). S pomočjo tehnik strojnega učenja je možno napovedi še dodatno izboljšati.

### 1.3.3 Normativni modeli življenjske vrednosti strank

Normativni modeli CLV se uporabljajo za razumevanje in interpretiranje CLV, njihov namen pa je podpora pri odločanju. Z njihovo pomočjo lahko na primer ocenimo veljavnost prepričanj, da so dolgoročne stranke bolj profitabilne, kar je bilo dolgo sprejeto dejstvo. Avtorji zadnjih raziskav s tega področja o takšnem prepričanju dvomijo (Reinartz & Kumar, 2000).

**Model CE** so Blattberg, Getz in Thomas (2001) predlagali kot osnovni model vsote celotne vrednosti strank podjetja. Z njegovo pomočjo lahko podjetje analizira učinke različnih trženjskih elementov in aktivnosti na dolgoročno vrednost strank podjetja ter izbere tiste, ki to vrednost povečujejo.

**Dinamični cenovni model** na osnovi CLV sta predlagala Blattberg in Deighton (1996), in sicer naj podjetje za maksimiranje CLV uporabi dinamično cenovno strategijo. Rezultat modela sta optimalna cena za pridobitev in za obdržanje stranke. Če so na primer na trgu potencialne stranke, za katere ugotovimo, da imajo nizko dolgoročno CLV in so cenovno občutljive, je bolje zvišati ceno za pridobitev stranke.

Medtem ko se ostali modeli osredotočajo bolj na operativno uporabnost CLV, na primer na izbiro posameznih strank, segmentacijo in management trženjskih kampanj, se **difuzijski modeli** posvečajo širši, strateški vrednosti CLV. Vsota CLV vseh trenutnih in prihodnjih strank predstavlja pomembno metriko vrednotenja podjetja. Ti modeli se tako posebej osredotočajo na napovedovanje pridobitve strank v prihodnosti.

## 1.4 Izzivi modeliranja življenjske vrednosti strank

Kot že rečeno, je modeliranje CLV kompleksna naloga predvsem zaradi velikega števila dejavnikov, ki lahko vplivajo na posamezne komponente izračuna CLV. Posamezni modeli se največkrat osredotočajo na določeno komponento v določenem kontekstu. Ne poznamo pa modela, ki bi bil tako splošen, da bi bil primeren v vseh situacijah in bi omogočal celovit pristop in izračun ob upoštevanju vseh spremenljivk. V splošnem lahko izpostavimo naslednje pomanjkljivosti, ki v določeni meri veljajo za vse vrste modelov (Jain & Singh, 2002; Gupta et al., 2006):

- osredotočanje na specifičen kontekst, dejavnost ali podatke pomeni, da je treba biti pri uporabi modela izredno previden in se prepričati, da so predpostavke v želenem kontekstu uporabe enake.
- Veliko modelov ni podprtih z empiričnimi raziskavi iz prakse, zato je njihove rezultate težko oceniti.
- Večina modelov ne vključuje heterogenosti, ki je posledica spremenljivk, kot so na primer razlike v posameznih izdelkih ali demografske značilnosti posameznih strank, kar lahko pomembno vpliva na napovedi.
- Modeli ne upoštevajo vpliva posameznih trženjskih aktivnosti na nakupe strank ali na primer vpliva različnih programov zvestobe. Prav tako modeli ne upoštevajo učinka mreženja, ki je v današnjem povezanem svetu izredno prisotno.
- Za izračun se uporablja pretekle podatke o transakcijah, ki pa nam ne povedo nič o motivih, ki so do teh transakcij privedli. Prav tako ne upoštevajo podatkov o konkurenci (na primer koliko nakupov določena stranka opravi pri konkurentu in koliko pri nas).

- Večina modelov se osredotoča na CLV posameznih strank, kar lahko podjetje zavede do te mere, da zanemari stranke, ki se ne zdijo dobičkonosne, a v praksi še vedno predstavljajo 80 % vseh strank.
- CLV je ocena in tudi temelji na ocenah stroškov (pridobitve, ohranitve, servisiranja itd.) za izračun. V praksi podjetja redko spremljajo stroške na nivoju posameznih strank, zato je priprava takšnih ocen izredno težavna.
- Vsako podjetje je specifično, zato je treba razumeti, da imajo teoretični modeli svoje meje in da je treba specifične lastnosti podjetja modelirati posebej.

Podjetje mora pred odločitvijo o izbiri ustreznega modela razumeti vse našteje pomanjkljivosti. CLV kot merilo uspešnosti je uporabna samo, če jo lahko razumno izračunamo s podatki, ki jih imamo na voljo. Dodajanje spremenljivk bi verjetno izboljšalo napovedno sposobnost modela, a ključno vprašanje je, ali takšna izboljšava upraviči dodatno kompleksnost. Nekateri avtorji se s tem ne strinjajo in trdijo, da bi dodatne spremenljivke samo zameglile zgodbo in razumljivost ter da dodatna kompleksnost ne upraviči minimalnih izboljšav v napovedi (Fader et al., 2006). Dodatno so predložili konkretne empirične dokaze, ki na mnogih primerih potrjujejo ustreznost verjetnostnih modelov, kakršni so Pareto/NBD in njegove izpeljanke, zato je uporaba teh modelov, v pravilnem kontekstu, za podjetja priporočljiva (Jain & Singh, 2010).

V nadaljevanju se osredotočam na verjetnostne modele, saj so ti najprimernejši za obravnavo primera v empiričnem delu magistrske naloge.

## **2 VERJETNOSTNI NAPOVEDNI MODELI ZA ANALIZO BAZE STRANK**

Verjetnostni modeli temeljijo na zavedanju, da je nakupno vedenje strank kompleksno in da podatki, ki jih podjetje o tem vedenju zbira (na primer čas in znesek posameznih nakupov, frekvenca nakupov ipd.), odražajo ogromno število faktorjev. Četudi bi bilo obnašanje strank popolnoma deterministično, bi bilo za podjetje praktično nemogoče zajeti vse spremenljivke, ki vplivajo na nakupno vedenje posamezne stranke. Nakupno vedenje ti modeli zato izražajo v obliki verjetnosti, naključij oziroma stohastičnosti, ki odražajo pomanjkljivo vedenje in podatke o teh spremenljivkah (Fader, Hardie & Sen, 2014).

Namesto modeliranja vplivov vseh teh spremenljivk na nakupno vedenje posameznika verjetnostni modeli upoštevajo stohastičnost in nakupno vedenje obravnavajo kot izid enega verjetnostnega procesa ali več le-teh. To ne pomeni, da nakupna vedenja strank jemljemo kot naključna, temveč gre zgolj za odražanje naše negotovosti glede spremenljivk, ki vplivajo na nakupno vedenje posameznika. Verjetnostni proces zajema vpliv vseh dejavnikov, ki v izbranem modelu niso specifično določeni. Takšni verjetnostni modeli obnašanja strank imajo v tem kontekstu dve komponenti. Prvič, verjetnost, da je stranka v prihodnjem obdobju še aktivna, sledi verjetnostni porazdelitvi (v bolj kompleksnih modelih

je lahko uporabljenih več različnih verjetnostnih porazdelitev), katere parametri odražajo vedenjska nagnjenja posamezne stranke (angl. *behavioral propensities*). In drugič, razlike v vedenjskih nagnjenjih posameznih strank so zajete v dodatni verjetnostni porazdelitvi. Ti komponenti skupaj tvorita mešano porazdelitev, ki označuje obnašanje naključno izbrane stranke (Fader & Hardie, 2015).

Preteklo nakupno vedenje stranke predstavlja osnovo za napovedovanje prihodnjega nakupnega vedenja te stranke, ni pa preteklo vedenje zagotovilo, da bo stranka tako nakupovala tudi v prihodnosti. Če je na primer v preteklosti nakupovala dvakrat na leto, to še ne pomeni, da bo tako tudi v prihodnosti. Pretekle in potencialne prihodnje transakcije stranke lahko zapišemo kot funkcijo vedenjskih nagnjenj, ki jih označujemo kot  $\theta$ . Funkcijo preteklih nakupov zapišemo kot preteklost =  $f(\theta)$ , funkcijo prihodnjih nakupov pa kot prihodnost =  $f_I(\theta)$ . Začetni korak verjetnostnega modeliranja je definicija matematičnega modela, kjer je opazovano obnašanje stranke funkcija njenih latentnih vedenjskih nagnjenj, torej preteklost =  $f(\theta)$ . To storimo z izbiro ene (lahko tudi kombinacije) enostavne verjetnostne porazdelitve (Poisson, binomska, eksponentna), ki najbolje odraža opazovano obnašanje. Pri tem nas, kot že rečeno, neposredno ne zanimajo značilnosti vedenjskih nagnjenj (torej  $\theta$ ). V naslednjem koraku zato predpostavljamo, kako variirajo razlike vedenjskih nagnjenj v naši bazi strank. To ponovno dosežemo z izbiro ustrezne verjetnostne porazdelitve (kriterij ustreznosti niha med fleksibilnostjo in matematično uporabnostjo), ki odraža heterogenost  $\theta$ . Ti dve združeni porazdelitvi tvorita mešano porazdelitev, ki karakterizira obnašanje naključno izbrane stranke (Fader & Hardie, 2009).

Različni verjetnostni modeli uporabljajo različne verjetnostne porazdelitve, s pomočjo katerih razlagajo obnašanje strank, kar pomembno vpliva na rezultate modela. Značilnost poslovnega odnosa med podjetjem in stranko je ključna pri razlagi in izbiri ustreznih verjetnostnih porazdelitev. Izbrano podjetje deluje v kontekstu nezavezujočih odnosov s stranko, kjer se lahko transakcije izvedejo kadar koli. V tem kontekstu je treba primarno odgovoriti na dve ključni vprašanji, in sicer:

- ali bo stranka v obdobju aktivna?
- Koliko nakupov bo stranka v obdobju izvedla?

Standardni verjetnostni model za modeliranje ponovnih nakupov je model NBD.

## 2.1 Model NBD

Model negativne binomske porazdelitve (angl. *negative binomial distribution*, v nadaljevanju NBD) je leta 1959 predlagal Ehrenberg (Ehrenberg, 1959). Ehrenberg je nakupe posameznih strank modeliral kot Poissonov proces, razlike med posameznimi

strankami pa kot porazdelitev gama. Takšna mešana Poisson-gama porazdelitev je znana tudi kot negativna binomska porazdelitev.

Tabela 4: Predpostavke modela NBD

Predpostavka 1	Število transakcij sledi Poissonovemu procesu s parametrom $\lambda$ .
Predpostavka 2	Heterogenost v $\lambda$ med strankami sledi porazdelitvi gama, kjer je $r$ parameter oblike in $\alpha$ parameter merila.

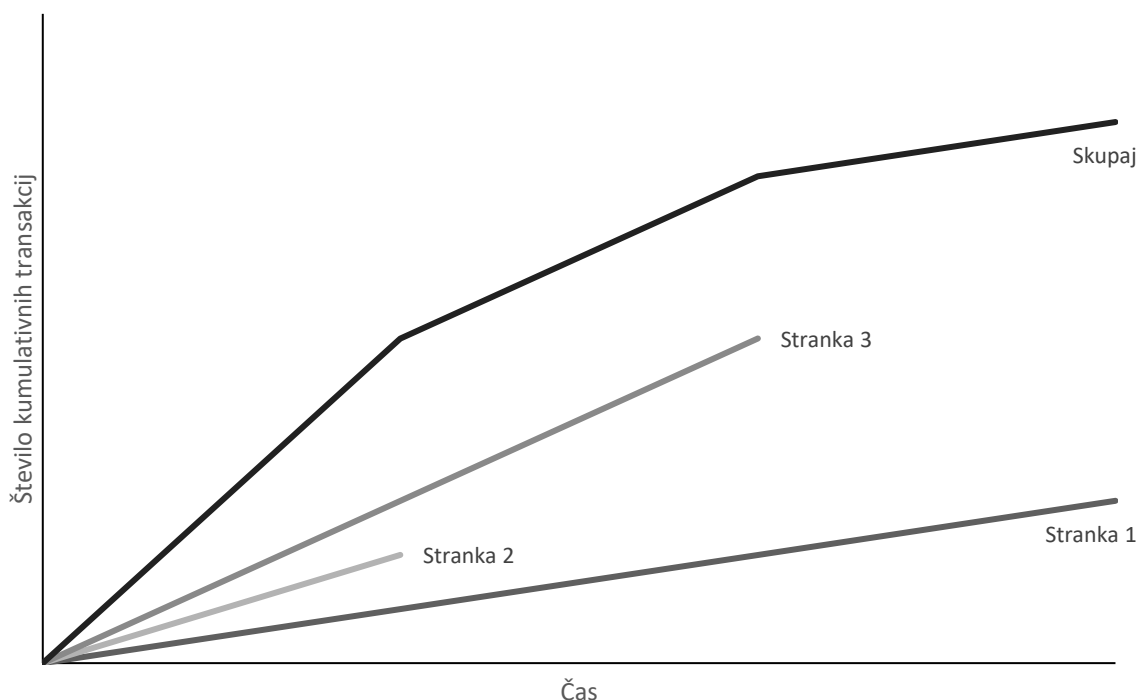
Predpostavke modela NBD prikazuje Tabela 4. Model predpostavlja, da posamezna stranka v opazovanem časovnem obdobju nakupuje naključno, z znano pogostostjo in neodvisno od časa, ki je potekel od zadnjega dogodka. Nadalje model predpostavlja, da pogostost teh nakupov med posameznimi strankami variira. Iz teh predpostavk izhaja, da se od strank pričakuje konstantna pogostost nakupov. Model je v desetletjih prestal in dokazal svojo uporabnost, hkrati pa so v tem času odkrili tudi pomanjkljivosti (Schmittlein, Bemmaor & Morrison, 1985).

Upoštevajoč Poissonovo porazdelitev števila transakcij (predpostavka 1), je mogoče dokazati, da čas med dvema transakcijama sledi eksponentni porazdelitvi, ki pa je samo posebna oblika porazdelitve gama, kjer je parameter oblike  $r = 1$ . Pomembna lastnost eksponentno porazdeljenih naključnih spremenljivk je, da so brez spomina (angl. *memoryless*). To pomeni, da nobena informacija o času prejšnjega nakupa ne spremeni verjetnosti prihodnjega nakupa. Takšna lastnost je matematično koristna, ko govorimo o nakupih strank, pa implicira, da se stranke odločajo za nakupe ne glede na svoje nakupe v preteklosti. Še več, takšna brezspominskost celo implicira, da je najverjetnejši čas za naslednji nakup takoj po tem, ko nakup opravimo. Model NDB torej ne prepozna aktivnosti stranke (oziroma se vse stranke podjetja štejejo kot aktivne stranke). Ehrenberg se je teh pomanjkljivosti delno zavedal in kot rešitev postavil pogoj, da obdobje opazovanja ni prekratko (Platzer, 2008).

## 2.2 Model Pareto/NBD

Konstantna pogostost nakupov, kot jo predpostavlja osnovni model NBD, bi pomenila, da se število kumulativnih nakupov naših strank linearno povečuje. V realnosti pa to ne velja, saj se rast števila kumulativnih nakupov s časom največkrat upočasnjuje. Za to upočasnitev lahko obstaja vrsta razlogov, ki so jih Schmittlein, Morrison in Colombo (1987) opisali kot »kupuj do smrti« (angl. *Buy 'Til You Die*). Fader in Hardie (2009) to ilustrirata, kot je prikazano na Sliki 3. Namesto naključnih nakupov z znano pogostostjo, kot to predpostavlja model NBD, posamezne stranke nakupujejo z različnimi stopnjami pogostosti.

Slika 3: Tipično gibanje števila kumulativnih transakcij v scenariju »kupuj do smrti«



Vir: P. Fader & B. Hardie, *Probability Models for Customer-Base Analysis*, 2009, str. 63, slika 5.

V različnih časovnih obdobjih posamezne stranke nato prenehajo z nakupi in postanejo neaktivne ali »umrejo«. Do te neaktivnosti oziroma »smrti« lahko pride zaradi različnih razlogov, na primer sprememb v potrebah stranke, osebnih finančnih okoliščin, slabih izkušenj s prejšnjimi nakupi ali celo resnične smrti. Z vidika modeliranja nas dejanski vzrok ne zanima. Pomembno je samo, da je ta neaktivnost v model zajeta in upoštevana. To pa je izredno zahtevna naloga, saj v kontekstu nezavezujočih odnosov s stranko ta podjetja ne obvesti, da bo postala neaktivna. Dodatno kompleksnost predstavlja dejstvo, da lahko vsaka stranka postane neaktivna samo enkrat in še takrat podjetje ne more preveriti oziroma zagotoviti, da je od tistega trenutka dalje ta stranka res neaktivna.

Vsako razmerje med stranko in podjetjem ima tako dve fazi: stranka je najprej neko nedoločljivo obdobje aktivna, nato postane neaktivna. Ko je stranka aktivna, nakupuje tako, kot to predvideva osnovni model NBD. Obdobje aktivnosti pa se obravnava kot naključno in je karakterizirano kot eksponentna porazdelitev. Heterogenost med strankami se ponovno karakterizira kot porazdelitev gama. Gama porazdelitev eksponentov je znana kot Paretova porazdelitev, zato ta model imenujemo Pareto/NBD (Fader & Hardie, 2009). Predpostavke modela so prikazane v Tabeli 5.

Tabela 5: Predpostavke modela Pareto/NBD

Predpostavka 1	V obdobju aktivnosti število transakcij, ki jih opravi stranka, sledi Poissonovemu procesu s stopnjo $\lambda$ .
Predpostavka 2	Heterogenost v $\lambda$ med strankami sledi porazdelitvi gama, kjer je $r$ parameter oblike in $\alpha$ parameter merila.
Predpostavka 3	Vsaka stranka ima »življenjsko dobo«, ki traja $\tau$ . Verjetnost, da stranka postane neaktivna, sledi eksponentni porazdelitvi s parametrom stopnje $\mu$ in se lahko zgodi kadarkoli, neodvisno od transakcij.
Predpostavka 4	Heterogenost v $\mu$ med strankami sledi porazdelitvi gama, kjer je $s$ parameter oblike in $\beta$ parameter merila.
Predpostavka 5	Parametra $\lambda$ in $\mu$ variirata neodvisno med strankami.

Model Pareto/NBD nam na eleganten način omogoča izračun verjetnosti, da je stranka aktivna, in izračun števila pričakovanih nakupov v opazovanem obdobju. Za takšen izračun model zahteva samo dva podatka o preteklih nakupih posamezne stranke, in sicer:

- koliko transakcij je stranka izvedla v opazovanem preteklem obdobju (angl. *frequency*),
- čas zadnje transakcije v opazovanem preteklem obdobju (angl. *recency*).

To lahko zapišemo kot  $X = x, t_x, T$ , kjer je  $x$  število transakcij v obdobju  $(0, T]$  in  $t_x$  ( $0 < t_x < T$ ) čas od zadnjega nakupa. Pričakovano število transakcij stranke v obdobju  $T$  lahko zapišemo kot:

$$E[X | r, \alpha, s, \beta, T] = \frac{r\beta}{\alpha(s-1)} \left[ 1 - \left( \frac{\beta}{\beta+T} \right)^{s-1} \right]. \quad (3)$$

Verjetnost, da je stranka, ki je opravila  $X$  transakcij, aktivna v obdobju  $(0, T)$  od zadnje transakcije v točki  $t$ , pa zapišemo kot (ko je  $\alpha > \beta$ ):

$$P(\text{aktivna} | r, s, \alpha, \beta, X=x, t, T) = \left\{ 1 + \frac{s}{r+x+s} \int \left( \frac{\alpha+T}{\alpha+t} \right)^{r+x} \left( \frac{\beta+T}{\alpha+t} \right)^s F(a_1, b_1; c_1; z_1(t)) \right\} \left\{ - \left( \frac{\beta+T}{\alpha+T} \right)^s F(a_1, b_1; c_1; z_1(T)) \right\}, \quad (4)$$

kjer je  $a_1 = r + x + s$ ;  $b_1 = s + 1$ ;  $c_1 = r + x + s + 1$ ;  $z_1(y) = (\alpha - \beta)/(\alpha + y)$ .

Čeprav je vsebinska razlaga modela enostavna in za izračun ne potrebujemo veliko podatkov, je verjetnostna funkcija modela izredno kompleksna in zahteva številna ocenjevanja Gaussove hipergeometrične funkcije ( $F(\cdot)$  v enačbi 4). Dodatno je natančnost nekaterih numeričnih ocen, ki jih uporabimo za ocenjevanje največjega verjetja (angl. *maximum likelihood estimation*, v nadaljevanju MLE) funkcije, izredno variabilna med



parametri. Vse to povzroča težave pri numeričnem iskanju optimuma in zahteva veliko virov pri izvajanju teh matematičnih operacij (Fader, Hardie & Lee, 2005a).

## 2.3 Model BG/NBD

Čeprav je bil model Pareto/NBD med teoretiki izredno priljubljen, je bil v praksi redko uporabljen. To gre pripisati predvsem matematični zahtevnosti in posledično potrebi po veliki procesorski moči za izvedbo potrebnih matematičnih operacij. Fader et al. (2005a) zato razvijejo in predlagajo prilagojen model, ki ga poimenujejo BG/NBD. Predpostavke modela prikazuje Tabela 6.

*Tabela 6: Predpostavke modela BG/NBD*

Predpostavka 1	V obdobju aktivnosti število transakcij, ki jih opravi stranka, sledi Poissonovemu procesu s stopnjo $\lambda$ .
Predpostavka 2	Heterogenost v $\lambda$ med strankami sledi porazdelitvi gama, kjer je $r$ parameter oblike in $\alpha$ parameter merila.
Predpostavka 3	Po vsakem strankinem nakupu obstaja konstanta verjetnost $p$ , da stranka postane neaktivna.
Predpostavka 4	Heterogenost v $p$ med strankami sledi porazdelitvi beta s parametroma $a$ in $b$ .
Predpostavka 5	Parametra $\lambda$ in $p$ variirata neodvisno med strankami.

Za razliko od modela Pareto/NBD, za katerega velja, da lahko stranka postane neaktivna kadarkoli, model BG/NBD predpostavlja (predpostavka 3), da lahko stranka postane neaktivna samo takoj zatem, ko opravi zadnji nakup. To avtorji utemeljijo s pojasnilom, da zadovoljstvo stranke po opravljenem nakupu diktira njeno nadaljnje obnašanje. Če stranka z izdelkom ali storitvijo ni zadovoljna, bo zelo verjetno odšla drugam. Predpostavka 3 tudi pomeni, da je verjetnost neaktivnosti karakterizirana kot geometrična porazdelitev. Skupaj z beta porazdelitvijo, ki karakterizira heterogenost v verjetnostih neaktivnosti med strankami, dobimo mešano beta-geometrično porazdelitev, ki daje modelu ime.

Pričakovano število transakcij stranke v obdobju  $t$  lahko zapišemo kot:

$$E(Y(t)|X=x, t_x, T, r, \alpha, a, b) = \frac{\frac{a+b+x-1}{a-1} \int_0^1 \left(1 - \left(\frac{\alpha+T}{\alpha+T+t}\right)^{r+x} {}_2F_1\left(r+x, b+x; a+b+x-1; \frac{t}{\alpha+T+t}\right)\right)}{1 + \delta_{x>0} \frac{a}{b+x-1} \left(\frac{\alpha+T}{\alpha+t_x}\right)^{r+x}}, \quad (5)$$

kjer je  ${}_2F_1(\cdot)$  Gaussova hipergeometrična funkcija, ki pa jo je treba oceniti samo enkrat, kar izredno poenostavi zahtevnost in potrebne resurse za izvedbo izračuna (Fader et al., 2005a).

## 2.4 Ostale izpeljanke modelov »kupuj do smrti«

Tako model Pareto/NBD kot model BG/NBD sta zanimiva, saj podjetje za napoved prihodnjih nakupov potrebuje samo dva enostavna podatka o vsaki stranki. Tudi izvedba zahtevnih matematičnih operacij, predvsem v modelu Pareto/NBD, z vidika razpoložljivih računskih virov danes ni več tako problematična. Vendar pa enostavnost v teh predpostavkah neposredno vpliva na rezultate modelov in njihovo interpretacijo. Fader et al. (2005a) sami opozarjajo na nekatere pomanjkljivosti, ki so posledica teh poenostavitev:

- pri uporabi modelov je treba stranke segmentirati glede na obdobje prvega nakupa in prodajnega kanala.
- Parametre modela za te posamezne skupine strank je treba oceniti ločeno. Če res želimo uporabiti enotne parametre za različne skupine, se moramo prepričati, da so te skupine primerljive.
- V modelih niso eksplicitno upoštevane različne trženjske aktivnosti, ki lahko pomembno vplivajo na nakupno vedenje posameznikov. Modeli tako predpostavljajo, da bo učinek trženjskih aktivnosti v prihodnosti enak, kot je bil v preteklosti.
- Čeprav je možno modele razširiti z različnimi spremenljivkami (na primer trženjske aktivnosti, demografski podatki ipd.), Fader et al. (2005) zagovarjajo enostavne modele, ki jih je možno enostavno razložiti in izračunati. Izboljšanje napovedne sposobnosti z vključitvijo dodatnih spremenljivk po njihovem mnenju v večini primerov ne upraviči dodatne kompleksnosti, ki jo prinaša vpeljava novih spremenljivk.

Oba modela imata tudi določene predpostavke, ki v določenih situacijah niso primerne. Če na primer obnašanje naših strank v realnosti ne sledi porazdelitvi, ki je uporabljena v modelu, bo napovedna sposobnost modela močno vprašljiva. V realnosti pa zaradi nezavezujočega konteksta odnosov s stranko niti nikdar ne vemo, katera porazdelitev je dejansko najbolj primerna (Singh, Borle & Jain, 2009).

Zaradi opisanih pomanjkljivosti so mnogi avtorji osnovna modela razširili, bodisi s spremenjenimi predpostavkami, specifičnim kontekstom ali dodatnimi spremenljivkami. V veliki večini je tem razširitvam skupno to, da so modeli bolj zahtevni za uporabo in zahtevajo več vhodnih podatkov. Prav tako mnoge (še) nimajo razvitih enačb, s katerimi lahko ugotovimo, ali je stranka še aktivna, in napovemo, koliko nakupov bo opravila. To pa je ključna informacija za analiziranje in napovedovanje prihodnjih nakupov strank in izračuna CLV (Fader et al., 2005a). V nadaljevanju zato predstavljam le modele, ki so del R paketa BTYD in jih je mogoče enakovredno primerjati z osnovnima modeloma. Tabela 7 prikazuje primere ustreznih modelov BTYD za posamezne kontekste poslovnih odnosov.

Tabela 7: Matrika poslovnih odnosov s primeri ustreznih modelov

Priložnost za transakcije	Tip odnosa s stranko	
	Nezavezujoč	Pogodben
Zvezna – transakcija se lahko izvede kadarkoli	Nakup potrošnega blaga – Pareto/NBD*	Telefonska naročnina – exponential-gamma**
Diskretna – transakcija se lahko izvede samo ob določenih trenutkih	Donacije – BG/BB***	Zavarovalna polica – sBG****

**Legenda:** \* Negativna binomska porazdelitev (angl. negative binomial distribution); \*\* Eksponentna gama porazdelitev (angl. exponential gamma distribution); \*\*\* Beta geometrična/beta binomska porazdelitev (angl. beta-geometric/beta-binomial distribution); \*\*\*\* Premaknjena beta geometrična porazdelitev (angl. shifted beta-geometric distribution)

Vir: D. C. Schmittlein, D. G. Morrison & R. Colombo, *Counting Your Customers: Who-Are They and What Will They Do Next?*, 1987, str. 16, tabela 3.

#### 2.4.1 Model CBG/NBD

Model CBG/NBD (angl. *central variant beta geometric*) uvaja možnost, da stranka postane neaktivna tudi v času 0 (Hoppe & Wagner, 2007). Na ta način poizkuša omiliti zahtevo modela BG/NBD, da lahko postane stranka neaktivna šele, ko opravi najmanj drugi nakup. Posledica te predpostavke je namreč to, da model BG/NBD vse stranke, ki so opravile samo en (prvi) nakup, šteje kot aktivne. To v realnosti ni nujno res, saj je povsem verjetno, da bo stranka kupila samo enkrat. Predpostavke modela prikazuje Tabela 8. Razen predpostavke 3 se ostale od modela BG/NBD ne razlikujejo. V istem času je bil predlagan tudi model MBG/CBG (Batislam, Denizel & Filiztekin, 2007), ki problem aktivnosti rešuje na enak način.

Tabela 8: Predpostavke modela CBG/NBD

Predpostavka 1	V obdobju aktivnosti število transakcij, ki jih opravi stranka, sledi Poissonovemu procesu s stopnjo $\lambda$ .
Predpostavka 2	Heterogenost v $\lambda$ med strankami sledi porazdelitvi gama, kjer je $r$ parameter oblike in $\alpha$ parameter merila.
Predpostavka 3	Po prvem in vsakem naslednjem strankinem nakupu obstaja konstantna verjetnost $p$ , da stranka postane neaktivna.
Predpostavka 4	Heterogenost v $p$ med strankami sledi porazdelitvi beta s parametroma $a$ in $b$ .
Predpostavka 5	Parametra $\lambda$ in $p$ variirata neodvisno med strankami.

Verjetnost, da je stranka aktivna, lahko zapišemo kot:

$$P(\text{aktivna} | x, t_x, T, r, \alpha, a, b) = \frac{1}{\left\{ 1 + \frac{a}{b+x} \left( \frac{a+T}{a+t_x} \right)^{r+x} \right\}} \quad (6)$$

ki je manj kot 1 za stranke, ki niso opravile nakupov v obdobju  $(0, T]$ .

#### 2.4.2 Model CBG/CNBD-k

Model CBG/CNBD-k (angl. *condensed negative binomial distribution*) se posveča rednosti nakupov strank (Platzer, 2008). Osnovni model NBD in njegove izpeljanke predpostavljajo, da stranke v času nakupujejo naključno, vendar takšna predpostavka pod določenimi pogoji ni primerna. Če na primer podjetje prodaja izdelke trajne potrošnje, ki imajo določeno življenjsko dobo, lahko z večjo verjetnostjo pričakuje, da bo rednost nakupov stranke, ki tak izdelek kupuje, temu sledila. Predpostavke modela prikazuje Tabela 9.

Tabela 9: Predpostavke modela CBG/CNBD-k

Predpostavka 1	V obdobju aktivnosti število transakcij, ki jih opravi stranka, sledi Erlangovi porazdelitvi s parametrom merila $\lambda$ in parametrom oblike $k$ .
Predpostavka 2	Heterogenost v $\lambda$ med strankami sledi porazdelitvi gama, kjer je $r$ parameter oblike in $\alpha$ parameter merila.
Predpostavka 3	Po prvem in vsakem naslednjem strankinem nakupu obstaja konstantna verjetnost $p$ , da stranka postane neaktivna.
Predpostavka 4	Heterogenost v $p$ med strankami sledi porazdelitvi beta s parametroma $a$ in $b$ .
Predpostavka 5	Parametra $\lambda$ in $p$ variirata neodvisno med strankami.
Predpostavka 6	Opazovano obdobje vsake posamezne stranke se začne s prvo transakcijo v času $0$ .

Model CBG/CNBD-k se od ostalih modelov NBD razlikuje v predpostavki 1. Namesto naključnih nakupov, ki sledijo eksponentni porazdelitvi (Poisson), model predpostavlja, da si nakupi redno sledijo v obliki Erlangove porazdelitve. Pomembno je dejstvo, da parameter  $k$  ni določen v samem modelu, ampak ga moramo izbrati sami. Ker predpostavka 1 pomeni, da transakcije med seboj v času niso več neodvisne, je bilo treba modelu dodati predpostavko 6. Na ta način je modeliranje prihodnjih transakcij bolj natančno, saj poznamo čas prejšnje transakcije. Problem tega modela je, da avtorju zaradi matematične zahtevnosti (še) ni uspelo izpeljati formule za napovedovanje prihodnjih nakupov. Razvil je samo približek, a pokazal, da je tudi samo z uporabo tega približka napovedna sposobnost v določenih primerih boljša kot pri ostalih NBD modelih.

$$\widehat{E}(Y(T, T+t)|x, t_x, T, r, \alpha, a, b) = \frac{1}{k} * \frac{a+b+x}{a-1} * G(r+kx, b+x, b+x, a+T/\alpha, t) * P(\tau > T | x, t_x, T, r, \alpha, a, b). \quad (7)$$

### 2.4.3 Model HB Pareto/NBD

Vsi do sedaj predstavljeni modeli NBD ne dovoljujejo uporabe parametrov na ravni posamezne stranke in ne upoštevajo značilnosti posameznih strank. Abe (2009) je predlagal hierarhično Bayes (angl. *hierarchical Bayes*) razširitev modela Pareto/NBD, pri kateri so parametri ocenjeni na nivoju posamezne stranke in sledijo multivariantni logaritemsko normalni porazdelitvi (angl. *multivariate lognormal distribution*). Srednja vrednost (angl. *mean*) porazdelitve je odvisna od značilnosti posamezne stranke. Predpostavke modela prikazuje Tabela 10.

Tabela 10: Predpostavke modela HB Pareto/NBD

Predpostavka 1	V obdobju aktivnosti število transakcij, ki jih opravi stranka, sledi Poissonovemu procesu s stopnjo $\lambda$ .
Predpostavka 2	Vsaka stranka ima »življenjsko dobo«, ki traja $\tau$ . Trenutek, ko stranka postane neaktivna, sledi eksponentni porazdelitvi s parametrom stopnje $\mu$ in se lahko zgodi kadarkoli, neodvisno od transakcij.
Predpostavka 3	Heterogenost v $\lambda$ in $\mu$ med strankami sledi multivariantni logaritemsko normalni porazdelitvi

Model ohranja predpostavke o Poissonovem procesu nakupovanja, brezspominskosti procesa neaktivnosti in heterogenosti med strankami, dovoljuje pa povezavo med procesom nakupovanja in aktivnostjo. Abe (2009) navaja naslednje prednosti: enostavna razumljivost, enostavno ocenjevanje parametrov modela, nezahtevne računske operacije, fleksibilnost modela, ocenjevanje latentnih spremenljivk, ocenjevanje pravih metrik za napake, enostavne razširitve in sledenje Bayesovi paradigmi.

Slabost modela je, da ne omogoča direktnih izračunov metrik, na primer pričakovanega števila prihodnjih nakupov. Tudi klasična tehnika ocenjevanja največjega verjetja v tem modelu ni primerna za iskanje optimalnih parametrov modela. Abe je zato predlagal uporabo simulacije s pomočjo tehnike Markovskih verig Monte Carlo (angl. *Markov chain Monte Carlo – MCMC*) za ocenjevanje parametrov in izračun različnih metrik (Korkmaz, Fok & Kuik, 2014). Abe (2009) se je teh pomanjkljivosti zavedal in predlagal, da se njegov model uporabi kot komplement osnovnemu modelu Pareto/NBD. Osnovni model naj služi kot orodje za strateško analizo baze strank, model HB pa je primeren za operativno trženje, pri katerem se osredotočamo na posamezne stranke.

## 2.5 Primerjava modelov in programska oprema za implementacijo modelov

Za napoved prihodnjih nakupov strank ima podjetje na razpolago vrsto modelov. Eden od izzivov je, kako izbrati pravega. Vsak model ima določene predpostavke glede obnašanja strank, kot to prikazuje Tabela 11.

*Tabela 11: Primerjava modelov z vidika predpostavk in ocenjevanja parametrov*

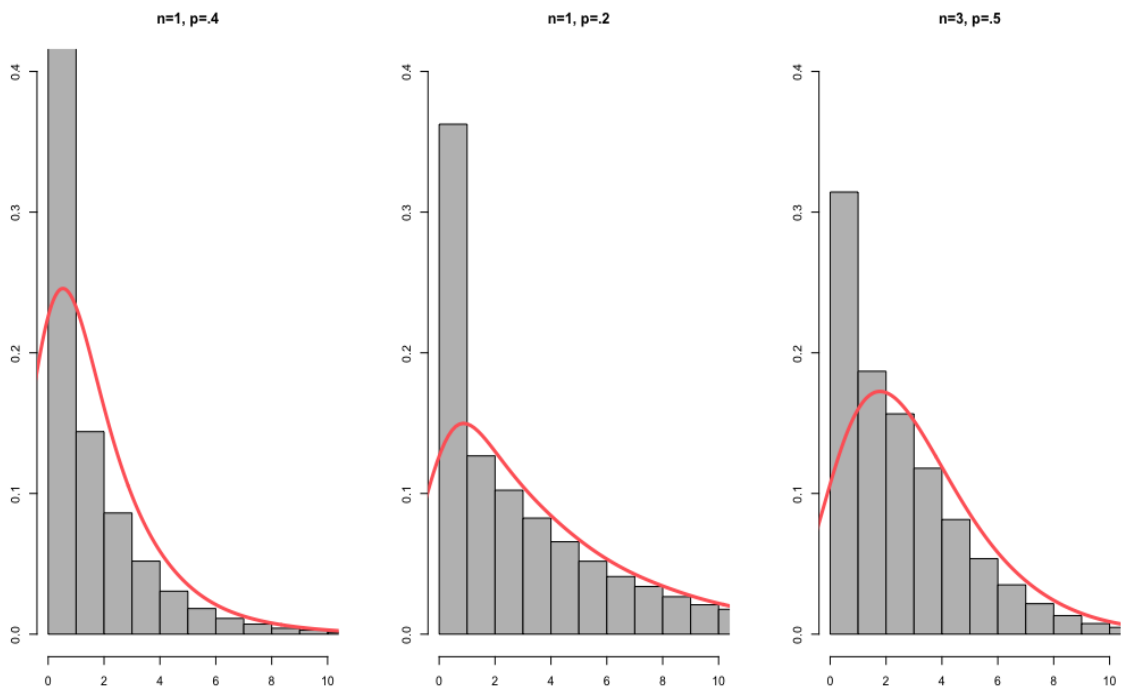
<b>Področje primerjave</b>	<b>Pareto/NBD</b>	<b>BG/NBD</b>	<b>CBG/NBD</b>	<b>CBG/CNBD-k</b>	<b>HB Pareto/NBD</b>
Nakupni proces	Poisson	Poisson	Poisson	Erlang	Poisson
Proces neaktivnosti (prebega)	EkspONENTNA	Zamaknjena geometrična	Zamaknjena geometrična	Zamaknjena geometrična	EkspONENTNA
Čas neaktivnosti	Zvezno	V trenutkih nakupa	V trenutkih nakupa in po prvem nakupu	V trenutkih nakupa in po prvem nakupu	Zvezno
Porazdelitev nakupov	Gama	Gama	Gama	Gama	Multi-variantna logaritemsko normalna
Porazdelitev stopnje osipa	Gama	Beta	Beta	Beta	
Ocenjeni parametri	Hiper-parametri	Hiper-parametri	Hiper-parametri	Hiper-parametri	Hiper-parametri in posamezni parametri
Metoda ocenjevanja	MLE	MLE	MLE	MLE	MCMC

Predpostavke modela glede obnašanja strank morajo v čim večji meri odražati obnašanje strank v realnosti, da lahko dobimo zanesljive napovedi. Slike 4, 5, 6 in 7 prikazujejo primere najpogostejših standardnih porazdelitev, na katerih temeljijo modeli. Te mi bodo v nadaljevanju, med raziskovalno analizo podatkov podjetja, v pomoč pri izbiri najprimernejših modelov za obravnavani primer.

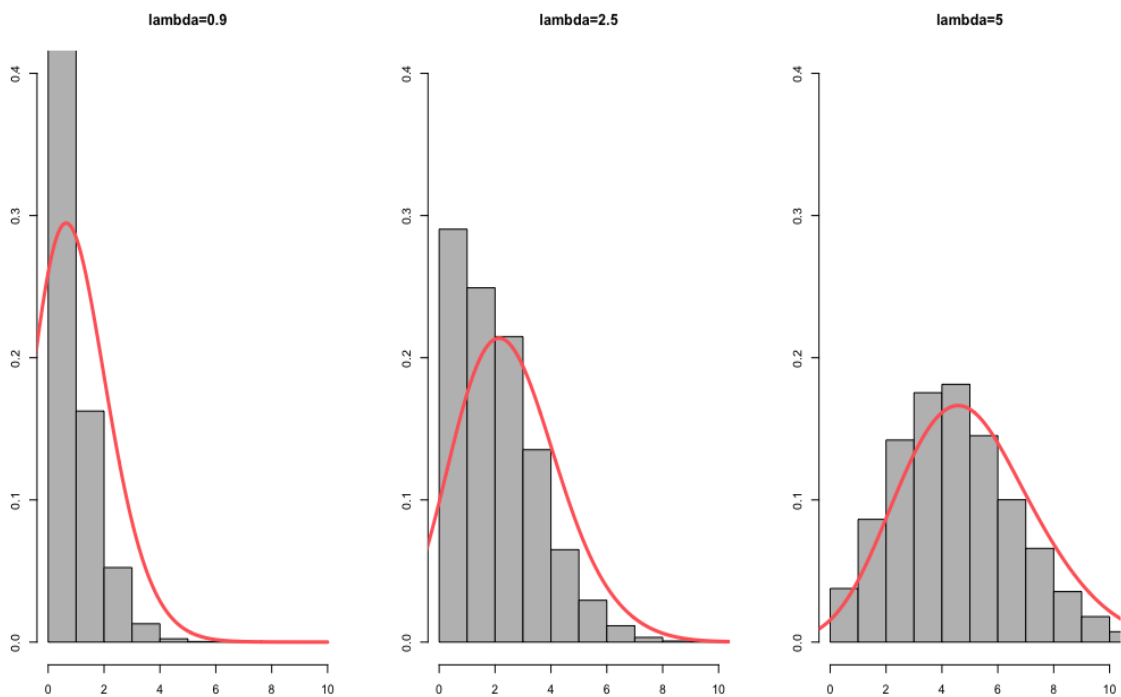
Leta 2012 je Wharton Customer Analytics Initiative, raziskovalni center Univerze v Pensilvaniji, izdal odprtokodni R paket BTYD, ki olajša implementacijo modelov Pareto/NBD in BG/NBD. Platzler (2008) je paket razširil z ostalimi zgoraj omenjenimi

modeli. Paketa vključujeta funkcije za pripravo podatkov, ocenjevanje parametrov, točkovanje in grafično prikazovanje.

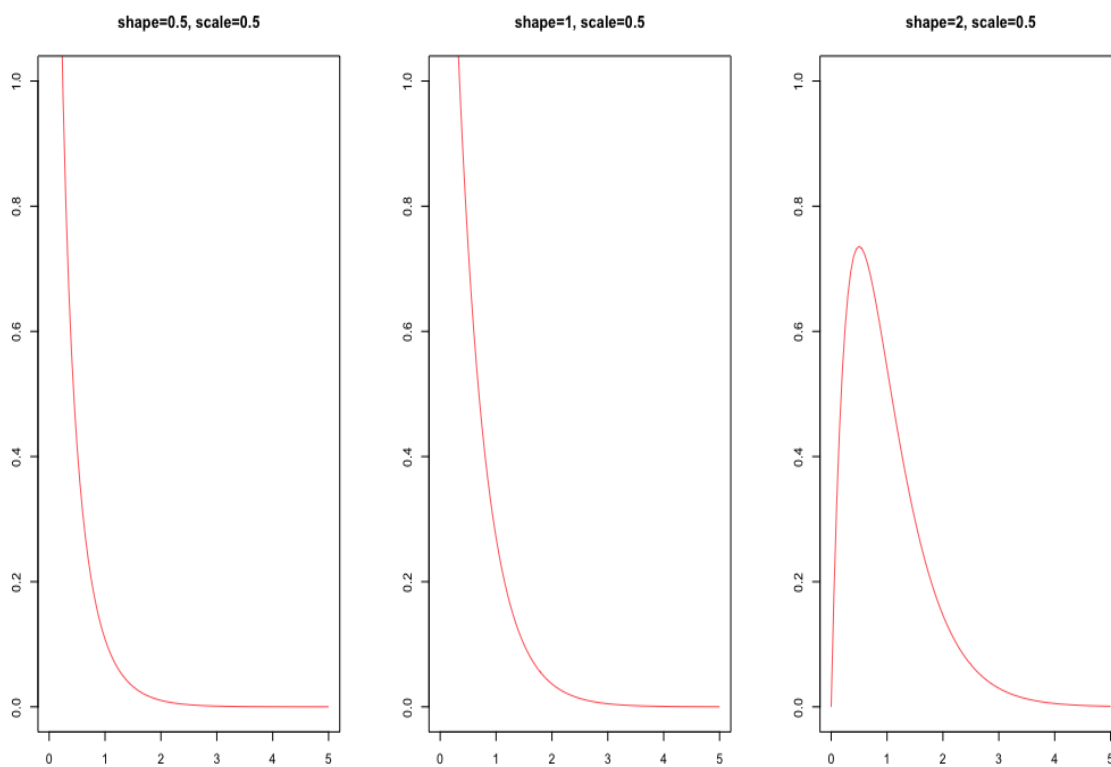
*Slika 4: Negativna binomska porazdelitev z različnimi parametri*



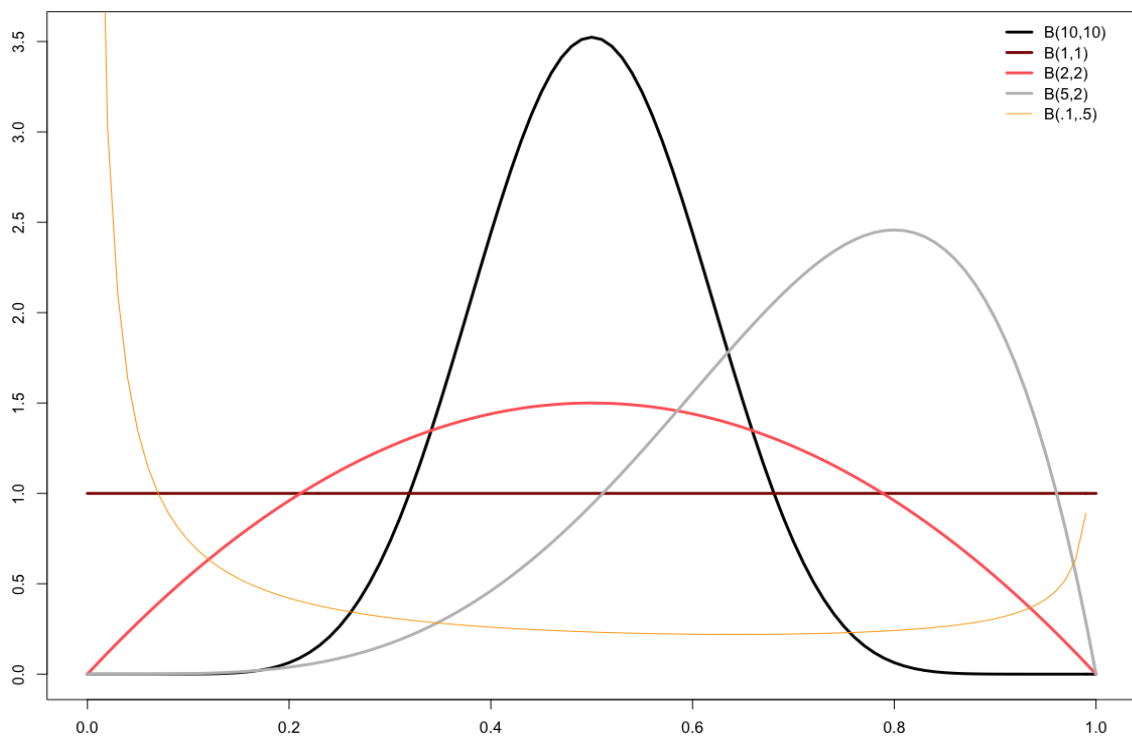
*Slika 5: Poissonova porazdelitev z različnimi parametri*



Slika 6: Gama porazdelitev z različnimi parametri



Slika 7: Beta porazdelitev z različnimi parametri





Uporaba napovednih modelov je kompleksna in zahtevna naloga, pri čemer dodaten problem predstavlja zaupanje odločevalcev v rezultate modela. Po drugi strani je ocenjevanje in razumevanje nakupnega vedenja strank aktivnost, ki poteka nepretrgoma. Te aktivnosti temeljijo na hevristični (angl. *heuristic*) analizi preteklih nakupov, intuiciji sodelavcev in poznavanju značilnosti posameznih strank, ki izhaja iz razmerja med podjetjem in stranko. Analiza, ki sta jo opravila Wubben in Wangenheim (2008), je pokazala, da je lahko takšno hevristično odločanje v določenih primerih bolj učinkovito kot napovedni modeli. To se pokaže predvsem pri napovedovanju aktivnosti oziroma neaktivnosti strank, medtem ko se pri napovedovanju števila prihodnjih nakupov bolje odrežejo napovedni modeli. Takšen rezultat lahko pojasnimo z osnovno značilnostjo napovednih modelov, ki ne upoštevajo vseh značilnosti nakupnega vedenja posameznih strank. Zaposleni sodelavci imajo po drugi strani neposredne odnose s strankami, poznajo njihove značilnosti in bolje razumejo njihovo delovanje.

### **3 BLACK DIAMOND EQUIPMENT**

Black Diamond Equipment je podjetje, ki načrtuje, proizvaja in prodaja inovativno športno opremo in oblačila za plezanje, gornišvo, alpinizem, smučanje in ostale športne aktivnosti v naravi.

#### **3.1 Trženje v podjetju**

Black Diamond Equipment je eden vodilnih svetovnih proizvajalcev inovativne športne opreme in oblačil za plezanje, gornišvo, alpinizem, smučanje in ostale športne aktivnosti v naravi. Svojo konkurenčno prednost je podjetje tradicionalno videlo v inovativnosti in kakovosti svojih izdelkov. Podjetje se je primarno osredotočalo na izdelke in temu podrejalo vsa področja poslovanja – od strategije, poslovnega planiranja, organizacije in procesov, merjenja učinkovitosti in zaposlenih. Najbolj cenjene stranke so bile tiste, ki so kupovale najbolj tehnične in napredne izdelke. Po pravilu je podjetje prodajalo izdelke preko visoko specializiranih manjših trgovcev z ozkim fokusom, ki prodajajo nišni skupini uporabnikov, ki točno vedo, kaj želijo.

Poslovno okolje se v zadnjih letih na trgih, na katerih podjetje deluje, spreminja. Manjše specializirane trgovine težko preživijo, prevladujejo večje mednarodne trgovske verige in prodaja prek spleta. Plezanje in gornišvo postajata zelo priljubljena med širšo populacijo, tipičen potrošnik pa po pravilu ni več strokovnjak za poznavanje izdelkov. Podjetje se mora, če želi ostati dobičkonosno, takšnemu trgu prilagoditi. V procesu strateškega in poslovnega planiranja je podjetje zavestno stopilo na pot osredotočanja na stranke. Seveda je treba posebej poudariti, da se osredotočenost na izdelke in osredotočenost na stranke med seboj ne izključujeta.

Eden od prvih korakov na tej poti je razumevanje in segmentacija strank. Podjetje svojim strankam prodaja po različnih distribucijskih kanalih. Velik del prodaje je sezonske narave, pri čemer posredniki naročila oddajajo za eno leto vnaprej, z zavezujočimi pogodbami. Podjetje na osnovi podatkov iz preteklih let ve, da je okrog 30 %–40 % naročil ad-hoc naročil končnih potrošnikov prek spletne trgovine. Gre za popolnoma nezavezujoča naročila, ki jih lahko stranke opravijo kadarkoli in ki pomembno vplivajo na uspeh podjetja, a predstavljajo veliko neznanko. V preteklosti se je podjetje za napovedovanje teh nakupov zanašalo na pretekle deleže ad-hoc naročil, vendar se ti niso izkazali kot zelo zanesljivi pokazatelji. Verjetnostni napovedni modeli predstavljajo alternativno možnost napovedovanja prihodnjih nakupov.

V podpoglavju 2.5 sem ugotovil, da različni napovedni modeli predpostavljajo različno obnašanje strank, kar pomembno vpliva na modeliranje in rezultate modela. Raziskovalna analiza podatkov o nakupih naših strank mi bo omogočila, da bolje spoznam obnašanje letih in ugotovim, kateri napovedni modeli so za podjetje primerni. V nadaljevanju bom za raziskovalno analizo uporabljal podatke podjetja o preteklih **ad-hoc naročilih končnih potrošnikov v letu 2014**. Iste podatke bom kasneje uporabil tudi za modeliranje prihodnjih nakupov s pomočjo izbranih modelov. Raziskovalna analiza bo ponudila pomemben vpogled v podatke in služila kot pomoč pri interpretaciji in validaciji modeliranja prihodnjih nakupov.

### **3.2 Vedenje strank in nakupno odločanje v primeru nakupa športnih rekvizitov**

Čeprav nakupno vedenje napovedni modeli obravnavajo kot izid enega verjetnostnega procesa ali več le-teh, pri čemer nas ne zanimajo posamezni dejavniki, je razumevanje nakupnega vedenja pomembno. Kot sta pokazala Wubben in Wangenheim (2008), lahko razumevanje vedenja posameznih strank in njihovega nakupnega odločanja služi kot dopolnilo ali kontrola pri interpretaciji rezultatov napovednih modelov.

Vedenje strank (angl. *consumer behavior*) je področje trženja, ki se ukvarja s tem, kako posamezniki, skupine in organizacije iščejo, kupujejo, uporabljajo in opustijo izdelke, storitve, ideje ali doživetja, da zadovoljijo svoje želje in potrebe (Kotler et al., 2015).

Ker podjetje Black Diamond Equipment deluje v športni panogi, je smiselno, da na vedenje strank gledamo v tem kontekstu. Bistvo vedenja strank v športu pa je pot, ne cilj (Funk, Alexandris & McDonald, 2008). Ne glede na to, ali gre za spremljanje ali dejansko udeleževanje v športni aktivnosti, je vedenje strank, ko govorimo o športu, usmerjeno v izkušnjo. Temu sledijo tudi »ne-športne« podporne aktivnosti, kot je nakupovanje opreme ali servisiranje. Naloga tržnikov in celotnega podjetja je, da strankam pomaga na njihovi poti.

Za zadovoljevanje svojih želja in potreb stranke potrebujejo čas in denar. Vsaka stranka se mora odločiti, koliko časa in denarja bo namenila ukvarjanju s športom. Čas, ki ga stranka nameni športu, je lahko dejanski čas, ki ga sama nameni izvajanju športa, ali pa gre za čas, ki ga posveti spremljanju športa po televiziji, sledenju po internetu in socialnih omrežjih ipd. Čas, namenjen športu, ima lahko neposredne posledice tudi na nakupno vedenje, ki ni direktno povezano z izbranim športom. Tukaj gre za izbiro življenjskega stila, ki je posebej med plezalci in alpinisti, ki predstavljajo glavni profil strank podjetja, izrazito prisoten. Gre na primer za izbiro oblačil, avtomobilov, hrane in podobno. Drugi faktor, ki strankam omogoča zadovoljevanje želja in potreb po športnem udejstvovanju, je denar. Ta denar lahko stranka porabi za udeležbo na prireditvah, športne počitnice, opremo in tudi naročnine na revije, televizijske programe ipd.

Na vedenje strank vplivajo različni kulturni, socialni in osebni dejavniki. Tako kot v ostalih panogah morajo skušati tržniki tudi v športni boljše spoznati in razumeti nakupno vedenje strank. Na ta način lahko podjetje ohranja in ustvarja potrebo po svojih izdelkih. Razumevanje strank omogoča tudi njihovo segmentacijo in zagotavlja, da so njihove želje in potrebe ustrezno upoštevane. Pri tem se je treba zavedati nekaterih posebnosti športne panoge, še posebej pri športih, ki jih podjetje podpira (Funk et al., 2008):

- športni potrošniki so strokovnjaki. Alpinisti, plezalci in smučarji se na opremo zanašajo s svojimi življenji, zato je njihovo poznavanje opreme v povprečju izredno dobro.
- Možnosti udejstvovanja so odvisne od klimatskih razmer in sezonske narave. Smučanje je na primer možno pozimi, (športno) plezanje večinoma v toplejših mesecih.
- Šport je sam po sebi neoprijemljiv. Če podjetje želi, da bo nova stranka kupila izdelek, jo je treba najprej prepričati v lepote in prednosti ukvarjanja s športom.
- Šport je socialna aktivnost. Plezanje in turno smučanje se po pravilu vedno izvajata v skupinah.
- Neuspeh v športu je nekaj normalnega in ni nujno posledica okvar ali slabe kakovosti opreme. Slabo vreme in razmere, slabo počutje, precenjevanje lastnih sposobnosti ipd. lahko vodijo v nezadovoljstvo, ki ni neposredno povezano z izdelki podjetja.

Z vidika nakupnega vedenja strank v športni panogi sta ključni dve vprašanji (Funk et al., 2008), in sicer:

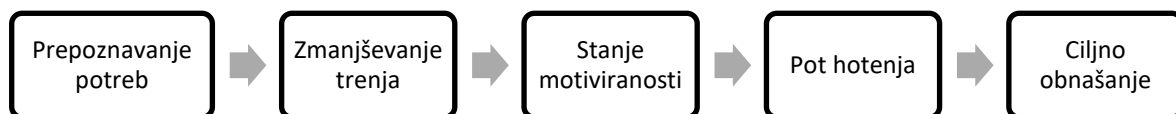
- zakaj stranka postane motivirana, da kupi točno določen izdelek?
- Kako je stranka sploh postala športnik, ki se zanima za takšen izdelek, oziroma kako udejstvovanje v športni aktivnosti vpliva na nakupno odločanje?

V nadaljevanju predstavljam procesa, ki izhajata iz zgornjih vprašanj – motivacijski proces in proces nakupnega odločanja.

### 3.2.1 Motivacijski proces

Motivacijski proces športnega potrošnika je proces, ki povzroča, da se posameznik obnaša, kot se. Enostavno povedano je posameznik motiviran, da se udeleži oziroma spremlja športno aktivnost zato, ker mu ta aktivnost nudi nekaj pozitivnega. Motivacija je notranji dejavnik, ki spodbuja in aktivira ciljno usmerjeno obnašanje. Motivacijski proces je sestavljen iz petih zaporednih korakov, kot to prikazuje Slika 8 (Funk et al., 2008).

*Slika 8: Motivacijski proces potrošnika*



*Vir: D. C. Funk et al., Consumer Behaviour in Sport and Events, 2008, str. 17, slika 2.1.*

Prvi korak v motivacijskem procesu je prepoznavanje potrebe, ko posameznik ugotovi, da ima potrebo. Ta potreba je posledica dejanskega stanja in nekega idealnega stanja. Potrebe so lahko biološke (na primer lakota, žeja, nelagodje) ali psihološke (na primer potreba po prepoznavnosti, pripadanju). Potreba postane motiv, ko je stimulirana do te mere, da nas privede do tega, da ukrepamo. Ta proces motivacije in prepoznavanja potreb so podrobno obravnavali različni avtorji, na primer Freud, Maslov in Herzberg (Kotler et al., 2015).

Potreba ustvarja trenje, ki je posledica med dejanskim in idealnim stanjem. Takšno trenje je neprijetno stanje, ki traja, dokler potreba, ki je to trenje povzročila, ni zadovoljena. Posamezniki poizkušajo trenja zmanjšati ali povsem odpraviti s potrošnjo izdelkov in storitev. Stopnja prisotnega trenja vpliva na nujnost oziroma intenzivnost, s katero želi posameznik ukrepati.

Ko je posameznik stimuliran do te mere, da želi potrebo zadovoljiti in odpraviti trenje, postane motiviran. Motiviran posameznik je pripravljen ukrepati. Kako bo ukrepal, je odvisno od njegovega zaznavanja situacije. Zaznavanje (angl. perception) je proces, s katerim izberemo, organiziramo in interpretiramo informacije iz okolja in si ustvarimo sliko sveta okrog nas (Kotler et al., 2015). Zaznavanje je izrazito subjektiven proces, saj vsak posameznik informacije interpretira drugače. Posledično bo vsak motiviran posameznik ukrepal v skladu s svojim zaznavanjem in izbral pot za zadovoljitev potrebe, ki mu najbolj ustreza.

Pot hotenja predstavlja korake do izbrane oblike potrošnje, ki jo je posameznik izbral za zadovoljitev svoje potrebe, in odpravo trenja, ki jo je ta potreba ustvarila. Pot, ki jo

posameznik izbere, je odvisna od mnogih dejavnikov – preteklih izkušenj in spoznanj, socialnih in kulturnih dejavnikov iz okolja, v katerem je odraščal in v katerem se nahaja, osebnih karakteristik ipd.

Trenje, ki ga potreba ustvari, potrošnika potiska (angl. *push*) na pot do zadovoljitve potrebe. Naloga tržnikov je, da potrošnike povlečejo (angl. *pull*) na pot do svojih izdelkov in storitev, ki lahko te potrebe zadovoljijo.

Ciljno obnašanje je zadnji korak v motivacijskem procesu in predstavlja nakup, s katerim zmanjšamo oziroma odpravimo trenje in zadovoljimo potrebo.

V športni panogi velja, da imajo tržniki omejene možnosti pri ustvarjanju potrebe po športu v okviru motivacijskega procesa. Lahko pa so izredno učinkoviti pri zagotavljanju in komuniciranju poti do športnih aktivnosti in izdelkov, ki zadovoljujejo takšne potrebe po športu. Obstaja pet glavnih motivov, ki nam pomagajo razumeti, zakaj se posamezniki odločajo za športne aktivnosti, tržniki pa jih lahko uporabijo za oblikovanje svojih strategij (Funk et al., 2008):

- socializacija (angl. *socialisation*) – predstavlja potrebo po druženju. V procesu športnega udejstvovanja se krepijo odnosi z drugimi udeleženci.
- Zmogljivost (angl. *performance*) – predstavlja potrebo po estetskem in fizičnem ugodju.
- Navdušenje (angl. *excitement*) – predstavlja potrebo po intelektualnem stimuliranju.
- Spoštovanje (angl. *esteem*) – predstavlja potrebo po tekmovanju in dokazovanju.
- Preusmeritev (angl. *diversion*) – predstavlja potrebo po mentalnem ugodju. Šport predstavlja odmik od rutinskega vsakdana in stresa.

### **3.2.2 Nakupno odločanje**

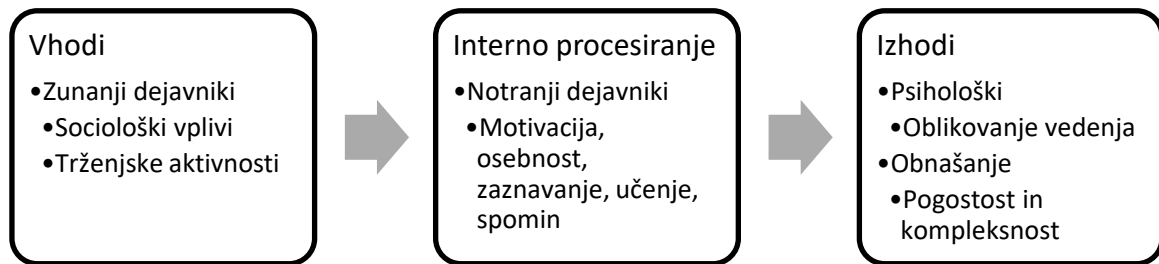
V kontekstu športne panoge so Funk et al. (2008) nakupno odločanje organizirali v tri glavne faze, kot to prikazuje Slika 9. Vhodni (zunanji) dejavniki vplivajo na ocenjevanje in vrednotenje možnosti, ki se odvija v drugi fazi. Ta predstavlja procese, kot so motivacija, osebnost, zaznavanje in spomin, ki vplivajo na izhode celotnega procesa odločanja. Ti se manifestirajo kot odnos do določenega športnega izdelka oziroma storitve in jih ni mogoče neposredno opazovati. Direktno opazovan izhod procesa pa je dejansko obnašanje, ki lahko vključuje na primer nakup ali različne druge aktivnosti, povezane s sprejeto odločitvijo.

Kot vhodi v procesu nakupnega odločanja nastopajo različni zunanji dejavniki, ki jih lahko združimo v dve glavni kategoriji:

- trženjske aktivnosti podjetja – naloga podjetja je, da potrošnike privabi in jih prepriča, da je njihov izdelek ali storitev pravi način za zadovoljitev njihove potrebe.

- Sociološki vplivi so druga kategorija dejavnikov iz okolja, ki vplivajo na potrošnika. Sem štejemo informacije in priporočila družine, prijateljev, sodelavcev, športnih partnerjev ipd. V to skupino lahko uvrstimo tudi geografske in klimatske vplive, socialni sloj, spol, kulturo in subkulturo, vrednote ipd.

Slika 9: Nakupno odločanje



Vir: D. C. Funk et al., *Consumer behaviour in sport and events*, 2008, str. 31, slika 3.2.

Zunanji dejavniki vplivajo na interno procesiranje v drugi fazi. V splošnem lahko notranje dejavnike kategoriziramo v tri skupine:

- spoznavni (angl. *cognitive*),
- čustveni (angl. *affective*),
- vedenjski (angl. *behavioral*).

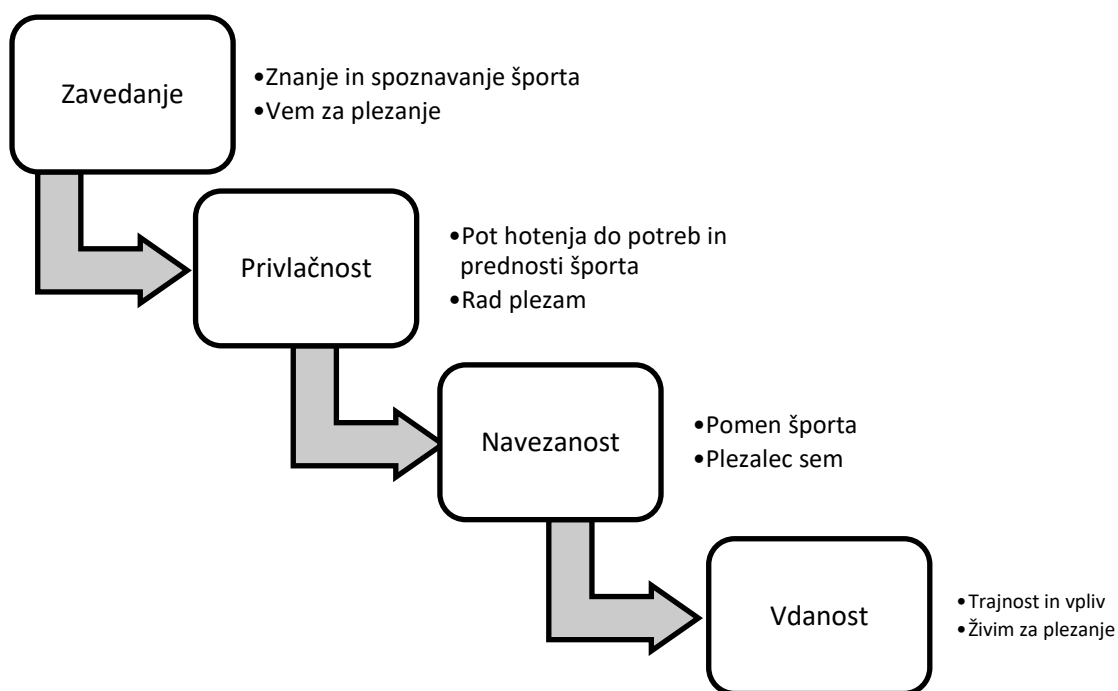
V to fazo spadajo aktivnosti, kot so prepoznavanje potreb, iskanje alternativ in njihovo ocenjevanje. Proces ocenjevanja je izrazito subjektiven in temelji na psiholoških in osebnostnih značilnostih posameznika. Cilj tega internega procesiranja je ugotoviti, ali lahko neka alternativa zadovolji določeno potrebo ali vsaj v zadostni meri zmanjša trenje, ki ga potreba povzroča. V veliki meri se to ocenjevanje nanaša na znanje, tj. informacije, ki so hranjene v spominu. Obstoječe znanje o določenem športnem rekvizitu oziroma še posebej pridobivanje novega znanja o rekvizitu sta zato v procesu ocenjevanja ključnega pomena. Pridobivanje novega znanja poteka s pomočjo vrste notranjih procesov, kot so motivacija, osebnostne značilnosti, zaznavanje, učenje in spomin. Pridobljeno znanje posamezniku pomaga pri izbiri ustrezne alternative in vodi v tretjo, zadnjo fazo.

Faza izhodov predstavlja končne psihološke in vedenjske rezultate. Psihološki rezultati se kažejo kot ustvarjanje mnenja o določeni športni aktivnosti oziroma izdelku in jih lahko prav tako kategoriziramo v spoznavne, čustvene in vedenjske. Teh rezultatov ni mogoče direktno opisati in se kažejo skozi čustva, znanje in pripravljenost na udeleževanje v športni aktivnosti. Vedenjske rezultate lahko po drugi strani opazujemo direktno, mednje pa štejemo

na primer nakupno vedenje, aktivnosti, ki sledijo odločitvi o nakupu, in vedenje, ki sledi nakupni izkušnji.

Vsa do sedaj predstavljena spoznanja so Funk et al. (2008) združili v svojem modelu PCM (angl. *psychological continuum model*), s katerim poizkušajo pojasniti vedenje potrošnikov v športu. Model pojasnjuje, kako osebnostni, psihološki in okolijski dejavniki skupaj vplivajo na razvoj vpletenosti potrošnika v športu. Ta vpletenost se opira na nakupno odločanje, ki karakterizira stopnjo vpletenosti v športu (tu pojem športa obsega zelo široko pojmovno področje – panoga, ekipa, igralec, dogodek, aktivnost, rekvizit). Model razloži, kako vhodi, interno procesiranje in izhodi ustvarjajo progresivno vpletenost in jih karakterizira v štirih zaporednih fazah: zavedanje (angl. *awareness*), privlačnost (angl. *attraction*), navezanost (angl. *attachment*) in vdanost (angl. *allegiance*). Te faze so hkrati tudi glavni mejniki v odločitvenem procesu nakupnega odločanja in določajo ocenjevanje osebnostnih, psiholoških in okolijskih vhodnih dejavnikov. Bolj je posameznik vpleten, bolj je interno procesiranje zahtevno in kompleksno. Model PCM prikazuje Slika 10.

Slika 10: PCM model



Vir: D. C. Funk et al., *Consumer behaviour in sport and events*, 2008, str. 43, slika 3.5.

### 3.3 Raziskovalna analiza podatkov

Podatki o vseh spletnih nakupih strank so zabeleženi v podatkovnih bazah podjetja. Za potrebe raziskovalne analize in napovedovanja prihodnjih nakupov sem imel na voljo

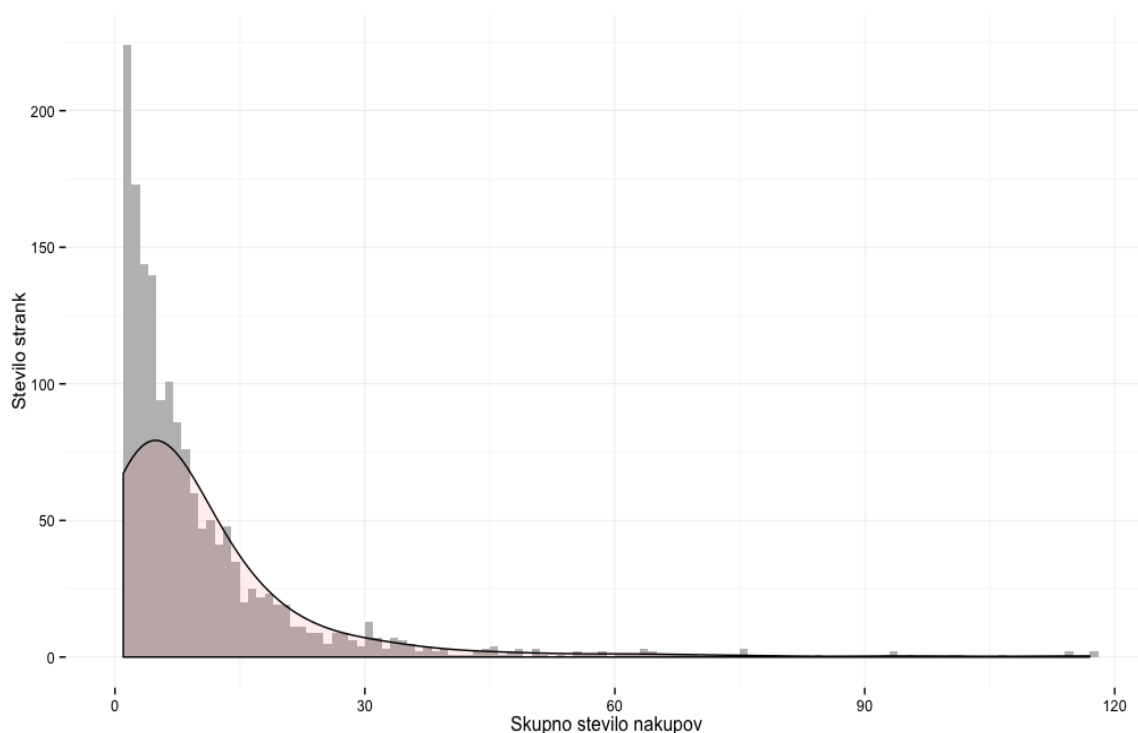
podatke o 16.779 nakupih, ki jih je prek spleta opravilo 1.633 končnih strank, in sicer v obdobju med 1. januarjem 2014 in 31. decembrom 2014. Vključene so samo tiste stranke, ki so prvi nakup opravile v prvem kvartalu leta 2014. Na podlagi podatkov, prikazanih v Tabeli 12, sem ugotovil, da so si stranke med seboj precej različne.

*Tabela 12: Pregled osnovnih značilnosti razpoložljivih podatkov*

Število nakupov; število strank	16.779; 1.633
Obdobje	1. 1. 2014–31. 12. 2014
Število nakupov: min; max; srednja vrednost	1; 117; 10,27
Število strank z enim nakupom	224; 13,7 %
Čas med nakupi v dnevih: min; max; srednja vrednost	1; 344; 43,81
Čas od zadnjega nakupa v dnevih: min; max; srednja vrednost	1; 365; 135
Znesek nakupa v USD: min; max; srednja vrednost	0; 30.304; 643,3

Te razlike so opazne pri številu in znesku nakupov in tudi pri času, ki preteče med posameznimi nakupi. 224 oziroma 13,7 % strank je v opazovanem obdobju opravilo samo en nakup. Ta delež je relativno nizek in nakazuje dobro lojalnost strank. Po drugi strani so nekatere stranke kupovale tudi do 117-krat. V povprečju je vsaka stranka opravila 10 nakupov. Slika 11 prikazuje nakupe in njihovo porazdelitev. Zneski nakupa se gibljejo med 0 in 30.304 USD, v povprečju pa je bil nakup vreden 643 ameriških dolarjev. Na tem mestu velja pojasniti, da znesek 0 pomeni 100 % popust.

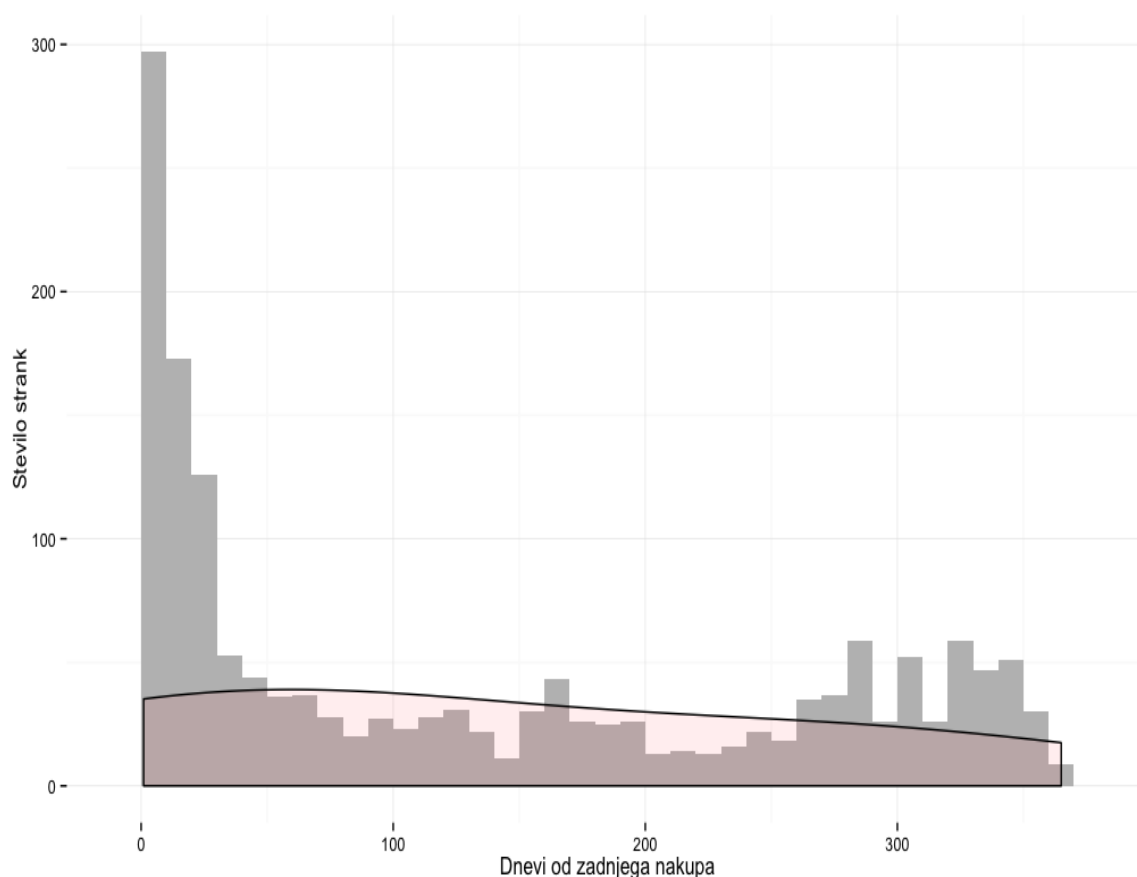
*Slika 11: Skupno število nakupov*





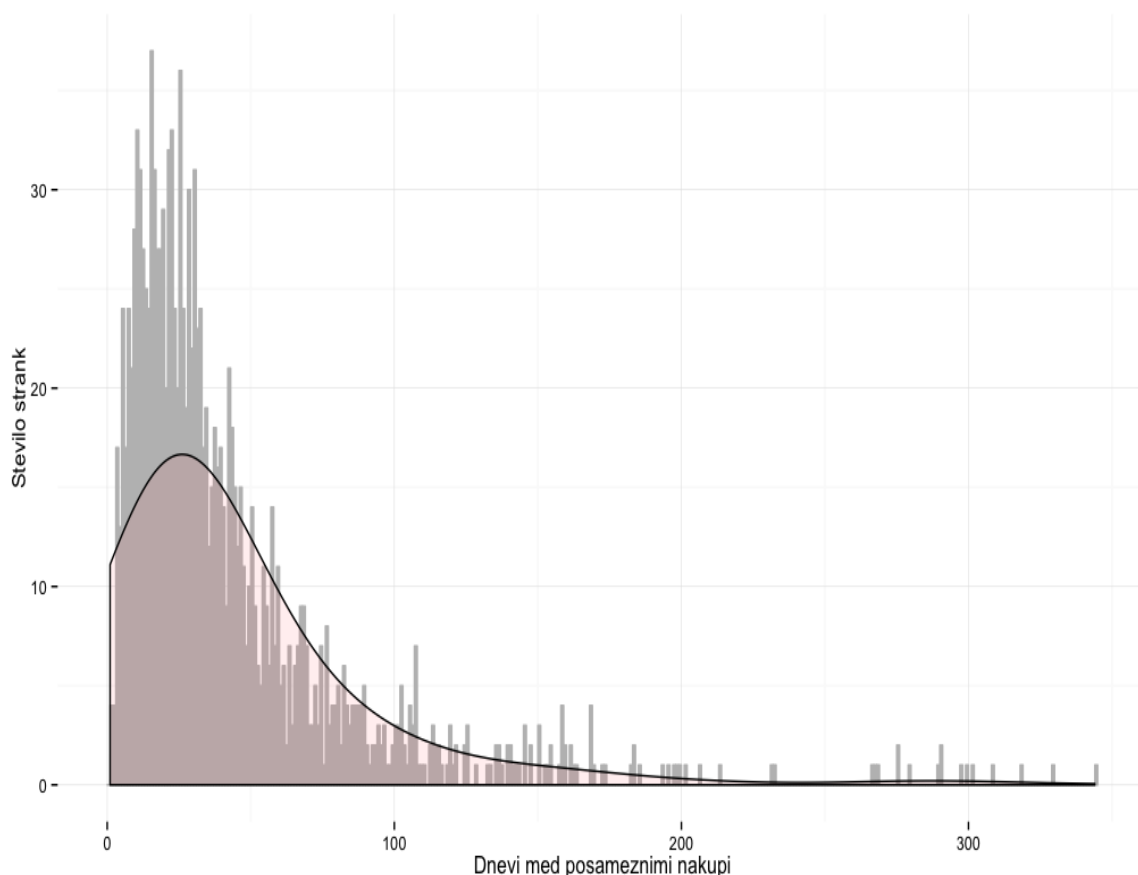
Opazne so tudi razlike v času od zadnjega nakupa, čeprav te niso tako izrazite. Tako so nekatere stranke prvi in edini nakup opravile na začetku leta, nekatere stranke pa konec leta. V povprečju je od zadnjega nakupa minilo 135 dni. Slika 12 prikazuje število dni od zadnjega nakupa.

*Slika 12: Število dni od zadnjega nakupa*



Velike razlike so tudi v času, ki preteče med posameznimi nakupi. Nekatere stranke so nakupe opravile v zaporednih dnevih, druge v 344 dneh, v povprečju pa je med nakupi preteklo 44 dni. Slika 13 prikazuje povprečno število dni med posameznimi nakupi. Čeprav večina strank naslednji nakup opravi v manj kot 100 dneh, tudi daljša neaktivnost ne pomeni, da nas je stranka zapustila. Veliko strank namreč potrebuje več kot 100 dni za ponovni nakup. Po drugi strani je iz podatkov razvidno, da zelo malo strank nakupuje v zaporednih dnevih in da obstaja obdobje nekaj dni po zadnjem nakupu, ko stranke niso pripravljene kupovati. Pri modeliranju prihodnje aktivnosti to predstavlja potencialno težavo, ker v podatkih ne obstaja jasen vzorec, kdaj stranka dejansko postane neaktivna oziroma nas zapusti, ampak samo hibernira in lahko naslednji nakup opravi v kateremkoli trenutku.

Slika 13: Povprečni čas med posameznimi nakupi

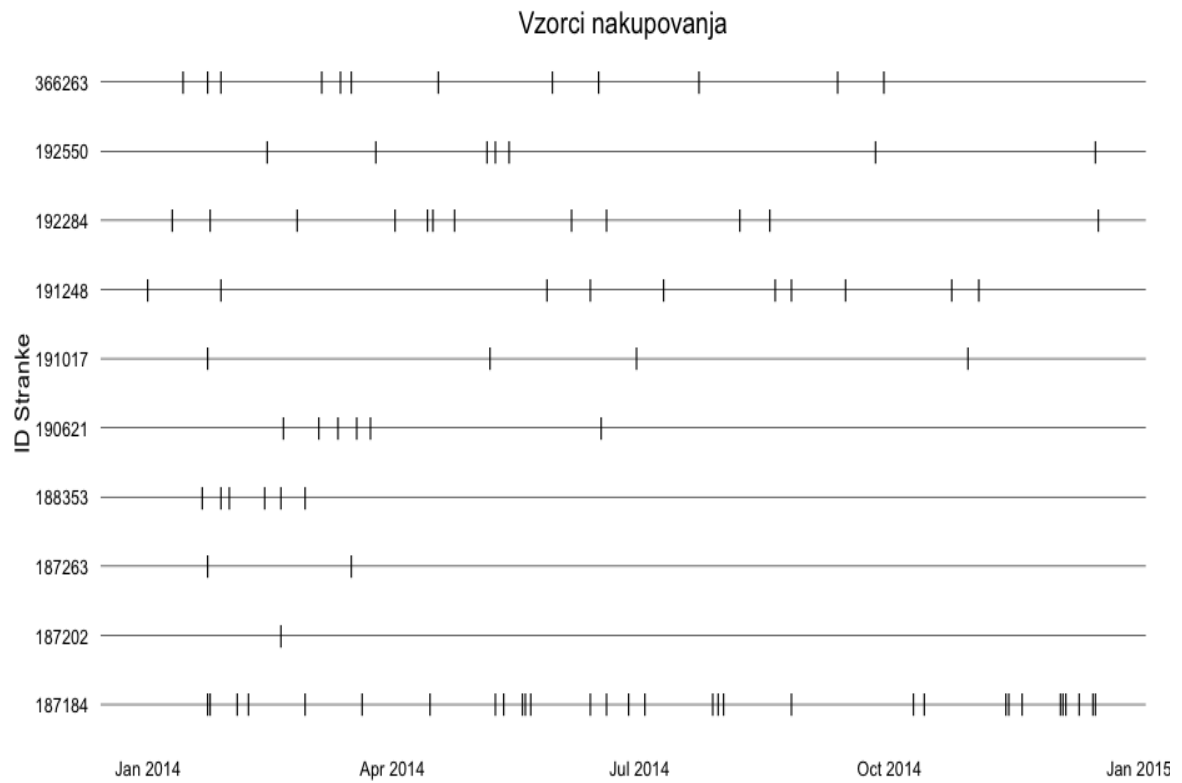


Slika 14 prikazuje časovne vzorce nakupovanja desetih naključno izbranih strank. Vidimo primer stranke, ki je kupila samo enkrat (ID 187202), primer redne stranke (ID 187184) in primer stranke, ki je verjetno neaktivna (ID 188353).

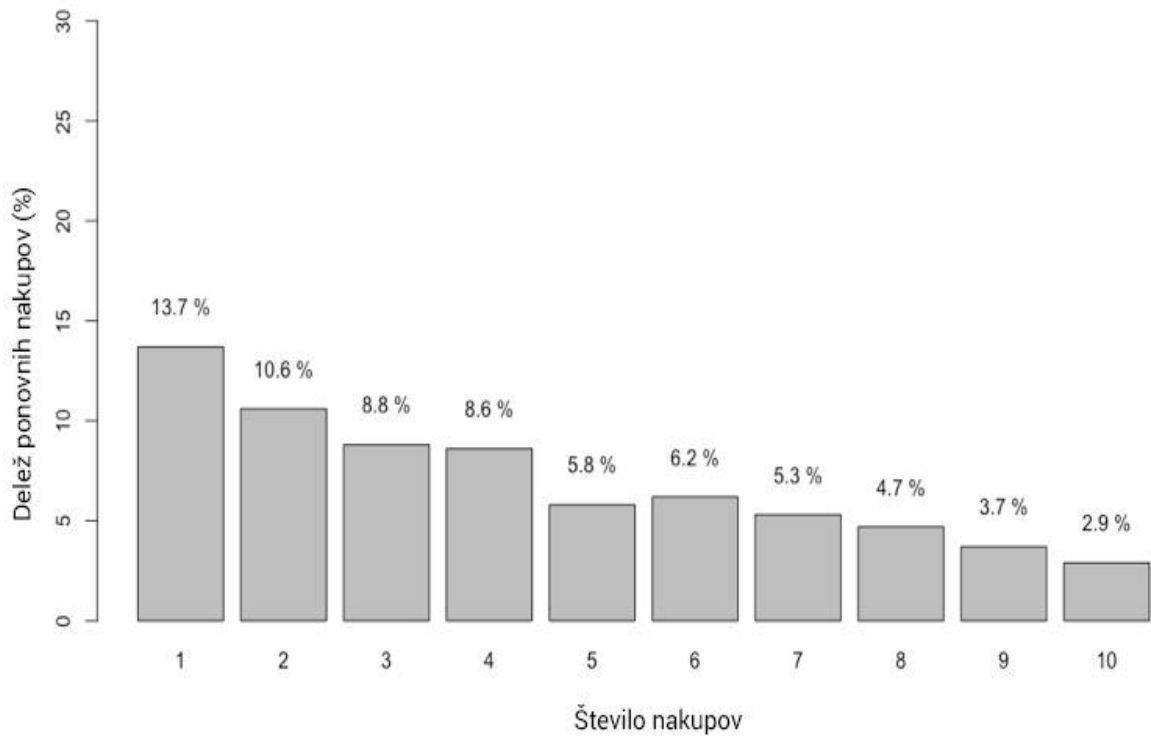
Kot rečeno, je 224 strank opravilo samo en nakup. Te stranke predstavljajo 13,7 % vseh strank, kot to prikazuje Slika 15. Stranke, ki so opravile 10 ali manj nakupov, predstavljajo 70,3 % vseh strank, vendar te stranke skupaj ustvarijo samo 23 % celotne vrednosti nakupov. 29,7 % strank, ki opravi več kot 10 nakupov, ustvari 67 % celotne vrednosti. To pomeni, da so stranke, ki opravijo več kot 10 nakupov, za podjetje izredno pomembne.

Slika 16 prikazuje gibanje števila transakcij v letu 2014. Dobro so vidni štirje vrhovi, ki predstavljajo sezonsko komponento. Konec zime in poletja se tipično poveča količina ad-hoc naročil zaradi posezonskih akcij in razprodaj. Vrhova sredi zime in sredi poletja sta bolj negotova in zelo odvisna od vremenskih razmer. Če so te ugodne, se ad-hoc prodaja sezonskih izdelkov tipično poveča in obratno.

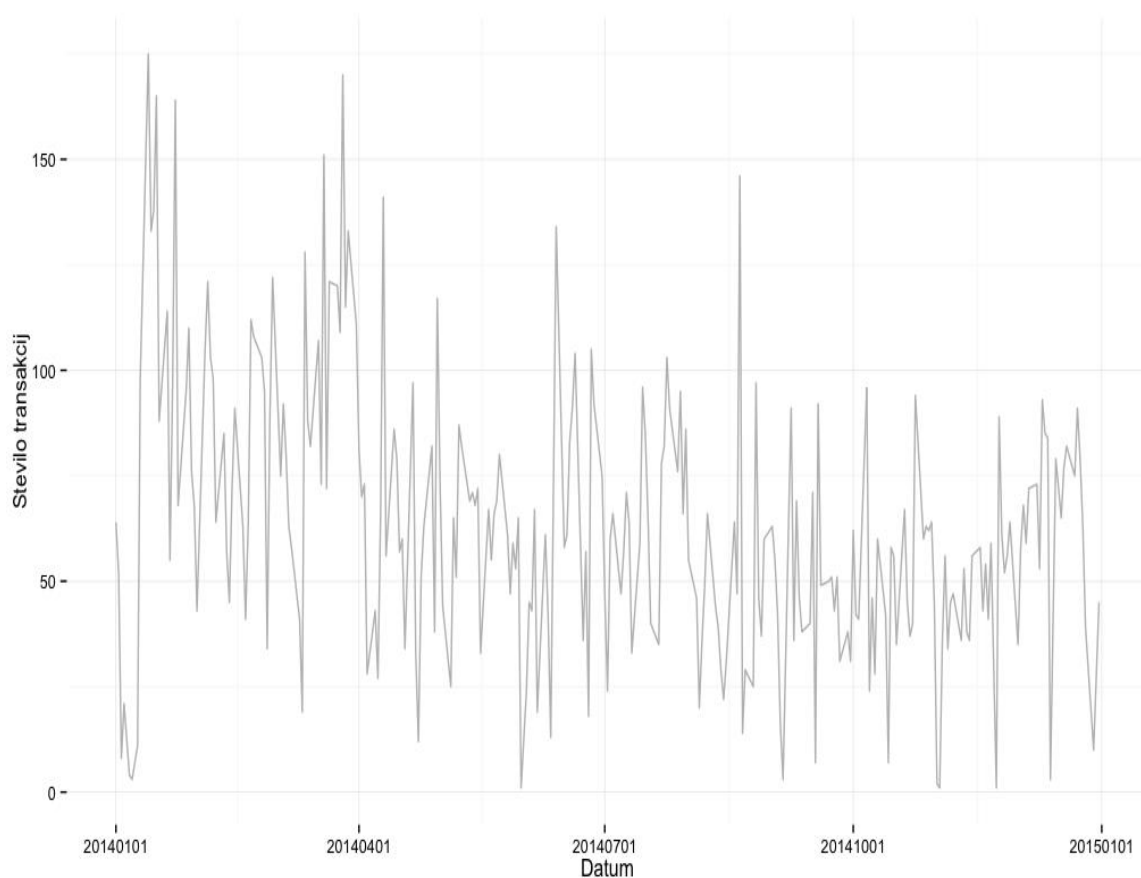
Slika 14: Vzorci nakupovanja v času desetih naključno izbranih strank



Slika 15: Deleži strank po številu nakupov



Slika 16: Število transakcij v letu 2014



### 3.4 Izbira modelov in kriteriji primerjave

Na tem mestu lahko ponovno izpostavim ugotovitev, da podjetje deluje v kontekstu nezavezujočih odnosov s strankami, ki lahko nakupe opravljajo kadarkoli. Dosedanja analiza je pokazala, da so modeli BTYD primerni za takšno analizo. Pomembno je izbrati tiste, ki so uporabni za naš kontekst. Glede na predstavitev modelov v kontekstu poslovnih odnosov (Tabela 2) morajo izbrani modeli temeljiti na osnovnem modelu NBD.

Dodatno lahko ugotovim, da imam na voljo vse podatke, ki jih potrebujem za uporabo teh modelov: koliko transakcij je stranka izvedla v opazovanem preteklem obdobju, čas zadnje transakcije v opazovanem preteklem obdobju in trajanje opazovanega obdobja.

S pomočjo R paketov BTYD in BTYDplus bom na razpoložljivih podatkih uporabil modele Pareto/NBD, BG/NBD, CBG/NBD in CBG/CNBD-k. Gre za najbolj razširjene in preverjene modele v strokovni literaturi. Njihova uporaba je s pomočjo R paketov relativno enostavna, obsežna strokovna literatura pa mi je olajšala razumevanje in interpretacijo.

Rezultate modelov bom primerjal z dveh vidikov, pri čemer je pomembno poudariti, da sem podatke razdelil na dva dela. Prva polovica leta predstavlja obdobje kalibracije (angl. *calibration period*), druga polovica leta pa obdobje validacije (angl. *validation period*). Podatke iz obdobja kalibracije vedno uporabimo za ocenjevanje parametrov modelov, na podlagi katerih nato napovemo, kaj se bo dogajalo v prihodnosti, v obdobju validacije. Dejanske podatke iz obdobja validacije uporabimo samo, ko želimo ugotoviti razliko med napovedjo modelov in dejanskim stanjem v obdobju validacije.

Prvi kriterij primerjave je skladnost (angl. *goodness of fit*) modelov z dejanskimi podatki. Tukaj gre za enostavno primerjavo med podatki o številu transakcij, ki jih napovejo modeli, in dejanskim številom nakupov, ki ga imamo na voljo v podatkih v obdobju validacije. Pri tem velja naslednja izpeljava – bolj sta si ti vrednosti bližji, bolj je napoved skladna z realnostjo.

Načeloma velja, da skladnost sama po sebi še ne zagotavlja tudi sposobnosti napovedovanja, zato bom dodatno primerjal tudi točnost napovedi (angl. *forecast accuracy*). Obstaja več mer točnosti in pravimo, da je model uspešnejši, če napove vrednost, ki je bližja pravi vrednosti, in manj uspešen, če je napoved modela bolj oddaljena od dejanske vrednosti števila nakupov. Tukaj nas zanima točnost modela kot celote. Vsi izbrani modeli namreč napovedujejo prihodnje nakupe posameznih strank (spomnimo, teh je 1.633) in vsaka posamezna vrednost lahko odstopa od dejanske. Da bi ocenili točnost modela kot celote, uporabljamo agregatne mere točnosti:

- srednja kvadratna napaka (angl. *mean squared error*, v nadaljevanju MSE), ki je definirana kot povprečje kvadratov razlik med napovedanimi vrednostmi  $\hat{y}_i$  in resničnimi vrednostmi  $y_i$  in je zapisana z enačbo 8.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (8)$$

- Koren srednje kvadratne napake (angl. *root mean squared error*, v nadaljevanju RMSE), ki je enostavno kvadratni koren MSE in ga uporabimo, kadar želimo, da je napaka izražena v enakih enotah kot podatki, na podlagi katerih je bila izračunana. Zapisana je z enačbo 9. RMSE daje večjo težo skladnosti in je občutljiva na vrednosti, ki zelo odstopajo od povprečja.

$$RMSE = \sqrt{MSE} \quad (9)$$

- Srednja absolutna napaka (angl. *mean absolute error*, v nadaljevanju MAE), ki je definirana kot povprečje absolutnih vrednosti razlik med napovedanimi vrednostmi  $\hat{y}_i$  in resničnimi vrednostmi  $y_i$  in je zapisana z enačbo 10.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (10)$$

- Srednja absolutna napaka v odstotkih (angl. *mean absolute percentage error*, v nadaljevanju MAPE), ki je definirana kot odstotkovno izražena vrednost povprečja absolutnih vrednosti razlik med napovedanimi vrednostmi  $\hat{y}_i$  in resničnimi vrednostmi  $y_i$  in je zapisana z enačbo 11.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad (11)$$

- Srednja kvadratna logaritemska napaka (angl. *mean squared logarithmic error*, v nadaljevanju MSLE), ki je definirana kot kvadrat relativne logaritemske razlike med resničnimi vrednostmi  $y_i$  in napovedanimi vrednostmi  $\hat{y}_i$  in je zapisana z enačbo 12.

$$MSLE = \frac{1}{n} \sum_{i=1}^n \left( \log \left( \frac{y_i + 1}{\hat{y}_i + 1} \right) \right)^2 \quad (12)$$

MSLE večjo težo namenja točnosti napovedi dominantni skupini in je kot taka manj občutljiva na vrednosti, ki zelo odstopajo od povprečja. Iz tega razloga bo moj glavni kriterij za ocenjevanje točnosti MSLE. Izbral bom model, kjer je MSLE najnižja.

### 3.5 Uporaba modelov za napovedovanje prihodnjih nakupov strank

Na voljo imamo podatke o 16.779 posameznih nakupih 1.633 strank obravnavanega podjetja v letu 2014. Del razpoložljivih podatkov v izvornem stanju prikazuje Tabela 13. Stolpec »Stranka« prikazuje enolično številko stranke (ID), stolpec »Datum« prikazuje datum transakcije in stolpec »Vrednost nakupa« prikazuje znesek nakupa v USD.

Tabela 13: Razpoložljivi podatki

Stranka	Datum	Vrednost nakupa
187614	2014-01-01	21,83
187784	2014-01-01	264,95
188032	2014-01-01	1.312,08
188993	2014-01-01	873,45

Preden sem lahko podatke dejansko uporabil v posameznih modelih, sem jih moral pripraviti. Priprava vključuje na primer združevanje nakupov, ki jih je stranka opravila v istem dnevu, saj je dan tista časovna enota, ki me je v tem primeru zanimala. Modeli zahtevajo samo tri podatke za vsako stranko: število nakupov v obdobju, čas od zadnjega

nakupa in trajanje obdobja. Te spremenljivke sem lahko enostavno pripravil na osnovi podatkov, ki sem jih imel na voljo. Koristnost paketov BTYD pride v tem koraku še posebej do izraza, saj vključujejo funkcije, ki te spremenljivke izračunajo.

Naslednji ključni korak je razdelitev podatkov v kalibracijski in validacijski set. V obravnavanem primeru sem za kalibracijo uporabil podatke prve polovice leta, do vključno 30. junija 2014. Prav tako sem lahko pripravil podatke za primerjavo kalibracijskega in validacijskega obdobja.

Tabela 14 prikazuje tako pripravljene podatke, ki sem jih lahko uporabil v izbranih modelih. Vidimo, da so podatki pripravljene za vsako od 1.633 strank, za razliko od osnovnih podatkov, ki so bili organizirani za vsak posamezni nakup. Stolpec »Stranka« prikazuje ID stranke, stolpec »Št. ponavljajočih nakupov v kalibracijskem obdobju« število ponavljajočih nakupov v kalibracijskem obdobju, »Čas med prvim in zadnjim nakupom« čas med prvim in zadnjim nakupom v kalibracijskem obdobju, »Vsota logaritmov obdobja med nakupi« vsoto logaritmov obdobja med nakupi (pomaga oceniti regularnost nakupov), »Št. dni med prvim nakupom in koncem kalibracijskega obdobja« število dni med prvim nakupom in koncem kalibracijskega obdobja, »Št. dni v validacijskem obdobju« število dni v validacijskem obdobju in »Št. nakupov validacijskem obdobju« število nakupov v validacijskem obdobju.

*Tabela 14: Pripravljene podatki*

Stranka	Št. ponavljajočih nakupov v kalibracijskem obdobju	Čas med prvim in zadnjim nakupom	Vsota logaritmov obdobja med nakupi	Št. dni med prvim nakupom in koncem kalibracijskega obdobja	Št. dni v validacijskem obdobju	Št. nakupov validacijskem obdobju
186523	7	147	17,6	159	184	3
186529	0	0	0	137	184	4
186538	0	0	0	145	184	0
186542	2	68	7,1	144	184	0

Stranka 186523, ki jo bom imenoval Nina, je na primer v prvem polletju leta 2014 opravila 8 nakupov (prvega in 7 ponovitev). Te nakupe je izvedla v 147 dneh, prvi nakup pa je opravila 25. dan leta (184 – 159). V validacijskem obdobju je izvedla 3 nakupe. V nadaljevanju bom lahko s pomočjo modelov spremljal, kolikšno število nakupov naj bi Nina opravila v validacijskem obdobju.

### 3.5.1 Uporaba modela Pareto/NBD

Prvi korak implementacije modela je ocenjevanje parametrov s pomočjo metode največjega verjetja (angl. *maximum likelihood estimation*) na kalibracijskih podatkih. Spomnimo,

model Pareto/NBD je mešanica dveh gama porazdelitev s parametri  $r$ ,  $\alpha$ ,  $s$  in  $\beta$ , katerih ocenjene vrednosti so v našem primeru sledeče:

$$r = 1,44267, \alpha = 4,316691 \text{ in} \\ s = 0,356792, \beta = 13,037282.$$

Paket BTYD vključuje tudi funkcije za optimizacijo parametrov, ki jih uporabimo, da preverimo, ali vrednosti konvergirajo. Tabela 15 prikazuje rezultat te funkcije na zgornjih parametrih, na osnovi katerega lahko zaključim, da sem našel optimalne parametre za naše podatke.

*Tabela 15: Optimizacija in konvergiranje parametrov*

$r$	$\alpha$	$s$	$\beta$	LL
1,144266991	4,3166909313	0,35679199473	13,037281704	-16662,73053975
1,144276659	4,3166355657	0,35681289544	13,036903118	-16662,73053915
1,442720254	4,3166614287	0,35681350297	13,036902673	-16662,73053908

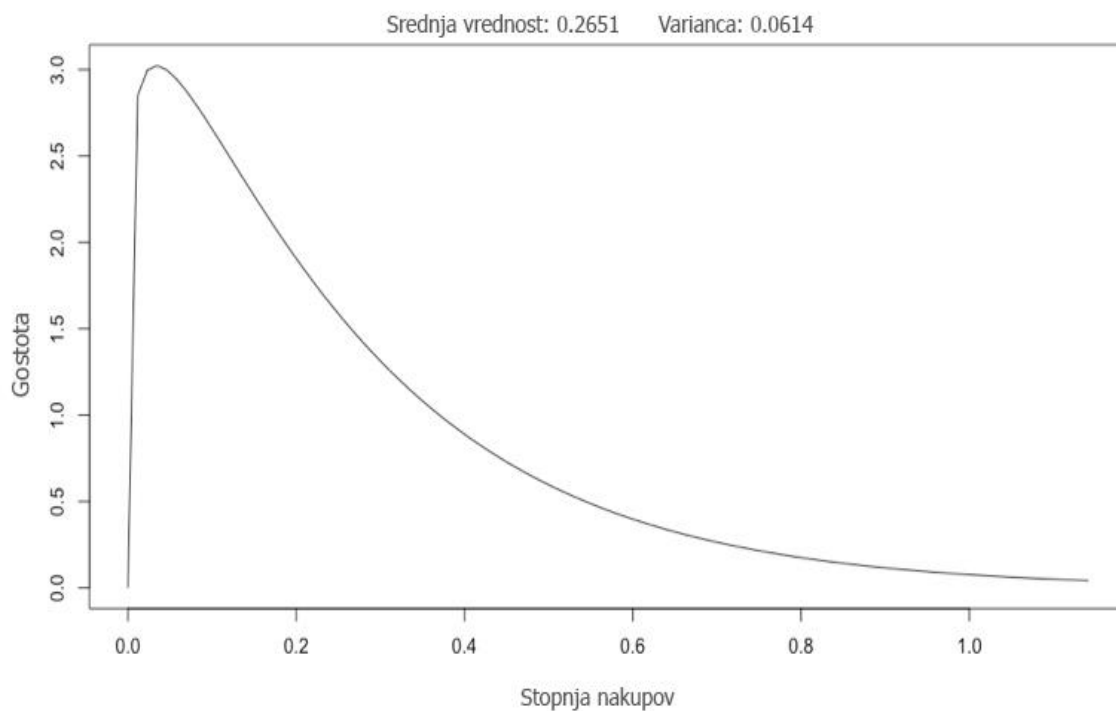
Parametra  $r$  in  $\alpha$  opisujeta gama porazdelitev NBD (Poisson) nakupnega procesa. Parametra  $s$  in  $\beta$  pa opisujeta gama porazdelitev Pareto procesa osipa. Obe gama distribuciji je mogoče predstaviti grafično (Slika 17 in Slika 18), njuno obliko pa primerjati z reprezentativnimi porazdelitvami, prikazanimi na Sliki 4. Medtem ko je parameter  $s$  precej manjši od nič in torej porazdelitev osipa ustreza levemu grafu na Sliki 4, je parameter  $r$  skoraj 1,5 in je porazdelitev nakupov bolj podobna desnemu grafu na Sliki 4. To pomeni, da je bolj verjetno, da imajo posamezne stranke majhne vrednosti parametrov osipa, parametri nakupnega procesa pa nakazujejo, da je malo verjetno, da stranke ne ponavljajo nakupov. To sem v raziskovalni analizi že zaznal z relativno nizkim deležem strank, ki so opravile samo en nakup.

Zgornje ugotovitve sem s pomočjo ocenjenih parametrov kvantificiral. Povprečni čas med dvema transakcijama ( $\alpha/r$ ) je 21 dni, povprečna življenjska doba stranke ( $\beta/s$ ) pa 259 dni. Predhodna analiza podatkov je pokazala, da je dejanski povprečni čas med transakcijami 44 dni in da model torej močno podcenjuje čas, ki preteče med dvema zaporednima nakupoma.

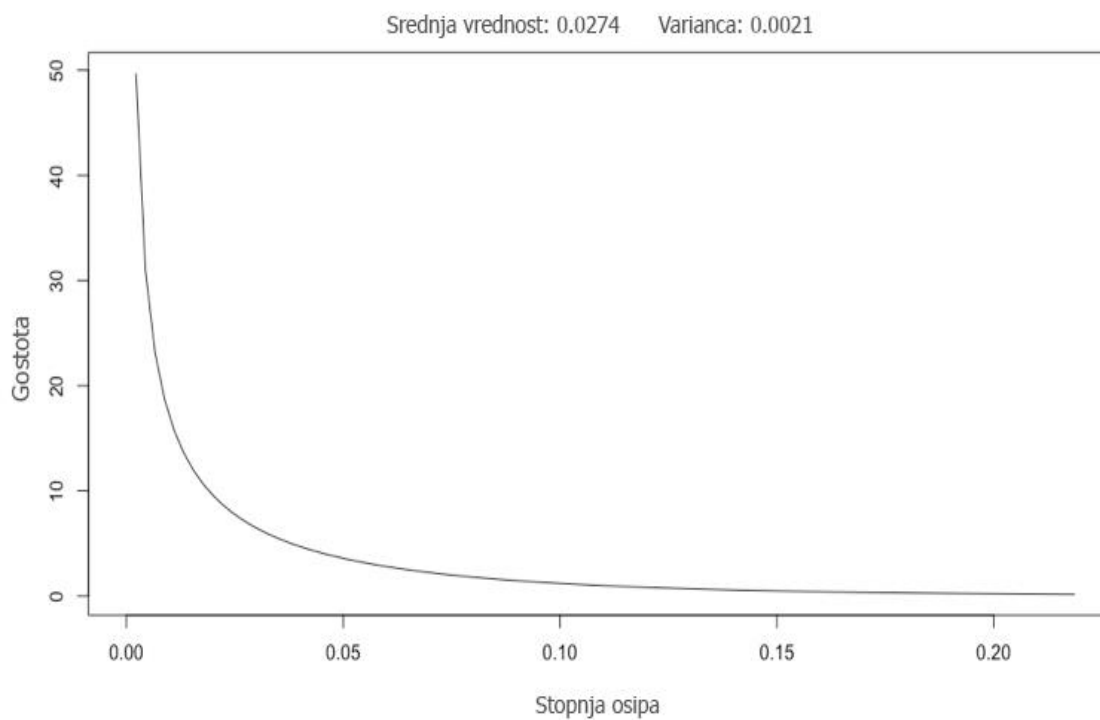
Ugotovljene vrednosti parametrov so mi omogočile uporabo napovednih funkcij modela. Na nivoju posamezne stranke je mogoče napovedati število nakupov, ki jih bo nova stranka opravila v izbranem obdobju. Model predvideva, da bi nova stranka v obdobju enega leta opravila 9,73 nakupa. Spomnimo se raziskovalne analize, kjer sem ugotovil, da je povprečna stranka opravila 10 nakupov. Glede na to, da ne poznam podrobnosti potencialnih novih strank, se zdi takšna predpostavka povsem verjetna.



*Slika 17: Heterogenost v nakupnem procesu*



*Slika 18: Heterogenost v procesu osipa*

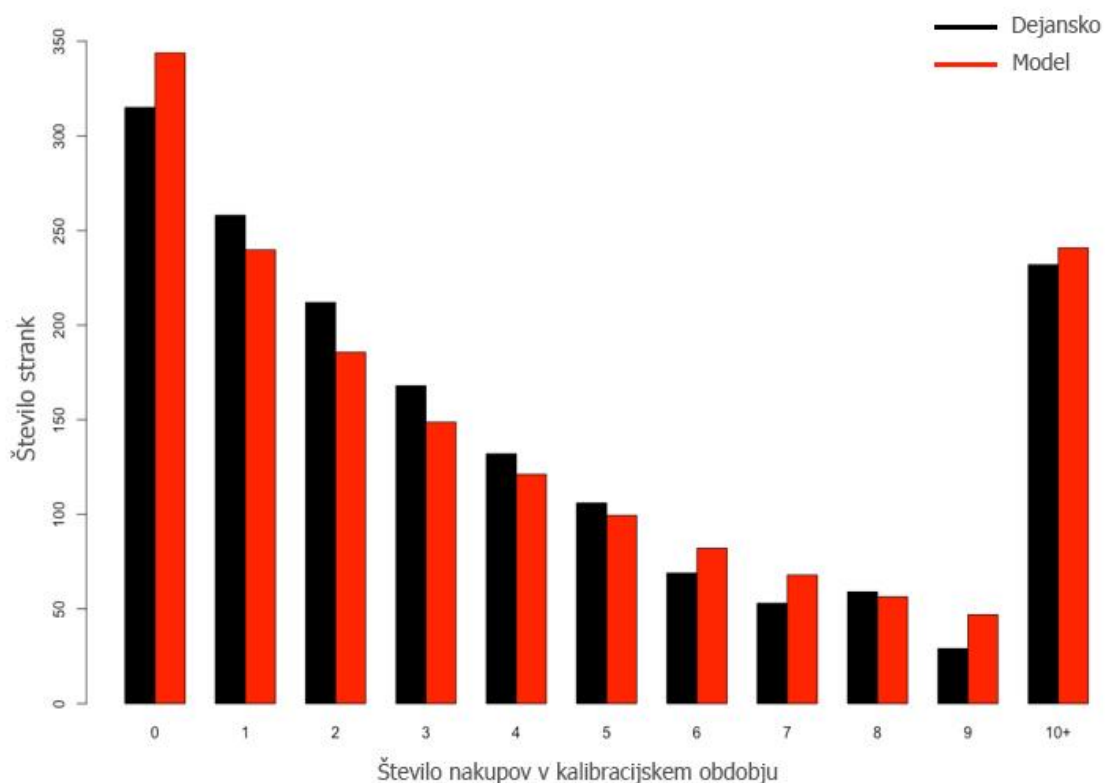


Model omogoča tudi napovedovanje prihodnjih nakupov obstoječih strank, ki temelji na podatkih kalibracijskega obdobja. Najprej model omogoča izračun verjetnosti, da je stranka po koncu kalibracijskega obdobja še aktivna. Če za primer vzamem stranko Nino, model

predvideva 97,72-odstotno verjetnost, da je po koncu kalibracijskega obdobja še aktivna. Prav tako model predvideva, da bo v validacijskem obdobju opravila 6,65 nakupa. Med procesom priprave podatkov se je izkazalo, da je Nina v validacijskem obdobju dejansko opravila 3 nakupe. Visoka verjetnost aktivnosti je torej pravilna, vendar pa je model zelo precenil prihodnje število nakupov.

Iz do sedaj povedanega je očitno, da model za posamezno stranko precenjuje prihodnje število nakupov in podcenjuje čas, ki mine med dvema nakupoma. Zanimalo pa me je, kako se to odraža na nivoju celotne baze strank, ki sem jo preučeval. Skladnost in točnost modela sta moja odločitvena kriterija. Najprej se bom posvetil skladnosti med pričakovanimi nakupi in dejanskimi nakupi v kalibracijskem obdobju, kot prikazuje Slika 19, s katere je razvidno, da skladnost napovedi z dejanskim stanjem ni najboljša, ni pa niti povsem zgrešena. Posebej dobro se odreže v primerih, ko so stranke opravile najmanj 10 nakupov (podjetje tam ustvarja največ prometa), najslabše pa zna določiti stranke, ki opravijo samo prvi osnovni nakup. Glede na to, da sem v fazi raziskovalne analize ugotovil, da je ugotavljanje neaktivnosti stranke zelo nespecifično, tak rezultat ni presenetljiv.

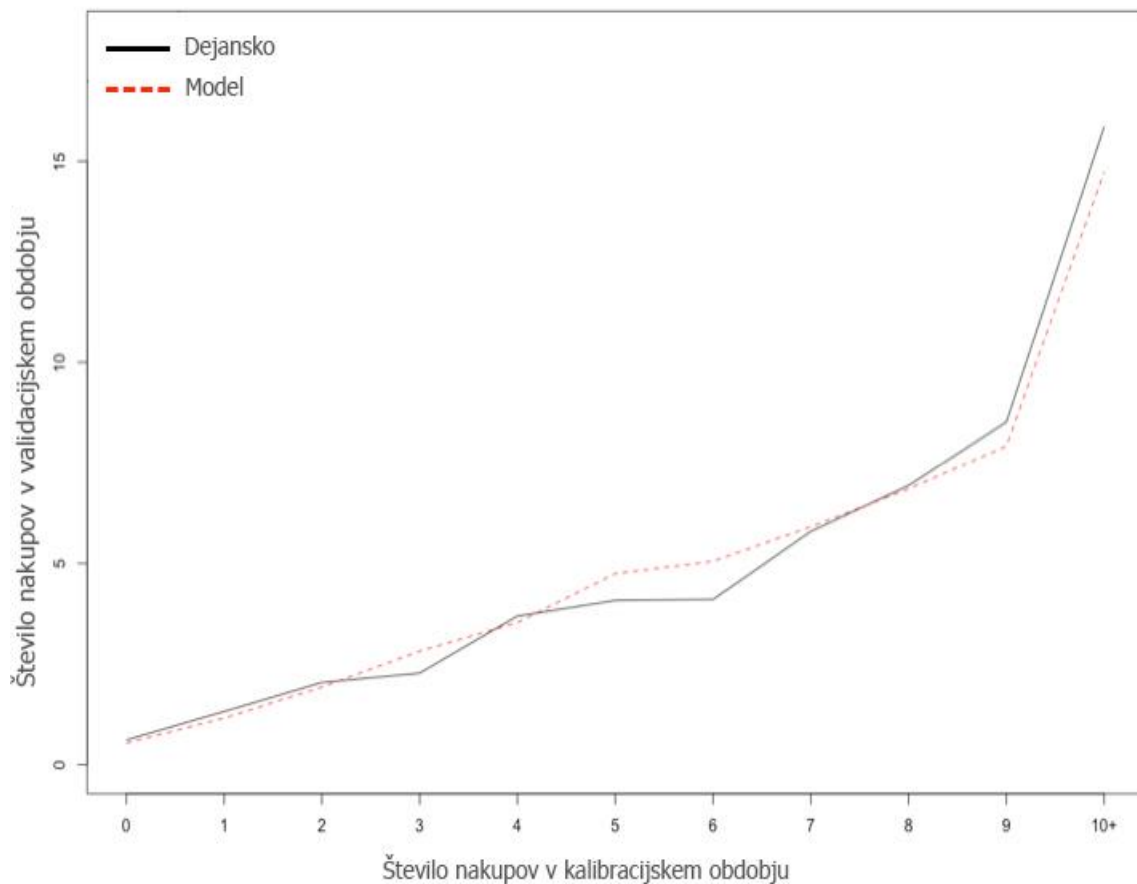
*Slika 19: Pogostost ocenjenih in dejanskih ponovnih nakupov v kalibracijskem obdobju v modelu Pareto/NBD*



Slika 19 kaže podatke samo za kalibracijsko obdobje. Za ocenjevanje modela pa je ključno ugotoviti, ali je model sposoben točno napovedati dogajanje v prihodnosti, torej v

validacijskem obdobju. V ta namen sem podatke na začetku razdelil na kalibracijski in validacijski del. Zdaj lahko napovedi modela v validacijskem obdobju primerjam z dejanskimi podatki iz tega obdobja. Na Sliki 20 sem primerjal število transakcij v validacijskem obdobju na osi y in število transakcij v kalibracijskem obdobju na osi x. Model je v določenih primerih precejšnje število transakcij (skupine 3, 5 in 6), medtem ko je pri ostalih skupinah napovedoval relativno dobro.

*Slika 20: Pričakovani nakupi v validacijskem obdobju glede na število nakupov v kalibracijskem obdobju v modelu Pareto/NBD*



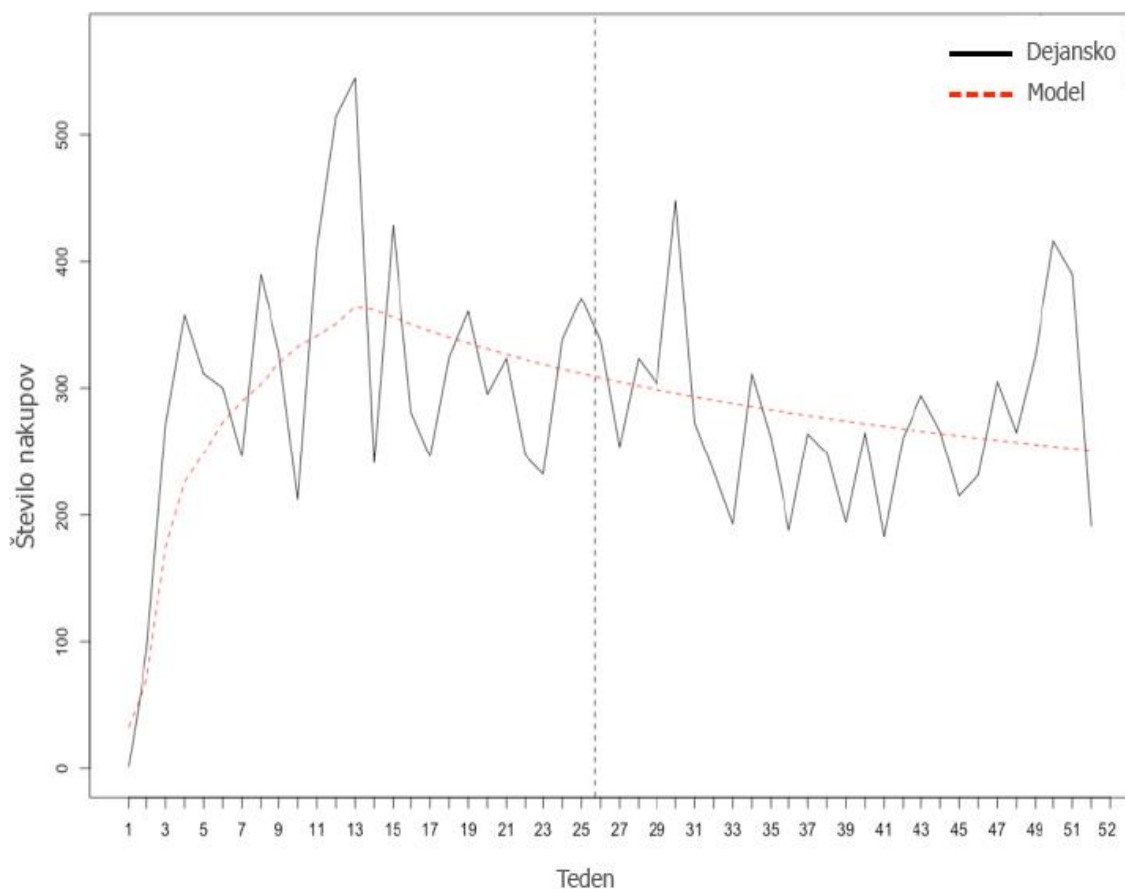
Te podatke sem prikazal tudi v Tabeli 16 in dodal število strank v posamezni skupini. Na ta način je lepo vidna pomembnost posameznih skupin. Razlike v skupini 0, kjer je 315 strank, so na primer veliko bolj pomembne kot razlike v skupini 7, ki vključuje samo 53 strank.

Tabela 16: Dejanski in pričakovani povprečni nakupi ter število strank po skupinah v modelu Pareto/NBD

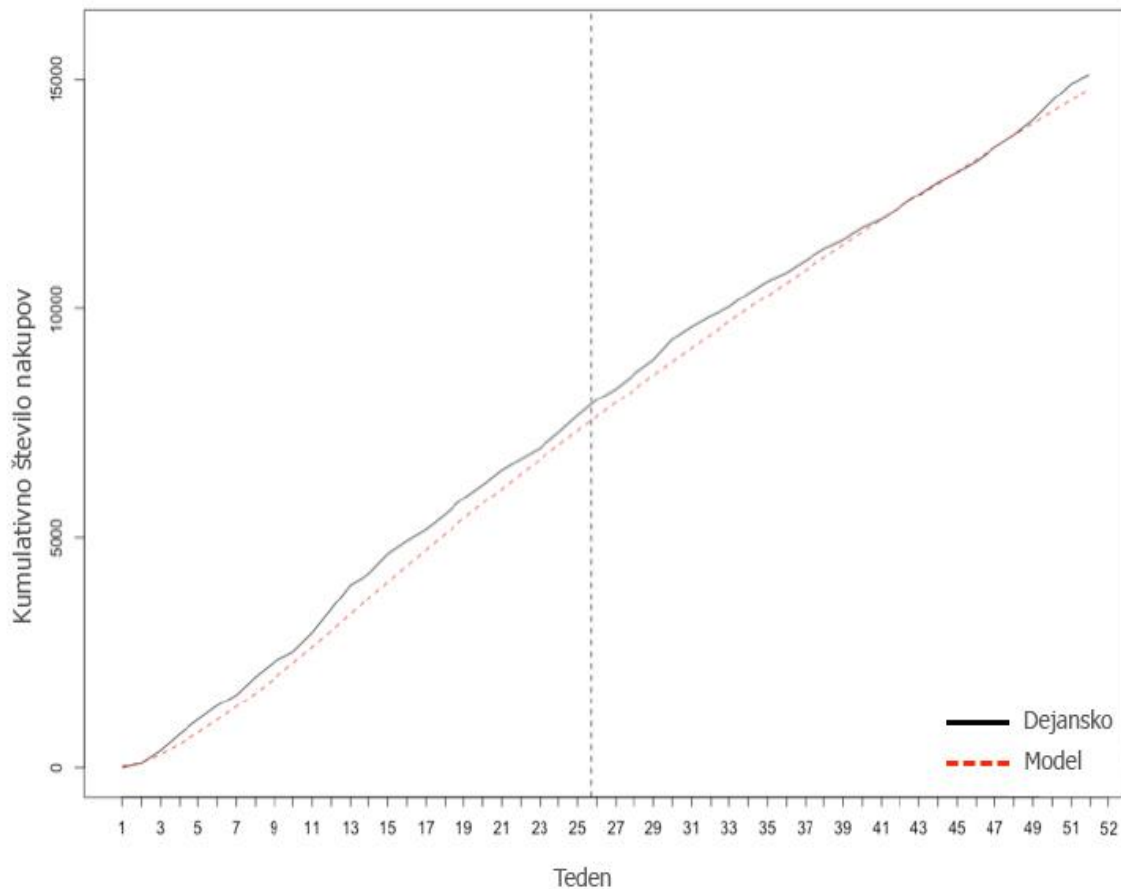
Skupina	0	1	2	3	4	5	6	7	8	9	10+
Dejanski nakupi	0,61	1,32	2,04	2,27	3,69	4,08	4,10	5,79	6,93	8,52	15,86
Pričakovani nakupi	0,54	1,16	1,93	2,82	3,53	4,74	5,05	5,91	6,85	7,91	14,74
Število strank	315	258	212	168	132	106	69	53	59	29	232

Nadalje sem lahko podatke o dejanskih in pričakovanih nakupih primerjal v času. Slika 21 prikazuje in primerja gibanje števila dejanskih in napovedanih nakupov po tednih v opazovanem letu. Navpična prekinjena črta označuje mejo med kalibracijskim in validacijskim obdobjem. Dejanske transakcije so zelo variabilne, vendar pa model zelo dobro zajame njihov trend. Kljub temu je takšen prikaz nezadosten in pomanjkljiv, ker nakupi preveč variirajo v času, to pa onemogoči kakovostno oceno rezultatov modela. Boljša rešitev je spremljanje celotnih kumulativnih nakupov, kot je to prikazano na Sliki 22, s katere je razvidno, da model večinoma podcenjuje skupno število nakupov.

Slika 21: Gibanje ocenjenih in dejanskih nakupov po tednih v modelu Pareto/NBD



Slika 22: Gibanje ocenjenih in dejanskih kumulativnih nakupov po tednih v modelu Pareto/NBD



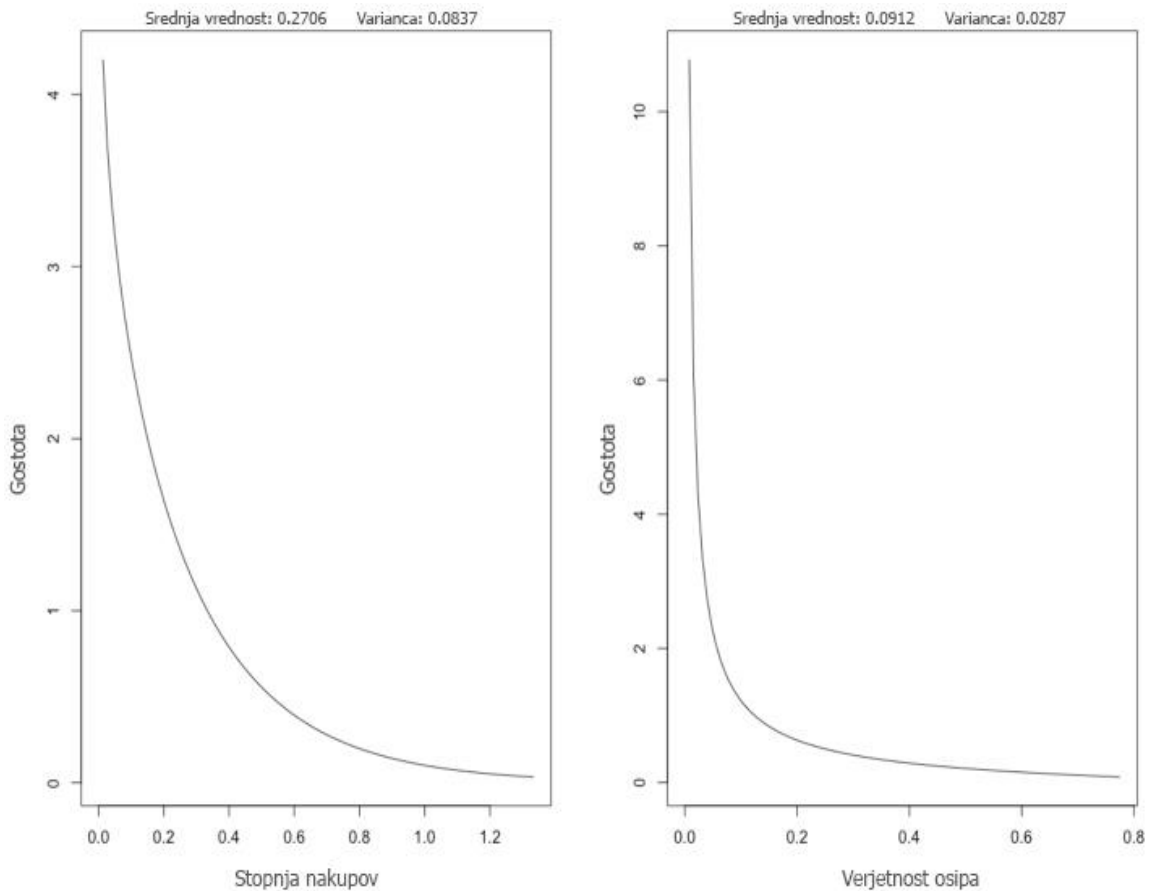
### 3.5.2 Uporaba modela BG/NBD

Model BG/NBD je mešanica gama porazdelitve s parametri  $r$  in  $\alpha$  in beta porazdelitve s parametroma  $a$  in  $b$ , katerih ocenjene vrednosti so v mojem primeru sledeče (seveda sem ponovno uporabil funkcije za optimizacijo, katerih rezultatov pa tu ne prikazujem):

$$r = 0,8748876, \alpha = 3,2326066 \text{ in} \\ a = 0,1725999, b = 1,7190372.$$

Porazdelitev nakupov  $\alpha$  se ne razlikuje občutno od modela Pareto/NBD, kar je bilo glede na njuno tesno povezanost tudi pričakovati. Povprečni čas med transakcijama je ocenjen na 25 dni, kar je še vedno precej manj, kot kažejo dejanski podatki. Porazdelitev stopnje osipa  $p$  v modelu BG/NBD sledi beta porazdelitvi. Ta stopnja  $p$  variira okrog aritmetične sredine ( $a / (a + b)$ ) pri 9,12 %, kar predstavlja 149 strank. Raziskovalna analiza je pokazala, da je 13,7 % (oziroma 224) strank opravilo samo en nakup. Razlika med parametri modela in dejanskimi podatki je tudi tukaj očitna. Obe porazdelitvi parametrov sta prikazani na Sliki 23.

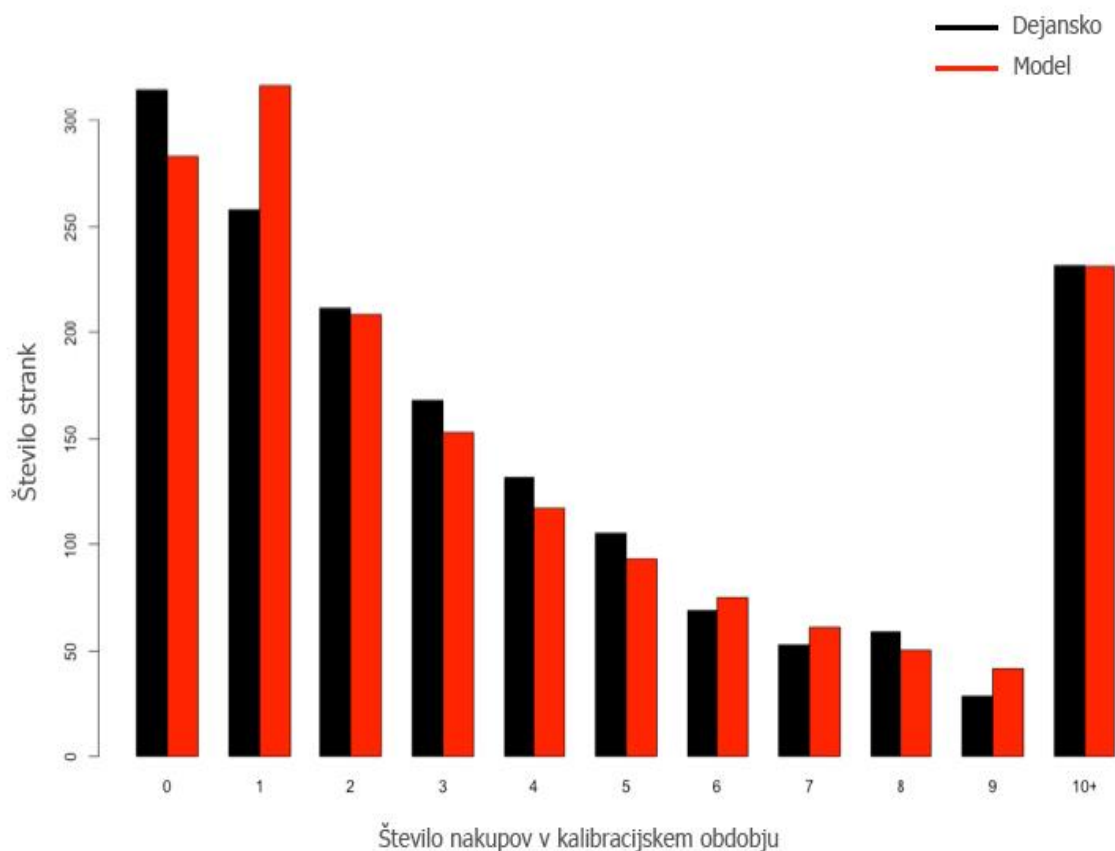
Slika 23: Porazdelitev stopnje nakupov  $\alpha$  in stopnje osipa  $p$  v modelu BG/NBD



Po določitvi parametrov sem lahko ponovno napovedal število nakupov posameznih strank. Model predvideva, da bi nova stranka v obdobju enega leta opravila 10,03 nakupa, kar je nekoliko več, kot je napovedal model Pareto/NBD (9,73), in je bližje temu, kar sem ugotovil z raziskovalno analizo (10). Če podrobneje pogledamo stranko Nina, model BG/NBD predvideva 96,31-odstotno verjetnost, da je Nina po koncu kalibracijskega obdobja aktivna, in vemo, da to drži. Model predvideva, da bo Nina v validacijskem obdobju opravila 6,85 nakupa, kar je ponovno nekoliko več kot v modelu Pareto/NBD (6,65) in precej več od treh nakupov, ki jih je Nina dejansko opravila.

Skladnost med pričakovanimi in dejanskimi nakupi v kalibracijskem obdobju je prikazana na Sliki 24. Vidimo, da je model izredno dober, ko govorimo o več kot desetih transakcijah. Po drugi strani pa precej podcenjuje število strank brez ponovnih nakupov in močno precenjuje število strank z enim samim ponovnim nakupom. To potrjuje moje predhodne ugotovitve glede stopnje osipa, ki je nižja od dejansko ugotovljene.

Slika 24: Pogostost ocenjenih in dejanskih ponovnih nakupov v kalibracijskem obdobju v modelu BD/NBD

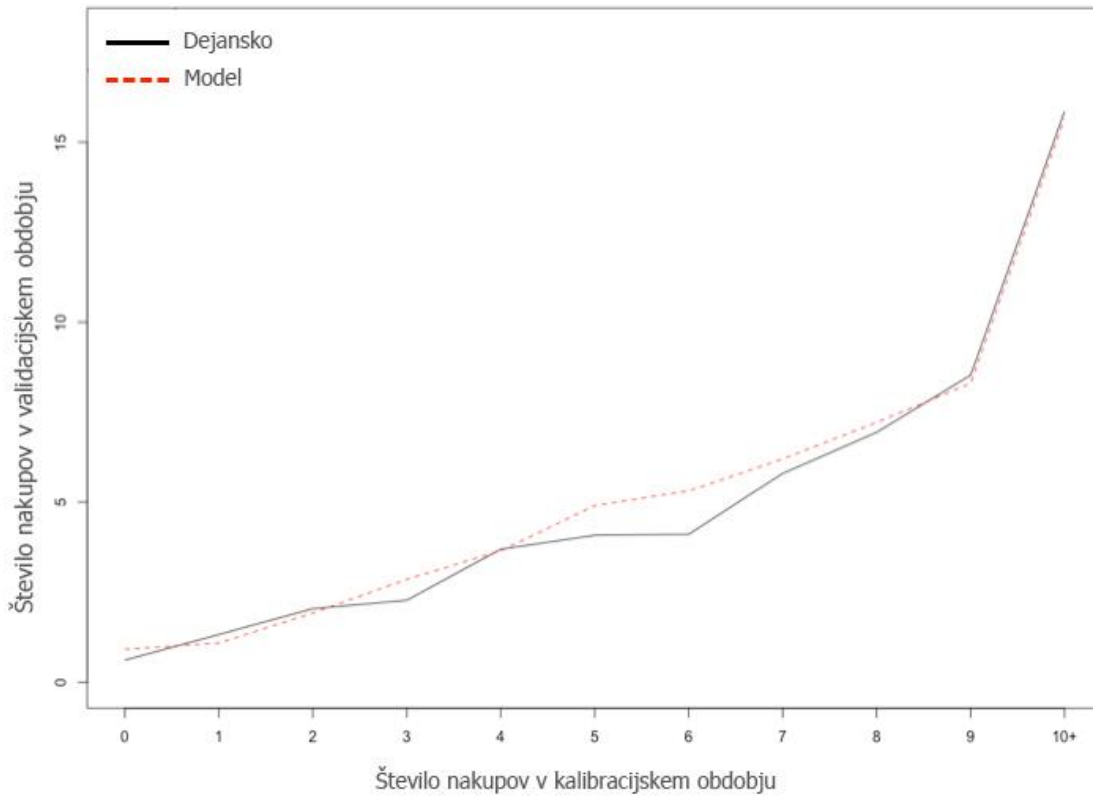


Na Sliki 25 sem primerjal število nakupov v validacijskem obdobju na osi y in število nakupov v kalibracijskem obdobju na osi x. Rezultati so zelo podobni kot pri modelu Pareto/NBD, prikazanem na Sliki 20, s to razliko, da model bolje napoveduje pri deset in več nakupih in slabše, ko gre za en nakup ali pa ponavljajočih se nakupov celo ni. Isti podatki so prikazani tudi v Tabeli 17.

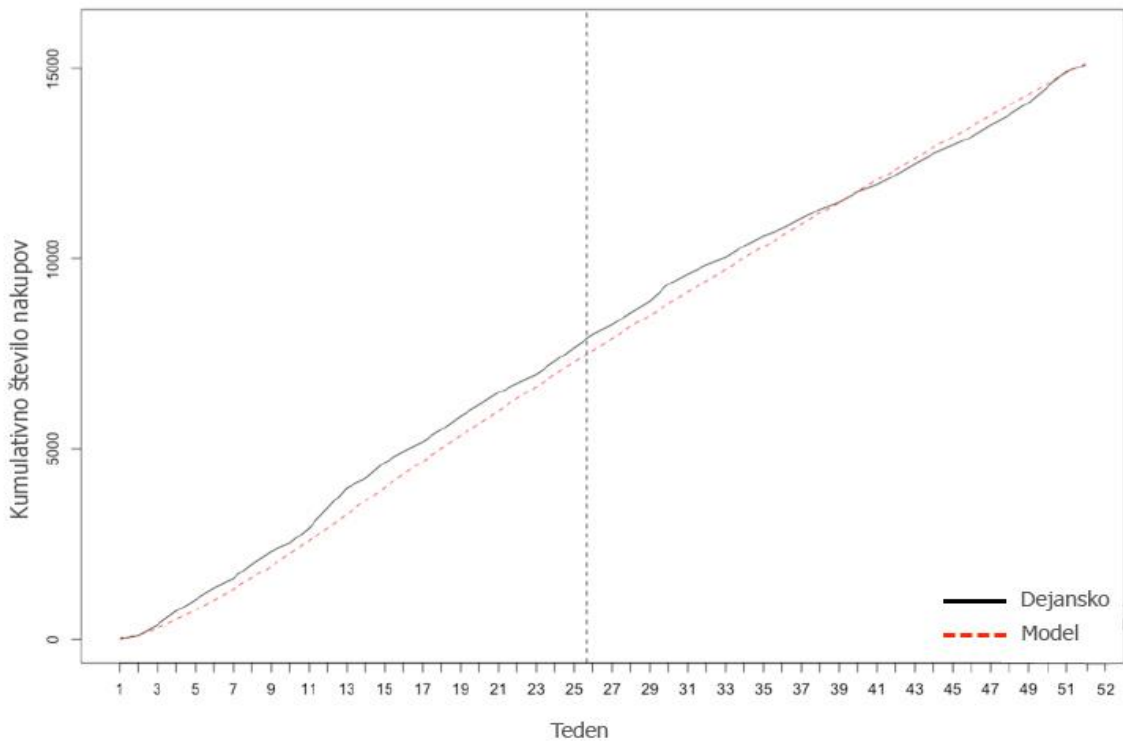
Tabela 17: Dejanski in pričakovani povprečni nakupi ter število strank po skupinah v modelu BD/NBD

Skupina	0	1	2	3	4	5	6	7	8	9	10+
Dejanski nakupi	0,61	1,32	2,04	2,27	3,69	4,08	4,10	5,79	6,93	8,52	15,86
Pričakovani nakupi	0,92	1,09	1,92	2,86	3,65	4,90	5,31	6,19	7,21	8,29	15,66
Število strank	315	258	212	168	132	106	69	53	59	29	232

Slika 25: Pričakovani nakupi v validacijskem obdobju glede na število nakupov v kalibracijskem obdobju v modelu BD/NBD



Slika 26: Gibanje ocenjenih in dejanskih kumulativnih nakupov po tednih v modelu BG/NBD





Prav tako kot za model Pareto/NBD lahko tudi za model BG/NBD rečem, da večinoma podcenjuje kumulativno število nakupov, kot je to razvidno s Slike 26. Zanimivo je videti, da je napoved skupnega števila nakupov na koncu leta zelo blizu dejanskemu številu nakupov. Kljub vidnim razlikam med letom je torej model sposoben relativno točno napovedati skupno število nakupov v celotnem letu.

### 3.5.3 Uporaba modela CBG/NBD

Model CBG/NBD uvaja dodatno možnost, da nas stranka zapusti v času 0, torej takoj ko opravi prvi nakup. Ta sprememba ima pomembne posledice in je precej bolj realistična. Predhodni modeli so vsako stranko, ki je opravila prvi nakup, šteli kot aktivno. Dejansko pa lahko stranke povsem legitimno opravijo samo en nakup. Raziskovalna analiza je pokazala, da je delež strank, ki so opravile samo en nakup, relativno nizek (13,7 %), zato bo zanimivo videti, kako ta sprememba vpliva na skladnost in točnost modela.

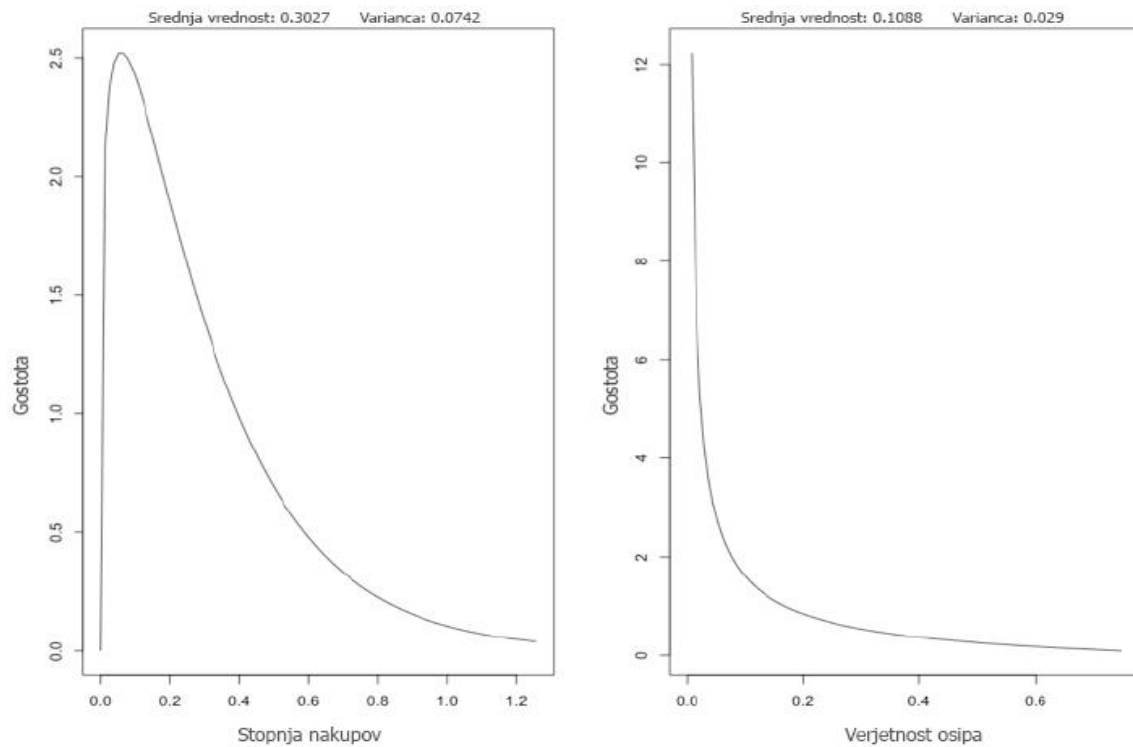
Model CBG/NBD je ponovno mešanica gama porazdelitve s parametri  $r$  in  $\alpha$  in beta porazdelitve s parametroma  $a$  in  $b$ . Ocenjene vrednosti parametrov po optimizaciji so sledeče:

$$r = 1,2348841, \alpha = 4,0801686 \text{ in} \\ a = 0,2544955, b = 2,0853499.$$

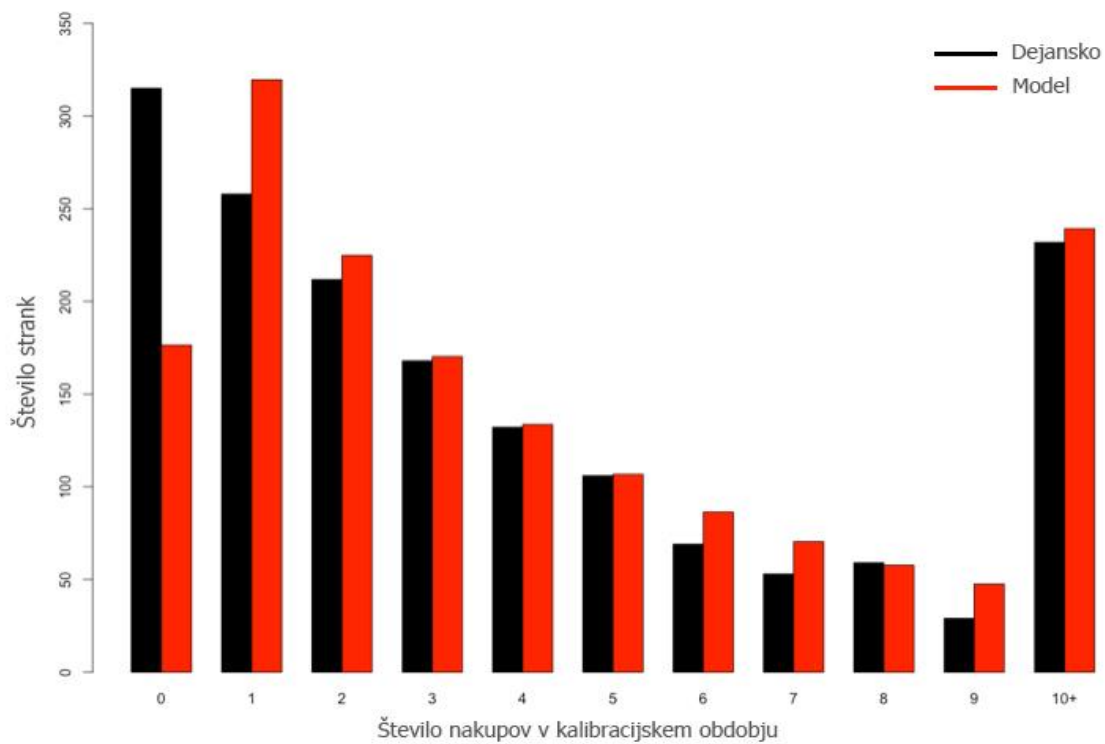
Povprečni čas med nakupi je ocenjen na 23 dni, kar je še vedno precej manj, kot kažejo dejanski podatki. Porazdelitev stopnje osipa  $p$  je ocenjena na 10,88 %, kar predstavlja 178 strank. Iz raziskovalne analize vemo, da je 13,7 % strank opravilo samo en nakup. Zdi se torej, da ocenjeni parametri v modelu CBG/NBD bolje odražajo dejanske podatke kot na primer v modelu BG/NBD. Glede na dodatno predpostavko modela CBG/NBD je bilo tudi pričakovati, da bo ocenjevanje strank z opravljenim enim nakupom bolj točno. Porazdelitvi parametrov stopnje nakupov in stopnje osipa sta prikazani na Sliki 27.

Za stranko Nino model CBG/NBD predvideva 95,39-odstotno verjetnost, da bo po koncu kalibracijskega obdobja še aktivna. Število pričakovanih nakupov je 6,75. Te napovedi so ponovno zelo podobne modeloma Pareto/NBD in BG/NBD in še vedno precej odstopajo od dejanskih nakupov, ki jih je Nina opravila. Slika 28 prikazuje ocenjene in dejanske nakupe v kalibracijskem obdobju. Za celotno bazo strank model močno podcenjuje število strank, ki so opravile samo prvi nakup, in po drugi strani zelo precenjuje število strank, ki so opravile vsaj en dodaten nakup.

Slika 27: Porazdelitev stopnje nakupov  $\alpha$  in stopnje osipa  $p$  v modelu CBG/NBD



Slika 28: Pogostost ocenjenih in dejanskih ponovnih nakupov v kalibracijskem obdobju v modelu CBG/NBD

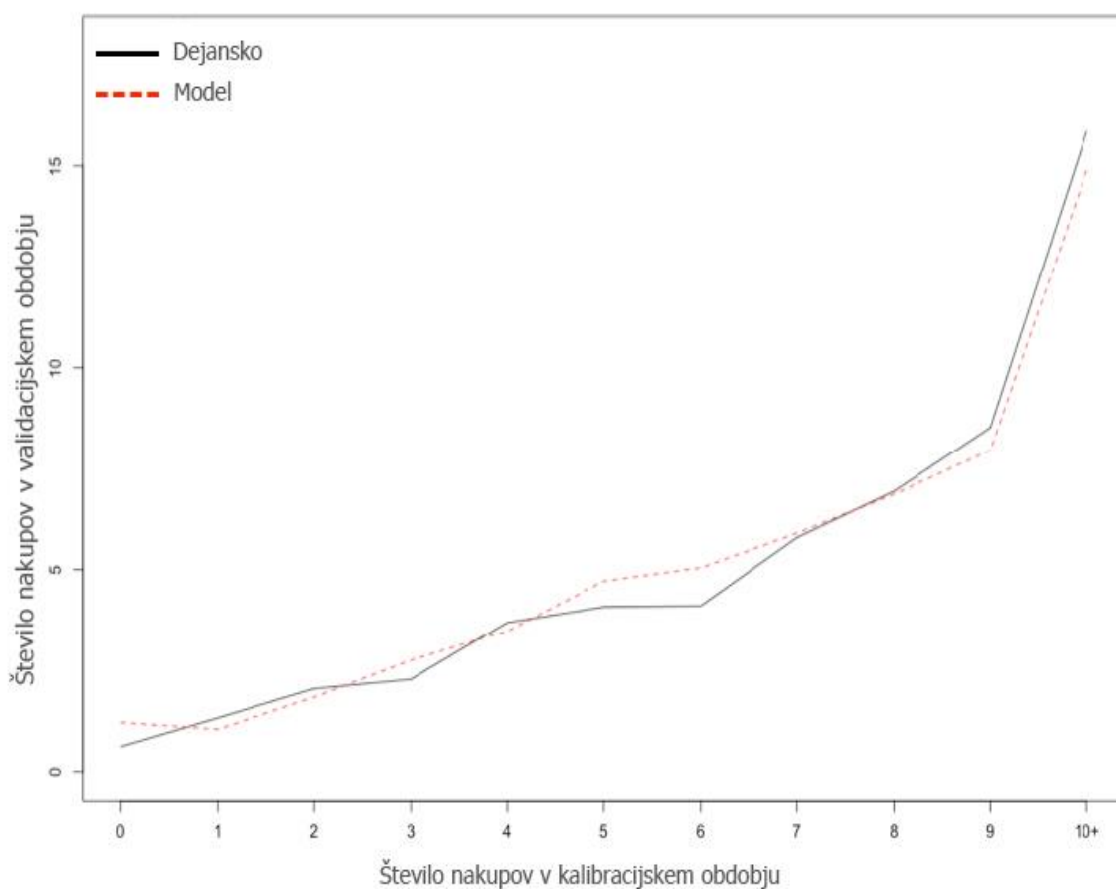


Na Sliki 29 prikazujem primerjavo število transakcij v validacijskem obdobju na osi y in število transakcij v kalibracijskem obdobju na osi x. Opazna so precejšnja odstopanja in zdi se, da je model CBG/NBD slabša izbira kot na primer BD/NBD. Podatki so prikazani tudi v Tabeli 18, kjer so razlike med dejanskim in napovedanim stanjem še bolj jasno vidne. Končno lahko na Sliki 30 prikažem gibanje celotnih kumulativnih nakupov. Za razliko od modela BD/NBD, ki je dobro zadel število celotnih nakupov konec leta, na Sliki 26, model CBG/NBD te očitno precenjuje.

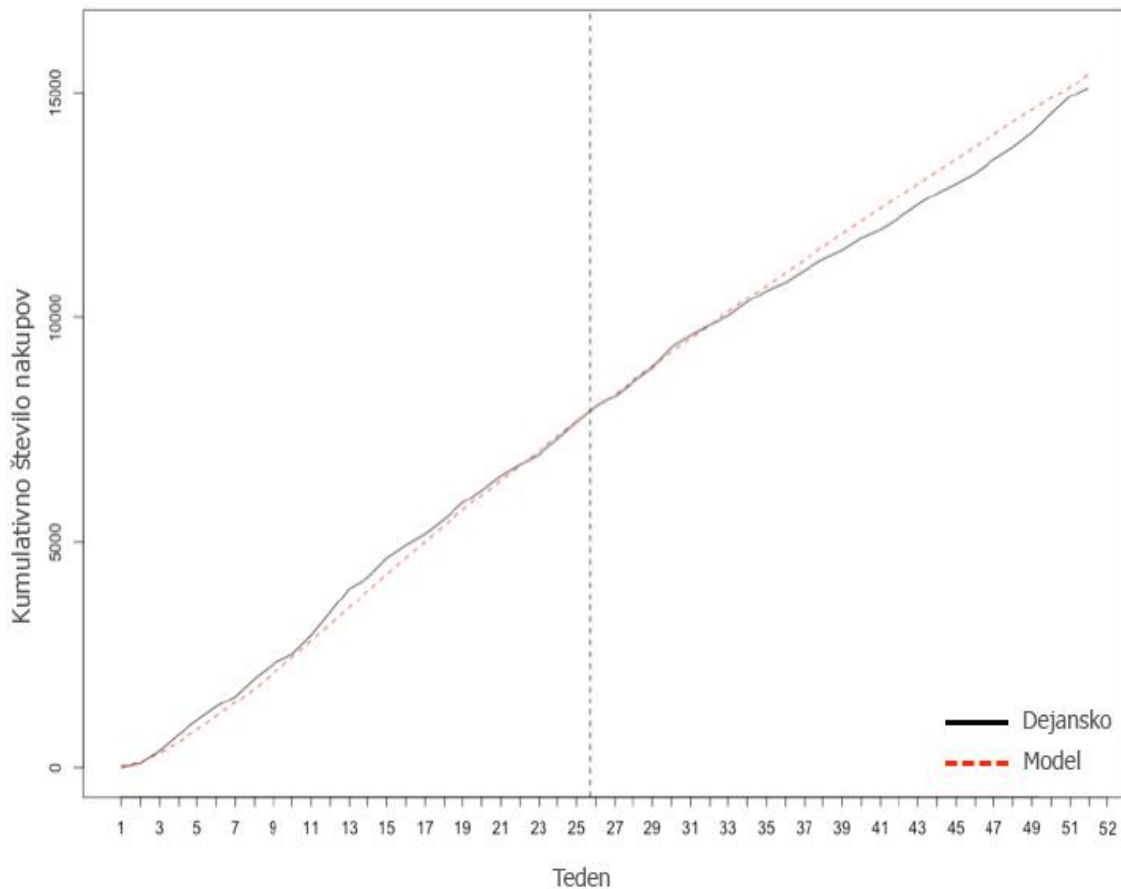
*Tabela 18: Dejanski in pričakovani povprečni nakupi ter število strank po skupinah v modelu CBG/NBD*

Skupina	0	1	2	3	4	5	6	7	8	9	10+
Dejanski nakupi	0,61	1,32	2,04	2,27	3,69	4,08	4,10	5,79	6,93	8,52	15,86
Pričakovani nakupi	1,21	1,04	1,83	2,75	3,48	4,71	5,04	5,90	6,86	7,95	14,90
Število strank	315	258	212	168	132	106	69	53	59	29	232

*Slika 29: Pričakovane transakcije v validacijskem obdobju glede na število transakcij v kalibracijskem obdobju v modelu CBG/NBD*



Slika 30: Gibanje ocenjenih in dejanskih kumulativnih transakcij po tednih v modelu CBG/NBD



### 3.5.4 Uporaba modela CBG/NBD-k

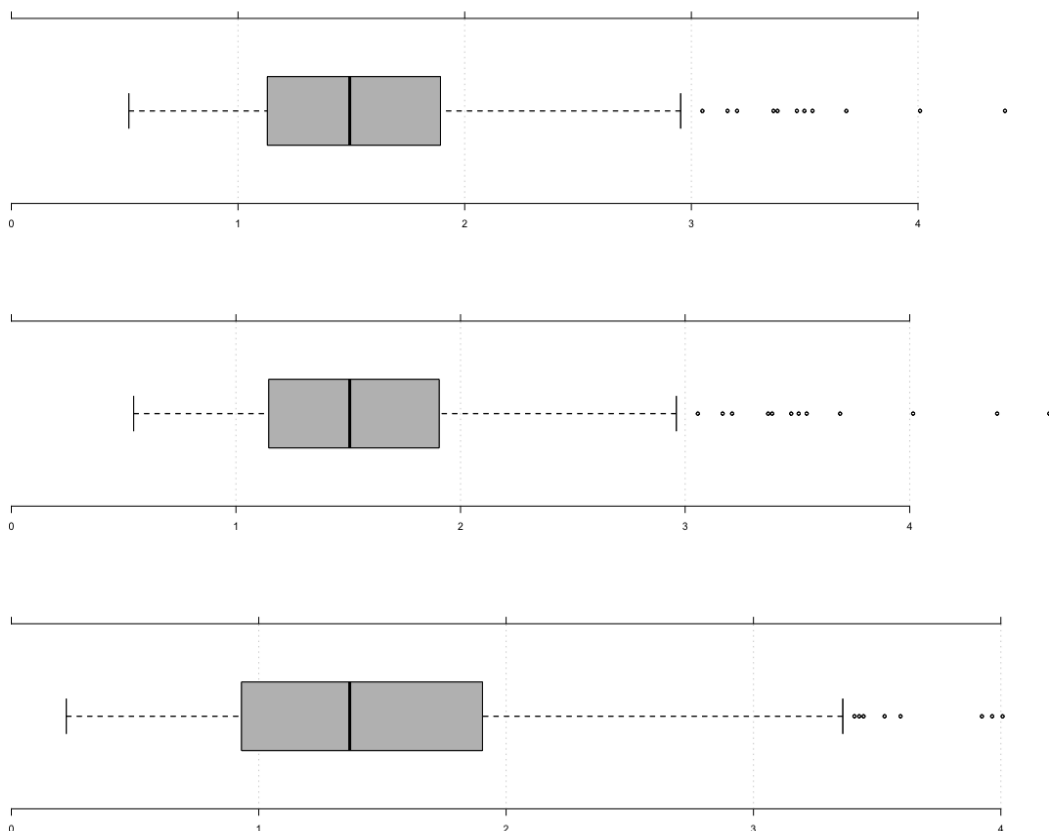
Zadnji uporabljeni model je CBG/NBD-k, ki se posveča rednosti nakupov strank. Dejstvo je, da imajo izdelki, ki jih prodaja Black Diamond, določeno življenjsko dobo. Prav tako se lahko obrabijo in načeloma velja, da je opremo po določenem času oziroma ob veliki obrabi dobro tudi zamenjati. Predpostavim torej lahko, da posamezne stranke izkazujejo določeno rednost. Model namesto naključnih nakupov, ki sledijo eksponentni (Poisson) porazdelitvi, predpostavlja, da si nakupi redno sledijo v obliki Erlangove porazdelitve. Erlangova porazdelitev je posebna oblika gama porazdelitve in vključuje parameter oblike  $k$ , ki ga moramo izbrati sami. Če je parameter  $k = 1$ , potem je model CBG/NBD-k enak modelu CBG/NBD.

Prva naloga pri uporabi modela CBG/NBD-k je torej določitev vrednosti parametra  $k$ . Ta lahko zavzame vrednost kateregakoli pozitivnega celega števila. Kot rečeno, vrednost 1 pomeni, da ni mogoče opaziti rednosti nakupov oziroma so nakupi povsem naključni. Vrednosti, večje od 1, pa nakazujejo določeno rednost. Paket BTYDplus vključuje funkcijo, ki s pomočjo dveh različnih metod olajša izbiro parametra rednosti. Prva metoda

predpostavlja, da je parameter rednosti enak za vse stranke, medtem ko vse ostale metode rednost ocenjujejo na nivoju vsake posamezne stranke in za parameter izberejo srednjo vrednost vseh posameznih ocen.

Za zanesljivo oceno parametra  $k$  na nivoju posamezne stranke potrebujemo zadostno število (najmanj 5) opravljenih transakcij. V nasprotnem primeru so rezultati lahko pristranski. Raziskovalna analiza je pokazala, da so stranke v povprečju izvedle 10 nakupov, vendar pa je precej takih, ki so jih opravile manj kot 5. Konkretno imamo 858 strank, ki so opravile več kot 5 nakupov, kar predstavlja 52 % vseh strank. Te stranke so skupaj opravile 14.747 nakupov. Slika 31 prikazuje porazdelitev ocenjenega  $r$  parametra oblike, ki sem ga uporabil kot osnovo za določitev parametra  $k$ . Srednja vrednost se nahaja nekje okrog 1,5, z relativno velikim razponom in precej odstopanji. Na podlagi tega sem vrednost parametra rednosti  $k$  določil pri 2.

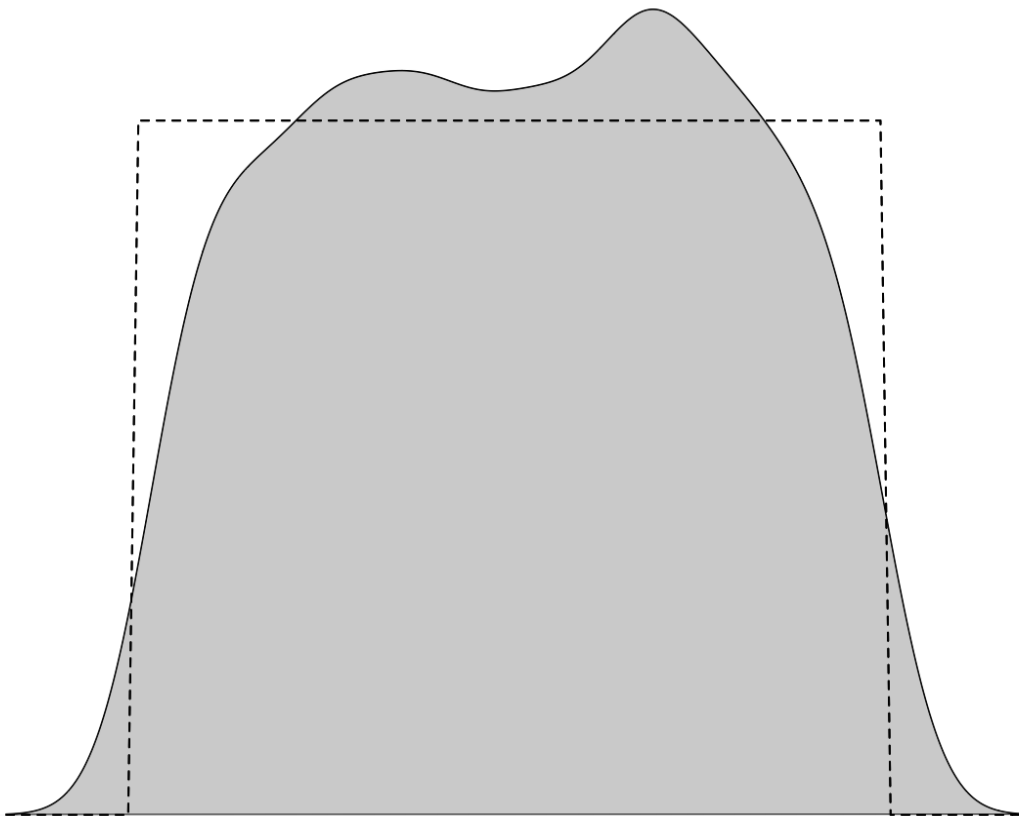
*Slika 31: Porazdelitev ocenjenega parametra oblike za stranke, ki so opravile najmanj 5 nakupov*



Druga metoda za določitev parametra rednosti predpostavlja, da je ta za vse stranke enak. Ta metoda je enostavnejša, saj ni potrebe, da imamo vsaj 5 transakcij za vsako posamezno stranko. Potrebne so namreč samo 3, kar v mojem primeru pomeni 1092 strank oziroma 66,8

% celotne baze strank. Slika 32 prikazuje zglajen histogram mere regularnosti celotne baze strank. Dodatno je prikazana teoretična porazdelitev, ko je mera regularnosti enaka 1. Ocena regularnosti po tej metodi znaša 1,172278, kar bi pomenilo, da v podatkih ni zaznati regularnosti nakupov.

Slika 32: Zglajen histogram mere regularnosti za stranke, ki so opravile najmanj 3 nakupe



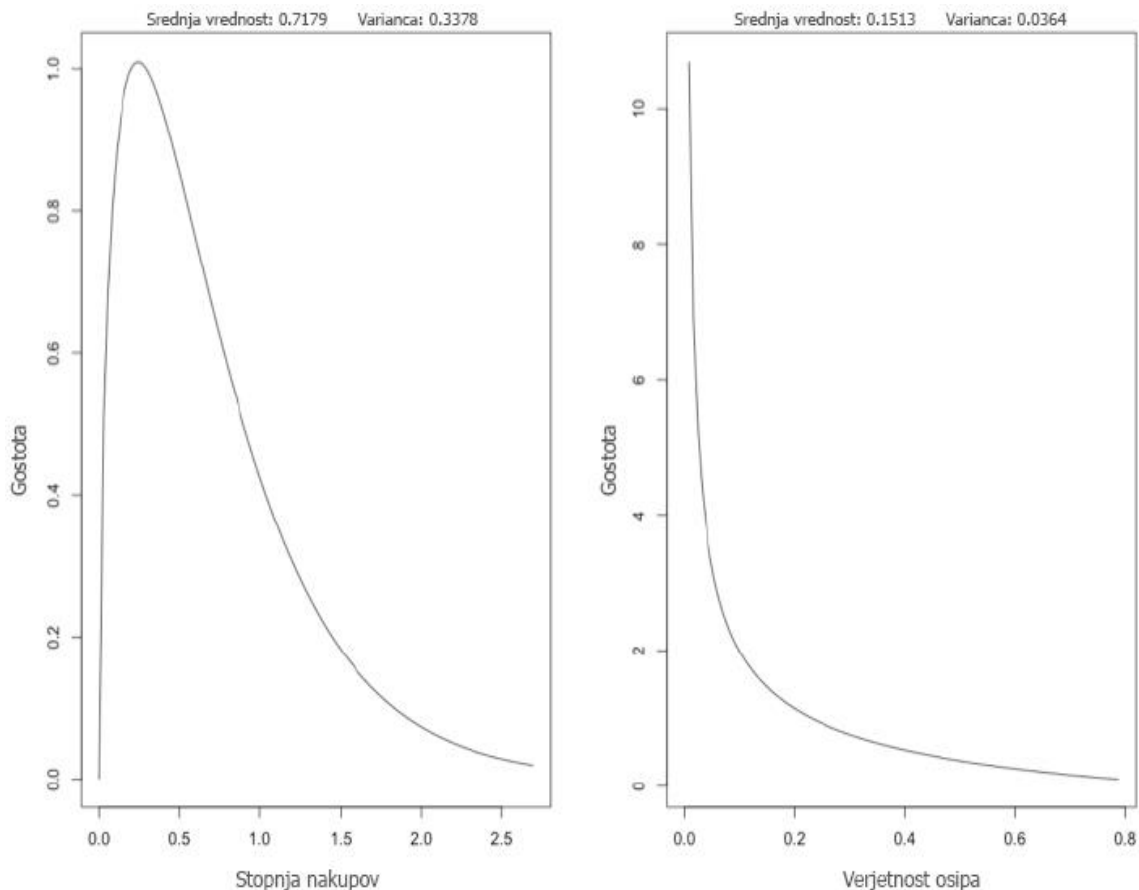
Dve metodi torej ponudita dve različni rešitvi. Glede na to, da lahko zaradi narave izdelkov pričakujemo določeno regularnost pri lojalnih strankah, sem za parameter  $k$  določil 2, model pa označil kot CBG/CNBD-2. Ob tem pa ne smemo pozabiti, da je, vsaj glede na podatke, ta regularnost dokaj vprašljiva.

Ko imamo določen parameter  $k$ , lahko ocenimo tudi ostale parametre modela. V osnovi gre še vedno za mešanico gama porazdelitve s parametri  $r$  in  $\alpha$  ter beta porazdelitve s parametroma  $a$  in  $b$ . Ocenjene vrednosti parametrov po optimizaciji so sledeče:

$$\begin{aligned} r &= 1,5254085, \alpha = 2,1249471 \text{ in} \\ a &= 0,3828819, b = 2,1477801 \text{ in} \\ k &= 2. \end{aligned}$$

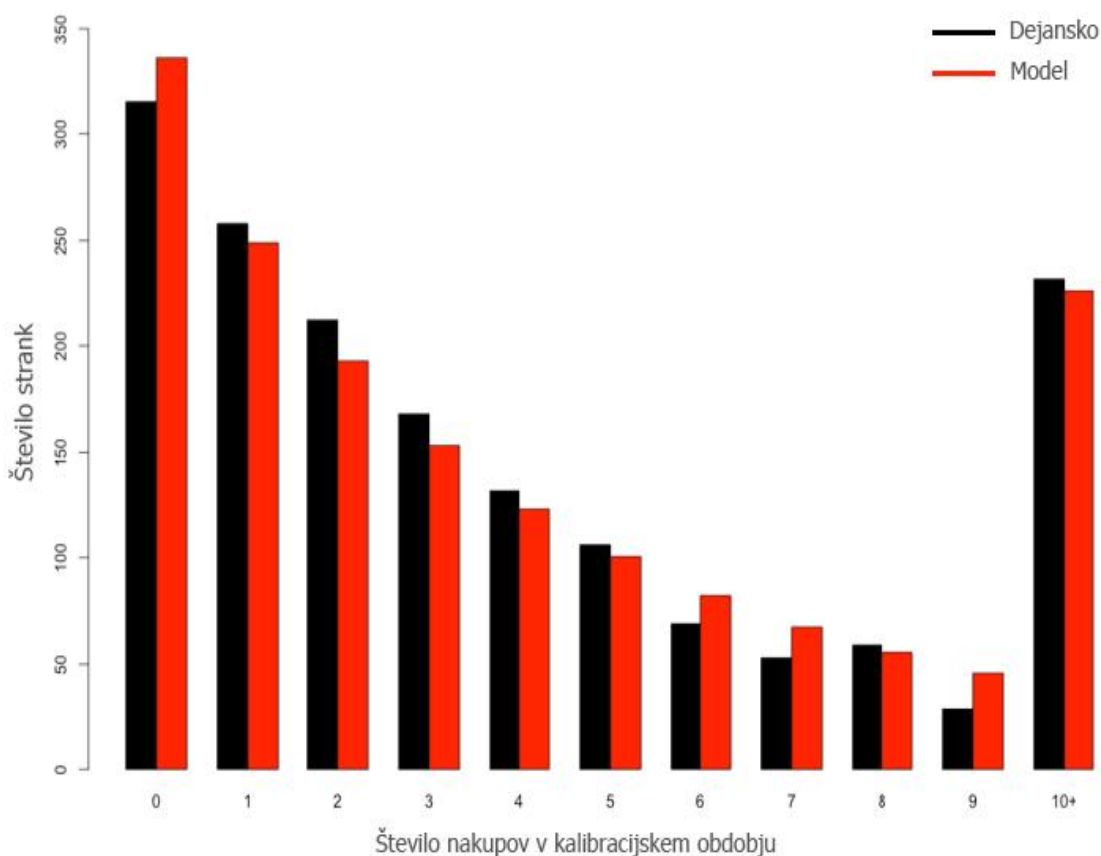
Povprečni čas med transakcijama je v modelu ocenjen na 20 dni, kar je še veliko manj, kot kažejo dejanski podatki, in precej manj kot v ostalih modelih. Porazdelitev stopnje osipa  $p$  je ocenjena na 15,13 %, kar predstavlja 247 strank. Model torej celo precenjuje število strank s samo enim nakupom, vendar pa je bližje dejanskemu številu teh strank kot ostali trije do sedaj obravnavani modeli. Porazdelitvi parametrov stopnje nakupov in stopnje osipa sta prikazani na Sliki 33.

Slika 33: Porazdelitev stopnje nakupov  $\alpha$  in stopnje osipa  $p$  v modelu CBG/CNBD-2



Za Nino model CBG/CNBD-2 predvideva 94,22-odstotno verjetnost, da bo po koncu kalibracijskega obdobja še aktivna. Število pričakovanih nakupov je 3,77. To je izredno dober rezultat, saj vemo, da je Nina v validacijskem obdobju dejansko opravila 3 nakupe. Še posebej pa je ta rezultat odličen v luči dejstva, da avtorjema modela ni uspelo izpeljati končne formule za napoved posameznih nakupov in gre zgolj za približek. Na Sliki 34 prikazujemo skladnost modela z dejanskimi podatki v kalibracijskem obdobju. Tudi tukaj se zdi, da je model relativno dobro napovedal dejansko stanje, še posebej v primerjavi z ostalimi obravnavanimi modeli.

Slika 34: Pogostost ocenjenih in dejanskih ponovnih nakupov v kalibracijskem obdobju v modelu CBG/CNBD-2



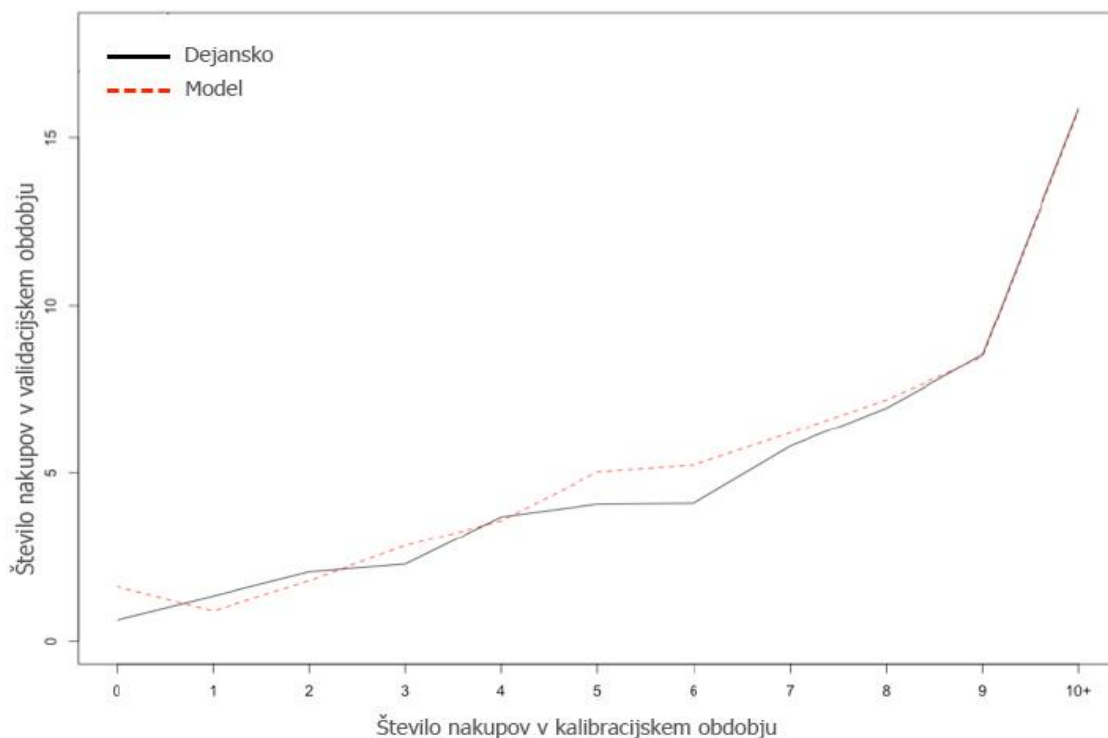
Na Sliki 35 prikazujem primerjavo števila nakupov v validacijskem obdobju na osi y in števila nakupov v kalibracijskem obdobju na osi x. Tudi v tem modelu so opazna precejšnja odstopanja, vendar pa model skoraj popolnoma napove število nakupov v primeru najmanj 9 transakcij. Podatki so prikazani tudi v Tabeli 19. Končno lahko na Sliki 36 prikažem gibanje celotnih kumulativnih nakupov. Presenetljivo je dejstvo, da model močno podcenjuje celotne kumulativne nakupe.

Tabela 19: Dejanski in pričakovani povprečni nakupi, ter število strank po skupinah v modelu CBG/CNBD-2

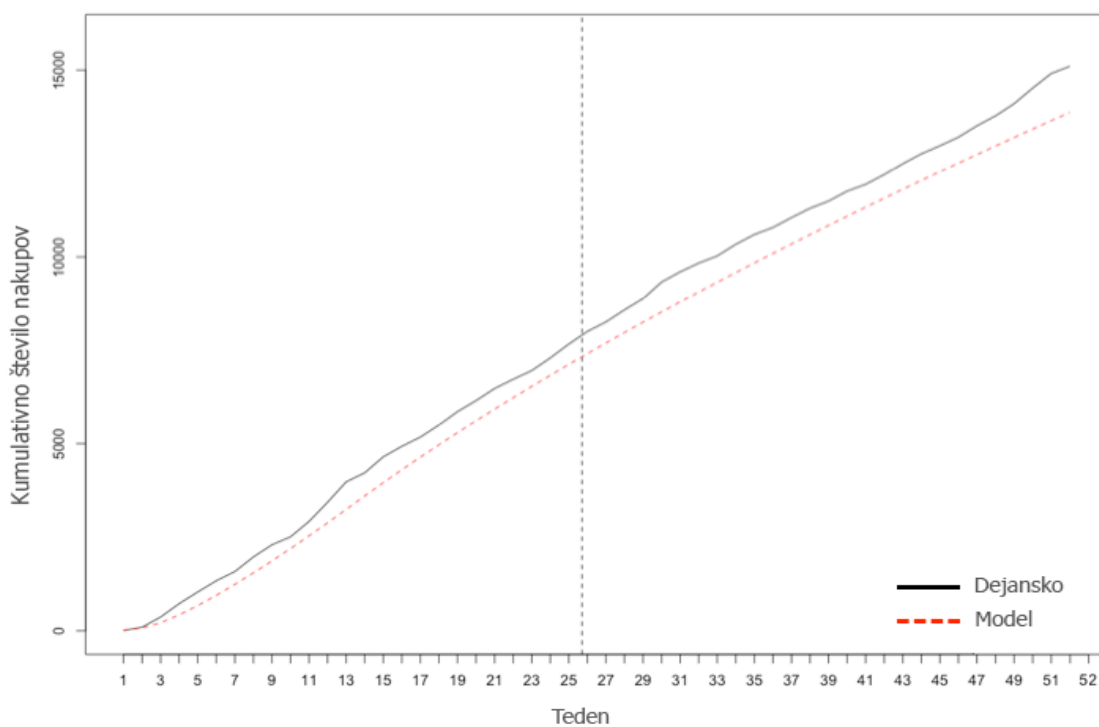
Skupina	0	1	2	3	4	5	6	7	8	9	10+
Dejanski nakupi	0,61	1,32	2,04	2,27	3,69	4,08	4,10	5,79	6,93	8,52	15,86
Pričakovani nakupi	1,61	0,88	1,78	2,82	3,58	5,02	5,23	6,19	7,18	8,45	15,79
Število strank	315	258	212	168	132	106	69	53	59	29	232



Slika 35: Pričakovani nakupi v validacijskem obdobju glede na število nakupov v kalibracijskem obdobju v modelu CBG/NBD



Slika 36: Gibanje ocenjenih in dejanskih kumulativnih nakupov po tednih v modelu CBG/CNBD-2



### 3.6 Primerjava posameznih modelov

Do sedaj sem se osredotočal na posamezne modele in le na kratko opravil določene primerjave na osnovi podatkov o skladnosti, prikazanih grafično. Zdaj pa se bomo posvetil bolj celostni primerjavi rezultatov skladnosti in predvsem točnosti, ki mi bo služila kot kriterij odločitve, kateri model je najprimernejši.

Tabela 20 prikazuje ocene vrednosti parametrov posameznih modelov, ki sem jim dodal dejanske in izpeljane povprečne vrednosti števila nakupov in stopnje osipa. Kot sem že ugotovil, vsi modeli precej podcenjujejo povprečni čas med nakupi, prav tako pa tudi stopnjo osipa. Izjema je model CBG/CNBD-2, ki stopnjo osipa precenjuje.

Tabela 20: Ocenjene vrednosti parametrov

Parameter	Pareto/NBD	BG/NBD	CBG/NBD	CBG/ CNBD-2
$r$	1,44	0,88	1,24	1,53
$\alpha$	4,32	3,23	4,08	2,13
$s$	0,36			
$\beta$	13,04			
$a$		0,17	0,26	0,38
$b$		1,72	2,09	2,15
<b>Dejanski povp. čas med nakupi</b>				
	44			
<b>Ocenjen povp. čas med nakupi</b>	21	25	23	20
<b>Dejanska stopnja osipa</b>				
	13,7 %			
<b>Ocenjena stopnja osipa</b>		9,1 %	10,9 %	15,1 %

#### 3.6.1 Skladnost

Kot že rečeno, modeli parametre ocenjujejo s pomočjo metode največjega verjetja. Cilj modela je poiskati vrednosti parametrov, ki maksimizirajo funkcijo verjetja oziroma, povedano drugače, poiskati parametre, ki kar najbolje ustrezajo dejanskim podatkom. Uporaba logaritemske funkcije verjetja je, ko primerjamo modele, največkrat uporabljena zaradi enostavnosti. Logaritem verjetja (angl. *log-likelihood*, v nadaljevanju LL) je vedno negativen, model z največjo vrednostjo logaritma verjetja pa je tisti, ki je najbolj skladen s podatki. Tabela 21 prikazuje logaritemske funkcije verjetja posameznih modelov. Največjo vrednost ima model BG/NBD, kar bi pomenilo, da najbolj ustreza dejanskim podatkom.

Tabela 21: Logaritmi verjetja

Model	Logaritem verjetja	Rang
Pareto/NBD	-16662,73	3
BG/NBD	<b>-16657,67</b>	<b>1</b>
CBG/NBD	-16661,78	2
CBG/CNBD-2	-16577,62	4

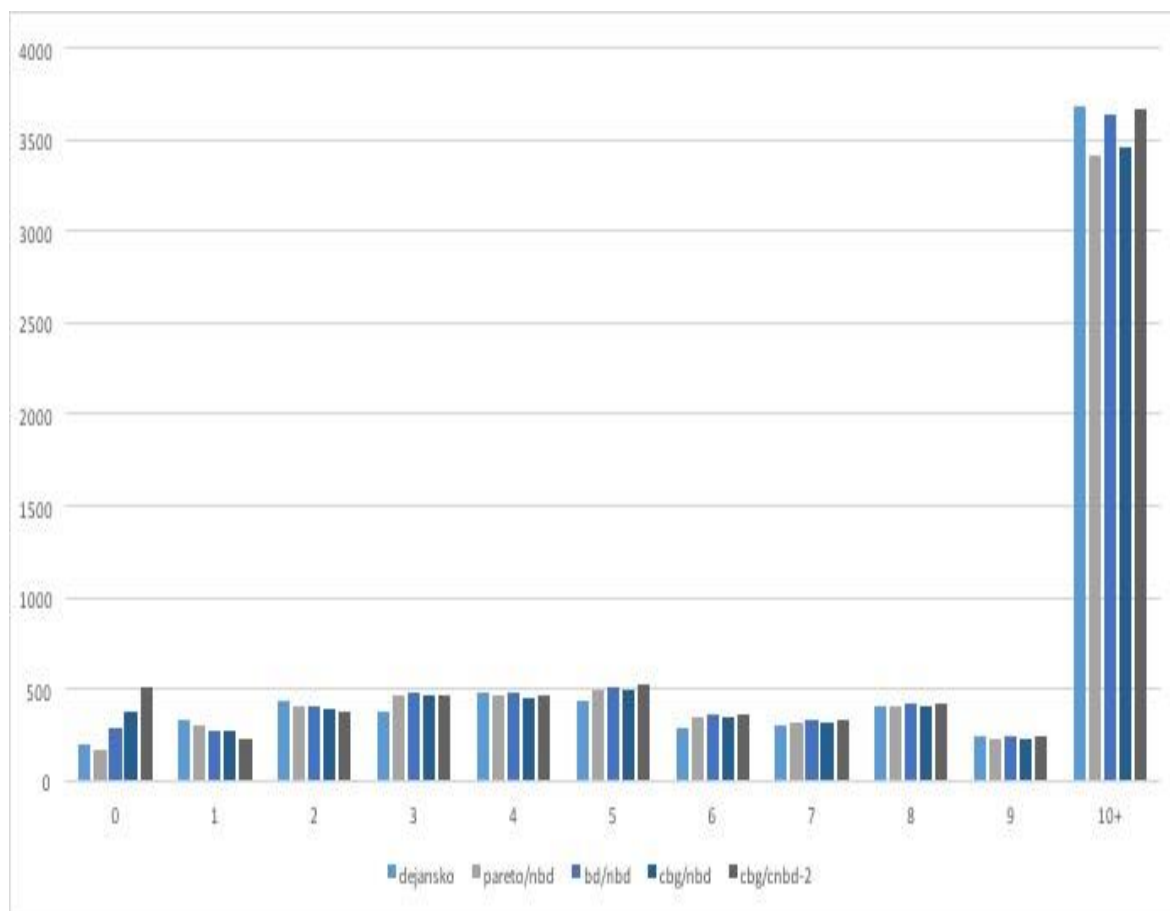
Metoda največjega verjetja deluje na posameznih vrednostih, zanimala pa me je tudi skladnost modelov za celotno bazo strank. Tabela 22 prikazuje celotne dejanske in pričakovane nakupe vseh 1.633 strank v kalibracijskem obdobju. Če pogledam skupne nakupe v tem obdobju, se najbolje odreže model CBG/NBD, ki je skupno število nakupov preцениl za zgolj 18 nakupov. Kot je razvidno iz Tabela 22, pa obstajajo precejšnje razlike pri razporeditvi števila ponovnih nakupov. Dejstvo je, da večina (58 %) strank opravi med nič in tremi ponovnimi nakupi. Presenetljivo se tukaj najbolje odreže model Pareto/NBD, ki je natančno zadel število nakupov. Če dodam še stranke z najmanj 10 nakupi, kar skupaj predstavlja 73 % strank, pa se najbolje odreže model BG/NBD.

Tabela 22: Dejanski in pričakovani celotni nakupi v kalibracijskem obdobju

Skupina	0	1	2	3	4	5	6	7	8	9	10+	Skupaj
Dejanski nakupi	193	341	433	382	487	432	283	307	409	247	3679	7193
Napoved Pareto/NBD	169	298	409	473	466	503	349	313	404	230	3419	7032
Napoved BG/NBD	288	280	407	480	481	519	366	328	425	240	3633	7449
Napoved CBG/NBD	381	268	389	461	459	500	348	313	405	231	3457	7211
Napoved CBG/CNBD-2	507	228	378	474	472	532	361	328	423	245	3663	7611
Število strank	315	258	212	168	132	106	69	53	59	29	232	1633

Podatki iz Tabele 22 so prikazani tudi na grafu na Sliki 37. Takoj so opazne velike razlike pri strankah brez ponovnega nakupa in pri strankah z več kot desetimi ponovnimi nakupi. Zdi se, da je model Pareto/NBD najbolje ocenil stranke brez ponovnih nakupov, model CBG-CNBD-2 pa stranke z najmanj 10 ponovnimi nakupi.

Slika 37: Dejanski in pričakovani celotni nakupi v kalibracijskem obdobju



### 3.6.2 Točnost

Končni cilj uporabe modelov je napovedovanje števila nakupov v prihodnosti. Če hočemo primerjati točnost teh napovedi, je treba podatke razdeliti na dva dela, kalibracijski in validacijski del. S podatki iz kalibracijskega obdobja sem ocenil parametre, podatke iz validacijskega obdobja pa sem uporabil pri ocenjevanju razlik med dejanskimi in ocenjenimi vrednostmi. Za ocenjevanje točnosti sem uporabil metode, opisane v pod poglavju 3.4.

Ena od prednosti obravnavanih modelov je zagotovo sposobnost napovedovanja na nivoju posamezne stranke, vendar pa prav tako želimo, da so modeli sposobni napovedati celotno število nakupov za celotno bazo strank. Tabela 23 prikazuje celotno dejansko in ocenjeno število nakupov na koncu validacijskega obdobja. Vidimo, da se je tu najboljše odrezal model CBG/NBD, najslabše pa model CBG/CNBD-2. Ti rezultati se ujemajo z ugotovitvami glede skladnosti, v zvezi s katerimi je model CBG/NBD prav tako najboljši.

Tabela 23: Dejansko in ocenjeno število nakupov na koncu validacijskega obdobja

Model	Dejansko	Napoved Pareto/NBD	Napoved BG/NBD	Napoved CBG/NBD	Napoved CBG/CNBD-2
Število nakupov	15.101	14.787	15.155	15.396	13.872
Relativna razlika v %		-2,1	2,4	1,6	-10,1
Absolutna razlika		-314	368	241	-1524

Omenil sem že, da je model uspešnejši, če napove vrednost, ki je bližja pravi vrednosti. Ker so modeli te vrednosti sposobni napovedovati na nivoju posamezne stranke, me je za ocenjevanje točnosti modela kot celote zanimala agregatna mera točnosti. Tabela 24 prikazuje tri agregatne mere točnosti, ki jih uporabljamo za ocenjevanje točnosti modelov. To so srednja absolutna napaka v odstotkih (MAPE), koren srednje kvadratne napake (RMSE) in srednja kvadratna logaritemska napaka (MSLE).

Na podlagi podatkov v Tabeli 24 sem ugotovil, da so modeli glede na uporabljeno mero točnosti razvrščeni različno. Najnižje vrednosti za posamezno mero točnosti so v tabeli označene odebeljeno. To pomeni, da noben od modelov dejanskih podatkov ne opisuje dovolj celovito. Ker sem kot kriterij odločitve izbral MSLE, bi lahko na tem mestu izbral model Pareto/NBD, pri katerem je vrednost MSLE med vsemi obravnavanimi modeli najnižja. Dodati velja, da so razlike med modeli precej majhne, in vprašanje je, ali je zanašanje samo na mere točnosti primerno.

Tabela 24: Mere točnosti uporabljenih modelov

Model	MAPE	RMSE	MSLE
Napoved Pareto/NBD	0,605	4,320	<b>0,563</b>
Napoved BG/NBD	0,613	<b>4,289</b>	0,592
Napoved CBG/NBD	0,609	4,310	0,576
Napoved CBG/CNBD-2	<b>0,600</b>	4,461	0,585

### 3.6.3 Hevristično odločanje in linearna regresija

Rezultati modelov me niso prepričali in zdi se, da nobeden med njimi ni sposoben popolnoma opisati naših podatkov. Spomnim naj, da nekateri avtorji (Wubben & Wangenheim, 2008) opozarjajo, da takšni kompleksni verjetnostni modeli velikokrat niso dosti boljši od enostavnega hevrističnega odločanja. V Tabeli 25 zato za primerjavo dodajam mere točnosti za predpostavko, da bodo stranke podjetja v povprečju povečale svoje število nakupov za 5 %, in enostavni linearni regresijski model.

Tabela 25: Mere točnosti za enostavno hevristično odločanje in linearno regresijo

Model	MAPE	RMSE	MSLE
Napoved Pareto/NBD	0,605	4,320	<b>0,563</b>
Napoved BG/NBD	0,613	4,289	0,592
Napoved CBG/NBD	0,609	4,310	0,576
Napoved CBG/CNBD-2	<b>0,600</b>	4,461	0,585
Hevristična (+ 5 %) napoved	0,622	4,414	0,663
Napoved z linearno regresijo	0,693	<b>4,269</b>	0,593

Vidimo, da sta takšna enostavna modela primerljiva z ostalimi, linearna regresija pa ima celo najnižjo vrednost RMSE. Postavlja se vprašanje, ali je torej uporaba teh kompleksnih verjetnostnih modelov sploh smiselna.

### 3.7 Ugotovitve in nadaljnji koraki

Dejstvo je, da mi z nobenim od obravnavanih modelov ni uspelo povsem zadovoljivo opisati podatkov, kar sem pokazal z analizo skladnosti. Dodatno sem pokazal, da tudi enostavno ugibanje poda primerljive rezultate glede točnosti napovedi, linearna regresija pa je celo bolj točna kot vsi ostali obravnavani verjetnostni modeli, če za primerjavo izberemo koren srednje kvadratne napake (RMSE).

Takšni rezultati so po eni strani razočaranje, po drugi strani pa sem skozi ta proces bolje spoznal nakupne vzorce strank podjetja. Skupaj z dobrim poznavanjem teh strank, izdelkov podjetja in trga sem oblikoval nekaj predlogov, s katerimi bi lahko potencialno izboljšali skladnost in točnost teh modelov in jih navajam v naslednjih točkah.

- Treba je uporabiti podatke daljšega obdobja. Eno leto je relativno kratko obdobje, še posebej, ker vemo, da podjetje deluje v izrazito sezonskem okolju.
- Skupaj z izbiro daljšega obdobja bi lahko podatke analizirali na mesečni ali tedenski ravni in se na ta način izognili variacijam na nivoju posameznih dni.
- Stranke je treba segmentirati v večji meri. Dejstvo je, da so si stranke podjetja med seboj zelo različne, posledica česar so različni vzorci njihovega nakupovanja. Ker modeli delujejo predvsem na nivoju posameznih strank, se zdi takšna segmentacija smiselna. Predvsem bi bila s tega vidika zanimiva segmentacija po državah, saj vemo, da se navade in obnašanje strank po državah zelo razlikujejo. Vemo tudi, da so med končnimi strankami gorski vodniki, gorski reševalci in druge profesionalne skupine uporabnikov, ki nakupujejo drugače kot ostali uporabniki. Te skupine bi bilo treba obravnavati ločeno.
- Mislim, da je smiselna tudi ločena analiza napovedi za posamezne skupine izdelkov. Podjetje prodaja izdelke za celo vrsto aktivnosti. Značilnosti in uporaba teh izdelkov je

zelo različna, zato se mi zdi smiselno osredotočanje na posamezne skupine, v določenih primerih celo podskupine izdelkov.

Napovedovanje prihodnjih nakupov je prvi korak pri izračunu CLV. Smiselno je, da se z nadaljnjimi koraki počaka, dokler nismo zadovoljni z rezultati prvega koraka. Z izboljšanjem rezultatov pri napovedovanju prihodnjih nakupov bo tudi kasnejši izračun CLV bolj zanesljiv in natančen. Po drugi strani pa je treba biti pripravljen tudi na možnost, da modeli enostavno ne prinašajo dodane vrednosti in ne opravičujejo vložkov, ki so potrebni za njihovo uporabo. S tega vidika ni smiselno izračunavati CLV, dokler nismo o tej ustreznosti prepričani.

### **3.8 Pomanjkljivosti in omejitve koncepta življenjske vrednosti strank**

Na koncu je treba opozoriti na nekatere pomembne pomanjkljivosti pri uporabi in modeliranju prihodnjih nakupov in CLV (povzeto po Gupta et al., 2006 in Singh & Jain, 2010):

- treba se je zavedati, da govorimo o napovedovanju prihodnosti, ki bo vedno negotova.
- CLV je dinamičen koncept in ne statičen model. Ne samo, da se izračunana življenjska vrednost strank oziroma napovedano število prihodnjih nakupov skozi čas spreminja, spreminjajo se tudi razmerja med podjetjem in stranko, obnašanje strank, širše poslovno okolje itd. Vse to pomeni, da lahko izbrani modeli v vsakem trenutku postanejo neuporabni ali celo napačni. To od podjetij, ki uporabljajo verjetnostne modele, zahteva, da nenehno spremljajo predpostavke, na katerih modeli temeljijo, in se ustrezno prilagajajo.
- Modeli so zastavljeni zelo široko, vemo pa, da je vsako podjetje specifično. Vprašati se je treba, ali je modeliranje takšnih posebnosti smiselno in ali so modeli sploh uporabni, če takšnih posebnosti ne upoštevamo. Hkrati se je treba zavedati, da je lahko modeliranje specifičnih lastnosti izredno kompleksno in dolgotrajno, vpliv teh sprememb na točnost in natančnost modelov pa minimalen.
- Modeli večinoma predpostavljajo neodvisnost med strankami, vemo pa, da so v današnjem svetu stranke med seboj povezane bolj kot kadarkoli prej. Nепrestano se je treba spraševati, kako te direktne in indirektne povezave vplivajo na nakupe in kako takšno kompleksnost modelirati.
- Modeli se načeloma posvečajo prihodkovni strani, vemo pa, da je s pridobivanjem, servisiranjem in prodajo strankam povezanih veliko (pogosto kompleksnih) stroškov. Da bi lahko natančno izračunali CLV, bi bilo treba te stroške upoštevati.
- Navzkrižna prodaja in prodaja dražjih različic izdelkov nista upoštevani, a lahko pri obstoječih strankah predstavljata pomembne priložnosti.

- Modeli ne upoštevajo aktivnosti konkurentov. Če na primer naš konkurent zniža cene ali pride na trg z novim izdelkom, se bo to na naši prodaji verjetno poznalo. Takšnih scenarijev modeli ne upoštevajo.
- Modeli delujejo na osnovi podatkov o obstoječih strankah in se tudi osredotočajo predvsem na obstoječe stranke. Podjetje pa samo na podlagi uporabljenih modelov ne ve, kako se obstoječe stranke primerjajo s potencialnimi novimi strankami. Stranke podjetja so lahko na primer koncentrirane v določenem segmentu, ki je izredno majhen in drag za vzdrževanje.
- Modeli ne upoštevajo trženjskih aktivnosti, ki lahko stranke z nizko življenjsko vrednostjo spremenijo v stranke z visoko življenjsko vrednostjo.

Dejstvo je, da je v literaturi mogoče najti veliko modelov, ki se bolj ali manj celovito posvečajo tem pomanjkljivostim. Problem teh rešitev je, da so večinoma izredno kompleksne in kot take nezanimive za praktično uporabo v podjetjih (Fader et al., 2005):

- povprečni analitični uporabniki, ki modele implementirajo, ne razumejo, kako ti modeli delujejo, saj je zato potrebno napredno poznavanje matematike in statistike.
- Če analitični uporabniki delovanje razumejo, je kompleksne modele težko razumljivo predstaviti ostalim sodelavcem in predvsem odločevalcem.
- Takšni kompleksni modeli zahtevajo veliko več časa in sredstev, ki pa jih je težko pridobiti, če prednosti in uporaba modelov niso enostavno razumljive.

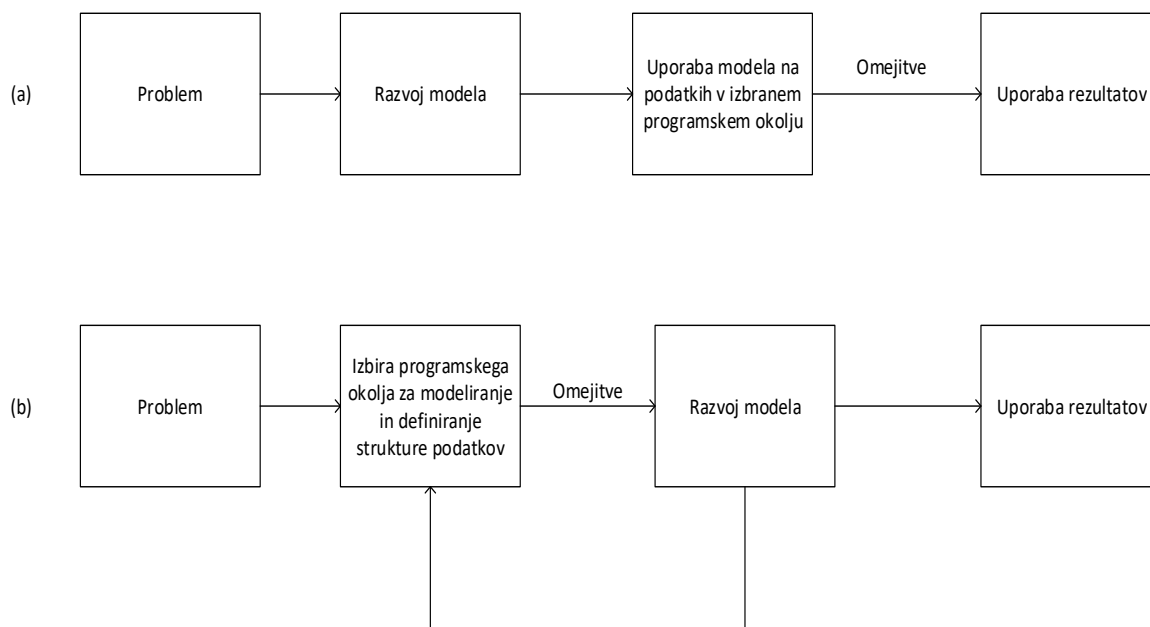
Osnovni modeli BTYD so s tega vidika primerni, saj je njihova implementacija mogoča celo neposredno v Excelu. Fader in Hardie (2005) takšne modele imenujeta »enostavni modeli« in navajata več argumentov, zakaj in kako naj se jih podjetja lotijo:

- ker so modeli enostavni, jih je mogoče uporabiti celo v Excelu. Ogromno prednost takšnega načina predstavlja dejstvo, da jih lahko uporabniki sami zgradijo od začetka in na ta način pridobijo dobro razumevanje delovanja modela. To močno vpliva na zaupanje v rezultate modela in sposobnost argumentiranja le-teh.
- Samo po sebi je takšno logično in strukturirano grajenje modela koristno za razumevanje podatkov in obravnavanega problema.
- Uporaba takšnih enostavnih modelov je veliko cenejša in hitrejša.

Če se izkaže, da so enostavni modeli za podjetje koristni, uporabniki sami začnejo zahtevati in zastavljati kompleksnejša vprašanja. Ker v podjetju zdaj že obstaja določeno znanje glede modeliranja, je naslednji korak h kompleksnejšim rešitvam lažji. Enostavni modeli so torej odlična vstopna točka v svet modeliranja prihodnjih nakupov strank. Fader in Hardie (2005) zagovarjata takšen evolucijski pristop k modeliranju, ki je prikazan na Sliki 38.



Slika 38: Pristopi k modeliranju verjetnostnih modelov



Vir: P. Fader & B. Hardie, *The value of simple models in new product forecasting and customer-base analysis*, 2005, str. 471, slika 3.

Klasični pristop modeliranja (a) se začne z obravnavo problema. Razvijalec modela tega razvije in v izbranem programskem okolju uporabi na podatkih podjetja. Problem nastane, ko želi podjetje rezultate uporabiti v praksi. Specifični programi, ki jih je uporabil razvijalec, in kompleksne podatkovne strukture, ki podpirajo model, omejujejo uporabo rezultatov v praksi. Modeliranje je preveč odtujeno od končnih uporabnikov, da bi mu ti zaupali.

Pristop enostavnih modelov (b) je drugačen. Izbira programskega okolja in določitev potrebnih podatkovnih struktur se izvedeta na začetku, v sodelovanju s končnimi uporabniki. Na ta način uporabniki dobro spoznajo podrobnosti modela, kar močno olajša uporabo v praksi. Po drugi strani je takšen razvoj modela bolj dolgotrajen, saj so večinoma potrebne iteracije in revizije podatkovnih struktur ter matematične logike, na kateri model temelji (Fader & Hardie, 2005).

## SKLEP

Razumevanje in vrednotenje strank v konceptu osrediščenosti na stranko kot ključno metriko izpostavlja CLV. Za izračun CLV je treba najprej napovedati prihodnje nakupe strank. Izmed vrste različnih pristopov napovedovanja prihodnjih nakupov se verjetnostni modeli osredotočajo neposredno na to problematiko. Izbira ustreznega modela je zahtevna naloga

in je v prvi vrsti odvisna od poslovnega konteksta med podjetjem in stranko. Različni verjetnostni modeli uporabljajo različne verjetnostne porazdelitve, s pomočjo katerih razlagajo obnašanje strank, kar pomembno vpliva na rezultate modela.

Na podlagi analize strokovne literature, ki obravnava CLV, sem se odločil za modele BTYD, saj imajo le-ti preverjeno napovedno sposobnost, so relativno enostavni za uporabo in ker je, kar je morda najpomembnejše, zgodba, ki jo pripovedujejo, enostavna in lahko razumljiva. Namen tega magistrskega dela je tako izpolnjen, saj sem temeljito raziskal uporabo teh modelov za napovedovanje prihodnjih nakupov strank. Ugotovim lahko, da so modeli relativno enostavni za uporabo in relativno lahko razumljivi. Opozoriti pa je treba, da so modeli zasnovani za splošno uporabo in kot taki niso nujno primerni za vsako podjetje. Modele je sicer možno razširiti z dodatnimi spremenljivkami, ki bolj točno zajemajo specifičnosti v posameznih situacijah, vendar to prinaša dodatno kompleksnost in zahteva poglobljeno znanje modeliranja, statistike in programskega jezika R.

V magistrskem delu sem predstavil projekt preverbe uporabnosti izbranih modelov BTYD v podjetju Black Diamond. Black Diamond je podjetje, ki izdeluje in prodaja izdelke za plezanje, gornišstvo, alpinizem, smučanje in ostale športne aktivnosti v naravi. Podjetje brez svojih strank ne bi obstajalo, zato je razumevanje strank za podjetje ključno. V želji, da bi postali bolj osredinjeni na stranke, podjetje raziskuje različne možnosti. Ena med njimi je izračun in uporaba CLV. Prvi korak pri izračunu CLV je napovedovanje prihodnjih nakupov strank. Med različnimi modeli, ki jih lahko uporabimo za napovedovanje prihodnjih nakupov strank, sem izbral verjetnostne napovedne modele in skupino modelov, ki so poznani kot modeli BTYD.

Izpolnil sem tudi osnovni cilj magistrskega dela in na podlagi odločitvenega kriterija najnižje MSLE kot za podjetje najbolj ustreznega izbral model Pareto/NBD. Vendar pa sem hkrati na podlagi temeljite analize skladnosti, točnosti in predpostavk modelov prišel do zaključka, da nobeden od izbranih modelov v resnici ni zadostno pojasnil naših podatkov. Na podlagi spoznanj, do katerih sem prišel v magistrskem delu, sem predlagal vrsto korakov, ki bi lahko izboljšali točnost in skladnost izbranih modelov. Hkrati sem opozoril na nekatere pomanjkljivosti verjetnostnih napovednih modelov in dejstvo, da se lahko tudi enostavna linearna regresija ali celo hevristično odločanje enakovredno uporablja pri napovedovanju prihodnjih nakupov.

Pokazal sem, da je uporaba verjetnostnih modelov za napovedovanje prihodnjih nakupov strank zanimiva in primerna metoda. Mislim pa, da zanašanje zgolj na verjetnostne modele ni primerno. Znanje o posameznih strankah, njihovem vedenju in nakupnem odločanju, ki je prisotno v podjetjih, je preveč obsežno, kompleksno in dinamično, da bi ga lahko uspešno modelirali. Hevristično odločanje in ocenjevanje bo ostalo pomemben del pri vrednotenju strank, verjetnostni modeli pa so lahko samo koristno orodje za sprejemanje boljših odločitev.

## LITERATURA IN VIRI

1. Abe, M. (2009). "Counting your customers" one by one: A hierarchical Bayes extension to the Pareto/NBD model. *Marketing Science*, 28(3), 541–553.
2. Batislam, E. P., Denizel, M., & Filiztekin, A. (2007). Empirical validation and comparison of models for customer base analysis. *International Journal of Research in Marketing*, 24(3), 201–209.
3. Bell, D., Deighton, J., Reinartz, W. J., Rust, R. T., & Swartz, G. (2002). Seven Barriers to Customer Equity Management. *Journal of Service Research*, 5(1), 77–85.
4. Berger, P. D., & Nasr, N. I. (1998). Customer lifetime value: Marketing models and applications. *Journal of Interactive Marketing*, 12(1), 17–30.
5. Blattberg, R. C., & Deighton, J. (1996). Manage marketing by the customer equity test. *Harvard business review*, 74(4), 136.
6. Blattberg, R. C., Getz, G., & Thomas, J. S. (2001). *Customer equity: Building and managing relationships as valuable assets*. Boston: Harvard Business Press.
7. Calciu, M. (2009). Deterministic and stochastic Customer Lifetime Value models. Evaluating the impact of ignored heterogeneity in non-contractual contexts. *Journal of Targeting, Measurement and Analysis for Marketing*, 17(4), 257–271.
8. Cheng, C. H., & Chen, Y. S. (2009). Classifying the segmentation of customer value via RFM model and RS theory. *Expert systems with applications*, 36(3), 4176-4184.
9. Dwyer, F. R. (1997). Customer lifetime valuation to support marketing decision making. *Journal of interactive marketing*, 11(4), 6-13.
10. Ehrenberg, A. S. (1959). The pattern of consumer purchases. *Applied Statistics*, 8(1), 26–41.
11. Esmailigookeh, M., & Tarokh, M. J. (2013). Customer Lifetime Value Models : A literature Survey. *International Journal of Industrial Engineering and Production Research*, 24(4), 317–336.
12. Fader, P. (2012). *Customer centricity: Focus on the right customers for strategic advantage*. Philadelphia: Wharton digital press.
13. Fader, P. S., & Hardie, B. G. (2005). The value of simple models in new product forecasting and customer-base analysis. *Applied Stochastic models in business and industry*, 21(4-5), 461–474.
14. Fader, P. S., & Hardie, B. G. (2008, december). Creating an RFM Summary Using Excel. *Bruce G. S. Hardie*. Najdeno 1. januarja 2015 na spletnem naslovu [http://brucehardie.com/notes/022/RFM\\_summary\\_in\\_Excel.pdf](http://brucehardie.com/notes/022/RFM_summary_in_Excel.pdf).
15. Fader, P. S., & Hardie, B. G. (2009). Probability models for customer-base analysis. *Journal of interactive marketing*, 23(1), 61–69.
16. Fader, P. S., & Hardie, B. G. (2015). Simple probability models for computing CLV and CE. V. Kumar & D. Shah (ur.), *Handbook of Research on Customer Equity in Marketing* (str. 77-100). Cheltenham: Edward Elgar Publishing.

17. Fader, P. S., Hardie, B. G., & Beger, P. D. (2011, januar). Implementing the BG/BB Model for Customer-Base Analysis in Excel. *Bruce G. S. Hardie*. Najdeno 1. januarja 2015 na spletnem naslovu [http://www.brucehardie.com/notes/010/bgbb\\_spreadsheet\\_2011-01-20.pdf](http://www.brucehardie.com/notes/010/bgbb_spreadsheet_2011-01-20.pdf).
18. Fader, P. S., Hardie, B. G., & Lee, K. L. (2005a). "Counting your customers" the easy way: An alternative to the Pareto/NBD model. *Marketing science*, 24(2), 275–284.
19. Fader, P. S., Hardie, B. G. S., & Lee, K. L. (2005b). RFM and CLV: Using Iso-Value Curves for Customer Base Analysis. *Journal of Marketing Research*, 42(4), 415–430.
20. Fader, P. S., Hardie, B. G., & Sen, S. (2014). Stochastic models of buyer behavior. R. S. Winer & S. A. Neslin (ur.), *The History of Marketing Science* (str. 165-205). Singapore: World Scientific Publishing.
21. Fader, P. S., Hardie, B. G. S., & Shang, J. (2010). Customer-Base Analysis in a Discrete-Time Noncontractual Setting. *Marketing Science*, 29(6), 1086–1108.
22. Funk, D., Alexandris, K., & McDonald, H. (2008). *Consumer behaviour in sport and events*. Abingdon-on-Thames: Routledge.
23. Galbraith, J. R. (2011). *Designing the customer-centric organization: A guide to strategy, structure, and process*. San Francisco: John Wiley & Sons.
24. Glady, N., Baesens, B., & Croux, C. (2009). A modified Pareto / NBD approach for predicting customer lifetime value. *Expert Systems With Applications*, 36(2), 2062–2071.
25. Gupta, S., & Lehmann, D. R. (2003). Customers as assets. *Journal of Interactive Marketing*, 17(1), 9–24.
26. Gupta, S., Lehmann, D. R., & Stuart, J. A. (2004). Valuing Customers. *Journal of Marketing Research*, 41(1), 7–18.
27. Gupta, S., Hanssens, D., Hardie, B., Kahn, W., Kumar, V., Lin, N., Ravishanker, N., Sriram, S. (2006). Modeling Customer Lifetime Value. *Journal of Service Research*, 9(2), 139–155.
28. Gupta, S., & Lehmann, D. (2005). *Managing customers as investments the strategic value of customers in the long run*. Philadelphia: Wharton School Publishing.
29. Hoppe, D., & Wagner, U. (2007). Customer base analysis: The case for a central variant of the Betageometric/NBD model. *Marketing ZFP*, 29(JRM 2), 75–90.
30. Jain, D., & Singh, S. S. (2002). Customer lifetime value research in marketing: A review and future directions. *Journal of Interactive Marketing*, 16(2), 34–46.
31. Korkmaz, E., Kuik, R., & Fok, D. (2013). "Counting Your Customers": When will they buy next? An empirical validation of probabilistic customer base analysis models based on purchase timing. Rotterdam: Erasmus Research Institute of Management (ERIM).
32. Korkmaz, E., Fok, D., & Kuik, R. (2014). *The Need for Market Segmentation in Buy-Till-You-Defect Models*. Rotterdam: Erasmus Research Institute of Management (ERIM).

33. Kotler, P., Keller, K. L., Manceau, D., & Hémonnet-Goujot, A. (2015). *Marketing management* (14<sup>th</sup> ed.). Englewood Cliffs: Prentice Hall.
34. Kumar, V. (2008). *Customer lifetime value: the path to profitability*. Hanover: Now Publishers Inc.
35. Kumar, V., Lemon, K. N., & Parasuraman, a. (2006). Managing Customers for Value: An Overview and Research Agenda. *Journal of Service Research*, 9(2), 87–94.
36. Kumar, V., Pozza, I. D., Petersen, J. A., & Shah, D. (2009). Reversing the Logic: The Path to Profitability through Relationship Marketing. *Journal of Interactive Marketing*, 23(2), 147–156.
37. Kumar, V., & Rajan, B. (2009). Profitable Customer Management: Measuring and Maximizing Customer Lifetime Value. *Management Accounting Quarterly*, 10(3), 1–19.
38. Lieber, J. (2014). *Osrediščenost v stranki : obravnava koncepta v teoriji in praksi* (magistrsko delo). Ljubljana: Ekonomska fakulteta.
39. Linoff, G. S. (2015). *Data analysis using SQL and Excel*. San Francisco: John Wiley & Sons.
40. Malthouse, E. C., & Blattberg, R. C. (2005). Can we predict customer lifetime value? *Journal of Interactive Marketing*, 19(1), 2–16.
41. Pfeifer, P. E., & Carraway, R. L. (2000). Modeling customer relationships as Markov chains. *Journal of interactive marketing*, 14(2), 43.
42. Platzer, M. (2008). *Stochastic models of noncontractual consumer relationships*. Dunaj: Vienna University of Economics and Business Administration.
43. Reinartz, W. J., & Kumar, V. (2000). On the Profitability of Long-Life Customers in a Noncontractual Setting: An Empirical Investigation and Implications for Marketing. *Journal of Marketing*, 64(4), 17–35.
44. Reinartz, W., & Kumar, V. (2002). The mismanagement of customer loyalty - Response. *Harvard Business Review*, 80(7), 86-94.
45. Reinartz, W., Krafft, M., & Hoyer, W. D. (2004). The Customer Relationship Management Process: Its Measurement and Impact on Performance. *Journal of Marketing Research*, 41(3), 293–305.
46. Schmittlein, D. C., Bemmaor, A. C., & Morrison, D. G. (1985). Technical Note-Why Does the NBD Model Work? Robustness in Representing Product Purchases, Brand Purchases and Imperfectly Recorded Purchases. *Marketing Science*, 4(3), 255–266.
47. Schmittlein, D. C., Morrison, D. G., & Colombo, R. (1987). Counting Your Customers: Who Are They and What Will They Do Next? *Management Science*, 33(1), 1–24.
48. Singh, S. S., Borle, S., & Jain, D. C. (2009). A generalized framework for estimating customer lifetime value when customer lifetimes are not observed. *QME*, 7(2), 181-205.

49. Singh, S. S., & Jain, D. C. (2010). Measuring Customer Lifetime Value. N. K. Malhotra (ur.), *Review of Marketing Research* (str. 37-62). Bradford: Emerald Group Publishing Limited.
50. Wadsworth, E. (2014, november). Buy'Til You Die-A Walkthrough. *Cran.r-project*. Najdeno 1. januarja 2015 na spletnem naslovu <https://cran.r-project.org/web/packages/BTYD/vignettes/BTYD-walkthrough.pdf>.
51. Wübben, M., & Wangenheim, F. V. (2008). Instant Customer Base Analysis: Managerial Heuristics Often "Get It Right." *Journal of Marketing*, 72(3), 82–93.

## **PRILOGE**





## **PRILOGA 1: Seznam v delu uporabljenih kratic**

BB – beta-binomial  
BG – beta-geometric  
BTYD – Buy 'Til You Die  
CBG – central variant beta geometric  
CE – customer equity  
CLV – customer lifetime value  
CNBD – condensed negative binomial distribution  
HB – hierarchical Bayes  
LL – log likelihood  
MAE – mean absolute error  
MAPE – mean absolute percentage error  
MCMC – Markov chain Monte Carlo  
MLE – maximum likelihood estimation  
MSE – mean squared error  
MSLE – mean squared logarithmic error  
NBD – negative binomial distribution  
PCM – psychological continuum model  
RMSE – root mean squared error  
sBG – shifted beta-geometric