

UNIVERZA V LJUBLJANI
FAKULTETA ZA DRUŽBENE VEDE

Manca Jesenko

Alternativne metode modeliranja relativnega preživetja

Doktorska disertacija

Ljubljana, 2014

UNIVERZA V LJUBLJANI
FAKULTETA ZA DRUŽBENE VEDE

Manca Jesenko

Mentor: red. prof. dr. Janez Stare

Alternativne metode modeliranja relativnega preživetja

Doktorska disertacija

Ljubljana, 2014

Zahvala

Ob tej priložnosti bi se najprej rada zahvalila celotnemu kolektivu podiplomskega študija Statistika. S pridobljenimi znanji in izkušnjami, ki so mi jih posredovali sodelujoči profesorji, sem dobila življenjsko popotnico, ki mi omogoča lažje širjenje obzorij in mi daje možnost, da resnično uživam v delu, ki ga opravljam.

Še posebej toplo se zahvaljujem mentorju prof. dr. Janezu Staretu, ki mi je pri ustvarjanju doktorske disertacije, s svojim strokovnim znanjem in izkušnjami veliko pomagal in mi stal ob strani tudi v najtežjih trenutkih mojega življenja.

Za spodbudo in veliko mere motivacije bi se rada zahvalila tudi svojim sodelavcem kolektiva Analitsko raziskovalnega centra Banke Slovenije.

Največja zahvala gre moji družini, hčerkama Nali in Lizi, ki sta marsikdaj morali potrpeti tudi brez moje navzočnosti, Kemalu, bratoma Mihu in Robertu in njegovi družini in seveda mami, ki je bila zlasti na koncu mojega študija ne preveč tihi, a učinkoviti glas vesti, ki je pognal kolesje v trenutkih, ko se je nameravalo ustaviti. Na koncu bi se rada zahvalila mojemu očetu. Za veselje do učenja, pridobivanja novih znanj in statistike. Ati, kjerkoli že si, hvala za tvojo ljubezen, ki mi je to omogočila. To delo posvečam tebi in tvoji neodtujljivi polovici, mami.



IZJAVA O AVTORSТVУ

doktorske disertacije

Podpisani/-a Manca Jesenko, z vpisno številko 74500043, sem avtor/-ica doktorske disertacije z naslovom:
Alternativne metode modeliranja relativnega preživetja.

S svojim podpisom zagotavljam, da:

- je predložena doktorska disertacija izključno rezultat mojega lastnega raziskovalnega dela;
- sem poskrbel/-a, da so dela in mnenja drugih avtorjev oz. avtoric, ki jih uporabljam v predloženem delu, navedena oz. citirana v skladu s fakultetnimi navodili;
- sem poskrbel/-a, da so vsa dela in mnenja drugih avtorjev oz. avtoric navedena v seznamu virov, ki je sestavni element predloženega dela in je zapisan v skladu s fakultetnimi navodili;
- sem pridobil/-a vsa dovoljenja za uporabo avtorskih del, ki so v celoti prenesena v predloženo delo in sem to tudi jasno zapisal/-a v predloženem delu;
- se zavedam, da je plagiatorstvo – predstavljanje tujih del, bodisi v obliki citata bodisi v obliki skoraj dobesednega parafraziranja bodisi v grafični obliki, s katerim so tuje misli oz. ideje predstavljene kot moje lastne – kaznivo po zakonu (Zakon o avtorski in sorodnih pravicah (UL RS, št. 16/07-UPB3, 68/08, 85/10 Skl.US: U-I-191/09-7, Up-916/09-16)), prekršek pa podleže tudi ukrepom Fakultete za družbene vede v skladu z njenimi pravili;
- se zavedam posledic, ki jih dokazano plagiatorstvo lahko predstavlja za predloženo delo in za moj status na Fakulteti za družbene vede;
- je elektronska oblika identična s tiskano obliko doktorske disertacije ter soglašam z objavo doktorske disertacije v zbirki »Dela FDV«.

Povzetek

Alternativne metode modeliranja relativnega preživetja

Klasična analiza preživetja se ukvarja s proučevanjem časa, ki preteče med dvema dobro definiranimi dogodkoma. Čeprav je proučevani čas razmernostna spremenljivka, pa se v analizi preživetja soočamo z nepopolnimi ali krnjjenimi podatki, ki jih ne moremo analizirati z običajnimi statističnimi metodami, ampak so bile za njihovo analizo razvite posebne metode, ki jih poznamo pod skupnim imenom analiza preživetja, v družboslovju bolj znana kot analiza dogodkov. Kadar enote opazujemo skozi daljše obdobje, se lahko zgodi, da končnega dogodka ne moremo dobro definirati oziroma ga pripisati proučevanemu pojavu. V takih primerih klasična analiza preživetja ni mogoča, saj bi z upoštevanjem vseh dogodkov preživetje enot podcenili, izločitev dogodkov, ki jih ne moremo pripisati proučevanemu pojavu, pa zaradi nezanesljivih ali celo nerazpoložljivih informacij ni mogoče. Tovrstno problematiko lahko rešujemo v okviru koncepta relativnega preživetja, s katerim iz preživetja opazovanih enot skušamo izločiti vse tiste dogodke, ki jih ne moremo pripisati proučevanemu pojavu. V okviru tega nastope dogodkov, povezane s proučevanim pojavom, primerjamo z enakovrednimi dogodki populacije (recimo ji splošna populacija), ki poleg proučevanega pojava vključujejo tudi vse druge vzroke, ki lahko vplivajo na nastop končnega dogodka. Uporaba metod relativnega preživetja je trenutno aplicirana predvsem na medicinsko področje, kjer paciente pogosto opazujemo skozi daljše časovno obdobje, natančnega vzroka smrti pa mnogokrat ne moremo opredeliti. Poleg tega statistični uradi redno posodabljajo tudi podatke o smrtnosti splošne populacije, kar omogoča primerjavo preživetja opazovanih enot s preživetjem splošne populacije, in s tem uporabo metodologije relativnega preživetja.

Formalno gledano je bilo relativno preživetje najprej definirano s funkcijo kumulativnega relativnega preživetja ($r(t)$) kot razmerje med preživetjem skupine opazovanih enot in preživetjem primerljive skupine splošne populacije. Funkcija $r(t)$ kot taka predstavlja skupinsko mero relativnega preživetja, ki pove, kolikšen je delež preživelih opazovanih enot v primerjavi z deležem preživelih v splošni populaciji. Za namene

modeliranja $r(t)$ obstajata dva pristopa, ki se med seboj razlikujeta v predpostavki o povezanosti med opazovano in populacijsko ogroženostjo. Prvi je aditivni pristop, ki predpostavlja, da opazovano ogroženost lahko izrazimo kot vsoto populacijske in presežne ogroženosti (pri čemer je presežna ogroženost odraz proučevane bolezni). Drugi pristop je multiplikativni pristop, v katerem je opazovana ogroženost definirana kot produkt populacijske ogroženosti in tako imenovane relativne umrljivosti. V literaturi so kljub dvema pristopoma skoraj izključno uporabljeni aditivni modeli, kar se utemeljuje s primernostjo za biološke podatke.

Prostor proučevanja relativnega preživetja se je razširil s pojavom individualne mere preživetja, ki relativno preživetje v nasprotju s skupinsko mero definira za vsako enoto posebej. Temelji na preprosti transformaciji časov preživetja, s katero proučevani čas opazovane enote s pomočjo pogojnih porazdelitvenih funkcij časov preživetja, določenih na splošni popoulaciji, preslikamo v individualno mero relativnega preživetja ali transformirani čas. Ta nam pove kolikšen delež primerljive splošne populacije je preživila opazovana enota. Nova mera relativnega preživetja tako že sama po sebi prinaša drugačen pogled na relativno preživetje, poleg tega pa nam opredilitev relativnega preživetja na individualnih enotah omogoča, da za njegovo proučevanje uporabimo metodologijo klasične analize preživetja. Doktorska disertacija je osredotočena predvsem na empirično modeliranje relativnega preživetja s pomočjo metodologije klasične analize preživetja, ki je v praksi sicer redkeje uporabljena. Jedro naloge tako predstavlja rezultati Aalenove linearne regresije ogroženosti in Buckley-Jamesove linearne regresije transformiranih časov preživetja. Predvsem zaradi primerjave z Aalenovim modelom pa so analize narejene še z v praksi najpogosteje uporabljeno Coxovo regresijo sorazmernih ogroženosti. Vsi modeli so preizkušeni na podatkih o raku na pljučih in raku na dojki, ki se med seboj najbolj razlikujejo glede na delež krnjениh enot, s tem pa tudi glede na njihovo preživetje. V doktorski disertaciji obravnavan pristop k relativnemu preživetju je popolnoma nov in kot tak prinaša možnosti poglobljenega vedenja o relativnem preživetju, njegova uporaba pa omogoča možnosti pridobivanja novih informacij o relativnem preživetju in s tem vsebinsko obogatitev področja relativnega preživetja. Prednost uporabe novega pristopa k modeliranju rel-

ativnega preživetja je poleg tega tudi v tem, da v primerjavi z aditivnimi in multiplikativnimi regresijskimi metodami v tradicionalni analizi relativnega preživetja ne zahteva nikakršnih predpostavk o povezanosti opazovane in populacijske ogroženosti.

Empirična analiza podatkov je pokazala, da z ocenjenimi modeli lahko pridemo do konsistentnih informacij o učinku neodvisnih spremenljivk na relativno preživetje. Kljub temu pa je potrebno biti pri uporabi modelov in njihovi interpretaciji previden, še zlasti v primeru ko se učinki neodvisnih spremenljivk v času spreminjajo. V nalogi je pokazano, da so v primeru, ko podatke generira Aalenov model, parametri Buckley-Jamesove regresije odvisni tako od parametrov Aalenovega modela, kot tudi od kombinacije neodvisnih spremenljivk. V primeru časovno odvisnih učinkov neodvisnih spremenljivk pa so parametri Buckley-Jamesovega modela pri enakih vrednostih neodvisnih spremenljivk odvisni tudi od pogojnih porazdelitev proučevanega časa na časovnih intervalih z različnimi vrednostmi parametrov Aalenovega modela. V primeru nespreminjajočih parametrov Aalenovega modela je torej uporaba Buckley-Jamesovega modela odvisna od stopnje nelinearnosti in heteroskedastičnosti, medtem ko v primeru spremnjajočih parametrov Aalenovega modela parametri Buckley-Jamesove regresije teh sprememb ne zaznajo. Kljub temu je s simulacijami mogoče pokazati, da v primeru ko podatke generiramo s klasično linearno regresijo tako, da vrednosti parametrov pri enakih vrednostih neodvisnih spremenljivk sprememimo glede na izbrane časovne točke, te spremembe lahko zaznamo z Aalenovim modelom. Ker v praksi ne vemo in pogosto niti s prilagoditvenimi testi ne zaznamo, kateremu modelu se podatki bolj prilegajo, s konsistentnostjo modelov lahko potrdimo uporabnost dobljenih rezultatov. V primeru ko se učinki neodvisnih spremenljivk v času spreminjajo, Buckley-Jamesovo regresijo lahko naredimo na intervalih, na katerih so vrednosti parametrov Aalenovega modela vsaj približno konstantni.

Rezultati ocenjenih modelov so potrdili prednost velike prilagodljivosti Aalenove regresije, s katero smo tako na primeru podatkov o raku na pljučih kot tudi na primeru podatkov o raku na dojki zaznali, da se učinki nekaterih neodvisnih spremenljivk na relativno preživetje v času spreminjajo. Čeprav je modeliranje časovno odvisnih

učinkov možno tudi v Coxovem modelu, je njegova slabost ta, da moramo za modeliranje ustrezzo preurediti vhodne podatke, na drugi strani pa je izbira prelomnih točk popolnoma arbitrarna. Prednost Coxovega modela je predvsem v tem, da učinke neodvisnih spremenljivk lahko kvantitativno ovrednotimo. Buckley-Jamesov model je kljub nezmožnosti zaznavanja sprememb učinkov v času pokazal na rezultate, ki so konsistenti z rezultati Aalenove in Coxove regresije. Ker je do sprememb učinkov različnih neodvisnih spremenljivk prišlo ob približno istem času, smo z uporabo Buckley-Jamesove metode na časovnih intervalih med prelomnimi točkami pokazali, da vsaj glede na smer učinka pridemo do podobnih rezultatov kot z ostalima uporabljenima metodama. S tem lahko tudi upravičimo uporabo ocenjenih pričakovanih vrednosti relativnega preživetja in pripadajočih dejanskih časov preživetja, do katerih pridemo z inverzno transformacijo individualne mere relativnega preživetja.

Na podlagi analize podatkov o raku na dojki, za katere je značilna visoka stopnja krnjjenja in s tem preživetja bolnic, se v kombinaciji z lastnostmi alternativne mere relativnega preživetja izkaže predvsem uporaba Aalenovega modela. Velja namreč, da s transformacijo opazovanega časa v individualno mero relativnega preživetja odstranimo razlike v pričakovanem populacijskem preživetju, kar pomeni, da so vse razlike, ki ostanejo, posledica proučevane bolezni. To pomeni, da z modeliranjem individualne mere relativnega preživetja dejansko proučujemo tisti del preživetja, ki je posledica presežne ogroženosti zaradi proučevane bolezni. Ker je individualna mera relativnega preživetja določena na podlagi populacijskih spremenljivk, se bo tudi presežna ogroženost pri modeliranju odrazila na učinku teh spremenljivk. V primeru podatkov o raku na dojki se izkaže, da starost in leto diagnoze na intervalu vrednosti individualne mere relativnega preživetja $0,1 < Y < 0,2$ postaneta statistično neznačilna. Na podlagi tega sklepamo, da bolnice, ki preživijo 10 do 20 % primerljive splošne populacije, statistično gledano ozdravijo. To predpostavko smo potrdili tudi s primerjavo na podlagi modelov in dejansko ocenjenih krivulj preživetja za te bolnice, s katerimi smo pokazali, da se njihova porazdelitev približa enakomerni porazdelitvi. Do podobnih zaključkov pridemo tudi na podlagi ocenjenih parametrov Coxove in Buckley-Jamesove regresije. Poleg tega pa ocenjene pričakovane vrednosti Buckley-Jamesove

regresije pokažejo, da obravnavane bolnice lahko pričakujejo, da bodo v povprečju preživele 50 % primerljive splošne populacije, kar lahko obravnavamo kot ozdravitev.

Ključne besede: preživetje, relativno preživetje, kumulativna funkcija relativnega preživetja, individualna mera relativnega preživetja, transformacijski pristop, transformiran čas preživetja, aditivni regresijski model, multiplikativni regresijski model, Aalenov aditivni model ogroženosti, Buckley-Jamesova linearna regresija transformiranih časov preživetja, Coxov model sorazmernih ogroženosti

Abstract

Alternative methods in modeling relative survival

Classic Survival Analysis deals with study of elapsed time between two well-defined events. Even though the studied time is a ratio variable, we are facing imperfect and censored data in survival analysis, which cannot be analysed by standard statistical methods. Instead, specific methods, which are known by the common term survival analysis and as event analysis in social sciences, have been developed to analyse this data. When observing units through longer period of time it can occur that end event cannot be well defined or linked to studied phenomenon (occurrence). In these cases classic survival analysis becomes useless, since considering all events would underestimate survival of units, while exclusion of events, which cannot be linked to the studied phenomenon, is impossible, because of unreliable or even unavailable information. Problems of this type can be solved inside of concept of relative survival, which attempts to exclude all of those events, which cannot be linked to the studied phenomenon. In this context occurrences of events linked to the studied phenomenon are compared to equivalent population (we will term it general population) events, which include all other causes that can affect the occurrence of end event beside the studied phenomenon. Relative survival methods are currently applied mainly in medicine, where patients are often studied through longer periods of time, but the actual cause of death often cannot be defined. In addition statistical offices are regularly updating mortality data, which enables comparison of observed survival with population survival and consequently application of relative survival methodology. Relative survival was first defined by cumulative relative survival function ($r(t)$) as a ratio between observed and population survival. Function $r(t)$ represents a group measure of relative survival, which for each value of observed survival time tells us, what is a share of survived observed units compared to a share of survived in general population. Two approaches exist for the purpose of modelling $r(t)$, which differ among each other in assumption of relationship between observed and population hazard. The first is additive regression approach, which presupposes that observed hazard can be expressed as a sum of population and excess hazard (excess hazard being reflection of studied

disease). The second is multiplicative regression approach, which defines observed hazard as product of population hazard and so called relative mortality. Despite two approaches, additive models in literature are almost exclusive, what is justified by adequacy argument for biological data. The scope of studying relative survival has expanded with individual measure of relative survival, which opposed to cumulative relative survival function provides a measure of survival relative to the background for each individual. It is founded on a simple transformation of observed survival time. The transformation is based on population mortality tables from which we can compute conditional distribution functions for residual lifetime for given age, sex and cohort year. Based on these CDFs we can convert observed survival time to the associated value on the CDF. The outcome we get through this transformation can be viewed as a random variable (as observed survival time is a random variable), which measures how long after the event and relative to the background population a person has lived. New measure of relative survival inherently treats relative survival differently. Besides that the definition of relative survival for individual units enables the use of classic survival analysis methodology for its study. The thesis is focused mainly on empirical modelling of relative survival through classic survival analysis methodology, which is otherwise rarely used in practice. The core of the thesis consists of Aalens linear regression for modelling hazards and Buckley-James linear regression for modelling transformed survival times. Mainly because of comparison to Aalens model the analysis are also done by Cox regression of proportional hazards, which is the most frequently used model for analysing survival data. All models have been tested with data on lung and breast cancer. The difference between data are most obvious with respect to the share of censored units and consequently to their survival. The approach to relative survival in the thesis is entirely new, and as such makes the possibility of in depth knowledge about relative survival, while its use enables the possibility to collect new information on relative survival and consequently contextual enrichment in the field of relative survival. An advantage of new approach to modelling relative survival is also that it doesn't demand any assumptions of relationship between observed and population hazard compared to additive and multiplicative regression methods in traditional analysis of relative survival. Empirical data analysis

has shown evaluated models can produce consistent information on the effect of independent variables to relative survival. Despite, caution is necessary when using these models, and interpret their results, especially when effects of independent variables are changing in time. It was shown in the thesis that when data are generated by Aalens model the parameters of Buckley-James regression depend on Aalens model parameters as well as combination of independent variables. In case of time-dependent effects of independent variables the parameters of Buckley-James model also depend on conditional distribution of studied time on time intervals with different parameter values of Aalens model. In case of unchanging parameters of Aalens model the use of Buckley-James model depends on degree of nonlinearity and heteroskedasticity, while in case of changing parameters of Aalens model the parameters of Buckley-James regression dont detect these changes. Despite, the simulations show that changes can be detected by Aalens model, when data are generated with classic linear regression in such a manner that values of parameters are changed at the same values of independent variables. While in practice we dont know and often even dont detect by goodness-of-fit tests, which model fits data the best, applicability of model's results can be justified by consistency of different models. In case of time dependent effects of independent variables, Buckley-James regression can be applied on intervals where parameters of Aalens model are at least approximately constant. The results of evaluated models have affirmed an advantage of great adjustability of Aalens regression, by which we have detected that the effects of some independent variables on relative survival change in time when applied to lung and breast cancer data. Although modelling of time-depending effects is possible in Cox model as well, its disadvantage is the necessity to adequately rearrange input data, while on the other hand a choice of turning points is completely arbitrary. The main advantage of Cox model is that the effects of independent variables can be quantitatively evaluated. Despite disadvantage of inability to detect changes of effects in time, Buckley-James model has given results, which are consistent with results of Aalens and Cox regression. Because effects of different independent variables were changed at approximately similar time, the application of Buckley-James method on time intervals between turning points has shown that at least by direction of effect we can achieve similar results as with other two ap-

plied methods. Consequently we can justify the use of estimated expected values of individual measure of relative survival and corresponding survival times, which we get as a result of inversed transformation of individual measure of relative survival. On the basis of breast cancer data analysis, which are characterised by high rate of censoring and consequently the survival of patients, in combination with alternative measure of relative survival properties Aalens model proves best. It is considered that transformation of observed time to individual measure of relative survival removes the differences in expected population survival, which means all differences that remain are the consequence of studied disease. Because individual measure of relative survival is defined on the basis of population variables, excessive hazard in modelling will also be reflected on the effect of these variables. It can be demonstrated in case of breast cancer data that age and year of diagnosis on the interval of individual measure of relative survival value $0,1 < Y < 0,2$ become statistically insignificant. Accordingly we can conclude that patients, who survive 10 to 20% of comparable general population, can statistically be considered cured. This assumption has also been confirmed by comparing survival curves estimated on the basis of models and observed data, which have demonstrated that their distribution is approximately uniform. Estimated parameters of Cox and Buckley-James regression produces similar conclusion. Beside, expected values of individual measure of relative survival estimated with Buckley-James regression show that treated patients can expect to survive 50% of comparable general population on average, which can be considered as being cured.

Keywords: survival, relative survival, cumulative relative survival function, individual measure of relative survival, transformation approach, transformed survival time, additive regression model, multiplicative regression model, Aalen's additive hazards model, Buckley-James linear regression of transformed survival times, Cox proportional hazards model

Kazalo

Kazalo	13
Slike	15
Tabele	19
1 Uvod	24
2 Relativno preživetje	28
2.1 Analiza preživetja in relativno preživetje	28
2.2 Definicija relativnega preživetja	32
2.3 Individualna mera relativnega preživetja	35
2.3.1 Lastnosti individualne mere relativnega preživetja	38
3 Obstojče regresijske metode v relativnem preživetju	40
3.1 Aditivni model	40
3.1.1 Kratek opis metod ocenjevanja aditivnega modela	42
3.2 Multiplikativni model	46
4 Alternativne regresijske metode v relativnem preživetju	47
4.1 Buckley-Jamesova regresija najmanjših kvadratov za krnjene podatke	48
4.1.1 Opis metode	49
4.2 Aalenov aditivni regresijski model	56
4.2.1 Opis metode	57
4.2.2 Testiranje hipotez o učinkih neodvisnih spremenljivk	62
4.2.3 Prileganje modela	64
4.3 Coxov model sorazmernih ogroženosti	65
4.3.1 Opis metode	66
4.3.2 Testiranje hipotez o učinkih neodvisnih spremenljivk	69
4.3.3 Prileganje modela	69
4.4 Prednosti uporabe alternativnih metod v analizi relativnega preživetja	73
4.5 Možne pomankljivosti proučevanih modelov v relativnem preživetju	75
5 Primerjava Aalenovega in Buckley-Jamesovega regresijskega modela	78
5.1 Teoretična primerjava	78
5.2 Simulacijska primerjava	81

6	Uporaba alternativnih regresijskih metod na dejanskih podatkih	95
6.1	Opis podatkov	95
6.2	Uporaba Aalenovega modela v relativnem preživetju	103
6.2.1	Aalenov model na podatkih o pljučnem raku	103
6.2.2	Aalenov model na podatkih o raku na dojki	124
6.3	Uporaba Coxovega modela sorazmernih ogroženosti v relativnem preživetju	149
6.3.1	Coxov model na podatkih pljučnega raka	149
6.3.2	Coxov model na podatkih o raku na dojki	172
6.4	Uporaba Buckley-Jamesovega linearnega modela v relativnem preživetju .	193
6.4.1	Buckley-Jamesova linearna regresija na podatkih o pljučnem raku .	193
6.4.2	Buckley-Jamesova linearna regresija na podatkih o raku na dojki .	215
7	Zaključek	235
8	Literatura	242
9	Priloga	252
9.1	Števni procesi in martingali	252
9.1.1	Osnove števnih procesov	252
9.1.2	Modeliranje števnih procesov	254
9.1.3	Osnove martingalov	255

Slike

Slika 2.1: Hipotetična klinična študija v desetletnem obdobju. Čas od vstopa pacientov v študijo do nastopa želenega dogodka v dejanskem koledarskem času (a); časi, ki jih upoštevamo za namene statističnih analiz, pri čemer privzamemo, da so vsi vstopni časi pacientov enaki nič (b). Zapolnjeni krogi označujejo nekrnjene podatke, prazni pa krnjene.	30
Slika 2.2: Porazdelitvena funkcija časa do smrti za moške (modra barva) in ženske (rdeča barva), ki so bili leta 1990 stari 58 let. Puščice označujejo izračun individualne mere relativnega preživetja po desetih letih opazovanja.	37
Slika 6.1: Porazdelitev neodvisnih spremenljivk, uporabljenih pri modeliranju podatkov o pljučnem raku	98
Slika 6.2: Funkcije preživetja transformiranega časa, ocenjene po Kaplan-Meierjevi metodi	99
Slika 6.3: Porazdelitev neodvisnih spremenljivk, uporabljenih pri modeliranju podatkov o raku na dojki	101
Slika 6.4: Funkcije preživetja transformiranega časa, ocenjene po Kaplan-Meierjevi metodi	102
Slika 6.5: Grafični prikaz posplošenih ostankov za (a) 1. stadij pljučnega raka, (b) 2. stadij pljučnega raka in (c) 3. stadij pljučnega raka . .	104
Slika 6.6: Ocijene kumulativne regresijske funkcije s 95-odstotnim intervalom zaupanja za podatke o pljučnem raku	108
Slika 6.7: Nadaljevanje slike 6.6	109
Slika 6.8: Nadaljevanje slike 6.6	110
Slika 6.9: Grafični prikaz posplošenih ostankov v enostavnem modelu za (a) 1. stadij pljučnega raka, (b) 2. stadij pljučnega raka in (c) 3. stadij pljučnega raka	113
Slika 6.10: Ocijene kumulativne regresijske funkcije s 95-odstotnim intervalom zaupanja za podatke pljučnega raka - enostavni model . .	114
Slika 6.11: Nadaljevanje slike 6.10	115
Slika 6.12: Ocijene kumulativne regresijske funkcije enostavnega modela v diskretni obliki originalnega vzorca (črna okrepljena črta) in desetih bootstrap vzorcev (barvne črte)	117

Slika 6.13:Na podlagi modelov ocenjene funkcije preživetja moških pacientov, diagnosticiranih v 1. stadiju pljučnega raka, za spremenljajoče vrednosti starosti v letu 1995 za (a) model z interakcijami in (b) enostavni model	119
Slika 6.14:Na podlagi modelov ocenjene funkcije preživetja moških pacientov, diagnosticiranih v 2. stadiju pljučnega raka, za spremenljajoče vrednosti starosti v letu 1995 za (a) model z interakcijami in (b) enostavni model	120
Slika 6.15:Na podlagi modelov ocenjene funkcije preživetja moških pacientov, diagnosticiranih v 3. stadiju pljučnega raka, za spremenljajoče vrednosti starosti v letu 1995 za (a) model z interakcijami in (b) enostavni model	120
Slika 6.16:Na podlagi modelov ocenjene funkcije preživetja moških pacientov, diagnosticiranih v 1. stadiju pljučnega raka, za spremenljajoče vrednosti leta diagnoze pri starosti 64 let za (a) model z interakcijami in (b) enostavni model	121
Slika 6.17:Na podlagi modelov ocenjene funkcije preživetja moških pacientov, diagnosticiranih v 2. stadiju pljučnega raka, za spremenljajoče vrednosti leta diagnoze pri starosti 64 let za (a) model z interakcijami in (b) enostavni model	121
Slika 6.18:Na podlagi modelov ocenjene funkcije preživetja moških pacientov, diagnosticiranih v 3. stadiju pljučnega raka, za spremenljajoče vrednosti leta diagnoze pri starosti 64 let za (a) model z interakcijami in (b) enostavni model	122
Slika 6.19:Na podlagi modelov ocenjene funkcije preživetja moških in ženskih pacientov, diagnosticiranih v 1. stadiju pljučnega raka, za starost 64 let in leto diagnoze 1995 za (a) model z interakcijami in (b) enostavni model	122
Slika 6.20:Na podlagi modelov ocenjene funkcije preživetja moških in ženskih pacientov, diagnosticiranih v 2. stadiju pljučnega raka, za starost 64 let in leto diagnoze 1995 za (a) model z interakcijami in (b) enostavni model	123
Slika 6.21:Na podlagi modelov ocenjene funkcije preživetja moških in ženskih pacientov, diagnosticiranih v 3. stadiju pljučnega raka, za starost 64 let in leto diagnoze 1995 za (a) model z interakcijami in (b) enostavni model	123
Slika 6.22:Grafični prikaz posplošenih ostankov za (a) 1. stadij raka na dojki, (b) 2. stadij raka na dojki in (c) 3. stadij raka na dojki	127
Slika 6.23:Ocenjene kumulativne regresijske funkcije s 95-odstotnim intervalom zaupanja za podatke o raku na dojki	128
Slika 6.24:Nadaljevanje slike 6.23	129
Slika 6.25:Nadaljevanje slike 6.23	130

Slika 6.26: Grafični prikaz posplošenih ostankov v enostavnem modelu za (a) 1. stadij raka na dojki, (b) 2. stadij raka na dojki in (c) 3. stadij raka na dojki	133
Slika 6.27: Ocenjene kumulativne regresijske funkcije s 95-odstotnim intervalom zaupanja za podatke raka na dojki - enostavni model	134
Slika 6.28: Nadaljevanje slike 6.27	135
Slika 6.29: Ocenjene kumulativne regresijske funkcije v diskretni obliki originalnega vzorca (črna okrepljena črta) in desetih bootstrap vzorcev (barvne črte) - model z interakcijami stadija s starostjo in letom diagnoze	137
Slika 6.30: Nadaljevanje slike 6.29	138
Slika 6.31: Ocenjene kumulativne regresijske funkcije v diskretni obliki originalnega vzorca (črna okrepljena črta) in desetih bootstrap vzorcev (barvne črte) - enostavni model	139
Slika 6.32: Na podlagi modelov ocenjene funkcije preživetja operativno zdravljenih pacientk za spremenjajoče vrednosti starosti v letu 1996 (neprekinjena črta - model z interakcijami, prekinjena črta - enostavni model) za (a) 1. stadij raka na dojki, (b) 2. stadij raka na dojki in (c) 3. stadij raka na dojki	141
Slika 6.33: Na podlagi modelov ocenjene funkcije preživetja operativno zdravljenih pacientk za spremenjajoče vrednosti leta diagnoze pri starosti 60 let (neprekinjena črta - model z interakcijami, prekinjena črta - enostavni model) za (a) 1. stadij raka na dojki, (b) 2. stadij raka na dojki in (c) 3. stadij raka na dojki	142
Slika 6.34: Ocenjene krivulje preživetja za operativno zdravljene paciente, diagnosticirane v prvem stadiju raka na dojki, za vrednosti transformiranega časa (a) $Y > 0,1$ in (b) $Y > 0,2$	144
Slika 6.35: Ocenjene krivulje preživetja za operativno zdravljene paciente, diagnosticirane v prvem stadiju raka na dojki, za vrednosti transformiranega časa (a) $Y > 0,1$ - dejanski podatki, (b) $Y > 0,1$ - model, (c) $Y > 0,2$ - dejanski podatki in (d) $Y > 0,2$ - model.	146
Slika 6.36: Ocenjene krivulje preživetja za operativno zdravljene paciente, diagnosticirane v drugem stadiju raka na dojki, od vrednosti transformiranega časa $Y > 0,4$	148
Slika 6.37: Zglajene krivulje vrednosti $U_{kj}^* + \hat{\beta}_j$ v času Y za podatke o pljučnem raku z vrisanimi mejami $\pm 2 \times$ standardna odklona	152
Slika 6.38: Nadaljevanje slike 6.37	153
Slika 6.39: Nadaljevanje slike 6.37	154
Slika 6.40: Grafični prikaz ocenjenih koeficientov Coxovega modela za podatke o pljučnem raku	161
Slika 6.41: Nadaljevanje slike 6.40	162
Slika 6.42: Nadaljevanje slike 6.40	163

Slika 6.43: Zglajene krivulje vrednosti $U_{kj}^* + \hat{\beta}_j$ v času Y za podatke pljučnega raka z vrisanimi mejami $\pm 2 \times$ standardna odklona	165
Slika 6.44: Nadaljevanje slike 6.43	166
Slika 6.45: Grafični prikaz ocenjenih koeficientov Coxovega modela za podatke o pljučnem raku	170
Slika 6.46: Nadaljevanje slike 6.45	171
Slika 6.47: Zglajene krivulje vrednosti $U_{kj}^* + \hat{\beta}_j$ v času Y za podatke o raku na dojki z vrisanimi mejami $\pm 2 \times$ standardna odklona	173
Slika 6.48: Nadaljevanje slike 6.47	174
Slika 6.49: Nadaljevanje slike 6.47	175
Slika 6.50: Grafični prikaz ocenjenih koeficientov Coxovega modela za podatke o raku na dojki	181
Slika 6.51: Nadaljevanje slike 6.50	182
Slika 6.52: Nadaljevanje slike 6.50	183
Slika 6.53: Zglajene krivulje vrednosti $U_{kj}^* + \hat{\beta}_j$ v času Y za podatke o raku na dojki z vrisanimi mejami $\pm 2 \times$ standardna odklona	185
Slika 6.54: Nadaljevanje slike 6.53	186
Slika 6.55: Grafični prikaz ocenjenih koeficientov Coxovega modela za podatke o raku na dojki	190
Slika 6.56: Nadaljevanje slike 6.55	191
Slika 9.1: Grafični prikaz procesov $N(t)$ in $Y(t)$ za nekrnjeno in krnjeno enoto	253

Tabele

Tabela 5.1: Parametri Buckley-Jamesovega modela pri v času konstantnih parametrih Aalenovega modela (Model 1)	87
Tabela 5.2: Parametri Buckley-Jamesovega modela pri v času konstantnih parametrih Aalenovega modela (Model 2)	88
Tabela 5.3: Parametri Buckley-Jamesovega modela pri časovno odvisnih parametrih Aalenovega modela, pri čemer je 20 % enot manjših od prelomne točke t^* (Model 1)	89
Tabela 5.4: Parametri Buckley-Jamesovega modela pri časovno odvisnih parametrih Aalenovega modela, pri čemer je 20 % enot manjših od prelomne točke t^* (Model 2)	90
Tabela 5.5: Parametri Buckley-Jamesovega modela pri časovno odvisnih parametrih Aalenovega modela, pri čemer je 50 % enot manjših od prelomne točke t^* (Model 1)	91
Tabela 5.6: Parametri Buckley-Jamesovega modela pri časovno odvisnih parametrih Aalenovega modela, pri čemer je 50 % enot manjših od prelomne točke t^* (Model 2)	92
Tabela 5.7: Parametri Buckley-Jamesovega modela pri časovno odvisnih parametrih Aalenovega modela, pri čemer je 80 % enot manjših od prelomne točke t^* (Model 1)	93
Tabela 5.8: Parametri Buckley-Jamesovega modela pri časovno odvisnih parametrih Aalenovega modela, pri čemer je 80 % enot manjših od prelomne točke t^* (Model 2)	94
 Tabela 6.1: TST testne statistike za Aalenov model na podatkih o pljučnem raku	106
Tabela 6.2: TST testne statistike za Aalenov enostavni model na podatkih o pljučnem raku	111
Tabela 6.3: TST testne statistike za Aalenov model na podatkih o raku na dojki	125
Tabela 6.4: TST testne statistike za Aalenov enostavni model na podatkih o raku na dojki	132
Tabela 6.5: Testne statistike za preverjanje predpostavke o sorazmernih ogroženostih	150
Tabela 6.6: Testne statistike za preverjanje predpostavke o sorazmernih ogroženostih	155
Tabela 6.7: Rezultati ocenjenega modela za podatke o pljučnem raku	158

Tabela 6.8: Rezultati ocenjenega modela za podatke o pljučnem raku - nadaljevanje tabele 6.7	159
Tabela 6.9: Rezultati ocenjenega modela za podatke o pljučnem raku - nadaljevanje tabele 6.7	160
Tabela 6.10: Testne statistike za preverjanje predpostavke sorazmernih ogroženosti	164
Tabela 6.11: Testne statistike za preverjanje predpostavke sorazmernih ogroženosti	167
Tabela 6.12: Rezultati ocenjenega modela za podatke o pljučnem raku	169
Tabela 6.13: Testne statistike za preverjanje predpostavke sorazmernih ogroženosti	172
Tabela 6.14: Testne statistike za preverjanje predpostavke sorazmernih ogroženosti	176
Tabela 6.15: Rezultati ocenjenega modela za podatke o raku na dojki	179
Tabela 6.16: Nadaljevanje tabele 6.15	180
Tabela 6.17: Testne statistike za preverjanje predpostavke o sorazmernih ogroženostih	184
Tabela 6.18: Testne statistike za preverjanje predpostavke o sorazmernih ogroženostih	187
Tabela 6.19: Rezultati ocenjenega modela za podatke o raku na dojki	189
Tabela 6.20: Waldove statistike za $\log(Y)$	195
Tabela 6.21: Rezultati ocenjenega modela za podatke o pljučnem raku	197
Tabela 6.22: Rezultati ocenjenega modela za podatke o pljučnem raku, ($Y \leq 0,03$)	198
Tabela 6.23: Rezultati ocenjenega modela za podatke o pljučnem raku, ($Y > 0,03$)	199
Tabela 6.24: Waldove statistike za $\log(Y)$, ($Y \leq 0,03$)	200
Tabela 6.25: Waldove statistike za $\log(Y)$, ($Y > 0,03$)	201
Tabela 6.26: Ocenjeni pričakovani transformirani časi za moške paciente, diagnosticirane v 1. stadiju pljučnega raka	205
Tabela 6.27: Ocenjeni pričakovani transformirani časi za moške paciente, diagnosticirane v 2. stadiju pljučnega raka	205
Tabela 6.28: Ocenjeni pričakovani transformirani časi za moške paciente, diagnosticirane v 3. stadiju pljučnega raka	206
Tabela 6.29: Ocenjeni pričakovani transformirani časi za patientke, diagnosticirane v 1. stadiju pljučnega raka	206
Tabela 6.30: Ocenjeni pričakovani transformirani časi za patientke, diagnosticirane v 2. stadiju pljučnega raka	207
Tabela 6.31: Ocenjeni pričakovani transformirani časi za patientke, diagnosticirane v 3. stadiju pljučnega raka	207
Tabela 6.32: Izvirni čas T , ki ustreza transformiranemu času $Y = 0,03$, za izbrano kombinacijo starosti, spola in leta diagnoze	208
Tabela 6.33: Ocenjeni pričakovani transformirani časi za moške paciente, diagnosticirane v 1. stadiju pljučnega raka; parametri modela, ocenjeni na podatkih pod prelomno točko $Y \leq 0,03$	209
Tabela 6.34: Ocenjeni pričakovani transformirani časi za moške paciente, diagnosticirane v 2. stadiju pljučnega raka; parametri modela, ocenjeni na podatkih pod prelomno točko $Y \leq 0,03$	209

Tabela 6.35: Ocenjeni pričakovani transformirani časi za moške paciente, diagnosticirane v 3. stadiju pljučnega raka; parametri modela, ocenjeni na podatkih pod prelomno točko $Y \leq 0,03$	210
Tabela 6.36: Ocenjeni pričakovani transformirani časi za patientke, diagnosticirane v 1. stadiju pljučnega raka; parametri modela, ocenjeni na podatkih pod prelomno točko $Y \leq 0,03$	210
Tabela 6.37: Ocenjeni pričakovani transformirani časi za patientke, diagnosticirane v 2. stadiju pljučnega raka; parametri modela, ocenjeni na podatkih pod prelomno točko $Y \leq 0,03$	211
Tabela 6.38: Ocenjeni pričakovani transformirani časi za patientke, diagnosticirane v 3. stadiju pljučnega raka; parametri modela, ocenjeni na podatkih pod prelomno točko $Y \leq 0,03$	211
Tabela 6.39: Ocenjeni pričakovani transformirani časi za moške paciente, diagnosticirane v 1. stadiju pljučnega raka; parametri modela, ocenjeni na podatkih pod prelomno točko $Y > 0,03$	212
Tabela 6.40: Ocenjeni pričakovani transformirani časi za moške paciente, diagnosticirane v 2. stadiju pljučnega raka; parametri modela, ocenjeni na podatkih pod prelomno točko $Y > 0,03$	212
Tabela 6.41: Ocenjeni pričakovani transformirani časi za moške paciente, diagnosticirane v 3. stadiju pljučnega raka; parametri modela, ocenjeni na podatkih pod prelomno točko $Y > 0,03$	213
Tabela 6.42: Ocenjeni pričakovani transformirani časi za patientke, diagnosticirane v 1. stadiju pljučnega raka; parametri modela, ocenjeni na podatkih pod prelomno točko $Y > 0,03$	213
Tabela 6.43: Ocenjeni pričakovani transformirani časi za patientke, diagnosticirane v 2. stadiju pljučnega raka; parametri modela, ocenjeni na podatkih pod prelomno točko $Y > 0,03$	214
Tabela 6.44: Ocenjeni pričakovani transformirani časi za patientke, diagnosticirane v 3. stadiju pljučnega raka; parametri modela, ocenjeni na podatkih pod prelomno točko $Y > 0,03$	214
Tabela 6.45: Waldove statistike za $\log(Y)$	216
Tabela 6.46: Rezultati ocenjenega modela za podatke o raku na dojki	217
Tabela 6.47: Waldove statistike za $\log(Y)$, $Y \leq 0,15$	218
Tabela 6.48: Waldove statistike za $\log(Y)$, $Y > 0,15$	219
Tabela 6.49: Rezultati ocenjenega modela za podatke o raku na dojki, $Y \leq 0,15$	219
Tabela 6.50: Rezultati ocenjenega modela za podatke o raku na dojki, $Y > 0,15$	220
Tabela 6.51: Waldove statistike za $\log(Y)$, $Y > 0,15$, enostavni model	220
Tabela 6.52: Rezultati ocenjenega modela za podatke o raku na dojki, $Y > 0,15$, enostavni model	221
Tabela 6.53: Ocenjeni pričakovani transformirani časi za patientke brez operativnega zdravljenja, diagnosticirane v 1. stadiju raka na dojki	224
Tabela 6.54: Ocenjeni pričakovani transformirani časi za patientke brez operativnega zdravljenja, diagnosticirane v 2. stadiju raka na dojki	224

Tabela 6.55: Ocenjeni pričakovani transformirani časi za pacientke brez operativnega zdravljenja, diagnosticirane v 3. stadiju raka na dojki	225
Tabela 6.56: Ocenjeni pričakovani transformirani časi za operativno zdravljene patientke, diagnosticirane v 1. stadiju raka na dojki	225
Tabela 6.57: Ocenjeni pričakovani transformirani časi za operativno zdravljene patientke, diagnosticirane v 2. stadiju raka na dojki	226
Tabela 6.58: Ocenjeni pričakovani transformirani časi za operativno zdravljene patientke, diagnosticirane v 3. stadiju raka na dojki	226
Tabela 6.59: Izvirni čas T , ki ustreza transformiranemu času $Y = 0,15$, za izbrano kombinacijo starosti in leta diagnoze	228
Tabela 6.60: Ocenjeni pričakovani transformirani časi za patientke brez operativnega zdravljenja, diagnosticirane v 1. stadiju raka na dojki, $Y \leq 0,15$	228
Tabela 6.61: Ocenjeni pričakovani transformirani časi za patientke brez operativnega zdravljenja, diagnosticirane v 2. stadiju raka na dojki, $Y \leq 0,15$	229
Tabela 6.62: Ocenjeni pričakovani transformirani časi za patientke brez operativnega zdravljenja, diagnosticirane v 3. stadiju raka na dojki, $Y \leq 0,15$	229
Tabela 6.63: Ocenjeni pričakovani transformirani časi za operativno zdravljene patientke, diagnosticirane v 1. stadiju raka na dojki, $Y \leq 0,15$	230
Tabela 6.64: Ocenjeni pričakovani transformirani časi za operativno zdravljene patientke, diagnosticirane v 2. stadiju raka na dojki, $Y \leq 0,15$	230
Tabela 6.65: Ocenjeni pričakovani transformirani časi za operativno zdravljene patientke, diagnosticirane v 3. stadiju raka na dojki, $Y \leq 0,15$	231
Tabela 6.66: Ocenjeni pričakovani transformirani časi za patientke brez operativnega zdravljenja, diagnosticirane v 1. stadiju raka na dojki, $Y > 0,15$	231
Tabela 6.67: Ocenjeni pričakovani transformirani časi za patientke brez operativnega zdravljenja, diagnosticirane v 2. stadiju raka na dojki, $Y > 0,15$	232
Tabela 6.68: Ocenjeni pričakovani transformirani časi za patientke brez operativnega zdravljenja, diagnosticirane v 3. stadiju raka na dojki, $Y > 0,15$	232
Tabela 6.69: Ocenjeni pričakovani transformirani časi za operativno zdravljene patientke, diagnosticirane v 1. stadiju raka na dojki, $Y > 0,15$	233
Tabela 6.70: Ocenjeni pričakovani transformirani časi za operativno zdravljene patientke, diagnosticirane v 2. stadiju raka na dojki, $Y > 0,15$	233

Tabela 6.71: Ocenjeni pričakovani transformirani časi za operativno zdravljene pacientke, diagnosticirane v 3. stadiju raka na dojki, $Y > 0,15$	234
--	-----

1. Uvod

Analiza preživetja obsega metode, s katerimi proučujemo čas, ki je potreben do nastopa nekega dogodka in ga običajno imenujemo čas preživetja. Posebne metode za proučevanje časov preživetja so bile razvite, ker so podatki zaradi narave njihovega zbiranja običajno nepopolni oziroma krnjeni, kar zahteva drugačna orodja za njihovo analizo. Klasična analiza preživetja predpostavlja, da sta začetni in končni dogodek dobro definirana, vendar je zlasti kontrola končnega dogodka v praksi mnogokrat onemogočena. Nadzor končnega dogodka je še zlasti težaven v primerih, ko enote spremljamo skozi daljše časovno obdobje in končnega dogodka ne moremo z gotovostjo pripisati proučevanemu pojavu. V takih primerih so rezultati klasične analize preživetja pristranski. Ob upoštevanju vseh dogodkov namreč preživetje enot podcenimo, izločitev dogodkov, ki jih ne moremo pripisati proučevanemu pojavu, pa je zaradi nezanesljivih ali pogosto celo nerazpoložljivih informacij onemogočena. Pri proučevanju preživetja si v tem primeru lahko pomagamo s konceptom relativnega preživetja. V okviru tega nastope dogodkov, povezane s proučevanim pojavom, primerjamo z enakovrednimi dogodki populacije (recimo ji splošna populacija), ki poleg proučevanega pojava vključujejo tudi vse druge vzroke, ki lahko vplivajo na nastop končnega dogodka. V praksi je uporaba relativnega preživetja na eni strani odvisna od spremeljanja oziroma zbiranja podatkov o proučevanem dogodku na celotni populaciji, na drugi strani pa je zaradi trenutno razvite metodologije relativnega preživetja omejena z relativno velikostjo populacije, na kateri proučujemo določeni pojav, v primerjavi s splošno populacijo, na kateri se zgodi opazovan dogodek. Predpostavlja se namreč, da je opazovana populacija v primerjavi s splošno populacijo dovolj majhna, tako da statistične parametre, izračunane na splošni populaciji, lahko obravnavamo kot populacijske vrednosti.

Proučevanje relativnega preživetja je bilo dolgo vezano na primerjavo med skupino opazovanih enot in primerljivo skupino splošne populacije. Pojav individualne mere relativnega preživetja (Stare in drugi 2005a) pa je omogočil proučevanje relativnega preživetja na posameznih enotah, s čimer se je prostor proučevanja relativnega preživetja razširil. Individualna mera relativnega preživetja v primerjavi s prvotno definirano skupinsko mero že sama po sebi prinaša drugačen vpogled v relativno preživetje, poleg tega pa omogoča, da za njeno proučevanje uporabimo metode klasične analize preživetja. Doktorska disertacija se osredotoča predvsem na modeliranje relativnega preživetja na podlagi metodologije klasične analize preživetja, ki je sicer v praksi redkeje uporabljenega. Tako v nalogi proučujem uporabo Aalenovega linearne modela ogroženosti in Buckley-Jamesove linearne regresije v relativnem preživetju; predvsem zaradi primerjave z Aalenovim modelom pa so v nalogi predstavljeni tudi rezultati v praksi najpogosteje uporabljenega Coxovega modela. Vsi modeli so preizkušeni na empiričnih podatkih o raku na pljučih in raku na dojki, ki se med seboj najbolj razlikujejo glede na delež krnjениh enot. S tehničnega vidika je prednost modeliranja relativnega preživetja z individualno mero relativnega preživetja predvsem ta, da v primerjavi s prvotno razvitimi metodami (aditivni in multiplikativni modeli relativnega preživetja) ne predpostavlja povezanosti med opazovano in populacijsko ogroženostjo. Kršenje te predpostavke v primeru uporabe tradicionalnih metod povzroča, da so dobljeni rezultati pristranski. Z vsebinskega vidika uporaba različnih modelov prinaša nove informacije o relativnem preživetju. Modeliranje z individualno mero relativnega preživetja predstavlja popolnoma nov pristop k proučevanju relativnega preživetja, zato na eni strani ponuja poglobljeno vedenje o relativnem preživetju, na drugi strani pa vsebinsko obogatitev področja relativnega preživetja. Tako nam Buckley-Jamesova regresija omogoča napovedovanje pričakovanih vrednosti individualne mere relativnega preživetja, ki je mogoča le z uporabo ocenjenih parametrov modela. Pri ostalih regresijskih metodah relativnega preživetja je za napovedovanje treba oceniti tudi osnovno ogroženost, za kar pa je potrebna vsakokratna uporaba celotnega vzorca. S pomočjo inverzne transformacije napovedane vrednosti individualne mere relativnega preživetja lahko izračunamo tudi pripadajoče pričakovane čase preživetja, kar z obstoječimi metodami ni mogoče. Na drugi strani Aalenov model zaradi svoje prilagodljivosti

omogoča enostaven in hiter vpogled v časovno spreminjanje učinkov neodvisnih spremenljivk na relativno preživetje. Takšen pristop pa je v kombinaciji z nekaterimi lastnostmi individualne mere lahko zelo koristen. Velja namreč, da je porazdelitev individualne mere relativnega preživetja, v primeru ko na preživetje vplivajo samo populacijske ogroženosti, enakomerna na intervalu $[0, 1]$. V tem primeru spremenljivke, na podlagi katerih je definirana individualna mera relativnega preživetja, nimajo vpliva na opazovano ogroženost. To pomeni, da v primeru, ko z Aalenovim modelom ne zaznamo statistično značilnih vplivov obravnavanih spremenljivk, opazovane enote z vidika ogroženosti ne odstopajo od populacije. V primeru podatkov o preživetju bolnikov z določeno bolezni tako informacijo lahko interpretiramo kot čas statistične ozdravitve, saj se preživetje opazovanih bolnikov izenači s preživetjem splošne populacije.

Bistveni del naloge in izviren prispevek na področju modeliranja relativnega preživetja je proučitev možnosti uporabe Buckley-Jamesovega, Aalenovega in Coxovega modela v relativnem preživetju. Namen naloge ni zagotoviti nadomestke obstoječim metodam relativnega preživetja, temveč proučiti možnosti alternativnih načinov analiziranja podatkov. To bi omogočilo pridobitev novih informacij in osvetlitev problema relativnega preživetja z novega zornega kota. Analize podatkov, ki so prikazane v nalogi, niso namenjene preverjanju kliničnih predpostavk proučevanih bolezni, temveč metodološkemu prikazu delovanja uporabljenih modelov. Razlog za to je tudi v tem, da registrski podatki ne ponujajo zadostnih informacij, ki bi zadostovale za namene kliničnih raziskav.

Doktorska disertacija je sicer sestavljena iz sedmih poglavij. V drugem poglavju je najprej predstavljena osnovna ideja preživetja in povezava z relativnim preživetjem, v nadaljevanju pa je opisan koncept skupinske in individualne mere relativnega preživetja.

Tretje in četrto poglavje opisujeta "tradicionalne" regresijske metode za modeliranje relativnega preživetja in v nalogi uporabljene alternativne metode, s katerimi smo modelirali relativno preživetje. Poglavlje je namenjeno razumevanju razlik med obema pristopoma, na kratko pa so predstavljene tudi glavne prednosti in pomankljivosti

uporabljenih alternativnih metod.

V petem poglavju je predstavljena primerjava med Aalenovo in Buckley-Jamesovo regresijo. Zanima nas predvsem, kako spremenjajoči učinki parametrov v Aalenovem modelu vplivajo na rezultate Buckley-Jamesovega modela.

V šestem poglavju je predstavljena uporaba alternativnih metod v relativnem preživetju na empiričnih podatkih. V ta namen smo uporabili registrske podatke o raku na pljučih in raku na dojki. Registrski podatki o raku predstavljajo tipičen primer podatkov, na katerih uporabljamо metode relativnega preživetja. Rakave bolnike namreč spremjamо skozi daljše časovno obdobje, natančnega vzroka smrti pa mnogokrat ne poznamo. Uporabljeni podatki se med seboj najbolj razlikujejo glede na odstotek krnjenih enot, kar lahko vpliva na rezultate uporabljenih metod. Poglavlje je namenjeno predstavitvi možnosti uporabe alternativnih metod v praksi, v njem pa je predstavljeno, do kakšnih informacij o relativnem preživetju lahko pridemo z njimi.

2. Relativno preživetje

2.1 Analiza preživetja in relativno preživetje

Analiza preživetja (ang. *survival analysis*), v družboslovju bolj znana kot analiza dogodkov (ang. *event history analysis*), proučuje slučajne pojave v času. Za razliko od presečnih analiz, kjer slučajni dogodek opazujemo v eni časovni točki, ali longitudinalnih analiz, pri katerih slučajni dogodek opazujemo v več časovnih točkah, analiza preživetja proučuje čas, ki je potreben do nastopa nekega dogodka. Vzemimo za primer trajanje brezposelnosti opazovanih oseb. Pri presečni analizi lahko opazujemo trajanje brezposelnosti opazovanih oseb v določeni časovni točki, pri longitudinalnih študijah trajanje brezposelnosti opazovanih oseb opazujemo v več časovnih točkah, medtem ko analiza preživetja proučuje čas, v katerem se brezposelna oseba zaposli ali pa postane neaktivna.

Opazovana slučajna spremenljivka v klasični analizi preživetja je torej čas med dvema dobro definiranimi dogodkoma. Primeri podatkov klasične analize preživetja oziroma analize dogodkov so denimo:

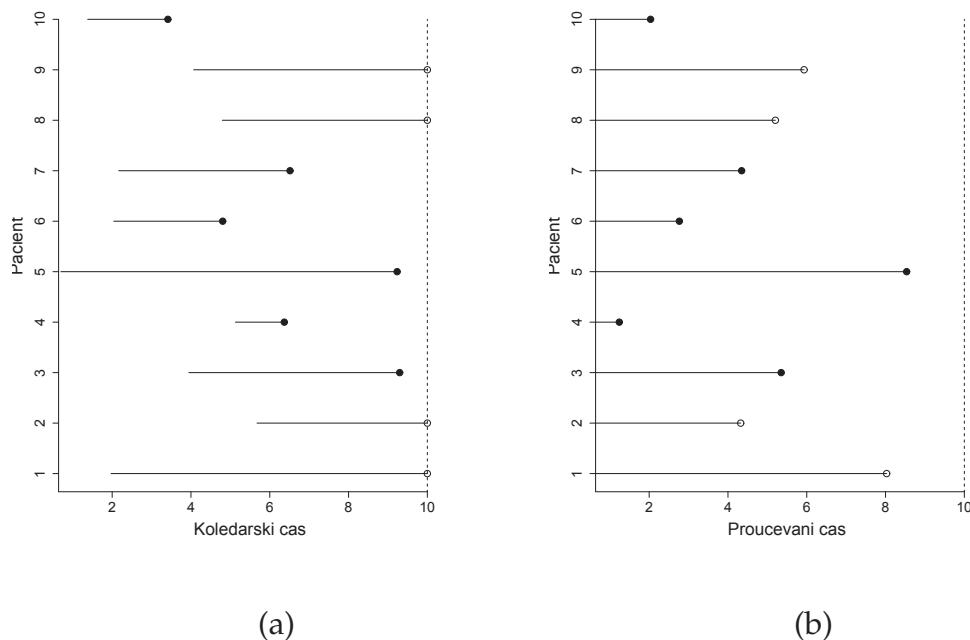
- čas od rojstva do smrti,
- čas od diagnoze bolezni do smrti,
- čas trajanja zaposlitve na določenem delovnem mestu,
- čas trajanja brezposelnosti,
- čas od poroke do ločitve
- in podobno.

Čas od začetnega do končnega oziroma želenega dogodka običajno imenujemo čas preživetja. Čeprav je čas preživetja dejansko razmernostna spremenljivka, so zbrani podatki zaradi narave njihovega zbiranja pogosto nepopolni. Pri pridobivanju podatkov moramo čakati do nastopa želenega dogodka, pri čemer je čas opazovanja običajno omejen s časovnicami zastavljenih raziskav. Po končanem opazovanju so podatki o časih preživetja opazovanih enot zbrani, dogodek pa v opazovanem času običajno ne nastopi pri vseh enotah. Tako na primer pri opazovanju brezposelnih nekateri v času opazovanja lahko najdejo zaposlitev ali preidejo v stanje neaktivnosti, drugi pa ne. Sicer se tisti, ki v času opazovanja ostanejo brezposelni, lahko zaposlijo ali postanejo neaktivni kasneje, vendar te informacije pri analiziranju zbranih podatkov nimamo. Zbrane podatke torej sestavljajo popolni in nepopolni podatki, v čemer se tudi bistveno razlikujejo od večine drugih statističnih podatkov. To pa zahteva drugačna orodja za njihovo analizo.

Nepopolni podatki se v analizi preživetja imenujejo krnjeni podatki (ang. *censored data*). Za primer je na sliki 2.1 prikazano, kako pride do krvnenja. Slika prikazuje hipotetično klinično študijo v obdobju desetih let. V tem obdobju je v študijo vstopilo deset pacientov, njihovi vstopi pa so se zgodili v različnih časih. Pacienti so bili opazovani do nastopa želenega dogodka (npr. smrt pacienta, ponovitev bolezni, konec bolezni in podobno) oziroma do zaključka študije. Vstopi pacientov in njihovi časi do nastopa želenega dogodka v koledarskem času so prikazani na sliki 2.1(a). Pri statističnih analizah se običajno osredotočimo na čas od vstopa v študijo do nastopa želenega dogodka. V ta namen privzamemo, da so vstopni časi pacientov nič in tako dobimo čase proučevanja (slika 2.1(b)). Vidimo lahko, da se je dogodek v času študije zgodil pri pacientih 3, 4, 5, 6, 7 in 10. Pri teh pacientih so podatki popolni oziroma nekrnjeni. Pri pacientih 1, 2, 8 in 9 pa dogodek do zaključka študije ni nastopil, njihovi opazovani časi preživetja pa so tako nepopolni oziroma krvnjeni. V prikazanem enostavnem primeru je krvnenje posledica zaključka študije, vendar do krvnenja lahko pride tudi zaradi prostovoljnega umika enot iz študije, enote pa v času opazovanja zaradi neznanih razlogov lahko tudi izgubimo. Krvnjene podatke obravnavamo kot pare vrednosti spremenljivk (Z_i, δ_i) , $i = 1, \dots, n$, kjer je Z_i opazovani čas enote, δ_i pa indikatorska spremenljivka, ki označuje, ali je pri opazovani enoti dogodek nastopil

$(\delta_i = 1)$ ali ne $(\delta_i = 0)$.

Slika 2.1: Hipotetična klinična študija v desetletnem obdobju. Čas od vstopa pacientov v študijo do nastopa želenega dogodka v dejanskem koledarskem času (a); časi, ki jih upoštevamo za namene statističnih analiz, pri čemer privzamemo, da so vsi vstopni časi pacientov enaki nič (b). Zapolnjeni krogi označujejo nekrnjene podatke, prazni pa krnjene.



Krnjeni podatki onemogočajo uporabo običajnih statističnih metod za analizo časov preživetja. Iz krnjениh podatkov namreč ne moremo izračunati niti enostavnega povprečja, torej je onemogočeno tudi ocenjevanje standardnega odklona, izvedba t-testa, regresijske analize in podobno. Prav zato so bile za analizo časov preživetja v primeru krnjениh podatkov razvite posebne metode, ki jih pod skupnim imenom poznamo kot analiza preživetja.

V analizi preživetja prevladujeta dva koncepta za analizo krnjениh podatkov in sicer funkcija preživetja (ang. *survival function*) in stopnja ogroženosti (ang. *hazard rate*) (Aalen in drugi 2008). Funkcija preživetja je definirana kot verjetnost, da se dogodek v času t še ni zgodil ozziroma kot delež enot, pri katerih dogodek v času t še ni nastopil:

$$S(t) = P(T > t) = 1 - F(t), \quad (2.1)$$

kjer je $F(t)$ porazdelitvena funkcija časa do nastopa dogodka.

Funkcija preživetja je definirana kot brezpogojna verjetnost, stopnjo ogroženosti pa definiramo s pomočjo pogojne verjetnosti. Predstavlja namreč verjetnost, da se dogodek zgodi v nekem majhnem časovnem intervalu $[t, t + dt]$, pod pogojem, da dogodek do časa t še ni nastopil. Natančneje, stopnja ogroženosti je definirana kot limita te verjetnosti, ko gre dt proti nič:

$$\lambda(t) = \lim_{\Delta t} \frac{1}{\Delta t} P(t \leq T < t + \Delta t | T \geq t). \quad (2.2)$$

Medtem ko vrednosti funkcije preživetja monotono padajo od ena proti nič (pri čemer vrednost ena zavzame, ko dogodka ni doživela še nobena enota, vrednost nič pa, ko dogodek nastopi pri vseh enotah), vrednosti funkcije ogroženosti lahko zavzamejo katerokoli nenegativno vrednost.

Klasična analiza preživetja predpostavlja, da sta začetni in končni dogodek dobro definirana. Medtem ko začetni dogodek lahko definiramo tako, da je ves čas pod nadzorom, je končni dogodek mnogokrat težje nadzorovati. Nadzor končnega dogodka je še toliko težavnejši v primerih, ko enote opazujemo skozi daljše časovno obdobje. Za analizo preživetja pacientov na primer, ki jih opazujemo od začetka diagnoze bolezni do smrti, je torej pomembno, da je vzrok smrti pacientov dobro opredeljen in znan. Vendar pa se v praksi pogosto soočimo z dejstvom, da so informacije o vzroku smrti pacientov nezanesljive ali celo nerazpoložljive. V tem primeru proučujemo paciente, ki so umirali iz različnih vzrokov, ocenjevanje preživetja ozira na smrtnosti zaradi bolezni pacientov pa postane nemogoče. Ob upoštevanju vseh smrti bi preživetje pacientov zaradi proučevane bolezni podcenili, izločitev vzrokov smrti, ki jih ne moremo pripisati bolezni pacientov, pa je zaradi nezanesljivih ali pogosto celo nerazpoložljivih podatkov onemogočena. Pogosto se tudi zgodi, da smrti pacientov ne moremo pripisati enemu samemu vzroku. Kot primer navedimo zdravljenje rakavih pacientov, ki povzroča hude stranske učinke in pogosto tudi smrt pacientov. V tem primeru denimo, je smrt le posredno povezana s proučevano boleznjijo. Podoben primer je samomor

rakavega pacienta. V takšnih situacijah se pojavi vprašanje, kako proučevati preživetje oziroma smrtnost kot posledico bolezni patientov. Odgovor na vprašanje ponuja analiza relativnega preživetja, ki poenostavljeno povedano smrtnost patientov primerja s smrtnostjo splošne populacije, s pomočjo katere iz ocene preživetja patientov skušamo izločiti vse tiste vzroke smrti, ki jih ne moremo pripisati proučevani bolezni. Relativno preživetje v tem smislu interpretiramo kot presežno smrtnost, ki je posledica bolezni proučevanih patientov.

2.2 Definicija relativnega preživetja

Formalno gledano relativno preživetje običajno definiramo kot razmerje med opazovanim preživetjem skupine patientov (označimo ga z $S_O(t)$) in pričakovanim preživetjem primerljive skupine splošne populacije ($S_P(t)$), ki ga imenujemo tudi populacijsko preživetje (Ederer in drugi 1961):

$$r(t) = \frac{S_O(t)}{S_P(t)}. \quad (2.3)$$

V zgornji enačbi je $r(t)$ kumulativna funkcija relativnega preživetja, ki nam pove, kakšen je delež preživelih patientov v primerjavi z deležem preživelih v splošni populaciji. V primerjavi s funkcijo preživetja ($S(t)$) kumulativna funkcija relativnega preživetja ni nujno monotono padajoča. Poleg tega njene vrednosti niso omejene z eno, kar pomeni, da je ne moremo interpretirati kot verjetnost preživetja. Vrednosti $r(t)$, ki so večje od ena pomenijo, da je preživetje opazovanih patientov boljše od preživetja splošne populacije, kar je lahko posledica izboljšanega zdravljenja in/ali izboljšane medicinske oskrbe bolnikov ali pa na primer izboljšanega življenjskega sloga, do katerega lahko pride zaradi izkušnje bolnikov z boleznijo.

Omenili smo že, da kumulativno funkcijo relativnega preživetja ocenimo tako, da opazovano preživetje skupine patientov primerjamo s podobno skupino splošne populacije v istem časovnem obdobju. Najprimernejša skupina splošne populacije bi se s skupino opazovanih patientov morala skladati v vseh dejavnikih, ki vplivajo na

preživetje, z izjemo same bolezni pacientov. Na ta način bi z oceno relativnega preživetja dobili neposredno informacijo o presežni smrtnosti kot posledici bolezni proučevanih pacientov. Dejansko so dostopni podatki le o nakaterih dejavnikih, ki vplivajo na preživetje splošne populacije. Tako imenovane tablice smrtnosti, na podlagi katerih izračunamo populacijsko preživetje, so namreč običajno razdeljene glede na starost, spol in kohortno leto. Tablice smrtnosti, ki so prikazane ločeno po spolu, prikazujejo kakšen delež ljudi, ki so v začetku izbranega leta dosegli določeno starost, je v tem letu tudi umrl. Ti deleži ocenjujejo verjetnost za smrt posameznika od začetka do konca izbranega leta, pod pogojem, da je posameznik živel do izbranega leta. Deleže obravnavamo kot populacijske vrednosti, kar pomeni, da varianco relativnega preživetja izračunamo kot:

$$var(r(t)) = \frac{var(S_O(t))}{S_P^2(t)}. \quad (2.4)$$

Pri oceni relativnega preživetja torej predpostavljam, da opazovani pacienti predstavljajo vzorec istospolne populacije, ki je v določenem letu isto stara in je z vidika vseh ostalih dejavnikov, ki vplivajo na preživetje, slučajen. Temu seveda v praksi pogosto ni zadoščeno. Za primer vzemimo paciente, ki so doživeli akutni miokardialni infarkt. Med pomembnejšimi dejavniki, ki vplivajo na preživetje teh pacientov, so poleg starosti in spola tudi dejavniki, kot so kajenje, povišana raven slabega holesterola v krvi, diabetes, kronične bolezni ledvic in podobno. Predpostavljam lahko, da je v primerjavi s splošno populacijo med pacienti, ki so doživeli miokardialni infarkt, znatno višji delež kadilcev, diabetikov, ljudi s povišano ravnjo slabega holesterola, ljudi s kroničnimi boleznimi ledvic in tako dalje. Pri tem moramo upoštevati, da ti dejavniki ne povzročajo le miokardialnega infarkta, temveč tudi druge bolezni, kar pomeni, da je preživetje zaradi naštetih dejavnikov v primerjavi s splošno populacijo manjše tudi pri tistih, ki niso doživeli miokardialnega infarkta. Pri izračunu pričakovanega preživetja teh dejavnikov zaradi nerazpoložljivih podatkov ne moremo upoštevati, s čimer pa oceno relativnega preživetja podcenimo. Kljub tej pomankljivosti metode relativnega preživetja ponujajo koristen vpogled v podatke preživetja, če smo pri interpretaciji rezultatov pozorni na dano strukturo tablic smrtnosti.

Pričakovano preživetje lahko iz tablic smrtnosti izračunamo za vsakega pacienta pose-

bej ali pa za skupino pacientov kot celoto. Za posameznega pacienta to naredimo tako, da za splošno populacijo, ki je bila v času začetnega dogodka (npr. diagnoze pacienta) iste starosti in spola kot pacient, izračunamo verjetnost preživetja do končnega dogodka pacienta. Glede na to, da tablice smrtnosti podajajo verjetnost za smrt od začetka (1. januar) do konca (31. december) izbranega leta ($S_P(1Y)$), verjetnost preživetja na intervalu dolžine $t < 365$ dni določimo kot:

$$S_P(t) = S_P(1Y)^{t/365.24} \quad (\text{Pohar Perme 2007}). \quad (2.5)$$

Intervale z dolžino, krajšo od enega leta, upoštevamo, kadar imamo podrobne podatke o nastopu začetnega dogodka in rojstnem datumu pacienta.

Za izračun pričakovanega preživetja skupine pacientov obstaja več metod, med najbolj znanimi pa so prva Edererjeva metoda, druga Edererjeva metoda in Hakulinenova metoda. Pričakovano preživetje najenostavnejše izračunamo kot povprečje pričakovanih preživetij posameznih opazovanih pacientov (prva Edererjeva metoda) (Ederer in drugi 1961). V primeru ko je krnjenje neodvisno od spremenljivk, na podlagi katerih so določene tablice umrljivosti, nam ta način da nepristransko oceno populacijskega preživetja za skupino opazovanih pacientov. V primeru ko paciente opazujemo skozi daljše časovno obdobje, pa predpostavka o neodvisnem krnjenu pogosto ne drži več. Kadar se spolna in starostna struktura opazovanih, ki so v študijo vstopili v začetnem obdobju, in tistih, ki so v študijo vstopili proti koncu obdobja, bistveno razlikuje, so potencialni časi krnjenga (časi do konca študije) odvisni od starosti in spola. Posledica pa je pristranska ocena populacijskega preživetja. To pomankljivost skuša odpraviti druga Edererjeva metoda (Ederer in Heise 1959), ki pričakovano preživetje namesto na vseh opazovanih pacientih računa samo na podlagi tistih, ki so v času t še v tveganju (do tedaj še niso umrli ali bili krnjeni). Čeprav ta metoda odpravlja pristranskost populacijskega preživetja zaradi odvisnosti krnjenga od starosti in spola opazovanih enot, pa je odvisna od dejanskih smrti pacientov. To povečuje varianco populacijskega preživetja, zaradi česar ta ni več zanemarljiva v primerjavi z varianco opazovanega preživetja. Hakulinenova metoda (Hakulinen 1982), podobno kot druga Edererjeva, upošteva možnost heterogenih časov vstopov v študijo, vendar hkrati ne upošteva dejanskih časov smrti. Temu se izogne tako, da dejanske čase opazovanj

nadomesti s potencialnimi časi opazovanj. Povedano drugače, časi krnjenja ostanejo nespremenjeni, časi smrti pa se nadomestijo s časi od diagnoze do zaključka študija. Izračun pričakovanega preživetja je potem podoben kot pri drugi Edererjevi metodi, pri čemer množico enot v tveganju določimo na podlagi potencialnih časov opazovanj. Na ta način se upošteva heterogenost vstopnih časov v študijo, hkrati pa pričakovano preživetje ni odvisno od opazovanih smrti, zato se varianca pričakovanega preživetja ne povečuje z naraščanjem števila umrlih pacientov. Hakulinenova metoda trenutno velja za najboljšo metodo računanja pričakovanega preživetja za oceno relativnega preživetja v obliki (2.3).

2.3 Individualna mera relativnega preživetja

Informacija o relativnem preživetju, ki jo dobimo z oceno kumulativne funkcije relativnega preživetja, se nanaša na skupino pacientov. Tako na primer vrednost kumulativne funkcije relativnega preživetja $r(t = 1 \text{ leto}) = 0.4$ pove, da delež preživelih pacientov predstavlja 0.4 deleža preživelih, ki bi ga pričakovali v splošni populaciji. Kot taka kumulativna funkcija relativnega preživetja podaja koristno opisno mero in je trenutno najbolj znana metoda za ocenjevanje relativnega preživetja. Med uporabniki manj znana metoda pa je individualna mera relativnega preživetja (Stare in drugi 2005a). Z njo, v nasprotju z $r(t)$, relativno preživetje definiramo na individualni ravni. Izračun individualne mere relativnega preživetja temelji na preprosti transformaciji opazovanega časa do končnega dogodka (smrti), zato so jo avtorji poimenovali tudi transformirani čas, sam pristop k relativnemu preživetju pa transformacijski pristop. Transformacijo opazovanega časa posameznih enot izvedemo s pomočjo tablic smrtnosti, ki so za dane vrednosti spola, starosti in obdobja (leta) začetka opazovanja (npr. čas postavitve diagnoze) podane v obliki pogojnih verjetnosti smrti. Na podlagi teh verjetnosti lahko po principu računanja pričakovanega preživetja za posamezno enoto (2.5) izračunamo pogojno porazdelitveno funkcijo časa do smrti za splošno populacijo (označimo jo z F_{ASY} , kjer je A starost (age), S spol (sex) in Y leto začetka opazovanja (cohort year)). Čas preživetja opazovanih enot lahko izrazimo s pomočjo F_{ASY} preprosto tako, da pogledamo, kakšna je vrednost F_{ASY} pri opazovanem času t . Na ta

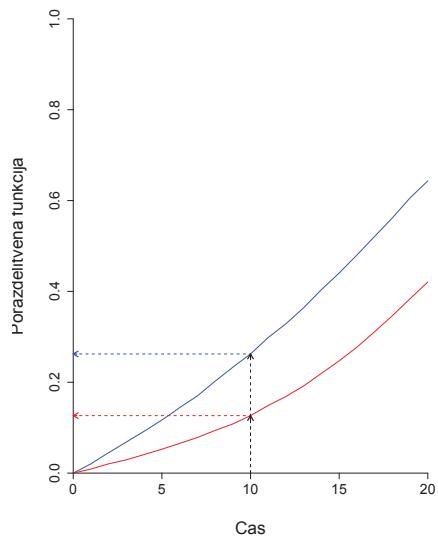
način izvedemo transformacijo opazovanega časa preživetja: $y = F_{ASY}(t)$. Kot novo slučajno spremenljivko vzamemo torej

$$Y = F_{ASY}(T), \quad (2.6)$$

ki meri, kako dolgo, relativno na primerljivo splošno populacijo, je posameznik živel (ob upoštevanju njegovega spola in starosti ob začetku opazovanja).

Za lažjo predstavo in interpretacijo individualne mere relativnega preživetja oziroma transformiranega časa si poglejmo preprost primer. Vzemimo, da imamo moškega in žensko, ki sta bila leta 1990 stara 58 let. Na sliki 2.2 je za oba spola in starost 58 let v letu 1990 prikazana pogojna porazdelitvena funkcija časov do smrti ($F_{A=58,S=\text{moški},Y=1990}$ in $F_{A=58,S=\text{ženski},Y=1990}$). Vidimo lahko, da pogojna porazdelitvena funkcija za ženske leži nižje kot za moške. To odraža dejstvo, da ženske v istem času preživijo manjši delež svojih sovrstnic v splošni populaciji kot moški. Na tem mestu naj opozorimo, da se pogojna porazdelitvena funkcija računa za vsako kombinacijo starosti, spola in leta začetka opazovanja posebej. Sedaj predpostavimo, da proučevano žensko in moškega opazujemo deset let. Kljub enaki vrednosti opazovanega časa pa je zaradi različnih pogojnih porazdelitvenih funkcij individualna mera relativnega preživetja oziroma transformirani čas za obe proučevani enoti različen. Po desetih letih je transformirani čas za proučevanega moškega tako enak $y_{\text{moški}} = 0.26$, za proučevano žensko pa $y_{\text{ženska}} = 0.13$. To pomeni, da je proučevani moški po desetih letih opazovanja preživel 26% splošne populacije moških, ki je bila leta 1990 stara 58 let, proučevana ženska pa je v istem obdobju preživila 13% splošne populacije žensk, ki je bila leta 1990 iste starosti.

Slika 2.2: Porazdelitvena funkcija časa do smrti za moške (modra barva) in ženske (rdeča barva), ki so bili leta 1990 stari 58 let. Puščice označujejo izračun individualne mere relativnega preživetja po desetih letih opazovanja.



2.3.1 Lastnosti individualne mere relativnega preživetja

Individualna mera relativnega preživetja $Y = F_{ASY}(T)$ je nenegativna numerična spremenljivka, ki ohranja mehanizem krnjenja opazovanega časa. Pomembnejše lastnosti mere so zlasti naslednje:

- Pri enakih vrednostih populacijskih spremenljivk (spola, starosti in leta začetka opazovanja) Y narašča monotno z naraščanjem opazovanega časa T . Z drugimi besedami, če primerjamo dve enoti, ki sta bili v izbranem letu istega spola in starosti, njun opazovani čas pa je različen (npr. $t_1 < t_2$), je njuna pogojna porazdelitvena funkcija časa do smrti enaka ($F_{ASY..1}(t) = F_{ASY..2}(t)$), vrednost funkcije pa je višja pri enoti z daljšim opazovanim časom. Enota z daljšim časom opazovanja preživi večji del primerljive splošne populacije (populacije, ki je bila v izbranem letu enake starosti in spola).
- Pri različnih vrednostih populacijskih spremenljivk in enakem času opazovanja (t) je individualna mera relativnega preživetja višja za tisto kombinacijo populacijskih spremenljivk, pri kateri je pričakovana verjetnost smrti do časa t višja. Dve enoti, ki ju opazujemo enako časa ($T_1 = T_2 = t$) in imata različni kombinaciji populacijskih spremenljivk, imata različni pogojni porazdelitveni funkciji ($F_{ASY..1}(t) \neq F_{ASY..2}(t)$). Individualna mera relativnega preživetja je višja za tisto enoto, pri kateri pogojna porazdelitvena funkcija v času t zavzame višjo vrednost. To pomeni, da je v enakem času opazovanja enota z višjo vrednostjo Y -a preživila večji delež primerljive splošne populacije kot enota z nižjo vrednostjo Y -a.
- Če je opazovani čas porazdeljen po F_{ASY} , je individualna mera relativnega preživetja porazdeljena enakomerno na intervalu $[0, 1]$:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(F_{ASY}(T) \leq y) \end{aligned}$$

$$\begin{aligned}
&= P(T \leq F_{ASY}^{-1}(y)) \\
&= F_{ASY}(F_{ASY}^{-1}(y)) \\
&= y. \tag{2.7}
\end{aligned}$$

Funkcijo ogroženosti torej zapišemo kot:

$$\begin{aligned}
\lambda_Y(y) &= -\frac{\partial}{\partial y} \log S_Y(y) \\
&= -\frac{\partial}{\partial y} \log(1-y) \\
&= \frac{1}{1-y}. \tag{2.8}
\end{aligned}$$

S transformacijo opazovanega časa na enotah, ki so izpostavljene samo populacijskim tveganjem, torej dosežemo, da imajo vse enote isto porazdelitev (enakomerno na intervalu $[0, 1]$). To pomeni, da so opazovani pacienti z vidika tveganj splošne populacije med seboj izenačeni, razlike, ki ostanejo, pa lahko pripisemo bolezni proučevanih pacientov.

3. Obstojče regresijske metode v relativnem preživetju

Regresijske metode, ki temeljijo na kumulativni funkciji relativnega preživetja (2.3), so usmerjene v modeliranje funkcije ogroženosti in temeljijo na predpostavkah o povezanosti med opazovano ($\lambda_O(t)$) in pričakovano ali populacijsko ogroženostjo ($\lambda_P(t)$). Pričakovana ogroženost, ki jo izračunamo neposredno iz tablic smrtnosti, predstavlja ogroženost, katere je izpostavljena vsaka oseba zaradi njene starosti, spola in kohortnega leta. Opazovano ogroženost ocenimo na podlagi zbranih podatkov o časih preživetja proučevanih enot. V grobem v literaturi zasledimo aditivne in multiplikativne modele, ki se med seboj razlikujejo po predpostavljeni povezanosti pričakovane in opazovane ogroženosti.

3.1 Aditivni model

Aditivni model (Hakulinen in Tenkanen 1987; Estève in drugi 1990; Dickman in drugi 2004) je v praksi najpogosteje uporabljen regresijska metoda za modeliranje relativnega preživetja. Glavna predpostavka modela je v razčlenitvi opazovane ogroženosti ($\lambda_O(t)$) v dve aditivni komponenti:

$$\lambda_O(t) = \lambda_P(t) + \lambda_E(t), \quad (3.1)$$

kjer je $\lambda_E(t)$ tako imenovana presežna ogroženost in predstavlja tveganje za smrt zaradi bolezni proučevanih pacientov. Če upoštevamo, da funkcijo preživetja lahko zapišemo v obliki

$$S(t) = e^{- \int \lambda(t) dt}, \quad (3.2)$$

ugotovimo, da je opazovana funkcija preživetja enaka

$$\begin{aligned} S_O(t) &= e^{- \int \lambda_O(t) dt} \\ &= e^{- \int (\lambda_P(t) dt + \lambda_E(t) dt)} \\ &= e^{- \int \lambda_P(t) dt} e^{- \int \lambda_E(t) dt} \\ &= S_P(t) \times S_E(t) \end{aligned} \quad (3.3)$$

ozziroma

$$S_E(t) = \frac{S_O(t)}{S_P(t)} = r(t), \quad (3.4)$$

kar pa je ravno kumulativna funkcija relativnega preživetja (2.3). Čeprav je zgornji rezultat enostaven za razumevanje in interpretacijo modela, pa je treba opozoriti, da velja samo v primeru, ko je opazovana ogroženost večja od populacijske, saj ogroženost lahko zavzame samo nenegativne vrednosti. V primeru $S_O(t) > S_P(t)$ ima kumulativna funkcija relativnega preživetja ($r(t)$) lastnosti funkcije preživetja, kar pomeni, da se začne pri 1, njene vrednosti pa potem monotono padajo proti nič. Prileganje modela bo dobro samo v primeru, ko je umrljivost med opazovanimi pacienti višja kot v splošni populaciji. V primerih kjer obstaja možnost ozdravitve pacientov, kumulativna funkcija relativnega preživetja pa na določenem časovnem intervalu zavzame vrednosti, večje od 1, potem presežna ogroženost v modelu mora zavzeti vrednost nič, prileganje modela pa je zaradi tega slabo.

V aditivnem modelu modeliramo funkcijo presežne ogroženosti v obliki, ki je podobna Coxovemu modelu sorazmernih ogroženosti:

$$\lambda_E(t) = \lambda_0(t) e^{\beta^T \mathbf{x}}. \quad (3.5)$$

Čeprav model spominja na Coxov model, pa se od njega razlikuje po tem, da najpogosteje uporabljeni metodi ocenjevanja modela upoštevajo tudi osnovno ogroženost (ang. *baseline hazard*) $\lambda_0(t)$. To pomeni, da moramo za njegovo ocenjevanje predpostaviti neko funkcionalno obliko.

3.1.1 Kratek opis metod ocenjevanja aditivnega modela

V literaturi zasledimo več metod za ocenjevanje koeficientov modela. Najpogosteje se uporablja prevedba na posplošene linearne modele (Hakulinen in Tenkanen 1987; Dickman in drugi 2004) in metoda največjega verjetja (Estève in drugi 1990), pri katerih je presežna ogroženost v celoti podana v parametrični obliki (tudi osnovna ogroženost). Potreba po določitvi funkcionalne oblike osnovne ogroženosti, ki nas običajno niti ne zanima, je problematična zaradi tega, ker izbira funkcionalne oblike vpliva na oceno modela. Do pred kratkim je bila edina metoda, ki pri ocenjevanju ne upošteva osnovne ogroženosti, Sasienijeva metoda (Sasieni 1996), ki pa je teoretično zapletena in povezana z obsežnimi izračunavanji. Zato se je v praksi skoraj ne uporablja. Druga metoda, ki pri ocenjevanju ne upošteva osnovne ogroženosti, temelji na uporabi EM algoritma, pri čemer vzroke smrti obravnava kot manjkajoče podatke (Pohar Perme in drugi 2009). EM algoritem iterira med ocenjevanjem Coxovega modela pod predpostavko istega vzroka smrti za umrle in ocenjevanjem verjetnosti smrti zaradi proučevane bolezni. Metoda, ki je implementirana tudi v programske okolje R, je v primerjavi s Sasienijevim metodo precej enostavnejša, po pričakovanjih pa verjetneje, da zaživi tudi v praksi.

Metode, ki temeljijo na posplošenih linearnih modelih

Metodo ocenjevanja aditivnega modela, ki temelji na teoriji posplošenih linearnih modelov, sta prva predlagala Hakulinen in Tekanen (Hakulinen in Tenkanen 1987; Dickman in drugi 2004). Za namene ocenjevanja enote najprej razvrstimo v L stratumov (indeks stratumov označimo z l), kjer so stratumi določeni na podlagi vseh možnih kombinacij neodvisnih spremenljivk vključenih v model. Čas preživetja razdelimo na manjše intervale (indeks časovnih intervalov označimo z r) in predpostavimo, da je funkcija

$\lambda_E(t)$ znotraj posameznega časovnega intervala konstantna. Nato v časovnih intervalih posameznega stratuma ocenimo opazovane in pričakovane deleže preživetja in model zapišemo v obliki:

$$\log(-\log(\frac{p_{lr}}{p_{lr}^*})) = \beta^T \mathbf{X} + \gamma_r, \quad (3.6)$$

kjer je p_{lr} opazovani, p_{lr}^* pa pričakovani delež preživetja v stratumu l na časovnem intervalu r .

Zgornji model z log-log povezovalno funkcijo (ang. *link function*) ima binomsko strukturo napake. Dickman in drugi (Dickman in drugi 2004) so menili, da je predpostavka o binomski strukturi napake malo verjetna, in predpostavili, da ima število smrti d_{lr} Poissonovo porazdelitev s parametrom $\mu_{lr} = \lambda_{O,lr} \times y_{lr}$, kjer je y_{lr} število enot znotraj stratuma l , ki so v opazovanih časih znotraj intervala r še v tveganju. Če pričakovano število smrti označimo z $d_{lr}^* = \lambda_{P,lr} \times y_{lr}$, enačbo (3.1) lahko zapišemo v obliki:

$$\begin{aligned} \frac{\mu_{lr}}{y_{lr}} &= \frac{d_{lr}^*}{y_{lr}} + e^{\beta^T \mathbf{X} + \gamma_r} \\ \log(\mu_{lr} - d_{lr}^*) &= \beta^T \mathbf{X} + \gamma_r + \log(y_{lr}). \end{aligned} \quad (3.7)$$

Zgornji model s povezovalno funkcijo $\log(\mu_{lr} - d_{lr}^*)$ predstavlja posplošen linearni model s Poissonovo strukturo napake, izidom d_{lr} in „offset“-om $\log(y_{lr})$.

Metode, ki temeljijo na funkciji največjega verjetja

Estève (Estève in drugi 1990) aditivni model ocenjuje z metodo največjega verjetja, pri čemer funkcijo največjega verjetja definira kot:

$$\begin{aligned} L(\beta) &= \prod_{i=1}^n (\lambda_O(t_i))^{\delta_i} e^{-\int_0^{t_i} \lambda_O(s) ds} \\ &= \prod_{i=1}^n (\lambda_{P_i} + \lambda_0(t_i) e^{\beta^T \mathbf{X}_i})^{\delta_i} e^{-\int_0^{t_i} (\lambda_{P_i}(s) + \lambda_0(s) e^{\beta^T \mathbf{X}_i}) ds}. \end{aligned} \quad (3.8)$$

Logaritem funkcije največjega verjetja je potem enak:

$$l(\beta) = \sum_{i=1}^n \delta_i \log(\lambda_{P_i} + \lambda_0(t_i)e^{\beta^T \mathbf{X}_i}) - \sum_{i=1}^n \int_0^{t_i} \lambda_0(s)e^{\beta^T \mathbf{X}_i} ds - \sum_{i=1}^n \int_0^{t_i} \lambda_{P_i}(s)ds. \quad (3.9)$$

Zadnji člen v zgornji enačbi ni odvisen od parametrov modela, zato ga pri ocenjevanju lahko izpustimo. Z odvajanjem logaritma funkcije največjega verjetja po parametrib modela dobimo:

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^n \delta_i \frac{X_{i,j}\lambda_0(t_i)e^{\beta^T \mathbf{X}_i}}{\lambda_{P_i}(t_i) + \lambda_0(t_i)e^{\beta^T \mathbf{X}_i}} - \sum_{i=1}^n \int_0^{t_i} X_{i,j}\lambda_0(u)e^{\beta^T \mathbf{X}_i} du. \quad (3.10)$$

Preostali del funkcije, na podlagi katere ocenimo parametre modela, dobimo z odvajanjem po parametrib, s katerimi je definirana funkcija osnovne ogroženosti λ_0 . V izvirnem Estèvovem članku je funkcija osnovne ogroženosti definirana kosoma konstantno, Giorgi in drugi (Giorgi in drugi 2003a) pa so model razširili tako, da je mogoča tudi uporaba zlepkov (ang. *splines*).

Metode, ki pri ocenjevanju ne upoštevajo osnovne ogroženosti

Prva znana metoda, ki pri ocenjevanju parametrov aditivnega modela ne upošteva osnovne ogroženosti oziroma njene funkcijске oblike, ki lahko vpliva na oceno modela, je Sasienijeva metoda (Sasieni 1996). Ideja metode je, da z iteracijskim postopkom dobimo ocene uteži, s katerimi ocenjujemo verjetnost smrti zaradi proučevane bolezni (v nasprotju z verjetnostjo smrti zaradi populacijskih tveganj). Te ocene pa uporabljamo tudi za ocenjevanje parametrov modela. Zaradi uporabe števnih procesov (ang. *counting process*) je logika ocenjevanja parametrov modela podobna Nelson-Aalenovi metodi. Razlika nastopi, ker je pričakovana oziroma populacijska kumulativna ogroženost, v nasprotju z opazovano ogroženostjo, zvezna funkcija, kar pomeni, da je integral različen od nič za vse paciente v množici tveganj, in ne le za enoto, pri kateri nastopi smrt. Integracija je sicer mogoča, če predpostavimo, da je populacijska ogroženost kosoma konstantna, vendar je izračunavanje še vedno obsežno. Metoda se v praksi, verjetno prav zaradi obsežnih izračunavanj in teoretične zapletenosti, redko uporablja.

Pohar Perme in drugi (Pohar Perme in drugi 2009) so predstavili novo metodo, ki pri ocenjevanju parametrov aditivnega modela ne predpostavlja funkcijске oblike os-

novne ogroženosti. Metoda, ki je enostavna za razumevanje in uporabo, vzroke smrti obravnava kot manjkajoče podatke, parametre modela pa ocenjuje z EM algoritmom (ang. *Estimation-Maximization algorithm*).

Metoda indikatorsko spremenljivko δ_i , s katero označujemo, ali je smrt pri enoti nastopila ($\delta_i = 1$) ali ne ($\delta_i = 0$), razdeli v dva dela:

$$\delta_i = \delta_{P_i} + \delta_{E_i}. \quad (3.11)$$

V enačbi (3.11) δ_{E_i} označuje smrt zaradi proučevane bolezni, δ_{P_i} pa smrt zaradi drugih razlogov. Če bi bili vzroki smrti znani za vse paciente, bi za ocenjevanje parametrov modela uporabili običajen Coxov model z δ_{E_i} kot indikatorsko spremeljivko:

$$\lambda_E(t) = \lambda_0(t)e^{\beta Z}, \quad (3.12)$$

parametre modela pa bi ocenili na podlagi običajne delne funkcije največjega verjetja (ang. *partial maximum likelihood*), ki ne upošteva osnovne ogroženosti:

$$L(\beta|X) = \prod_{i=1}^n \left(\frac{e^{\beta Z_i}}{\sum_{j \in R_i} e^{\beta Z_j}} \right)^{\delta_{E_i}} \quad (3.13)$$

V zgornji enačbi je n število proučevanih enot, $Z = \min(T, C)$, pri čemer je T čas preživetja, C pa knjeni čas in R_i množica enot v tveganju pri dogodku (smrti) i -tega pacienta.

Izkaže se, da je verjetnost smrti zaradi proučevane bolezni za i -to enoto, pri danih vrednostih opazovanega časa t_i in indikatorsko spremenljivko smrti δ_i , enaka:

$$\begin{aligned} P(\delta_{E_i} = 1 | t_i, \delta_i) &= E(\delta_{E_i} | t_i, \delta_i) \\ &= \delta_i \frac{\lambda_{E_i}(t_i)}{\lambda_{P_i}(t_i) + \lambda_{E_i}(t_i)} \\ &= \delta_i \frac{\lambda_0(t_i)e^{\beta Z_i}}{\lambda_{P_i}(t_i) + \lambda_0(t_i)e^{\beta Z_i}} \end{aligned} \quad (3.14)$$

EM algoritem je tako osnovan na iteracijskem postopku maksimizacije delne funkcije največjega verjetja (3.13) in obnavljanju vrednosti indikatorske spremenljivke δ_{E_i} . Prednost modela je, da z ocenami parametrov β popolnoma neparametrično ocenimo tudi

osnovno ogroženost ($\lambda_0(t)$). Metoda predstavlja posplošitev Coxovega modela, zato jo lahko enostavno razširimo (na primer z vključitvijo zlepkov ali krhkosti v model), kar je tudi bistvena prednost pred Sasienijevo metodo. Prav zaradi svoje izjemne prilagodljivosti pa ima metoda, v nasprotju s Sasienijevo, večjo možnost, da zaživi tudi v praksi.

3.2 Multiplikativni model

Multiplikativni model predpostavlja, da sta opazovana in populacijska ogroženost povezani multiplikativno:

$$\lambda_O(t) = \lambda_P(t)\nu(t), \quad (3.15)$$

kjer $\nu(t)$ lahko interpretiramo kot relativno umrljivost. Najbolj znan multiplikativni model je Andersenov model (Andersen in drugi 1985), ki predpostavlja, da relativno umrljivost lahko zapišemo kot

$$\nu(t) = \nu_0(t)e^{\beta^T \mathbf{X}}. \quad (3.16)$$

To omogoča, da osnovni model prevedemo v Coxov model sorazmernih ogroženosti z dodatno časovno odvisno spremenljivko, ki ima koeficient fiksiran na 1:

$$\lambda_O(t) = \nu_0(t)e^{\beta^T \mathbf{X} + 1 \times \log(\lambda_P(t))}. \quad (3.17)$$

Prevedba na Coxov model omogoča uporabo že razvite teorije ocenjevanja parametrov modela.

Multiplikativni model je v primerjavi z aditivnim modelom precej bolj prilagodljiv, saj ne zahteva, da je opazovana ogroženost večja od populacijske. Kljub temu je v praksi manj uporabljen.

4. Alternativne regresijske metode v relativnem preživetju

Regresijske metode, ki smo jih opisali v prejšnjem poglavju, temeljijo na predpostavkah o povezanosti opazovane in populacijske ogroženosti. Če predpostavke niso izpolnjene, so rezultati metod in posledično tudi njihova interpretacija vprašljivi.

Alternativni oziroma tako imenovani transformacijski pristop (Stare in drugi 2005a) omogoča nove možnosti na področju modeliranja relativnega preživetja. S transformacijo opazovanega časa lahko preživetja relativno preživetje definiramo za vsako posamezno enoto. Tako dobimo novo, nenegativno razmernostno slučajno spremenljivko Y , ki so jo avtorji poimenovali transformacijski čas oziroma individualna mera relativnega preživetja (Stare in drugi 2005a), in je krnjena, če je krnjen tudi opazovani čas. S transformacijskim časom definirano relativno preživetje omogoča, da za namene modeliranja lahko uporabimo katerokoli regresijsko metodo, ki se tudi sicer uporablja v klasični analizi preživetja.

V primerjavi s tradicionalnimi regresijskimi metodami v analizi relativnega preživetja (aditivni in multiplikativni modeli) je prednost modeliranja transformiranega časa Y v tem, da ne zahteva nikakršnih predpostavk o povezanosti med opazovano in populacijsko ogroženostjo. Ko opazovani čas transformiramo z ustrezeno (pogojno) porazdelitveno funkcijo časa do smrti, namreč avtomatično zajamemo vsa populacijska tveganja. Ostane nam torej le presežna ogroženost, ki jo tudi neposredno modeliramo. V praksi se v okviru klasične analize preživetja najpogosteje uporablja Coxov model, ki je sicer zelo prilagodljiv, vendar kljub temu ne omogoča vsega, kar bi si želeli. Poleg tega se za Coxov model s transformiranimi časi izkaže, da so rezultati modela zelo podobni rezultatom multiplikativnega regresijskega modela v relativnem preživetju,

torej je podobna tudi interpretacija obeh modelov (Pohar Perme 2007).

V tem poglavju se bomo posvetili predvsem dvema drugima metodama, ki se v klasični analizi preživetja sicer redkeje uporabljata. Prva je Buckley-Jamesova linearna regresija za modeliranje časov preživetja, druga pa je Aalenov linearni regresijski model za modeliranje funkcije ogroženosti. Uporaba obeh metod na transformiranih časih ali individualnih merah relativnega preživetja odpira nove možnosti v analizi relativnega preživetja. Pričakujemo lahko, da bomo z uporabo metod dobili nove informacije in drugačen vpogled v relativno preživetje. To bi pomenilo tudi obogatitev področja relativnega preživetja nasploh.

4.1 Buckley-Jamesova regresija najmanjših kvadratov za krnjene podatke

Čas preživetja je nenegativna numerična spremenljivka in v primeru izbire regresijske metode za nekrnjene podatke bi se verjetno brez večjih pomislekov najprej odločili za navadno linearno regresijo. Zaradi možnosti, da v modeliranje vključimo tudi ne-linearne člene in s tem opustimo predpostavko o strogi linearnosti neodvisnih spremenljivk, je linearna regresija postala še toliko bolj privlačna. Kljub temu pa lahko opazimo, da je linearna regresija v primeru krnjenih vrednosti opazovanega časa le redko uporabljena regresijska metoda. V praksi se za modeliranje krnjenih podatkov daleč najpogosteje uporablja Coxov model, čeprav predpostavke sorazmernih ogroženosti, na katerih model temelji, pogosto niso izpolnjene. V takih primerih so ocenjeni parametri in njihova interpretacija seveda vprašljivi. Heller in Simonoff (Heller in Simonoff 1992) na primer predlagata, da bi bilo v primeru nesorazmernih ogroženosti namesto Coxovega modela bolje uporabiti linearno regresijo za modeliranje krnjenih podatkov.

Čeprav v literaturi lahko zasledimo več metod za ocenjevanje koeficientov linearne regresijskega modela s krnjenimi podatki (na primer (Miller 1976; Koul in drugi 1981)), se bomo v delu osredotočili na Buckley-Jamesovo metodo (Buckley in James 1979), ki temelji na metodi najmanjših kvadratov in je prilagojena krnjenim podatkom. Metodologijo so po prvi objavi preverjali številni avtorji. Miller in Halpern (Miller in Halpern 1982),

Weissfeld in Schneider (Weissfeld in Schneider 1987), Heller in Simonoff (Heller in Simonoff 1990) ter Hillis (Hillis 1993) so različne metode za ocenjevanje parametrov v linearni regresiji primerjali s pomočjo simulacij. Stare in drugi (Stare in drugi 2000) so proučevali občutljivost ocen Buckley-Jamesovega modela na kršitev predpostavk o linearnosti in homoskedastičnosti modela. James in Smith (James in Smith 1984) sta dokazala šibko konsistentnost regresijskih cenilk, Lai in Ying (Lai in Ying 1991) sta predlagala ocenjevanje varianc parametrov na podlagi splošnejšega modela, Smith in Zhang (Smith in Zhang 1995) ter Currie (Currie 1996) pa so nadgradili razsevni diagram, da bi bil primeren tudi za krnjene podatke. Jin in drugi (Jin in drugi 2006) so razvili nov način ocenjevanja parametrov, ki temelji na metodologiji najmanjših kvadratov, s katerim dobimo skupino konsistentnih cenilk, ki so asimptotično normalno porazdeljene. Kovariančno matriko cenilk parametrov dobimo s pomočjo ponovnega vzorčenja (ang. *resampling approach*).

Izkaže se, da je Buckley-Jamesova metoda pod običajnimi pogoji regularnosti konsistentna, v primerjavi z drugimi linearimi regresijskimi metodami za krnjene podatke pa Buckley-Jamesova metoda da najboljše rezultate (Miller in Halpern 1982; Heller in Simonoff 1992). Opis metode je večinoma povzet po (Smith in Zhang 1995) in (Smith 2002).

4.1.1 Opis metode

V enostavnem modelu linearne regresije opazujemo pare vrednosti (x_i, t_i) , kjer gre $i = 1, \dots, n$ in predpostavljam, da je odvisna spremenljivka časov preživetja T ali njena monotona transformacija v linearni povezavi z neodvisno spremenljivko X :

$$T_i = \beta_0 + \beta_1 x_i + \epsilon_i; \quad \epsilon_i \sim F; \quad E(\epsilon_i) = 0; \quad Var(\epsilon_i) = \sigma^2. \quad (4.1)$$

V enačbi (4.1) F ni določena, kar pomeni, da so napake modela porazdeljene po neznani funkciji preživetja $S = 1 - F$. Poleg tega so napake modela neodvisne od X .

V primeru krnjjenja ne opazujemo le časov preživetja oziroma časov do smrti, temveč za krnjene enote dobimo samo informacijo o tem, do kdaj so še živele, ne pa tudi, kdaj so umrle. Opazovani čas v primeru krnjjenja torej zapišemo:

$$z_i = t_i \delta_i + c_i(1 - \delta_i), \quad (4.2)$$

kjer je $\delta_i = I(T_i \leq C_i)$ indikatorska spremenljivka, ki označuje, ali je opazovani čas čas preživetja (t_i) enote ali krnjeni čas (c_i) enote. Očitno bi z modelom (4.1), v katerem bi za odvisno spremenljivko namesto dejanskega časa preživetja T uporabili opazovani čas Z , podcenili ocenjene pričakovane čase preživetja \hat{T} , saj so krnjeni časi krajši od pripadajočih dejanskih časov preživetja. Povezava, določena v (4.1), torej za Z ne velja.

Opazimo pa lahko, da se v primeru, če krnjene vrednosti prerazporedimo na njihove prave pričakovane vrednosti $E(T_i|T_i > Z_i)$, linearnost modela (4.1) ne spremeni. Tako lahko definiramo novo odvisno spremenljivko

$$V_i = T_i \delta_i + E(T_i|T_i > Z_i)(1 - \delta_i), \quad (4.3)$$

za katero je mogoče pokazati, da je njen povprečje enako povprečju dejanskih časov preživetja:

$$\begin{aligned} E(V_i) &= E(T_i)P(\delta_i = 1) + E[E(T_i|T_i > Z_i)]P(\delta_i = 0) = \\ &= E(T_i)P(\delta_i = 1) + E(T_i)P(\delta_i = 0) = \\ &= E(T_i)[P(\delta_i = 1) + P(\delta_i = 0)] = \\ &= E(T_i). \end{aligned}$$

Ideja Buckleyja in Jamesa je, da krnjene vrednosti opazovanih časov z_i nadomestimo z ocenjenimi pričakovanimi časi preživetja, ki so pod določenimi pogoji nepristranski. Za oceno parametra β_1 v enostavni linearni regresiji moramo torej najprej oceniti pričakovane čase preživetja za krnjene podatke $E(T_i|T_i > Z_i)$:

$$\begin{aligned} E(T_i|T_i > Z_i) &= E(\beta_0 + \beta_1 x_i + \epsilon_i | \beta_0 + \beta_1 x_i + \epsilon_i > Z_i) = \\ &= \beta_0 + \beta_1 x_i + E(\epsilon_i | \epsilon_i > Z_i - (\beta_0 + \beta_1 x_i)). \end{aligned} \quad (4.4)$$

To pomeni, da moramo dejansko oceniti pričakovane napake modela za krnjene enote $E(\epsilon_i | \epsilon_i > Z_i - (\beta_0 + \beta_1 x_i))$:

$$E(\epsilon_i | \epsilon_i > Z_i - (\beta_0 + \beta_1 x_i)) = \int_{Z_i - (\beta_0 + \beta_1 x_i)}^{\infty} \epsilon \frac{dF}{1 - F(Z_i - (\beta_0 + \beta_1 x_i))}, \quad (4.5)$$

kjer je F porazdelitvena funkcija napak modela ϵ .

Zapišimo sedaj napako modela (4.1) kot $\epsilon_i^1 = \epsilon_i + \beta_0$. Odvisno spremenljivko potem lahko izrazimo v obliki

$$T_i = \beta_1 x_i + \epsilon_i^1; \quad \epsilon_i^1 \sim F; \quad E(\epsilon_i^1) = \beta_0; \quad Var(\epsilon_i^1) = \sigma^2, \quad (4.6)$$

s katero najprej ocenimo naklon β_1 in oblikujemo novo odvisno spremenljivko, ki je odvisna od naklona b :

$$T_i^*(b) = bx_i + [\epsilon_i(b)\delta_i + \hat{E}_b(\epsilon_i(b)|\epsilon_i(b) > d_i(b))(1 - \delta_i)]. \quad (4.7)$$

V oglatih oklepajih enačbe (4.7) so običajni ostanki za nekrnjene enote $\epsilon_i(b) = T_i - bx_i$ in ocjenjeni ostanki za krnjene enote $\hat{E}_b(\epsilon_i(b)|\epsilon_i(b) > d_i(b))$, pri čemer je $d_i(b) = c_i - bx_i$. Ostanke za krnjene enote izračunamo kot uteženo vsoto nekrnjениh ostankov, večjih od $d_i(b)$:

$$\hat{E}_b(\epsilon_i(b)|\epsilon_i(b) > d_i(b)) = \sum_{\epsilon_k > d_i(b)} w_{ik} \epsilon_k(b). \quad (4.8)$$

V (4.8) so uteži w_{ik} določene na podlagi ocenjene porazdelitvene funkcije \hat{F} , ki jo dobimo s Kaplan-Meierjevo oceno funkcije preživetja opazovanih ostankov modela $e_i(b) = Z_i - bx_i$, torej $\hat{F} = 1 - \hat{S}$. Uteži potem izračunamo kot je prikazano v enačbi (4.9)

$$w_{ik}(b) = \begin{cases} \frac{d\hat{F}(e_k(b))\delta_k(1-\delta_i)}{\hat{S}(e_i(b))} & k > i \\ 0 & \text{sicer,} \end{cases} \quad (4.9)$$

kjer smo privzeli, da so ostanki urejeni ($e_i = e_{(i)}$). Zaradi urejenosti ostankov velja, da je $k > i$ samo v primeru $e_{(k)} > e_{(i)}$.

Postopek ocenjevanja parametrov v Buckley-Jamesovi regresiji je od tu naprej podoben ocenjevanju v navadni linearnej regresiji. Če namesto dejanskih časov preživetja, ki jih zaradi krnjjenja ne moremo v celoti opazovati, upoštevamo nove čase T_i^* , modificirano različico normalne enačbe najmanjših kvadratov lahko zapišemo kot:

$$\sum_{i=1}^n \frac{(x_i - \bar{x})T_i^*(b)}{(x_i - \bar{x})^2} - b = 0. \quad (4.10)$$

Ker pravih vrednosti spremenljivke T_i^* ne poznamo, njene vrednosti pa so odvisne od ocjenjenega naklona b , Buckley-Jamesovo cenilko regresijskega koeficenta $\hat{\beta}_n$ za parameter β dobimo z iterativno rešitvijo enačbe (4.10) na b .

Rešitev enačbe (4.10) lahko nastopi v obliki funkcije $\Phi_n(b)$

$$\sum_{i=1}^n \frac{(x_i - \bar{x})T_i^*(b)}{(x_i - \bar{x})^2} - b = \Phi_n(b), \quad (4.11)$$

Buckley-Jamesovo oceno parametra β_1 v enostavni linearnej regresiji pa nato določimo v točki, kjer funkcija $\Phi_n(b)$ spremeni predznak.

$$\Phi_n(\hat{\beta}_n^+) \Phi_n(\hat{\beta}_n^-) \leq 0. \quad (4.12)$$

Buckley in James sta ugotovila, da je funkcija $\Phi_n(b)$ kosoma linearnej. Nezvezni skoki funkcije nastopajo takrat, ko se spremeni b . Za dano velikost vzorca n funkcija $\Phi_n(b)$ ni nujno monotono padajoča, kar pomeni, da ima enačba (4.10) lahko več kot eno rešitev (Buckley in James 1979), ki je konsistentna (James in Smith 1984).

Cenilko regresijske konstante $\hat{\beta}_{0n}$ za β_0 poiščemo na podoben način kot v običajni metodi najmanjših kvadratov:

$$\hat{\beta}_{0n} = \bar{T}^*(b) - \hat{\beta}_n \bar{x}. \quad (4.13)$$

V primeru multiple linearnej regresije lahko linearnej model zapišemo v matrični obliki:

$$\mathbf{T} = \mathbf{X}\beta + \epsilon, \quad (4.14)$$

kjer je \mathbf{X} $n \times (p + 1)$ razsežna podatkovna matrika s prvim stolpcem enic, β je $(p + 1) \times 1$ razsežni vektor parametrov, ϵ pa je $n \times 1$ razsežni vektor napak modela.

Novo odvisno spremenljivko T^* zapišemo v obliki

$$\mathbf{T}^*(\mathbf{b}) = \mathbf{X}\mathbf{b} + \mathbf{W}(\mathbf{b})(\mathbf{Z} - \mathbf{X}\mathbf{b}), \quad (4.15)$$

v kateri je $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)$ vektor opazovanih časov, $\mathbf{W}(\mathbf{b})$ pa matrika uteži, ki jo zapišemo kot:

$$\mathbf{W}(\mathbf{b}) = \text{diag}(\delta) + [w_{ik}(\mathbf{b})] = \begin{pmatrix} \delta_1 & w_{12}(\mathbf{b}) & w_{13}(\mathbf{b}) & \dots & w_{1n}(\mathbf{b}) \\ 0 & \delta_2 & w_{23}(\mathbf{b}) & \dots & w_{12}(\mathbf{b}) \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & w_{n-1\ n}(\mathbf{b}) \\ 0 & 0 & 0 & \dots & \delta_n \end{pmatrix}.$$

Iz enačbe (4.15) je lepo razvidno, da novo oblikovana spremenljivka nekrnjene vrednosti ohranja nespremenjene, krnjene vrednosti pa se povečajo za $\mathbf{W}(\mathbf{b})(\mathbf{Z} - \mathbf{X}\mathbf{b})$.

Buckley-Jamesove ocene parametrov β dobimo, če v multivariatno različico normalne enačbe najmanjših kvadratov vstavimo spremenljivko T^* :

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{T}^*(\mathbf{b}). \quad (4.16)$$

Če v zgornji enačbi $\mathbf{T}^*(\mathbf{b})$ nadomestimo z izrazom (4.15), dobimo

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T [\mathbf{X}\mathbf{b} + \mathbf{W}(\mathbf{b})(\mathbf{Z} - \mathbf{X}\mathbf{b})], \quad (4.17)$$

ozziroma

$$\Phi_n(\mathbf{b}) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(\mathbf{b})(\mathbf{Z} - \mathbf{X}\mathbf{b}) = \mathbf{0}. \quad (4.18)$$

Rešitev enačbe (4.18) poiščemo iterativno, pri čemer za začetne vrednosti $\mathbf{b} = \mathbf{b}^0$ lahko izberemo na primer regresijske koeficiente, ocenjene na nekrnjencih podatkih. Nato oblikujemo spremenljivko $\mathbf{T}^*(\mathbf{b}^0)$ in preidemo na naslednji korak iteracije. Koraka ponavljamo, dokler ne dosežemo željene natančnosti, čeprav se lahko zgodi, da vrednosti

ne skonvergirajo (Buckley in James 1979; James in Smith 1984). V primeru da funkcija $\Phi_n = \mathbf{0}$ predznak spremeni večkrat, ima enačba (4.18) več rešitev, za oceno $\hat{\beta}_n$ pa se običajno vzame povprečje dobljenih rešitev.

Buckley-Jamesova cenilka tudi ni vedno konsistentna. James in Smith sta za iterativno rešitev cenilke $\hat{\beta}_n$ pod določenimi pogoji regularnosti (ang. *regularity conditions*) dokazala šibko konsistentnost (James in Smith 1984). Dokazala sta, da za neko vrednost $\delta > 0$, za $n \rightarrow \infty$ s $P \rightarrow 1$ v $(\beta - \delta, \beta + \delta)$ obstaja točka $b = \hat{\beta}_n$, v kateri funkcija $\Phi_n(b)$ spremeni predznak. Ta lastnost zagotavlja obstoj Buckley-Jamesove cenilke blizu prave vrednosti parametra. Eden pomembnejših potrebnih pogojev za tovrsten rezultat je, da je krnjene po regresijski funkciji enakomerno. Z drugimi besedami, če definiramo pričakovano število krnjениh in nekrnjениh vrednosti ostankov $e_i(b)$, ki so večji od u z $\mathcal{N}_b(u)$, potem velja

$$\mathcal{N}_b(u) \rightarrow \infty, \text{ ko gre } n \rightarrow \infty \text{ za vse } u \in (-\infty, U_b). \quad (4.19)$$

V (4.19) je U_b na skrajni desni polovici porazdelitve ostankov. S tem zagotovimo konsistentnost Kaplan-Meierjeve ocene funkcije preživetja S . Nestabilnost ocen funkcije preživetja S za vrednosti u , kadar je $S(u)$ blizu 0, onemogoča, da z

$$\hat{E}_b(\epsilon_i(b) | \epsilon_i(b) > d_i(b)) = \sum_{k=1}^n w_{ik}(b) \epsilon_k(b)$$

ocenimo pravo vrednost

$$E(\epsilon_i(b) | \epsilon_i(b) > d_i(b)).$$

Lai in Ying (Lai in Ying 1991) sta tudi dokazala, da je asimptotična varianca cenilke $\hat{\beta}_n$ odvisna od neznane funkcije ogroženosti ostankov, zaradi česar jo je težko oceniti.

Jin in drugi (Jin in drugi 2006) so razvili nov postopek ocenjevanja parametrov v linearne regresiji, ki ravno tako temelji na metodi najmanjših kvadratov, z njim pa naj bi obšli nekatere probleme Buckley-Jamesove cenilke parametrov, še zlasti problem kon-

vergence predlaganega algoritma za iskanje rešitev enačbe (4.18), problem konsistentnosti in problem ocenjevanja kovariančne matrike cenilk. Z novim postopkom dobimo skupino konsistetnih cenilk, ki so asimptotično normalno porazdeljene, kovariančno matriko pa ocenimo s postopkom ponovnega vzorčenja (ang. *resampling*).

Novi postopek je sledeč. Vpeljimo

$$U(\beta, b) = \sum_{i=1}^n (X_i - \bar{X})(T_i^*(b) - X_i^T \beta), \quad (4.20)$$

ozziroma

$$U(\beta, b) = \sum_{i=1}^n (X_i - \bar{X})(T_i^*(b) - \bar{T}^*(b) - (X_i - \bar{X})^T \beta), \quad (4.21)$$

kjer je X_i $(p+1) \times 1$ razsežni vektor neodvisnih spremenljivk, β je $(p+1) \times 1$ razsežni vektor neznanih regresijskih parametrov in $\bar{T}^*(b) = \frac{\sum_{i=1}^n T_i^*}{n}$.

V primeru $U(b, \beta) = 0$ je

$$\beta = L(b) = \left(\sum_{i=1}^n (X_i - \bar{X})^{\otimes 2} \right)^{-1} \left(\sum_{i=1}^n (X_i - \bar{X})(T_i^*(b) - \bar{T}^*(b)) \right). \quad (4.22)$$

Za vektor a velja $a^{\otimes 2} = aa^T$.

(4.22) predstavlja osnovo za iterativni algoritem

$$\hat{\beta}_{(m)} = L(\hat{\beta}_{(m-1)}), \quad m \geq 1. \quad (4.23)$$

Jin in drugi (Jin in drugi 2006) so za začetno oceno $\hat{\beta}_{(0)}$ uporabili Gehanovo ranžirno cenilko (ang. *Gehan-type rank estimator*) $\hat{\beta}_G$, ki jo dobimo z maksimiziranjem konveksne funkcije

$$\sum_{i=1}^n \sum_{j=1}^n \delta_i |e_i(\beta) - e_j(\beta)| + \left| M - \beta^T \sum_{k=1}^n \sum_{l=1}^n \delta_k (X_l - X_k) \right|, \quad (4.24)$$

v kateri je M neko vnaprej določeno zelo veliko število. V tem primeru je cenilka $\hat{\beta}_{(m)}$ za vsak m konsistentna in asimptotično normalno porazdeljena. Na ta način

torej dobimo skupino konsistentnih in asimptotično normalno porazdeljenih cenilk $\hat{\beta}_{(m)} : m \geq 1$. Kot so pokazali Jin in drugi (Jin in drugi 2006), lahko kovariančno matriko cenilke $\beta_{(m)}$ dobimo s postopkom ponovnega vzorčenja (ang. *resampling approach*), s katerim najprej generiramo n identičnih neodvisnih slučajnih spremenljivk Z_i s povprečjem in varianco 1 ($E(Z_i) = Var(Z_i) = 1$). S pomočjo teh spremenljivk nato definiramo $\hat{S}_b^{*Z}(t)$, ki je podobna Kaplan-Meierjevi oceni funkcije preživetja:

$$\hat{S}_b^{*Z}(t) = \prod_{i:e_i(b) < t} \left[1 - \frac{Z_i \delta_i}{\sum_{j=1}^n Z_j 1(e_j(b) \geq e_i(b))} \right]. \quad (4.25)$$

Podobno kot v (4.15) oblikujemo novo odvisno spremenljivko T_i^{*Z} , pri čemer so uteži določene na podlagi $\hat{F}_b^{*Z}(t) = 1 - \hat{S}_b^{*Z}(t)$ (za podrobnosti glej enačbo (4.9)). Za določitev točkaste ocene nadaljujemo z iterativnim procesom

$$L^{*Z}(b) = \left(\sum_{i=1}^n Z_i (X_i - \bar{X})^{\otimes 2} \right)^{-1} \left(\sum_{i=1}^n Z_i (X_i - \bar{X}) (T_i^{*Z}(b) - \bar{T}^{*Z}(b)) \right), \quad (4.26)$$

in $\hat{\beta}_{(m)}^{*Z} = L(\hat{\beta}_{(m-1)}^{*Z})$, $m \geq 1$.

Z generiranjem N slučajnih vzorcev spremenljivke Z_i ($i = 1, \dots, n$) dobimo N realizacij ocen $\hat{\beta}_{(m)}^{*Z}$. Če posamezno realizacijo označimo z $\hat{\beta}_{(m),j}^{*Z}$ ($j = 1, \dots, N$), kovariančno matriko za $\hat{\beta}_{(m)}$ izračunamo kot:

$$s^2 = \frac{1}{N-1} \sum_{j=1}^N (\hat{\beta}_{(m),j}^{*Z} - \bar{\beta}_{(m)}^{*Z})(\hat{\beta}_{(m),j}^{*Z} - \bar{\beta}_{(m)}^{*Z})^T, \quad (4.27)$$

pri čemer je $\bar{\beta}_{(m)}^{*Z} = \frac{\sum_{j=1}^N \hat{\beta}_{(m),j}^{*Z}}{N}$.

4.2 Aalenov aditivni regresijski model

Aalenov aditivni model, poznan tudi pod imenom Aalenov aditivni model ogroženosti ali Aalenova linearna regresija, je neperametrični model, s katerim modeliramo funkcijo ogroženosti. Idejo modela je Aalen prvič objavil že leta 1980 (Aalen 1980), poglobljeno in dopolnjeno razlago pa je objavil kasneje (Aalen 1989, 1993; Aalen in drugi 2004). V tem času so model proučevali številni avtorji. Različne metode ocenjevanja so poleg Aalena (Aalen 1980) proučevali tudi Huffer in McKeague (Huffer in McKeague

1991a) ter Fosen in drugi (Fosen in drugi 2006). Postopke za ocenjevanje učinkov neodvisnih spremenljivk so proučevali Aalen (Aalen 1989), McKeague in Utikal (McKeague in Utikal 1991a), Grønnesby (Grønnesby 1997), Lee in Weissfeld (Lee in Weissfeld 1998), Bhattacharyya in Klein (Bhattacharyya in Klein 2005) ter Fosen in drugi (Fosen in drugi 2006), s prileganjem modela pa so se temeljiteje ukvarjali Aalen (Aalen 1993), Grønnesby in Borga (Grønnesby in Borga 1996) ter Gandy in Jensen (Gandy in Jensen 2005). Asimptotične lastnosti modela so podrobnejše proučevali McKeague (McKeague 1988) ter Huffer in McKeague (Huffer in McKeague 1991a).

Osnovna ideja Aalenovega modela je, da funkcijo ogroženosti modeliramo v vsakem času, v katerem je nastopil dogodek. Model prav zato dopušča modeliranje časovno odvisnih spremenljivk ter možnost spremicanja učinkov neodvisnih spremenljivk v opazovanem času, kar omogoča veliko prilagodljivost. Zlasti v primeru manjših vzorcev pa je prilagodljivost Aalenovega modela, ki dopušča spremicanje učinkov vseh neodvisnih spremenljivk, lahko potratna. V takšnih primerih bi bilo regresijske koeficiente, ki se v času ne spreminja, smiselnopravnavati kot konstantne. Tako sta McKeague in Sasieni razvila tako imenovani semiparametrični aditivni model ogroženosti (McKeague in Sasieni 1994), Lin in Ying (Lin in Ying 1994a) pa sta na njegovi osnovi vzpostavila paralelo s Coxovim modelom sorazmernih ogroženosti.

Namen naloge je proučiti vsebinski doprinos k relativnemu preživetju, ki ga dobimo z Aalenovim modelom. V nalogi se bomo, tudi zaradi velikih vzorcev proučevanih podatkov, posvetili prvotnemu modelu, v katerem se vsi regresijski koeficienti modelirajo v vsaki vrednosti opazovanega časa, v katerem nastopi dogodek.

4.2.1 Opis metode

Izhodišče Aalenove linearne regresije je modeliranje funkcije ogroženosti kot linearne funkcije, v kateri posamezne komponente merijo (linearne) učinke neodvisnih spremenljivk, vključenih v model. Model je bil v celoti razvit na podlagi teorije števnih procesov (ang. *counting process theory*), zato tudi razlago zanj podajamo v tej obliki. Osnove teorije števnih procesov so podane v Prilogi.

V primeru regresijskih modelov moramo števni proces definirati za vsako posamezno

enoto. Za n vzorčnih enot tako proučujemo števne procese N_1, N_2, \dots, N_n , pri čemer $N_i(t)$ šteje število opazovanih dogodkov na časovnem intervalu $[0, t]$ za i -to enoto. V primeru da pri vsaki enoti opazujemo le en dogodek, ima proces vrednost ena ($N_i(t) = 1$), če se je dogodek pri i -ti enoti zgodil do ali v času t in vrednost nič ($N_i(t) = 0$), če v tem času dogodek pri i -ti enoti še ni nastopil.

Čeprav se regresijski modeli za podatke števnih procesov lahko določijo na različne načine, pa se običajno modelira proces funkcije ogroženosti števnega procesa. Naj bo $\mathbf{x}(t) = (x_1(t), \dots, x_p(t))^T$ vektor neodvisnih spremenljivk i -te enote, funkcijo ogroženosti (ang. *intensity process*) za števni proces N_i pa zapišimo v obliki

$$\lambda_i(t) = Y_i(t)\alpha(t|\mathbf{x}_i). \quad (4.28)$$

V (4.28) je $Y_i(t)$ indikatorska spremenljivka, ki zavzame vrednost ena, če se pri i -ti enoti do časa t dogodek še ni zgodil (oziroma je i -ta enota v tem času še v množici tveganih enot) in vrednost nič sicer ($Y_i(t) = I(Z_i > t)$), pri čemer je Z_i opazovani, nekrnjeni ali krnjeni čas i -te enote). Stopnja ogroženosti $\alpha(t|\mathbf{x}_i)$ za i -to enoto je definirana pogojno na vrednosti neodvisnih spremenljivk, ki so bodisi določene na začetku opazovanja, lahko pa se v času tudi spreminja.

V Aalenovem modelu ima stopnja ogroženosti za i -to enoto v času t naslednjo obliko:

$$\alpha(t|\mathbf{x}_i) = \beta_0(t) + \beta_1(t)x_{i1}(t) + \dots + \beta_p(t)x_{ip}(t), \quad (4.29)$$

pri čemer je $\mathbf{x}_i(t) = (x_{i1}(t), x_{i2}(t), \dots, x_{ip}(t))^T$ pripadajoči vektor neodvisnih spremenljivk, $\beta_0(t)$ pa je osnovna ogroženost, ki je enaka stopnji ogroženosti, kadar so vrednosti vseh neodvisnih spremenljivk enake nič. Regresijski koeficient $\beta_j(t)$ pove, za koliko se v času t spremeni stopnja ogroženosti, če se j -ta neodvisna spremenljivka poveča za eno enoto. V tem smislu $\beta_j(t)$ lahko interpretiramo kot presežno ogroženost j -te neodvisne spremenljivke v času t . Parametri v enačbi (4.29) so arbitrarne regresijske funkcije, ki omogočajo, da se učinki neodvisnih spremenljivk v času spreminjajo. S tega vidika je model povsem neparametričen.

Na podlagi enačbe (4.28) funkcijo ogroženosti števnega procesa $N_i(t)$ zapišemo v naslednji obliki

$$\lambda_i(t) = Y_i(t)(\beta_0(t) + \beta_1(t)x_{i1}(t) + \dots + \beta_p(t)x_{ip}(t)). \quad (4.30)$$

Poenostavljeni povedano se ocenjevanje izvede v vsakem času, v katerem je nastopil dogodek, tako da regresiramo $dN_i(t)$ na neodvisne spremenljivke enot, ki so v času t v množici tveganih enot (ozioroma do časa t pri njih dogodek še ni nastopil). Posamezne regresije sicer slabo ocenijo prirastke $dB_q(t) = \beta_q(t)dt$, vendar stabilnost zagotovimo s kumulativnimi regresijskimi funkcijami $B_j(t) = \int_0^t \beta_j(u)du$, ki jih dobimo z agregiranjem ocenjenih prirastkov.

Če definiramo $dN_i(t) = \lambda_i(t)dt + dM_i(t)$, potem velja

$$dN_i(t) = Y_i(t)dB_0(t) + \sum_{j=1}^p Y_i(t)x_{ij}(t)dB_j(t) + dM_i(t), \quad i = 1, \dots, n. \quad (4.31)$$

Za dano vrednost opazovanega časa t ima povezava v (4.31) obliko običajnega regresijskega modela, v katerem so $dN_i(t)$ vrednosti odvisne spremenljivke, $Y_i(t)x_{ij}(t)$ proučevane neodvisne spremenljivke, $dB_j(t)$ parametri, ki jih ocenjujemo in $dM_i(t)$ slučajne napake.

Zgoraj opisani model pa lahko zapišemo tudi v matrični obliki

$$d\mathbf{N}(t) = \mathbf{X}(t)d\mathbf{B}(t) + d\mathbf{M}(t), \quad (4.32)$$

v kateri posamezni členi predstavljajo:

- vektor števnih procesov $\mathbf{N}(t) = (N_1(t), \dots, N_n(t))^T$,
- pripadajoči vektor martingalov (ang. *martingales*) $\mathbf{M}(t) = (M_1(t), \dots, M_n(t))^T$,
- vektor kumulativnih regresijskih funkcij $\mathbf{B}(t) = (B_0(t), B_1(t), \dots, B_p(t))^T$,
- in $n \times (p+1)$ razsežno matriko $\mathbf{X}(t)$ z elementi $Y_i(t), Y_i(t)x_{i1}(t), \dots, Y_i(t)x_{ip}(t)$.

Naj opozorimo, da se matrika $\mathbf{X}(t)$ spreminja na vsakem koraku ocenjevanja. Tako so vrstice matrike za enote, ki so v času t še v množici tveganih enot, enake pripadajočim vrednostim neodvisnih spremenljivk, vrstice za preostale enote pa se nadomestijo z ničelnimi vektorji.

Enačba (4.32) ima obliko linearrega regresijskega modela v matrični obliki s podatkovno matriko $\mathbf{X}(t)$. Kadar ima matrika $\mathbf{X}(t)$ polni rang, je rezultat običajne regresije najmanjših kvadratov enak

$$d\hat{\mathbf{B}}(t) = (\mathbf{X}(t)^T \mathbf{X}(t))^{-1} \mathbf{X}(t)^T d\mathbf{N}(t) \quad (4.33)$$

in je $(\mathbf{X}(t)^T \mathbf{X}(t))^{-1} \mathbf{X}(t)^T$ posplošena inverzna matrika, definirana na osnovi metode najmanjših kvadratov. V nadaljevanju jo bomo označili z $\mathbf{X}^-(t)$.

V matriki $\mathbf{X}(t)$ se vrednosti neodvisnih spremenljivk za enote, pri katerih do časa t nastopi dogodek, oziroma so do tega časa krmjene, nadomestijo z ničlami, zato proti koncu opazovanih časov matrika lahko izgubi polni rang. V ta namen vpeljemo indikator $J(t)$ z vrednostjo ena, kadar ima $\mathbf{X}(t)$ polni rang.

Z agregiranjem ocenjenih prirastkov $d\hat{\mathbf{B}}(t)$ po opazovanih časih dogodkov $T_1 < T_2 < \dots$, v katerih ima matrika $\mathbf{X}(t)$ polni rang, dobimo cenilko za vektor kumulativnih regresijskih funkcij

$$\hat{\mathbf{B}}(t) = \int_0^t J(u) \mathbf{X}^-(u) d\mathbf{N}(u) = \sum_{T_j \leq t} J(T_j) \mathbf{X}^-(T_j) \Delta \mathbf{N}(T_j). \quad (4.34)$$

$\Delta \mathbf{N}(T_j)$ v zgornji enačbi je vektor, ki ima na mestu enote, pri kateri se je v času T_j zgodil dogodek, vrednost ena, ostali elementi vektorja pa so enaki nič.

Ocenjene kumulativne regresijske funkcije $\hat{B}_q(t)$ za $q > 0$ opisujejo učinke proučevanih neodvisnih spremenljivk. Njihova interpretacija temelji na grafičnem prikazu, v katerem prikažemo gibanje posamezne funkcije $\hat{B}_q(t)$ v času. O učinku proučevane neodvisne spremenljivke na funkcijo ogroženosti $\lambda(t)$ pa sklepamo na podlagi naklona pripadajoče kumulativne regresijske funkcije. Če se učinek neodvisne spremenljivke v času ne spreminja, bo oblika ustrezne ocenjene kumulativne regresijske funkcije podobna premici. Pozitivni naklon premice odraža povečanje funkcije ogroženosti, negativni naklon pa zmanjšanje. Če neodvisna spremenljivka nima vpliva na funkcijo ogroženosti, bo naklon ustrezne kumulativne regresijske funkcije v vseh opazovanih časih dogodkov enak nič.

Za izpeljavo kovariančne matrike cenilke $\hat{\mathbf{B}}(t)$ najprej vpeljimo

$$\mathbf{B}^*(t) = \int_0^t J(u) d\mathbf{B}(u), \quad (4.35)$$

ki je zelo podobna $\mathbf{B}(t)$, kadar so verjetnosti za polni rang matrike $\mathbf{X}(u)$ za vse vrednosti $u \in [0, t]$ blizu ena. Če v (4.32) vstavimo (4.34), dobimo

$$\begin{aligned}\hat{\mathbf{B}}(t) &= \int_0^t J(u) \mathbf{X}^-(u) (\mathbf{X}(u) d\mathbf{B}(u) + d\mathbf{M}(u)) \\ &= \int_0^t J(u) (\mathbf{X}(u)^T \mathbf{X}(u))^{-1} \mathbf{X}(u)^T \mathbf{X}(u) d\mathbf{B}(u) + \int_0^t J(u) \mathbf{X}^-(u) d\mathbf{M}(u) \\ &= \int_0^t J(u) d\mathbf{B}(u) + \int_0^t J(u) \mathbf{X}^-(u) d\mathbf{M}(u),\end{aligned} \quad (4.36)$$

s čimer pridemo do ključne zveze

$$\hat{\mathbf{B}}(t) - \mathbf{B}^*(t) = \int_0^t J(u) \mathbf{X}^-(u) d\mathbf{M}(u). \quad (4.37)$$

Desna stran enačbe (4.37) predstavlja transformacijo martingala in je torej tudi martingal s povprečjem nič (ang. *a mean zero martingale*). Zaradi $E [\hat{\mathbf{B}}(t) - \mathbf{B}^*(t)] = \mathbf{0}$ je $\hat{\mathbf{B}}(t)$ nepristranska cenilka $\mathbf{B}^*(t)$ in skoraj nepristranska cenilka $\mathbf{B}(t)$.

Napovedljiv variacijski proces (ang. *predictable variation process*) martingala (4.37) zapišemo z vektorjem funkcij (procesov) ogroženosti (ang. *vector of intensity processes*) $\lambda(t) = (\lambda_1(t), \dots, \lambda_n(t))^T$

$$\langle \hat{\mathbf{B}} - \mathbf{B}^* \rangle (t) = \int_0^t J(u) \mathbf{X}^-(u) \text{diag}(\lambda(u) du) \mathbf{X}^-(u)^T, \quad (4.38)$$

pri čemer $\text{diag}(\mathbf{v})$ predstavlja diagonalno matriko z vektorjem \mathbf{v} na diagonali.

Cenilko kovariančne matrike za $\hat{\mathbf{B}}(t) - \mathbf{B}^*(t)$ in obenem tudi za $\hat{\mathbf{B}}(t)$ dobimo, če ocene za $\lambda(u) du$ vstavimo v izraz (4.38). $\lambda(u) du$ lahko ocenimo na dva načina. Če za oceno uporabimo $d\mathbf{N}(u)$, kovariančno matriko zapišemo kot

$$\hat{\Sigma}(t) = \sum_{T_j \leq t} J(T_j) \mathbf{X}^-(T_j) \text{diag}(\Delta \mathbf{N}(T_j)) \mathbf{X}^-(T_j)^T. \quad (4.39)$$

Druga možnost je, da $\lambda(t) du$ ocenimo z $\mathbf{X}(u) d\hat{\mathbf{B}}(u)$ in kovariančno matriko tako ocenimo na podlagi modela

$$\hat{\Sigma}_{mod}(t) = \sum_{T_j \leq t} J(T_j) \mathbf{X}^-(T_j) \text{diag}(\mathbf{X}(T_j) \Delta \hat{\mathbf{B}}(T_j)) \mathbf{X}^-(T_j)^T. \quad (4.40)$$

S pomočjo centralnega limitnega izreka za martingale (ang. *martingale central limit theorem*) je mogoče pokazati, da porazdelitev $\sqrt{n}(\hat{\mathbf{B}} - \mathbf{B}^*)$ konvergira k multivariatnemu normalnemu martingalu s povprečjem nič (ang. *a mean zero multivariate Gaussian martingale*) in neodvisnimi prirastki. Asimptotično med \mathbf{B}^* in \mathbf{B} ni razlike, tako da tudi $\sqrt{n}(\hat{\mathbf{B}} - \mathbf{B})$ šibko konvergira k navedenemu martingalu. Za dane vrednosti t je torej cenilka $\hat{\mathbf{B}}(t)$ porazdeljena približno normalno okoli pravih vrednosti $\mathbf{B}(t)$, kovariančno matriko pa lahko ocenimo s (4.39) ali (4.40).

Približni $100(1 - \alpha)\%$ interval zaupanja za q -to kumulativno regresijsko funkcijo $B_q(t)$ lahko ocenimo z

$$\hat{B}_q(t) \pm z_{1-\alpha/2} \sqrt{\hat{\sigma}_{qq}(t)}, \quad (4.41)$$

kjer je $z_{1-\alpha/2}$ $(1 - \alpha/2)$ -ti kvantil standardizirane normalne porazdelitve, $\hat{\sigma}_{qq}(t)$ pa q -ti diagonalni element kovariančne matrike $\hat{\Sigma}(t)$ oziroma $\hat{\Sigma}_{mod}(t)$.

4.2.2 Testiranje hipotez o učinkih neodvisnih spremenljivk

S testiranjem hipotez o učinkih neodvisnih spremenljivk v aditivnem modelu so se ukvarjali različni avtorji (McKeague in Utikal 1991a; Grønnesby 1997; Lee in Weissfeld 1998; Bhattacharyya in Klein 2005; Fosen in drugi 2006), tu pa bomo predstavili teste, ki jih izpeljemo z integriranjem uteženih funkcij (ang. *weight functions*), na podlagi katerih ocenimo kumulativne regresijske funkcije.

Ničelno domnevo, da posamezna neodvisna spremenljivka nima vpliva na funkcijo ogroženosti, zapišemo kot

$$H_0 : \beta_q(t) = 0, \quad \text{za vse } t \in [0, t_0],$$

pri čemer je t_0 neka ustrezno izbrana točka opazovanega časa. Za interval $[0, t_0]$ lahko izberemo celotno proučevano obdobje $[0, \tau]$, lahko pa se omejimo tudi na kraši interval.

V primeru pravilne ničelne hipoteze se vrednosti prirastkov $\Delta\hat{B}_q(T_j)$ gibljejo okrog ničle, če pa je $\beta_q(t) > 0$ ali $\beta_q(t) < 0$, bodo prirastki pozitivni oziroma negativni. Test hipoteze je torej zasnovan na stohastičnem integralu

$$Z_q(t_0) = \int_0^{t_0} L_q(t) d\hat{B}_q(t) = \sum_{T_j \leq t_0} L_q(T_j) \Delta \hat{B}_q(T_j), \quad (4.42)$$

kjer je $L_q(t)$ nenegativen napovedljiv "utežni proces" (ang. *predictable "weight process"*), za katerega se predpostavlja, da je enak nič, kadar je tudi $J(t) = 0$. Kot proces v t_0 je $Z_q(t_0)$ v primeru veljavne ničelne hipoteze martingal z $E(Z_q(t)) = 0$ in napovedljivim variacijskim procesom (ang. *predictable variation process*)

$$\langle Z_q \rangle (t_0) = \int_0^{t_0} L_q^2(t) d\langle \hat{B}_q \rangle (t). \quad (4.43)$$

Varianco za $Z_q(t_0)$ lahko ocenimo, če v izrazu (4.43) $d\langle \hat{B}_q \rangle (t)$ nadomestimo $d\hat{\sigma}_{qq}(t)$, ki predstavlja prirastek q -tega diagonalnega elementa v (4.39). Varianco torej ocenimo z

$$V_{qq}(t_0) = \int_0^{t_0} L_q^2(t) d\hat{\sigma}_{qq}(t) = \sigma_{T_j \leq t_0} L_q^2(T_j) \Delta \hat{\sigma}_{qq}(T_j), \quad (4.44)$$

kjer je $\Delta \hat{\sigma}_{qq}(T_j)$ prirastek q -tega diagonalnega elementa v (4.39) v času T_j . Testna statistika ničelne hipoteze v obliki $Z_q(t_0)V_{qq}(t_0)^{-1/2}$ pa je v primeru veljavnosti ničelne hipoteze porazdeljena standardizirano normalno.

Aalen je utežni proces $L_q(t)$ določil na podlagi matrike

$$\mathbf{K}(t) = \left(\left[(\mathbf{X}(t)^T \mathbf{X}(t))^{-1} \right] \right)^{-1}, \quad (4.45)$$

svojo izbiro pa utekeljil na podlagi navadne regresije najmanjših kvadratov, kjer so variance cenilk sorazmerne diagonalnim elementom matrike $\mathbf{X}^T \mathbf{X}^{-1}$. V tej matriki je \mathbf{X} podatkovna matrika, uteži ocen pa so v obrtnem sorazmerju z varianco. Pri testiranju domneve o ničelnem vplivu izbrane regresijske funkcije $\beta_q(t)$ na ogroženost, funkcijo uteži nato določimo s q -tim diagonalnim elementom matrike $\mathbf{K}(t)$: $L_q(t) = K_{qq}(t)$.

Tako dobljeno statistiko je Aalen imenoval *TST* statistika.

Utežni proces $L_q(t)$ v praksi sicer lahko določimo na različne načine, vendar na splošno ne moremo reči, katera izbira je najboljša. To je namreč odvisno od postavljenih alternativnih hipotez in od tega, kakšen pomen pripišemo posameznim časovnim intervalom.

4.2.3 Prileganje modela

Prileganje aditivnega modela lahko preverjamo na različne načine. V nalogi bomo predstavili grafične načine preverjanja prileganja modela, ki so tudi v praksi najpogosteje uporabljeni.

Na splošno razlikujemo štiri grafične načine prevrejanja prileganja, ki temeljijo na ostankih modela:

1. Ostanki martingala (ang. *martingale residuals*), ki jih definiramo na podlagi martingalov števnih procesov (ang. *martingale residual processes*):

$$\begin{aligned}\mathbf{M}_{res}(t) &= \int_0^t J(u)d\mathbf{N}(u) - \int_0^t J(u)\mathbf{X}(u)d\hat{\mathbf{B}}(u) = \\ &= \mathbf{N}^*(t) - \hat{\Lambda}^*(t),\end{aligned}\tag{4.46}$$

kjer so $\mathbf{N}^*(t)$ in $\hat{\Lambda}^*(t)$ vektorji števnih procesov in ocenjenih kumulativnih ogroženosti do časa, v katerem ima matrika $\mathbf{X}(t)$ še polni rang.

Ostanki martingala so določeni s procesi ostankov martingala v času $t = \tau$, to je v zadnji točki proučevanega časa preživetja:

$$\hat{M}_i = M_{res,i}(\tau).\tag{4.47}$$

Ostanke martingala lahko prikažemo grafično v odnosu do nedovisnih spremenljivk, vključenih v model. Če je model pravilno določen, ostanki ne bodo pokazali nikakršnih linearnih trendov s proučevanimi spremenljivkami.

2. Namesto ostankov martingala, ki jih gledamo v zadnji opazovani časovni točki, lahko namesto tega opazujemo procese ostankov martingala (ang. *martingale residual process plots*) (4.46) v času. Če se model dobro prilega podatkom, se prikazani ostanki gibljejo okoli osi x . Procese ostankov martingala lahko proučujemo tudi znotraj podskupin enot vzorca, ki so določene na podlagi kategoričnih neodvisnih spremenljivk.

3. Posplošeni ostanki za aditivni model so podobni Cox-Snellovim ostankom v Coxovem modelu. V Aalenovem linearinem modelu kumulativno ogroženost lahko ocenimo do časa, v katerem ima matrika $\mathbf{X}(t)$ še polni rang. Čas, v katerem matrika $\mathbf{X}(t)$ izgubi polni rang (označimo ga z R), predstavlja zadnji čas krnjenja. Za oblikovanje ostankov potem vse enote, ki so po tem času še v množici tveganih enot, krnimo v času R . Uvedba krnjenega časa R v statistične analize ne prinaša pristranskosti, saj z vidika martingala dejansko predstavlja čas prekinitve.

Označimo čas preživetja oziroma krnjeni čas i -te enote z S_i in predpostavimo, da se vrednosti neodvisnih spremenljivk v času ne spreminjajo (torej oznako za časovno odvisnost neodvisnih spremenljivk lahko izpustimo). Ocenjena vrednost kumulativne ogroženosti za izbrano enoto v času t zapišemo kot

$$\Lambda_i^*(t) = B^*(t)\mathbf{X}^i, \quad (4.48)$$

kjer je \mathbf{X}^i vektor vrednosti neodvisnih spremenljivk za i -to enoto. Ostanke definiramo kot $\Lambda_i^*(S_i)$ in so krnjeni, če je krnjen tudi opazovani čas. Če se model dobro prilega podatkom, potem so tako definirani krnjeni ostanki porazdeljeni po eksponentni porazdelitvi s parametrom 1. Ostanke lahko opazujemo tudi znotraj podskupin enot vzorca, ki so določene na podlagi kategoričnih neodvisnih spremenljivk.

4. Arjasov diagram (ang. *Arjas plot*) (Arjas 1988) je tesno povezan s procesi ostankov martingala (4.46). Namesto da procese ostankov prikažemo v času, lahko primerjamo $\mathbf{N}^*(t)$ in $\hat{\Lambda}^*(t)$, s čimer primerjamo število opazovanih in pričakovanih dogodkov. Tudi tako definirane ostanke je smiselno opazovati znotraj podskupin enot vzorca, ki jih določimo na podlagi kategoričnih neodvisnih spremenljivk.

4.3 Coxov model sorazmernih ogroženosti

Model sorazmernih ogroženosti za krnjene podatke je prvi predstavil Cox (Cox 1972), kasneje pa sta ga Andersen in Gill predstavila še v obliki števnih procesov in martingalov (Andersen in Gill 1982). Postavitev modela v okvir števnih procesov je omogočila

izjemno prilagodljivost modela, s čimer se je originalni model lahko obogatil z analizo ostankov, časovno odvisnimi koeficienti, večkratnimi/odvisnimi opazovanji, časovno odvisnimi stratumi in ocenjevanjem modelske funkcije ogroženosti. Coxov model je tako postal vodilno orodje na področju modeliranja krnjenih podatkov.

4.3.1 Opis metode

V Coxovem modelu stopnjo ogroženosti $\alpha(t|\mathbf{x}_i)$ za i -to enoto zapišemo z zvezo:

$$\alpha(t|\mathbf{x}_i) = \alpha_0(t)e^{\mathbf{x}_i^T \beta}, \quad (4.49)$$

kjer je \mathbf{x}_i vektor neodvisnih spremenljivk za i -to enoto, β je $p \times 1$ razsežni vektor regresijskih koeficientov in α_0 nedoločena nenegativna funkcija časa, ki ji pravimo tudi osnovna ogroženost. Model (4.49) je semiparametričen, saj je sestavljen tako iz neparametričnega (osnovna ogroženost) kot tudi parametričnega ($e^{\mathbf{x}_i^T \beta}$) dela. Eksponiranje parametričnega dela v Coxovem modelu zagotavlja nenegativnost stopnje ogroženosti.

Značilnost Coxovega modela predstavlja v času konstantno razmerje ogroženosti dveh enot z vektorjem fiksnih neodvisnih spremenljivk \mathbf{x}_i in \mathbf{x}_j :

$$\frac{\alpha(t|\mathbf{x}_i)}{\alpha(t|\mathbf{x}_j)} = \frac{\alpha_0(t)e^{\mathbf{x}_i^T \beta}}{\alpha_0(t)e^{\mathbf{x}_j^T \beta}} = \frac{e^{\mathbf{x}_i^T \beta}}{e^{\mathbf{x}_j^T \beta}}, \quad (4.50)$$

kar predstavlja tudi glavno predpostavko modela. Zaradi konstantnega razmerja ogroženosti se Coxov model imenuje tudi model sorazmernih ogroženosti.

Privlačnost Coxovega modela je tudi v njegovi preprosti interpretaciji ocenjenih koeficientov, kar je lepo ponazorjeno v naslednjem primeru. Predpostavimo dva vektorja neodvisnih spremenljivk, $\mathbf{x}_1(t)$ in $\mathbf{x}_2(t)$, ki se razlikujeta samo v j -ti komponenti. Če je $\mathbf{x}_{2j}(t) = \mathbf{x}_{1j}(t) + 1$, razmerje ogroženosti (4.50) postane enako

$$\frac{\alpha(t|\mathbf{x}_2)}{\alpha(t|\mathbf{x}_1)} = \exp \left\{ (\mathbf{x}_2(t) - \mathbf{x}_1(t))^T \beta \right\} = e^{\beta_j},$$

kar pomeni, da povečanje j -te neodvisne spremenljivke za eno enoto, ceteris paribus, spremeni njen učinek za zmnožek stopnje ogroženosti z e^{β_j} .

Semiparametrična oblika Coxovega modela onemogoča uporabo običajnih metod verjetja, značilnih za parametrične regresijske modele s krnjjenimi podatki. Cox je za ocenjevanje zato predlagal metodo delnega verjetja, ki je predstavljena v nadaljevanju.

Najprej zapišimo funkcijo ogroženosti za števni proces $N_i(t)$

$$\lambda_i(t) = Y_i(t)\alpha_0(t)e^{\mathbf{x}_i(t)^T\beta}, \quad (4.51)$$

in uvedimo agregirani števni proces $N_\bullet = \sum_{l=1}^n N_l(t)$, ki šteje število opazovanih dogodkov za vse enote. Funkcija ogroženosti za tako definirani števni proces je enaka:

$$\lambda_\bullet(t) = \sum_{l=1}^n Y_l(t)\alpha_0(t)e^{\mathbf{x}_l(t)^T\beta}. \quad (4.52)$$

Funkcijo ogroženosti števnega procesa $N_i(t)$ lahko faktoriziramo v obliki $\lambda_i(t) = \lambda_\bullet(t)\pi(i|t)$, kjer je $\pi(i|t)$ pogojna verjetnost opazovanega dogodka za i -to enoto v času t , pod pogojema, da se je dogodek zgodil v času t in je preteklost znana:

$$\pi(i|t) = \frac{\lambda_i(t)}{\lambda_\bullet(t)} = \frac{Y_i(t)e^{\mathbf{x}_i(t)^T\beta}}{\sum_{l=1}^n Y_l(t)e^{\mathbf{x}_l(t)^T\beta}}. \quad (4.53)$$

Delno verjetje za ocenjevanje regresijskih koeficentov dobimo z množenjem pogojnih verjetnosti (4.53) pri vseh časih, v katerih nastopi dogodek. Označimo s $T_1 < T_2 < \dots$ čase opazovanih dogodkov in predpostavimo, da se vsi dogodki zgodijo v različnih časih. Delno verjetje tako lahko zapišemo v naslednji obliki:

$$L(\beta) = \prod_{T_j} \pi(i_j|T_j) = \prod_{T_j} \frac{Y_{i_j}(T_j)e^{\mathbf{x}_{i_j}(T_j)^T\beta}}{\sum_{l=1}^n Y_l(T_j)e^{\mathbf{x}_l(T_j)^T\beta}}, \quad (4.54)$$

pri čemer indeks i_j označuje enoto, ki je dogodek doživela v času T_j .

Bolj poznano obliko delnega verjetja lahko zapišemo z uvedbo pojma množice enot v tveganju v času T_j , $\mathfrak{R}_j = \{l | Y_l(T_j) = 1\}$:

$$L(\beta) = \prod_{T_j} \frac{e^{\mathbf{x}_{i_j}(T_j)^T\beta}}{\sum_{l \in \mathfrak{R}_j} e^{\mathbf{x}_l(T_j)^T\beta}}. \quad (4.55)$$

Za ocenjevanje regresijskih koeficentov zgoraj definirano funkcijo delnega verjetja obravnavamo kot običajno funkcijo verjetja in parametre ocenimo z metodo največjega verjetja. Logaritem delnega verjetja (4.55) je enak

$$l(\beta) = \sum_{T_j} \left[\mathbf{x}_{i_j}(T_j)^T \beta - \log \sum_{l \in \mathfrak{R}_j} e^{\mathbf{x}_l(T_j)^T \beta} \right], \quad (4.56)$$

z odvajanjem po vektorju regresijskih koeficientov pa dobimo 'score' funkcijo

$$U(\beta) = \frac{\partial l(\beta)}{\partial \beta} = \sum_{T_j} \left[\mathbf{x}_{i_j}(T_j) - \sum_{l \in \mathfrak{R}_j} \mathbf{x}_l(T_j) \frac{e^{\mathbf{x}_l(T_j)^T \beta}}{\sum_{i \in \mathfrak{R}_j} e^{\mathbf{x}_i(T_j)^T \beta}} \right]. \quad (4.57)$$

Izraz $\sum_{l \in \mathfrak{R}_j} \mathbf{x}_l(T_j) \frac{e^{\mathbf{x}_l(T_j)^T \beta}}{\sum_{i \in \mathfrak{R}_j} e^{\mathbf{x}_i(T_j)^T \beta}}$ = $\bar{\mathbf{x}}(\beta, T_j)$ v (4.57) predstavlja uteženo povprečje \mathbf{x} -a na množici enot v tveganju \mathfrak{R}_j z utežmi $e^{\mathbf{x}_l(T_j)^T \beta}$.

Opazovano informacijsko matriko dobimo z drugim odvodom logaritmirane funkcije delnega verjetja:

$$I(\beta) = -\frac{\partial^2 l(\beta)}{\partial \beta \partial \beta^T} = \sum_{T_j} \left(\frac{\sum_{l \in \mathfrak{R}_j} e^{\mathbf{x}_l(T_j)^T \beta} \mathbf{x}_l(T_j) [\mathbf{x}_l(T_j) - \bar{\mathbf{x}}(\beta, T_j)]^T [\mathbf{x}_l(T_j) - \bar{\mathbf{x}}(\beta, T_j)]}{\sum_{i \in \mathfrak{R}_j} e^{\mathbf{x}_i(T_j)^T \beta}} \right), \quad (4.58)$$

kjer je izraz $\frac{\sum_{l \in \mathfrak{R}_j} e^{\mathbf{x}_l(T_j)^T \beta} \mathbf{x}_l(T_j) [\mathbf{x}_l(T_j) - \bar{\mathbf{x}}(\beta, T_j)]^T [\mathbf{x}_l(T_j) - \bar{\mathbf{x}}(\beta, T_j)]}{\sum_{i \in \mathfrak{R}_j} e^{\mathbf{x}_i(T_j)^T \beta}}$ = $V(\beta, T_j)$ utežena varianca \mathbf{x} -a v času T_j .

Cenilko regresijskih koeficientov $\hat{\beta}$ po metodi največjega verjetja dobimo z rešitvijo enačbe

$$U(\hat{\beta}) = 0. \quad (4.59)$$

Tako dobljena cenilka je konsistentna in asimptotično normalno porazdeljena s popvrečjem β (vektorjem pravih vrednosti regresijskih koeficientov) in varianco, dobljeno z inverzom pričakovane informacijske matrike $(E(I(\beta)))^{-1}$. Za izračun pričakovane informacijske matrike moramo poznati porazdelitev krnjjenja vseh enot (tudi tistih, pri katerih se dogodek zgodi), kar pa običajno ni mogoče. Zaradi tega se za varianco ocenjenih regresijskih koeficientov uporablja inverz opazovane informacijske matrike $(I^{-1}(\hat{\beta}))$.

4.3.2 Testiranje hipotez o učinkih neodvisnih spremenljivk

V Coxovem modelu učinke neodvisnih spremenljivk testiramo z običajnimi testi osnovanimi na funkciji verjetja. Za testiranje ničelne hipoteze

$$H_0 : \beta = \beta_0$$

lahko uporabimo:

- test razmerja verjetij $\chi_{LR}^2 = 2(l(\beta_0) - l(\hat{\beta}))$
- Waldov test $\chi_W^2 = (\hat{\beta} - \beta_0)^T I(\hat{\beta})(\hat{\beta} - \beta_0)$
- 'score' test $\chi_{SC}^2 = U(\beta_0)^T I(\beta_0)^{-1} U(\beta_0)$.

Za omenjene teste velja, da so asimptotično enakovredni, ob veljavni ničelni hipotezi pa so vsi porazdeljeni približno po χ^2 porazdelitvi s p stopinjami prostosti.

4.3.3 Prileganje modela

V tem delu predstavljamo dve možnosti preverjanja glavne predpostavke Coxovega modela, torej predpostavke sorazmernih ogroženosti modela. Obe možnosti sta osnovani na tako imenovanih Schoenfeldovih ostankih, ki so na podlagi 'score' funkcije definirani za vsako spremenljivko posebej, za vse čase T_k , v katerih je dogodek nastopil:

$$U_k(\beta) = \mathbf{x}_{i_k}(T_k) - \sum_{l \in \mathfrak{R}_k} \mathbf{x}_l(T_k) \frac{e^{\mathbf{x}_l(T_k)^T \beta}}{\sum_{i \in \mathfrak{R}_k} e^{\mathbf{x}_i(T_k)^T \beta}}. \quad (4.60)$$

Ob upoštevanju da drugi člen zgornje enačbe obravnavamo kot uteženo povprečje \mathbf{x} -a na množici enot v tveganju \mathfrak{R}_j , Schoenfeldove ostanke zapišemo kot

$$U_k(\beta) = \mathbf{x}_{i_k}(T_k) - \bar{\mathbf{x}}(\beta, T_k). \quad (4.61)$$

Na podlagi enačbe (4.61) Schoenfeldove ostanke za r -to neodvisno spremenljivko interpretiramo kot odstopanje dejanske vrednosti spremenljivke od modelske napovedi, pri čemer dejansko vrednost spremenljivke predstavlja vrednost pri enoti, pri kateri je dogodek nastopil v času T_k , modelsko napoved pa povprečje te spremenljivke, izračunano

za enote, ki so bile v času T_k še v tveganju.

Prvi od dveh obravnavanih testov za preverjanje predpostavke sorazmernih ogroženosti idejno izhaja iz postavitve alternativnega modela

$$\lambda(t) = \lambda_0(t)e^{x(t)\beta(t)}, \quad (4.62)$$

ki dovoljuje, da se regresijski parametri v času spreminjajo. V primeru sorazmernih ogroženosti je $\beta(t)$ v (4.62) enak β . Grambsch in Therneau (Grambsch in Therneau 1994) sta pokazala, da za regresijske koeficiente $\hat{\beta}$ ocnjene s Coxovim modelom velja

$$E(U_{kj}^*) + \hat{\beta}_j \approx \beta_j(T_k),$$

kjer so $U_k^* = V^{-1}(\hat{\beta}, T_k)U_k$ skalirani Schoenfeldovi ostanki z uteženo varianco x -a v času T_k , $V(\hat{\beta}, T_k)$. Skalirane Schoenfeldove ostanke tako lahko prikažemo grafično v odvisnosti od časa oziroma neke funkcije časa $g(t)$. Testna statistika temelji na zapisu regresijskih koeficentov v obliki:

$$\beta_j(t) = \beta_j + \theta_j(g_j(t) - \bar{g}_j), \quad j = 1, \dots, p, \quad (4.63)$$

kjer je \bar{g}_j povprečje posameznih vrednosti $g_j(T_k)$. Ničelni predpostavki sorazmernih ogroženosti je torej zadoščeno v primeru, ko je $\theta_j = 0$ za vse $j = 1, \dots, p$. Naj bo G_k $p \times p$ razsežna diagonalna matrika, v kateri je (j, j) -ti element enak $g_j(t_k) - \bar{g}_j$. Ob upoštevanju zvez $E(U_k^*) \approx G_k\theta$ in $Var(U_k^*) \approx V^{-1}(\beta, T_k)$ (Grambsch in Therneau 1994) parameter θ ocenimo z

$$\hat{\theta} = Q^{-1} \sum G_k s_k, \quad (4.64)$$

kjer je

$$Q = \sum G_k \hat{V}_k G_k - (\sum G_k \hat{V}_k)(\sum \hat{V}_k)^{-1} (\sum G_k \hat{V}_k)^T \quad (4.65)$$

z $\hat{V}_k = V(\hat{\beta}, T_k)$. Matrika Q^{-1} , ob veljavni ničelni predpostavki sorazmernih ogroženosti, predstavlja konsistenstno cenilko variance za θ , testna statistika ničelne hipoteze pa je enaka

$$T(G) = \hat{\theta}^T Q \hat{\theta} = (\sum G_k U_k)^T Q^{-1} (\sum G_k U_k) \quad (4.66)$$

in se ob veljavni ničelni predpostavki asimptotično približuje χ^2 porazdelitvi s p stopinjam prostosti.

Izkaže se (Grambsch in Therneau 1994), da so ocene varianc stabilne in se počasi spreminja, proti koncu opazovanega časa pa postanejo nestabilne, zato se v praksi pogosto uporablja njihova povprečna vrednost. Na podlagi povedanega in ob dejstvu, da je $\sum_k \hat{V}_k = I(\hat{\beta})$, lahko zapišemo povprečno vrednost informacijske matrike $\bar{V} = I/d$, kjer je d število opazovanih dogodkov. Sedaj matriko Q lahko izrazimo kot

$$Q = d^{-1} \sum G_k I G_k,$$

testna statistika pa se poenostavi. Naj bo U $d \times p$ razsežna matrika običajnih Schoenfeldovih ostankov, $U^* = dUI^{-1}$ pa matrika skaliranih Schoenfeldovih ostankov pod predpostavko nespreminjajočih se varianc. Testno statistiko za preverjanje predpostavke sorazmernih ogroženosti za vse spremenljivke hkrati sedaj lahko zapišemo kot

$$T = \frac{(g - \bar{g})^T U I^{-1} U^T (g - \bar{g})}{\sum (g_k - \bar{g})^2 / d} = \frac{(g - \bar{g})^T U^* I U^{*T} (g - \bar{g})}{d \sum (g_k - \bar{g})^2}. \quad (4.67)$$

Test sorazmernih ogroženosti za j -to neodvisno spremenljivko je osnovan na naklonu linearne povezave med časom oziroma neko funkcijo časa g_k in vrednostmi skaliranih Schoenfeldovih ostankov U_{kj}^* , ocenjenih z metodo najmanjših kvadratov:

$$\hat{\theta}_j = \frac{\sum_k (g_k - \bar{g}) U_{kj}^*}{\sum_k (g_k - \bar{g})^2} = \frac{\sum_k (g_k - \bar{g})(U_{kj}^* - \bar{U}_j^*)}{\sum_k (g_k - \bar{g})^2}, \quad (4.68)$$

testna statistika pa je enaka:

$$T_j = \frac{\left[\sum_k (g_k - \bar{g}) U_{kj}^* \right]^2}{d I_{jj} \sum_k (g_k - \bar{g})^2}, \quad (4.69)$$

pri čemer je v (4.68) in (4.69) $I^{jk} = I_{jk}^{-1}$ (j, k)-ti element matrike I^{-1} .

Drugi test (Pohar Perme 2007) temelji na standardiziranih Schoenfeldovih ostankih

$$R_k(\beta) = \frac{U_k(\beta)}{\sqrt{V_k(\beta)}}, \quad (4.70)$$

kjer je pričakovana vrednost U_k enaka 0, V_k pa je nepristranska cenilka variance. Po zgledu (O'Quigley 2003) in (O'Quigley in Stare 2003) se na podlagi ostankov (4.70) določi proces kumulativnih vsot

$$B_n(\beta, \frac{k}{n}) = \frac{1}{\sqrt{n}} \sum_{i=1}^k R_i(\beta), \text{ za } k = 1, \dots, n \text{ in } B_n(\beta, 0) = 0. \quad (4.71)$$

V (4.71) n ne predstavlja velikosti vzorca, temveč število nekrnjeneh enot. Čeprav slučajne spremenljivke R_i v zgornjem procesu niso niti enako porazdeljene niti neodvisne, je mogoče pokazati (Pohar Perme 2007), da proces pri pravih vrednostih parametrov ($B_n(\beta^0)$) konvergira k Brownovemu gibanju. Ker v praksi pravih vrednosti parametrov ne poznamo, jih nadomestimo z njihovimi ocenami $\hat{\beta}$. V tem primeru pa pogojna neodvisnost ostankov ne velja več in proces (4.71) ne konvergira k Brownovemu gibanju. Pohar (Pohar Perme 2007) zato namesto slučajnega procesa (4.71) uporabi

$$BP_n(\beta, \frac{k}{n}) = \frac{1}{\sqrt{n}} \left[\sum_{i=1}^k R_i(\beta) - \frac{k}{n} \sum_{i=1}^n R_i(\beta) \right], \quad (4.72)$$

ki pri pravih vrednostih parametrov β^0 konvergira k Brownovemu mostu, in pokaže, da je pri vzorčnih vrednostih $BP_n(\hat{\beta})$ dober približek $BP_n(\beta^0)$.

Testna statistika temelji na ideji, da se ob veljavni predpostavki sorazmernih ogroženosti Coxovega modela proces (4.72) na celotnem časovnem intervalu giblje blizu nič. V primeru, ko β v času ni konstantna, se to odrazi v gibanju procesa $BP_n(\beta, \frac{k}{n})$. Za testno statistiko se torej lahko uporabi

$$T = \max_k |BP_n(\beta, \frac{k}{n})|, \quad (4.73)$$

pri čemer je porazdelitev maksimalne absolutne vrednosti Brownovega mostu ($BB(t)$) enaka

$$P\left(\max_{t \in [0,1]} (|BB(t)|) \leq x\right) = 1 + 2 \sum_{k=1}^{\infty} (-1)^k e^{-2k^2x^2}, \quad x > 0. \quad (4.74)$$

4.4 Prednosti uporabe alternativnih metod v analizi relativnega preživetja

Relativno preživetje s transformacijskim pristopom dobi nove razsežnosti. S transformacijo opazovanega časa relativno preživetje definiramo na individualni ravni, kot delež primerljive splošne populacije, ki je umrla pred opazovano enoto. Transformirani čas oziroma individualna mera relativnega preživetja tako že sama po sebi prinaša drugačen pogled na relativno preživetje. Poleg tega opredelitev relativnega preživetja na individualni ravni omogoča, da za proučevanje uporabimo metode klasične analize preživetja in na zelo enostaven način pridemo do novih informacij, s pomočjo katerih lahko pogolobimo razumevanje in vedenje o relativnem preživetju proučevanih populacij. Tako nam na primer že enostavna ocena funkcije preživetja transformiranega časa pove, kolikšen delež proučevane populacije (denimo pacientov diagnosticiranih za rakom na pljučih) je preživel določen kvantil (npr. 5%, 10%, ...) splošne populacije. Po drugi strani možnost uporabe metod klasične analize preživetja omogoča tudi uporabo novih regresijskih metod v relativnem preživetju, ki se bistveno razlikujejo od trenutno obstoječih, kot sta denimo aditivni in multiplikativni regresijski model.

V primerjavi s "tradicionalnimi" metodami je neposredna prednost modeliranja s transformiranim časom v tem, da predpostavke o povezanosti med opazovano in populacijsko ogroženostjo niso potrebne, ne glede na to, katero regresijsko metodo klasične analize preživetja uporabimo. S transformacijo opazovanega časa namreč upoštevamo vso populacijsko ogroženost in ostane nam le presežna ogroženost, ki jo tudi neposredno modeliramo. Z različnimi modeli lahko pridobimo tudi nove informacije, s pomočjo katerih poglobimo vedenje o relativnem preživetju in tako vsebinsko obogatimo področje relativnega preživetja. Poleg tega, da pri modeliranju z obravnavanima modeloma ne potrebujemo predpostavk o povezanosti med opazovano in populacijsko ogroženostjo, ostale prednosti izhajajo iz lastnosti obravnavanih modelov. S pričakovanimi rezultati se tako ponuja možnost pridobitve informacij, s katerimi bi vsebinsko lahko obogatili področje relativnega preživetja. Predvsem je potrebno poudariti, da je tovrsten pristop modeliranja relativnega preživetja popolnoma nov, uporabnost informacij pa je odvisna od prileganja modelov podatkom transformiranega časa.

Prednosti Buckley-Jamesovega modela

V nasprotju z obstoječimi metodami v relativnem preživetju, ki so usmerjene v modeliranje funkcije presežne ogroženosti, Buckley-Jamesov model omogoča neposredno modeliranje opazovanega transformiranega časa oziroma individualne mere relativnega preživetja. Z vidika uporabe in interpretacije je model enostaven, njegova prednost pa je predvsem v tem, da ponuja vpogled v pričakovane vrednosti transformiranega časa.

Z vsebinskega vidika lahko doprinos Buckley-Jamesovega modela v relativnem preživetju strnemo v naslednjih točkah:

- Napovedovanje transformiranih časov oziroma individualne mere relativnega preživetja je možno le z uporabo ocenjenih parametrov modela. Pri ostalih regresijskih metodah relativnega preživetja je za napovedovanje potrebno oceniti tudi osnovno ogroženost (ang. *baseline hazard*) λ_0 , za kar je potrebna vsakokratna uporaba celotnega vzorca.
- Na podlagi napovedanih vrednosti transformiranih časov z inverzno transformacijo ($F^{-1}(Y) = T$) lahko izračunamo tudi pripadajoče pričakovane čase preživetja pacientov, kar z obstoječimi regresijskimi metodami relativnega preživetja ni mogoče. Transformacija individualne mere relativnega preživetja v čas preživetja pa interpretacijo rezultatov približa tudi laični javnosti.
- Transformirani časi povejo, kakšen delež splošne populacije je opazovanec preživel. Z ocenjevanjem pričakovanih transformiranih časov pacientov tako lahko ugotovimo, kolikšen delež splošne populacije bo pacient v povprečju preživel.
- Za dani delež splošne populacije, ki naj bi ga pacient v povprečju preživel, z Buckley-Jamesovim modelom lahko poiščemo kombinacijo neodvisnih spremenljivk modela, pri kateri pacient ta delež doseže.

Prednosti Aalenovega linearnega modela

Podobno kot večina ostalih regresijskih metod v klasični analizi preživetja tudi z Aalenovo regresijo modeliramo funkcijo ogroženosti. Glavna prednost Aalenovega modela

pa je zlasti njegova prilagodljivost, saj funkcijo ogroženosti neparametrično ocenjuje v vsaki točki opazovanega časa preživetja. Doprinos Aalenovega modela v relativnem preživetju je predvsem naslednji:

- Ocenjevanje v vsaki točki opazovanega transformiranega časa omogoča neposredno modeliranje časovno odvisnih spremenljivk, učinki neodvisnih spremenljivk pa se v času lahko spreminjajo. Ker učinke neodvisnih spremenljivk v obliki ocenjenih kumulativnih regresijskih funkcij prikažemo grafično, hitro lahko vidimo, kdaj pride do spremembe učinka v času.
- Z vidika transformiranega časa oziroma individualne mere relativnega preživetja so opazovani pacienti enakovredni glede na splošno populacijo, kar pomeni, da je porazdelitev transformiranega časa v primeru, ko na preživetje vplivajo samo populacijska tveganja, enakomerna na intervalu $[0, 1]$. Drugače povedano, v tem primeru spremenljivke, na podlagi katerih je definirana individualna mera relativnega preživetja, nimajo vpliva na opazovano ogroženost. Možnost spremeljanja sprememb učinkov v času bi torej lahko omogočila iskanje "točke ozdravitev" pacientov, ki jo lahko definiramo kot vrednost transformiranega časa, v katerem omenjene spremenljivke izgubijo vpliv na opazovano ogroženost pacientov.

4.5 Možne pomankljivosti proučevanih modelov v relativnem preživetju

Buckley-Jamesova linearna regresija

Na tem mestu naj spomnimo, da odvisna spremenljivka transformiranega časa Y leži na intervalu $[0, 1]$ in da so pri Buckley-Jamesovi regresiji vrednosti spremenljivke Y porazdeljene okoli pričakovanih vrednosti $E(Y|X)$. To pomeni, da pri modeliranju vrednosti transformiranega časa lahko padejo izven intervala $[0, 1]$.

Praktično gledano vrednosti transformiranega časa nad 1 ne bi smele predstavljati bistvenega problema. V tem primeru se namreč lahko zgodi dvoje:

1. Vrednost $Y > 1$ se lahko pojavi znotraj opazovanega obdobja, kar pomeni, da je pacient preživel celotno populacijo, ki je bila v času diagnoze iste starosti. Tablice

smrtnosti zajemajo verjetnosti za smrt splošne populacije do starosti 110 let. Torej pacient, ki doseže vrednost $Y > 1$ znotraj opazovanega obdobja, predstavlja ekstremno vrednost transformiranega časa, ki je v resnični situaciji malo verjetna.

2. Druga možnost je, da pacient doseže vrednost transformiranega časa $Y > 1$ po opazovanem obdobju. V tem primeru bo vrednost transformiranega časa, čeprav ekstremna, krnjena, še preden jo pacient dejansko doseže.

Večji problem predstavljajo modelsko nakazane negativne vrednosti transformiranega časa Y , saj je transformirani čas nenegativna spremenljivka. Problem lahko obidemo, če namesto transformiranega časa vzamemo na primer njegove logaritmirane vrednosti, $\log(Y)$.

Ostale pomankljivosti Buckley-Jamesovega modela so povezane s predpostavkami modela, ki so podobne tistim v navadni linearni regresiji:

- model je linearen v parametrih (predpostavka o linearnosti modela),
- porazdelitev ostankov je neodvisna od vrednosti neodvisnih spremenljivk v modelu (predpostavka o homoskedastičnosti).

V primeru krnjениh vrednosti odvisne spremenljivke je preverjanje teh predpostavk oteženo, če ne celo nemogoče (Stare in drugi 2000). Medtem ko predpostavko o strogi linearnosti modela lahko opustimo z uvedbo nelinearnih členov v model, pa problem preverjenja predpostavke o homoskedastičnosti zaenkrat ostaja nerešen. Prav zato je priporočilo za "varno" uporabo Buckley-Jamesove linearne regresije nizka stopnja krnjjenja podatkov, ki omogoča, da predpostavke modela preverimo z običajnimi grafičnimi metodami linearne regresije (Stare in drugi 2000).

Aalenova linearna regresija

Enačba (4.30) kaže na najočitnejšo pomankljivost Aalenovega modela. Izraz za funkcijo opazovane ogroženosti $\lambda(t)$ namreč ni omejen na nenegativne vrednosti. To pomeni, da na podlagi modela ocenjene funkcije preživetja niso nujno monotono padajoče, temveč občasno lahko naraščajo. Kljub temu Aalen ugotavlja (Aalen 1989), da je v primeru dovolj velikega vzorca in neekstremnih vrednosti neodvisnih spremenljivk

problem zanemarljiv.

Pomankljivost modela je tudi v tem, da je ocenjevanje možno le do časa, ko matrika $\mathbf{X}(t)$ izgubi polni rang. V primeru modeliranja kategoričnih spremenljivk, po katerih se preživetje pacientov bistveno razlikuje, to lahko povzroči izgubo informacij za paciente z višjim preživetjem. Za primer vzemimo bolnike, ki so bili diagnosticirani v različnih stadijih bolezni. Zgodi se lahko, da vse smrti bolnikov, diagnosticiranih v višjih stadijih, nastopijo precej pred smrtmi bolnikov, ki so bili diagnosticirani v nižjih stadijih bolezni. Matrika $\mathbf{X}(t)$ izgubi polni rang v primeru, ko umre zadnji bolnik diagnosticiran v določenem stadiju. To pomeni, da bi kakršnekoli spremembe v preživetju bolnikov, diagnosticiranih v nižjih stadijih, lahko ostale neopažene. V takih primerih je potrebno biti previden in preveriti, koliko enot preživi zadnji čas ocenjevanja in koliko dogodkov še nastopi po tem času. Kadar je število takih dogodkov relativno veliko, je smiselno bolnike, diagnosticirane v nižjih stadijih bolezni, proučevati ločeno.

5. Primerjava Aalenovega in Buckley-Jamesovega regresijskega modela

5.1 Teoretična primerjava

V tem delu prikazujemo primerjavo med Aalenovim in Buckley-Jamesovim modelom, s čimer želimo ugotoviti, ali spreminjači se učinki v Aalenovem modelu vplivajo na parametre Buckley-Jamesovega modela. Za lažje razumevanje navedimo najprej nekaj osnovnih zvez povezanih s proučevanim časom preživetja T .

Gostota verjetnosti časa T je enaka

$$f(t) = \frac{dF(t)}{dt},$$

pri čemer porazdelitveno funkcijo $F(t)$ lahko izrazimo s pomočjo funkcije preživetja $S(t)$:

$$F(t) = 1 - S(t).$$

Gostoto verjetnosti proučevanega časa tako lahko zapišemo v obliki

$$f(t) = \frac{-dS(t)}{dt}.$$

Spomnimo, da je stopnja ogroženosti definirana kot

$$\begin{aligned}\alpha(t) &= \lim_{h \rightarrow 0} P(t \leq T < t+h | T \geq h) / h \\ &= f(t) / S(t) \\ &= -d \log S(t) dt.\end{aligned}$$

Z integriranjem α po t in upoštevanjem $S(0) = 1$ dobimo

$$\int_0^t \alpha(s)ds = -\log S(t)$$

ozziroma

$$S(t) = e^{-\int_0^t \alpha(s)ds}. \quad (5.1)$$

Na podlagi (5.1) gostoto verjetnosti lahko zapišemo kot

$$f(t) = \alpha(t)e^{-\int_0^t \alpha(s)ds}.$$

V Aalenovem modelu je stopnja ogroženosti definirana kot linearna funkcija neodvisnih spremenljivk:

$$\alpha(t_i | \mathbf{x}_i, \beta_0^A, \dots, \beta_p^A) = \beta_0^A(t) + \dots + \beta_p^A(t)x_{ip},$$

kjer označka β_\bullet^A predstavlja označko za parametre β Aalenovega modela.

Za poenostavitev predpostavljajmo, da je odvisna spremenljivka pri posameznih vrednostih neodvisnih spremenljivk \mathbf{x}_i eksponentna slučajna spremenljivka $T_i | \mathbf{x}_i$ s parametrom $\lambda = \beta_0^A + \dots + \beta_p^A x_{ip}$. Pričakovana vrednost in varianca odvisne spremenljivke $T_i | \mathbf{x}_i$ sta potem enaki

$$E(T_i | \mathbf{x}_i, \beta_0^A, \dots, \beta_p^A) = \frac{1}{\beta_0^A + \dots + \beta_p^A x_{ip}}$$

ozziroma

$$V(T_i | \mathbf{x}_i, \beta_0^A, \dots, \beta_p^A) = \frac{1}{(\beta_0^A + \dots + \beta_p^A x_{ip})^2}.$$

Po definiciji parameter β_k^B linearnega Buckley-Jamesovega modela ustreza odvodu pričakovane vrednosti $E(T_i | \mathbf{x}_i, \beta_0^B, \dots, \beta_p^B)$ po ustrejni neodvisni spremenljivki x_k :

$$\beta_k = \frac{dE(T_i | \mathbf{x}_i, \beta_0^B, \dots, \beta_p^B)}{dx_k}. \quad (5.2)$$

Pri tem je vrednost parametra neodvisna od vrednosti neodvisnih spremenljivk. V primeru ko podatke generira Aalenov model, pa so parametri linearnega modela odvisni

tako od parametrov Aalenovega modela, kot tudi od kombinacije vrednosti neodvisnih spremenljivk. Parameter Buckley-Jamesovega modela s pomočjo (5.2) lahko zapišemo kot

$$E(T_i | \mathbf{x}_i, \beta_0^B, \dots, \beta_p^B) = E(T_i | \mathbf{x}_i, \beta_0^A, \dots, \beta_p^A)$$

$$\beta_0^B + \dots + \beta_p^B x_{ip} = \frac{1}{\beta_0^A + \dots + \beta_p^A x_{ip}},$$

$$\beta_{ik}^B | \mathbf{x}_i = \frac{-\beta_k^A}{(\beta_0^A + \dots + \beta_k^A x_{ik} + \dots + \beta_p^A x_{ip})^2}.$$

To sicer onemogoča neposredno prevedbo Aalenovega modela v stogo linearen Buckley-Jamesov model (kršeni predpostavki o linearnosti in homoskedastičnosti), omogoča pa intuitivni razmislek o vplivu spremenljajočih vrednosti parametrov Aalenovega modela na parametre Buckley-Jamesovega modela.

V primeru nespreminjajočih parametrov Aalenovega modela je $\beta(t)^A = \beta^A$, pričakovana vrednost opazovanega časa $E(T | \mathbf{x}_i, \beta_0^A, \beta_1^A, \dots, \beta_p^A)$ pa konstantna za vsako vrednost neodvisne spremenljivke x_i oziroma za vsako kombinacijo vrednosti neodvisnih spremenljivk \mathbf{x}_i , kar je tudi predpostavka linearnega regresijskega modela. Če zanemarimo, da so parametri Buckley-Jamesovega modela, v primeru ko podatke generira Aalenov model, odvisni od vrednosti neodvisnih spremenljivk, je pričakovana vrednost $E(T | \mathbf{x}_i, \beta_0^A, \dots, \beta_p^A x_{ip})$ hkrati tudi pričakovana vrednost opazovanega časa, modeliranega z Buckley-Jamesovim modelom:

$$E(T_i | \mathbf{x}_i, \beta_0^A, \dots, \beta_p^A) = E(T_i | \mathbf{x}_i, \beta_{i0}^B, \dots, \beta_{ip}^B) = \beta_{i0}^B + \dots + \beta_{ip}^B x_i.$$

Vse vrednosti spremenljivke $T | \mathbf{x}_i$ so torej porazdeljene z enakim povprečjem in enako varianco.

V primeru spremenljajočih se vrednosti parametrov Aalenovega modela je pričakovana vrednost opazovanega časa pri vrednosti \mathbf{x}_i odvisna od vrednosti parametrov Aalen-

ovega modela pri različnih časih t , kar pomeni, da sta pričakovana vrednost in varianca opazovanega časa (odvisne spremenljivke) pri enaki vrednosti neodvisnih spremenljivk x_i različni. Klasična linearna regresija (Buckley-Jamesov model v primeru krnjene podatkov) torej ne zazna spreminjačega učinka parametrov Aalenovega modela, saj predpostavlja, da so podatki pri x_i generirani z enako pričakovano vrednostjo in varianco. Parameter klasične linearne regresije bo torej odvisen od pogojne porazdelitve odvisne spremenljivke pri x_i , ki v tem primeru predstavlja mešanico različnih pogojnih porazdelitev, odvisnih od vrednosti parametrov Aalenovega modela pri določenih vrednostih odvisne spremenljivke.

5.2 Simulacijska primerjava

V nadaljevanju je s pomočjo simulacij prikazano, kako se obnašajo ocjenjeni parametri modela klasične linearne regresije, ko podatke generira Aalenov model. S simulacijami želimo pokazati:

- kako se obnašajo parametri modela klasične linearne regresije, ko so parametri Aalenovega modela konstantni,
- kako se obnašajo parametri modela klasične linearne regresije, ko se parametri Aalenovega modela spreminjajo,
- kakšen je vpliv krnjenga na parametre Buckley-Jamesove linearne regresije za zgoraj navedena primera.

Spodaj prikazane simulacije so zasnovane na naslednji način. Podatke generira Aalenov model z dvema zveznima, na intervalu $(1, 10)$ enakomerno porazdeljenima spremenljivkama:

$$\alpha(t_i|x_{i1}, x_{i2}, \beta_1^A, \beta_2^A) = \beta_1^A(t)x_{i1} + \beta_2^A(t)x_{i2},$$

pri čemer predpostavljamo, da je pogojna porazdelitev odvisne spremenljivke eksponentna s parametrom $\alpha(t_i|x_{i1}, x_{i2}, \beta_1^A, \beta_2^A)$. V primeru ko se parametri Aalenovega modela v času ne spreminjajo, so izbrane vrednosti parametrov enake $\beta_1 = \beta_2 = 1$. V

primeru spreminjajočih parametrov β_1 v času t^* spremeni vrednost v $\beta_1(t > t^*) = 5$. Čas t^* je izbran tako, da delež enot z vrednostjo odvisne spremenljivke $t \leq t^*$ (v povprečju) ustreza vnaprej izbranemu deležu vseh enot, ki je v simulacijah enak 0,2, 0,5 in 0,8. Za proučevanje učinka krvnenja je v simulaciji uporabljen slučajno krvnenje. Izbrani deleži krvnenih enot so enaki 0,05, 0,30 in 0,50, pri čemer smo predvsem prvo in zadnjo vrednost izbrali zaradi lažje primerjave z realnimi podatki o raku na pljučih in raku na dojki. Vse simulacije so izpeljane za tri različne velikosti vzorcev, $n = 100, 500, 1000$, vsaka simulacija pa je ponovljena 500-krat.

Koeficiente Buckley-Jamesove regresije ocenujemo z dvema modeloma. V prvi model sta vključeni obe neodvisni spremenljivki na podlagi katerih generiramo podatke

$$E(T|x_1, x_2) = \beta_0^{BJ} + \beta_1^{BJ}x_1 + \beta_2^{BJ}x_2,$$

v drugi model pa je vključena le prva neodvisna spremenljivka

$$E(T|x_1) = \beta_0^{BJ} + \beta_1^{BJ}x_1.$$

V vseh simulacijah ocenjene koeficiente Buckley-Jamesove regresije primerjamo s teoretičnimi koeficenti, ki jih računamo za vsakokratno izbrani vzorec. Spomnimo, da so pri prevedbi Aalenovega linearne modela v model klasične linearne regresije parametri slednjega odvisni tako od parametrov Aalenovega modela, kot tudi od vrednosti neodvisnih spremenljivk. Ocjenjeni parameter v strogo linearinem klasičnem regresijskem modelu je torej neko povprečje parametrov pri različnih vrednostih neodvisnih spremenljivk. Tako imenovane teoretične koeficiente Buckley-Jamesovega modela ($\beta_k^{BJ.T}$) za izbrani vzorec pa smo izračunali kot koeficiente klasične linearne regresije z vzorčnimi vrednostmi neodvisnih spremenljivk in pogojnimi pričakovanimi vrednostmi Aalenovega modela kot odvisno spremenljivko:

$$E(T|\mathbf{x}, \beta^A) = \mathbf{x}\beta^{BJ.T} + \epsilon_T.$$

Primerjava med ocenjenimi in tako imenovanimi teoretičnimi koeficenti je narejena s t-testom za testiranje razlik med dvema povprečjema, pri čemer predpostavljamo, da varianci ocenjenih in teoretičnih cenilk parametrov (tudi ti so izračunani na podlagi

vzorca) nista enaki. Če varianco teoretičnega parametra označimo z s_1^2 in varianco ocjenjenega parametra z s_2^2 , je cenilka variance t-statistike enaka

$$s_{\hat{\beta}^{BJ} - \hat{\beta}^{BJ.T}}^2 = \frac{s_1^2 + s_2^2}{n},$$

stopinje prostosti pa so izračunane z enačbo

$$df = \frac{(s_1^2 + s_2^2)^2 / n^2}{((s_1^2)^2 + (s_2^2)^2) / (n^2(n-1))}.$$

Simulacije pri nespreminjajočih vrednostih parametrov Aalenovega modela

V tem delu prikazujemo rezultate simulacij, kjer podatke generiramo z Aalenovim modelom, v katerem se parametri v času ne spreminja:

$$\beta_1^A(t) = \beta_2^A(t) = 1.$$

Predpostavljamo, da je odvisna spremenljivka (T) eksponentna slučajna spremenljivka s parametrom

$$\begin{aligned}\alpha(t_i | x_{i1}, x_{i2}, \beta_1^A, \beta_2^A) &= \beta_1^A(t_i)x_{i1} + \beta_2^A(t_i)x_{i2} \\ &= 1 \cdot x_{i1} + 1 \cdot x_{i2}.\end{aligned}$$

Pogojne pričakovane vrednosti odvisne spremenljivke so torej enake

$$\begin{aligned}E(T_i | x_{i1}, x_{i2}, \beta_1^A, \beta_2^A) &= \frac{1}{\beta_1^A(t_i)x_{i1} + \beta_2^A(t_i)x_{i2}} \\ &= \frac{1}{1 \cdot x_{i1} + 1 \cdot x_{i2}},\end{aligned}$$

teoretični parametri Buckley-Jamesovega modela pa so določeni kot parametri klasične linearne regresije

$$E(T | x_1, x_2, \beta_1^A, \beta_2^A) = \beta_0^{BJ.T} + \beta_1^{BJ.T}x_1 + \beta_2^{BJ.T}x_2 + \epsilon_T.$$

Tabeli 5.1 in 5.2 prikazujeta povprečni vrednosti teoretičnih in z Buckley-Jamesovim modelom ocjenjenih parametrov, izračunanih na podlagi 500 vzorcev. Rezultati kažejo, da se povprečni vrednosti ocjenjenih in teoretičnih vrednosti parametrov v primeru

nekrnjenih podatkov statistično značilno ne razlikujeta, kar pomeni, da so cenilke Buckley-Jamesovega modela na nekrnjenih podatkih, generiranih z zgoraj opisanim Aalenovim modelom, nepristranske. S krnjem Buckley-Jamesova cenilka parametrov postane pristranska. Absolutne vrednosti ocenjenih parametrov postanejo pričakovano podcenjene, saj so višje vrednosti odvisne spremenljivke ob sočasni nelinearnosti v parametrih krnjene pogoste. Višja stopnja krnjena povzroči tudi višjo pristranskost, velikost vzorca pa nanjo bistveno ne vpliva.

Simulacije pri spremnjajočih vrednostih parametrov Aalenovega modela

Za primer ko se parametri Aalenovega modela v času spremnjajo, so v simulacijah podatki generirani tako, da je odvisna spremenljivka do določenega časa (t^*) porazdeljena eksponentno s parametrom

$$\alpha_1 = \alpha(t_i | t_i \leq t^*, x_{i1}, x_{i2}, \beta_1^A, \beta_2^A) = \beta_1^A(t \leq t^*)x_{i1} + \beta_2^A x_{i2},$$

po času t^* pa se vrednost prvega parametra Aalenovega modela spremeni, s čimer se spremeni tudi parameter eksponentne slučajne spremenljivke:

$$\alpha_2 = \alpha(t_i | t_i > t^*, x_{i1}, x_{i2}, \beta_1^A, \beta_2^A) = \beta_1^A(t > t^*)x_{i1} + \beta_2^A x_{i2}.$$

Vrednosti parametrov Aalenovega modela so v simulacijah enake

$$\begin{aligned}\beta_1^A(t \leq t^*) &= 1 \\ \beta_1^A(t > t^*) &= 5 \\ \beta_2^A &= 1.\end{aligned}$$

Odvisna spremenljivka T je tako mešanica dveh eksponentnih slučajnih spremeljivk,

$$T_1 \sim \exp(\beta_1^A(t \leq t^*)x_{i1} + \beta_2^A x_{i2})$$

in

$$T_2 \sim \exp(\beta_1^A(t > t^*)x_{i1} + \beta_2^A x_{i2})$$

njena pogojna pričakovana vrednost pa je enaka

$$E(T|x_i, \beta_1(t), \beta_2) = \phi E(T_1|T_1 \leq t^*) + (1 - \phi)E(T_2|T_2 > t^*).$$

V zgornji enačbi je

$$E(T_1|T_1 \leq t^*) = \frac{1}{\alpha_1} - t^* \frac{S(t^*; \alpha_1)}{F(t^*; \alpha_1)}$$

in

$$E(T_2|T_2 > t^*) = \frac{1}{\alpha_2} + t^*.$$

Utež ϕ je določena kot

$$\phi = \frac{P(T_1 \leq t^*)}{P(T_1 \leq t^*) + P(T_2 > t^*)},$$

pri čemer je $P(T_1 \leq t^*) = F(t^*; \alpha_1)$ in $P(T_2 > t^*) = S(t^*; \alpha_2)$.

Preломna točka t^* , v kateri se vrednost parametra β_1^A v Aalenovem modelu spremeni, je izbrana tako, da je delež časov, ki so v povprečju manjši od t^* , enak 0,2, 0,5 in 0,8.

Rezultati simulacij zgoraj opisanega postopka generiranja podatkov in računanja ocenjenih in teoretičnih vrednosti parametrov klasične linearne regresije so prikazani v tabelah od 5.3 do 5.8, iz njih pa je razvidno, da so ocene Buckley-Jamesovega modela prisranske ne glede na izbrano točko preloma in stopnjo krnjenja. Na smer pristranskosti pričakovano vpliva tudi točka preloma, ki določa, kakšen delež vrednosti odvisne spremenljivke pri posameznih vrednostih spremenljivk je generiran z različnima vrednostima parametra v Aalenovem modelu.

Pri zgornji interpretaciji je potrebno upoštevati, da z linearnim modelom modeliramo nelinearne in heteroskedastične podatke. V konkretnem primeru, se s časovno spremembijo parametrov v Aalenovem modelu, učinek nelinearnosti še poveča, kar vpliva na parametre Buckley-Jamesovega modela. Poleg tega na ocnjene parametre vpliva tudi delež enot pod in nad prelomno točko. Pri tem velja, da več kot je enot pod prelomno točko, manjši je vpliv na ogroženost nastopa dogodka. Ugotovimo lahko, da v konkretnem primeru višja stopnja krnjenja vpliva na podcenjenost ocenjenih parametrov,

medtem ko povečevanje deleža enot pod prelomno točko povečuje stopnjo nelinearnosti, s tem pa negativni učinek ocenjenega parametra. Učinka pristranskosti na ocenjene parametre tako vplivata v obratni smeri, zaradi česar v določenih primerih razlika med ocenjenimi in teoretičnimi parametri postane neznačilna.

Tabela 5.1: Parametri Buckley-Jamesovega modela pri v času konstantnih parametrih Aalenovega modela (Model 1)

	n		Pravi	Ocenjeni	t	df	p-vrednost
nekrnjeni	100	β_1	-0.012	-0.012	-0.115	113.520	0.909
		β_2	-0.012	-0.012	0.368	112.063	0.714
	500	β_1	-0.012	-0.012	-0.986	560.485	0.325
		β_2	-0.012	-0.012	-0.138	564.192	0.891
krnjeni (0.05)	1000	β_1	-0.012	-0.012	0.042	1131.113	0.966
		β_2	-0.012	-0.012	0.076	1127.920	0.939
	100	β_1	-0.012	-0.011	-2.323	113.915	0.022
		β_2	-0.012	-0.011	-2.155	115.037	0.033
krnjeni (0.3)	500	β_1	-0.012	-0.011	-10.254	571.892	0.000
		β_2	-0.012	-0.011	-10.364	575.729	0.000
	1000	β_1	-0.012	-0.011	-20.604	1148.307	0.000
		β_2	-0.012	-0.011	-20.314	1152.821	0.000
krnjeni (0.5)	100	β_1	-0.012	-0.006	-15.906	129.972	0.000
		β_2	-0.012	-0.007	-15.746	132.640	0.000
	500	β_1	-0.012	-0.007	-72.890	663.846	0.000
		β_2	-0.012	-0.007	-78.128	693.665	0.000
	1000	β_1	-0.012	-0.007	-147.474	1301.103	0.000
		β_2	-0.012	-0.007	-140.858	1304.288	0.000
	100	β_1	-0.012	-0.004	-27.608	150.126	0.000
		β_2	-0.012	-0.004	-27.574	154.665	0.000
	500	β_1	-0.012	-0.004	-142.957	779.259	0.000
		β_2	-0.012	-0.004	-147.238	788.521	0.000
	1000	β_1	-0.012	-0.004	-289.158	1611.433	0.000
		β_2	-0.012	-0.004	-276.200	1548.421	0.000

Tabela 5.2: Parametri Buckley-Jamesovega modela pri v času konstantnih parametrih Aalenovega modela (Model 2)

	n		Pravi	Ocenjeni	t	df	p-vrednost
nekrnjeni	100	β_1	-0.012	-0.012	-0.121	127.501	0.904
	500	β_1	-0.012	-0.012	-0.913	625.521	0.362
	1000	β_1	-0.012	-0.012	0.069	1242.834	0.945
krnjeni (0.05)	100	β_1	-0.012	-0.011	-2.227	129.864	0.028
	500	β_1	-0.012	-0.011	-9.567	629.236	0.000
	1000	β_1	-0.012	-0.011	-19.010	1289.204	0.000
krnjeni (0.3)	100	β_1	-0.012	-0.006	-14.571	157.452	0.000
	500	β_1	-0.012	-0.007	-66.949	801.960	0.000
	1000	β_1	-0.012	-0.007	-134.622	1580.156	0.000
krnjeni (0.5)	100	β_1	-0.012	-0.004	-25.352	183.706	0.000
	500	β_1	-0.012	-0.004	-126.575	944.392	0.000
	1000	β_1	-0.012	-0.004	-254.183	1963.715	0.000

Tabela 5.3: Parametri Buckley-Jamesovega modela pri časovno odvisnih parametrih Aalenovega modela, pri čemer je 20 % enot manjših od prelomne točke t^* (Model 1)

	n	Pravi	Ocenjeni	t	df	p-vrednost	
nekrnjeni	100	β_1	-0.007	-0.006	-2.594	110.462	0.011
		β_2	-0.002	-0.002	-0.953	109.336	0.343
	500	β_1	-0.007	-0.006	-12.438	562.214	0.000
		β_2	-0.002	-0.002	-4.233	558.270	0.000
	1000	β_1	-0.007	-0.006	-21.248	1122.296	0.000
		β_2	-0.002	-0.002	-6.483	1122.583	0.000
krnjeni (0.05)	100	β_1	-0.007	-0.006	-4.420	112.126	0.000
		β_2	-0.002	-0.002	-1.993	113.983	0.049
	500	β_1	-0.007	-0.006	-22.856	563.217	0.000
		β_2	-0.002	-0.002	-9.334	572.168	0.000
	1000	β_1	-0.007	-0.006	-44.485	1125.235	0.000
		β_2	-0.002	-0.002	-14.746	1130.174	0.000
krnjeni (0.3)	100	β_1	-0.007	-0.004	-20.269	129.046	0.000
		β_2	-0.002	-0.001	-7.247	127.439	0.000
	500	β_1	-0.007	-0.004	-96.297	635.965	0.000
		β_2	-0.002	-0.001	-34.880	624.958	0.000
	1000	β_1	-0.007	-0.004	-194.304	1266.992	0.000
		β_2	-0.002	-0.001	-71.135	1276.462	0.000
krnjeni (0.5)	100	β_1	-0.007	-0.002	-36.470	145.167	0.000
		β_2	-0.002	-0.001	-11.973	135.107	0.000
	500	β_1	-0.007	-0.003	-174.948	670.915	0.000
		β_2	-0.002	-0.001	-57.992	654.441	0.000
	1000	β_1	-0.007	-0.003	-354.923	1461.985	0.000
		β_2	-0.002	-0.001	-105.782	1319.851	0.000

Tabela 5.4: Parametri Buckley-Jamesovega modela pri časovno odvisnih parametrih Aalenovega modela, pri čemer je 20 % enot manjših od prelomne točke t^* (Model 2)

	n		Pravi	Ocenjeni	t	df	p-vrednost
nekrnjeni	100	β_1	-0.007	-0.006	-2.480	114.408	0.015
	500	β_1	-0.007	-0.006	-12.132	586.467	0.000
	1000	β_1	-0.007	-0.006	-20.808	1166.576	0.000
krnjeni (0.05)	100	β_1	-0.007	-0.006	-4.347	117.188	0.000
	500	β_1	-0.007	-0.006	-22.196	584.111	0.000
	1000	β_1	-0.007	-0.006	-43.114	1177.654	0.000
krnjeni (0.3)	100	β_1	-0.007	-0.004	-19.725	141.460	0.000
	500	β_1	-0.007	-0.004	-93.556	677.134	0.000
	1000	β_1	-0.007	-0.004	-190.112	1369.416	0.000
krnjeni (0.5)	100	β_1	-0.007	-0.002	-35.508	161.331	0.000
	500	β_1	-0.007	-0.003	-168.070	742.773	0.000
	1000	β_1	-0.007	-0.003	-337.785	1584.587	0.000

Tabela 5.5: Parametri Buckley-Jamesovega modela pri časovno odvisnih parametrih Aalenovega modela, pri čemer je 50 % enot manjših od prelomne točke t^* (Model 1)

	n	Pravi	Ocenjeni	t	df	p-vrednost	
nekrnjeni	100	β_1	-0.007	-0.006	-4.548	114.742	0.000
		β_2	-0.003	-0.004	3.453	112.592	0.001
	500	β_1	-0.007	-0.006	-20.272	569.825	0.000
		β_2	-0.003	-0.004	19.068	565.799	0.000
	1000	β_1	-0.007	-0.007	-38.258	1149.521	0.000
		β_2	-0.003	-0.004	37.929	1122.101	0.000
krnjeni (0.05)	100	β_1	-0.007	-0.006	-6.902	114.308	0.000
		β_2	-0.003	-0.003	2.453	112.946	0.016
	500	β_1	-0.007	-0.006	-31.517	568.153	0.000
		β_2	-0.003	-0.004	13.459	576.728	0.000
	1000	β_1	-0.007	-0.006	-64.797	1146.041	0.000
		β_2	-0.003	-0.004	30.206	1134.393	0.000
krnjeni (0.3)	100	β_1	-0.007	-0.004	-16.383	122.276	0.000
		β_2	-0.003	-0.003	-0.133	116.521	0.895
	500	β_1	-0.007	-0.004	-80.106	608.410	0.000
		β_2	-0.003	-0.003	0.491	580.087	0.624
	1000	β_1	-0.007	-0.004	-145.113	1179.150	0.000
		β_2	-0.003	-0.003	-0.210	1167.218	0.834
krnjeni (0.5)	100	β_1	-0.007	-0.003	-24.395	124.884	0.000
		β_2	-0.003	-0.002	-2.848	119.070	0.005
	500	β_1	-0.007	-0.003	-116.657	629.675	0.000
		β_2	-0.003	-0.003	-12.608	629.204	0.000
	1000	β_1	-0.007	-0.003	-237.395	1279.800	0.000
		β_2	-0.003	-0.003	-27.016	1230.593	0.000

Tabela 5.6: Parametri Buckley-Jamesovega modela pri časovno odvisnih parametrih Aalenovega modela, pri čemer je 50 % enot manjših od prelomne točke t^* (Model 2)

	n	Pravi	Ocenjeni	t	df	p-vrednost
nekrnjeni	100	β_1	-0.007	-0.006	-4.380	120.008
	500	β_1	-0.007	-0.006	-19.584	596.100
	1000	β_1	-0.007	-0.007	-37.584	1199.827
krnjeni (0.05)	100	β_1	-0.007	-0.006	-6.687	120.296
	500	β_1	-0.007	-0.006	-30.594	589.741
	1000	β_1	-0.007	-0.006	-62.564	1206.162
krnjeni (0.3)	100	β_1	-0.007	-0.004	-16.104	129.826
	500	β_1	-0.007	-0.004	-77.171	640.058
	1000	β_1	-0.007	-0.004	-141.795	1250.972
krnjeni (0.5)	100	β_1	-0.007	-0.003	-23.811	135.744
	500	β_1	-0.007	-0.003	-113.582	676.626
	1000	β_1	-0.007	-0.003	-231.006	1360.830

Tabela 5.7: Parametri Buckley-Jamesovega modela pri časovno odvisnih parametrih Aalenovega modela, pri čemer je 80 % enot manjših od prelomne točke t^* (Model 1)

	n		Pravi	Ocenjeni	t	df	p-vrednost
nekrnjeni	100	β_1	-0.005	-0.008	8.999	104.395	0.000
		β_2	-0.003	-0.007	12.114	104.797	0.000
	500	β_1	-0.005	-0.008	40.484	534.478	0.000
		β_2	-0.003	-0.007	55.353	528.515	0.000
krnjeni (0.05)	1000	β_1	-0.005	-0.008	84.628	1076.791	0.000
		β_2	-0.003	-0.007	117.067	1066.464	0.000
	500	β_1	-0.005	-0.007	7.059	107.479	0.000
		β_2	-0.003	-0.006	11.035	106.457	0.000
krnjeni (0.3)	1000	β_1	-0.005	-0.007	32.195	535.579	0.000
		β_2	-0.003	-0.006	53.371	533.246	0.000
	500	β_1	-0.005	-0.007	66.944	1082.546	0.000
		β_2	-0.003	-0.006	103.743	1072.098	0.000
krnjeni (0.5)	100	β_1	-0.005	-0.005	1.420	108.418	0.158
		β_2	-0.003	-0.005	7.276	106.282	0.000
	500	β_1	-0.005	-0.006	8.496	546.436	0.000
		β_2	-0.003	-0.005	39.264	536.483	0.000
krnjeni (0.5)	1000	β_1	-0.005	-0.006	21.527	1106.144	0.000
		β_2	-0.003	-0.005	83.133	1094.919	0.000
	500	β_1	-0.005	-0.004	-4.251	112.966	0.000
		β_2	-0.003	-0.004	2.980	107.515	0.004
krnjeni (0.5)	1000	β_1	-0.005	-0.004	-20.424	564.087	0.000
		β_2	-0.003	-0.004	17.555	546.998	0.000
	500	β_1	-0.005	-0.004	-38.305	1128.372	0.000
		β_2	-0.003	-0.004	36.611	1095.524	0.000

Tabela 5.8: Parametri Buckley-Jamesovega modela pri časovno odvisnih parametrih Aalenovega modela, pri čemer je 80 % enot manjših od prelomne točke t^* (Model 2)

	n		Pravi	Ocenjeni	t	df	p-vrednost
nekrnjeni	100	β_1	-0.005	-0.008	8.761	106.856	0.000
	500	β_1	-0.005	-0.008	39.026	553.116	0.000
	1000	β_1	-0.005	-0.008	81.215	1109.338	0.000
krnjeni (0.05)	100	β_1	-0.005	-0.007	6.800	111.203	0.000
	500	β_1	-0.005	-0.007	31.341	554.063	0.000
	1000	β_1	-0.005	-0.007	64.180	1124.404	0.000
krnjeni (0.3)	100	β_1	-0.005	-0.005	1.216	113.141	0.227
	500	β_1	-0.005	-0.006	7.264	565.832	0.000
	1000	β_1	-0.005	-0.006	18.573	1156.358	0.000
krnjeni (0.5)	100	β_1	-0.005	-0.004	-4.628	119.656	0.000
	500	β_1	-0.005	-0.004	-20.558	594.501	0.000
	1000	β_1	-0.005	-0.004	-39.431	1172.705	0.000

6. Uporaba alternativnih regresijskih metod na dejanskih podatkih

Uporaba alternativnih regresijskih metod v relativnem preživetju omogoča pridobivanje novih informacij in osvetlitev problema relativnega preživetja iz novega zornega kota. V tem poglavju bomo predstavili rezultate Aalenovega linearne modela, Buckley-Jamesove linearne regresije in Coxovega modela sorazmernih ogroženosti na dejanskih podatkih pljučnega raka in raka na dojki. Namen poglavja je preučitev rezultatov različnih metod na praktičnih primerih v kontekstu relativnega preživetja in ugotoviti, na kakšen način lahko predstavimo pridobljene informacije. Izpostavili bomo tudi nekatere probleme modelov in podali predloge, kako lahko z njimiravnamo v praksi. Rezultati analiz so za vsak regresijski model prikazani v ločenih poglavjih, pri čemer so poglavja napisana tako, da jih lahko beremo kot samostojne celote. Zato na določenih mestih, predvsem pri interpretaciji rezultatov lahko prihaja do podvajanja informacij.

6.1 Opis podatkov

V nalogi so uporabljeni podatki o pljučnem raku in raku na dojki, pridobljeni iz Slovenskega registra raka. Z metodološkega vidika se podatki med seboj najbolj razlikujejo glede na odstotek krnjenih enot. Medtem ko je odstotek krnjenih enot pri podatkih o pljučnem raku slabih 5 %, je ta odstotek pri podatkih o raku na dojki skoraj 50 %.

Informacija o vzroku smrti je v študijah, kjer se paciente opazuje skozi daljše časovno obdobje, pogosto nezanesljiva ali celo nerazpoložljiva. To velja tudi za registrske po-

datke, kjer se je analiza preživetja tudi začela razvijati. Uporaba alternativnih metod relativnega preživetja na registrskih podatkih o raku je torej smiselna in upravičena. Neodvisne spremenljivke, uporabljene v obeh modelih, so spremenljivke, na podlagi katerih opazovani čas transformiramo v individualno mero relativnega preživetja oziroma v transformirani čas Y (starost, spol in leto diagnoze) in stadij, v katerem je bila bolezen odkrita. Za opredelitev stadijev, ki so v besedilu in na slikah predstavljeni s številkami 1, 2, 3, 4, smo uporabili poenostavljeno definicijo stadijev, ki upošteva vse preiskovalne metode, vključno z operacijo. Če pacient pred tem ni bil zdravljen, se za preiskovalno metodo upošteva tudi obdukcija. Štirje stadiji so opredeljeni na naslednji način:

- *omejeni stadij* predstavlja bolezen, pri kateri niso prizadete regionalne bezgavke in ni zasevkov v oddaljenih organih (v besedilu in na slikah je označen s številko 1);
- *stadij regionalne razširitve* predstavlja bolezen, pri kateri so prizadete tudi regionalne bezgavke, zasevkov v oddaljenih bezgavkah in organih pa ni (v besedilu in na slikah je označen s številko 2);
- *stadij oddaljene razširitve* je bolezen, pri kateri so zasevki že v oddaljenih bezgavkah ali organih (v besedilu in na slikah je označen s številko 3);
- v primeru ko je podatek o stadiju nerazpoložljiv ali neznan, je stadij v besedilu označen s številko 4.

Podatki o pljučnem raku

Podatki o pljučnem raku zajemajo paciente, ki so bili s pljučnim rakom diagnosticirani v obdobju med 1. januarjem 1983 in 29. decembrom 2004, pacienti pa so bili opazovani do 30. septembra 2007. V opazovanem obdobju je bilo s proučevanim rakom diagnosticiranih 17639 patientov. Samo 0,1 % patientov je v tem obdobju emigriral, kar pomeni, da skoraj vse paciente sledimo do smrti, oziroma je krnjenoje podatkov povezano s preživetjem patientov.

Neodvisne spremenljivke, ki smo jih uporabili za modeliranje transformiranega časa,

so spremenljivke, ki jih potrebujemo za transformacijo originalnega časa T (starost, spol in leto diagnoze) in stadij, v katerem je bila bolezen odkrita. V nadaljevanju podajamo opisne statistike proučevanih spremenljivk, njihova porazdelitev pa je prikazana na sliki 6.1.

- *STAROST*: starost ob diagnozi pljučnega raka (v letih)

Min	Q_1	Mediana	Q_3	Max	Povprečje	Std. odklon
12	56	64	71	98	63.9	10.6

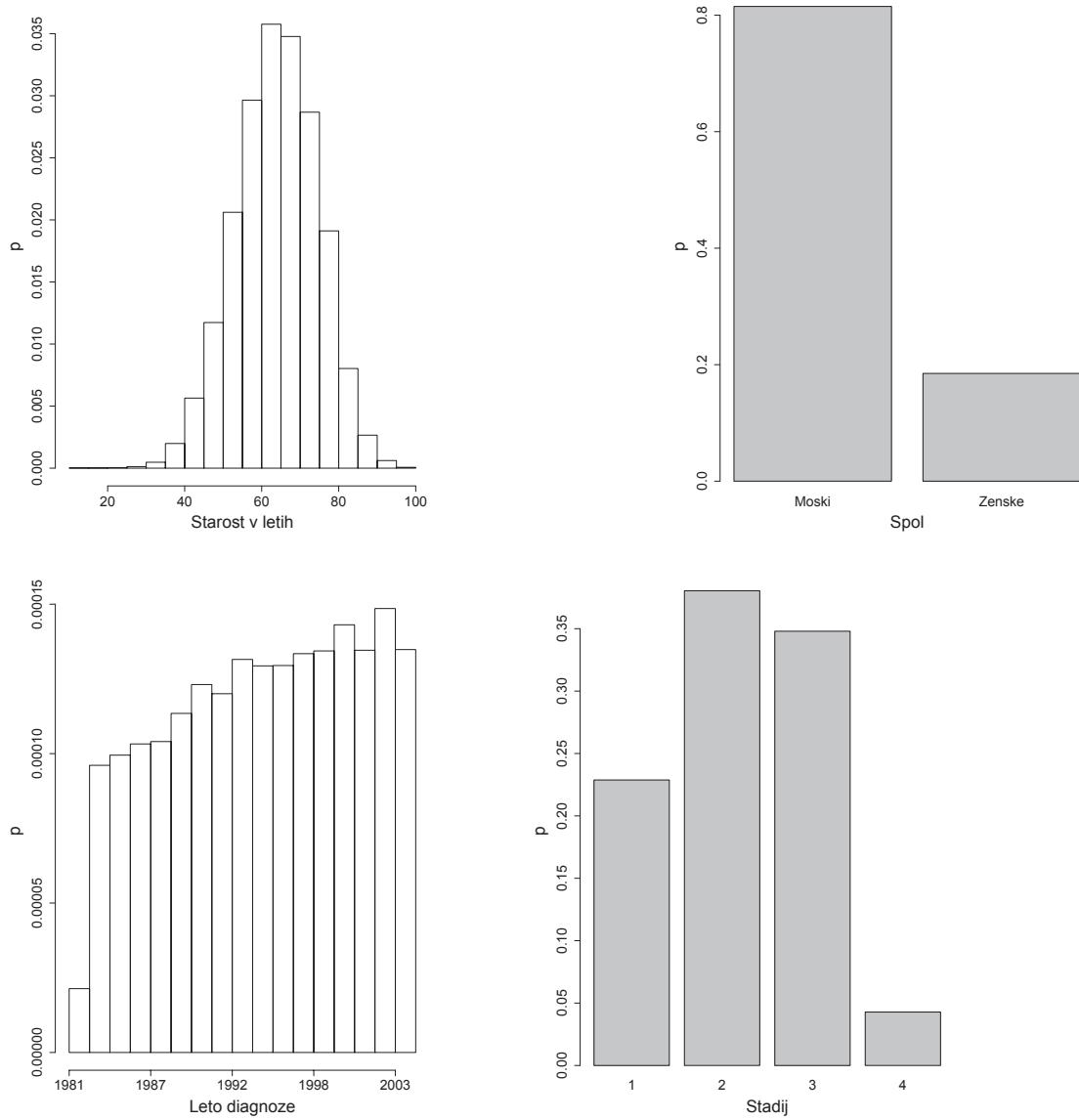
- *SPOL*: spol pacientov (približno 81 % moških)
- *LETO DIAGNOZE*: leto, v katerem so bili diagnosticirani pacienti s pljučnim rakom (med 1. 1. 1983 in 31. 12. 2004)

Min	Q_1	Mediana	Q_3	Max	Povprečje	Std. odklon
1Jan83	2Sep89	9Mar95	10Apr00	29Dec04	3Nov94	2289 dni

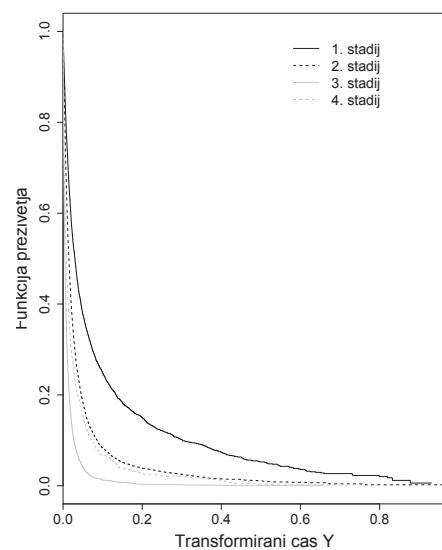
- *STADIJ*: stadij, v katerem je bila bolezen odkrita (samo četrtina pacientov je bila diagnosticirana v prvem stadiju pljučnega raka)

Porazdelitev transformiranega časa po stadijih je v obliki funkcij preživetja, ocenjenih po Kaplan-Meierjevi metodi, prikazana na sliki 6.2. Slika kaže, da se preživetje pacientov v višjih stadijih zmanjšuje, vidimo pa lahko tudi, da je preživetje pacientov, za katere podatek o stadiju ni bil razpoložljiv, oziroma je le ta neznan, bližje preživetju pacientov, diagnosticiranih v drugem stadiju pljučnega raka, čeprav bi intuitivno pričakovali, da bo njihovo preživetje bolj podobno preživetju pacientov, diagnosticiranih v tretjem stadiju bolezni.

Slika 6.1: Porazdelitev neodvisnih spremenljivk, uporabljenih pri modeliranju podatkov o pljučnem raku



Slika 6.2: Funkcije preživetja transformiranega časa, ocenjene po Kaplan-Meierjevi metodi



Podatki o raku na dojki

Rak na dojki je bolezen, ki v splošnem lahko prizadane oba spola, vendar se precej pogosteje pojavlja pri ženskah kot pri moških. V Sloveniji tako beležijo le okrog deset primerov raka na dojki pri moških. V nalogi smo se zato osredotočili samo na proučevanje raka na dojki pri ženskah.

Podatki zajemajo pacientke, ki so bile diagnosticirane z rakom na dojki v obdobju med 1. januarjem 1983 in 30. decembrom 2004. Pacientke so bile opazovane do 30. junija 2007. V opazovanem časovnem obdobju je bilo z rakom na dojki diagnosticiranih 15622 žensk, med katerimi jih je emigriralo približno 0,1 %. To pomeni, da so bile skoraj vse pacientke opazovane do smrti oziroma je krnjenoje podatkov povezano s preživetjem patientk.

Pri modeliranju transformiranega časa smo se osredotočili na starost in leto, v katerem so bile pacientke diagnosticirane z bolezni, torej na spremenljivki, ki ju potrebujemo za izračun transformiranega časa¹. Poleg tega smo, podobno kot pri pljučnem raku, za modeliranje uporabili tudi stadij, v katerem je bila bolezen odkrita, saj se preživetje patientk med stadiji zelo razlikuje. Porazdelitev proučevanih spremenljivk je prikazana na sliki 6.3, v nadaljevanju pa je predstavljenih še nekaj opisnih statistik.

- **STAROST:** Starost ob diagnozi raka na dojki (v letih)

Min	Q ₁	Mediana	Q ₃	Max	Povprečje	Std. odklon
21	49	60	70	99	60.2	13.6

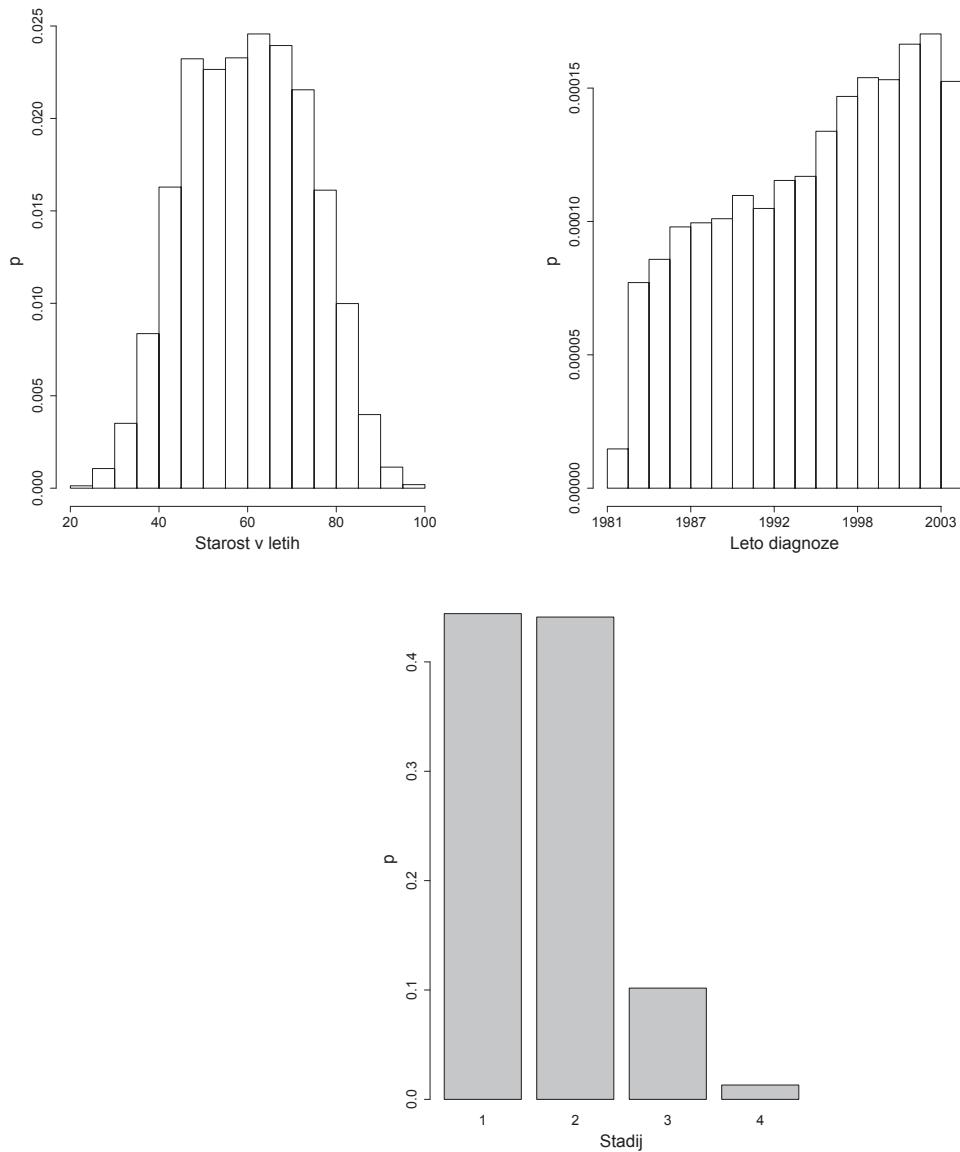
- **LETO DIAGNOZE:** leto, v katerem so bile diagnosticirane pacientke z rakom na dojki (med 1. 1. 1983 in 30. 12. 2004)

Min	Q ₁	Mediana	Q ₃	Max	Povprečje	Std. odklon
1Jan83	22May90	30May96	17Dec00	30Dec04	21Jul95	2272 dni

- **STADIJ:** stadij, v katerem je bila bolezen odkrita (skoraj 90 % patientk je bilo diagnosticiranih v prvih dveh stadijih, približno polovica med njimi v prvem stadiju)

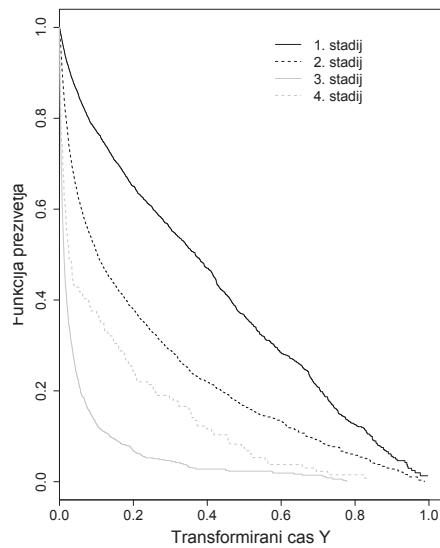
¹Ker so v analizi vključene le ženske patientke, za transformacijo originalnega časa v transformirani čas spremenljivka spola ni potrebna.

Slika 6.3: Porazdelitev neodvisnih spremenljivk, uporabljenih pri modeliranju podatkov o raku na dojki



Porazdelitev transformiranega časa po stadijih je prikazana v obliki funkcij preživetja, ki so ocenjene po Kaplan-Meierjevi metodi (slika 6.4). Preživetje se v višjih stadijih zmanjšuje. V primerjavi s podatki o pljučnem raku, kjer je bilo preživetje patientov, pri katerih je podatek o stadiju nerazpoložljiv oziroma neznan, bliže preživetju patientov v drugem stadiju, pa je preživetje patientk, diagnosticiranih z rakom na dojki, v neznanem stadiju (4. stadij) med drugim in tretjim stadijem.

Slika 6.4: Funkcije preživetja transformiranega časa, ocenjene po Kaplan-Meierjevi metodi



Na podlagi ocenjenih funkcij preživetja smo pri obeh proučevanih boleznih odkrili, da je preživetje pacientov z neznanim oziroma nerazpoložljivim podatkom o stadiju bolezni med preživetjem pacientov, diagnosticiranih v drugem in tretjem stadiju bolezni. Glede na to lahko sklepamo, da skupino teh pacientov sestavljajo pacienti z boleznjijo v različnih stadijih. Na podlagi predhodnih analiz smo ugotovili, da rezultati za podatke o neznanem stadiju ne podajajo vsebinsko smiselnih ugotovitev, zato smo jih iz nadaljnjih analiz izključili.

6.2 Uporaba Aalenovega modela v relativnem preživetju

Delovanje Aalenovega linearnega modela ogroženosti bomo preučili na podatkih o pljučnem raku in raku na dojki. Proučevani podatki se z metodološkega vidika analize preživetja med seboj najbolj razlikujejo glede na odstotek krnjenih enot. Prednost Aalenovega modela je predvsem diagnosticiranje sprememb učinkov neodvisnih spremenljivk na ogroženost v času. V tem smislu rezultati Aalenovega modela niso koristni le z vidika doprinosu novih informacij, temveč te rezultate lahko uporabimo tudi kot vhodne informacije pri oblikovanju modelov z drugimi metodami.

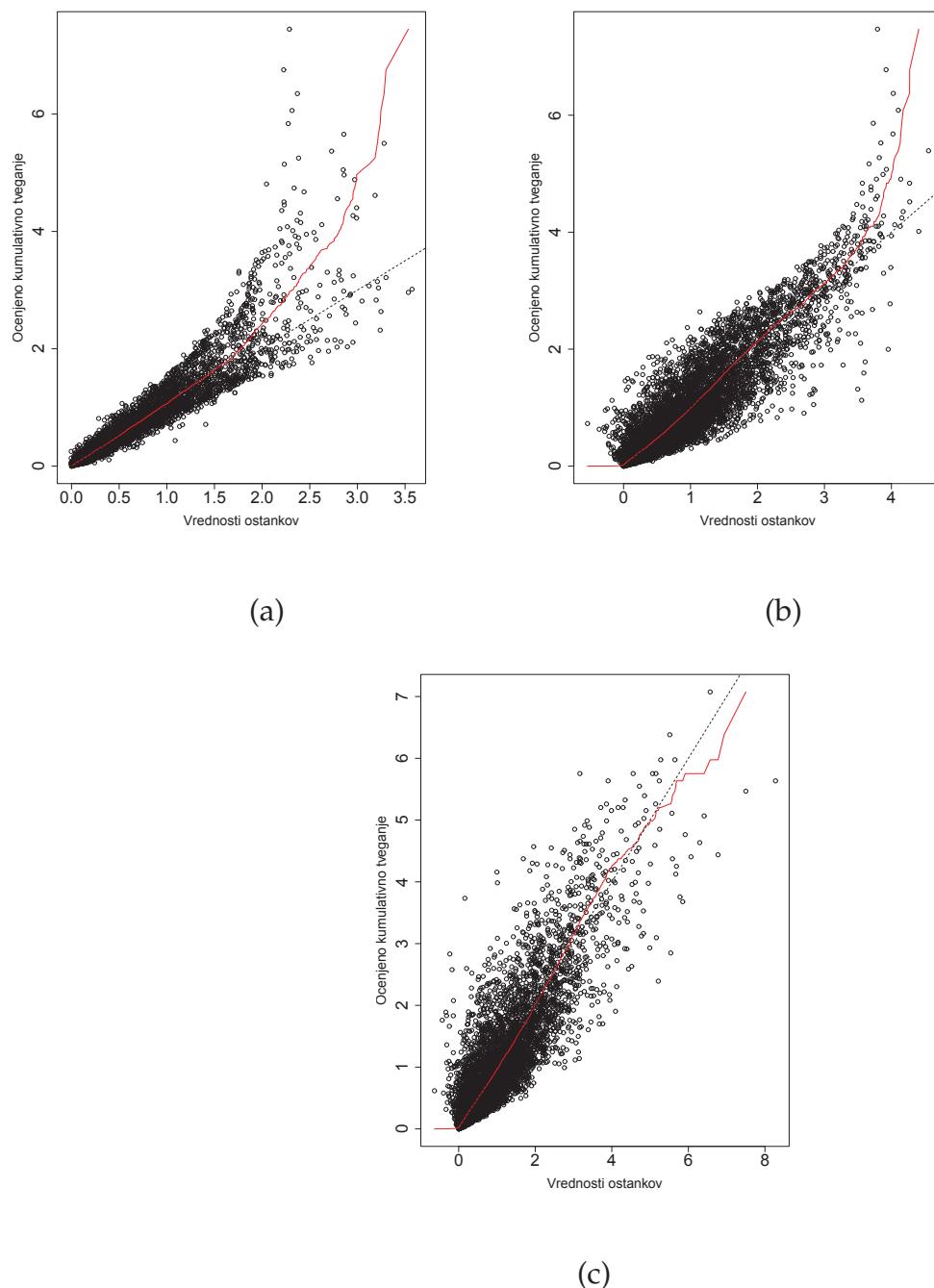
6.2.1 Aalenov model na podatkih o pljučnem raku

Pri proučevanju Aalenovega modela na podatkih o pljučnem raku smo v model vključili neodvisne spremenljivke: starost, spol, leto diagnoze in stadij, v katerem so bili pacienti diagnosticirani s pljučnim rakom. Predpostavljamo, da se vplivi starosti, spola in leta diagnoze razlikujejo glede na stadij, v katerem so bili bolniki diagnosticirani za boleznijo, zato so v analizo vključene tudi interakcije med tremi neodvisnimi spremenljivkami in stadijem pljučnega raka.

Pri ocenjevanju opisanega modela opazimo, da se ocenjevanje modela ustavi v času $Y \approx 0,387$, v katerem matrika $\mathbf{X}(y)$ izgubi polni rang. Vrednost transformiranega časa $Y \approx 0,387$ predstavlja zadnji dogodek pri moških, diagnosticiranih v tretjem stadiju pljučnega raka. Zaradi visoke umrljivosti patientov s pljučnim rakom pa končni čas ocenjenega modela ne vpliva bistveno na interpretacijo rezultatov.

Prileganje modela smo preverjali s pomočjo posplošenih ostankov (Aalen 1989), ki so za vsakega od treh stadijev prikazani na sliki 6.5. Na podlagi grafičnega prikaza posplošenih ostankov lahko rečemo, da se model zadovoljivo prilega analiziranim podatkom.

Slika 6.5: Grafični prikaz posplošenih ostankov za (a) 1. stadij pljučnega raka, (b) 2. stadij pljučnega raka in (c) 3. stadij pljučnega raka



V tabeli 6.1 so prikazane TST statistike (Aalen 1989), s katerimi preverjamo, ali imajo neodvisne spremenljivke statistično značilen vpliv na ogroženost:

$$H_0 : \beta_j(y) = 0, \quad \text{za vse } y \in [0, y_0]. \quad (6.1)$$

Za preverjanje zgornje ničelne domneve smo za opazovani časovni interval izbrali celotno obdobje, v katerem je bil izbrani model ocenjen ($[0, 0, 387]$). Na podlagi prikazanih rezultatov je v tretjem stadiju pljučnega raka vpliv spola in leta diagnoze na ogroženost neznačilen. Ostale spremenljivke imajo pri 5 % stopnji tveganja značilen vpliv na ogroženost. Na tem mestu velja opozoriti, da moramo biti previdni pri interpretaciji statistično značilnih vplivov na ogroženost, izračunanih na podlagi vrednosti TST statistik. Iz ocenjenih kumulativnih regresijskih funkcij, ki so prikazane na slikah 6.6, 6.7 in 6.8, namreč lahko razberemo, da se vpliv starosti in spola na ogroženost v času spremeni ob približno enaki vrednosti transformiranega časa ($\gamma \approx 0,03$). Izračun TST statistik na časovnem intervalu $[0, 0,03]$ pokaže na statistično značilen vpliv spola (tabela 6.1), kar je pomembno zlasti za interpretacijo rezultatov.

Učinki neodvisnih spremenljivk na ogroženost za smrt zaradi pljučnega raka so prikazani na slikah 6.6, 6.7 in 6.8. Iz njih lahko razberemo naslednje:

- Učinek starosti in spola na ogroženost se v času spremeni, sprememba pa se pri obeh spremenljivkah pojavi ob približno enaki vrednosti transformiranega časa ($\gamma \approx 0,03$).
 - Starost ima v začetnem obdobju ($\gamma < 0,03$) pozitiven učinek na preživetje, kasneje pa negativen. Pozitiven učinek starosti na preživetje odraža dejstvo, da je transformirani čas mlajših pacientov ob enakem dejanskem času preživetja krajši kot transformirani čas starejših pacientov. To pomeni, da je razlika med preživetjem mlajših bolnikov s pljučnim rakom in njihovo primerljivo splošno populacijo večja kot v primeru starejših bolnikov s pljučnim rakom.
- Sprememba pozitivnega učinka starosti na preživetje v negativnega pa pomeni, da po določenem obdobju preživetja pacientov ($\gamma \approx 0,03$), postane razlika v času preživetja med bolniki in primerljivo splošno populacijo večja pri starejših bolnikih. To pomeni uspenejše zdravljenje in/ali medicinsko oskrbo po času $\gamma \approx 0,03$ pri mlajših pacientih.

Tabela 6.1: TST testne statistike za Aalenov model na podatkih o pljučnem raku

	$y \in [0, 0.387]$		$y \in [0, 0.03]$	
	TST*	P	TST*	P
Konstanta	13.736	0.000	15.29	0.000
STAROST	-1.988	0.047	-8.975	0.000
SPOL=ženski	0.178	0.858	3.453	0.001
LETO	-9.396	0.000	-6.306	0.000
STADIJ=2	11.917	0.000	13.412	0.000
STADIJ=3	23.298	0.000	23.879	0.000
STAROST (STADIJ=2)	-15.122	0.000	-16.741	0.000
STAROST (STADIJ=3)	-24.21	0.000	-24.481	0.000
SPOL=ženski (STADIJ=2)	8.159	0.000	8.622	0.000
SPOL=ženski (STADIJ=3)	8.22	0.000	9.036	0.000
LETO (STADIJ=2)	3.737	0.000	3.315	0.001
LETO (STADIJ=3)	1.786	0.074	0.894	0.371

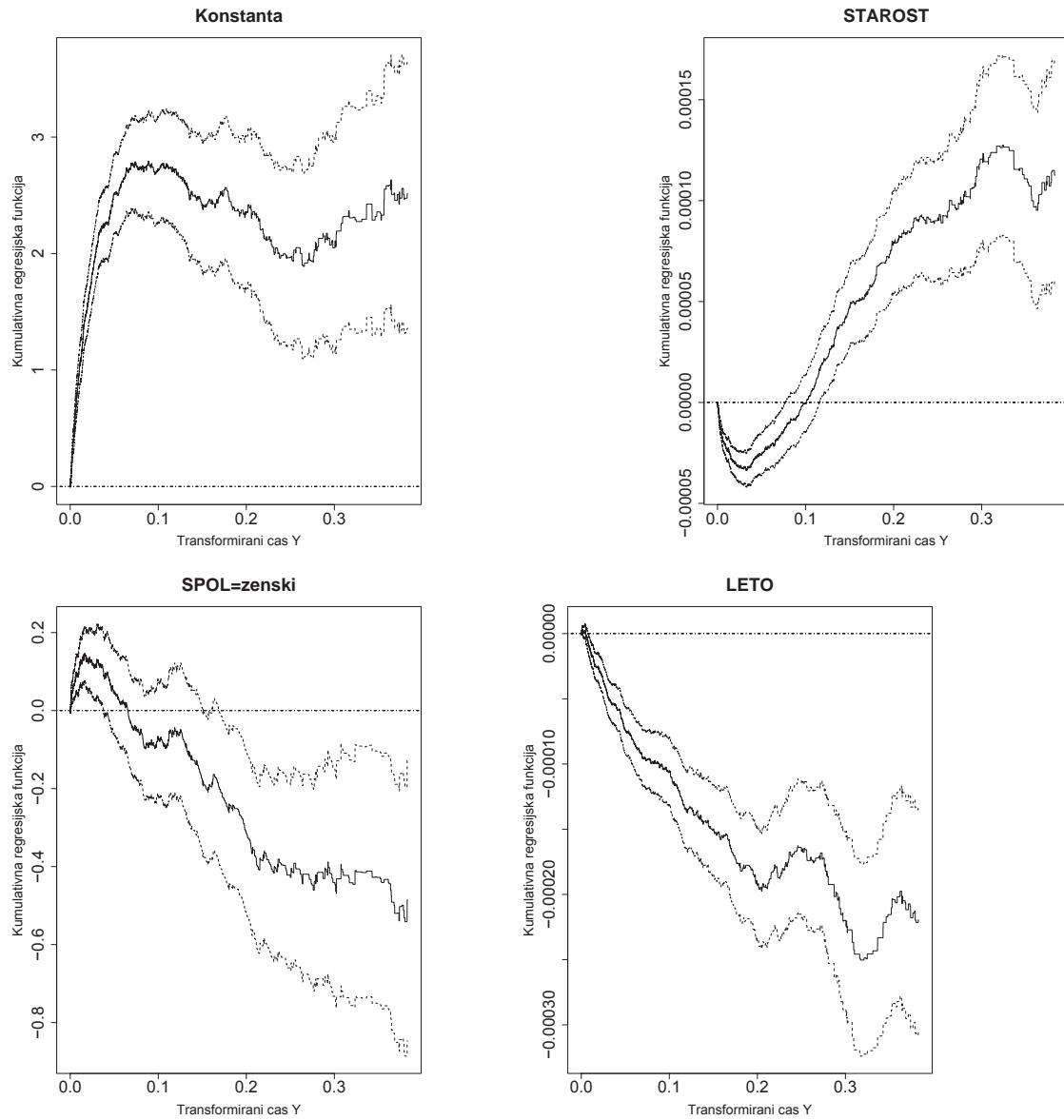
*Ob veljavni ničelni domnevi so TST testne statistike porazdeljene po standardizirani normalni porazdelitvi.

- Podobno kot pri starosti do spremembe učinka pride tudi pri spolu, pri čemer sprememba nastopi ob približno enakem času ($Y < 0,03$). V začetnem obdobju ($Y < 0,03$) je ogroženost za smrt za pljučnim rakom večja za ženske, kasneje pa za moške. Čeprav je preživetje žensk v splošni populaciji večje od preživetja moških, se ta prednost v začetnem obdobju zaradi bolezni izgublja. Preživetje žensk s spremembou negativnega učinka spola v pozitivnega postane večje od preživetja moških.
- Vpliv leta diagnoze na ogroženost je skozi celotno ocenjeno obdobje negativen, kar pomeni, da se preživetje pacientov, diagnosticiranih s pljučnim rakom v času zaradi izboljšanega zdravljenja in/ali medicinske oskrbe stalno izboljšuje.
- Po pričakovanjih se ogroženost za smrt za pljučnim rakom v višjih stadijih bistveno poveča.

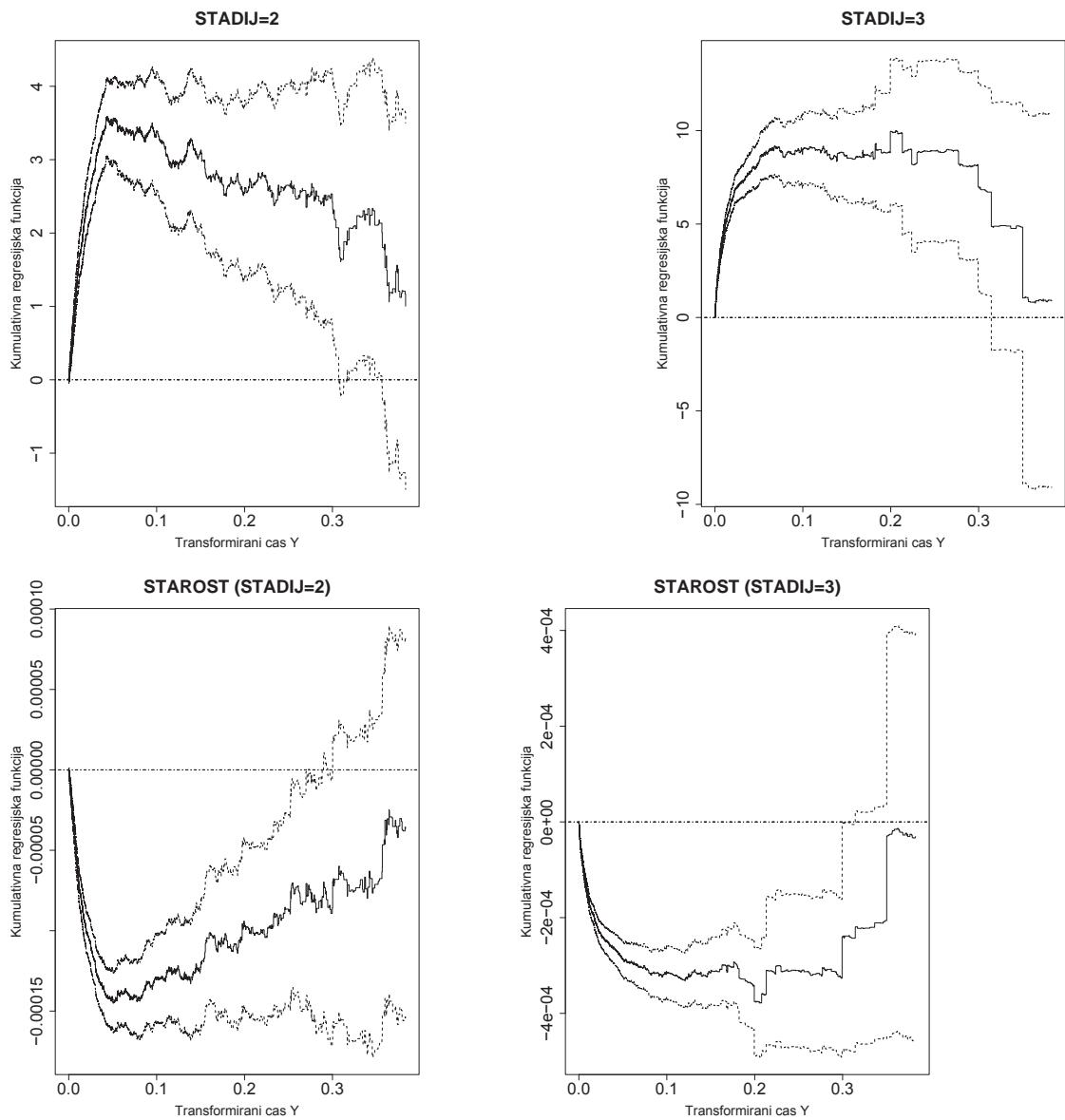
- Interakcije starosti, spola in leta diagnoze s stadijem pokažejo, da se glavni vplivi starosti in spola v višjih stadijih, zlasti v začetnem obdobju, poudarijo, medtem ko ni opaziti bistvenih sprememb v spreminjanju teh učinkov v času v primerjavi z glavnimi vplivi. Na drugi strani pri interakciji leta diagnoze s stadijem po drugi strani lahko opazimo, da se pozitivni vpliv leta diagnoze na preživetje v višjih stadijih zmanjšuje oziroma izginja. Sklepamo lahko, da je zdravljenje in/ali medicinska oskrba pacientov, diagnosticiranih v višjih stadijih pljučnega raka bistveno manj uspešna kot pri pacientih, diagnosticiranih v prvem stadiju pljučnega raka.

Na tem mestu naj spomnimo, da modeliranje ogroženosti s transformiranim časom v primerjavi z ogroženostjo splošne populacije dejansko predstavlja modeliranje presežne ogroženosti, ki je posledica bolezni proučevanih pacientov. Del presežne ogroženosti se odrazi na spremenljivkah, na podlagi katerih je izračunan transformirani čas Y . Zgoraj prikazani model podrobno opiše obnašanje glavnih vplivov, kot tudi interakcij starosti, spola in leta diagnoze s stadijem pljučnega raka. Ugotovimo lahko, da se del presežne ogroženosti v višjih stadijih pljučnega raka odrazi na proučevanih spremenljivkah. Prednost grafičnega prikaza učinkov kovariat na ogroženost je v opazovanju sprememb teh učinkov v času, kar se odraža na gibanju ocenjenih kumulativnih regresijskih funkcij, medtem ko kvantitativne učinke na ogroženost precej težje ovrednotimo. Doprinos dodatnih informacij o interakcijah demografskih spremenljivk s stadijem je zato majhen in je primernejša izbira enostavnnejšega modela. V nadaljevanju je prikazan model, ki vključuje le glavne vplive proučevanih spremenljivk (starosti, spola, leta diagnoze in stadija v katerem je bil pacient s pljučnim rakom diagnostican).

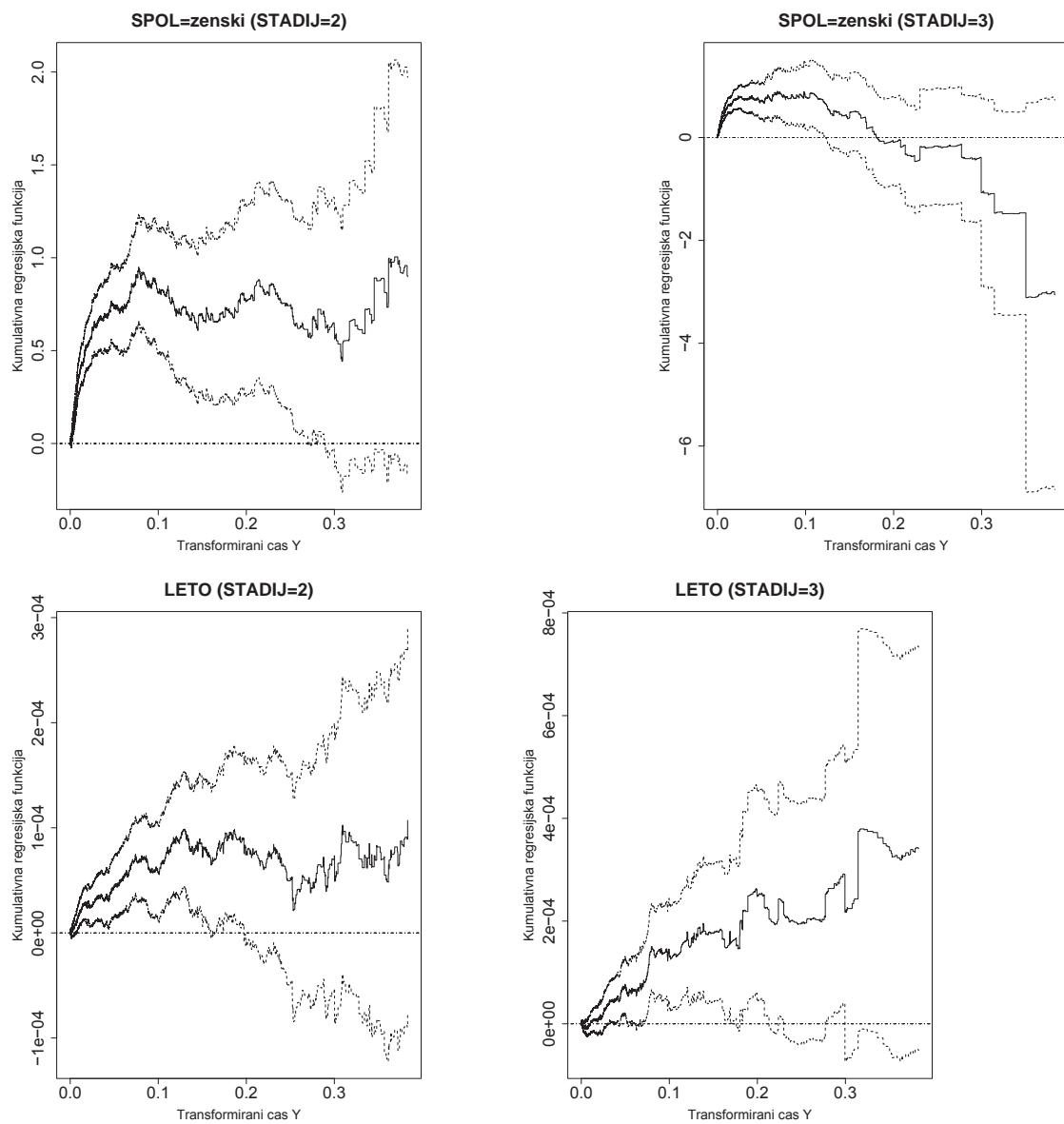
Slika 6.6: Ocenjene kumulativne regresijske funkcije s 95-odstotnim intervalom zaupanja za podatke o pljučnem raku



Slika 6.7: Nadaljevanje slike 6.6



Slika 6.8: Nadaljevanje slike 6.6



V enostavnem modelu se obdobje v katerem model še lahko ocenimo (do časa, ko ima matrika $\mathbf{X}(y)$ še polni rang), podaljša z $Y \approx 0,387$ na $Y \approx 0,660$, kar predstavlja tudi čas zadnje smrti v tretjem stadiju pljučnega raka.

Posplošeni ostanki enostavnega modela za vsakega od treh proučevanih stadijev pljučnega raka so prikazani na sliki 6.9, iz katere lahko razberemo, da je prileganje enostavnega modela v vseh treh stadijih zadovoljivo. TST statistike, izračunane na celotnem ocenjenem obdobju opazovanega transformiranega časa Y ($[0,0,657]$), kažejo na statistično značilen vpliv vseh neodvisnih spremenljivk, vključenih v model (tabela 6.2).

Tabela 6.2: TST testne statistike za Aalenov enostavni model na podatkih o pljučnem raku

	TST*	P
Konstanta	38.99	0.000
STAROST	-35.282	0.000
SPOL=ženski	14.451	0.000
LETO	-6.151	0.000
STADIJ=2	23.28	0.000
STADIJ=3	50.27	0.000

*Ob veljavni ničelni domnevi so TST testne statistike porazdeljene po standardizirani normalni porazdelitvi.

Vplivi nedovisnih spremenljivk na ogroženost za smrt za pljučnim rakom, ki so prikazani na slikah 6.10 in 6.11, so podobni kot glavni vplivi v modelu z interakcijami, zato je tudi njihova interpretacija podobna.

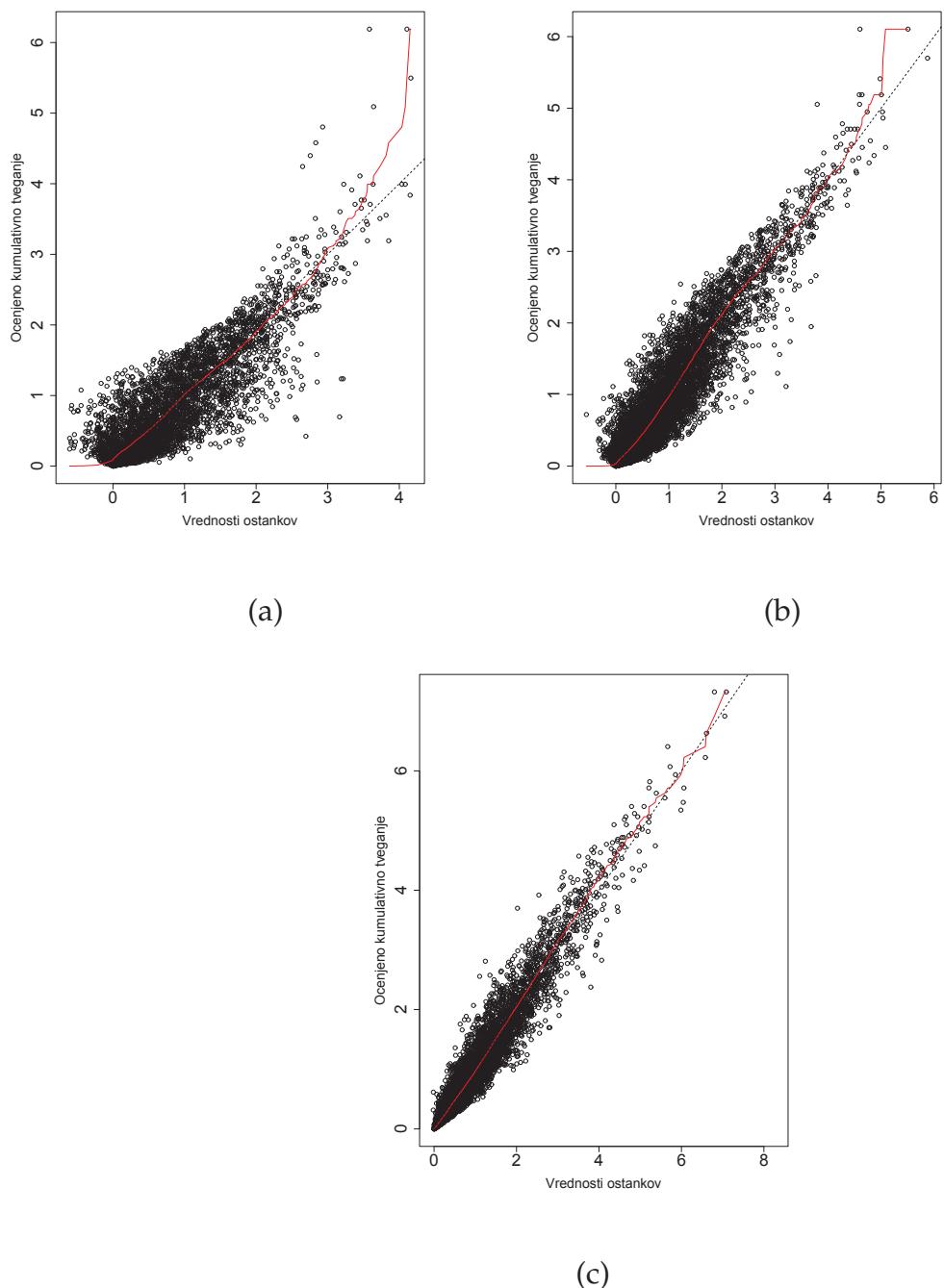
- Vpliv starosti na ogroženost je v začetnem obdobju negativen, kasneje pa postane pozitiven. Sprememba učinka nastopi okoli vrednosti transformiranega časa $Y \approx 0,03$.
- Podobno kot pri starosti tudi pri spolu v času $Y \approx 0,03$ pride do spremembe učinka. V začetnem obdobju je preživetje žensk slabše, kasneje pa boljše od preživetja moških.

- Negativen vpliv leta diagnoze na ogroženost za smrt za pljučnim rakom je prisoten v celotnem ocenjenem opazovanem obdobju.
- Preživetje pacientov se v višjih stadijih zmanjšuje.

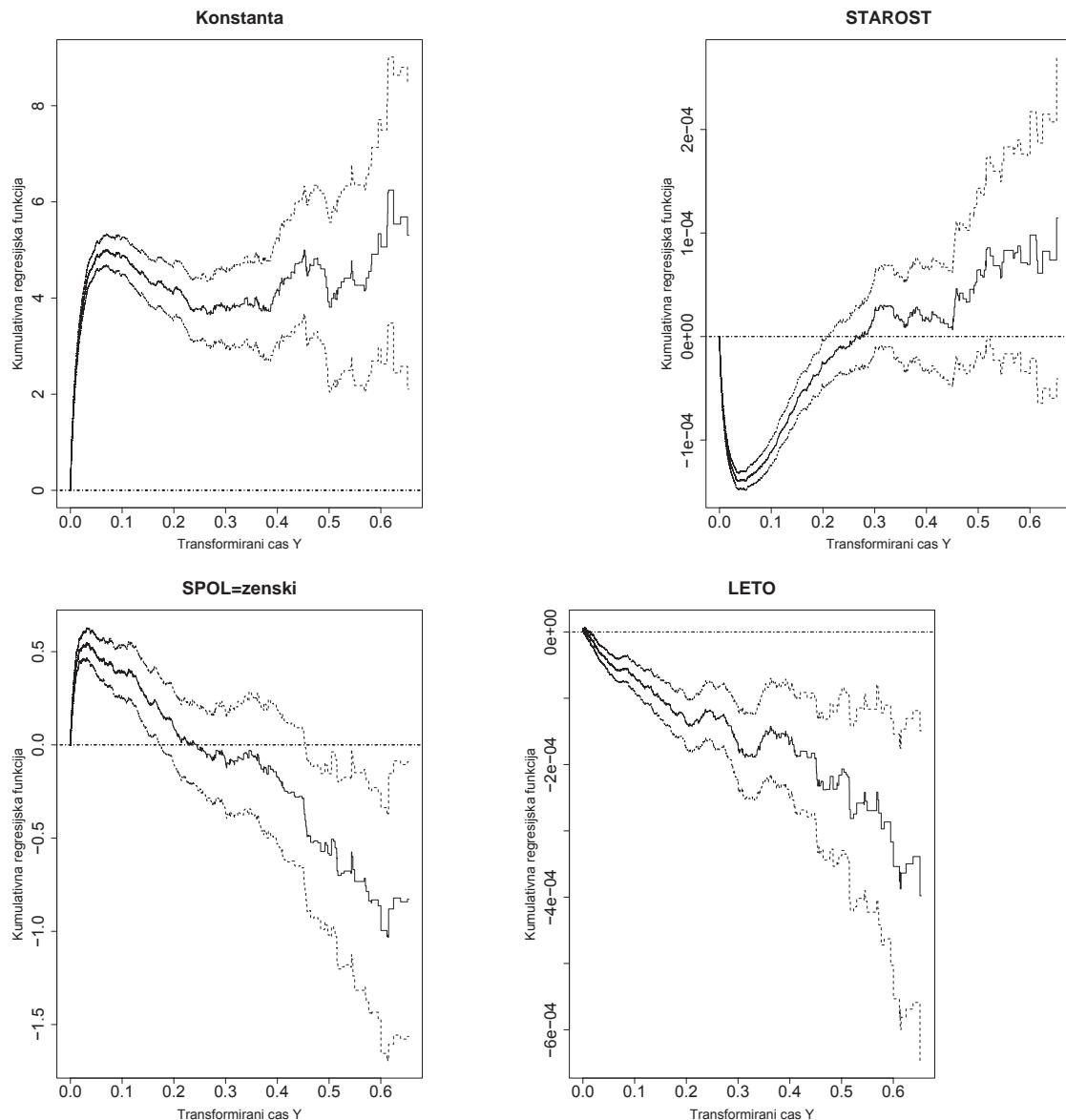
Informacije, ki jih pridobimo na podlagi enostavnega modela, so torej primerljive s tistimi, ki jih dobimo iz kompleksnejšega modela. Izjemo predstavlja leto, v katerem so bili pacienti s pljučnim rakom diagnosticirani, kjer nam model z interakcijami posreduje informacijo, da se preživetje pacientov, diagnosticiranih v prvem stadiju pljučnega raka izboljšuje, medtem ko je izboljšanje preživetja pacientov, diagnosticiranih v višjih stadijih, bistveno manjše. Kakšen vpliv ima leto diagnoze na ogroženost oziroma preživetje pacientov v različnih stadijih, je s pomočjo ocenjenih regresijskih funkcij kvantitativno težko ovrednotiti zaradi različnih standardnih napak v različnih stadijih. Ena od možnosti je uporaba na podlagi modela ocenjenih funkcij preživetja, ki so opisane v nadaljevanju.

Enostavni model nam pokaže tudi, kako se ogroženost pri glavnem vplivu drugega in tretjega stadija v primerjavi s kompleksnejšim modelom spremeni zaradi prenosa dela presežne ogroženosti na spremenljivke, na podlagi katerih računamo transformirani čas (starost, spol in leto diagnoze).

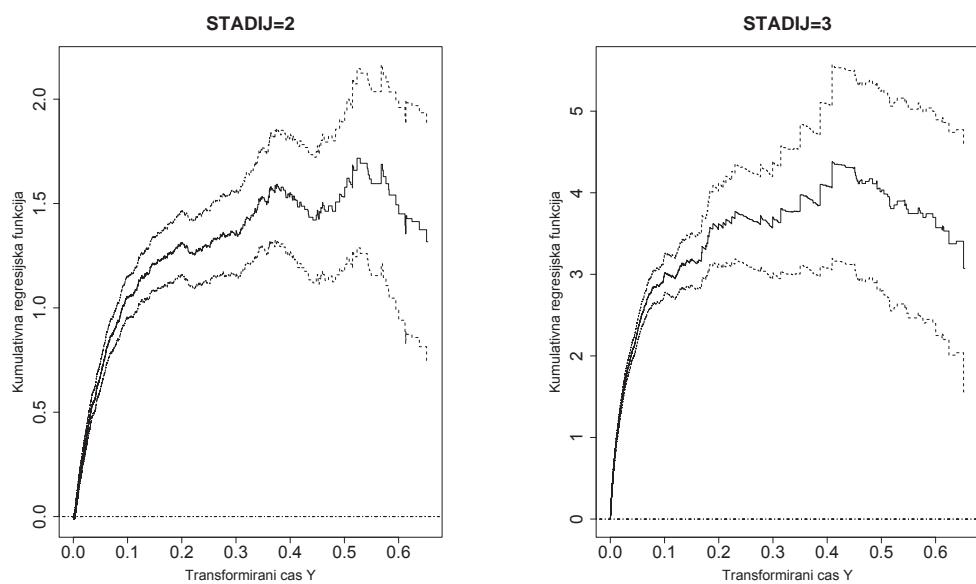
Slika 6.9: Grafični prikaz posplošenih ostankov v enostavnem modelu za (a) 1. stadij pljučnega raka, (b) 2. stadij pljučnega raka in (c) 3. stadij pljučnega raka



Slika 6.10: Ocjenjene kumulativne regresijske funkcije s 95-odstotnim intervalom zaupanja za podatke pljučnega raka - enostavni model



Slika 6.11: Nadaljevanje slike 6.10



Pri spremembi učinkov nedovisnih spremenljivk na ogroženost in ob predpostavki, da se ocenjeni model dobro prilega podatkom, se postavi vprašanje, ali so spremembe učinkov v času dejansko prisotne, ali so le posledica slučajnih vplivov. Formalnih testov, s katerimi bi lahko preverili značilnost sprememb učinkov v času, v literaturi ni zaslediti. Aalen namesto tega predлага bootstrap analizo (Aalen 1993), pri čemer kumulativne regresijske funkcije ocenimo na vsaki bootstrap ponovitvi. Informacijo o statistični negotovosti postopka ocenjevanja dobimo na podlagi variabilnosti med ocenjenimi kumulativnimi regresijskimi funkcijami. Bootstrap analiza predstavlja alternativo točkastim (pointwise) standardnim napakam in mejam zaupanja (confidence bands) (Huffer in McKeague 1991b), ki so bile razvite za grafični prikaz kumulativnih regresijskih funkcij. Za razliko od omenjenih mer, ki so precej grobe, bootstrap analiza ponuja podrobnejšo informacijo o vrsti funkcijskih oblik, ki bi lahko odstopale od ocenjene.

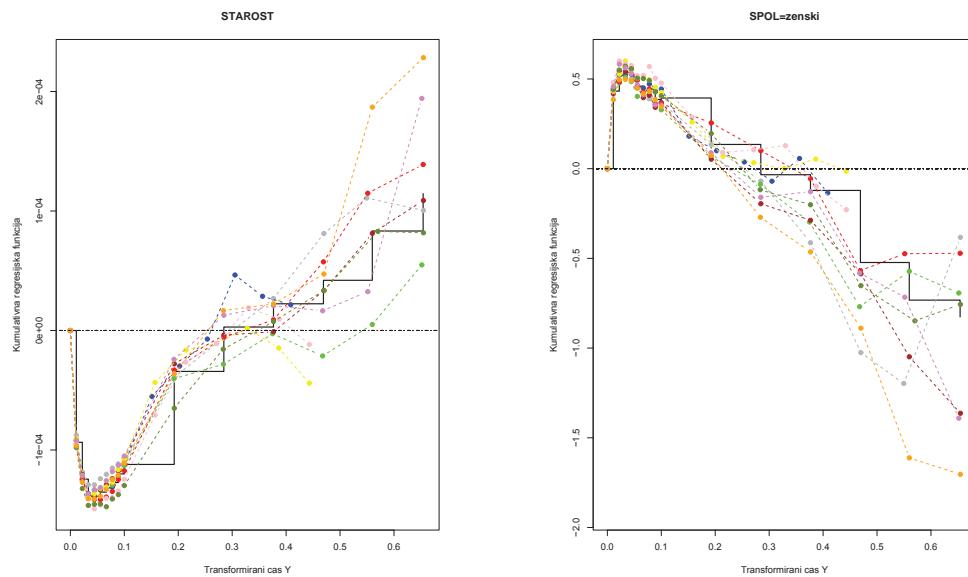
Pri izvedbi bootstrap analize smo uporabili enostavno bootstrap tehniko. Z vzorčenjem s ponavljanjem smo iz originalnega vzorca dane velikosti v bootstrap vzorcu izbrali enako število enot. Čeprav v analizi preživetja obstajajo tudi drugačne tehnike bootstrap vzorčenja, so te precej bolj odvisne od veljavnosti modela in zahtevnejše za izvedbo. Nasprotno je predstavljeni bootstrap enostaven in robusten obenem.

Aalen za lažje spremjanje konsistentnih značilnosti kumulativnih regresijskih funkcij, ocenjenih na originalnem vzorcu in bootstrap vzorcih, predлага, da funkcije prikažemo le na izbranih časovnih točkah (Aalen 1993). Takšno predstavitev lahko poimenujemo diskretna predstavitev in predstavlja zglajeno različico originalnih regresijskih funkcij.

Na primeru pljučnega raka smo z Aalenovim modelom odkrili, da se učinek starosti in spola pacientov v času spremeni. Ali je sprememba učinka za omenjeni spremenljivki značilna, smo preverili z zgoraj opisano bootstrap analizo, pri čemer smo zaradi velikega vzorca izvedli deset bootstrap ponovitev. Za neformalno presojo o konsistentnosti učinkov tudi Aalen predлага enako število ponovitev (Aalen 1993). Rezultati analize za enostavni model so prikazani na sliki 6.12, ki kaže, da učinek starosti in spola v času spremeni predznak v vseh izvedenih bootstrap ponovitvah, na tej podlagi pa lahko z veliko gotovostjo sklepamo, da se učinek starosti in spola na ogroženost v

času dejansko spremeni.

Slika 6.12: Ocenjene kumulativne regresijske funkcije enostavnega modela v diskretni obliki originalnega vzorca (črna okrepljena črta) in desetih bootstrap vzorcev (barvne črte)



Na tem mestu velja opozoriti na dodatno informacijo, ki jo dobimo z bootstrap analizo in je na sliki 6.12 na prvi pogled težje opazna. Čas, ko matrika $\mathbf{X}(y)$ izgubi polni rang je pri različnih bootstrap vzorcih različen. Na primeru desetih bootstrap vzorcev je matrika $\mathbf{X}(y)$ rang izgubila v naslednjih časih: 0.410, 0.448 (2 krat), 0.660 (3 krat), 0.668 (2 krat) in 0.673 (2 krat). Različni časi, v katerih se ocenjevanje modela ustavi, opozarjajo na dejstvo, da so ocenjene kumulativne regresijske funkcije proti koncu ocenjevalnega obdobja veliko bolj podvržene slučajnim vplivom, kar pomeni, da moramo biti pri interpretaciji rezultatov bolj previdni.

Vpliv kovariat na relativno preživetje pacientov s pljučnim rakom (Adjusted survival curves)

Primerjava modela z interakcijami starosti, spola in leta diagnoze s stadijem ter enostavnega modela bo v nadaljevanju predstavljena s funkcijami preživetja, ki jih izračunamo na podlagi ocenjenih kumulativnih regresijskih funkcij. Zanima nas predvsem vpliv starosti, spola in leta diagnoze na relativno preživetje pacientov, diagnosticiranih v posameznem stadiju pljučnega raka. Za spremenjajoče vrednosti starosti smo izbrali vrednosti 50, 60 in 70 let, za spremenjajoče vrednosti leta diagnoze pa 1. januar 1983, 1. januar 1993 in 1. januar 2003. Za fiksne vrednosti zveznih spremenljivk (starost in leto diagnoze) so izbrane njihove mediane, medtem ko smo za fiksno vrednost spola izbrali moške, ki predstavljajo približno 82 % celotnega vzorca. Pri spremenjajoči starosti smo za fiksno vrednost leta diagnoze izbrali 8. marec 1995, pri spremenjajočem letu diagnoze pa smo za fiksno starost izbrali vrednost 64 let.

Grafični prikaz vplivov kovariat na ocenjeno relativno preživetje pacientov s pljučnim rakom je prikazano na slikah 6.13-6.21, ki kažejo da se različna modela (model z interakcijami in enostavni model) najbolj razlikujeta v prvem in drugem stadiju. Razlike so vidne v legi ocenjenih krivulj preživetja, najbolj opazne pa so pri starosti in spolu, za katera smo ugotovili, da se njun vpliv na ogroženost v času spremeni.

Na podlagi grafičnih prikazov lahko razberemo, da ima starost v začetnem obdobju pozitiven vpliv na relativno preživetje, kar pomeni, da se s staranjem relativno preživetje pacientov izboljšuje. V prvem stadiju (slika 6.13) vpliv starosti po določenem času spremeni predznak, kar pomeni, da se s staranjem preživetje pacientov zmanjšuje. Čas, v katerem mlajši pacienti z vidika relativnega preživetja pridobijo prednost pred starejšimi pacienti, pa je z vidika modela z interakcijami in enostavnega modela različen.

Tudi v drugem in tretjem stadiju je v začetnem obdobju viden pozitivni vpliv starosti na relativno preživetje pacientov, s časom pa se ta vpliv v obeh stadijih izgubi (sliki 6.14 in 6.15).

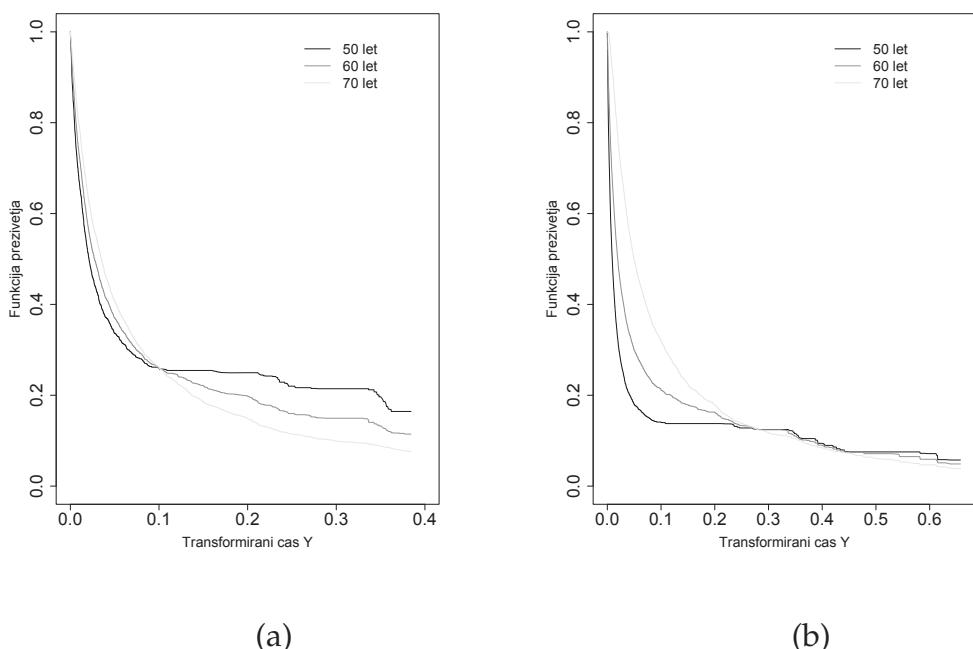
Pozitivni vpliv leta diagnoze je viden zlasti v prvem stadiju (slika 6.16), torej se preživetje

pacientov, diagnosticiranih v prvem stadiju pljučnega raka, zaradi izboljšanega zdravljenja in/ali medicinske oskrbe v času izboljšuje. V drugem stadiju je vpliv leta diagnoze zelo majhen (slika 6.17), medtem ko leto diagnoze v tretjem stadiju ne vpliva več na preživetje pacientov (slika 6.18).

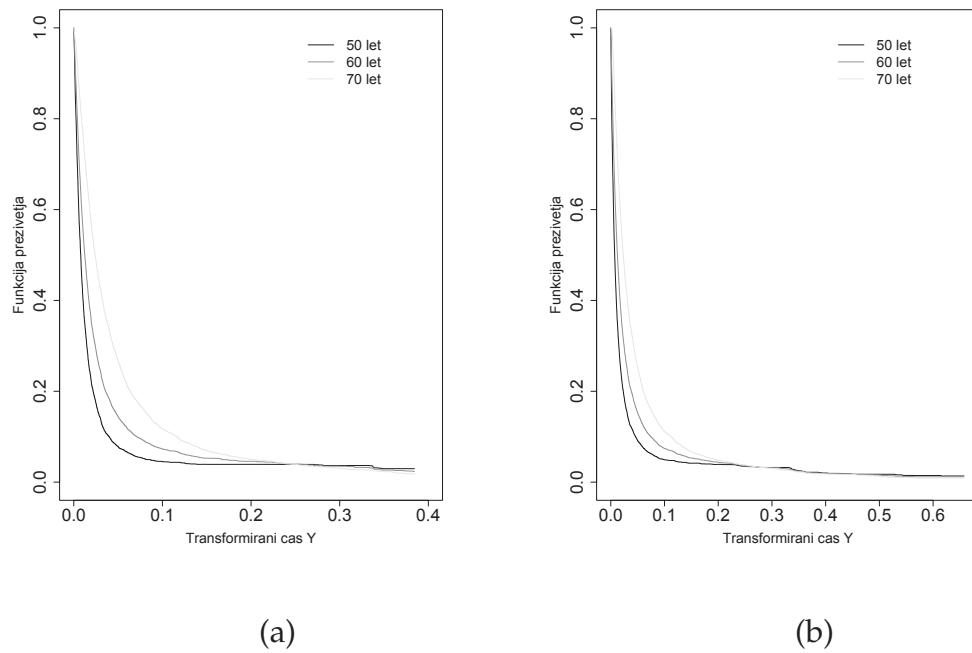
Primerjava relativnega preživetja po spolu pacientov pokaže na spremembo učinka spola v času. Najbolj je to opazno v prvem stadiju (slika 6.19), kjer je preživetje žensk v začetnem obdobju manjše od preživetja moških, v kasnejšem obdobju pa ženske z vidika preživetja pridobijo prednost pred moškimi. Čas, v katerem se relativno preživetje žensk v primerjavi s preživetjem moških izboljša, je podobno kot pri starosti v modelu z interakcijami in v enostavnem modelu različen, modela pa se razlikujeta tudi v razliki med ocenjenima krivuljama relativnega preživetja moških in žensk.

V drugem in tretjem stadiju podobno lahko opazimo, da je relativno preživetje žensk v začetnem obdobju manjše od preživetja moških, razlika med njima pa se v času izgublja (sliki 6.20 in 6.21).

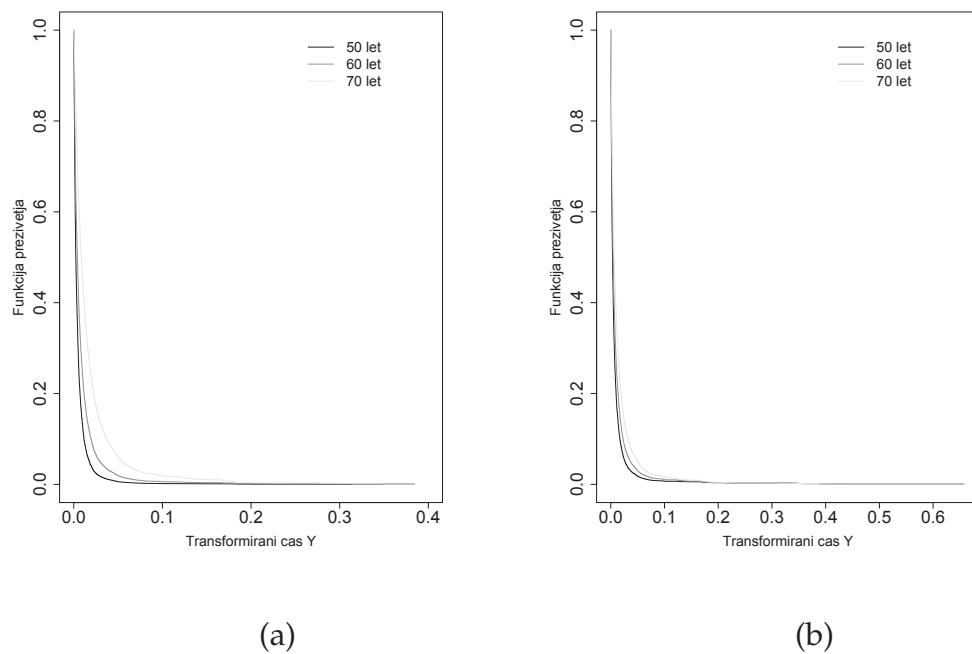
Slika 6.13: Na podlagi modelov ocenjene funkcije preživetja moških pacientov, diagnosticiranih v 1. stadiju pljučnega raka, za spremenljajoče vrednosti starosti v letu 1995 za (a) model z interakcijami in (b) enostavni model



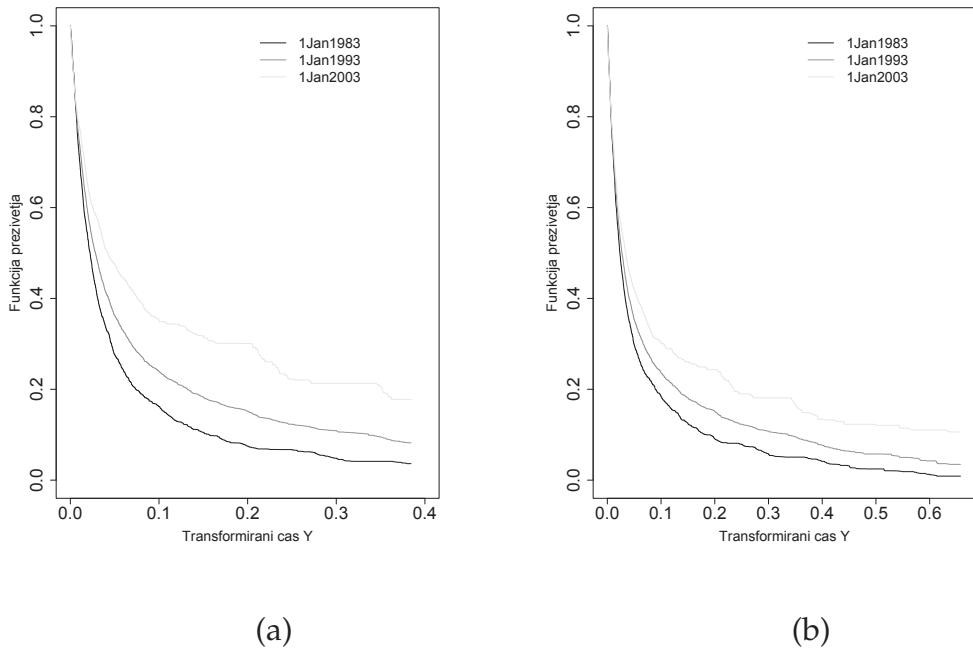
Slika 6.14: Na podlagi modelov ocenjene funkcije preživetja moških pacientov, diagnosticiranih v 2. stadiju pljučnega raka, za spremenljajoče vrednosti starosti v letu 1995 za (a) model z interakcijami in (b) enostavni model



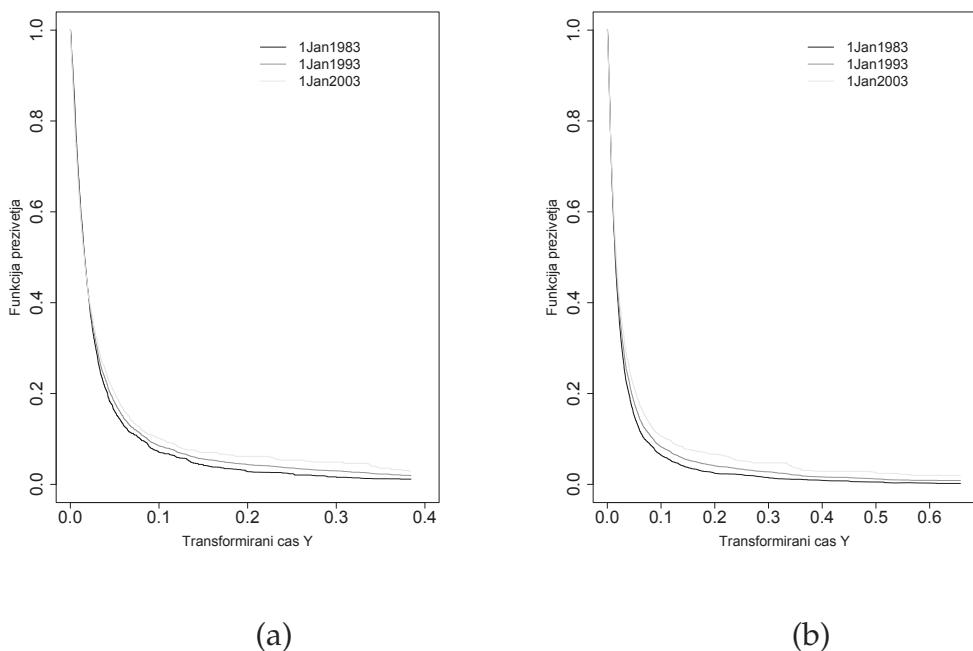
Slika 6.15: Na podlagi modelov ocenjene funkcije preživetja moških pacientov, diagnosticiranih v 3. stadiju pljučnega raka, za spremenljajoče vrednosti starosti v letu 1995 za (a) model z interakcijami in (b) enostavni model



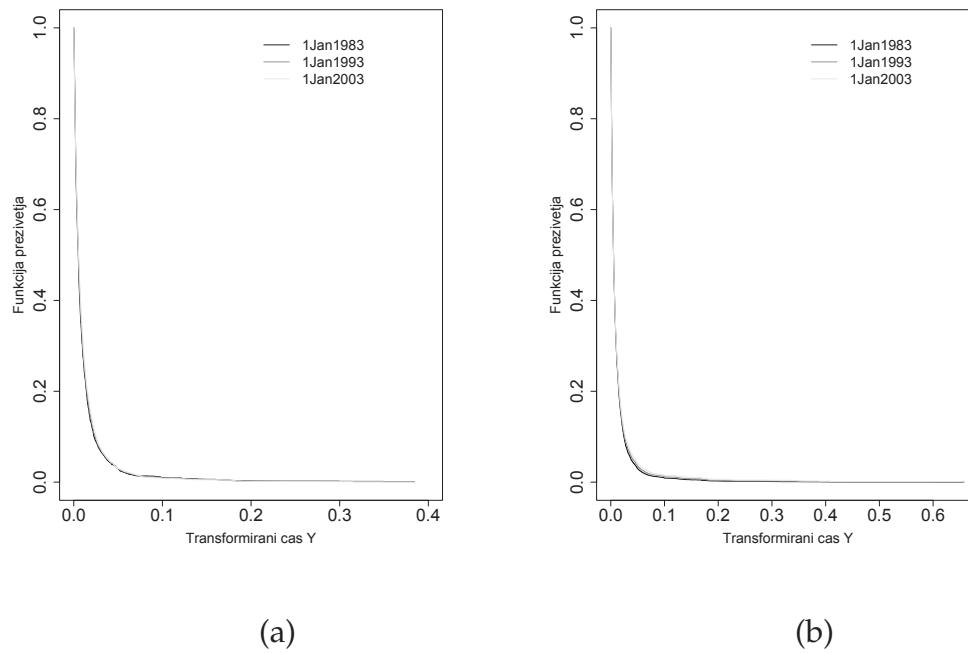
Slika 6.16: Na podlagi modelov ocenjene funkcije preživetja moških pacientov, diagnosticiranih v 1. stadiju pljučnega raka, za spremenljajoče vrednosti leta diagnoze pri starosti 64 let za (a) model z interakcijami in (b) enostavni model



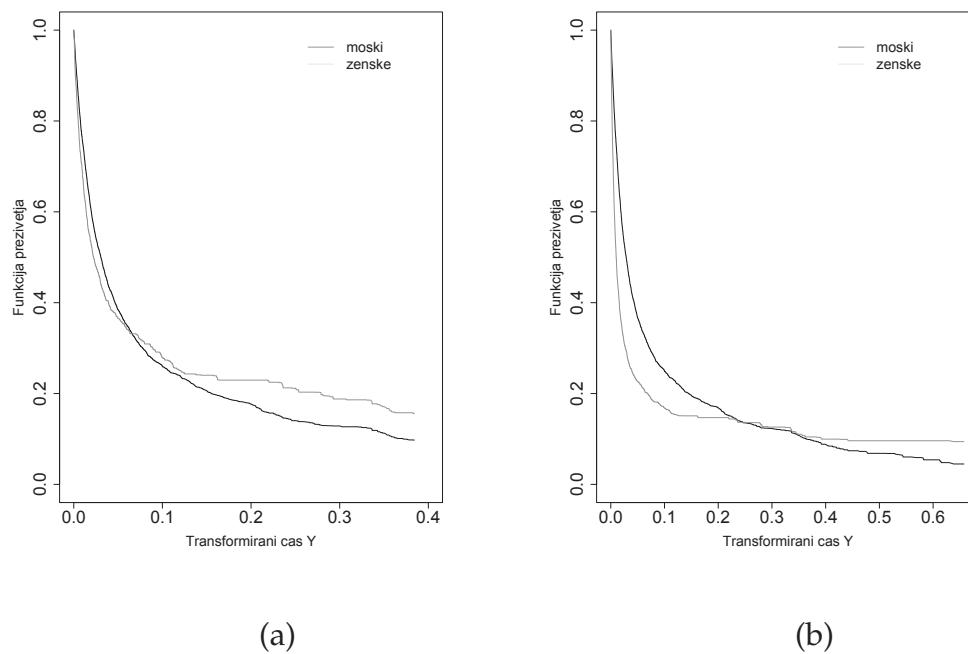
Slika 6.17: Na podlagi modelov ocenjene funkcije preživetja moških pacientov, diagnosticiranih v 2. stadiju pljučnega raka, za spremenljajoče vrednosti leta diagnoze pri starosti 64 let za (a) model z interakcijami in (b) enostavni model



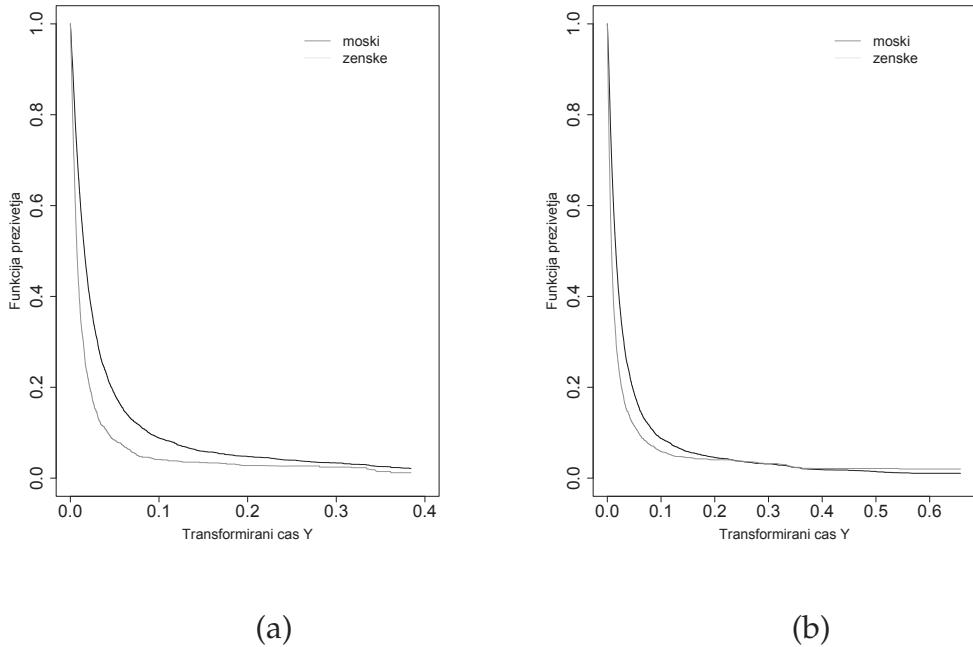
Slika 6.18: Na podlagi modelov ocenjene funkcije preživetja moških pacientov, diagnosticiranih v 3. stadiju pljučnega raka, za spremenljajoče vrednosti leta diagnoze pri starosti 64 let za (a) model z interakcijami in (b) enostavni model



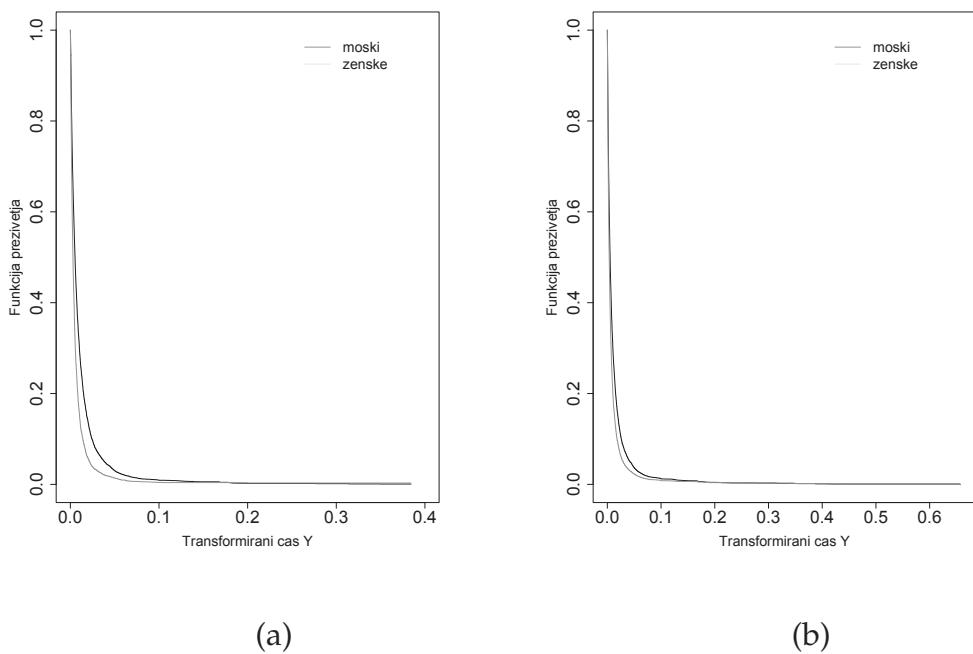
Slika 6.19: Na podlagi modelov ocenjene funkcije preživetja moških in ženskih pacientov, diagnosticiranih v 1. stadiju pljučnega raka, za starost 64 let in leto diagnoze 1995 za (a) model z interakcijami in (b) enostavni model



Slika 6.20: Na podlagi modelov ocenjene funkcije preživetja moških in ženskih pacientov, diagnosticiranih v 2. stadiju pljučnega raka, za starost 64 let in leto diagnoze 1995 za (a) model z interakcijami in (b) enostavni model



Slika 6.21: Na podlagi modelov ocenjene funkcije preživetja moških in ženskih pacientov, diagnosticiranih v 3. stadiju pljučnega raka, za starost 64 let in leto diagnoze 1995 za (a) model z interakcijami in (b) enostavni model



6.2.2 Aalenov model na podatkih o raku na dojki

Za namen proučevanja Aalenovega modela na podatkih o raku na dojki, smo kot neodvisne spremenljivke v model vključili starost, leto diagnoze, stadij in indikator operativnega zdravljenja (ki pove, ali je bila pacientka zdravljena z operativnim posegom). Predpostavljamo, da operativno zdravljenje, zlasti v prvem in drugem stadiju bolezni, predstavlja pomembno informacijo o stanju pacientk ob diagnozi in pomembnejše vpliva na preživetje pacientk.

Poleg glavnih vplivov so v model vključene tudi interakcije starosti in leta diagnoze s stadijem bolezni.

Kljub visokemu preživetju se ocenjevanje modela ustavi v času $Y \approx 0,777$, ko matrika $X(y)$ izgubi polni rang, kar predstavlja tudi zadnji dogodek pacientov, diagnostiranih v tretjem stadiju raka na dojki. Proti koncu opazovanja so 95-odstotni intervali zaupanja pri glavnem vplivu starosti ter starosti in leta diagnoze v tretjem stadiju zelo veliki, kar onemogoča preglednost grafičnega prikaza regresijskih kumulativnih funkcij pri teh spremenljivkah. Rezultati so zato grafično prikazani le do časa $Y = 0,7$, kar pa ne vpliva na interpretacijo rezultatov.

Prileganje modela smo preverjali s pomočjo posplošenih ostankov (Aalen 1989), ki so za vsakega od treh stadijev prikazani na sliki 6.22. Razen v prvem stadiju, kjer je opaziti rahlo odstopanje dela točk od simetrale prvega kvadranta, je prileganje modela zadovoljivo.

TST statistike, izračunane na celotnem časovnem intervalu, na katerem je bil model ocenjen, kažejo, da imajo z izjemo leta diagnoze v tretjem stadiju vse spremenljivke statistično značilen vpliv na ogroženost (tabela 6.3). Podoben rezultat pokažejo tudi rezultati TST statistik, izračunanih na krajših časovnih intervalih.

Učinki neodvisnih spremenljivk na ogroženost za smrt za rakom na dojki so prikazani na slikah 6.23, 6.24 in 6.25. Iz grafičnega prikaza ocenjenih kumulativnih regresijskih funkcij pa lahko razberemo naslednje:

- Vpliv starosti in leta diagnoze po določenem času izgubi vpliv na ogroženost.

Tabela 6.3: TST testne statistike za Aalenov model na podatkih o raku na dojki

	TST*	P
Konstanta	8.681	0.000
STAROST	-9.351	0.000
LETO	-3.973	0.000
STADIJ=2	15.284	0.000
STADIJ=3	15.796	0.000
BREZ OPERACIJE	19.719	0.000
STAROST (STADIJ=2)	-14.513	0.000
STAROST (STADIJ=3)	-15.436	0.000
LETO (STADIJ=2)	-4.196	0.000
LETO (STADIJ=3)	-1.491	0.136

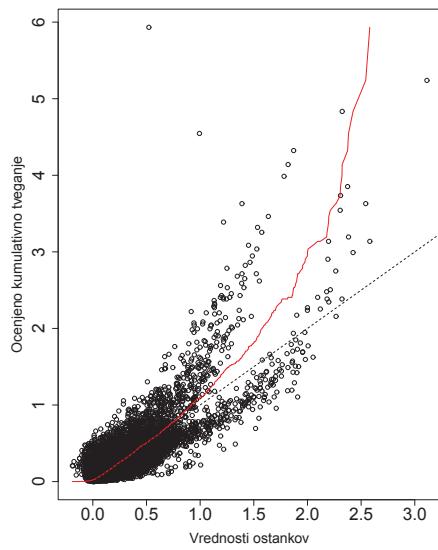
*Ob veljavni ničelni domnevi so TST testne statistike porazdeljene po standardizirani normalni porazdelitvi.

- Vpliv starosti na ogroženost za smrt za rakom na dojki je v začetnem obdobju negativen, kar pomeni da je za mlajše patientke ogroženost večja kot za starejše. Mlajše patientke v primerjavi z vrstnicami v splošni populaciji zaradi raka na dojki izgubijo več kot starejše patientke.
- V začetnem obdobju je vpliv leta diagnoze na ogroženost za smrt negativen, kar pomeni, da se preživetje patientk z rakom na dojki v času izboljšuje.
- Vpliv stadija na ogroženost za smrt se v višjih stadijih povečuje. Medtem ko je negativen vpliv tretjega stadija na preživetje prisoten skozi celotno časovno obdobje ocenjevanja, se negativni vpliv na preživetje v drugem stadiju po določenem obdobju izgubi. Vrednost transformiranega časa, kjer se to zgodi, se nahaja na intervalu med $Y = 0,3$ in $Y = 0,4$.
- Ogroženost patientk brez operativnega posega se skozi celotno ocenjevano obdobje povečuje.
- Interakcije starosti s stadijem kažejo, da se pozitiven vpliv starosti na preživetje v

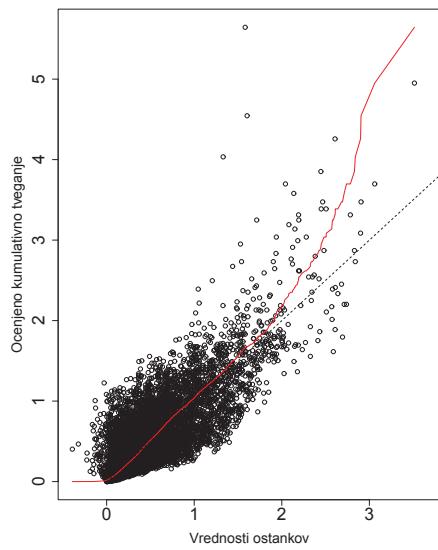
višjih stadijih poveča. Višji ko je stadij raka na dojki, močnejši je pozitiven vpliv starosti na preživetje. V tretjem stadiju je pozitiven vpliv starosti na preživetje prisoten domala skozi celotno ocenjevano časovno obdobje, medtem ko se vpliv starosti na preživetje v drugem stadiju približno v času $Y \approx 0,2$ izgubi.

- Pri interakcijah leta diagnoze s stadijem so TST statistike pokazale, da je vpliv leta diagnoze v tretjem stadiju neznačilen. V drugem stadiju se negativen vpliv leta diagnoze na ogroženost za smrt v začetnem obdobju poveča, kasneje (približno v času $Y \approx 0,2$) pa se ta vpliv izgubi.

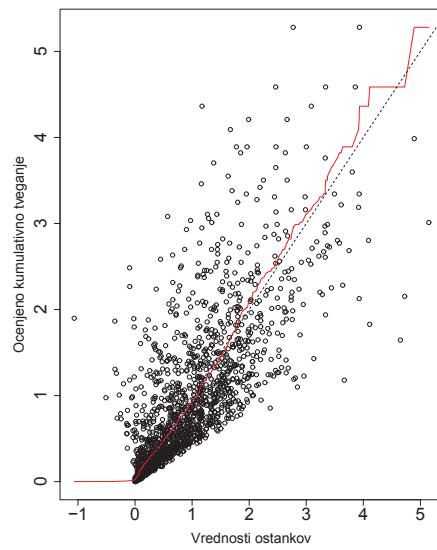
Slika 6.22: Grafični prikaz posplošenih ostankov za (a) 1. stadij raka na dojki, (b) 2. stadij raka na dojki in (c) 3. stadij raka na dojki



(a)

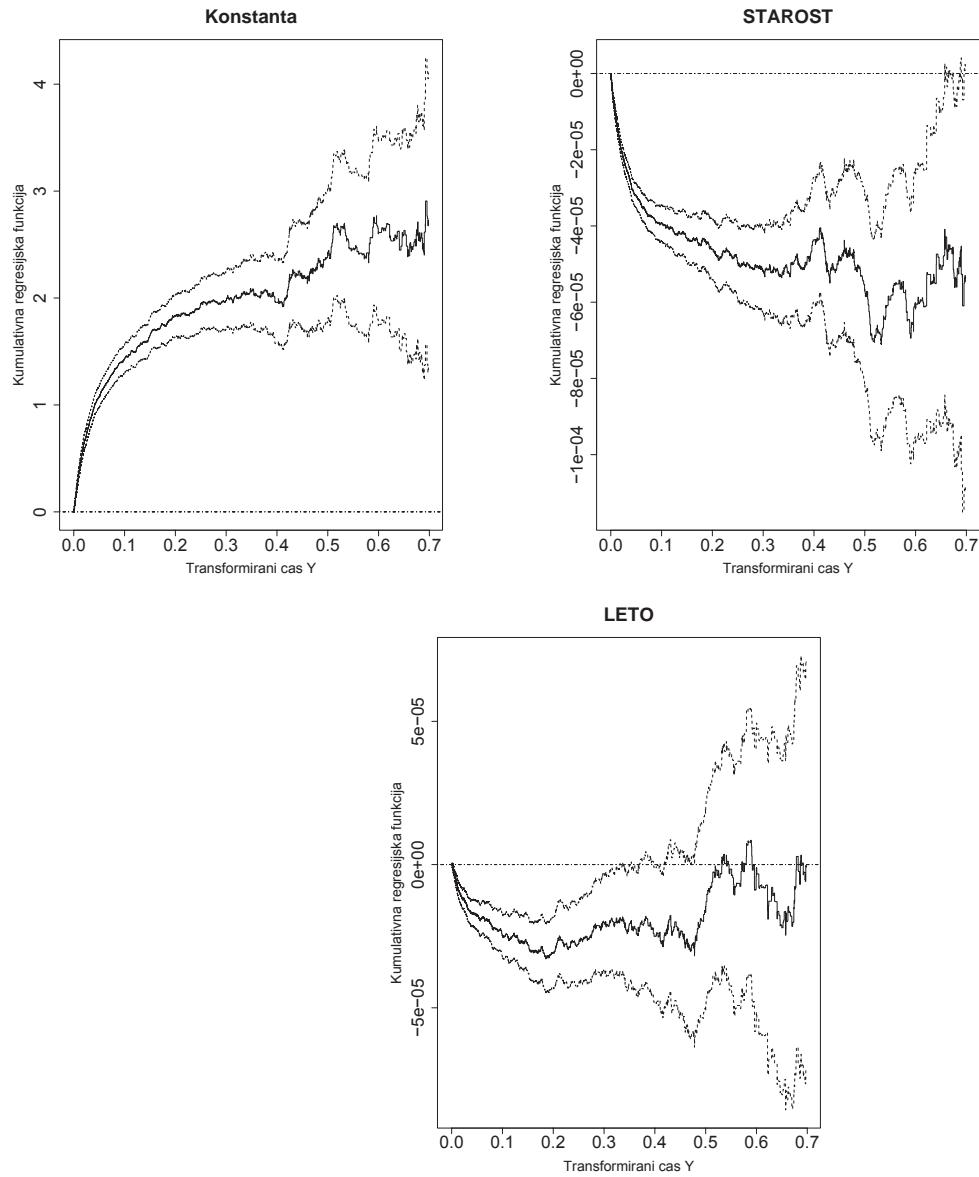


(b)

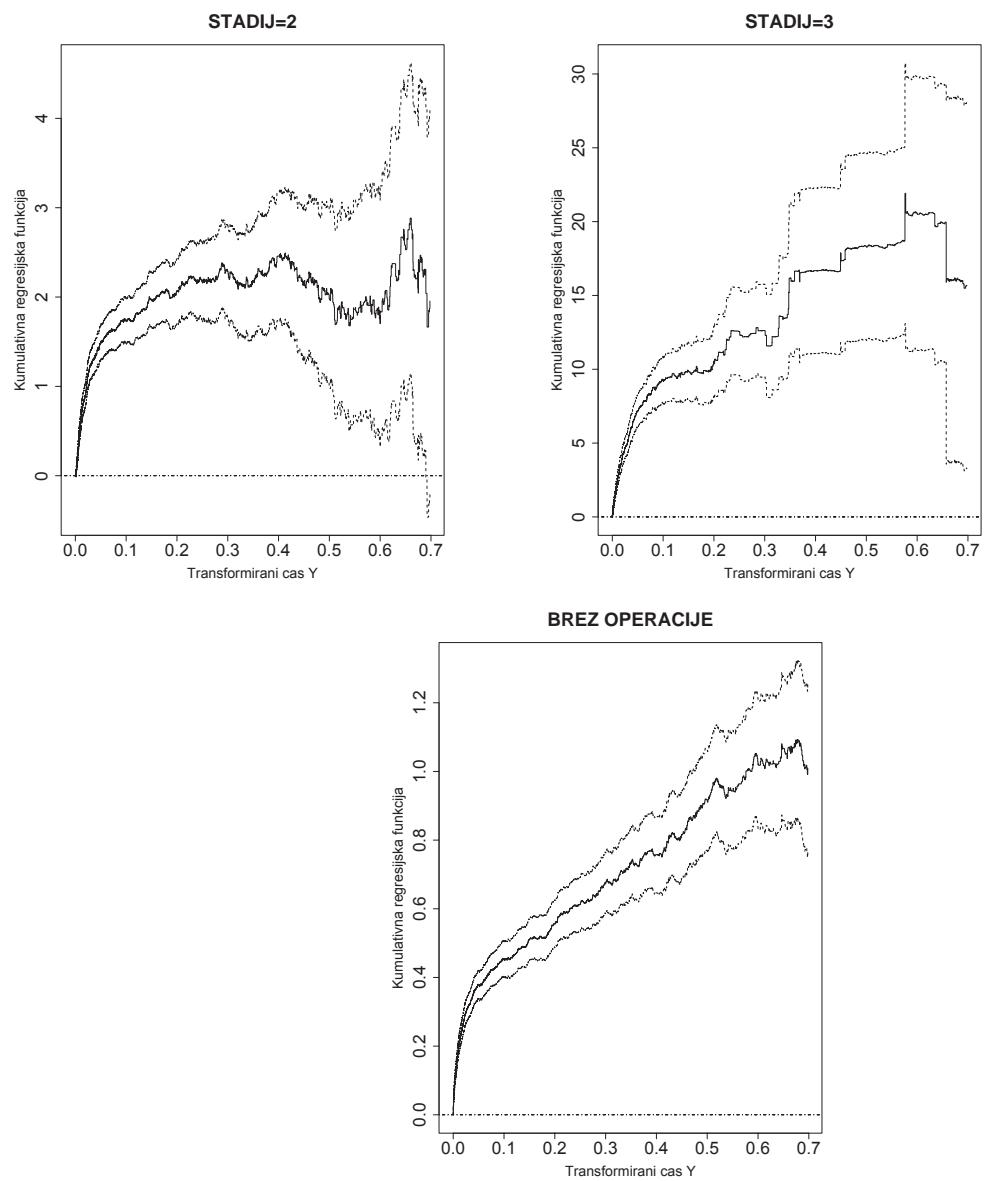


(c)

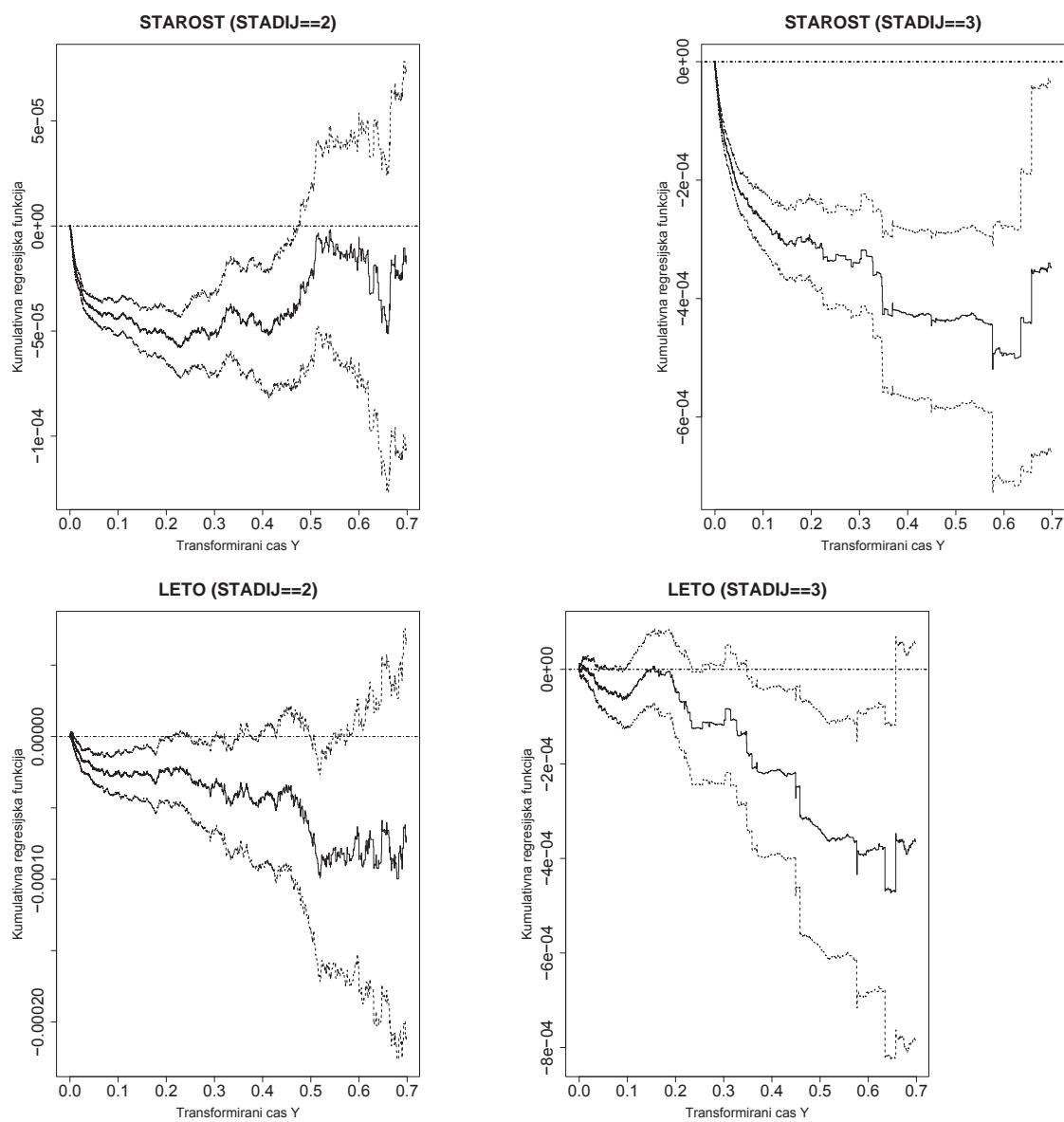
Slika 6.23: Ocjenjene kumulativne regresijske funkcije s 95-odstotnim intervalom zaupanja za podatke o raku na dojki



Slika 6.24: Nadaljevanje slike 6.23



Slika 6.25: Nadaljevanje slike 6.23



Podobno kot pri pljučnem raku tudi pri raku na dojki lahko ugotovimo, da je interpretacija modela z vključenimi interakcijami težavna. Del presežne ogroženosti v drugem in tretjem stadiju se prenese na spremenljivke, s katerimi določimo vrednosti transformiranih časov. Čeprav na podlagi ocenjenih kumulativnih regresijskih funkcij lahko ugotovimo, da se vpliv starosti in leta diagnoze po stadijih razlikuje, pa je kvantitativne učinke v posameznih stadijih težko ovrednotiti. Poleg tega do sprememb učinkov prihaja v vrednostih transformiranih časov, kjer so standardne napake ocenjenih kumulativnih regresijskih funkcij že velike, točkaste ocene pa precej bolj podvržene slučajnim vplivom. Interpretacija rezultatov je zaradi tega težja, zato je izbira enostavnega modela toliko primernejša.

V nadaljevanju je prikazan še enostavni model, ki ne vključuje interakcij starosti in leta diagnoze s stadijem raka na dojki. Čas v katerem se ocenjevanje modela ustavi oziroma matrika $\mathbf{X}(y)$ izgubi polni rang, je enak kot v modelu z interakcijami ($Y \approx 0,777$) in predstavlja zadnji dogodek pacientk v tretjem stadiju raka na dojki.

Pospoljeni ostanki, ki so za vsakega od treh stadijev prikazani na sliki 6.26, kažejo na zadovoljivo prileganje modela v vseh treh stadijih.

TST statistike, izračunane na celotnem ocenjevanem obdobju opazovanega transformiranega časa ($Y \in [0, 0,777]$), so prikazane v tabeli 6.4, iz njih pa lahko vidimo, da je vpliv vseh proučevanih spremenljivk na ogroženost za smrt za rakom na dojki značilen. Vplivi neodvisnih spremenljivk na ogroženost za smrt za rakom na dojki so prikazani na slikah 6.27 in 6.28, iz njih pa lahko razberemo naslednje:

- Starost pozitivno vpliva na preživetje pacientk z rakom na dojki, vendar se vpliv starosti po določenem obdobju izgubi. Na podlagi grafičnega prikaza kumulativne regresijske funkcije za spremenljivko starost vidimo, da se to zgodi približno v času $Y \approx 0,2$, opazimo pa lahko tudi, da je vpliv starosti na preživetje do transformiranega časa $Y \approx 0,1$ precej večji kot v obdobju $0,1 < Y < 0,2$.
- Leto diagnoze ima v začetnem obdobju pozitiven vpliv na preživetje, po času $Y \approx 0,1$ pa se ta vpliv izgubi.

Tabela 6.4: TST testne statistike za Aalenov enostavni model na podatkih o raku na dojki

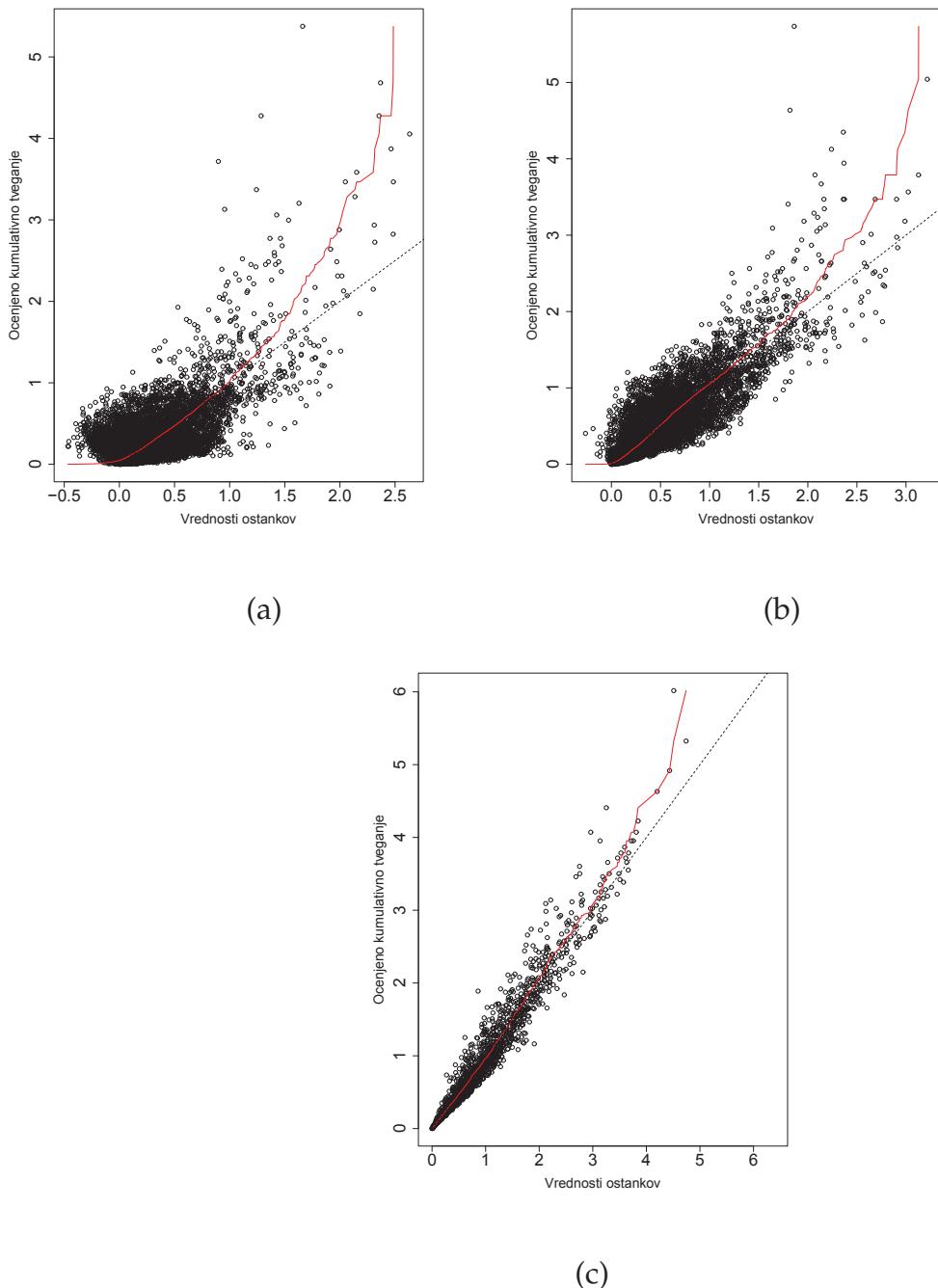
	TST*	P
Konstanta	38.497	0.000
STAROST	-37.006	0.000
LETO	-9.941	0.000
STADIJ=2	28.729	0.000
STADIJ=3	28.634	0.000
BREZ OPERACIJE	21.165	0.000

*Ob veljavni ničelni domnevi so TST testne statistike porazdeljene po standardizirani normalni porazdelitvi.

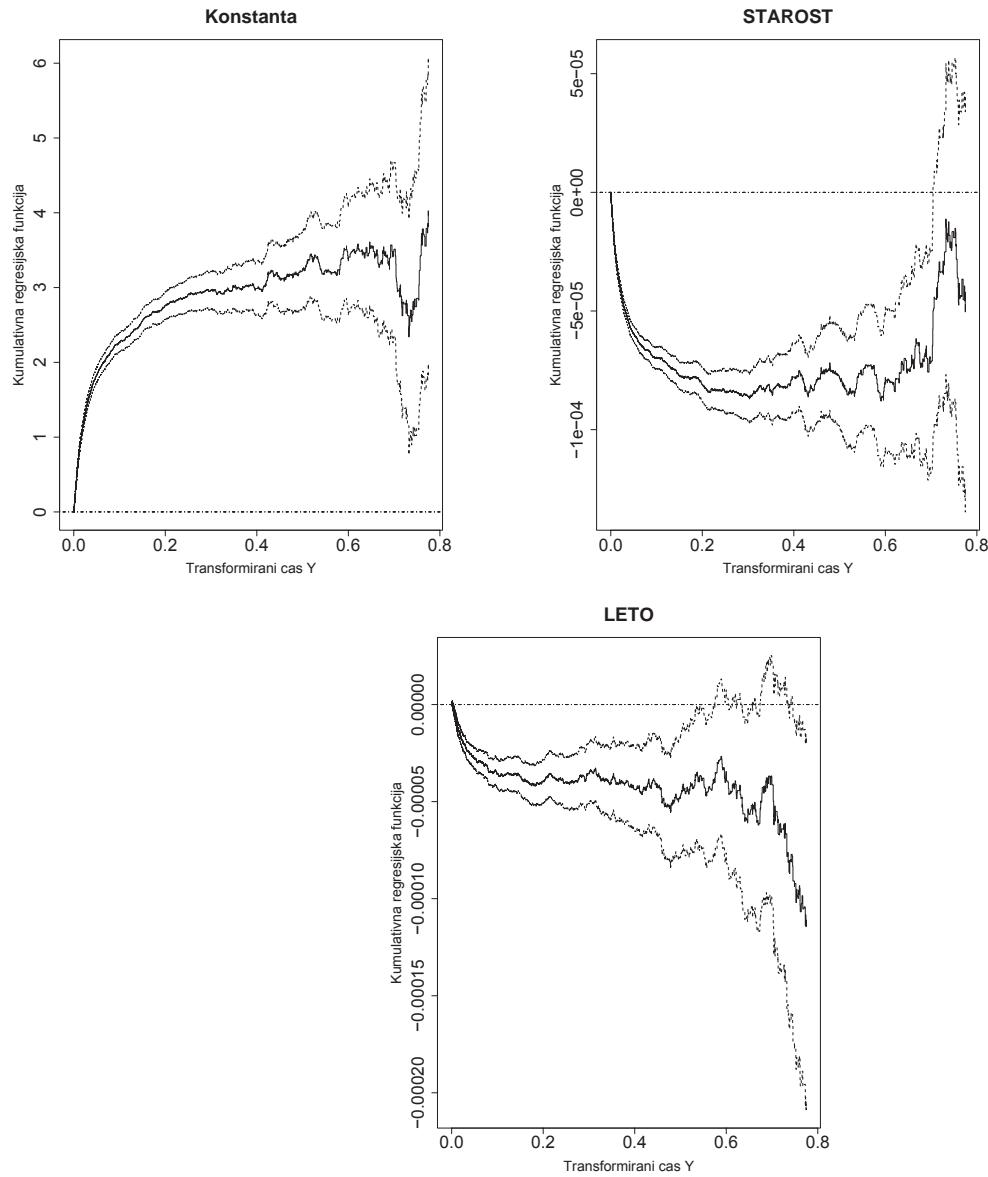
- Ogroženost za smrt za rakom na dojki se v višjih stadijih povečuje. V obeh stadijih (drugem in tretjem) lahko opazimo, da se pozitiven vpliv stadija na ogroženost izgubi približno v času $Y \approx 0,4$. Pri tem je treba poudariti, da do vrednosti transformiranega časa $Y \approx 0,4$ umre dobrih 97 % pacientov v tretjem stadiju, kar pomeni, da ocenjena kumulativna regresijska funkcija za tretji stadij po tem času ne prinaša koristnih informacij za interpretacijo.

Primerjava glavnih vplivov drugega in tretjega stadija raka na dojki pokaže, kako se zaradi negativnega dodatnega vpliva starosti in leta diagnoze v teh stadijih njun vpliv v modelu z interakcijami poveča. To potrjuje dejstvo, da se presežni vpliv višjih stadijev deloma prenese na spremenljivke, na podlagi katerih računamo transformirani čas.

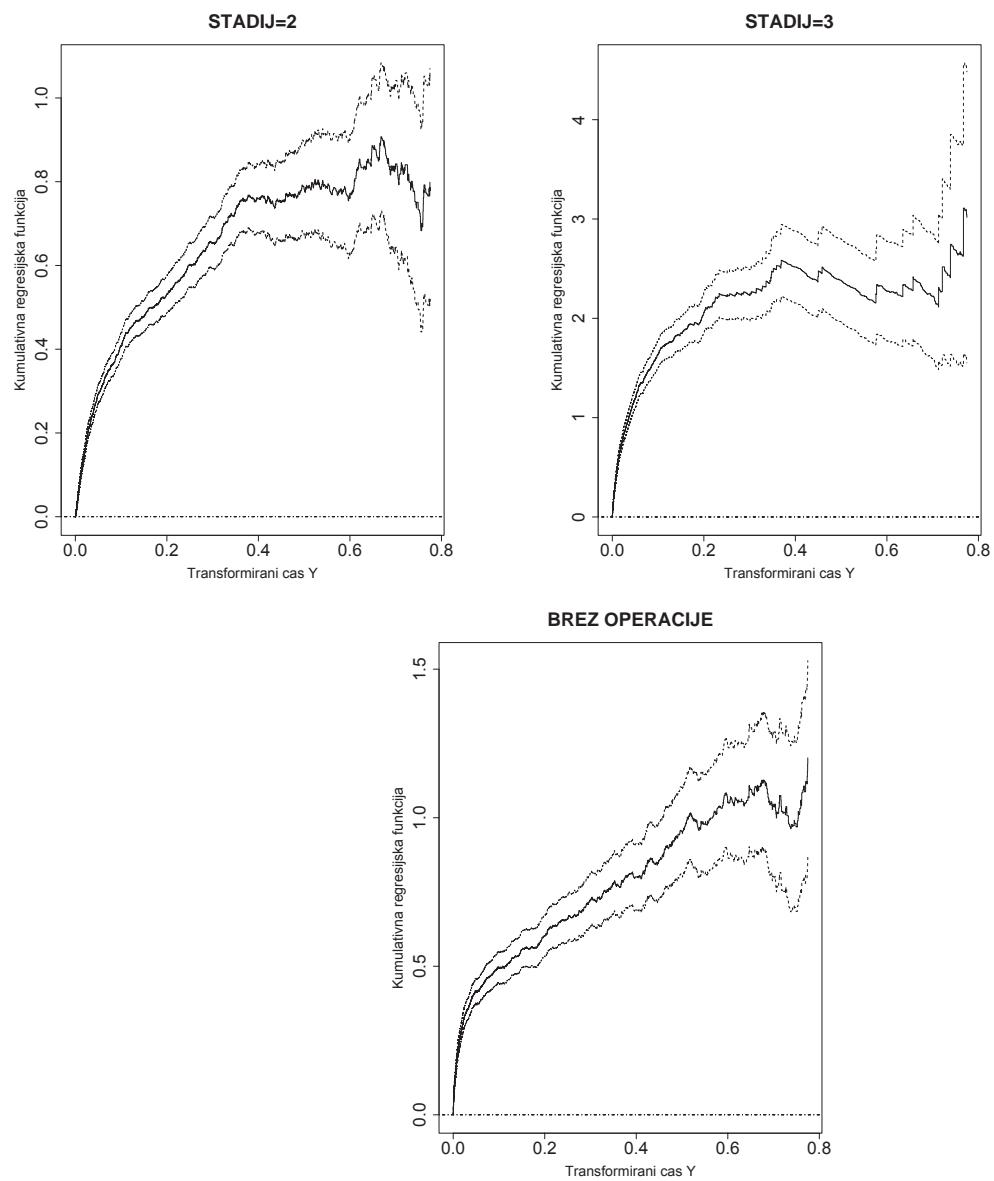
Slika 6.26: Grafični prikaz pospoljenih ostankov v enostavnem modelu za (a) 1. stadij raka na dojki, (b) 2. stadij raka na dojki in (c) 3. stadij raka na dojki



Slika 6.27: Ocjenjene kumulativne regresijske funkcije s 95-odstotnim intervalom zaupanja za podatke raka na dojki - enostavni model



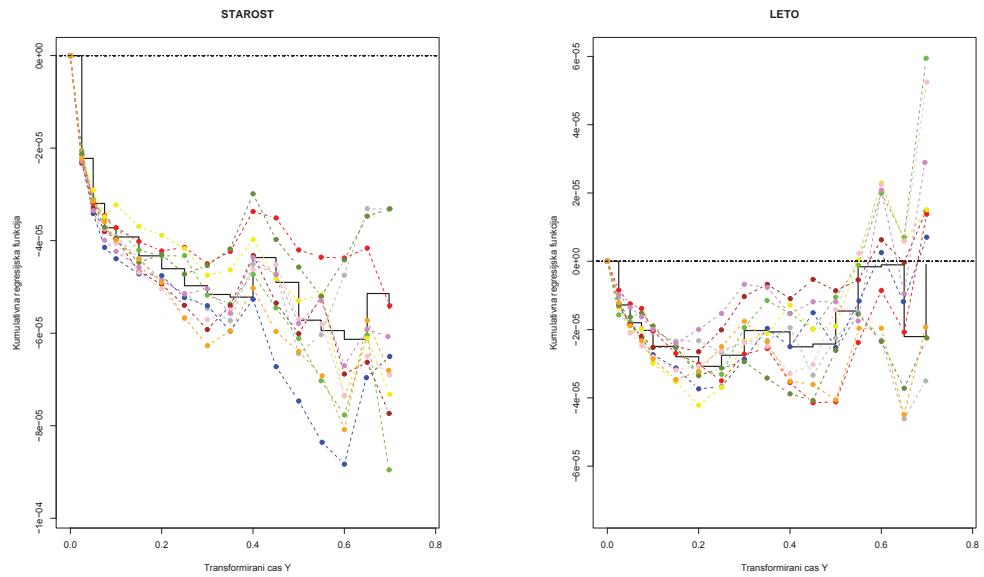
Slika 6.28: Nadaljevanje slike 6.27



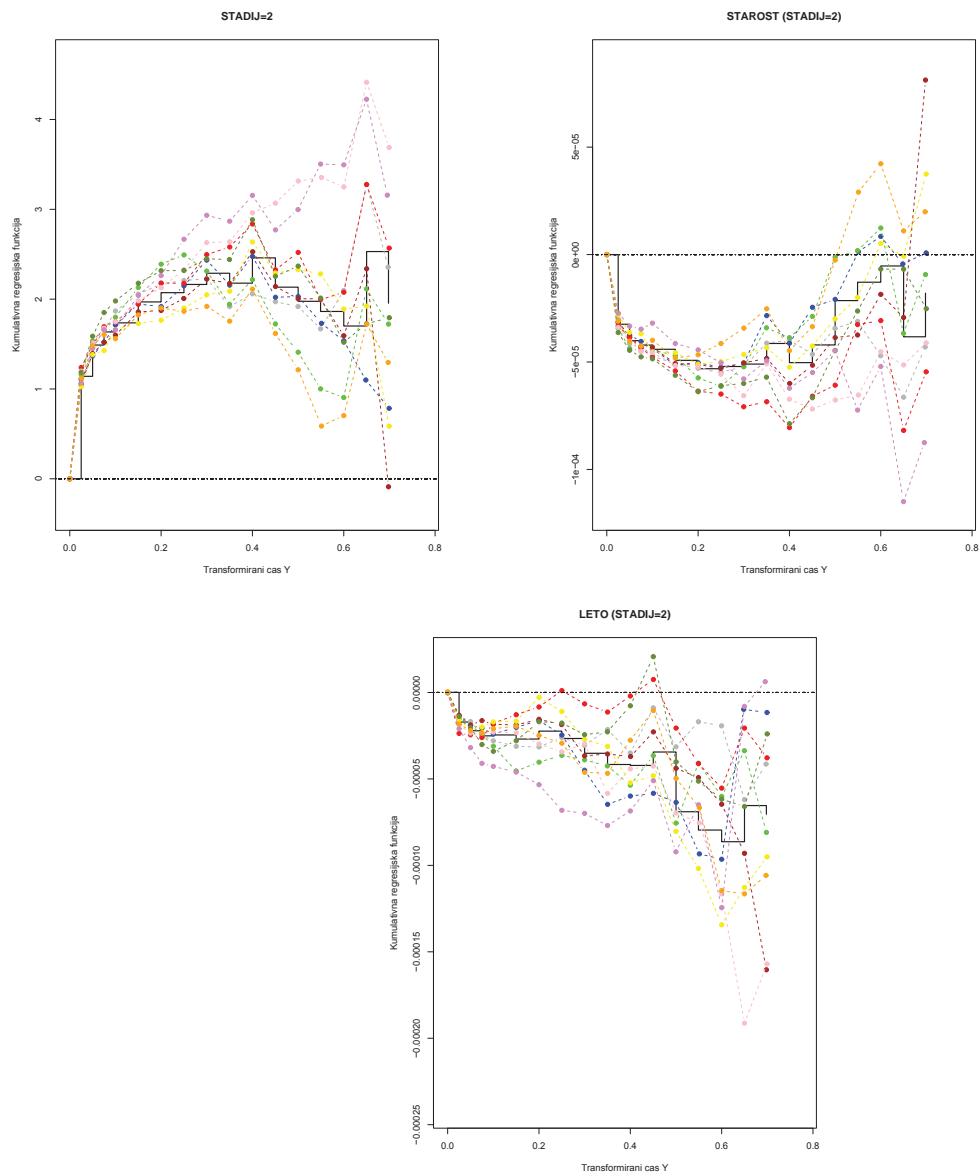
Aalenov model na podatkih o raku na dojki je tako v primeru modela z interakcijami kot v primeru enostavnega modela pokazal, da neodvisni spremenljivki starost in leto diagnoze po določenem času izgubita vpliv na ogroženost za smrt. Opazimo pa lahko, da je približna točka preloma v modelih različna vsaj za spremenljivko leto diagnoze. V modelu z interakcijami stadija s starostjo in letom diagnoze glavna vpliva starost in leto diagnoze izgubita vpliv na ogroženost za smrt ob približno enakem času ($Y \approx 0,2$). Podobno lahko zaznamo, da se dodatni vpliv starosti v drugem stadiju izgubi ob približno enakem času ($Y \approx 0,2$), dodatni vpliv leta diagnoze pa se izgubi že nekoliko prej (približno v času $Y \approx 0,1$). Presežni del ogroženosti v drugem stadiju, ki se ne porazdeli na dodatni vpliv spremenljivk starosti in leta diagnoze v drugem stadiju, pa se izgubi na intervalu $Y \in [0,3, 0,4]$.

Na drugi strani se v enostavnem modelu vpliv starosti (podobno kot v modelu z interakcijami) izgubi v času $Y \approx 0,2$, vpliv leta diagnoze pa se izgubi že nekoliko prej (približno v času $Y \approx 0,1$). Presežna ogroženost drugega stadija v enostavnem modelu, ki je prikazana na sliki 6.28 in se v modelu z interakcijami delno prenese na dodatni vpliv starosti in leta diagnoze v drugem stadiju, izgubi vpliv približno v času $Y \approx 0,4$. Da bi ugotovili, ali so opisane spremembe učinkov v času dejansko prisotne ali so zgolj posledica slučajnih vplivov, smo izvedli enostavno bootstrap analizo, pri čemer smo izvedli deset bootstrap ponovitev. Rezultati analize diskretnih kumulativnih regresijskih funkcij so prikazani na slikah 6.29, 6.30 in 6.31. Na podlagi grafičnega prikaza bootstrap analize ugotavljamo, da so rezultati skladni z intervali zaupanja ocenjenih kumulativnih regresijskih funkcij. Precej nazorneje se za glavni vpliv ogroženosti v drugem stadiju v modelu z interakcijami lahko razbere, da se ta vpliv izgubi približno v času $Y \approx 0,4$.

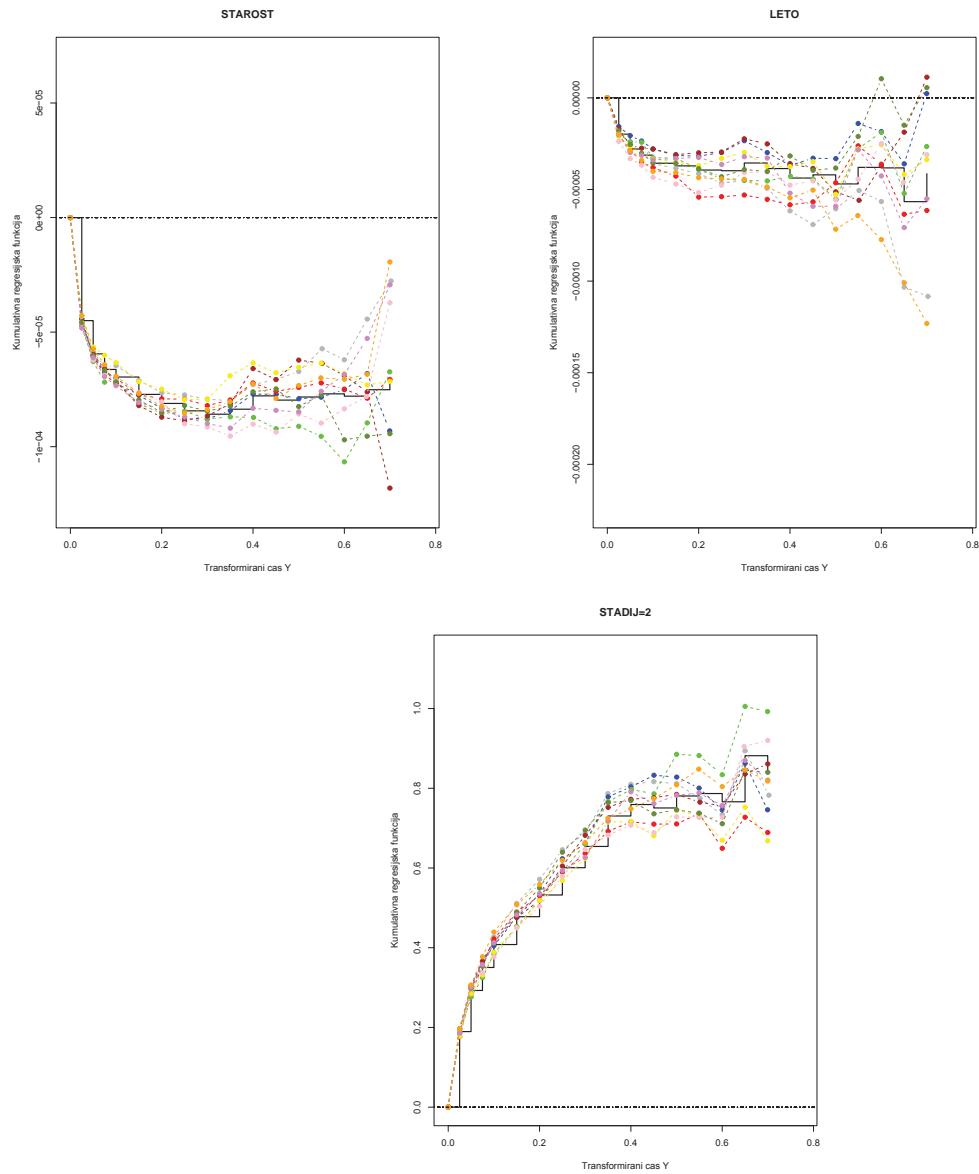
Slika 6.29: Ocenjene kumulativne regresijske funkcije v diskretni obliki originalnega vzorca (črna okrepljena črta) in desetih bootstrap vzorcev (barvne črte) - model z interakcijami stadija s starostjo in letom diagnoze



Slika 6.30: Nadaljevanje slike 6.29



Slika 6.31: Ocenjene kumulativne regresijske funkcije v diskretni obliki originalnega vzorca (črna okrepljena črta) in desetih bootstrap vzorcev (barvne črte) - enostavni model



Vpliv kovariat na preživetje pacientk z rakom na dojki (Adjusted survival curves)

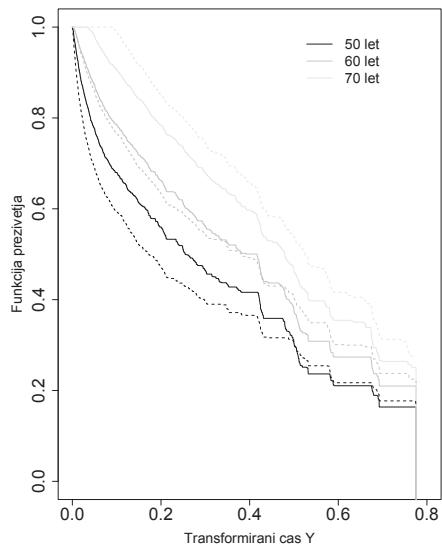
Primerjavo med modelom z interakcijami in enostavnim modelom bomo v nadaljevanju predstavili še s pomočjo funkcij preživetja, ki jih dobimo na podlagi ocenjenega Aalenovega modela. Proučili bomo vpliv starosti in leta diagnoze na relativno preživetje pacientk v vsakem od treh stadijev raka na dojki. Za spremenljajoče vrednosti starosti bomo vzeli vrednosti 50, 60 in 70 let, za spremenljajoče vrednosti leta diagnoze pa bomo vzeli vrednosti 1. januar 1983, 1. januar 1993 in 1. januar 2003. Za fiksne vrednosti zveznih spremenljivk smo izbrali njihove mediane. Tako smo pri spremenljajoči starosti za fiksno vrednost leta diagnoze izbrali 10. junij 1996, pri spremenljajočem letu diagnoze pa smo za fiksno vrednost starosti izbrali 60 let. V grafičnih prikazih smo se omejili le na patientke, ki so bile zdravljene z operacijo.

Slike 6.32 in 6.33 kažeta, da sta različna modela (model z interakcijami in enostavni model) najbolj neskladna v prvem in tretjem stadiju, kjer prihaja predvsem do manjših razlik v legi ocenjenih funkcij relativnega preživetja, pri vplivu starosti v tretjem stadiju pa je opazen tudi nekoliko večji vpliv starosti na relativno preživetje pacientk.

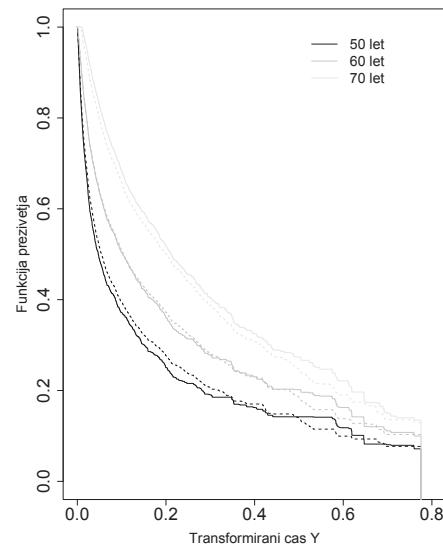
Na podlagi grafičnega prikaza vplivov kovariat na relativno preživetje v splošnem lahko zaključimo, da je pozitiven vpliv starosti na relativno preživetje pacientk prisoten v vseh treh stadijih (slika 6.32), vendar je vpliv starosti na preživetje v tretjem stadiju manjši kot v prvih dveh stadijih raka na dojki.

Na relativno preživetje pacientk vpliva tudi leto diagnoze (slika 6.33), kar kaže na to, da se zdravljenje in/ali medicinska oskrba patientk v času izboljšuje, vendar lahko opazimo, da je v primerjavi s prvima stadijema vpliv leta diagnoze na relativno preživetje pacientk, diagnosticiranih v tretjem stadiju, neznaten.

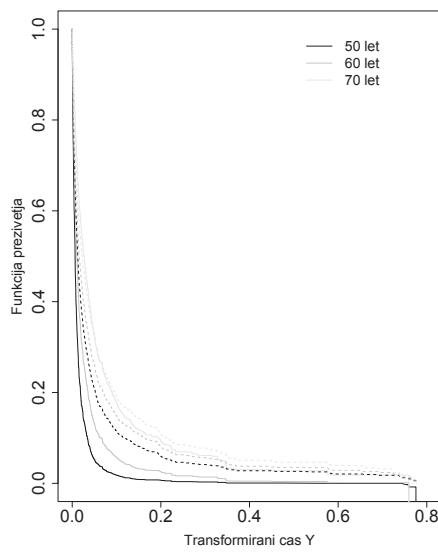
Slika 6.32: Na podlagi modelov ocenjene funkcije preživetja operativno zdravljenih pacientk za spremenjajoče vrednosti starosti v letu 1996 (neprekinjena črta - model z interakcijami, prekinjena črta - enostavni model) za (a) 1. stadij raka na dojki, (b) 2. stadij raka na dojki in (c) 3. stadij raka na dojki



(a)

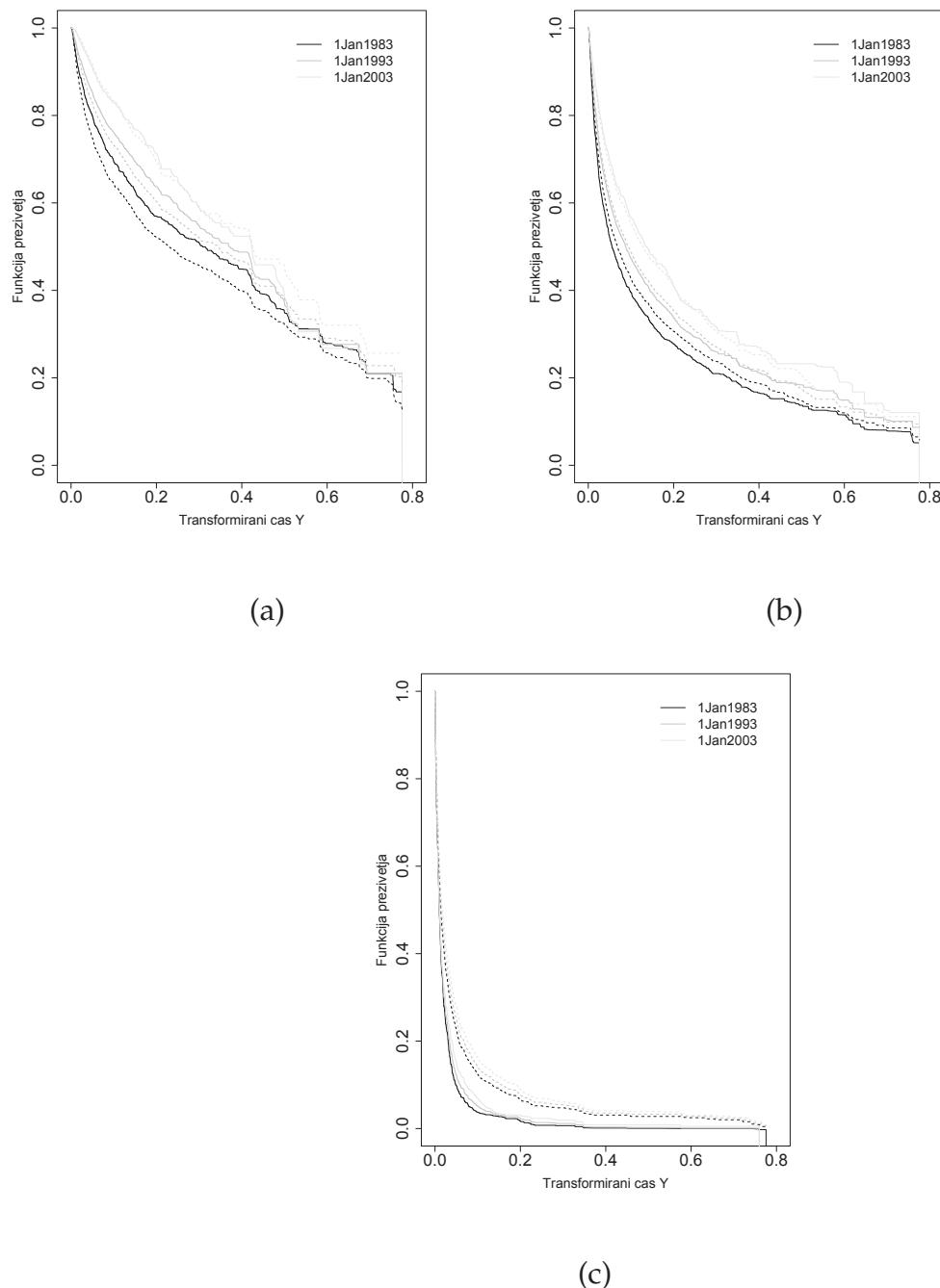


(b)



(c)

Slika 6.33: Na podlagi modelov ocenjene funkcije preživetja operativno zdravljenih pacientk za spremenjajoče vrednosti leta diagnoze pri starosti 60 let (neprekinjena črta - model z interakcijami, prekinjena črta - enostavni model) za (a) 1. stadij raka na dojki, (b) 2. stadij raka na dojki in (c) 3. stadij raka na dojki



Čas statistične ozdravitve patientk, diagnosticiranih z rakom na dojki

Aalenov model raka na dojki je pokazal, da spremenljivki starost in leto diagnoze po določenem času izgubita vpliv na ogroženost za smrt za rakom na dojki. Na tem mestu naj spomnimo na eno od značilnosti transformiranega časa, in sicer, da je ta na intervalu $[0, 1]$ enakomerno porazdeljen, torej funkcijo preživetja lahko zapišemo: $S_Y(y) = 1 - y$. Z vidika ogroženosti za smrt splošne populacije so si torej pacienti med seboj enakovredni, presežna ogroženost, ki je lastna proučevani bolezni, pa se odrazi v spremenjeni porazdelitvi transformiranega časa.

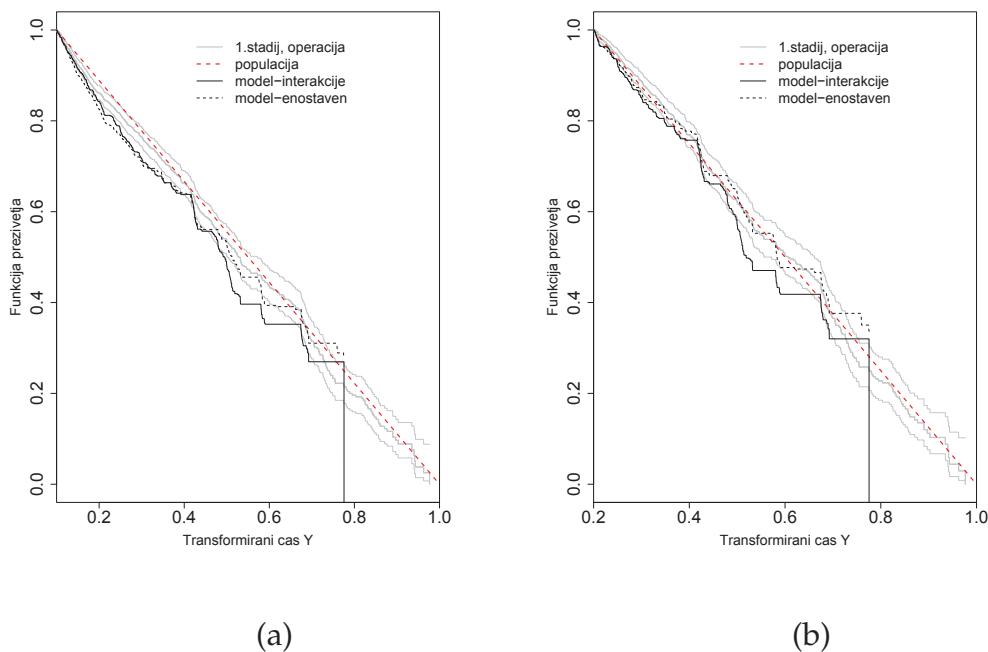
Presežna ogroženost se odrazi tudi na spremenljivkah, s katerimi računamo transformirani čas (pri raku na dojki sta to spremenljivki starost in leto diagnoze), na podlagi česar bi lahko sklepali, da se v primeru, ko te spremenljivke izgubijo vpliv na ogroženost, presežna ogroženost pacientov izgubi, preživetje pacientov pa se izenači s preživetjem splošne populacije.

Ocenjevanje vplivov neodvisnih spremenljivk na ogroženost smo proučevali z dvema modeloma: prvi model je vključeval interakcije starosti in leta diagnoze s stadijem, v drugem modelu pa smo proučevali le glavne vplive starosti, leta diagnoze in stadija na ogroženost za smrt za rakom na dojki. Oba modela pokažeta, da se po določenem času glavni vplivi starosti in leta diagnoze izgubijo. Z vsebinskega vidika bi lahko sklepali, da paciente, diagnosticirane v prvem stadiju raka na dojki in zdravljenje z operacijo (zdravljenje brez operacije skozi celotno obdobje proučevanja zvišuje ogroženost za smrt), po določenem času ozdravijo. Čas, v katerem se to zgodi, je čas, v katerem starost in spol izgubita vpliv na ogroženost. V modelu z interakcijami obe spremenljivki izgubita vpliv na ogroženost ob približno istem času ($Y \approx 0,2$). V enostavnem modelu pa se vpliv leta diagnoze izgubi že nekoliko prej ($Y \approx 0,1$).

Da bi ugotovili, kaj se dogaja s porazdelitvijo transformiranega časa, smo za paciente, ki so preživele čas $Y = 0,1$ oziroma čas $Y = 0,2$, na podlagi modelov in dejanskih podatkov ocenili funkcije relativnega preživetja. Grafična analiza ocenjenih krivulj preživetja je prikazana na spodnjih slikah (6.34, 6.35). Slika 6.34 kaže ocenjeno preživetje operativno zdravljenih pacientek, diagnosticiranih v prvem stadiju raka na dojki, ki so preživele čas $Y = 0,1$ oziroma čas $Y = 0,2$, in na podlagi modela ocen-

jeno preživetje od časa $Y = 0,1$ oziroma $Y = 0,2$ za operativno zdravljene patientke v starosti 60 let, diagnosticirane v prvem stadiju, pri čemer smo za leto diagnoze izbrali 28. julij 1995. Iz slike lahko razberemo, da je preživetje za patientke, ki so preživele vrednost transformiranega časa $Y = 0,1$, v začetnem obdobju nekoliko manjše od preživetja splošne populacije, medtem ko se preživetje patientk, ki so preživele vrednost transformiranega časa $Y = 0,2$ v celotnem obdobju približno izenači s preživetjem splošne populacije.

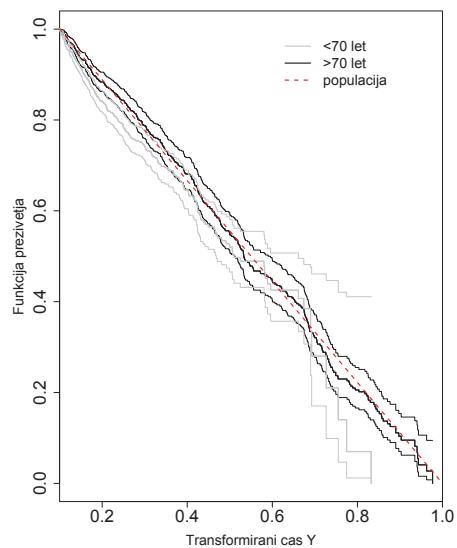
Slika 6.34: Ocenjene krivulje preživetja za operativno zdravljene patientke, diagnosticirane v prvem stadiju raka na dojki, za vrednosti transformiranega časa (a) $Y > 0,1$ in (b) $Y > 0,2$.



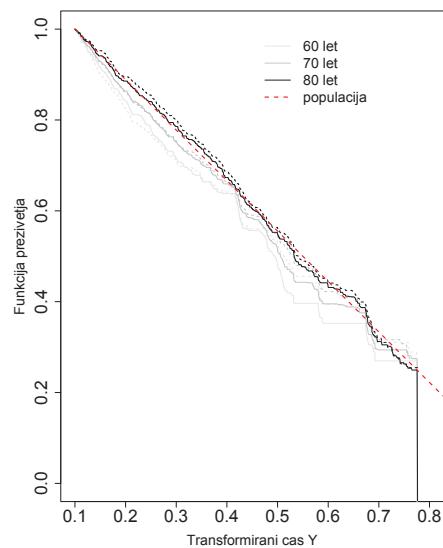
Enostavni model je pokazal, da se glavni vpliv leta diagnoze izgubi v času $Y \approx 0,1$, glavni vpliv starosti pa ima rahel vpliv na relativno preživetje še do časa $Y \approx 0,2$. Ugotovili smo, da se preživetje operativno zdravljenih patientk, diagnosticiranih v prvem stadiju raka na dojki, po času $Y = 0,2$ v povprečju približa populacijskemu preživetju. V nadaljevanju smo zato proučili še vpliv starosti na preživetje teh patientk. Pri dejanskih podatkih smo patientke starostno razdelili na mlajše in starejše od 70 let. Pri tem smo upoštevali, da je mediana patientk, ki so preživele čas $Y = 0,1$, približno 70 let. Pri modelskih ocenah krivulj preživetja smo proučevali starosti 60, 70 in 80

let. Grafična analiza vpliva starosti na preživetje je prikazana na sliki 6.35, zaradi lažje preglednosti pa so ocenjene krivulje preživetja za dejanske podatke in za oba modela prikazane ločeno. Iz slik je očitno, da za proučevani čas $Y > 0,1$ glede na starost prihaja do razlik v ocenjenih krivuljah preživetja. Preživetje starejših patientk se populacijskemu preživetju približa hitreje kot preživetje mlajših patientk. Na drugi strani razlika v preživetju glede na starost za patientke, ki so preživele čas $Y = 0,2$, izgine tako na dejanskih podatkih, kot tudi v modelih.

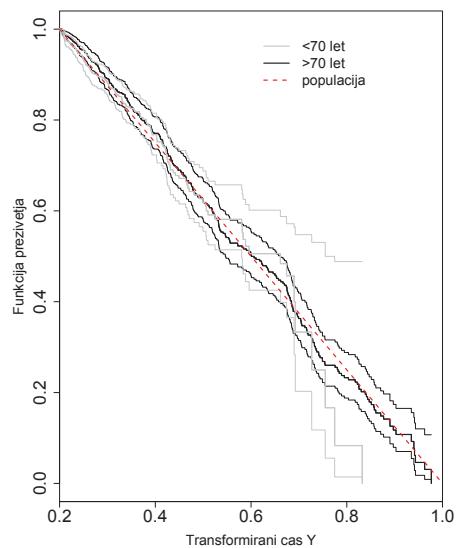
Slika 6.35: Ocnjene krivulje preživetja za operativno zdravljene patientke, diagnosticirane v prvem stadiju raka na dojki, za vrednosti transformiranega časa (a) $Y > 0,1$ - dejanski podatki, (b) $Y > 0,1$ - model, (c) $Y > 0,2$ - dejanski podatki in (d) $Y > 0,2$ - model.



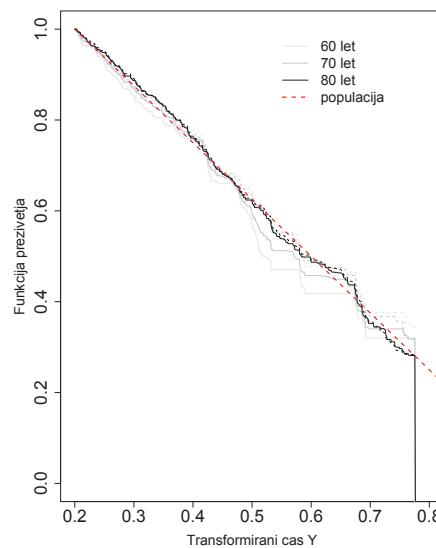
(a)



(b)



(c)



(d)

Grafične analize so pokazale, da se preživetje operativno zdravljenih patientk, diagnosticiranih v prvem stadiju, ki preživijo čas $Y = 0,2$, izenači s preživetjem splošne populacije. To pomeni, da so patientke po tem času izpostavljene samo še populacijskim tveganjem smrti. Ogroženost v Aalenovem modelu lahko zapišemo:

$$\lambda(y) = \beta_0(y) + \beta_1(y)STAROST + \beta_2(y)LETO \quad (6.2)$$

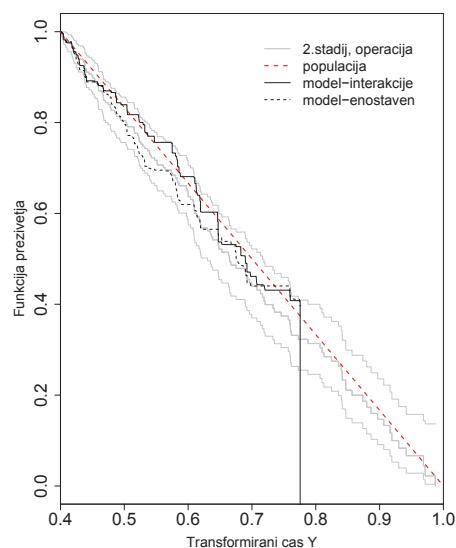
in po času $Y = 0,2$ koeficiente $\beta_1(t)$ in $\beta_2(t)$ izgubita vpliv na ogroženost. Ogroženost postane enaka:

$$\lambda(y) = \beta_0(y) = \lambda_P(y) = \frac{1}{1-y}, \quad \text{za } y > 0,2, \quad (6.3)$$

kjer je $\lambda_P(y)$ populacijska ogroženost za smrt. Statistično gledano operativno zdravljene patientke, diagnosticirane v prvem stadiju raka na dojki, po vrednosti transformiranega časa $Y = 0,2$ v povprečju ozdravijo.

Rezultati ocenjenega Aalenovega modela so pokazali tudi, da presežna ogroženost za operativno zdravljene patientke, diagnosticirane v drugem stadiju, izgine po času $Y \approx 0,4$. Po tem času se ogroženost za smrt za rakom na dojki v drugem stadiju (za operativno zdravljene patientke) izenači z ogroženostjo operativno zdravljenih patientk, diagnosticiranih v prvem stadiju, ki je v tem času enaka ogroženosti splošne populacije. Sklepamo torej lahko, da operativno zdravljene patientke, ki preživijo vrednost transformiranega časa $Y \approx 0,4$, statistično gledano ozdravijo. Funkcije preživetja, ocnjene na dejanskih podatkih in obeh modelih, so prikazane na sliki 6.36, pri čemer smo za oceno modelskih funkcij preživetij vzeli starost 70 let (ki predstavlja tudi približno mediano), za vrednost leta diagnoze pa smo izbrali 28. julij 1995. Vse ocnjene funkcije preživetja potrjujejo predpostavko o statistični ozdravitivi patientk, diagnosticiranih v drugem stadiju raka na dojki.

Slika 6.36: Ocenjene krivulje preživetja za operativno zdravljene patientke, diagnosti-
cirane v drugem stadiju raka na dojki, od vrednosti transformiranega časa $Y > 0,4$.



6.3 Uporaba Coxovega modela sorazmernih ogroženosti v relativnem preživetju

Na podatkih pljučnega raka in raka na dojki smo preizkusili tudi Coxov model sorazmernih ogroženosti, ki je v praksi analize preživetja najpogosteje uporabljen. Predvsem zaradi primerjave z Aalenovim modelom sta v nalogi predstavljena tako model z interakcijami, kot tudi enostavni model.

6.3.1 Coxov model na podatkih pljučnega raka

V tem delu prikazujemo Coxov model sorazmernih ogroženosti, ki poleg glavnih vplivov starosti, spola, leta diagnoze in stadija, v katerem je bil bolnik z rakom na pljučih diagnosticiran, predpostavlja tudi interakcije stadija z ostalimi neodvisnimi spremenljivkami. Predpostavko sorazmernih ogroženosti smo preverjali z dvema testoma. Prvi se osredotoča na linearno povezanost med skaliranimi Schoenfeldovimi ostanki in proučevanim časom. Drugi test temelji na stohastičnem procesu standardiziranih Schoenfeldovih ostankov, ki pod določenimi pogoji konvergirajo k procesu Brownovega mostu (Stare in drugi 2005b). Za testno statistiko smo uporabili maksimalno vrednost tako dobljenega stohastičnega procesa Brownovega mosta.

V tabeli 6.5 zadnja vrstica prikazuje ocenjeno vrednost testnih statistik za preverjanje predpostavke sorazmernih ogroženosti in pripadajoči p-vrednosti za celotni model. Pri prvem testu ostale vrstice prikazujejo ocnjene Pearsonove korelacijske koeficiente med skaliranimi Schoenfeldovimi ostanki in transformiranim časom (Y) za posamezne neodvisne spremenljivke v modelu, statistike s katerimi preverjamo značilnost korelacije oziroma predpostavko sorazmernih ogroženosti in pripadajoče p-vrednosti. Pri drugem testu je za posamezne neodvisne spremenljivke prikazana testna statistika maksimalne vrednosti Brownovega mostu in pripadajoča p-vrednost.

Oba globalna testa z nizkim tveganjem zavrneta predpostavko o sorazmernih ogroženostih modela. Pri preverjanju predpostavke o sorazmernih ogroženostih posameznih neodvisnih spremenljivk pa med testoma prihaja do razlik. Medtem ko drugi test zavrne predpostavko pri vseh spremenljivkah, prvi test ne zazna kršitve predpostavke pri

spremenljivkah LETO, STADIJ=2, pri interakciji starosti v drugem stadiju (STAROST*STADIJ=2) ter pri interakciji leta diagnoze s stadijem (LETO*STADIJ=2 in LETO*STADIJ=3).

Tabela 6.5: Testne statistike za preverjanje predpostavke o sorazmernih ogroženostih

	Test 1			Test 2	
	rho	chisq	p	max	p
STAROST	0.092	131.530	0.000	9.379	0.000
SPOL=ženski	-0.046	33.212	0.000	4.146	0.000
LETO	-0.005	0.338	0.561	1.560	0.015
STADIJ=2	0.006	0.511	0.475	4.075	0.000
STADIJ=3	0.032	16.707	0.000	4.072	0.000
STAROST(STADIJ=2)	-0.014	3.378	0.066	4.957	0.000
STAROST(STADIJ=3)	-0.052	44.298	0.000	3.154	0.000
SPOL=ženski(STADIJ=2)	0.020	6.323	0.012	2.035	0.001
SPOL=ženski(STADIJ=3)	0.029	13.822	0.000	2.712	0.000
LETO(STADIJ=2)	0.009	1.368	0.242	3.802	0.000
LETO(STADIJ=3)	0.009	1.387	0.239	3.795	0.000
GLOBAL		324.888	0.000	7.246	0.000

S pomočjo skaliranih Schoenfeldovih ostankov lahko približna gibanja vplivov neodvisnih spremenljivk v času prikažemo tudi grafično, saj na podlagi (Grambsch in Therneau 1994) velja

$$E(U_{kj}^*) + \hat{\beta}_j \approx \beta_j(t), \quad (6.4)$$

kjer so U_{kj}^* skalirani Schoenfeldovi ostanki. S pomočjo grafičnega prikaza $U_{kj}^* + \hat{\beta}_j$ v času dobimo občutek o obliki in moči nesorazmernih ogroženosti. Na sliki 6.37 so prikazane zglajene krivulje vrednosti izraza $U_{kj}^* + \hat{\beta}_j$ v času Y za vse neodvisne spremenljivke v modelu.

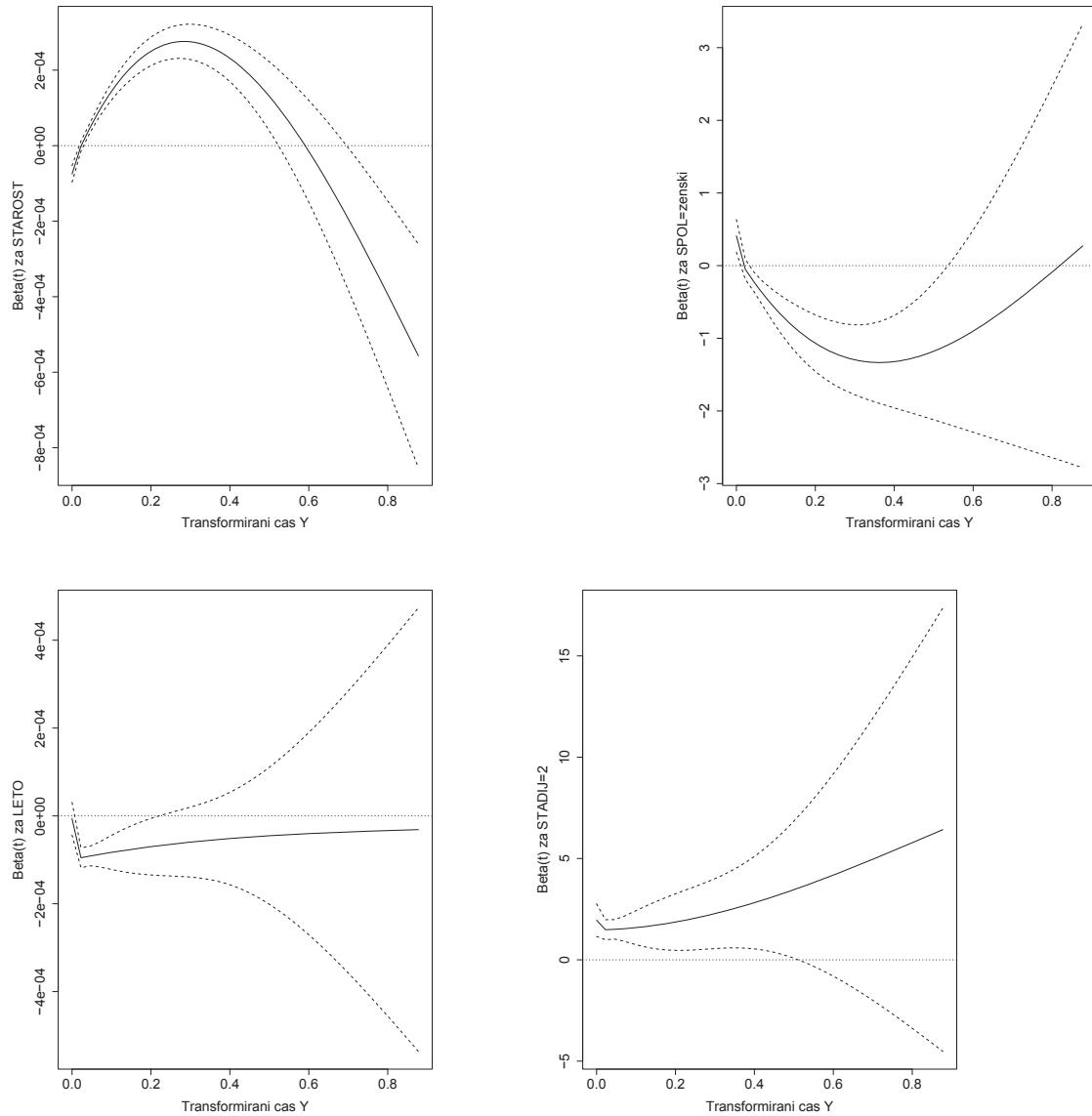
Iz grafičnega prikaza lahko razberemo, da so spremembe učinkov v času prisotne pri vseh spremenljivkah. Pri starosti in spolu lahko opazimo tudi, da se predznak učinkov spremeni.

Eden od načinov, s katerim lahko obidemo nesorazmerne ogroženosti v Coxovem modelu, je, da predpostavimo, da predpostavka o sorazmernih ogroženostih vsaj približno velja na krajših časovnih intervalih. V ta namen lahko za vsako neodvisno spremenljivko vpeljemo niz novih spremenljivk, ki imajo vrednosti spremenljivke na izbranem časovnem intervalu in vrednost nič drugje:

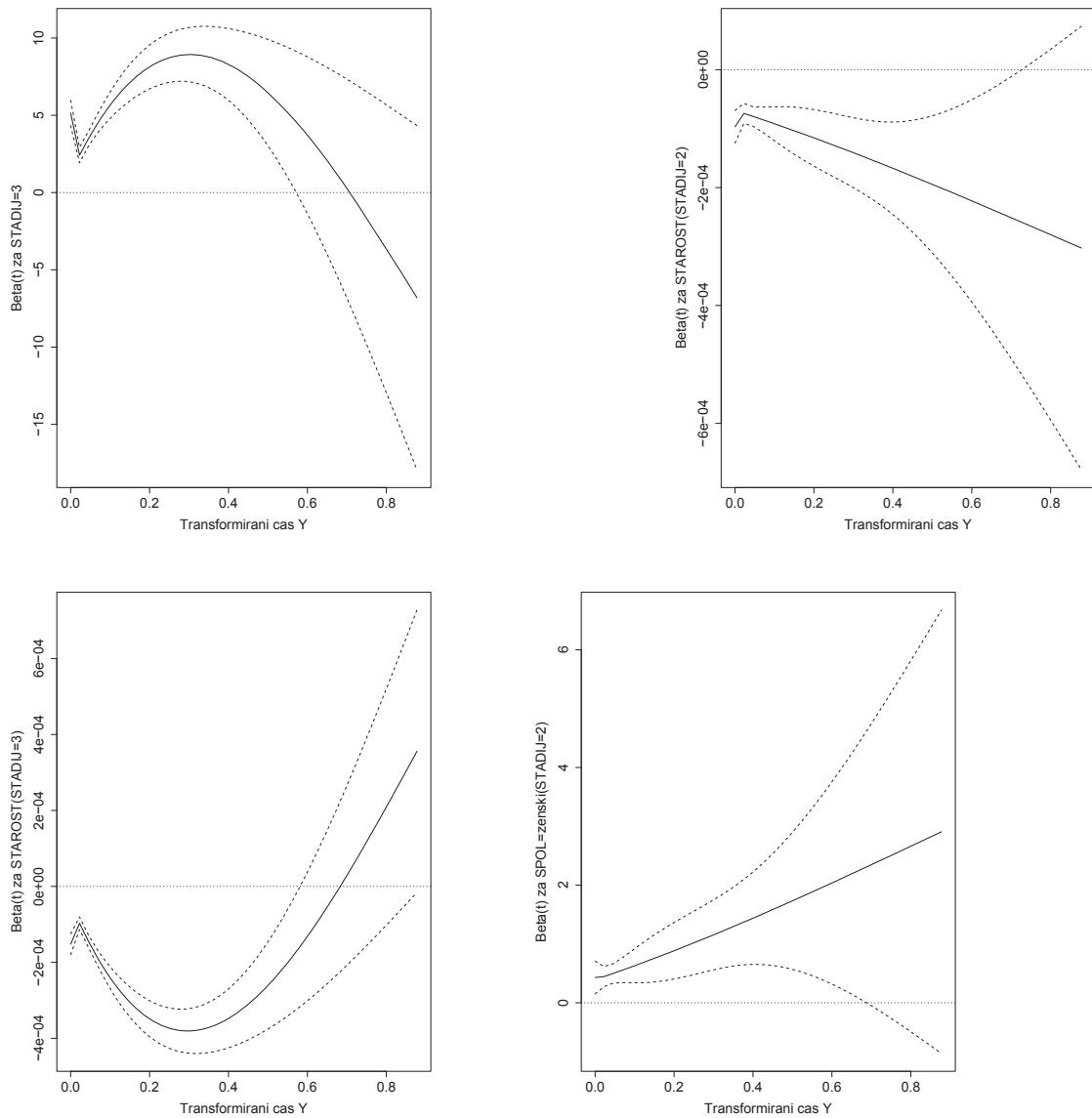
$$X : X_i = \begin{cases} X & (a_i < Y \leq b_i), i = 1, \dots, g \\ 0 & \text{drugje} \end{cases}$$

kjer je $a_1 = 0$ in g število podintervalov proučevanega časa Y . Velja tudi $a_{i+1} > b_i$.

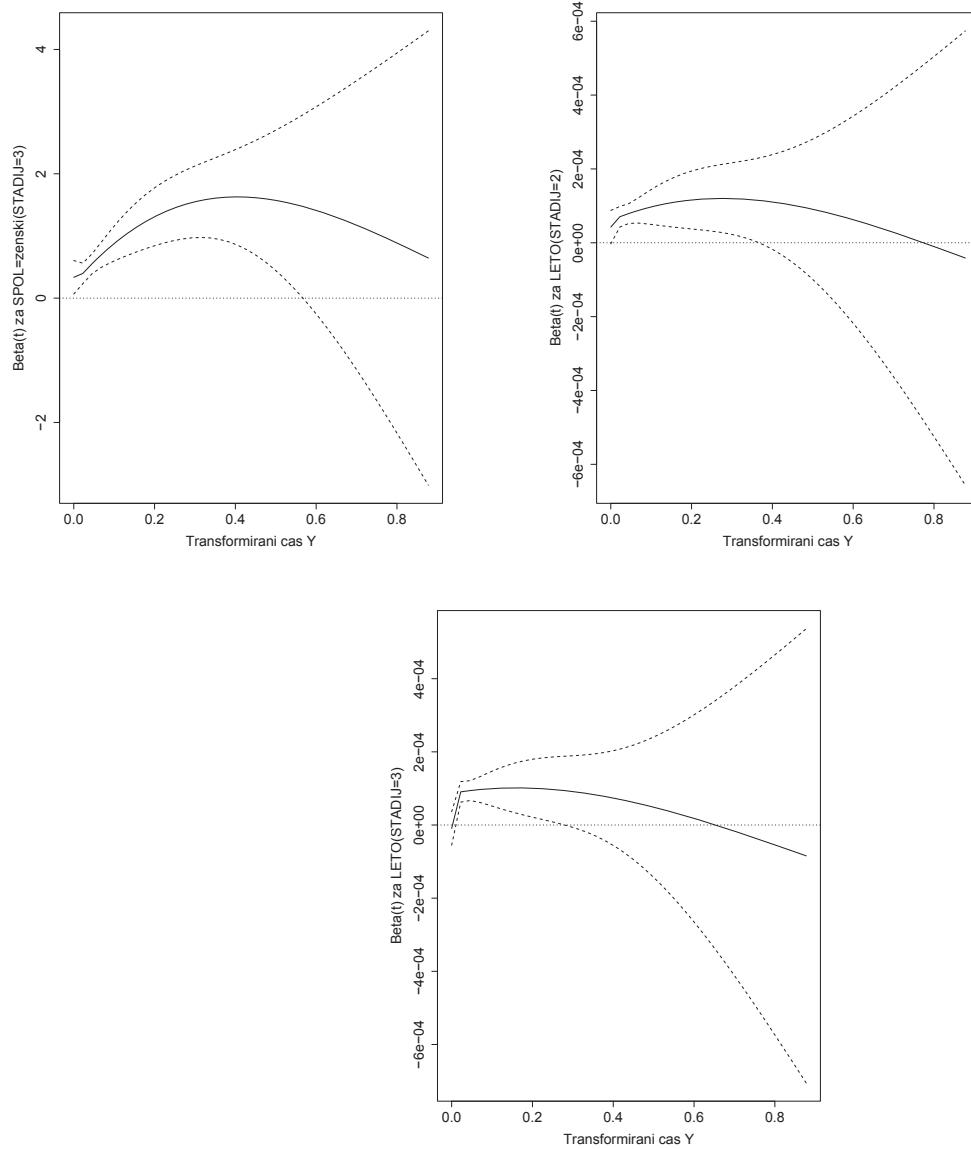
Slika 6.37: Zglajene krivulje vrednosti $U_{kj}^* + \hat{\beta}_j$ v času Y za podatke o pljučnem raku z vrisanimi mejami $\pm 2 \times$ standardna odklona



Slika 6.38: Nadaljevanje slike 6.37



Slika 6.39: Nadaljevanje slike 6.37



Na podlagi grafičnega prikaza sprememb učinkov v času in preizkušanja različnih možnosti izbora časovnih podintervalov je bil kot končni model izbran model, v katerem so vse neodvisne sremenljivke modelirane na šestih časovnih podintervalih: $0 < Y \leq 0,0025$, $0,0025 < Y \leq 0,005$, $0,005 < Y \leq 0,01$, $0,01 < Y \leq 0,025$, $0,025 < Y \leq 0,05$ in $0,05 < Y \leq 1$. Rezultati prileganja modela so prikazani v tabeli 6.6, ki kaže, da oba testa za preverjanje predpostavke o sorazmernih ogroženostih kažeta na ustrezeno prileganje izbranega modela.

Tabela 6.6: Testne statistike za preverjanje predpostavke o sorazmernih ogroženostih

	Test 1			Test 2	
	rho	chisq	p	max	p
STAROST	0.006	0.537	0.464	0.875	0.428
SPOL=ženski	0.006	0.608	0.436	0.561	0.911
LETO	-0.007	0.858	0.354	0.860	0.451
STADIJ=2	0.004	0.223	0.636	0.606	0.856
STADIJ=3	0.000	0.001	0.972	0.454	0.986
STAROST(STADIJ=2)	-0.005	0.306	0.580	0.506	0.960
STAROST(STADIJ=3)	-0.003	0.146	0.703	0.449	0.988
SPOL=ženski(STADIJ=2)	-0.002	0.093	0.761	0.611	0.850
SPOL=ženski(STADIJ=3)	-0.003	0.188	0.665	0.564	0.908
LETO(STADIJ=2)	-0.026	10.954	0.001	0.607	0.855
LETO(STADIJ=3)	-0.010	1.492	0.222	0.600	0.864
GLOBAL		14.480	0.208	0.675	0.753

Rezultati ocenjenega modela so prikazani v tabelah 6.7 in 6.8. Za hitrejši in enostavnejši pregled pa so na slikah 6.40, 6.41 in 6.42 rezultati prikazani tudi v grafični obliki. Čeprav je oblika sprememb koeficentov stopničaste oblike, so (zaradi ozkih podintervalov v začetnem obdobju) vrednosti koeficentov med časovnimi podintervalli zrisane zvezno, vrednosti koeficientov pa so vrisane na sredini ustreznega časovnega podin-

tervala.

Na podlagi rezultatov modela lahko razberemo naslednje:

- Starost v začetnem obdobju zmanjšuje ogroženost za smrt za pljučnim rakom, kasneje pa se ogroženost s starostjo povečuje. Prehod iz negativnega v pozitiven vpliv se zgodi na podintervalu ($0,025 < Y \leq 0,05$), na katerem ocenjeni koeficient ni statistično značilen. V začetnem obdobju umrljivost zaradi pljučnega raka tako bolj prizadane mlajše, kasneje pa starejše paciente. Vidimo lahko, da se na prvem intervalu ($0 < Y \leq 0,0025$) ogroženost za smrt na vsakih 10 let starosti zmanjša za faktor 0,659 ($\exp(-1,14e - 04 \times 3652,5) = 0,659$) oziroma za 34 %, medtem ko se na intervalu ($Y > 0,05$) ogroženost za smrt na vsakih 10 let starosti poveča za 36 %.
- Ocenjeni koeficienti za spol podobno kot pri starosti kažejo na spremembo predznaka učinka v času. Tako je ogroženost za smrt pri ženskah v začetnem obdobju večja kot pri moških, po času $Y = 0,05$ pa se ogroženost za smrt pri ženskah v primerjavi z moškimi zmanjša. Na časovnem intervalu $0 < Y \leq 0,0025$ je ocenjeni koeficient za spol enak 0,475, kar predstavlja za skoraj 61 % večjo ogroženost za smrt žensk v primerjavi z moškimi. Na intervalu ($Y > 0,05$) je ocenjeni koeficient za spol statistično značilno negativen. V tem obdobju je ogroženost za smrt pri ženskah za okoli 30% manjša kot pri moških.
- Glede na ocnjene koeficiente leto diagnoze v začetnem obdobju ($Y \leq 0,005$) nima statistično značilnega vpliva na ogroženost za smrt, po tem obdobju pa se ogroženost za smrt za pljučnim rakom v času zmanjšuje. Koeficienti se v tem obdobju bistveno ne razlikujejo med seboj, ogroženost za smrt pa se na vsakih deset let zmanjša za slabih 30 %.
- Ogroženost za smrt se za paciente, diagnosticirane v drugem stadiju, v primerjavi s pacienti, diagnosticiranimi v prvem stadiju pljučnega raka, poveča za približno 5-krat in po času $Y = 0,05$ izgubi statistično značilen vpliv. Pri pacientih, diagnosticiranih v tretjem stadiju pljučnega raka, se ogroženost v primerjavi s paciente, diagnosticiranimi v prvem stadiju, v začetnem obdobju ($Y < 0,01$) poveča za približno 10-krat, kasneje pa za približno 5-krat.

- Interakcije starosti s stadijem kažejo, da se v drugem stadiju okrepi negativni vpliv starosti na ogroženost, ki po času $Y = 0,05$ izzveni, v tretjem stadiju pa je krepitev negativnega vpliva starosti prisotna skozi celotno obdobje. Okrepitev negativnega vpliva starosti na ogroženost v višjih stadijih odraža še višje izgubljanje prednosti v preživetju mlajših pacientov pred starejšimi. Med vrednostmi statistično značilnih koeficientov ni bistvenih razlik, zato lahko rečemo, da se ogroženost za smrt s povečanjem starosti za 10 let v drugem stadiju v primerjavi s prvim zmanjša za približno 25 %, v tretjem stadiju pa za približno 20 %.

Pri interakciji spola s stadijem se pozitivni vpliv spola na ogroženost okrepi v drugem stadiju, ta vpliv pa izzveni po času $Y = 0,025$. Umrljivost žensk v primerjavi z umrljivostjo moških je v drugem stadiju večja kot v prvem stadiju, in sicer za približno 50 %. V tretjem stadiju so ocenjeni koeficienti za spol statistično neznačilni.

Interakcija leta diagnoze s stadijem kaže, da se negativni vpliv leta diagnoze v višjih stadijih zmanjšuje (ocenjeni koeficienti so pozitivni), kar pomeni, da se zdravljenje pacientov, diagnosticiranih v drugem in tretjem stadiju pljučnega raka, s časom izboljšuje v manjši meri kot pri pacientih, diagnosticiranih v prvem stadiju.

Tabela 6.7: Rezultati ocenjenega modela za podatke o pljučnem raku

	$\hat{\beta}$	$\hat{\beta} - 1.96\sqrt{S.E.}$	$\hat{\beta} + 1.96\sqrt{S.E.}$	S.E.	Z	Pr(> Z)
STAROST _{0.0025}	-1.140e-04	-1.400e-04	-8.750e-05	1.350e-05	-8.440	0.000
STAROST _{0.005}	-7.870e-05	-1.090e-04	-4.830e-05	1.550e-05	-5.080	0.000
STAROST _{0.01}	-4.340e-05	-6.830e-05	-1.850e-05	1.270e-05	-3.420	0.001
STAROST _{0.025}	-2.960e-05	-4.770e-05	-1.150e-05	9.240e-06	-3.200	0.001
STAROST _{0.05}	1.560e-05	-5.730e-06	3.700e-05	1.090e-05	1.430	0.152
STAROST ₁	8.480e-05	6.640e-05	1.030e-04	9.390e-06	9.030	0.000
SPOL=ž _{0.0025}	0.475	0.211	0.738	0.134	3.540	0.000
SPOL=ž _{0.005}	0.256	-0.049	0.560	0.155	1.650	0.100
SPOL=ž _{0.01}	0.268	0.020	0.515	0.126	2.120	0.034
SPOL=ž _{0.025}	0.058	-0.135	0.250	0.098	0.585	0.559
SPOL=ž _{0.05}	-0.207	-0.453	0.038	0.125	-1.650	0.098
SPOL=ž ₁	-0.373	-0.565	-0.181	0.098	-3.800	0.000
LETO _{0.0025}	1.760e-05	-3.350e-05	6.870e-05	2.610e-05	0.674	0.500
LETO _{0.005}	-2.290e-05	-7.710e-05	3.120e-05	2.760e-05	-0.830	0.406
LETO _{0.01}	-8.700e-05	-1.310e-04	-4.350e-05	2.220e-05	-3.920	0.000
LETO _{0.025}	-7.490e-05	-1.060e-04	-4.400e-05	1.580e-05	-4.750	0.000
LETO _{0.05}	-1.030e-04	-1.390e-04	-6.770e-05	1.820e-05	-5.690	0.000
LETO ₁	-9.190e-05	-1.220e-04	-6.180e-05	1.540e-05	-5.970	0.000
S=2 _{0.0025}	1.750	0.762	2.740	0.505	3.470	0.001
S=2 _{0.005}	1.700	0.620	2.780	0.552	3.080	0.002
S=2 _{0.01}	2.110	1.220	2.990	0.453	4.650	0.000
S=2 _{0.025}	1.460	0.777	2.130	0.346	4.200	0.000
S=2 _{0.05}	1.210	0.369	2.050	0.428	2.820	0.005
S=2 ₁	-0.037	-0.812	0.738	0.396	-0.093	0.926
S=3 _{0.0025}	3.240	2.340	4.130	0.456	7.100	0.000
S=3 _{0.005}	2.090	1.040	3.140	0.535	3.910	0.000
S=3 _{0.01}	2.500	1.590	3.410	0.464	5.400	0.000
S=3 _{0.025}	1.970	1.200	2.730	0.390	5.040	0.000
S=3 _{0.05}	1.110	0.008	2.220	0.564	1.970	0.049
S=3 ₁	1.380	0.102	2.670	0.654	2.120	0.034

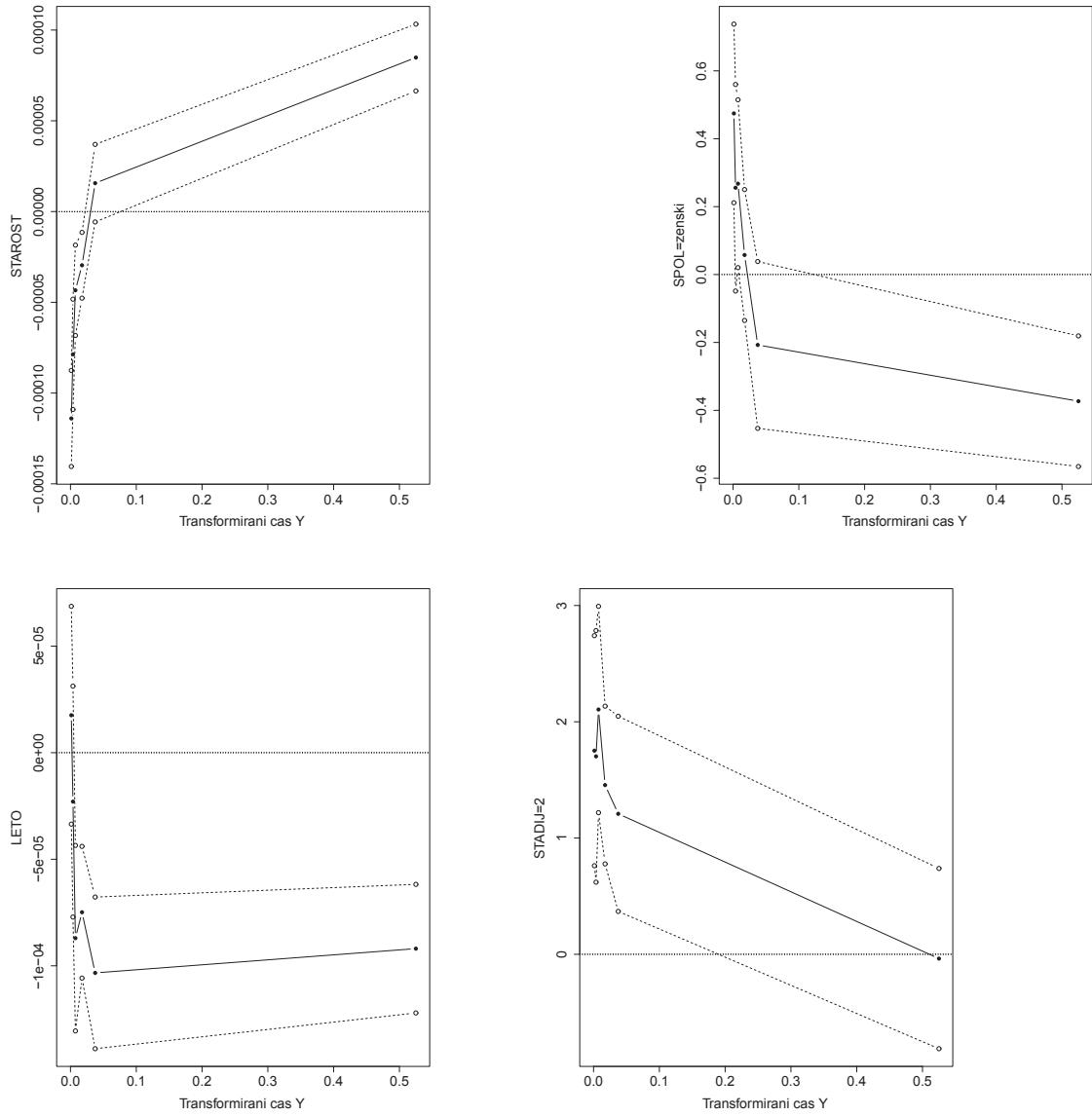
Tabela 6.8: Rezultati ocjenjenega modela za podatke o pljučnem raku - nadaljevanje tabele 6.7

	$\hat{\beta}$	$\hat{\beta} - 1.96\sqrt{S.E.}$	$\hat{\beta} + 1.96\sqrt{S.E.}$	S.E.	Z	Pr($> Z $)
STAROST(S=2) _{0.0025}	-7.270e-05	-1.050e-04	-4.020e-05	1.660e-05	-4.380	0.000
STAROST(S=2) _{0.005}	-8.350e-05	-1.200e-04	-4.670e-05	1.880e-05	-4.450	0.000
STAROST(S=2) _{0.01}	-1.070e-04	-1.370e-04	-7.640e-05	1.540e-05	-6.910	0.000
STAROST(S=2) _{0.025}	-6.690e-05	-9.000e-05	-4.380e-05	1.180e-05	-5.680	0.000
STAROST(S=2) _{0.05}	-6.350e-05	-9.200e-05	-3.500e-05	1.450e-05	-4.360	0.000
STAROST(S=2) ₁	-1.580e-05	-4.350e-05	1.200e-05	1.410e-05	-1.110	0.266
STAROST(S=3) _{0.0025}	-5.920e-05	-8.810e-05	-3.040e-05	1.470e-05	-4.020	0.000
STAROST(S=3) _{0.005}	-4.700e-05	-8.200e-05	-1.210e-05	1.780e-05	-2.640	0.008
STAROST(S=3) _{0.01}	-8.110e-05	-1.120e-04	-5.050e-05	1.560e-05	-5.200	0.000
STAROST(S=3) _{0.025}	-7.050e-05	-9.620e-05	-4.480e-05	1.310e-05	-5.370	0.000
STAROST(S=3) _{0.05}	-6.140e-05	-9.770e-05	-2.520e-05	1.850e-05	-3.320	0.001
STAROST(S=3) ₁	-7.100e-05	-1.110e-04	-3.060e-05	2.060e-05	-3.440	0.001
SPOL=ž(S=2) _{0.0025}	0.296	-0.013	0.606	0.158	1.880	0.061
SPOL=ž(S=2) _{0.005}	0.430	0.079	0.782	0.179	2.400	0.017
SPOL=ž(S=2) _{0.01}	0.530	0.242	0.817	0.147	3.610	0.000
SPOL=ž(S=2) _{0.025}	0.385	0.147	0.623	0.121	3.170	0.002
SPOL=ž(S=2) _{0.05}	0.316	-0.009	0.642	0.166	1.900	0.057
SPOL=ž(S=2) ₁	0.257	-0.037	0.551	0.150	1.720	0.086

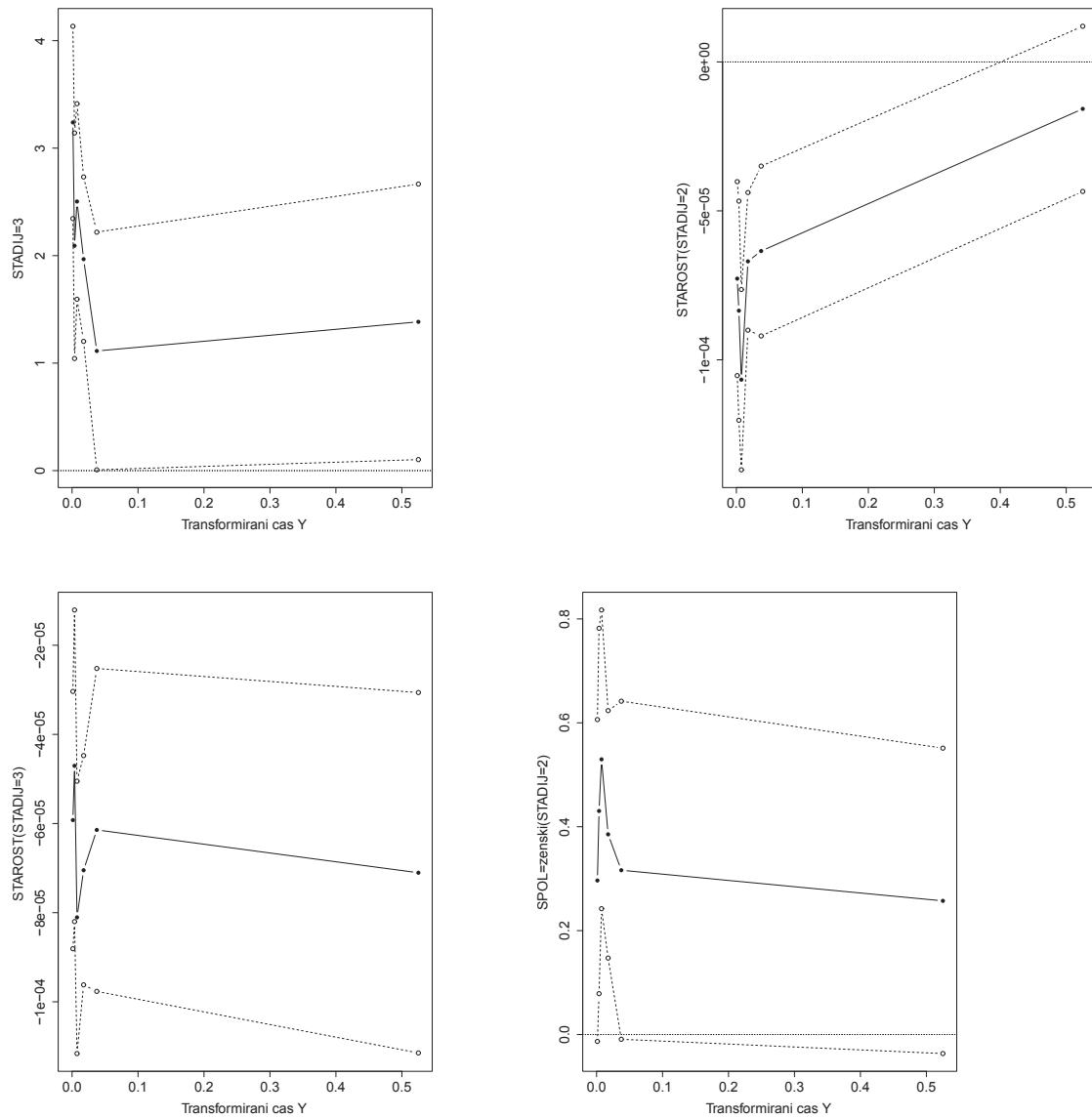
Tabela 6.9: Rezultati ocjenjenega modela za podatke o pljučnem raku - nadaljevanje tabele 6.7

	$\hat{\beta}$	$\hat{\beta} - 1.96\sqrt{S.E.}$	$\hat{\beta} + 1.96\sqrt{S.E.}$	S.E.	Z	$Pr(> Z)$
SPOL=ž(S=3) _{0.0025}	0.013	-0.268	0.294	0.143	0.090	0.928
SPOL=ž(S=3) _{0.005}	0.320	-0.015	0.655	0.171	1.870	0.061
SPOL=ž(S=3) _{0.01}	0.183	-0.106	0.472	0.147	1.240	0.214
SPOL=ž(S=3) _{0.025}	0.203	-0.049	0.455	0.128	1.580	0.114
SPOL=ž(S=3) _{0.05}	0.089	-0.291	0.469	0.194	0.457	0.648
SPOL=ž(S=3) ₁	0.085	-0.335	0.504	0.214	0.395	0.693
LETO(S=2) _{0.0025}	2.610e-06	-5.720e-05	6.240e-05	3.050e-05	0.086	0.932
LETO(S=2) _{0.005}	4.540e-05	-1.720e-05	1.080e-04	3.190e-05	1.420	0.155
LETO(S=2) _{0.01}	6.850e-05	1.810e-05	1.190e-04	2.570e-05	2.670	0.008
LETO(S=2) _{0.025}	5.130e-05	1.420e-05	8.840e-05	1.890e-05	2.710	0.007
LETO(S=2) _{0.05}	7.420e-05	2.990e-05	1.190e-04	2.260e-05	3.280	0.001
LETO(S=2) ₁	7.460e-05	3.290e-05	1.160e-04	2.130e-05	3.500	0.000
LETO(S=3) _{0.0025}	-3.220e-05	-8.680e-05	2.240e-05	2.790e-05	-1.160	0.248
LETO(S=3) _{0.005}	2.430e-05	-3.610e-05	8.470e-05	3.080e-05	0.787	0.431
LETO(S=3) _{0.01}	5.870e-05	8.210e-06	1.090e-04	2.570e-05	2.280	0.023
LETO(S=3) _{0.025}	7.320e-05	3.320e-05	1.130e-04	2.040e-05	3.590	0.000
LETO(S=3) _{0.05}	1.210e-04	6.630e-05	1.760e-04	2.790e-05	4.340	0.000
LETO(S=3) ₁	1.060e-04	4.250e-05	1.700e-04	3.270e-05	3.260	0.001

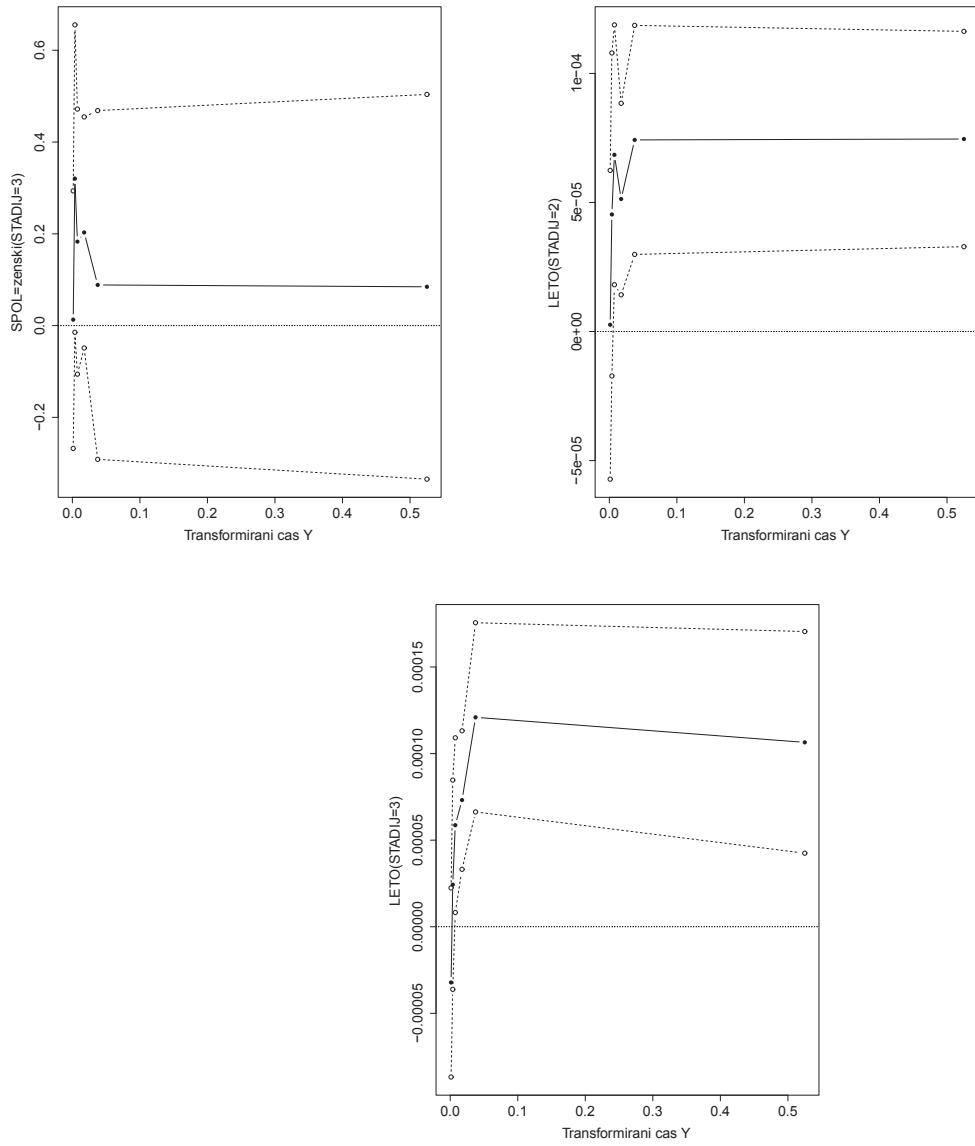
Slika 6.40: Grafični prikaz ocenjenih koeficientov Coxovega modela za podatke o pljučnem raku



Slika 6.41: Nadaljevanje slike 6.40



Slika 6.42: Nadaljevanje slike 6.40

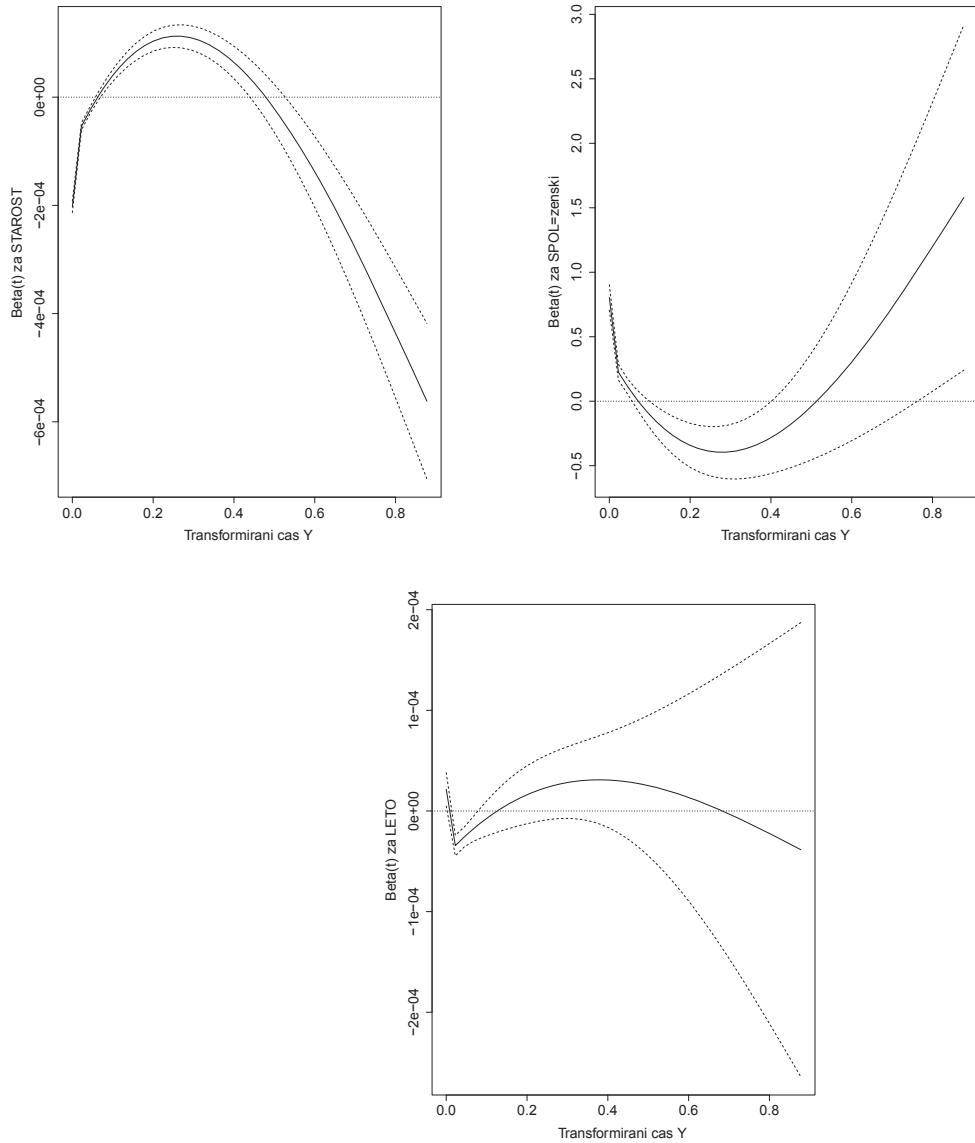


V nadaljevanju prikazujemo še rezultate enostavnega Coxovega modela za podatke o pljučnem raku, v katerega so vključeni le glavni vplivi izbranih neodvisnih spremenljivk (starost, spol, leto diagnoze in stadij, v katerem je bil bolnik z rakom na pljučih diagnosticiran). Predpostavko o sorazmernosti ogroženosti celotnega modela z nizkim tveganjem zavrneta globalni statistiki obeh izbranih testov, medtem ko prvi test pri letu diagnoze ne zazna kršitve (tabela 6.10). Sliki 6.43 in 6.44 kažeta zglajene vrednosti Schoenfeldovih ostankov (oziroma $U_{kj}^* + \hat{\beta}_j$) v času Y, ki nakazujejo moč in obliko spremenljajočih učinkov. Podobno kot pri modelu z interakcijami je tudi v enostavnem modelu iz grafičnega prikaza zglajenih Schoenfeldovih ostankov mogoče razbrati spremembo predznaka učinka starosti in spola v času Y.

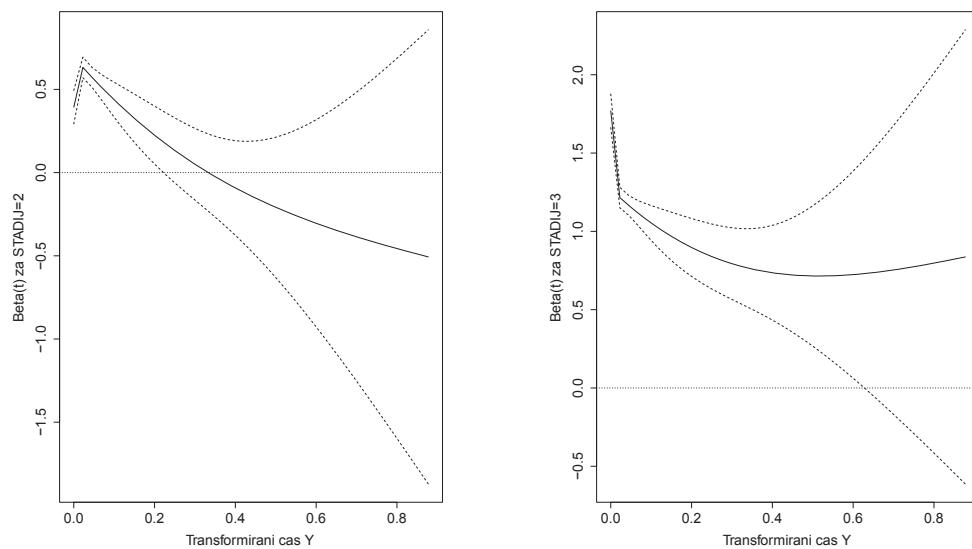
Tabela 6.10: Testne statistike za preverjanje predpostavke sorazmernih ogroženosti

	Test 1			Test 2	
	rho	chisq	p	max	p
STAROST	0.151	367.845	0.000	11.965	0.000
SPOL=ženski	-0.064	66.394	0.000	5.136	0.000
LETO	0.009	1.156	0.282	1.506	0.021
STADIJ=2	-0.030	14.583	0.000	4.650	0.000
STADIJ=3	-0.052	42.845	0.000	5.354	0.000
GLOBAL		433.574	0.000	11.370	0.000

Slika 6.43: Zglajene krivulje vrednosti $U_{kj}^* + \hat{\beta}_j$ v času Y za podatke pljučnega raka z vrisanimi mejami $\pm 2 \times$ standardna odklona



Slika 6.44: Nadaljevanje slike 6.43



Spreminjajče učinke neodvisnih spremenljivk v času smo tudi v enostavnem Coxovem modelu modelirali s pomočjo niza novo uvedenih spremenljivk na šestih časovnih podintervalih: $0 < Y \leq 0,0025$, $0,0025 < Y \leq 0,005$, $0,005 < Y \leq 0,01$, $0,01 < Y \leq 0,025$, $0,025 < Y \leq 0,05$, in $0,05 < Y \leq 1$. Rezultati prileganja modela, ki so prikazani v tabeli 6.11, kažejo na ustrezno prileganje izbranega modela.

Tabela 6.11: Testne statistike za preverjanje predpostavke sorazmernih ogroženosti

	Test 1			Test 2	
	rho	chisq	p	max	p
STAROST	0.002	0.040	0.842	0.883	0.416
SPOL=ženski	0.003	0.109	0.741	0.580	0.890
LETO	0.008	0.917	0.338	0.842	0.478
STADIJ=2	-0.018	5.440	0.020	0.616	0.842
STADIJ=3	-0.012	2.236	0.135	0.539	0.934
GLOBAL		6.690	0.245	0.662	0.774

Ocenjene statistike izbranega končnega enostavnega Coxovega modela za podatke o pljučnem raku so prikazane v tabeli 6.12. Za hitrejši in enostavnejši pregled so ocenjeni koeficienti modela na slikah 6.45 in 6.46 prikazani tudi grafično, pri čemer so kljub dejansko stopničasti obliko koeficientov ti zrisani zvezno z vrednostmi koeficientov na sredini ustreznega časovnega podintervala. Na podlagi ocenjenih koeficientov modela lahko rečemo naslednje:

- S starostjo se ogroženost za smrt za pljučnim rakom najprej zmanjšuje, po času $Y = 0,05$ pa ogroženost s starostjo začne naraščati. Negativni vpliv starosti na ogroženost, ki dejansko predstavlja izgubljanje prednosti v preživetju mlajših pred starejšimi, je najbolj izrazit pri začetnih vrednostih transformiranega časa Y , kasneje pa se negativni učinek starosti zmanjšuje in po času $Y = 0,05$ preide v pozitivnega. Ocenjeni vpliv starosti na prvem intervalu ($Y < 0,0025$) kaže, da se ogroženost s povečanjem starosti za 10 let zmanjša za približno 45 % ($\exp(-1,68e - 04 * 3652,5) = 0,541$), na intervalu $Y > 0,05$ pa se ogroženost na vsakih 10 let

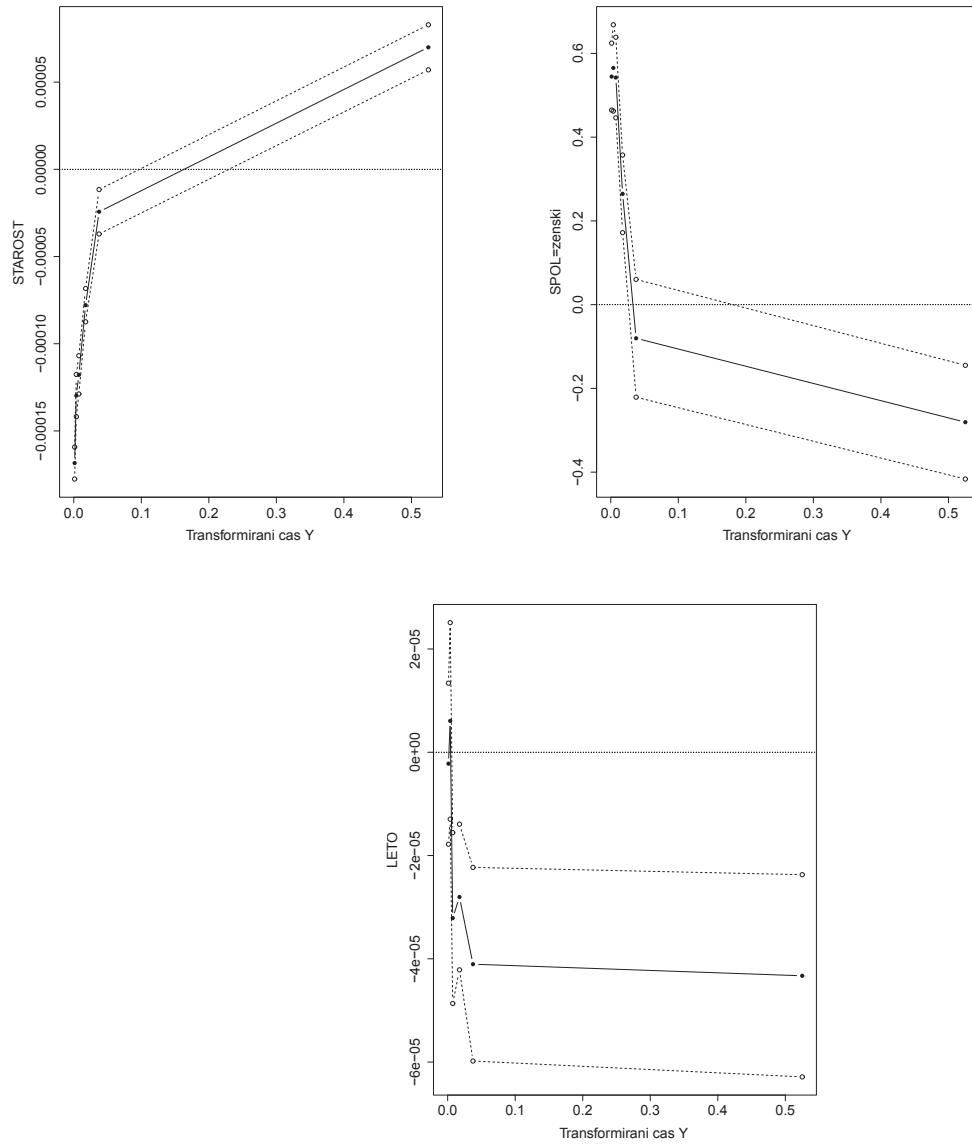
starosti poveča za slabih 30 %.

- Vpliv spola na ogroženost za smrt kaže, da je ogroženost pri ženskah v primerjavi z moškimi do časa $Y < 0,01$ večja za dobrih 70 %, v kasnejšem obdobju se ocenjeni vpliv spola zmanjšuje in na koncu preide v negativnega. Tako se po času $Y = 0,05$ ogroženost za smrt žensk v primerjavi z moškimi zmanjša za približno 25 %.
- Ocjenjeni koeficienti za leto diagnoze so do časa $Y \leq 0,005$ statistično neznačilni, kasneje pa postanejo statistično značilno negativni. V tem obdobju se ogroženost za smrt za pljučnim rakom na vsakih deset let zmanjša za približno 10-15 %.
- Ogroženost za smrt je v višjih stadijih statistično značilno večje kot v prvem stadiju skozi celotno obdobje. V drugem stadiju je vpliv na ogroženost na začetku manjši (ogroženost za smrt se v primerjavi s pacienti, diagnosticiranimi v prvem stadiju, poveča za približno 35 %), kasneje pa se ogroženost za smrt patientov, diagnosticiranimi v drugem stadiju, v primerjavi s tistimi, diagnosticiranimi v prvem stadiju, poveča za slabih 80 %.
Ogroženost za smrt se za paciente, diagnosticirane v tretjem stadiju pljučnega raka, v primerjavi s pacienti, diagnosticiranimi v prvem stadiju pljučnega raka, do časa $Y = 0,01$ poveča za približno 4-krat, kasneje pa za približno 3-krat.

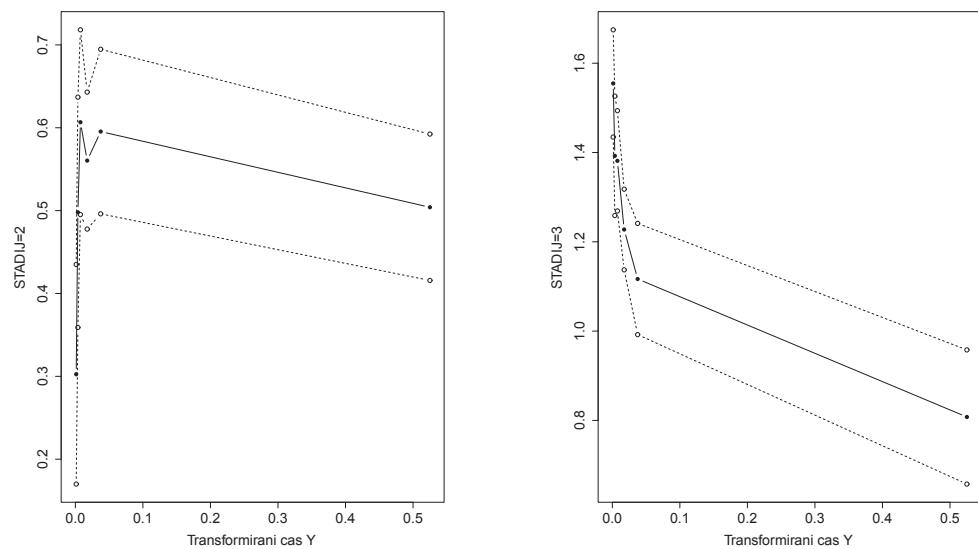
Tabela 6.12: Rezultati ocenjenega modela za podatke o pljučnem raku

	$\hat{\beta}$	$\hat{\beta} - 1.96\sqrt{S.E.}$	$\hat{\beta} + 1.96\sqrt{S.E.}$	S.E.	Z	Pr($> Z $)
STAROST _{0.0025}	-1.680e-04	-1.780e-04	-1.590e-04	4.670e-06	-36.000	0.000
STAROST _{0.005}	-1.300e-04	-1.420e-04	-1.180e-04	6.170e-06	-21.000	0.000
STAROST _{0.01}	-1.180e-04	-1.290e-04	-1.070e-04	5.600e-06	-21.000	0.000
STAROST _{0.025}	-7.790e-05	-8.740e-05	-6.830e-05	4.870e-06	-16.000	0.000
STAROST _{0.05}	-2.430e-05	-3.710e-05	-1.160e-05	6.500e-06	-3.740	0.000
STAROST ₁	6.990e-05	5.700e-05	8.290e-05	6.600e-06	10.600	0.000
SPOL=ženski _{0.0025}	0.545	0.465	0.625	0.041	13.300	0.000
SPOL=ženski _{0.005}	0.565	0.462	0.668	0.053	10.700	0.000
SPOL=ženski _{0.01}	0.542	0.446	0.639	0.049	11.000	0.000
SPOL=ženski _{0.025}	0.265	0.172	0.357	0.047	5.600	0.000
SPOL=ženski _{0.05}	-0.080	-0.221	0.060	0.072	-1.120	0.263
SPOL=ženski ₁	-0.281	-0.416	-0.145	0.069	-4.050	0.000
LETO _{0.0025}	-2.210e-06	-1.780e-05	1.340e-05	7.960e-06	-0.278	0.781
LETO _{0.005}	6.070e-06	-1.300e-05	2.510e-05	9.710e-06	0.626	0.532
LETO _{0.01}	-3.210e-05	-4.870e-05	-1.560e-05	8.440e-06	-3.800	0.000
LETO _{0.025}	-2.800e-05	-4.210e-05	-1.390e-05	7.200e-06	-3.890	0.000
LETO _{0.05}	-4.100e-05	-5.980e-05	-2.230e-05	9.570e-06	-4.290	0.000
LETO ₁	-4.330e-05	-6.290e-05	-2.370e-05	9.990e-06	-4.330	0.000
STADIJ=2 _{0.0025}	0.303	0.170	0.435	0.068	4.470	0.000
STADIJ=2 _{0.005}	0.498	0.359	0.637	0.071	7.020	0.000
STADIJ=2 _{0.01}	0.607	0.495	0.718	0.057	10.700	0.000
STADIJ=2 _{0.025}	0.560	0.478	0.643	0.042	13.300	0.000
STADIJ=2 _{0.05}	0.595	0.496	0.695	0.051	11.800	0.000
STADIJ=2 ₁	0.504	0.416	0.592	0.045	11.200	0.000
STADIJ=3 _{0.0025}	1.550	1.430	1.670	0.061	25.400	0.000
STADIJ=3 _{0.005}	1.390	1.260	1.530	0.068	20.400	0.000
STADIJ=3 _{0.01}	1.380	1.270	1.490	0.057	24.100	0.000
STADIJ=3 _{0.025}	1.230	1.140	1.320	0.046	26.700	0.000
STADIJ=3 _{0.05}	1.120	0.992	1.240	0.064	17.600	0.000
STADIJ=3 ₁	0.808	0.657	0.958	0.077	10.500	0.000

Slika 6.45: Grafični prikaz ocenjenih koeficientov Coxovega modela za podatke o pljučnem raku



Slika 6.46: Nadaljevanje slike 6.45



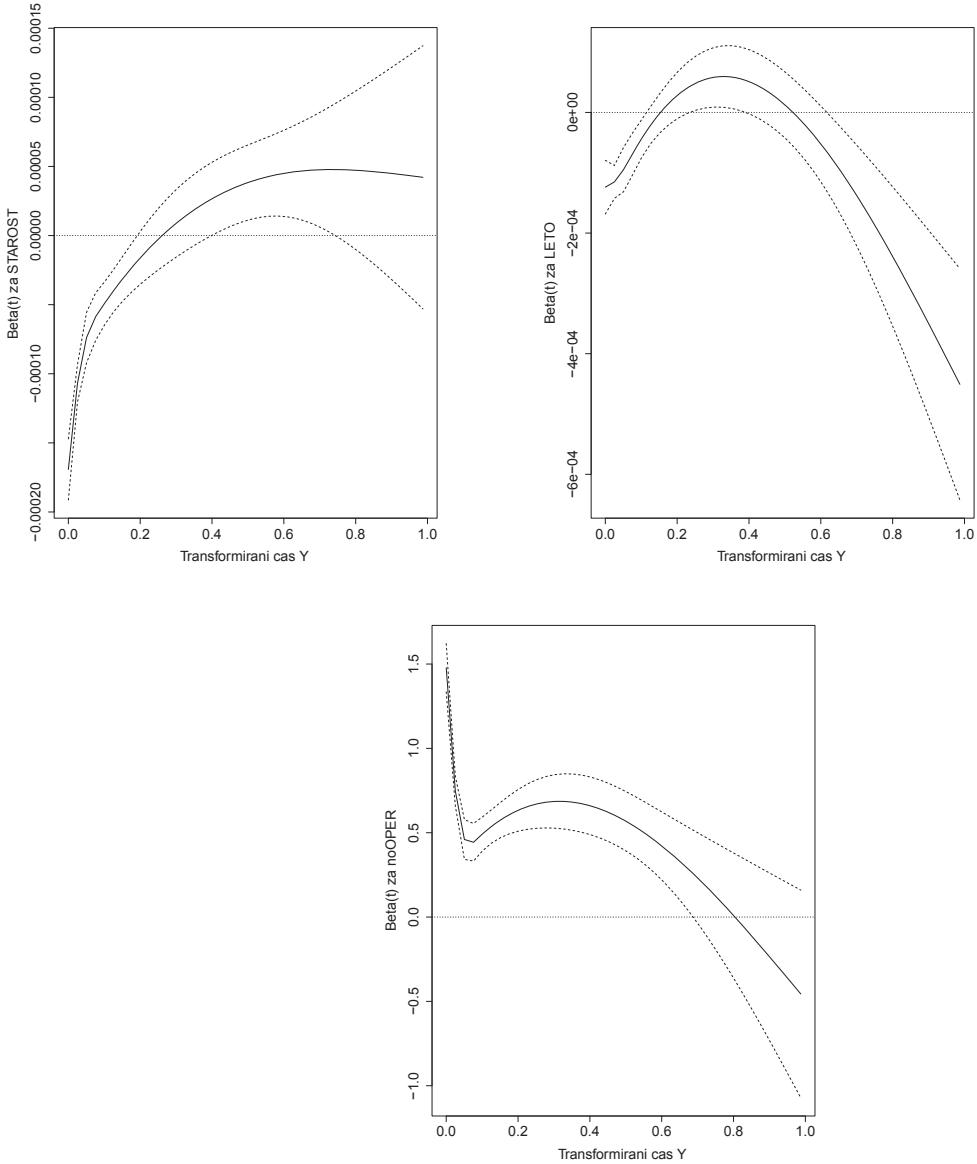
6.3.2 Coxov model na podatkih o raku na dojki

V Coxov model za podatke o raku na dojki je poleg starosti, leta diagnoze in stadija, v katerem so bile paciente z rakom na dojki diagnosticirane, vključen tudi indikator operativnega zdravljenja, ki se je v Aalenovem modelu pokazal kot pomemben dejavnik, ki vpliva na stanje statistične ozdravitve pacientek. Najprej si poglejmo model, v katerega so poleg glavnih vplivov neodvisnih spremenljivk vključene tudi interakcije starosti in leta diagnoze s stadijem raka na dojki. Rezultati prileganja modela, ki so prikazani v tabeli 6.13, kažejo na očitno kršenje predpostavke sorazmernih ogroženosti modela. Podobno lahko razberemo tudi iz grafičnega prikaza zglajenih krivulj vrednosti Schoenfeldovih ostankov, prikazanih na slikah 6.47, 6.48 in 6.49.

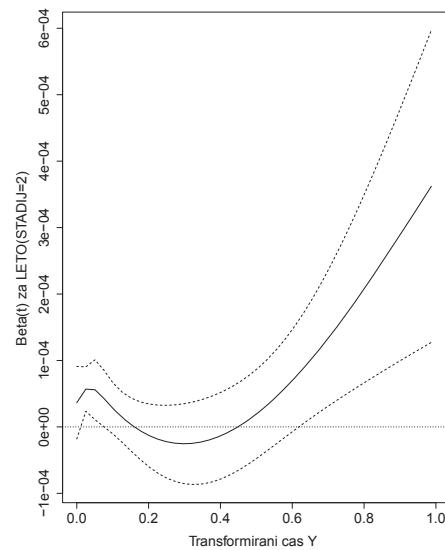
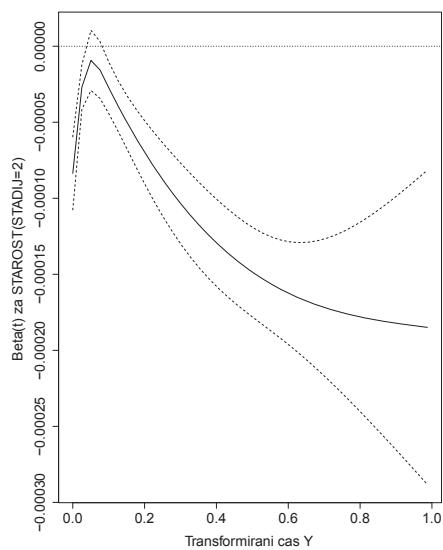
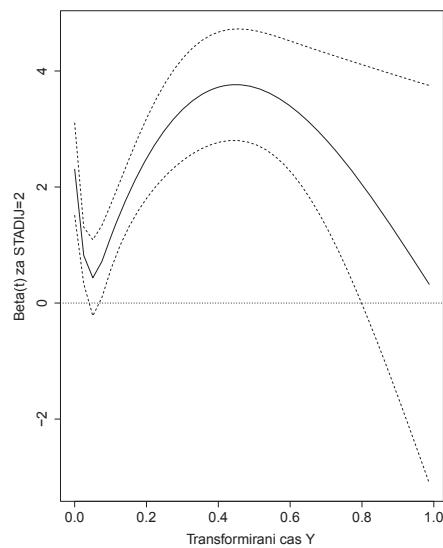
Tabela 6.13: Testne statistike za preverjanje predpostavke sorazmernih ogroženosti

	Test 1			Test 2	
	rho	chisq	p	max	p
STAROST	0.121	116.002	0.000	7.483	0.000
LETO	0.024	4.342	0.037	5.128	0.000
STADIJ=2	0.033	9.226	0.002	1.737	0.005
STADIJ=3	0.072	42.266	0.000	2.617	0.000
BREZ OPERACIJE	-0.068	38.549	0.000	3.087	0.000
STAROST(STADIJ=2)	-0.063	33.375	0.000	2.544	0.000
STAROST(STADIJ=3)	-0.085	60.459	0.000	2.811	0.000
LETO(STADIJ=2)	0.006	0.224	0.636	1.721	0.005
LETO(STADIJ=3)	-0.019	2.722	0.099	2.671	0.000
GLOBAL		223.751	0.000	7.649	0.000

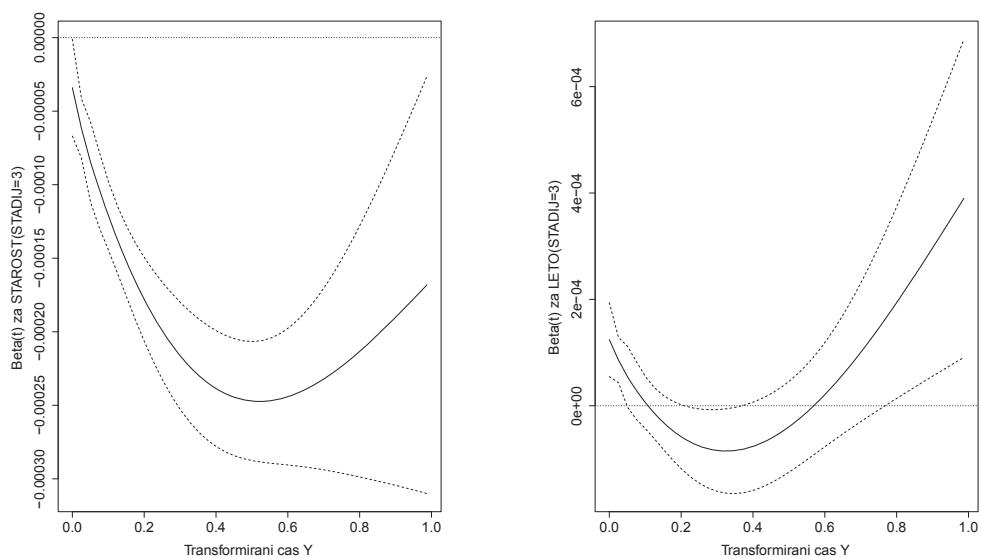
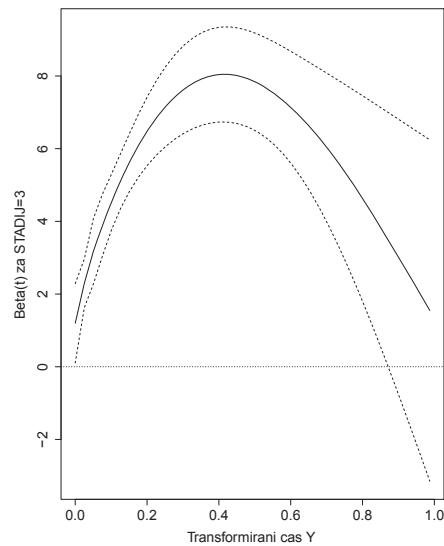
Slika 6.47: Zglajene krivulje vrednosti $U_{kj}^* + \hat{\beta}_j$ v času Y za podatke o raku na dojki z vrisanimi mejami $\pm 2 \times$ standardna odklona



Slika 6.48: Nadaljevanje slike 6.47



Slika 6.49: Nadaljevanje slike 6.47



Z vpeljavo niza novih spremenljivk, s pomočjo katerih lahko zajamemo spremištanje učinkov neodvisnih spremenljivk v času in na ta način zagotovimo veljavnost predpostavke sorazmernih ogroženosti, je bil kot končni model izbran model, v katerem so vse neodvisne spremenljivke modelirane na sedmih časovnih podintervalih: $0 < Y \leq 0,005$, $0,005 < Y \leq 0,025$, $0,025 < Y \leq 0,05$, $0,05 < Y \leq 0,10$, $0,10 < Y \leq 0,20$, $0,20 < Y \leq 0,40$ in $0,40 < Y \leq 1$. Rezultati, prikazani v tabeli 6.14, kažejo na zadovoljivo prileganje izbranega modela.

Tabela 6.14: Testne statistike za preverjanje predpostavke sorazmernih ogroženosti

	Test 1			Test 2	
	rho	chisq	p	max	p
STAROST	-0.001	0.014	0.904	1.246	0.090
LETO	0.001	0.006	0.939	0.538	0.935
STADIJ=2	-0.006	0.244	0.621	1.018	0.251
STADIJ=3	0.008	0.483	0.487	1.170	0.130
BREZ OPERACIJE	-0.002	0.026	0.871	1.617	0.011
STAROST(STADIJ=2)	-0.001	0.005	0.943	0.841	0.479
STAROST(STADIJ=3)	0.008	0.513	0.474	1.223	0.100
LETO(STADIJ=2)	0.006	0.249	0.618	0.917	0.369
LETO(STADIJ=3)	-0.003	0.097	0.755	1.162	0.134
GLOBAL		0.989	0.999	0.996	0.275

Ocenjene vrednosti parametrov zgoraj opisanega modela so prikazane v tabelah 6.15 in 6.16, za hitrejši in enostavnnejši pregled pa tudi grafično na slikah 6.50, 6.51 in 6.52. Na tem mestu naj ponovno opomnimo, da so ocenjene vrednosti parametrov, podobno kot pri podatkih o pljučnem raku, prikazane v zvezni obliki, čeprav se glede na izbrani model vrednosti parametrov v času spreminjajo stopničasto. Vrednosti parametrov so na slikah vrisane na sredini ustreznega časovnega podintervala.

Na podlagi prikazanih rezultatov lahko razberemo naslednje:

- Vpliv starosti na ogroženost za smrt je do vrednosti transformiranega časa $Y =$

0,1 negativen, kasneje pa se statistično značilen vpliv starosti na ogroženost izgubi. Vpliv starosti na ogroženost za smrt je najmočnejši v začetnem obdobju in s časom slabi. Na intervalu $Y < 0,005$ se s povečanjem starosti za 10 let ogroženost za smrt za rakom na dojki zmanjša za dobrih 50 % ($\exp(-1,96e-04 \cdot 3652.5) = 0,489$), na intervalu $0,05 < Y \leq 0,1$ pa se ogroženost za smrt s povečanjem starosti za 10 let zmanjša za slabih 20 %.

- Tudi leto diagnoze sprva negativno vpliva na ogroženost za smrt, podobno kot pri starosti pa vpliv leta diagnoze po času $Y = 0,1$ izgubi statistično značilen vpliv. Vpliv leta diagnoze je najmočnejši v začetnem obdobju in s časom slabi. Na intervalu $Y < 0,025$ se ogroženost za smrt za rakom na dojki na vsakih deset let zmanjša za približno 40 %, na intervalu $0,05 < Y \leq 0,1$ pa se ogroženost za smrt na vsakih deset let zmanjša le še za dobrih 20 %.
- Pri patientkah brez operativnega zdravljenja je ogroženost skozi celotno obdobje povečana, pri čemer je vpliv največji v začetnem obdobju in s časom slabi, po času $Y = 0,05$ pa se stabilizira na nižjih vrednostih. Na intervalu $Y < 0,005$ se ogroženost za smrt patientkam brez operativnega zdravljenja poveča za skoraj 5 krat, po času $Y = 0,05$ pa za približno 1.5-krat.
- Ocenjeni koeficienti višjih stadijev so pozitivni, kar pomeni, da se ogroženost za smrt pri patientkah, diagnosticiranih v višjih stadijih raka na dojki, v primerjavi s patientkami, diagnosticiranimi v prvem stadiju, poveča. Kljub temu lahko opazimo, da je vpliv drugega stadija v večjem delu opazovanega obdobja statistično neznačilen. Glede na ocene parametrov lahko rečemo, da se parametri do časa $Y = 0,4$ bistveno ne spreminjajo. Ogroženost patientk, diagnosticiranih v drugem stadiju raka na dojki, se v primerjavi s patientkami, diagnosticiranimi v prvem stadiju poveča za približno 2.5-krat.

Pri patientkah, diagnosticiranih v tretjem stadiju raka na dojki, se ogroženost za smrt v primerjavi s patientkami, diagnosticiranimi v prvem stadiju, do časa $Y = 0,4$ povečuje. Medtem ko je na intervalu $Y < 0,005$ ocenjeni koeficient statistično neznačilen, se na intervalu $0,005 > Y \leq 0,025$ ogroženost patientk, diagnosticiranih v tretjem stadiju, v primerjavi s patientkami, diagnosticiranimi

v prvem stadiju, poveča za faktor 2,8, kasneje pa za približno 10-krat.

- Ocenjeni koeficienti interakcij starosti s stadijem kažejo, da se negativni vpliv starosti na ogroženost za smrt v drugem stadiju na intervalu $Y < 0,005$ okrepi, nato pa postane statistično neznačilen.

V tretjem stadiju se negativni vpliv starosti na ogroženost patientk, diagnosticiranih v prvem stadiju, na časovnem intervalu $Y < 0,005$ zmanjša (ogroženost se s povečanjem starosti povečuje), kasneje postane vpliv starosti na ogroženost negativen, vendar statistično neznačilen.

Interakcije leta diagnoze s stadijem kažejo, da se negativni vpliv leta diagnoze na ogroženost za smrt v višjih stadijih v primerjavi s prvim stadijem zmanjšuje. V drugem stadiju je vpliv leta diagnoze statistično značilen na časovnem intervalu $Y < 0,005$. V tem obdobju se ogroženost patientkam, diagnosticiranim v drugem stadiju raka na dojki, na vsakih 10 let poveča za slabih 40 %, s čimer se negativni vpliv leta diagnoze na ogroženost za smrt v tem začetnem obdobju skoraj izniči. Podobno lahko opazimo, da je vpliv leta diagnoze na ogroženost pri patientkah, diagnosticiranih v tretjem stadiju, statistično značilen na časovnem intervalu $Y < 0,025$, na katerem se ogroženost na vsakih 10 let poveča za okoli 50 oziroma 60 %. To pomeni, da se ogroženost za smrt pri patientkah, diagnosticiranih v tretjem stadiju, v primerjavi s tistimi, diagnosticiranimi v prvem stadiju, v tem obdobju s časom celo povečuje.

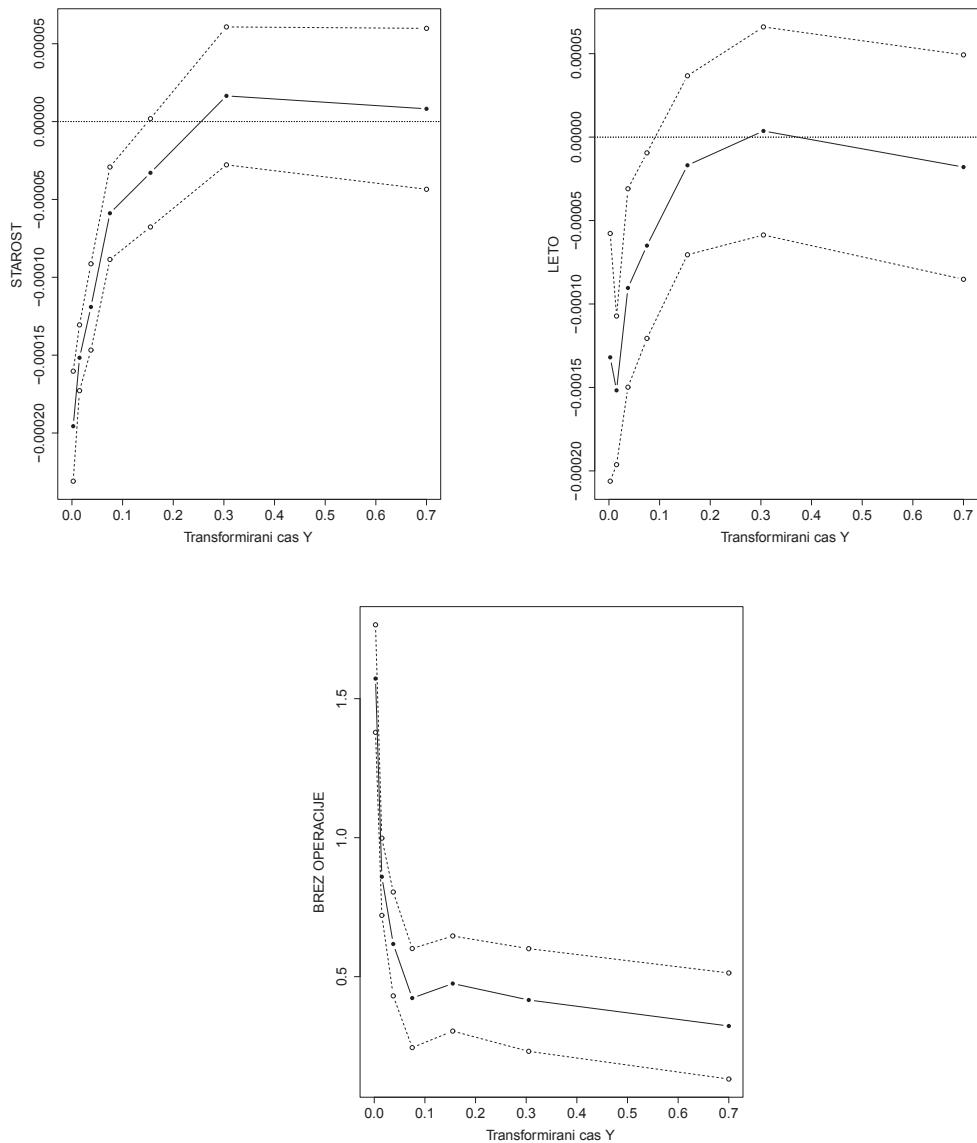
Tabela 6.15: Rezultati ocenjenega modela za podatke o raku na dojki

	$\hat{\beta}$	$\hat{\beta} - 1.96\sqrt{S.E.}$	$\hat{\beta} + 1.96\sqrt{S.E.}$	S.E.	Z	Pr($> Z $)
STAROST _{0.005}	-1.960e-04	-2.310e-04	-1.600e-04	1.800e-05	-10.900	0.000
STAROST _{0.025}	-1.520e-04	-1.730e-04	-1.310e-04	1.080e-05	-14.100	0.000
STAROST _{0.05}	-1.190e-04	-1.470e-04	-9.130e-05	1.410e-05	-8.420	0.000
STAROST _{0.1}	-5.890e-05	-8.850e-05	-2.920e-05	1.510e-05	-3.890	0.000
STAROST _{0.2}	-3.290e-05	-6.770e-05	1.890e-06	1.770e-05	-1.850	0.064
STAROST _{0.4}	1.650e-05	-2.780e-05	6.080e-05	2.260e-05	0.731	0.465
STAROST ₁	8.200e-06	-4.350e-05	5.990e-05	2.640e-05	0.311	0.756
LETO _{0.005}	-1.320e-04	-2.060e-04	-5.770e-05	3.790e-05	-3.490	0.000
LETO _{0.025}	-1.520e-04	-1.960e-04	-1.070e-04	2.270e-05	-6.680	0.000
LETO _{0.05}	-9.040e-05	-1.500e-04	-3.090e-05	3.030e-05	-2.980	0.003
LETO _{0.1}	-6.500e-05	-1.210e-04	-9.440e-06	2.840e-05	-2.290	0.022
LETO _{0.2}	-1.690e-05	-7.050e-05	3.670e-05	2.740e-05	-0.617	0.537
LETO _{0.4}	3.680e-06	-5.870e-05	6.600e-05	3.180e-05	0.116	0.908
LETO ₁	-1.800e-05	-8.520e-05	4.930e-05	3.430e-05	-0.524	0.601
S=2 _{0.005}	0.968	-0.262	2.200	0.627	1.540	0.123
S=2 _{0.025}	1.050	0.353	1.760	0.358	2.940	0.003
S=2 _{0.05}	0.274	-0.623	1.170	0.458	0.599	0.549
S=2 _{0.1}	0.389	-0.526	1.300	0.467	0.832	0.405
S=2 _{0.2}	1.240	0.188	2.300	0.538	2.310	0.021
S=2 _{0.4}	1.200	-0.213	2.620	0.724	1.670	0.096
S=2 ₁	-0.253	-2.080	1.580	0.934	-0.271	0.787
S=3 _{0.005}	-0.166	-1.420	1.090	0.639	-0.259	0.795
S=3 _{0.025}	1.030	0.145	1.920	0.452	2.280	0.023
S=3 _{0.05}	2.800	1.360	4.230	0.732	3.820	0.000
S=3 _{0.1}	2.560	0.691	4.420	0.953	2.680	0.007
S=3 _{0.2}	2.380	-0.265	5.030	1.350	1.760	0.078
S=3 _{0.4}	6.700	2.270	11.100	2.260	2.970	0.003
S=3 ₁	1.540	-9.840	12.900	5.810	0.266	0.790
BREZ OPER. _{0.005}	1.570	1.380	1.770	0.099	15.900	0.000
BREZ OPER. _{0.025}	0.859	0.720	0.998	0.071	12.100	0.000
BREZ OPER. _{0.05}	0.618	0.431	0.804	0.095	6.480	0.000
BREZ OPER. _{0.1}	0.423	0.245	0.601	0.091	4.650	0.000
BREZ OPER. _{0.2}	0.475	0.304	0.646	0.087	5.440	0.000
BREZ OPER. _{0.4}	0.416	0.231	0.601	0.094	4.410	0.000
BREZ OPER. ₁	0.322	0.132	0.513	0.097	3.310	0.001

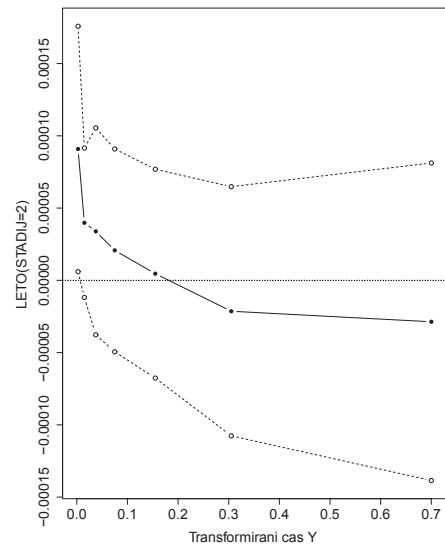
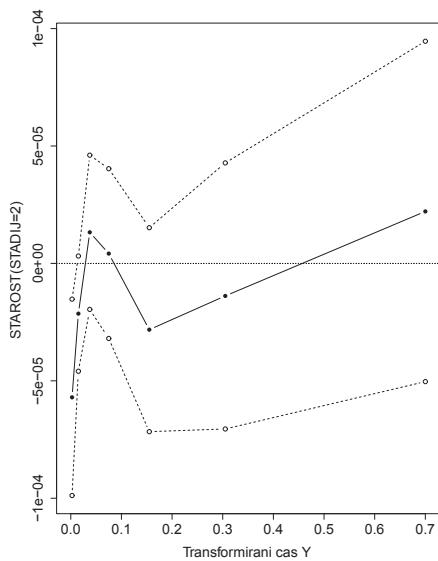
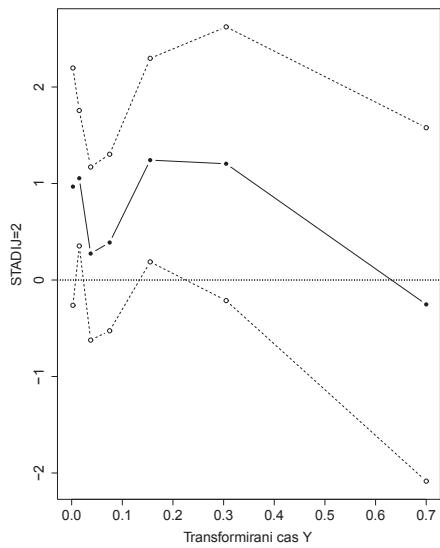
Tabela 6.16: Nadaljevanje tabele 6.15

	$\hat{\beta}$	$\hat{\beta} - 1.96\sqrt{S.E.}$	$\hat{\beta} + 1.96\sqrt{S.E.}$	S.E.	Z	$Pr(> Z)$
STAROST(S=2) _{0.005}	-5.700e-05	-9.880e-05	-1.520e-05	2.130e-05	-2.670	0.008
STAROST(S=2) _{0.025}	-2.140e-05	-4.590e-05	3.140e-06	1.250e-05	-1.710	0.087
STAROST(S=2) _{0.05}	1.320e-05	-1.960e-05	4.610e-05	1.680e-05	0.791	0.429
STAROST(S=2) _{0.1}	4.190e-06	-3.200e-05	4.030e-05	1.840e-05	0.227	0.820
STAROST(S=2) _{0.2}	-2.820e-05	-7.160e-05	1.520e-05	2.220e-05	-1.270	0.203
STAROST(S=2) _{0.4}	-1.390e-05	-7.050e-05	4.280e-05	2.890e-05	-0.479	0.632
STAROST(S=2) ₁	2.210e-05	-5.030e-05	9.460e-05	3.700e-05	0.599	0.549
STAROST(S=3) _{0.005}	4.530e-05	4.630e-06	8.590e-05	2.070e-05	2.180	0.029
STAROST(S=3) _{0.025}	-2.680e-05	-5.690e-05	3.350e-06	1.540e-05	-1.740	0.082
STAROST(S=3) _{0.05}	-3.780e-05	-8.810e-05	1.240e-05	2.560e-05	-1.480	0.140
STAROST(S=3) _{0.1}	-4.750e-05	-1.120e-04	1.740e-05	3.310e-05	-1.430	0.152
STAROST(S=3) _{0.2}	-6.500e-05	-1.570e-04	2.670e-05	4.680e-05	-1.390	0.165
STAROST(S=3) _{0.4}	-1.310e-04	-2.670e-04	4.940e-06	6.950e-05	-1.890	0.059
STAROST(S=3) ₁	2.990e-05	-2.630e-04	3.230e-04	1.490e-04	0.201	0.841
LETO(S=2) _{0.005}	9.090e-05	6.000e-06	1.760e-04	4.330e-05	2.100	0.036
LETO(S=2) _{0.025}	3.980e-05	-1.190e-05	9.150e-05	2.640e-05	1.510	0.131
LETO(S=2) _{0.05}	3.390e-05	-3.770e-05	1.050e-04	3.650e-05	0.928	0.354
LETO(S=2) _{0.1}	2.070e-05	-4.940e-05	9.090e-05	3.580e-05	0.579	0.563
LETO(S=2) _{0.2}	4.610e-06	-6.760e-05	7.690e-05	3.690e-05	0.125	0.900
LETO(S=2) _{0.4}	-2.140e-05	-1.080e-04	6.480e-05	4.400e-05	-0.486	0.627
LETO(S=2) ₁	-2.870e-05	-1.390e-04	8.120e-05	5.610e-05	-0.511	0.609
LETO(S=3) _{0.005}	1.090e-04	2.500e-05	1.930e-04	4.290e-05	2.540	0.011
LETO(S=3) _{0.025}	1.350e-04	7.550e-05	1.930e-04	3.010e-05	4.470	0.000
LETO(S=3) _{0.05}	5.350e-06	-8.660e-05	9.730e-05	4.690e-05	0.114	0.909
LETO(S=3) _{0.1}	2.720e-05	-7.280e-05	1.270e-04	5.100e-05	0.534	0.593
LETO(S=3) _{0.2}	4.450e-05	-8.170e-05	1.710e-04	6.440e-05	0.691	0.489
LETO(S=3) _{0.4}	-1.690e-04	-3.590e-04	2.110e-05	9.700e-05	-1.740	0.082
LETO(S=3) ₁	-2.040e-04	-5.570e-04	1.480e-04	1.800e-04	-1.140	0.256

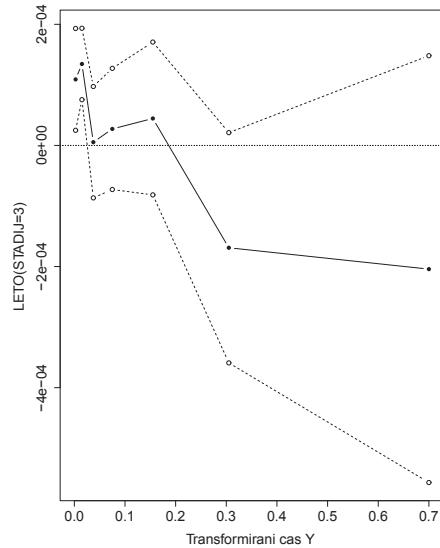
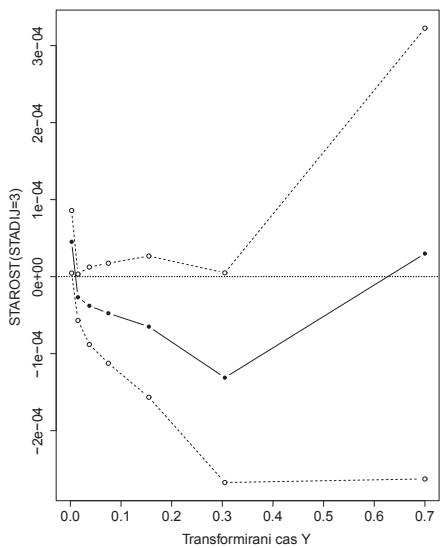
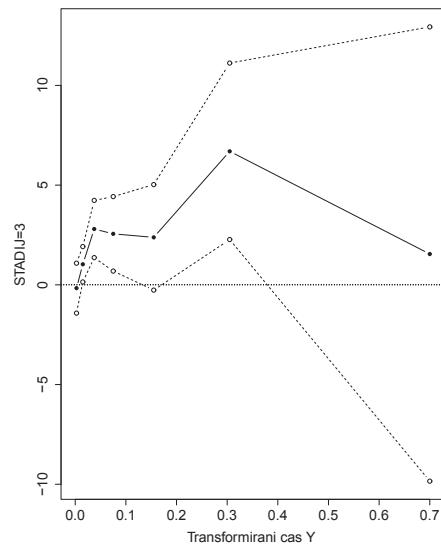
Slika 6.50: Grafični prikaz ocenjenih koeficientov Coxovega modela za podatke o raku na dojki



Slika 6.51: Nadaljevanje slike 6.50



Slika 6.52: Nadaljevanje slike 6.50



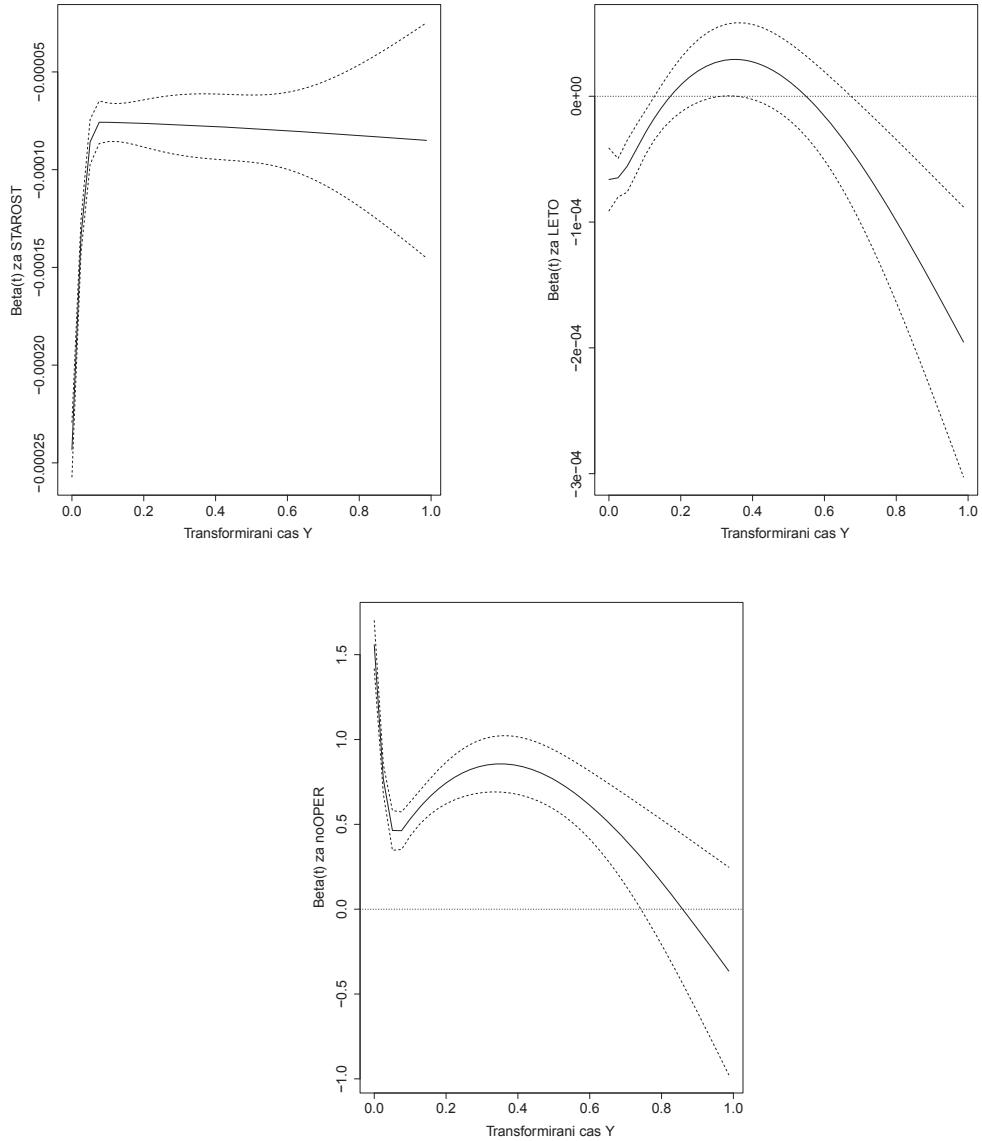
V prikazanem Coxovem modelu z vključenimi interakcijami stadija s starostjo in letom diagnoze na podatkih o raku na dojki se podobno kot v Aalenovem modelu izkaže, da se del presežne ogroženosti višjih stadijev prenese na spremenljivke, na podlagi katerih računamo transformirani čas Y . Zato interpretacija modela postane precej težavna. Na primeru rezultatov ocenjenih koeficientov drugega stadija denimo, vpliv stadija ni povsem jasno razviden (na intervalu $Y > 0,025$ je vpliv drugega stadija na ogroženost za smrt v primerjavi s prvim stadijem statistično neznačilen). To je dodatni razlog, da v nadaljevanju prikazujemo še enostavni Coxov model za podatke o raku na dojki, v katerega so vključeni le glavni vplivi starosti, leta diagnoze, stadija in indikatorja operativnega zdravljenja pacientk.

Testne statistike za preverjanje kakovosti prileganja modela (tabela 6.17) kažejo na kršenje predpostavke o sorazmernih ogroženostih oziroma na potrebo po modeliranju sprememb učinkov neodvisnih spremenljivk v času. Podobno lahko razberemo tudi na podlagi zglajenih krivulj gibanja Schoenfeldovih ostankov (ozioroma $s_{kj}^* + \hat{\beta}_j$) v času (slike 6.53 in 6.54).

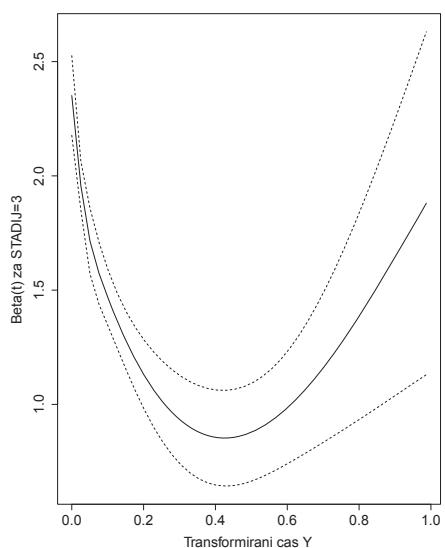
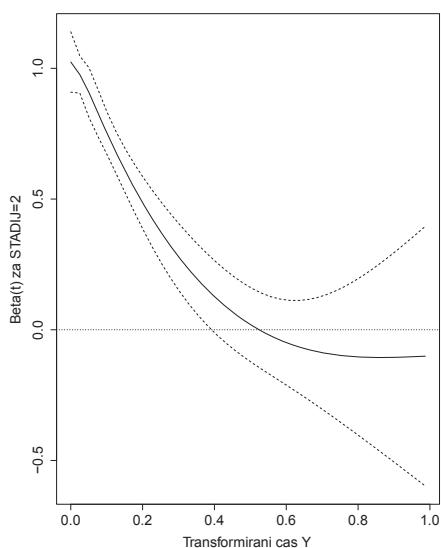
Tabela 6.17: Testne statistike za preverjanje predpostavke o sorazmernih ogroženostih

	Test 1			Test 2	
	rho	chisq	p	max	p
STAROST	0.118	102.064	0.000	7.785	0.000
LETO	0.030	7.257	0.007	5.059	0.000
STADIJ=2	-0.137	139.370	0.000	3.140	0.000
STADIJ=3	-0.099	77.254	0.000	4.495	0.000
BREZ OPERACIJE	-0.054	24.448	0.000	3.714	0.000
GLOBAL		304.492	0.000	11.010	0.000

Slika 6.53: Zglajene krivulje vrednosti $U_{kj}^* + \hat{\beta}_j$ v času Y za podatke o raku na dojki z vrisanimi mejami $\pm 2 \times$ standardna odklona



Slika 6.54: Nadaljevanje slike 6.53



Spreminjanje učinkov neodvisnih spremenljivk v času in s tem zagotovitev veljavnosti predpostavke o sorazmernih ogroženostih lahko rešimo z vpeljavo niza novih spremenljivk, s pomočjo katerih posamezno spremenljivko modeliramo na več različnih časovnih intervalih. Na podlagi grafičnega prikaza zglajenih krivulj Scoenfeldovih ostankov in zaradi primerljivosti z modelom, ki poleg glavnih vplivov starosti, leta diagnoze, stadija in indikatorja operativnega zdravljenja vključuje tudi interakcije starosti in leta diagnoze s stadijem, smo tudi enostavni model modelirali na sedmih časovnih podintervalih: $0 < Y \leq 0,005$, $0,005 < Y \leq 0,025$, $0,025 < Y \leq 0,05$, $0,05 < Y \leq 0,10$, $0,10 < Y \leq 0,20$, $0,20 < Y \leq 0,40$ in $0,40 < Y \leq 1$. Testne statistike za preverjanje kakovosti prileganja modela, prikazane v tabeli 6.18, kažejo, da se izbrani model zadovoljivo prilega podatkom.

Tabela 6.18: Testne statistike za preverjanje predpostavke o sorazmernih ogroženostih

	Test 1			Test 2	
	rho	chisq	p	max	p
STAROST	-0.003	0.085	0.770	1.322	0.061
LETO	0.003	0.063	0.802	0.550	0.923
STADIJ=2	-0.003	0.067	0.796	0.928	0.355
STADIJ=3	0.000	0.001	0.976	1.070	0.202
BREZ OPERACIJE	-0.003	0.100	0.751	1.573	0.014
GLOBAL		0.251	0.998	0.965	0.309

Ocenjeni koeficienti zgoraj opisanega enostavnega modela so prikazani v tabeli 6.19, grafično pa na slikah 6.55 in 6.56, ki sta prikazani podobno kot pri modelu z interakcijami starosti in leta diagnoze s stadijem.

Iz rezultatov modela razberemo naslednje:

- Starost negativno vpliva na ogroženost za smrt za rakom na dojki. Vpliv starosti je najmočnejši v začetnem obdobju, s časom pa postopoma slablji in na intervalu $Y > 0,2$ izgubi statistično značilen vpliv na ogroženost. Tako se denimo ogroženost za smrt za rakom na dojki na intervalu $Y \leq 0,005$ s povečanjem

starosti za 10 let zmanjša za dobrih 50 %, na intervalu $0,1 < Y \leq 0,2$ pa za dobrih 15 %.

- Negativni vpliv na ogroženost za smrt za raki na dojki ima tudi leto diagnoze, pri čemer vpliv leta diagnoze postane statistično neznačilen po času $Y = 0,1$. Glede na ocenjene koeficiente je vpliv leta diagnoze nekoliko manjši v začetnem obdobju ($Y < 0,005$), v katerem se ogroženost zaradi raka na dojki na vsakih 10 let zmanjša za približno 15 %, kasneje pa se najprej poveča (in na vsakih deset let zmanjša ogroženost za okoli 30 %), nato pa s časom slabí.
- Ogroženost za smrt je pri neoperativno zdravljenih patientkah povečana skozi celotno obdobje. Vpliv neoperativne obravnave patientk na ogroženost za smrt zaradi raka na dojki je tako najmočnejši na intervalu $Y < 0,005$, v katerem je ogroženost patientk, ki niso bile operirane, v primerjavi z operativno zdravljenimi povečano za približno 4,5-krat, kasneje ta vpliv slabí in se po času $Y = 0,05$ stabilizira. V tem obdobju je ogroženost neoperativno obravnavanih patientk večja za približno 1,5-krat.
- Podobno tudi diagnoza v višjem stadiju predstavlja povečano ogroženost za smrt za raki na dojki, ki je prisotna skozi celotno obdobje. Pri obeh stadijih lahko opazimo, da je vpliv na ogroženost najmočnejši v začetnem obdobju in s časom slabí.

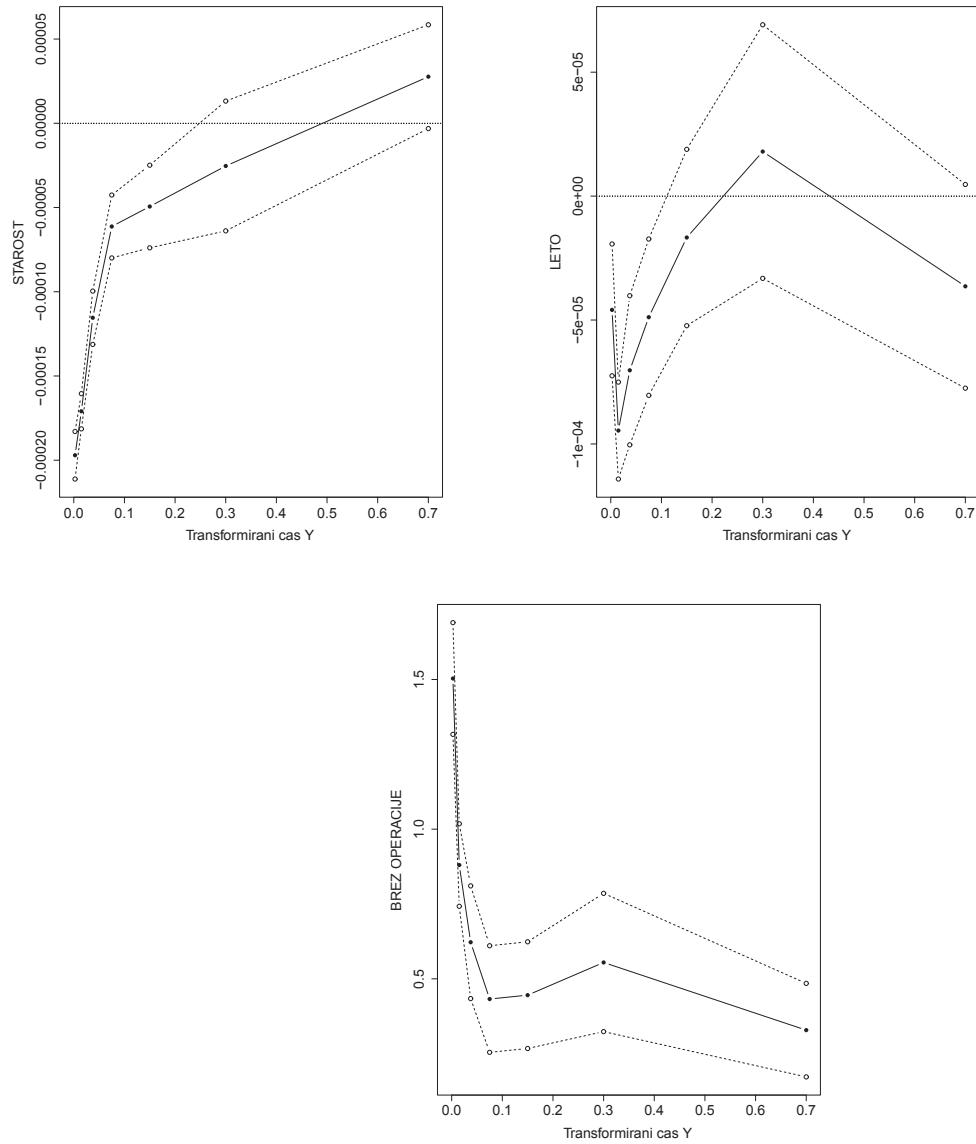
Pri patientkah, diagnosticiranih v drugem stadiju, je tako ogroženost za smrt v primerjavi s patientkami, diagnosticiranimi v prvem stadiju, v začetnem obdobju ($Y < 0,005$) povečana za približno 3-krat, na časovnem intervalu $0,1 < Y \leq 0,4$ pa za približno 1,8-krat.

Patientke, diagnosticirane v tretjem stadiju, imajo v začetnem obdobju v primerjavi s patientkami, diagnosticiranimi v prvem stadiju, ogroženost povečana za skoraj 9-krat, na časovnem intervalu $0,05 < Y < 0,1$ denimo pa za približno-5 krat.

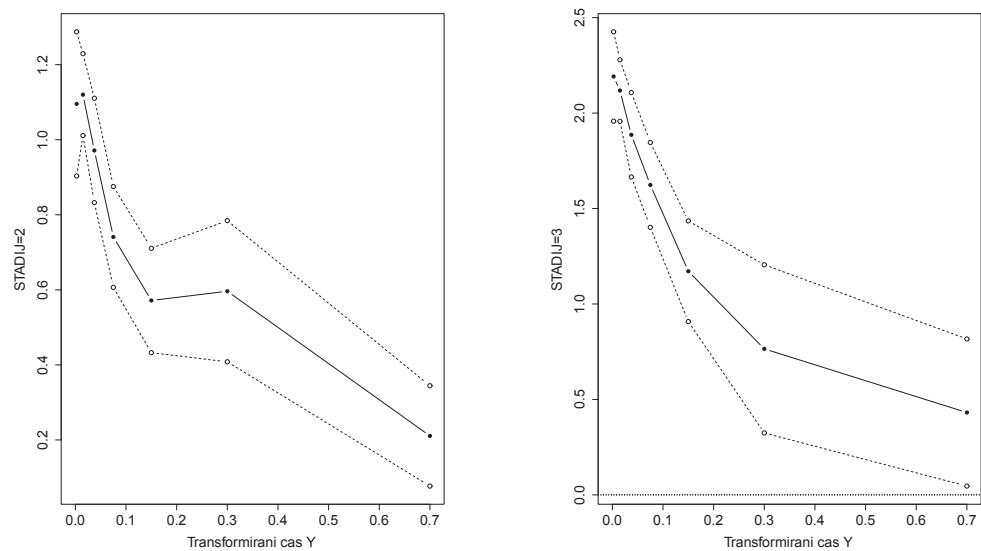
Tabela 6.19: Rezultati ocenjenega modela za podatke o raku na dojki

	$\hat{\beta}$	$\hat{\beta} - 1.96\sqrt{S.E.}$	$\hat{\beta} + 1.96\sqrt{S.E.}$	S.E.	Z	Pr(> Z)
STAROST _{0.005}	-1.970E-04	-2.110E-04	-1.830E-04	7.200E-06	-27.400	0.000
STAROST _{0.025}	-1.710E-04	-1.810E-04	-1.600E-04	5.330E-06	-32.100	0.000
STAROST _{0.05}	-1.150E-04	-1.310E-04	-9.960E-05	8.080E-06	-14.300	0.000
STAROST _{0.1}	-6.130E-05	-7.990E-05	-4.260E-05	9.520E-06	-6.430	0.000
STAROST _{0.2}	-4.940E-05	-7.390E-05	-2.490E-05	1.250E-05	-3.950	0.000
STAROST _{0.4}	-2.530E-05	-6.390E-05	1.320E-05	1.970E-05	-1.290	0.197
STAROST ₁	2.770E-05	-3.160E-06	5.850E-05	1.570E-05	1.760	0.079
LETO _{0.005}	-4.590E-05	-7.250E-05	-1.930E-05	1.360E-05	-3.380	0.001
LETO _{0.025}	-9.460E-05	-1.140E-04	-7.500E-05	9.980E-06	-9.480	0.000
LETO _{0.05}	-7.030E-05	-1.000E-04	-4.020E-05	1.540E-05	-4.580	0.000
LETO _{0.1}	-4.890E-05	-8.040E-05	-1.730E-05	1.610E-05	-3.040	0.002
LETO _{0.2}	-1.670E-05	-5.230E-05	1.890E-05	1.820E-05	-0.920	0.357
LETO _{0.4}	1.800E-05	-3.320E-05	6.910E-05	2.610E-05	0.689	0.491
LETO ₁	-3.640E-05	-7.750E-05	4.700E-06	2.100E-05	-1.740	0.083
STADIJ=2 _{0.005}	1.100	0.903	1.290	0.098	11.200	0.000
STADIJ=2 _{0.025}	1.120	1.010	1.230	0.056	20.100	0.000
STADIJ=2 _{0.05}	0.971	0.832	1.110	0.071	13.700	0.000
STADIJ=2 _{0.1}	0.741	0.606	0.875	0.069	10.800	0.000
STADIJ=2 _{0.2}	0.572	0.433	0.711	0.071	8.050	0.000
STADIJ=2 _{0.4}	0.596	0.408	0.784	0.096	6.210	0.000
STADIJ=2 ₁	0.211	0.077	0.344	0.068	3.090	0.002
STADIJ=3 _{0.005}	2.190	1.960	2.430	0.119	18.400	0.000
STADIJ=3 _{0.025}	2.120	1.960	2.280	0.082	25.800	0.000
STADIJ=3 _{0.05}	1.890	1.670	2.110	0.113	16.700	0.000
STADIJ=3 _{0.1}	1.620	1.400	1.850	0.113	14.300	0.000
STADIJ=3 _{0.2}	1.170	0.908	1.440	0.135	8.710	0.000
STADIJ=3 _{0.4}	0.765	0.325	1.210	0.225	3.400	0.001
STADIJ=3 ₁	0.431	0.046	0.817	0.197	2.200	0.028
BREZ OPER. _{-0.005}	1.500	1.320	1.690	0.095	15.800	0.000
BREZ OPER. _{-0.025}	0.880	0.742	1.020	0.070	12.500	0.000
BREZ OPER. _{-0.05}	0.622	0.434	0.810	0.096	6.480	0.000
BREZ OPER. _{-0.1}	0.432	0.254	0.611	0.091	4.760	0.000
BREZ OPER. _{-0.2}	0.445	0.267	0.624	0.091	4.900	0.000
BREZ OPER. _{-0.4}	0.555	0.324	0.786	0.118	4.710	0.000
BREZ OPER. ₋₁	0.329	0.172	0.485	0.080	4.120	0.000

Slika 6.55: Grafični prikaz ocenjenih koeficientov Coxovega modela za podatke o raku na dojki



Slika 6.56: Nadaljevanje slike 6.55



Rezultati Coxovega modela z modeliranimi spremenjajočimi učinki parametrov tako na primeru podatkov o raku na pljučih, kot tudi na primeru podatkov o raku na dojki kažejo na visoko konsistentnost z rezultati Aalenovega linearnega modela, čeprav modela nista neposredno primerljiva. Prednost Coxovega modela pred Aalenovim je predvsem v kvantitativnem ovrednotenju učinkov parametrov na posameznih časovnih intervalih. Slabost prikazanega modeliranja pa je zlasti arbitrarna izbira časovnih intervalov, s katerimi skušamo zajeti spremembe učinkov v času. V primeru raka na dojki smo s Coxovim modelom podobno kot z Aalenovim ugotovili, da neodvisni spremenljivki starost in leto diagnoze, na podlagi katerih računamo vrednosti transformiranega časa Y , po določenem času izgubita vpliv na ogroženost. Na primeru Aalenovega modela smo pokazali, da to pomeni, da operativno zdravljenje pacientke, diagnosticirane v prvem stadiju raka na dojki, ki preživijo vrednost transformiranega časa $Y = 0,2$, lahko obravnavamo kot statistično ozdravele. Podoben rezultat dobimo tudi s prikazanim Coxovim modelom, ki je v primeru modela z interakcijami pokazal, da se statistično značilen vpliv starosti in leta diagnoze izgubi po času $Y = 0,1$, v enostavnem modelu pa učinek starosti postane statistično neznačilen po času $Y = 0,2$. Po drugi strani s Coxovim modelom nismo uspeli potrditi ugotovitve Aalenovega modela, da operativno zdravljenje pacientke, diagnosticirane v drugem stadiju raka na dojki, statistično gledano ozdravijo, če preživijo vrednost transformiranega časa $Y = 0,4$. Rezultati Coxovega modela so pokazali še nekaj, kar pri rezultatih Aalenovega modela aditivnih ogroženosti ni tako očitno. Spremembe učinkov parametrov so pri obeh podatkovjih najhitrejše in pogosto tudi najmočnejše v začetnem obdobju, kar opozarja na začetno močnejšo agresivnost obeh vrst raka. Prav zaradi hitrega začetnega spremembe učinkov, pa so izbrani časovni intervali, znotraj katerih smo modelirali spremembe učinkov parametrov, bolj zgoščeni v začetnih časih.

6.4 Uporaba Buckley-Jamesovega linearnega modela v relativnem preživetju

V tem delu prikazujemo rezultate Buckley-Jamesovega modela na primerih podatkov o pljučnem raka in raka na dojki. Proučevani podatki se z metodološkega vidika analize preživetja med seboj najbolj razlikujejo glede na odstotek krnjenih enot. Večji ko je odstotek krnjenih enot, težje preverimo veljavnost predpostavk o linearnosti in homoskedastičnosti modela (Stare in drugi 2000). Poleg tega smo z Aalenovim modelom tako v primeru podatkov o raku na pljučih, kot tudi v primeru podatkov o raku na dojki ugotovili, da se učinki neodvisnih spremenljivk na ogroženost v času spreminjajo. Tako zaradi krnjenja, kot tudi zaradi ugotovljene časovne spremembe učinkov neodvisnih spremenljivk na ogroženost lahko pričakujemo, da bodo ocenjeni parametri Buckley-Jamesovega modela pristranski. Kljub temu smo z rezultati simulacij ugotovili, da ocenjeni parametri Buckley-Jamesovega modela v primeru, ko podatke generira Aalenov model, še vedno ponujajo vsebinsko smiselne rezultate, s katerimi lahko doprinesemo nove informacije in vsebinsko oplemenitimo problematiko relativnega preživetja na realnih podatkih. Poleg tega predpostavko modela o strogi linearnosti zveznih spremenljivk lahko opustimo z uvedbo nelinearnih členov, kar ponuja dodatno prilagodljivost klasične linearne regresije pri modeliranju podatkov.

6.4.1 Buckley-Jamesova linearna regresija na podatkih o pljučnem raku

V proučevanem obdobju (1. januar 1983 do 29. december 2004) je bilo diagnostičiranih 17639 pacientov s pljučnim rakom. Za namene analiz smo iz podatkov izključili 29 pacientov, ki so v tem obdobju emigrirali in 754 pacientov, za katere je podatek o stadiju bolezni, v katerem so bili diagnosticirani, nerazpoložljiv oziroma neznan. Pacienti so bili opazovani do 30. septembra 2007, v tem času pa jih je 95 % odstotkov umrlo. Neodvisne spremenljivke, ki smo jih vključili v modeliranje transformiranega časa Y (starost, spol, leto diagnoze in stadij), nimajo manjkajočih vrednosti.

Pri proučevanju Buckley-Jamesove linearne regresije na podatkih o pljučnem raku smo

za izbiro ustreznega modela izvedli podrobne analize, na podlagi katerih smo proučili možne interakcije, ki nastopajo v modelu, ter ustrezne funkcijске oblike zveznih spremenljivk in preverili, kako se izbrani model prilega proučevanim podatkom.

Za namene modeliranja smo kot odvisno spremenljivko uporabili naravni logaritem transformiranega časa $\log(Y)$. Namenske uporabljeni transformacije je dvojen: 1. z logaritmom transformiranega časa se izognemo negativnim vrednostim transformiranega časa v originalnem modelu in 2. v primeru heteroskedastičnih podatkov z logaritmiranjem odvisne spremenljivke lahko stabiliziramo varianco modela.

Predpostavili smo, da so učinki starosti, spola in leta diagnoze na transformirani čas v različnih stadijih pljučnega raka različni, zato smo v model vključili interakcije med njimi. Predpostavko o strogi linearnosti obeh zveznih spremenljivk (starosti in leta diagnoze) smo preverjali tako, da smo v model vključili nelinearne člene, natančneje kubične zlepke z omejitvami. Lokacije vozlišč, določene na podlagi kvantilov obeh neodvisnih spremenljivk, so izbrane, kot je predlagano v (Harrell 2001). Analize so pokazale, da leto diagnoze nima značilnega nelinearnega vpliva, medtem ko se je predpostavljen nelinearni učinek starosti izkazal za značilnega, tako v primeru podatkov o raku na pljučih, kot tudi v primeru podatkov o raku na dojki. V obeh primerih do nelinearnega učinka pride pri višjih vrednostih starosti, na kar bi lahko vplivalo tudi krnjjenje. Ker je nelinearni učinek majhen, ga v predstavljenem modelu nismo upoštevali. Za predpostavljene interakcije (starosti, spola in leta diagnoze s stadijem) smo ugotovili, da imajo pri nizki stopnji tveganja statistično značilen vpliv na transformirani čas (tabela 6.20).

Rezultati ocenjenega modela so prikazani v tabeli 6.21. Napaka modela je ocenjena s $\hat{\sigma} = \sqrt{\frac{\hat{Y}-Y}{n_{nc}-p-1}}$, kjer je n_{nc} število nekrnjjenih enot, p pa število neodvisnih spremenljivk, vključenih v model, in znaša $\hat{\sigma} = 1,338$.

Interpretacija modela

Tabela 6.21 prikazuje ocene parametrov Buckley-Jamesovega modela za podatke o

Tabela 6.20: Waldove statistike za $\log(Y)$

	χ^2	<i>d.f.</i>	<i>P</i>
STAROST (Factor+Higher Order Factors)	2361.74	3	< 0.0001
<i>All Interactions</i>	196.98	2	< 0.0001
SPOL (Factor+Higher Order Factors)	380.75	3	< 0.0001
<i>All Interactions</i>	32.27	2	< 0.0001
LETO (Factor+Higher Order Factors)	38.95	3	< 0.0001
<i>All Interactions</i>	31.20	2	< 0.0001
STADIJ (Factor+Higher Order Factors)	4133.42	8	< 0.0001
<i>All Interactions</i>	238.51	6	< 0.0001
STAROST \times STADIJ (Factor+Higher Order Factors)	196.98	2	< 0.0001
SPOL \times STADIJ (Factor+Higher Order Factors)	32.27	2	< 0.0001
LETO \times STADIJ (Factor+Higher Order Factors)	31.20	2	< 0.0001
TOTAL INTERACTION	238.51	6	< 0.0001
TOTAL	7638.19	11	< 0.0001

pljučnem raku. Ocenjeni parametri modela z vidika relativnega preživetja, izraženega s transformiranim časom ali individualno mero relativnega preživetja, povedo naslednje:

- Preživetje bolnikov, diagnosticiranih v drugem oziroma v tretjem stadiju, se v primerjavi s preživetjem bolnikov, diagnosticiranih v prvem stadiju pljučnega raka, bistveno poslabša. Višji ko je stadij odkrite bolezni, slabše so možnosti preživetja bolnikov.
- V primerjavi s splošno populacijo je preživetje žensk, diagnosticiranih s pljučnim rakom, slabše od preživetja moških, negativni učinek na preživetje žensk v primerjavi z moškimi pa se v višjih stadijih še poveča. Ob upoštevanju dejstva, da je preživetje žensk v splošni populaciji večje od preživetja moških, to pomeni, da ženske zaradi pljučnega raka z vidika preživetja izgubljajo populacijsko prednost pred moškimi.

- Vpliv starosti na preživetje bolnikov s pljučnim rakom je pozitiven, kar pomeni, da je v primerjavi z vrstniki splošne populacije preživetje starejših patientov večje od preživetja mlajših patientov. To pomeni, da starejši patienti ob isti vrednosti izvirnega časa $T = t$ preživijo večji delež svojih sovrstnikov kot mlajši patienti. Povedano drugače, čas do smrti je za mlajše paciente s pljučnim rakom, v primerjavi z vrstniki splošne populacije veliko krajši kot čas do smrti starejših patientov v primerjavi z vrstniki splošne populacije. V višjih stadijih se vpliv starosti na preživetje dodatno poveča.
- Pozitivni učinek leta diagnoze, ki ga lahko opazimo v prvem stadiju pljučnega raka, pove, da se relativno preživetje patientov, diagnosticiranih v prvem stadiju bolezni, v času izboljšuje. Ker izboljševanje proučevanega preživetja samodejno izključuje izboljševanje preživetja v splošni populaciji, to lahko pripisemo izboljšanju zdravljenja in/ali medicinske oskrbe patientov. V višjih stadijih pljučnega raka leto diagnoze izgubi vpliv na preživetje, kar pomeni, da zdravljenje in/ali medicinska oskrba bolnikov s pljučnim rakom, diagnosticiranih v drugem oziroma tretjem stadiju, s časom ni bistveno napredovala.

V Aalenovem modelu smo na podatkih o pljučnem raku zaznali, da se učinka spola in starosti na ogroženost v času spremenita. To pomeni, da parametri ocen, ocenjeni z Buckley-Jamesovim modelom, predstavljajo utežene povprečne učinke neodvisnih spremenljivk, ki so odvisni od moči učinkov na časovnih intervalih med prelomnimi točkami in od deleža enot na omenjenih časovnih intervalih (glej poglavje Primerjava Aalenovega in Buckley-Jamesovega regresijskega modela). Informacijo o časovnih spremembah učinkov neodvisnih spremenljivk na ogroženost oziroma čas preživetja pa v Buckley-Jamesovem modelu lahko 'zajamemo' tako, da regresijo izvedemo znotraj časovnih intervalov, določenih s prelomnimi točkami. S simulacijami je namreč mogoče pokazati, da v primeru ko podatke generiramo s klasično linearno regresijo, tako da so podatki pod in nad izbrano prelomno točko generirani z različnimi vrednostmi parametrov neodvisnih spremenljivk, spremembo učinka zazna tudi Aalenov model.

Tabela 6.21: Rezultati ocenjenega modela za podatke o pljučnem raku

	Value	Std. Error	Z	$Pr(> Z)$
Konstanta	-5.674	0.182	-31.126	0.000
STAROST	5.679e-5	6.243e-6	9.097	0.000
SPOL=ženski	-0.244	0.064	-3.804	0.000
LETO	6.735e-5	1.083e-5	6.220	0.000
STADIJ=2	-1.684	0.225	-7.475	0.000
STADIJ=3	-3.402	0.228	-14.932	0.000
STAROST * STADIJ=2	8.398e-5	7.774e-6	10.803	0.000
STAROST * STADIJ=3	1.059e-4	7.650e-6	13.847	0.000
SPOL=ženski * STADIJ=2	-0.445	0.079	-5.640	0.000
SPOL=ženski * STADIJ=3	-0.255	0.076	-3.335	0.001
LETO * STADIJ=2	-6.965e-5	1.308e-5	-5.325	0.000
LETO * STADIJ=3	-6.420e-5	1.325e-5	-4.844	0.000

Na podlagi Aalenovega modela ocenjenega na podatkih o pljučnem raku pri vrednosti transformiranega časa $Y \approx 0,03$ pride do spremembe učinka starosti in spola na ogroženost za preživetje opazovanih bolnikov. Podobno kažejo tudi rezultati modelov ocenjenih z Buckley-Jamesovo metodo na podatkih pod in nad prelomno točko v tabelah 6.22 in 6.23 (rezultati prileganja obeh modelov so prikazani v tabelah 6.24 in 6.25). Vidimo lahko, da sta parametra starosti in spola, v modelih ocenjenih na podatkih pod in nad prelomno točko, različnega predznaka in skladna z rezultati Aalenovega modela. Tako je vpliv starosti na preživetje opazovanih bolnikov do prelomne točke $Y = 0,03$ pozitiven, nad njo pa je negativen. Podobno je preživetje obolelih žensk v primerjavi s preživetjem obolelih moških v začetnem obdobju manjše, kasneje pa se ta učinek obrne v prid žensk. Ocenjena parametra starosti in spola sta v obeh modelih pri stopnji tveganja 0,05 statistično značilna. Opazimo lahko tudi, da vpliv leta diagnoze v začetnem obdobju ($Y < 0,03$) ni statistično značilen.

Tabela 6.22: Rezultati ocenjenega modela za podatke o pljučnem raku, ($Y \leq 0,03$)

	Value	Std. Error	Z	$Pr(> Z)$
Konstanta	-5.421	0.198	-27.348	0.000
STAROST	3.249E-05	6.818E-06	4.765	0.000
SPOL=ženski	-0.137	0.067	-2.064	0.039
LETO	-1.391E-05	1.204E-05	-1.156	0.248
STADIJ=2	-1.044	0.237	-4.415	0.000
STADIJ=3	-3.058	0.232	-13.195	0.000
STAROST*STADIJ=2	4.868E-05	8.204E-06	5.934	0.000
STAROST*STADIJ=3	9.539E-05	7.873E-06	12.116	0.000
SPOL=ženski*STADIJ=2	-0.254	0.079	-3.205	0.001
SPOL=ženski*STADIJ=3	-0.284	0.076	-3.748	0.000
LETO*STADIJ=2	-5.135E-06	1.402E-05	-0.366	0.714
LETO*STADIJ=3	1.571E-05	1.377E-05	1.141	0.254

$n=12384$, št. nekrnjnih enot=12256, $\hat{\sigma} = 1,122$

Tabela 6.23: Rezultati ocjenjenega modela za podatke o pljučnem raku, ($Y > 0,03$)

	Value	Std. Error	Z	$Pr(> Z)$
Konstanta	-1.883	0.159	-11.864	0.000
STAROST	-4.013E-05	5.627E-06	-7.132	0.000
SPOL=ženski	0.236	0.058	4.083	0.000
LETO	5.608E-05	8.628E-06	6.500	0.000
STADIJ=2	-0.537	0.227	-2.370	0.018
STADIJ=3	-1.061	0.355	-2.991	0.003
STAROST*STADIJ=2	2.603E-05	8.300E-06	3.136	0.002
STAROST*STADIJ=3	4.466E-05	1.224E-05	3.649	0.000
SPOL=ženski*STADIJ=2	-0.221	0.083	-2.660	0.008
SPOL=ženski*STADIJ=3	-0.120	0.109	-1.104	0.270
LETO*STADIJ=2	-4.229E-05	1.142E-05	-3.705	0.000
LETO*STADIJ=3	-6.032E-05	1.620E-05	-3.723	0.000

$n=4470$, št. nekrnjenih enot=3748, $\hat{\sigma} = 0,698$

Tabela 6.24: Waldove statistike za $\log(Y)$, ($Y \leq 0,03$)

	χ^2	d.f.	P
STAROST (Factor+Higher Order Factors)	1394.72	3	< 0.0001
<i>All Interactions</i>	162.91	2	< 0.0001
SPOL (Factor+Higher Order Factors)	223.55	3	< 0.0001
<i>All Interactions</i>	14.46	2	0.0007
LETO (Factor+Higher Order Factors)	8.43	3	0.0378
<i>All Interactions</i>	4.72	2	0.0945
STADIJ (Factor+Higher Order Factors)	1336.52	8	< 0.0001
<i>All Interactions</i>	184.73	6	< 0.0001
STAROST \times STADIJ (Factor+Higher Order Factors)	162.91	2	< 0.0001
SPOL \times STADIJ (Factor+Higher Order Factors)	14.46	2	0.0007
LETO \times STADIJ (Factor+Higher Order Factors)	4.72	2	0.0945
TOTAL INTERACTION	184.73	6	< 0.0001
TOTAL	2837.84	11	< 0.0001

Tabela 6.25: Waldove statistike za $\log(Y)$, ($Y > 0,03$)

	χ^2	d.f.	P
STAROST (Factor+Higher Order Factors)	56.37	3	< 0.0001
<i>All Interactions</i>	17.74	2	0.0001
SPOL (Factor+Higher Order Factors)	18.33	3	0.0004
<i>All Interactions</i>	7.09	2	0.0288
LETO (Factor+Higher Order Factors)	45.74	3	< 0.0001
<i>All Interactions</i>	19.67	2	0.0001
STADIJ (Factor+Higher Order Factors)	499.74	8	< 0.0001
<i>All Interactions</i>	38.21	6	< 0.0001
STAROST \times STADIJ (Factor+Higher Order Factors)	17.74	2	0.0001
SPOL \times STADIJ (Factor+Higher Order Factors)	7.09	2	0.0288
LETO \times STADIJ (Factor+Higher Order Factors)	19.67	2	0.0001
TOTAL INTERACTION	38.21	6	< 0.0001
TOTAL	546.11	11	< 0.0001

Pričakovane vrednosti individualnega relativnega preživetja pacientov s pljučnim rakom

V prejšnjem razdelku smo si pogledali interpretacijo učinkov neodvisnih spremenljivk na pričakovane vrednosti transformiranega časa oziroma individualne mere relativnega preživetja. Buckley in Jamesov model v nasprotju z ostalimi metodami v analizi relativnega preživetja ponuja neposredno napovedovanje pričakovanih vrednosti transformiranega časa. Napovedovanje preživetja s klasičnimi metodami relativnega preživetja je običajno možno le ob predhodni oceni osnovne ogroženosti (ang. *baseline hazard*), ki jo moramo oceniti za vsako enoto posebej. V Buckley-Jamesovi linearni regresiji pa pričakovano preživetje, podobno kot v navadni linearni regresiji, napovedujemo le na podlagi ocenjenih parametrov modela.

Napovedovanje transformiranih časov ponuja informacijo o povprečni individualni meri relativnega preživetja ali povprečnem transformiranem času za katerokoli kombinacijo neodvisnih spremenljivk v modelu. To pomeni, da lahko za vsakega pacienta ocenimo, kakšen delež primerljive splošne populacije v povprečju preživi oboleli.²

Na drugi strani nam napovedovanje transformiranih časov omogoča, da določimo kombinacijo neodvisnih spremenljivk, ob kateri bo oboleli v povprečju preživel neki izbrani delež splošne populacije.

V tabelah 6.26 do 6.31 so prikazani ocenjeni povprečni transformirani časi za različne vrednosti starosti in obdobja diagnoze pljučnega raka (1. januar 1983, 1. januar 1993 in 1. januar 2003). Tabele so prikazane ločeno po spolu in stadiju pljučnega raka ob postavitvi diagnoze. Za lažjo predstavo o pomenu transformiranih časov so poleg njih v tabelah prikazani še pripadajoči dejanski časi do smrti.

Tabele 6.26 do 6.28 prikazujejo pričakovane vrednosti transformiranih časov za moške, ki so bili diagnosticirani v posameznem stadiju pljučnega raka. Pri interpretaciji se bomo osredotočili na moškega, ki je bil v času diagnoze star 65 let, kar predstavlja približno povprečno starost pacientov, diagnosticiranih s pljučnim rakom.

²Primerljiva splošna populacija predstavlja ženske oziroma moške splošne populacije, ki so bili v letu diagnoze stari enako kot pacientka oziroma pacient.

Retzultati kažejo, da 65 let star moški, ki je bil s pljučnim rakom diagnosticiran leta 1983, v povprečju preživi 2,3 % primerljive splošne populacije. Relativno preživetje pacientov, diagnosticiranih v prvem stadiju pljučnega raka, se skozi čas v povprečju izboljšuje. Vidimo lahko, da moški, ki je bil diagnosticiran z bolezni jo v letu 2003, v svojem 65. letu starosti, lahko pričakuje, da bo v povprečju preživel 3,8 % primerljive splošne populacije. Čas preživetja 65-letnikov, ki so bili diagnosticirani v prvem stadiju pljučnega raka, se je s pričakovanim 215 dni v letu 1983 v povprečju povečal na 589 dni v letu 2003. Medtem ko dejanski čas preživetja (izražen v dnevih) odraža tako izboljšanje preživetja splošne populacije, kot tudi izboljšanje zdravljenja oziroma medicinske oskrbe pacientov s pljučnim rakom, pa transformirani čas odraža le izboljšanje preživetja, ki ga lahko pripišemo izboljšanemu zdravljenju oziroma medicinski oskrbi pacientov.

Pričakovano relativno preživetje pacientov diagnosticiranih v drugem in tretjem stadiju pljučnega raka, je glede na rezultate Buckley-Jamesovega modela precej slabše kot za paciente, diagnosticirane v prvem stadiju bolezni. Poleg tega se relativno preživetje teh pacientov skozi čas v povprečju ne izboljšuje.

Vidimo lahko, da 65 let star moški, diagnosticiran v drugem stadiju pljučnega raka, lahko pričakuje, da bo v povprečju preživel 1,8 % primerljive splošne populacije. Ocenjeni povprečni transformirani čas enako starega moškega, diagnosticiranega v tretjem stadiju bolezni, pa je 0,006, kar pomeni, da bo v povprečju preživel le 0,6 % primerljive splošne populacije. Na podlagi ocenjenih pričakovanih časov do smrti za moške, diagnosticirane v drugem in tretjem stadiju pljučnega raka, pa lahko vidimo, kakšen je vpliv na čas preživetja zaradi izboljšanega preživetja splošne populacije. Medtem ko transformirani čas skozi čas ostaja nespremenjen, se dejanski čas do smrti povečuje.

V tabelah 6.29 do 6.31 so podobno kot za moške prikazani rezultati za ženske, diagnosticirane s pljučnim rakom. Opazimo lahko, da so ocenjeni pričakovani transformirani časi za ženske nižji od tistih za moške za vse kombinacije nedovisnih spremenljivk. Rečemo torej lahko, da je relativno preživetje žensk slabše od relativnega preživetja moških in da ženske zaradi pljučnega raka izgubljam populačijsko prednost

pred moškimi. Če primerjamo dejanske čase do smrti, pa lahko opazimo, da so ti za ženske še vedno višji kot za moške, kar pomeni, da ženske zaradi pljučnega raka ne izgubijo vse populacijske prednosti, ki jo imajo pred moškimi.

Za interpretacijo rezultatov vzemimo žensko, ki je bila ob času diagnoze stara 65 let. 65-letna ženska, ki je bila diagnosticirana v prvem stadiju pljučnega raka leta 1983, lahko pričakuje, da bo v povprečju preživila 1,8 % primerljive splošne populacije. Glede na Buckley-Jamesov model se relativno preživetje pacientk v času izboljšuje in vidimo lahko, da se transformirani čas za žensko enake starosti, diagnosticirano v letu 2003, poveča na 0,03. To pomeni, da 65 letna ženska, ki je bila diagnosticirana v prvem stadiju pljučnega raka v letu 2003, lahko pričakuje, da bo v povprečju preživila 3 % primerljive splošne populacije.

Relativno preživetje žensk, ki so bile diagnosticirane v višjih stadijih pljučnega raka, je bistveno slabše in se v času ne izboljšuje. Tako lahko vidimo, da 65-letna ženska, diagnosticirana v drugem stadiju bolezni lahko pričakuje, da bo v povprečju preživila 0,9 % primerljive splošne populacije, enako stara ženska, diagnosticirana v tretjem stadiju bolezni, pa bo v povprečju preživila le še 0,3 % primerljive splošne populacije.

Tabela 6.26: Ocenjeni pričakovani transformirani časi za moške paciente, diagnostičirane v 1. stadiju pljučnega raka

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.014	1233	0.018	1566	0.023	2506
45	0.015	759	0.020	1199	0.025	1979
50	0.017	571	0.022	905	0.028	1367
55	0.019	376	0.024	587	0.031	1047
60	0.021	299	0.027	451	0.034	911
65	0.023	215	0.030	331	0.038	589
70	0.026	177	0.033	295	0.042	382
75	0.029	119	0.037	208	0.047	294
80	0.032	83	0.041	128	0.052	213

Tabela 6.27: Ocenjeni pričakovani transformirani časi za moške paciente, diagnostičirane v 2. stadiju pljučnega raka

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.005	447	0.005	425	0.005	690
45	0.006	343	0.006	402	0.006	532
50	0.008	294	0.008	360	0.008	366
55	0.011	210	0.010	258	0.010	346
60	0.014	194	0.014	227	0.013	347
65	0.018	163	0.018	193	0.017	262
70	0.023	156	0.023	201	0.022	201
75	0.030	123	0.029	166	0.029	181
80	0.038	100	0.038	119	0.038	153

Tabela 6.28: Ocenjeni pričakovani transformirani časi za moške paciente, diagnostičirane v 3. stadiju pljučnega raka

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.001	115	0.001	112	0.001	197
45	0.002	92	0.002	110	0.002	154
50	0.002	82	0.002	102	0.002	106
55	0.003	61	0.003	76	0.003	105
60	0.004	58	0.004	70	0.004	109
65	0.006	51	0.006	62	0.006	85
70	0.008	51	0.008	67	0.008	68
75	0.010	41	0.010	57	0.010	64
80	0.014	35	0.014	43	0.014	56

Tabela 6.29: Ocenjeni pričakovani transformirani časi za patientke, diagnosticirane v 1. stadiju pljučnega raka

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.011	2254	0.014	2941	0.018	4218
45	0.012	1484	0.015	2005	0.020	3013
50	0.013	1153	0.017	1592	0.022	2238
55	0.015	766	0.019	1308	0.024	1881
60	0.016	636	0.021	840	0.027	1609
65	0.018	390	0.023	693	0.030	1103
70	0.020	263	0.026	393	0.033	698
75	0.022	157	0.029	272	0.037	442
80	0.025	95	0.032	150	0.041	258

Tabela 6.30: Ocenjeni pričakovani transformirani časi za pacientke, diagnosticirane v 2. stadiju pljučnega raka

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.002	510	0.002	856	0.002	868
45	0.003	393	0.003	559	0.003	615
50	0.004	402	0.004	514	0.004	442
55	0.005	282	0.005	371	0.005	379
60	0.007	276	0.007	267	0.007	471
65	0.009	189	0.009	284	0.009	335
70	0.011	148	0.011	171	0.011	256
75	0.015	103	0.015	138	0.015	175
80	0.019	73	0.019	89	0.019	118

Tabela 6.31: Ocenjeni pričakovani transformirani časi za pacientke, diagnosticirane v 3. stadiju pljučnega raka

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.001	164	0.001	253	0.001	273
45	0.001	130	0.001	179	0.001	244
50	0.001	136	0.001	185	0.001	165
55	0.002	99	0.002	133	0.002	138
60	0.003	101	0.003	100	0.003	184
65	0.003	72	0.003	110	0.003	133
70	0.005	59	0.005	69	0.005	105
75	0.006	42	0.006	58	0.006	75
80	0.008	31	0.008	39	0.008	52

Podobne informacije kot so predstavljene v tabelah 6.26 do 6.31 na podlagi Buckley-Jamesovega modela, ocenjenega na vseh podatkih, kažejo tabele 6.33 do 6.44 na podlagi modelov, ocenjenih pod (tabele 6.33 do 6.38) in nad prelomno točko $Y \approx 0,03$ (tabele 6.39 do 6.44). Za pomoč pri branju rezultatov so v tabeli 6.32 prikazani tudi izvirni časi, ki pri izbrani vrednosti spola in leta diagnoze ustrezano vrednosti transformiranega časa $Y = 0,03$. Z vidika praktične uporabe so koristne zlasti informacije o pričakovani vrednosti časa preživetja nad prelomno točko, ki predstavljajo pričakovani čas preživetja pod pogojem, da je pacient preživel 3 % ($Y = 0,03$) ustrezne splošne populacije. Poleg tega je iz prikazanih rezultatov mogoče razbrati spremembo učinka starosti in spola na časovnih intervalih pod in nad prelomno točko.

Tabela 6.32: Izvirni čas T , ki ustreza transformiranemu času $Y = 0,03$, za izbrano kombinacijo starosti, spola in leta diagnoze

Starost v letih	moški			ženske		
	1983	1993	2003	1983	1993	2003
40	2315	2494	3060	4538	4852	5540
45	1429	1763	2284	3128	3433	4008
50	941	1230	1451	2357	2498	2954
55	579	724	1013	1565	1915	2304
60	435	503	812	1140	1154	1768
65	279	333	459	627	876	1109
70	206	268	269	391	459	637
75	124	170	187	211	284	364
80	78	94	121	115	141	189

Tabela 6.33: Ocjenjeni pričakovani transformirani časi za moške paciente, diagnostičirane v 1. stadiju pljučnega raka; parametri modela, ocjenjeni na podatkih pod prelomno točko $Y \leq 0,03$

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.006	581	0.006	525	0.006	797
45	0.007	365	0.006	409	0.006	520
50	0.007	256	0.007	301	0.006	292
55	0.008	150	0.007	176	0.007	227
60	0.008	113	0.008	127	0.007	186
65	0.009	78	0.008	89	0.008	115
70	0.009	61	0.009	75	0.008	72
75	0.010	39	0.009	51	0.009	53
80	0.010	26	0.010	30	0.009	36

Tabela 6.34: Ocjenjeni pričakovani transformirani časi za moške paciente, diagnostičirane v 2. stadiju pljučnega raka; parametri modela, ocjenjeni na podatkih pod prelomno točko $Y \leq 0,03$

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.004	397	0.004	357	0.004	551
45	0.005	274	0.005	302	0.004	386
50	0.006	210	0.005	242	0.005	231
55	0.007	134	0.006	155	0.006	196
60	0.008	111	0.007	122	0.007	176
65	0.009	83	0.009	93	0.008	119
70	0.011	71	0.010	87	0.009	81
75	0.012	50	0.011	64	0.011	65
80	0.014	36	0.013	41	0.012	49

Tabela 6.35: Ocenjeni pričakovani transformirani časi za moške paciente, diagnosticirane v 3. stadiju pljučnega raka; parametri modela, ocenjeni na podatkih pod prelomno točko $Y \leq 0,03$

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.001	124	0.001	121	0.001	210
45	0.002	93	0.002	111	0.002	155
50	0.002	78	0.002	97	0.002	100
55	0.003	54	0.003	68	0.003	92
60	0.003	49	0.004	58	0.004	90
65	0.004	40	0.004	48	0.004	66
70	0.006	37	0.006	49	0.006	49
75	0.007	28	0.007	39	0.007	43
80	0.009	22	0.009	27	0.009	36

Tabela 6.36: Ocenjeni pričakovani transformirani časi za patientke, diagnosticirane v 1. stadiju pljučnega raka; parametri modela, ocenjeni na podatkih pod prelomno točko $Y \leq 0,03$

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.006	1217	0.005	1617	0.005	1691
45	0.006	674	0.006	910	0.005	1019
50	0.006	600	0.006	726	0.006	577
55	0.007	350	0.006	437	0.006	439
60	0.007	281	0.007	260	0.006	443
65	0.007	158	0.007	228	0.007	257
70	0.008	101	0.007	112	0.007	161
75	0.008	58	0.008	74	0.008	90
80	0.009	33	0.008	39	0.008	49

Tabela 6.37: Ocenjeni pričakovani transformirani časi za pacientke, diagnosticirane v 2. stadiju pljučnega raka; parametri modela, ocenjeni na podatkih pod prelomno točko $Y \leq 0,03$

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.003	605	0.003	933	0.003	921
45	0.003	418	0.003	565	0.003	590
50	0.004	388	0.004	470	0.003	391
55	0.005	243	0.004	302	0.004	288
60	0.005	214	0.005	194	0.005	330
65	0.006	131	0.006	185	0.005	205
70	0.007	92	0.007	100	0.006	140
75	0.008	57	0.008	72	0.007	86
80	0.010	36	0.009	41	0.008	52

Tabela 6.38: Ocenjeni pričakovani transformirani časi za pacientke, diagnosticirane v 3. stadiju pljučnega raka; parametri modela, ocenjeni na podatkih pod prelomno točko $Y \leq 0,03$

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.001	191	0.001	294	0.001	316
45	0.001	142	0.001	195	0.001	265
50	0.001	140	0.001	189	0.001	168
55	0.002	96	0.002	128	0.002	132
60	0.002	91	0.002	90	0.002	165
65	0.003	61	0.003	93	0.003	112
70	0.004	47	0.004	54	0.004	83
75	0.005	32	0.005	43	0.005	55
80	0.006	22	0.006	27	0.006	36

Tabela 6.39: Ocenjeni pričakovani transformirani časi za moške paciente, diagnostirane v 1. stadiju pljučnega raka; parametri modela, ocenjeni na podatkih pod prelomno točko $Y > 0,03$

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.136	6771	0.166	8123	0.204	9365
45	0.126	4738	0.155	6381	0.190	7495
50	0.117	3075	0.144	4729	0.176	5752
55	0.109	2047	0.134	3177	0.164	4342
60	0.101	1486	0.124	1919	0.152	3284
65	0.094	960	0.115	1345	0.142	2103
70	0.087	612	0.107	883	0.132	1302
75	0.081	347	0.100	599	0.122	744
80	0.075	202	0.093	300	0.114	463

Tabela 6.40: Ocenjeni pričakovani transformirani časi za moške paciente, diagnostirane v 2. stadiju pljučnega raka; parametri modela, ocenjeni na podatkih pod prelomno točko $Y > 0,03$

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.081	4873	0.085	5472	0.090	6088
45	0.079	3409	0.083	4192	0.088	4691
50	0.077	2161	0.081	2949	0.085	3397
55	0.075	1459	0.079	1994	0.083	2501
60	0.073	1101	0.077	1254	0.081	1981
65	0.071	714	0.075	867	0.079	1200
70	0.070	485	0.073	623	0.077	721
75	0.068	288	0.071	416	0.075	464
80	0.066	176	0.070	223	0.073	303

Tabela 6.41: Ocjenjeni pričakovani transformirani časi za moške paciente, diagnostičirane v 3. stadiju pljučnega raka; parametri modela, ocjenjeni na podatkih pod prelomno točko $Y > 0,03$

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.054	3707	0.053	3931	0.053	4558
45	0.055	2483	0.054	2975	0.053	3396
50	0.055	1630	0.054	2148	0.053	2351
55	0.056	1065	0.055	1378	0.054	1723
60	0.056	842	0.055	910	0.054	1379
65	0.057	552	0.056	641	0.055	843
70	0.057	396	0.056	486	0.055	508
75	0.057	242	0.057	325	0.056	352
80	0.058	154	0.057	182	0.056	231

Tabela 6.42: Ocjenjeni pričakovani transformirani časi za pacientke, diagnosticirane v 1. stadiju pljučnega raka; parametri modela, ocjenjeni na podatkih pod prelomno točko $Y > 0,03$

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.172	11861	0.211	12829	0.258	13719
45	0.159	9502	0.196	10745	0.240	11663
50	0.148	7196	0.182	8821	0.223	9678
55	0.138	5312	0.169	6812	0.207	7793
60	0.128	3668	0.157	4811	0.193	6060
65	0.119	2270	0.146	3235	0.179	4368
70	0.111	1380	0.136	2013	0.166	2869
75	0.103	726	0.126	1169	0.155	1719
80	0.095	379	0.117	568	0.144	912

Tabela 6.43: Ocenjeni pričakovani transformirani časi za pacientke, diagnosticirane v 2. stadiju pljučnega raka; parametri modela, ocenjeni na podatkih pod prelomno točko $Y > 0,03$

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.082	8318	0.087	9051	0.091	9587
45	0.080	6253	0.085	7206	0.089	7786
50	0.078	4747	0.082	5768	0.087	6199
55	0.076	3551	0.080	4124	0.084	4904
60	0.074	2372	0.078	2874	0.082	3794
65	0.072	1409	0.076	1943	0.080	2612
70	0.071	904	0.074	1146	0.078	1547
75	0.069	486	0.072	655	0.076	889
80	0.067	263	0.071	340	0.074	469

Tabela 6.44: Ocenjeni pričakovani transformirani časi za pacientke, diagnosticirane v 3. stadiju pljučnega raka; parametri modela, ocenjeni na podatkih pod prelomno točko $Y > 0,03$

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.061	6816	0.060	7564	0.059	7955
45	0.061	5352	0.060	5915	0.060	6315
50	0.062	4094	0.061	4626	0.060	4928
55	0.062	3055	0.061	3324	0.061	3870
60	0.063	2070	0.062	2311	0.061	3108
65	0.063	1226	0.063	1648	0.062	2136
70	0.064	824	0.063	976	0.062	1257
75	0.065	456	0.064	578	0.063	744
80	0.065	255	0.064	308	0.063	401

6.4.2 Buckley-Jamesova linearja regresija na podatkih o raku na dojki

Pri podatkih o raku na dojki lahko ugotovimo, da je bilo v proučevanem obdobju (od 1. januarja 1983 do 30. decembra 2004) z boleznijo diagnosticiranih 15622 žensk. Za namene analiz smo iz podatkov izključili tri paciente, ki so v tem obdobju emigrirale, in 205 pacientek, za katere je podatek o stadiju bolezni neznan oziroma nerazpoložljiv. Pacienteke so bile opazovane do 30. junija 2007, v tem času pa jih je 49,7 % umrlo. V modeliranje transformiranega časa smo vključili spremenljivki, na podlagi katerih smo določili vrednosti transformiranega časa (starost in leto diagnoze), stadij, v katerem so bile paciente z rakom na dojki diagnosticirane, in indikator operativnega zdravljenja, ki se je v Aalenovem modelu izkazal kot pomemben dejavnik preživetja pacientek. Vse neodvisne spremenljivke so brez manjkajočih vrednosti.

Pri analizi podatkov o raku na dojki z Buckley-Jamesovim modelom smo kot odvisno spremenljivko uporabili naravni logaritem transformiranega časa ($\log(Y)$). Podobno kot pri modeliranju podatkov pljučnega raka je tudi tu transformacija izvedena zato, da se izognemo negativnim vrednostim transformiranih časov v izvirnem modelu, na drugi strani pa z logaritmiziranjem odvisne spremenljivke v primeru heteroskedastičnosti lahko stabiliziramo varianco modela.

Model je narejen pod predpostavko, da so učinki starosti in leta diagnoze v različnih stadijih različni, zato so vanj vključene interakcije med njimi. Na podlagi izvedenih testov, prikazanih v tabeli 6.45, lahko ugotovimo, da je prileganje modela ustrezno.

Interpretacija modela

Rezultati ocenjenega modela so prikazani v tabeli 6.46. Napaka modela je ocenjena z $\hat{\sigma} = \sqrt{\frac{\hat{Y}-Y}{n_{nc}-p-1}}$, kjer je n_{nc} število nekrnjениh enot, p pa število neodvisnih spremenljivk, vključenih v model, in znaša $\hat{\sigma} = 1,410$.

Interpretacija ocenjenega modela je povzeta v naslednjih odstavkih:

Tabela 6.45: Waldove statistike za $\log(Y)$

	χ^2	d.f.	P
STAROST (Factor+Higher Order Factors)	2561.69	3	< 0.0001
<i>All Interactions</i>	63.52	2	< 0.0001
LETO (Factor+Higher Order Factors)	99.91	3	< 0.0001
<i>All Interactions</i>	7.42	2	0.0244
STADIJ (Factor+Higher Order Factors)	1939.71	6	< 0.0001
<i>All Interactions</i>	66.87	4	< 0.0001
OPERACIJA	818.13	1	< 0.0001
STAROST × STADIJ (Factor+Higher Order Factors)	63.52	2	< 0.0001
LETO × STADIJ (Factor+Higher Order Factors)	7.42	2	0.0244
TOTAL INTERACTION	66.87	4	< 0.0001
TOTAL	7283.82	9	< 0.0001

- Preživetje bolnic, diagnosticiranih v drugem oziroma v tretjem stadiju, se v primerjavi s preživetjem bolnic, diagnosticiranih v prvem stadiju raka na dojki, poslabša. Višji ko je stadij odkrite bolezni, slabše so možnosti preživetja bolnic.
- Operativno zdravljenje bolnic z rakom na dojki povečuje njihovo preživetje.
- Vpliv starosti na preživetje bolnic z rakom na dojki je pozitiven, kar pomeni, da je v primerjavi z vrstnicami splošne populacije preživetje starejših bolnic večje od preživetja mlajših bolnic. V višjih stadijih se vpliv starosti na preživetje dodatno poveča.
- Učinek leta diagnoze je pozitiven, vendar ta vpliv v višjih stadijih slablji. Relativno preživetje bolnic, diagnosticiranih v prvem stadiju bolezni, se torej s časom izboljšuje, vendar je ta učinek v višjih stadijih manjši.

Aalenov model je na podatkih raka na dojki pokazal, da se učinka starosti in leta diagnoze med vrednostima transformiranega časa $0,1 < Y < 0,2$ izgubita. S pomočjo dodatnih analiz smo uspeli empirično potrditi predpostavko, da za operativno zdravljene bolnice, diagnosticirane v prvem stadiju bolezni, to statistično gledano predstavlja

Tabela 6.46: Rezultati ocenjenega modela za podatke o raku na dojki

	Value	Std. Error	Z	$Pr(> Z)$
Konstanta	-6.748	0.211	-31.918	0.000
STAROST	1.335e-4	5.991e-6	22.278	0.000
LETO	9.943e-5	1.430e-5	6.952	0.000
STADIJ=2	-1.893	0.248	-7.624	0.000
STADIJ=3	-3.043	0.338	-8.998	0.000
OPERACIJA=1	1.255	0.044	28.603	0.000
STAROST*STADIJ=2	5.277e-5	7.286e-6	7.243	0.000
STAROST*STADIJ=3	6.247e-5	9.882e-6	6.322	0.000
LETO*STADIJ=2	-2.548e-5	1.800e-5	-1.416	0.157
LETO*STADIJ=3	-5.946e-5	2.182e-5	-2.725	0.006

ozdravitev. Ali do podobnega sklepa lahko pridemo tudi z Buckley-Jamesovo metodo modeliranja podatkov, smo preverjali podobno kot pri podatkih o pljučnem raku. Podatke smo razdelili v dela pod in nad prelomno točko, pri čemer smo za prelomno točko privzeli vrednost transformiranega časa $Y = 0,15$. Rezultati prileganja modelov, ocenjenih na podatkih pod in nad prelomno točko transformiranega časa, so prikazani v tabelah 6.47 in 6.48, ocenjeni parametri modela pa v tabelah 6.49 in 6.50.

Rezultati kažejo, da v modelu, ocenjenem na podatkih nad prelomno točko transformiranega časa Y , ocenjena parametra starosti in leta diagnoze nista statistično značilna. Vendar pa je hkrati nejasen vpliv drugega stadija na preživetje bolnic, diagnosticiranih z rakom na dojki, saj so vsi ocenjeni parametri drugega stadija neznačilni. Da bi vpliv drugega stadija omejili zgolj na en parameter (v modelu z interakcijami se vpliv drugega stadija odraža tudi preko interakcij stadija s starostjo in letom diagnoze), smo na podatkih nad prelomno točko transformiranega časa preizkusili tudi model brez vključenih interakcij. Enostavni model se na podlagi testov, prikazanih v tabeli 6.51, ustrezeno prilega podatkom, ocenjeni parametri modela (tabela 6.52) podobno kot v modelu z interakcijami kažejo na neznačilen vpliv starosti in leta diagnoze, vpliv

drugega stadija pa se izkaže za statistično značilnega.

Tabela 6.47: Waldove statistike za $\log(Y)$, $Y \leq 0,15$

	χ^2	d.f.	P
STAROST (Factor+Higher Order Factors)	734.36	3	< 0.0001
<i>All Interactions</i>	159.91	2	< 0.0001
LETO (Factor+Higher Order Factors)	362.09	3	< 0.0001
<i>All Interactions</i>	78.73	2	< 0.0001
STADIJ (Factor+Higher Order Factors)	973.04	6	< 0.0001
<i>All Interactions</i>	199.64	4	< 0.0001
OPERACIJA	564.98	1	< 0.0001
STAROST \times STADIJ (Factor+Higher Order Factors)	159.91	2	< 0.0001
LETO \times STADIJ (Factor+Higher Order Factors)	78.73	2	< 0.0001
TOTAL INTERACTION	199.64	4	< 0.0001
TOTAL	3534.57	9	< 0.0001

Tabela 6.48: Waldove statistike za $\log(Y)$, $Y > 0, 15$

	χ^2	d.f.	P
STAROST (Factor+Higher Order Factors)	3.26	3	0.3539
<i>All Interactions</i>	0.35	2	0.8399
LETO (Factor+Higher Order Factors)	4.25	3	0.2354
<i>All Interactions</i>	4.10	2	0.1285
STADIJ (Factor+Higher Order Factors)	98.36	6	< 0.0001
<i>All Interactions</i>	5.73	4	0.2200
OPERACIJA	44.03	1	< 0.0001
STAROST \times STADIJ (Factor+Higher Order Factors)	0.35	2	0.8399
LETO \times STADIJ (Factor+Higher Order Factors)	4.10	2	0.1285
TOTAL INTERACTION	5.73	4	0.2200
TOTAL	151.43	9	< 0.0001

Tabela 6.49: Rezultati ocenjenega modela za podatke o raku na dojki, $Y \leq 0, 15$

	Value	Std. Error	Z	$Pr(> Z)$
Konstanta	-6.961	0.242	-28.817	0.000
STAROST	1.745e-5	7.701e-6	2.266	0.023
LETO	2.582e-4	1.761e-5	14.657	0.000
STADIJ=2	-1.273	0.280	-4.548	0.000
STADIJ=3	-2.040	0.354	-5.758	0.000
OPERACIJA=1	1.206	0.051	23.769	0.000
STAROST*STADIJ=2	9.071e-5	9.137e-6	9.928	0.000
STAROST*STADIJ=3	1.334e-4	1.107e-5	12.044	0.000
LETO*STADIJ=2	-1.194e-4	2.118e-5	-5.636	0.000
LETO*STADIJ=3	-2.129e-4	2.401e-5	-8.868	0.000

$n=11702$, št. nekrnjenih enot=5911, $\hat{\sigma} = 1,359$

Tabela 6.50: Rezultati ocenjenega modela za podatke o raku na dojki, $Y > 0,15$

	Value	Std. Error	Z	$Pr(> Z)$
Konstanta	-1.019	0.181	-5.625	0.000
STAROST	7.036e-6	6.585e-6	1.069	0.285
LETO	-2.873e-6	9.233e-6	-0.311	0.756
STADIJ=2	-0.317	0.241	-1.317	0.188
STADIJ=3	-1.566	0.728	-2.152	0.031
OPERACIJA=1	0.195	0.029	6.636	0.000
STAROST*STADIJ=2	1.923e-6	9.186e-6	0.209	0.834
STAROST*STADIJ=3	1.497e-5	2.585e-5	0.579	0.562
LETO*STADIJ=2	5.253e-6	1.337e-5	0.393	0.695
LETO*STADIJ=3	6.336e-5	3.128e-5	2.026	0.043

$n=3711$, št. nekrnjenih enot=1750, $\hat{\sigma} = 0,499$

Tabela 6.51: Waldove statistike za $\log(Y)$, $Y > 0.15$, enostavni model

	χ^2	d.f.	P
STAROST	2.36	1	0.1245
LETO	0.10	1	0.7518
STADIJ	89.55	2	< 0.0001
OPERACIJA	46.37	1	< 0.0001
TOTAL	141.32	5	< 0.0001

Tabela 6.52: Rezultati ocjenjenega modela za podatke o raku na dojki, $Y > 0, 15$, enos-tavni model

	Value	Std. Error	Z	$Pr(> Z)$
Konstanta	-1.095	0.144	-7.606	0.000
STAROST	7.608E-06	4.952E-06	1.536	0.125
LETO	2.055E-06	6.497E-06	0.316	0.752
STADIJ=2	-0.202	0.025	-8.172	0.000
STADIJ=3	-0.357	0.059	-6.037	0.000
OPERACIJA=1	0.199	0.029	6.809	0.000

$n=3711$, št. nekrnjnih enot=1750, $\hat{\sigma} = 0,499$

Pričakovane vrednosti individualnega relativnega preživetja pacientk z rakom na dojki

Bistvena prednost Buckley-Jamesovega modela je, da ocenjeni model lahko enostavno uporabimo za ocenjevanje pričakovanih transformiranih časov preživetja bolnic, diagnosticiranih z rakom na dojki. Tako za vsako kombinacijo neodvisnih spremenljivk modela lahko ocenimo pričakovano vrednost transformiranega časa, ki pove, kolikšen delež primerljive splošne populacije bo bolnica v povprečju preživila. Na drugi strani napovedovanje transformiranih časov omogoča, da določimo ustrezeno kombinacijo neodvisnih spremenljivk, ob kateri bo oboleli v povprečju preživel neki izbrani delež splošne populacije.

V tabelah 6.53 do 6.58 so prikazani ocenjeni transformirani časi za različne vrednosti starosti pacientk, ki so bile diagnosticirane z rakom na dojki v treh različnih obdobjih (1. januar 1983, 1. januar 1993 in 1. januar 2003). Tabele so ločene glede na to, ali so bile bolnice operativno zdravljene ali ne, prikazane pa so za vsak stadij raka na dojki posebej. Poleg ocenjenih vrednosti pričakovanih transformiranih časov so v tabelah prikazani tudi pripadajoči dejanski časi preživetja bolnic.

Interpretacije spodnjih tabel so podane za primer bolnice, ki je bila v času diagnoze stara 60 let, kar predstavlja približno povprečno starost bolnic, diagnosticiranih z rakom na dojki.

Na podlagi prikazanih rezultatov lahko rečemo, da 60-letna ženska, ki je bila leta 1983 diagnosticirana v prvem stadiju raka na dojki, v primeru brez operativnega zdravljenja lahko pričakuje, da bo v povprečju preživila 5 % primerljive splošne populacije ($Y=0,050$). V primeru operativnega zdravljenja se pričakovani transformirani čas (Y) poveča na 0,177. Ugotovili smo že, da se relativno preživetje žensk, diagnosticiranih v prvem stadiju bolezni, v času v povprečju izboljšuje. V tabelah 6.53 in 6.56 se to odrazi v povišanih vrednostih transformiranih časov pri višjih vrednostih leta diagnoze. Vidimo lahko, da se za 60-letno bolnico brez operativnega zdravljenja, ki je bila diagnosticirana z rakom na dojki leta 2003, pričakovani transformirani čas poveča na

0,104, za bolnico iste starosti z operativnim zdravljenjem pa na 0,365. To pomeni, da bo 60-letna bolnica brez operativnega zdravljenja, ki je bila diagnosticirana z boleznijo v letu 2003, v povprečju preživila 10,4 % primerljive splošne populacije, bolnica iste starosti, ki je bila zdravljena operativno, pa bo v povprečju preživila 36,5 % ustrezne splošne populacije.

Relativno preživetje žensk, diagnosticiranih v drugem stadiju bolezni, je slabše od relativnega preživetja žensk, diagnosticiranih v prvem stadiju bolezni. Za 60-letno bolnico brez operativnega zdravljenja, ki je bila leta 1983 diagnosticirana v drugem stadiju, tako lahko vidimo, da bo v povprečju preživila 1,9 % primerljive splošne populacije. Za operativno zdravljeno bolnico iste starosti se preživetje poveča na 6,8 % primerljive splošne populacije. Relativno preživetje žensk, diagnosticiranih v drugem stadiju bolezni, se skozi čas izboljšuje po podobni stopnji kot v prvem stadiju bolezni. Iz tabel 6.54 in 6.57 tako lahko vidimo, da se pričakovana vrednost transformiranega časa z 0,019 za bolnice brez operativnega zdravljenja, diagnosticirane leta 1983, poveča na 0,033 v letu 2003, za operativno zdravljene bolnice iste starosti pa se pričakovani transformirani čas v istem obdobju poveča z 0,068 na 0,117.

Tabeli 6.55 in 6.58 prikazujeta še ocenjene pričakovane transformirane čase za bolnice brez oziroma z operativnim zdravljenjem, ki so bile diagnosticirane v tretjem stadiju raka na dojki. Za 60-letno žensko, ki je bila diagnosticirana v tretjem stadiju bolezni leta 1983 in ni bila zdravljena operativno, je ocenjeni pričakovani transformirani čas 0,006, za operativno zdravljeno bolnico pa 0,02. To pomeni, da proučevana bolnica brez operativnega zdravljenja lahko pričakuje, da bo v povprečju preživila 0,6 % primerljive splošne populacije, operativno zdravljena bolnica pa 2 % primerljive splošne populacije. Na podlagi rezultatov je tudi pri bolnicah, diagnosticiranih v tretjem stadiju bolezni, opazno rahlo izboljševanje pričakovanih vrednosti opazovanega relativnega preživetja v času.

Tabela 6.53: Ocenjeni pričakovani transformirani časi za pacientke brez operativnega zdravljenja, diagnosticirane v 1. stadiju raka na dojki

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.019	3299	0.027	4504	0.039	6396
45	0.024	2713	0.035	3859	0.050	5740
50	0.031	2410	0.044	3490	0.064	5139
55	0.039	2002	0.057	3113	0.082	4793
60	0.050	1761	0.072	2676	0.104	4413
65	0.064	1243	0.092	2270	0.133	3616
70	0.082	1042	0.118	1804	0.170	2910
75	0.105	740	0.151	1412	0.216	2271
80	0.134	528	0.192	947	0.276	1695

Tabela 6.54: Ocenjeni pričakovani transformirani časi za pacientke brez operativnega zdravljenja, diagnosticirane v 2. stadiju raka na dojki

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.005	1125	0.007	1855	0.009	2839
45	0.007	814	0.009	1363	0.012	2065
50	0.010	894	0.013	1279	0.017	1781
55	0.014	720	0.018	1262	0.024	1843
60	0.019	748	0.026	995	0.033	1934
65	0.027	574	0.036	1037	0.047	1680
70	0.038	500	0.050	784	0.066	1330
75	0.054	384	0.071	641	0.093	1067
80	0.076	299	0.099	480	0.130	823

Tabela 6.55: Ocjenjeni pričakovani transformirani časi za pacientke brez operativnega zdravljenja, diagnosticirane v 3. stadiju raka na dojki

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.001	292	0.002	575	0.002	655
45	0.002	246	0.002	391	0.003	534
50	0.003	275	0.003	419	0.004	417
55	0.004	213	0.005	327	0.005	392
60	0.006	231	0.007	260	0.008	531
65	0.008	175	0.009	307	0.011	417
70	0.012	152	0.014	204	0.016	357
75	0.017	117	0.019	183	0.022	271
80	0.024	92	0.028	131	0.032	202

Tabela 6.56: Ocjenjeni pričakovani transformirani časi za operativno zdravljene paciente, diagnosticirane v 1. stadiju raka na dojki

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.067	7207	0.096	9511	0.138	11361
45	0.085	6480	0.122	8873	0.176	10486
50	0.109	5902	0.156	8281	0.224	9696
55	0.138	5330	0.199	7397	0.286	8880
60	0.177	4535	0.254	6506	0.365	8083
65	0.225	3685	0.324	5646	0.466	7196
70	0.288	3122	0.414	4839	0.595	6484
75	0.367	2483	0.528	4176	0.759	6051
80	0.468	1924	0.674	3537	0.968	6977

Tabela 6.57: Ocenjeni pričakovani transformirani časi za operativno zdravljene pacientke, diagnosticirane v 2. stadiju raka na dojki

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.018	3141	0.023	4074	0.030	5545
45	0.025	2738	0.032	3662	0.042	5050
50	0.035	2627	0.045	3545	0.059	4891
55	0.049	2457	0.064	3424	0.083	4863
60	0.068	2197	0.089	3222	0.117	4682
65	0.096	1904	0.126	2870	0.165	4155
70	0.135	1651	0.177	2515	0.231	3614
75	0.189	1305	0.248	2210	0.325	3106
80	0.266	1058	0.349	1712	0.457	2713

Tabela 6.58: Ocenjeni pričakovani transformirani časi za operativno zdravljene pacientke, diagnosticirane v 3. stadiju raka na dojki

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.005	1090	0.006	1677	0.006	2156
45	0.007	795	0.008	1202	0.009	1686
50	0.010	892	0.011	1152	0.013	1441
55	0.014	731	0.016	1156	0.019	1404
60	0.020	772	0.023	918	0.027	1610
65	0.029	602	0.033	967	0.039	1433
70	0.041	535	0.048	743	0.055	1128
75	0.059	417	0.068	617	0.079	917
80	0.084	333	0.097	470	0.113	710

Tabele 6.60 do 6.71 prikazujejo pričakovane vrednosti transformiranega časa in pripadajoče dejanske čase preživetja, ocenjene na podlagi modelov pod (tabele 6.60 do 6.65) in nad (tabele 6.66 do 6.71) prelomno točko $Y = 0,15$. Z Aalenovim modelom smo ugotovili, da učinka starosti in leta diagnoze med vrednostmi transformiranega časa $0,1 < Y < 0,2$ izgubita vpliv na opazovano relativno preživetje, na podlagi česar smo empirično uspeli pokazati, da za operativno zdravljene bolnice, diagnosticirane v prvem stadiju raka na dojki, to statistično pomeni ozdravitev. Tabela 6.69 kaže pričakovane vrednosti transformiranih časov za operativno zdravljene bolnice, diagnosticirane v prvem stadiju bolezni, iz katerih je razvidno, da je njihov pričakovani transformirani čas preživetja pri vseh starostih blizu 0,5, kar pomeni, da te bolnice lahko pričakujejo, da bodo preživele približno 50 % vrstnic v splošni populaciji. To bi lahko označili kot čas, ki s statističnega vidika predstavlja ozdravitev.

Tabela 6.59 prikazuje izvirne čase, ki ob izbrani kombinaciji starosti in leta diagnoze ustrezajo vrednosti transformiranega časa $Y = 0,15$. Ta čas smo izbrali kot prelomno točko na podlagi Aalenovega modela, pri kateri za operativno zdravljene bolnice, diagnosticirane v prvem stadiju bolezni, s statističnega vidika pride do ozdravitve. Izvirni časi ob prelomni točki $Y = 0,15$ nudijo pomoč pri interpretaciji pričakovanih časov opazovanega relativnega preživetja, ocjenjenega na podlagi modelov pod in nad prelomno točko, vendar jih v tem delu podrobnejše ne opisujemo.

Tabela 6.59: Izvirni čas T , ki ustreza transformiranemu času $Y = 0,15$, za izbrano kombinacijo starosti in leta diagnoze

Starost v letih	ženske		
	1983	1993	2003
40	11290	11530	11699
45	9159	9753	9933
50	7243	8149	8272
55	5599	6409	6718
60	4053	4659	5319
65	2726	3306	3925
70	1819	2182	2659
75	1052	1407	1672
80	594	734	953

Tabela 6.60: Ocenjeni pričakovani transformirani časi za pacientke brez operativnega zdravljenja, diagnosticirane v 1. stadiju raka na dojki, $Y \leq 0,15$

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.011	2231	0.027	4526	0.071	8534
45	0.011	1358	0.028	3269	0.073	6993
50	0.011	1006	0.029	2457	0.075	5719
55	0.012	614	0.03	1927	0.078	4612
60	0.012	479	0.031	1199	0.08	3723
65	0.013	269	0.032	938	0.083	2664
70	0.013	168	0.033	512	0.085	1670
75	0.013	93	0.034	326	0.088	1019
80	0.014	52	0.035	168	0.091	574

Tabela 6.61: Ocjenjeni pričakovani transformirani časi za pacientke brez operativnega zdravljenja, diagnosticirane v 2. stadiju raka na dojki, $Y \leq 0,15$

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.004	894	0.007	1904	0.011	3327
45	0.005	589	0.008	1256	0.014	2296
50	0.006	594	0.01	1058	0.017	1782
55	0.007	396	0.012	870	0.021	1556
60	0.009	367	0.015	614	0.025	1518
65	0.011	237	0.018	559	0.031	1132
70	0.014	175	0.022	340	0.037	781
75	0.016	115	0.027	259	0.045	544
80	0.02	77	0.033	158	0.055	354

Tabela 6.62: Ocjenjeni pričakovani transformirani časi za pacientke brez operativnega zdravljenja, diagnosticirane v 3. stadiju raka na dojki, $Y \leq 0,15$

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.002	348	0.002	729	0.002	816
45	0.002	270	0.003	442	0.003	594
50	0.003	278	0.003	430	0.004	434
55	0.004	198	0.004	310	0.005	377
60	0.005	197	0.006	227	0.007	477
65	0.006	138	0.008	246	0.009	346
70	0.009	110	0.01	151	0.012	269
75	0.011	78	0.013	124	0.016	188
80	0.015	56	0.017	81	0.021	129

Tabela 6.63: Ocenjeni pričakovani transformirani časi za operativno zdravljene pacientke, diagnosticirane v 1. stadiju raka na dojki, $Y \leq 0,15$

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.036	5009	0.092	9315	0.236	13381
45	0.037	3692	0.095	7719	0.243	11713
50	0.038	2873	0.098	6450	0.251	10096
55	0.039	1996	0.101	4946	0.259	8537
60	0.041	1479	0.104	3601	0.268	7059
65	0.042	844	0.108	2538	0.277	5553
70	0.043	563	0.111	1700	0.286	4134
75	0.045	318	0.115	1059	0.295	2895
80	0.046	179	0.119	575	0.304	1853

Tabela 6.64: Ocenjeni pričakovani transformirani časi za operativno zdravljene pacientke, diagnosticirane v 2. stadiju raka na dojki, $Y \leq 0,15$

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.014	2678	0.023	4076	0.038	6293
45	0.017	1976	0.028	3234	0.046	5430
50	0.021	1716	0.034	2799	0.057	4744
55	0.025	1291	0.042	2430	0.069	4228
60	0.03	1155	0.051	1888	0.084	3851
65	0.037	764	0.062	1629	0.102	3053
70	0.045	588	0.075	1158	0.125	2316
75	0.055	392	0.092	831	0.152	1691
80	0.067	264	0.111	539	0.185	1176

Tabela 6.65: Ocenjeni pričakovani transformirani časi za operativno zdravljene pacientke, diagnosticirane v 3. stadiju raka na dojki, $Y \leq 0,15$

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.005	1206	0.006	1839	0.008	2586
45	0.007	837	0.008	1270	0.01	1830
50	0.009	865	0.011	1135	0.013	1442
55	0.012	649	0.015	1056	0.017	1295
60	0.016	635	0.019	783	0.023	1401
65	0.022	458	0.026	753	0.03	1113
70	0.028	371	0.034	517	0.04	826
75	0.038	265	0.044	414	0.052	623
80	0.049	192	0.058	279	0.069	436

Tabela 6.66: Ocenjeni pričakovani transformirani časi za pacientke brez operativnega zdravljenja, diagnosticirane v 1. stadiju raka na dojki, $Y > 0,15$

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.391	15197	0.387	15218	0.382	15223
45	0.396	13360	0.392	13463	0.387	13475
50	0.401	11513	0.397	11750	0.392	11749
55	0.406	9619	0.402	9993	0.398	10047
60	0.411	7715	0.407	8235	0.403	8404
65	0.417	5872	0.412	6520	0.408	6752
70	0.422	4183	0.417	4870	0.413	5173
75	0.427	2852	0.423	3472	0.418	3728
80	0.433	1767	0.428	2126	0.424	2536

Tabela 6.67: Ocenjeni pričakovani transformirani časi za pacientke brez operativnega zdravljenja, diagnosticirane v 2. stadiju raka na dojki, $Y > 0,15$

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.306	14214	0.308	14333	0.311	14427
45	0.311	12363	0.314	12587	0.316	12691
50	0.316	10497	0.319	10914	0.322	10997
55	0.321	8565	0.324	9182	0.327	9334
60	0.327	6778	0.329	7442	0.332	7747
65	0.332	4964	0.335	5758	0.338	6170
70	0.337	3538	0.34	4189	0.343	4639
75	0.343	2325	0.346	2929	0.349	3279
80	0.349	1405	0.352	1727	0.355	2143

Tabela 6.68: Ocenjeni pričakovani transformirani časi za pacientke brez operativnega zdravljenja, diagnosticirane v 3. stadiju raka na dojki, $Y > 0,15$

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.173	11891	0.216	12930	0.269	13873
45	0.18	10041	0.225	11312	0.28	12228
50	0.187	8165	0.234	9763	0.291	10640
55	0.195	6517	0.243	8133	0.303	9085
60	0.203	4989	0.253	6495	0.316	7578
65	0.211	3515	0.264	4930	0.329	6078
70	0.22	2523	0.274	3607	0.342	4631
75	0.229	1563	0.286	2469	0.356	3325
80	0.238	945	0.297	1472	0.371	2236

Tabela 6.69: Ocjenjeni pričakovani transformirani časi za operativno zdravljene paciente, diagnosticirane v 1. stadiju raka na dojki, $Y > 0, 15$

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.475	15987	0.47	15989	0.465	15979
45	0.481	14170	0.476	14228	0.471	14223
50	0.487	12347	0.482	12494	0.477	12482
55	0.494	10491	0.488	10734	0.483	10763
60	0.5	8603	0.495	8974	0.49	9090
65	0.506	6714	0.501	7247	0.496	7421
70	0.513	4888	0.508	5576	0.502	5819
75	0.52	3432	0.514	4084	0.509	4338
80	0.526	2194	0.521	2624	0.515	3039

Tabela 6.70: Ocjenjeni pričakovani transformirani časi za operativno zdravljene paciente, diagnosticirane v 2. stadiju raka na dojki, $Y > 0, 15$

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.372	14987	0.375	15091	0.378	15177
45	0.378	13161	0.381	13347	0.385	13443
50	0.384	11325	0.388	11666	0.391	11736
55	0.391	9441	0.394	9924	0.397	10046
60	0.397	7571	0.4	8178	0.404	8414
65	0.404	5741	0.407	6474	0.411	6774
70	0.41	4095	0.414	4840	0.417	5202
75	0.417	2788	0.421	3455	0.424	3766
80	0.424	1727	0.428	2122	0.431	2577

Tabela 6.71: Ocenjeni pričakovani transformirani časi za operativno zdravljene pacientke, diagnosticirane v 3. stadiju raka na dojki, $Y > 0,15$

Starost v letih	1Jan1983		1Jan1993		1Jan2003	
	Y	T	Y	T	Y	T
40	0.21	12671	0.262	13697	0.327	14620
45	0.219	10862	0.273	12050	0.34	12960
50	0.228	9036	0.284	10481	0.354	11348
55	0.237	7307	0.296	8848	0.369	9784
60	0.247	5685	0.308	7204	0.384	8242
65	0.257	4066	0.321	5608	0.4	6688
70	0.268	2955	0.334	4132	0.416	5194
75	0.279	1893	0.347	2939	0.433	3827
80	0.29	1155	0.362	1776	0.451	2682

7. Zaključek

Kadar končnih dogodkov ne moremo z gotovostjo pripisati proučevanemu pojavu ali vzroka zanj ne poznamo, časov preživetja ne moremo proučevati z metodami klasične analize preživetja, ampak jih proučujemo v okviru metodologije relativnega preživetja. Relativno preživetje je bilo prvotno zasnovano v smislu primerjave preživetja med opazovanimi enotami in splošno populacijo, na tej podlagi pa so bile razvite tudi posebne metode za analizo. Alternativo prvotno definirani „skupinski“ meri relativnega preživetja predstavlja individualna mera relativnega preživetja (ali transformirani čas preživetja), ki, kot pove že ime samo, relativno preživetje definira na individualni ravni. Temelji na preprosti transformaciji časov preživetja, s katero proučevani čas opazovane enote s pomočjo pogojnih porazdelitvenih funkcij časov preživetja, določenih na splošni popoulaciji, preslikamo v individualno mero relativnega preživetja. Ta nam tako pove, kolikšen delež splošne populacije je opazovana enota preživila. Na ta način individualna mera relativnega preživetja že sama po sebi prinaša drugačen vpogled v relativno preživetje. Poleg tega je njena prednost v tem, da za analizo podatkov lahko uporabimo metode klasične analize preživetja, ker so vrednosti relativnega preživetja definirane za vsako enoto posebej. Doktorska disertacija je osredotočena predvsem na empirično modeliranje relativnega preživetja s pomočjo metodologije klasične analize preživetja, ki se v praksi sicer redkeje uporablja. Jedro naloge tako predstavljajo rezultati Aalenove linearne regresije ogroženosti in Buckley-Jamesove linearne regresije transformiranih časov preživetja. Predvsem zaradi primerjave z Aalenovim modelom pa so analize narejene še z v praksi najpogosteje uporabljenou Coxovo regresijo sorazmernih ogroženosti. Vsi modeli so preizkušeni na podatkih o raku na pljučih in raku na dojki, ki se med seboj najbolj razlikujejo glede na delež krnjениh enot, s tem pa tudi glede na preživetje opazovanih enot. V doktorski disertaciji obravna-

vani pristop k relativnemu preživetju je popolnoma nov in kot tak prinaša možnosti poglobljenega vedenja o relativnem preživetju, njegova uporaba pa omogoča možnosti pridobivanja novih informacij o relativnem preživetju, kot tudi vsebinsko obogatitev področja relativnega preživetja. Prednost uporabe novega pristopa k modeliranju relativnega preživetja je poleg tega tudi v tem, da v primerjavi s tradicionalnimi regresijskimi metodami za modeliranje relativnega preživetja (aditivni in multiplikativni modeli) ne zahteva nikakršnih predpostavk o povezanosti opazovane in populacijske ogroženosti.

Rezultati ocenjenih modelov so pokazali, da z vsemi tremi uporabljenimi modeli lahko pridemo do konsistentnih informacij o učinkih neodvisnih spremenljivk na relativno preživetje. Kljub temu je pri uporabi modelov in interpretaciji rezultatov treba biti previden, kar velja zlasti v primeru ko se učinki neodvisnih spremenljivk na relativno preživetje v času spreminjajo. V nalogi je pokazano, da je v primeru ko učinke generiramo z Aalenovim modelom, parameter klasične linearne regresije (Buckley-Jamesov model v primeru krnjениh podatkov) odvisen tako od parametrov Aalenovega modela, kot tudi od kombinacije neodvisnih spremenljivk. Za sam model to predstavlja kršenje predpostavke o linearnosti in homoskedastičnosti, vendar je treba upoštevati, da je uporabnost modela odvisna od stopnje nelinearnosti in heteroskedastičnosti. V primeru nespreminjajočih parametrov Aalenovega modela so vse vrednosti odvisne spremenljivke pri izbrani kombinaciji neodvisnih spremenljivk porazdeljene z enakim povprečjem in enako varianco. Na podlagi tega lahko sklepamo, da bodo v primeru majhne stopnje nelinearnosti in heteroskedastičnosti rezultati Buckley-Jamesove regresije z vidika praktične uporabe prinesli koristne informacije. V primeru časovno odvisnih učinkov Aalenovega modela pa je parameter klasične linearne regresije odvisen od pogojne porazdelitve odvisne spremenljivke (pri izbrani vrednosti neodvisne spremenljivke), ki v tem primeru predstavlja mešanico različnih pogojnih porazdelitev, odvisnih od vrednosti parametrov Aalenovega modela pri različnih vrednostih proučevanega časa. To pomeni, da sta pričakovana vrednost in varianca opazovanega časa (odvisne spremenljivke) pri enaki vrednosti neodvisnih spremenljivk x_i različni. Spreminjajočega učinka parametrov Aalenovega modela klasična linearna regresija torej ne zazna, saj

predpostavlja, da so podatki pri x_i generirani z enako pričakovano vrednostjo in varianco. Na drugi strani pa je s simulacijami mogoče pokazati, da v primeru ko podatke generiramo s klasično linearo regresijo tako, da vrednosti parametrov pri enakih vrednostih neodvisnih spremenljivk sprememimo glede na izbrane časovne točke, te spremembe lahko zaznamo z Aalenovim modelom. Ker v praksi ne vemo in pogosto niti s prilagoditvenimi testi ne zaznamo, kateremu modelu se podatki bolj prilegajo, s konsistentnostjo modelov lahko potrdimo uporabnost dobljenih rezultatov. V primeru ko se učinki neodvisnih spremenljivk v času spreminja, Buckley-Jamesovo regresijo lahko naredimo na intervalih, na katerih so vrednosti parametrov Aalenovega modela vsaj približno konstantni.

Analize prikazane v nalogi niso namenjene preverjanju kliničnih predpostavk, temveč metodološkemu prikazu delovanja uporabljenih modelov. Ocenjeni učinki populacijskih spremenljivk so odraz presežne ogroženosti za smrt proučevanih enot in kot taki logična posledica manjšega preživetja bolnikov v primerjavi s populacijskim preživetjem. Upoštevati je potrebno, da v primerih visoke umrljivosti, ki je pogosto večja prav v začetnem obdobju bolezni, populacijske značilnosti bistveno ne vplivajo na opazovani čas preživetja. Pri modeliranju individualne mere relativnega preživetja bo v takih primerih učinek starosti na ogroženost za smrt vedno negativen, učinek spola pa pozitivno večji za ženske v primerjavi z moškimi. To odraža zgolj dejstvo, da mlajši bolniki in ženske izgubljajo svojo populacijsko prednost v preživetju. Pri podatkih o pljučnem raku se denimo izkaže, da se pri vrednosti transformiranega časa $Y = 0.03$ učinek starosti na ogroženost za smrt spremeni iz negativnega v pozitivnega, učinek spola (žensk v primerjavi z moškimi) pa iz pozitivnega v negativnega. Znano je, da je smrtnost bolnikov, diagnosticiranih s pljučnim rakom, izredno velika. Večina bolnikov umre v prvih dveh letih od začetka diagnoze, petletno relativno preživetje za bolnike diagnosticirane v obdobju 2001-2005 pa je le 12 % za moške in 13 % za ženske. Incidenca pljučnega raka je največja v starostnih skupinah nad 60 let in pri moških. Za natančnejšo proučitev prelomne točke $Y = 0.03$ bi bile sicer potrebne podrobnejše klinične raziskave, vendar na podlagi proučevanega vzorca enot lahko sklepamo, da največji učinek predstavlja visoka smrtnost v prvih dveh letih opazovanja. Povprečni

dejanski časi preživetja pri vrednosti transformiranega časa $Y = 0.03$, se za moške z začetkom opazovanja v različnih obdobjih, gibljejo med dvemi in tremi leti, kar je razvidno tudi iz tabele 6.32.

Uporaba modelov na podatkih o raku na pljučih in raku na dojki je pokazala, da se v primeru, ko v model vključimo interakcije neodvisnih spremenljivk, na podlagi katerih določimo individualno mero relativnega preživetja, z drugimi neodvisnimi spremenljivkami, del presežne ogroženosti glavnih vplivov drugih neodvisnih spremenljivk prenese na interakcije. To je problematično zlasti v primeru Aalenovega modela, s katerim učinke neodvisnih spremenljivk prikažemo grafično, kvantitativno pa jih težko ovrednotimo. V tem primeru je zaradi interpretacije rezultatov bolj smiselno proučevati model, ki vključuje samo glavne vplive. Prednost Aalenovega modela je predvsem v njegovi prilagodljivosti. Tako smo z njim na obeh podatkovjih hitro zaznali, da se učinki določenih neodvisnih spremenljivk v času spremenijo. Čeprav je modeliranje časovno odvisnih učinkov možno tudi v Coxovem modelu, je njegova slabost ta, da moramo na eni strani za tako modeliranje ustrezno preurediti vhodne podatke, na drugi strani pa je izbira prelomnih točk popolnoma arbitralna. Kljub temu s Coxovim modelom lahko pridemo do dodatnih informacij o vplivu neodvisnih spremenljivk na relativno preživetje, saj učinke neodvisnih spremenljivk lahko kvantitativno ovrednotimo. Tako denimo lahko ugotovimo, da so v primeru obeh podatkovij spremembe učinkov parametrov najhitrejše in pogosto tudi najmočnejše v začetnem obdobju, kar opozarja na začetno močnejšo agresivnost obeh vrst raka. Prednost Buckley-Jamesove regresije je, da lahko z ocenjenimi parametri brez vsakokratne uporabe vzorca ocenimo pričakovane vrednosti individualne mere relativnega preživetja, posredno pa tudi pričakovane dejanske čase preživetja. Njegova največja slabost pa je v tem, da z modelom ne moremo zaznati časovno odvisnih učinkov neodvisnih spremenljivk. Kljub temu smo z modeliranjem podatkov na časovnih intervalih med prelomnimi točkami, na katerih so se učinki Aalenovega modela izkazali za vsaj približno konstatne, uspeli pokazati, da tudi ocenjeni parametri Buckley-Jamesovega modela odražajo rezultate, konsistentne z Aalenovim in Coxovim modelom. To pomeni, da ocnjene pričakovane vrednosti individualne mere relativnega preživetja in pri-

padajočih dejanskih časov preživetja lahko uporabimo kot koristno informacijo v medicinski diagnostiki.

Prilagodljivost Aalenovega modela v smislu zaznavanja časovno odvisnih učinkov neodvisnih spremenljivk je v primeru podatkov o raku na dojki³ pokazala, da z modeliranjem individualne mere relativnega preživetja lahko ugotavljamo tudi, kdaj se preživetje obravnavanih bolnikov izenači s preživetjem splošne populacije. Na tem mestu naj omenimo, da s transformacijo opazovanega časa v individualno mero relativnega preživetja upoštevamo vso populacijsko ogroženost. Ostane nam le presežna ogroženost, ki jo tudi neposredno modeliramo. Na drugi strani je ena od lastnosti individualne mere relativnega preživetja ta, da je njena porazdelitev na splošni populaciji enakomerna na intervalu $[0, 1]$, kar pomeni, da se presežna ogroženost, ki je lastna proučevani bolezni, odrazi v spremenjeni porazdelitvi individualne mere relativnega preživetja. Aalenov model je pokazal, da učinka starosti in leta diagnoze⁴ med vrednostima individualne mere relativnega preživetja 0,1 in 0,2 postaneta statistično neznačilna. To pomeni, da na tem časovnem intervalu presežna ogroženost za smrt za rakom na dojki izgine oziroma se ogroženost za smrt opazovanih bolnic izenači z ogroženostjo za smrt v splošni populaciji. Ta ugotovitev je v nalogi prikazana tudi s primerjavo krivulj preživetja operativno zdravljenih bolnic, diagnosticiranih v prvem stadiju raka na dojki, ki pokaže, da se porazdelitev modelskih ocen krivulj preživetja, kot tudi porazdelitev krivulj preživetja, ocenjenih na dejanskih podatkih, približa enakomerni porazdelitvi oziroma porazdelitvi individualne mere relativnega preživetja na splošni populaciji. Do podobnih zaključkov pridemo tudi s Coxovim modelom sorazmernih ogroženosti (kjer na omenjenem intervalu učinka starosti in leta diagnoze postaneta statistično neznačilna) in z Buckley-Jamesovim modelom. Z zadnjim smo to ugotovitev uspeli dokazati tako, da smo podatke modelirali pod in nad arbitrarно izbrano točko preloma⁵. Rezultati modela, ocenjeni na podatkih nad točko preloma, so pokazali

³Krnjenje in s tem tudi preživetje bolnic je v primeru raka na dojki zelo visoko (stopnja krnjenja je okoli 50 %).

⁴Starost in leto diagnoze sta spremenljivki, na podlagi katerih izračunamo individualno mero relativnega preživetja in na katerih se odrazi presežna ogroženost za smrt, ki je lastno proučevani bolezni.

⁵za točko preloma smo vzeli vrednost $Y=0,15$.

na statistično neznačilen vpliv starosti in leta diagnoze na relativno preživetje. Poleg tega smo z ocenjenimi pričakovanimi vrednostmi individualne mere relativnega preživetja prišli do ugotovitve, da je ocenjena pričakovanata vrednost alternativne mere relativnega preživetja operativno zdravljenih bolnic, diagnosticiranih v prvem stadiju bolezni, enaka $\gamma \approx 0,5$. To pomeni, da te bolnice lahko pričakujejo, da bodo v povprečju preživele 50 % vrstnic v splošni populaciji, kar lahko torej vzamemo za tisto vrednost, ki jo označimo kot točko ozdravitve.

Ugotovimo torej lahko, da z modeliranjem individualne mere relativnega preživetja s tremi proučevanimi modeli pridemo do konsistentnih informacij o učinku neodvisnih spremenljivk na relativno preživetje. Na podlagi tega lahko sklepamo o veljavnosti rezultatov, pri čemer vsak od modelov ponuja svojevrsten vpogled v relativno preživetje. Namen prikazanih analiz ni preverjanje kliničnih predpostavk proučevanih bolezni, temveč predvsem metodološki prikaz delovanja modelov. Razlog za to je tudi v tem, da registrski podatki o raku ne ponujajo informacij, ki bi zadostovale namenom kliničnih raziskav. Vendar na podlagi rezultatov lahko zaključimo, da z modeliranjem individualne mere relativnega preživetja, s katero iz preživetja izključimo vsa populacijska tveganja, pridemo do novih informacij o relativnem preživetju. Zaradi tega je uporaba novega pristopa primerna tudi za namene kliničnih raziskav.

Pri izbiri ustreznega modela se izkaže, da je uporaba Buckley-Jamesove regresije omejena. Čeprav bi bila v empiričnih analizah zaradi praktične uporabe najbolj zanimiva, ker ponuja možnost ocenjevanja pričakovanih časov relativnega preživetja in pripadajočih dejanskih časov preživetja samo na podlagi ocenjenih parametrov modela, pa z njo ne moremo zaznavati časovno odvisnih učinkov neodvisnih spremenljivk. V praksi namreč le redko naletimo na primere, v katerih bi bili učinki neodvisnih spremenljivk na ogroženost za smrt konstantni. Empiričnim raziskovalcem tako priporočamo, da pri analizi podatkov najprej uporabijo Aalenovo linearno regresijo, s katero lahko hitro ugotovimo, ali in pri katerih spremenljivkah prihaja do sprememb učinkov v času. Pomanjkljivost Aalenovega modela je, da ocenjene učinke lahko prikažemo le grafično, kvantitativno pa jih ne moremo ovrednotiti. Informacije, ki jih dobimo na podlagi rezultatov Aalenovega modela, zato lahko uporabimo kot okvirne vhodne informacije za modeliranje podatkov s Coxovim modelom, v katerem podatke priredimo za

namene analize časovno odvisnih učinkov neodvisnih spremenljivk. Coxov model je v empiričnih analizah zanimiv predvsem zato, ker njegove rezultate kvantitativno ovrednotimo in jih interpretiramo v kontekstu relativnih tveganj.

8. Literatura

- Aalen, O. 1980. A model for nonparametric regression analysis of counting processes. *Lecture Notes in Statistics* 2:1–25.
- . 1989. A linear regression model for the analysis of life times. *Statistics in Medicine* 8:907–925.
- . 1993. Further results on the non-parametric linear regression model in survival analysis. *Statistics in Medicine* 12:1569–1588.
- Aalen, O., J. Fosen, H. Weedon-Fekjaer, Ø. Borgan, in E. Husebye. 2004. Dynamic analysis of multivariate failure time data. *Biometrics* 60:764–773.
- Aalen, O. O., Ø. Borgan, in H.K. Gjessing. 2008. *Survival and event history analysis*. New York: Springer-Verlag.
- Andersen, P. K., K. Borch-Johnsen, T. Deckert, A. Green, P. Hougaard, N. Keiding, in S. Kreiner. 1985. A Cox regression model for relative mortality and its application to diabetes mellitus survival data. *Biometrics* 41:921–932.
- Andersen, P. K., O. Borgan, R. D. Gill, in N. Keiding. 1993. *Statistical models based on counting processes*. New York: Springer-Verlag.
- Andersen, P. K., L. S. Hansen, in N. Keiding. 1991. Non- and semi-parametric estimation of transition probabilities from censored observation of a non-homogeneous Markov process. *Scandinavian Journal of Statistics* 18:153–167.
- Andersen, P. K., in N. Keiding. 2002. Multi-state models for event history analysis. *Statistical Methods in Medical Research* 11:91–115.

- Andersen, P. K., in J. P. Klein. 2007. Regression analysis for multistate models based on a pseudo-value approach, with applications to bone marrow transplantation studies. *Scandinavian Journal of Statistics* 34:3–16.
- Andersen, P. K., J. P. Klein, in S. Rosthøj. 2003. Generalized linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika* 90:15–27.
- Andersen, P. K., in M. Vaeth. 1989. Simple parametric and nonparametric models for excess and relative mortality. *Biometrics* 45:523–535.
- Andersen, P.K., in R.D. Gill. 1982. Cox's regression model for counting processes: A large sample study. *Annals of Statistics* 10:1100–1120.
- Arjas, E. 1988. A graphical method for assesing goodnes of fit in cox's proportional hazards model. *Journal of the American Statistical Association* 83:204–212.
- Beran, R. 1981. Nonparametric regression with randomly censored survival data. Tech. Rep., University of California, Berkeley.
- Besag, Julian, Peter Green, David Higdon, in Kerrie Mengersen. 1995. Bayesian computation and stochastic systems. *Statist. Sci.* 10(1):3–66. With comments and a reply by the authors.
- Bhattacharyya, M., in J. P. Klein. 2005. A note on testing in aalen's additive hazards regression models. *Statistics in Medicine* 24:2235–2240.
- Blossfeld, H. P., K. Golsch, in G. Rohwer. 2007. *Event history analysis with stata*. New Jersey: Lawrence Erlbaum.
- Box-Steffensmeier, J. M., in B. S. Jones. 2004. *Event history modeling*. New York: Cambridge University Press.
- Breslow, N. E., J. H. Lubin, P. Marek, in B. Langholz. 1983. Multiplicative models and cohort analysis. *Journal of the American Statistical Association* 78:1–12.
- Buckley, J. D. 1984. Additive and multiplicative models for relative survival rates. *Biometrics* 40:51–62.

Buckley, J. J., in I. R. James. 1979. Linear regression with censored data. *Biometrics* 66: 661–664.

Cheng, S. C., J. P. Fine, in L. J. Wei. 1998. Prediction of cumulative incidence function under the proportional hazards model. *Biometrics* 54:219–228.

Cheuvart, B., in L. Ryan. 1991. Adjusting for age-related competing mortality in long-term cancer clinical trials. *Statistics in Medicine* 10:65–77.

Cleveland, W. S., E. Grosse, in W. M. Shyu. 1992. *Statistical models* in S. Chapman and Hall Computer Science series.

Commenges, D. 1999. Multi-state models in epidemiology. *Lifetime Data Analysis* 5: 315–327.

Cook, R. J., in J. F. Lawless. 2007. *The statistical analysis of recurrent events*. New York: Springer-Verlag.

Cox, D. R. 1972. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society — Series B* 34:187–220.

Csörgő, C., in J. J. Faraway. 1996. The exact and asymptotic distributions of Cramér - Von Mises statistics. *Journal of the Royal Statistical Society — Series B* 58:221–234.

Currie, I. E. 1996. A note on buckley-james estimators for censored data. *Biometrika* 83: 912–916.

Datta, S., in G. A. Satten. 2001. Validity of the Aalen-Johansen estimators of stage occupation probabilities and Nelson-Aalen estimators of integrated transition hazards for non-Markov models. *Statistics and Probability Letters* 55:403–411.

Dempster, A. P., N. M. Laird, in D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society* 39:1–38.

Dickman, P. W., A. Sloggett, M. Hills, in T. Hakulinen. 2004. Regression models for relative survival. *Statistics in Medicine* 23:51–64.

- Ederer, F., L. M. Axtell, in S. J. Cutler. 1961. *The relative survival rate: a statistical methodology*, vol. 6, 101–121. National Cancer Institute Monograph.
- Ederer, F., in H. Heise. 1959. Instructions to ibm 650 programmers in processing survival computations, methodological note no. 10, end results evaluation section. Tech. Rep., National Cancer Institute, Bethesda MD.
- Estève, J., E. Benhamou, M. Croasdale, in M. Raymond. 1990. Relative survival and the estimation of net survival: elements for further discussion. *Statistics in Medicine* 9:529–538.
- Fine, J. P. 2001. Regression modeling of competing crude failure probabilities. *Biostatistics* 2:85–97.
- Fine, J. P., in R. J. Gray. 1999. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association* 94:496–509.
- Fleming, T. R., in D. P. Harrington. 1991. *Counting processes and survival analysis*. John Wiley and Sons, Inc.
- Fosen, J., Ø. Borgan, H. Weedon-Fekjaer, in O. Aalen. 2006. Dynamic analysis of recurrent event data using the additive hazard model. *Biometrical Journal* 48:381–398.
- Frydman, H. 1992. A non-parametric estimation procedure for a periodically observed three-state Markov process, with application to AIDS. *Journal of the Royal Statistical Society — Series B* 54:853–866.
- . 1995. Nonparametric estimation of a Markov ‘illness-death’ process from interval- censored observations, with application to diabetes survival data. *Biometrika* 82:773–789.
- Gail, M. H., S. Wieand, in S. Piantadosi. 1984. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika* 71:431–444.
- Gandy, A., in U. W. E. Jensen. 2005. On goodnes-of-fit tests for aalen’s additive risk model. *Scandinavian Journal of Statistics* 32:425–445.

- Giorgi, R., M. Abrahamowicz, C. Quantin, P. Bolard, J. Estève, J. Gouvernet, in J. Faivre. 2003a. A relative survival regression model using b-spline functions to model non-proportional hazards. *Statistics in Medicine* 22:2767–84.
- Giorgi, R., G. Hédelin, in P. Schaffer. 2003b. Relative survival: comparison of regressive models and advice for the user. *Statistics in Medicine* 22:2767–84.
- Giorgi, R., J. Payan, in J. Gouvernet. 2005. RSURV: A function to perform relative survival analysis with S-PLUS or R 78:175–178.
- Grambsch, P. M., in T. M. Therneau. 1994. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 81:515–526.
- Grønnesby, J. K. 1997. Testing covariate effects in aalen's linear hazard model. *Scandinavian Journal of Statistics* 24:125–135.
- Grønnesby, J. K., in Ø. Borgan. 1996. A method for checking regression models in survival analysis based on thi risk score. *Lifetime Data Analysis* 2:315–328.
- Guilbaud, O. 1988. Exact Kolmogorov-Smirnov tests for left-truncated and/or right censored data. *Journal of the American Statistical Association* 83:213–221.
- Hakulinen, T. 1982. Cancer survival corrected for heterogeneity in patient withdrawal. *Biometrics* 38:933–942.
- Hakulinen, T., in K. H. Abeywickrama. 1985. A computer program package for relative survival analysis. *Computer Methods and Programs in Biomedicine* 19 (2-3):197–207.
- Hakulinen, T., in L. Tenkanen. 1987. Regression analysis of relative survival rates. *Journal of the Royal Statistical Society — Series C* 36:309–317.
- Halekoh, U., S. Højsgaard, in J. Yan. 2006. The R package geepack for generalized estimating equations. *Journal of Statistical Software* 15:1–11.
- Harrell, F. E. 2001. *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. Springer-Verlag.
- Heller, G., in J. S. Simonoff. 1990. A comparsion of estimators for regression with a censored response variable. *Biometrics* 77:515–520.

- . 1992. Prediction in censored survival data: a comparsion of the proportional hazards and linear regression models. *Biometrics* 48:101–115.
- Hillis, S. L. 1993. A comparison of three buckley-james variance estimators. *Communications in Statistics, B* 22:955–73.
- Hougaard, P. 1999. Multi-state models: A review. *Lifetime Data Analysis* 5:239–264.
- Huffer, F. W., in I. W. McKeague. 1991a. Weighted least squares estimation for aalen's additive risk model. *Journal of the American Statistical Association* 86:114–129.
- Huffer, F.W., in I.W. McKeague. 1991b. Weighted least squares estimation for aalen's additive risk model. *Journal of the American Statistical Association* 86:114–129.
- James, I. R., in P. J. Smith. 1984. Consistency results for linear regression with censored data. *Annals of Statistics* 12:590–600.
- Jin, Z., D. Y. Lin, in Z. Ying. 2006. On least-squares regression with censored data. *Biometrika* 93:147–161.
- Kay, R. A. 1986. Markov model for analysing cancer markers and disease states in survival models. *Biometrics* 42:855–865.
- Klein, J. P., in P. K. Andersen. 2005. Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics* 61:223–229.
- Klein, J. P., in Y. Shu. 2002. Multi-state models for bone marrow transplantation studies. *Statistical Methods in Medical Research* 11:117–139.
- Koul, H., V. Susarla, in J. Van Ryzin. 1981. Regression analysis with randomly right-censored data. *Annals of Statistics* 9:1276–1288.
- Lai, T. L., in Z. Ying. 1991. Large sample theory of a modified buckley-james estimator for regression analysis with censored data. *Annals of Statistics* 19:1370–1402.
- Lang, S, in A Brezger. 2002. BayesX: software for Bayesian inference based on Markov chain Monte Carlo simulation techniques. University of Munich.

- Le Teuff, F., M. Abrahamowicz, P. Bolard, in C. Quantin. 2006. Comparison of Cox's and relative survival models when estimating the effects of prognostic factors on disease-specific mortality: A simulation study under proportional excess hazards. *Statistics in Medicine* 24:3887–3909.
- Lee, E., in L. A. Weissfeld. 1998. Assesment of covariate effects in aalen's additive hazard model. *Statistics in Medicine* 17:983–998.
- Lin, D. Y., L. J. Wei, I. Yang, in Z. L. Ying. 2000. Robust inferences for counting processes under Andersen-Gill model. *Journal of the Royal Statistical Society — Series B* 62:711–730.
- Lin, D. Y., in Z. Ying. 1994a. Semiparametric analysis of the additive risk model. *Biometrika* 81:61–71.
- Lin, D. Y., in Z. L. Ying. 1994b. Semiparametric analysis of the additive risk model. *Biometrika* 81:61–71.
- Lin, H., in J. Zhenzhen. 2007. Lss: An s-plus/r program for the accelerated failure time model to right censored data based on least-squares principle. *Computer Methods and Programs in Biomedicine* 86:45–50.
- Lombard, M., B. Portmann, J. Neuberger, R. Williams, N. Tygstrup, L. Ranek, H. Ring-Larsen, J. Rodés, M. Navasa, C. Trepo, G. Pape, G. Schou, J. H. Badsberg, in P. K. Andersen. 1993. Cyclosporin A treatment in primary biliary cirrhosis: results of a long-term placebo controlled trial. *Gastroenterology* 104:519–526.
- Louis, T. A. 1982. Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society — Series B* 44:226–233.
- Martinussen, T., in T. H. Scheike. 2006. *Dynamic regression models for survival data*. New York: Springer-Verlag.
- Mathsoft. 1999. *S-plus 2000 guide to statistics* 2. Data Analysis Products Division of MathSoft, Seattle, WA.
- McKeague, I. W. 1988. Asymptotic theory for weighted least squares estimators in aalen's additive risk model. *Contemp. Math.* 80:139–152.

- McKeague, I. W., in P. D. Sasieni. 1994. A partly parametric additive risk model. *Biometrika* 81:501–514.
- McKeague, I. W., in K. J. Utikal. 1991a. Goodness-of-fit-tests for additive hazards and proportional hazards model. *Scandinavian Journal of Statistics* 18:177–195.
- . 1991b. Goodness-of-fit tests for additive hazards and proportional hazards models. *Scandinavian Journal of Statistics* 18:177–195.
- Meira-Machado, L., J. Úna-Álvarez, in C. Cadarso-Suárez. 2006. Nonparametric estimation of transition probabilities in a non-Markov illnessdeath model. *Lifetime Data Analysis* 12:325–344.
- Miller, R.G. 1976. Least squares regression with censored data. *Biometrics* 63:449–464.
- Miller, R.G., in J. Halpern. 1982. Regression with censored data. *Biometrics* 69:521–531.
- Müller, H. G., in J. L. Wang. 1994. Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics* 50:61–76.
- O’Quigley, J. 2003. Khmaladze-type graphical evaluation of the proportional hazards assumption. *Biometrika* 90:577–584.
- O’Quigley, J., in J. Stare. 2003. Cumulative empirical processes for survival models. In *Proceedings of the 25th international conference on information technology interfaces*, ed. L. Budin, L. Luar-Stiffler, in Z. Bekic, 205–210. SRCE, University of Zagreb.
- Pepe, M. S. 1991. Inference for events with dependent risks in multiple endpoint studies. *Journal of the American Statistical Association* 86:770–778.
- Pohar, M., in J. Stare. 2006a. EM algorithm based estimation in relative survival regression. In *Proceedings of the 27th isb conference*, 174–179. International society for clinical biostatistics.
- . 2006b. Relative survival analysis in R. *Computer Methods and Programs in Biomedicine* 81:272–278.
- Pohar Perme, M. 2007. *Goodnes of fit of relative survival models*. Ph.D. thesis, Medicinska fakulteta, Univerza v Ljubljani.

Pohar Perme, M., R Henderson, in J. Stare. 2009. An approach to estimation in relative survival regression. *Biostatistics* 10:136–146.

Putter, H., M. Fiocco, in R. B. Geskus. 2007. Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine* 26:2389–2430.

R Development Core Team. 2005. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramlau-Hansen, H. 1983. Smoothing counting process intensities by means of kernel functions. *Annals of Statistics* 11:453–466.

Resnick, S. I. 1992. *Adventures in stochastic processes*. Boston: Birkhäuser.

Sasieni, P. D. 1993. Maximum weighted partial likelihood estimators for the Cox model. *Journal of the American Statistical Association* 88(421):144–152.

———. 1996. Proportional excess hazards. *Biometrika* 83:127–141.

Sasieni, P. D., in A. Winnett. 2003. Martingale difference residuals as a diagnostic tool for the Cox model. *Biometrika* 90:899–912.

Scheike, T. H., in M. J. Zhang. 2003. Extensions and applications of the Cox-Aalen survival model. *Biometrics* 59:1036–1045.

———. 2007. Direct modelling of regression effects for transition probabilities in multistate models. *Scandinavian Journal of Statistics* 34:17–32.

Schoenfeld, D. 1982. Partial residuals for the proportional hazards regression model. *Biometrika* 69:239–241.

Shen, Y., in S. C. Cheng. 1999. Confidence bands for cumulative incidence curves under the additive risk model. *Biometrics* 55:1093–1100.

Shu, Y., in J. P. Klein. 2005. Additive hazards Markov regression models illustrated with bone marrow transplant data. *Biometrika* 92:283–301.

Shu, Y., J. P. Klein, in M. J. Zhang. 2007. Asymptotic theory for the Cox semi-Markov illness-death model. *Lifetime Data Analysis* 13:91–117.

- Silverman, B. W. 1986. *Density estimation for statistics and data analysis*. London: Chapman and Hall.
- Smith, P. J. 2002. *Analysis of failure and survival data*. Boca Raton: Chapman and Hall/CRC.
- Smith, P. J., in J. Zhang. 1995. Renovated scatterplots for censored data. *Biometrika* 82: 447–452.
- Stare, J., H. Heinzl, in F. Harrel. 2000. On the use of Buckley and James least squares regression for survival data. *Metodološki zvezki, New Approaches in Advanced Statistics* 16:125–134.
- Stare, J., R. Henderson, in M. Pohar. 2005a. An individual measure of relative survival. *Journal of the Royal Statistical Society — Series C* 54:115–126.
- Stare, J., M. Pohar, in R. Henderson. 2005b. Goodness of fit of relative survival models. *Statistics in Medicine* 24.
- Stone, C. J., in C. Y. Koo. 1985. Additive splines in statistics. In *Proceedings of the statistical computing section asa*, 45–48.
- Struthers, C. A., in J. D. Kalbfleisch. 1986. Misspecified proportional hazard models. *Biometrika* 73:363–369.
- Stute, W., in J. L. Wang. 1994. The jackknife estimate of a Kaplan-Meier integral. *Biometrika* 81:603–606.
- Therneau, T., in J. Offord. 1999. Expected survival based on hazard rates (update). Tech. Rep. 63, Section of Biostatistics, Mayo Clinic.
- Therneau, T. M., in P. M. Grambsch. 2000. *Modelling survival data: Extending the Cox model*. New York: Springer-Verlag.
- Therneau, T. M., P. M. Grambsch, in T. R. Fleming. 1990. Martingale-based residuals for survival models. *Biometrika* 77:147–160.
- Weissfeld, L. A., in H. Schneider. 1987. Inferences based on the buckley-james procedure. *Communications in Statistics, A* 16:177–87.

9. Priloga

9.1 Števni procesi in martingali

9.1.1 Osnove števnih procesov

Krnjene podatke v analizi preživetja sestavljajo časi do nastopa želenega dogodka (označimo jih s T_i^*) in krnjeni časi (označimo jih s C_i^*), za katere velja, da želeni dogodek v času opazovanja enot (še) ni nastopil. Predpostavljam, da so časi T_i^* in C_i^* med sabo neodvisni. Krnjene podatke zato obravnavamo kot pare vrednosti spremenljivk (T_i, δ_i) za $i = 1, \dots, n$, kjer je T_i opazovani čas enote, $\delta_i = I(T_i^* \leq C_i^*)$ pa indikatorska spremenljivka, ki zavzame vrednost 1, če je želeni dogodek pri opazovani enoti nastopil, v obratnem primeru pa vrednost 0.

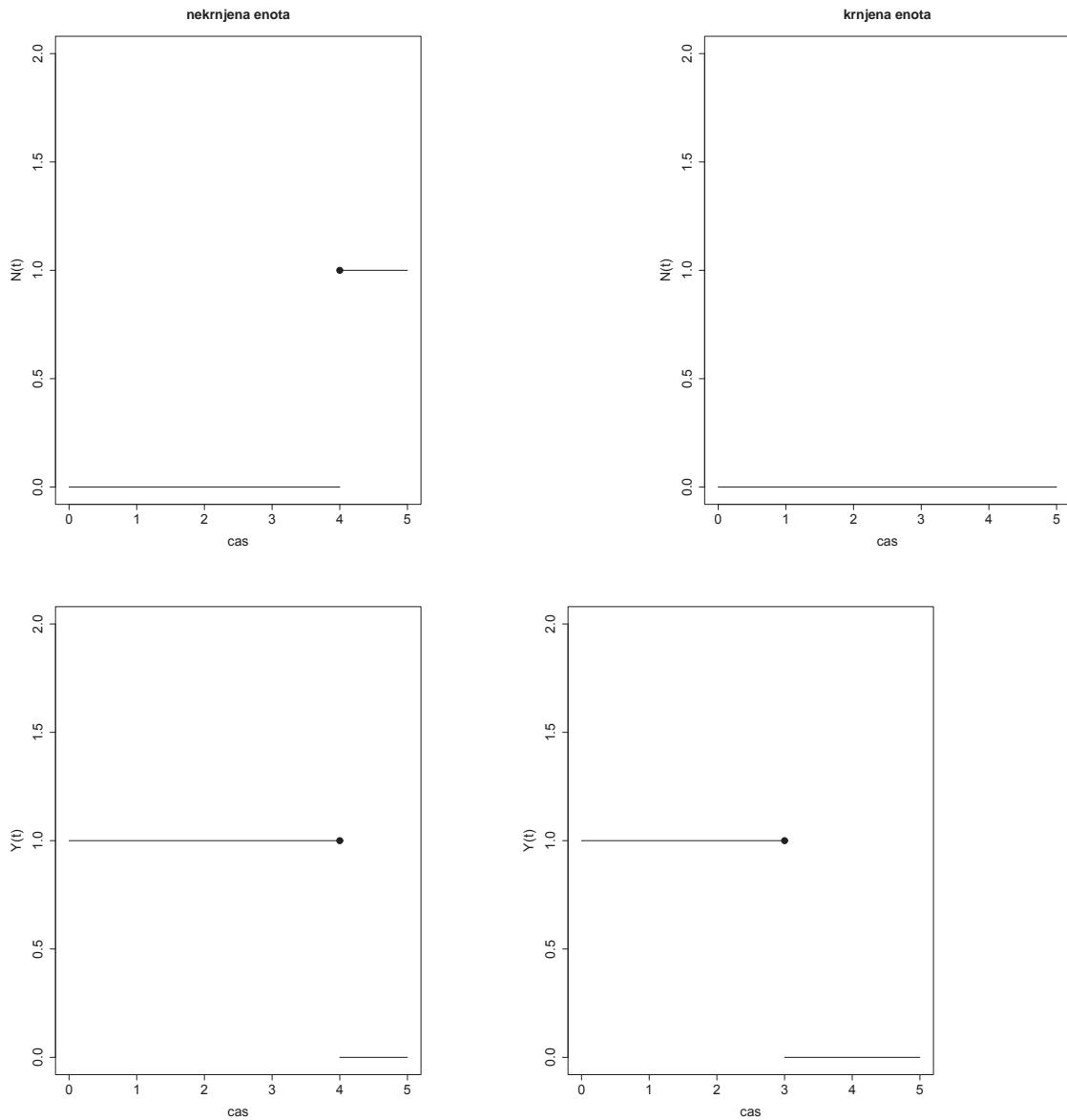
Pri števnih procesih pare vrednosti spremenljivk (T_i, δ_i) nadomestimo s pari funkcij $(N_i(t), Y_i(t))$, kjer je

$$\begin{aligned} N_i(t) &= \text{število opazovanih dogodkov na časovnem intervalu } [0, t] \text{ za } i\text{-to enoto} \\ Y_i(t) &= \begin{cases} 1, \text{ do nastopa dogodka oziroma krnjenja } i\text{-te enote} \\ 0, \text{ v obratnem primeru.} \end{cases} \end{aligned}$$

Števni proces za desno krnjene podatke tako lahko opišemo s funkcijama $N_i(t) = I(T_i \leq t, \delta_i = 1)$ in $Y_i(t) = I(T_i \geq t)$. $N_i(t)$ je stopničasta funkcija, ki ima vrednost 0, dokler se dogodek pri i -ti ne zgodi. Z nastopom dogodka, vrednost funkcije skoči na 1, pri krnjениh podatkih pa ostane enaka 0. Proses $N_i(t)$ je desno-zvezni proces. $Y_i(t)$ je indikatorski proces, ki ima vrednost 1, dokler je enota v tveganju. To pomeni, da je $Y_i(t)$ stopničasta funkcija, zvezna z leve, in je v času $t = 0$ enaka 1, z nastopom dogodka oziroma krnjenja pa njena vrednost pada na 0. Grafični prikaz obeh procesov

je za primer nekrnjene in krnjene enote prikazan na sliki 9.1.

Slika 9.1: Grafični prikaz procesov $N(t)$ in $Y(t)$ za nekrnjeno in krnjeno enoto



9.1.2 Modeliranje števnih procesov

V statistiki je vsak model temelji na podlagi določenih informacij. V primeru števnih procesov te informacije predstavljajo tako imenovano zgodovino opazovanja do časa t in vključno z njim. To zgodovino pogosto imenujemo filtracija (ang. *filtration*) in jo označimo z $\{\mathcal{F}_t; t \geq 0\}$. V splošnem \mathcal{F}_t predstavlja zgodovino procesov N_i, Y_i in procesov neodvisnih spremenljivk \mathbf{X}_i . Z uporabo oznak σ algebре filtracijo zapišemo v obliki

$$\mathcal{F}_t = \sigma(N_i(s), Y_i(s+), \mathbf{X}_i(s); i = 1, \dots, n; 0 \leq s \leq t).$$

Glede na \mathcal{F}_t sta procesa Y in \mathbf{X} napovedljiva, kar pomeni, da so z zgodovino, znano do časa t (kar označimo z \mathcal{F}_{t-}), znane tudi vrednosti $Y(t)$ in $(X)(t)$.

Stopnjo ogroženosti krnjениh podatkov ($\lim_{\Delta t} \frac{1}{\Delta t} P(t \leq T < t + \Delta t | T \geq t)$), ki jo na tem mestu označimo z $\alpha(t)$, pri števnih procesih nadomestimo s funkcijo (oziroma procesom) ogroženosti. Za števni proces $N_i(t)$ je funkcija ogroženosti enaka

$$\lambda_i(t) = Y_i(t)\alpha_i(t)$$

in je definirana kot

$$\lambda_i(t)dt = P(dN_i(t) = 1 | \mathcal{F}_{t-}) = E(dN_i(t) | \mathcal{F}_{t-}),$$

kjer $dN_i(t)$ predstavlja spremembo procesa $N_i(t)$ na intervalu $[t, t + dt]$. Zgornjo definicijo lahko utemeljimo z naslednjim razmišljanjem. V primeru odsotnosti neodvisnih spremenljivk velja, da je edini pomemben del zgodovine i -te enote v času t – zgodovina procesa $Y_i(t)$. Verjetnost $P(dN_i(t) = 1 | \mathcal{F}_{t-})$ tako lahko zapišemo kot $P(dN_i(t) = 1 | \mathcal{F}_{t-}) = P(dN_i(t) = 1 | Y_i(t))$. Če je $Y_i(t) = 0$, je dogodek pri enoti bodisi že nastopil, bodisi je bila enota krnjena, in je $dN_i(t) = 0$ z verjetnostjo 1. Če pa je $Y_i(t) = 1$, enota tvega, da dogodek pri njej nastopi in velja

$$P(dN_i(t) = 1 | Y_i(t) = 1) = P(t \leq T_i^* \leq t + dt | t \leq T_i^*, t \leq C_i^*).$$

Zaradi neodvisnosti časov T_i^* in C_i^* po definiciji stopnje ogroženosti velja

$$P(t \leq T_i^* \leq t + dt | t \leq T_i^*, t \leq C_i^*) = P(t \leq T_i^* \leq t + dt | t \leq T_i^*) = \alpha_i(t)dt.$$

Na podlagi teh dveh možnosti torej lahko zapišemo:

$$P(dN_i(t) = 1 | Y_i(t)) = Y_i(t)\alpha_i(t)dt.$$

9.1.3 Osnove martingalov

Vzemimo proces

$$M_i(t) = N_i(t) - \int_0^t \lambda_i(s)ds, \quad (9.1)$$

ki predstavlja martingal števnega procesa z zgodovino, opisano na koncu prejšnjega poglavja. Glede na filtracijo \mathcal{F}_t je martingal ($M(t)$) stohastični proces, za katerega ob določenih pogojih velja

$$E(dM(t)|\mathcal{F}_{t-}) = 0, \text{ za vsak } t > 0. \quad (9.2)$$

Podobno za katerikoli $0 \leq s \leq t$ velja

$$E(M(t)|\mathcal{F}_s) = M(s), \quad (9.3)$$

iz česar sledi

$$E(M(t)|M(u); 0 \leq u \leq s) = M(s). \quad (9.4)$$

To pomeni, da je odvisno od preteklosti najboljša napoved prihodnje vrednosti njena sedanja vrednost. Pomembni lastnosti martingalov sta tudi naslednji:

- Povprečje martingala je konstantno, ker je povprečje prirastkov martingala enako 0. Martingal števnega procesa je potem martingal s povprečjem 0, saj velja $M(0) = 0$.
- Prirastki martingalov so nekorelirani, vendar ne nujno tudi neodvisni. Za $t, u, s > 0$ velja

$$\begin{aligned} Cov[M(t), M(t+u) - M(t)] &= 0 \\ Cov[M(t) - M(t-s), M(t+u) - M(t)] &= 0. \end{aligned}$$

Po teoremu Doob Meyerjeve razčlenitve lahko vsak števni proces razčlenimo kot vsoto martingala in tako imenovanega kompenzatorja (označimo ga z $\Lambda(t)$), za katerega velja, da je napovedljiv desno zvezni proces in je $\Lambda(0) = 0$. Kompenzator števnega procesa v (9.1) je $\int_0^t \lambda(s)ds = \int_0^t Y_i(s)\alpha_i(s)ds$. Števni proces torej lahko zapišemo v obliki:

$$N_i(t) = M_i(t) + \int_0^t \lambda_i(s)ds.$$

Če označimo $\int_0^t \lambda_i(s)ds$ z $E_i(t)$, Doob Meyerjevo razčlenitev preprosto lahko zapišemo kot

$$N_i(t) = E_i(t) + M_i(t), \quad (9.5)$$

kar spominja na statistično razčlenitev: podatki=model+napaka. Kljub podobnosti pa moramo pri uporabi zgornjega zapisa števnega procesa upoštevati, da za razliko s statistično razčlenitvijo, kjer imajo napake ničelno avtokorelacijo in konstantno varianco, za martingal velja, da ima pozitivno avtokorelacijo $Cov[M(t+s), M(t)] = Var[M(t)]$ in naraščajočo varianco. Boljšo analogijo s statistično razčlenitvijo dosežemo z razčlenitvijo diferencialov števnih procesov

$$dN_i(t) = dE_i(t) + dM_i(t), \quad (9.6)$$

vendar je uporaba teh diferencialov v praksi mnogo težja.

Z zamenjavo enakosti $' ='$ v enačbah (9.2) in (9.3) dobimo proces tako imenovanega nepopolnega martingala (ang. *submartingale*), kar bi lahko primerjali s procesom kockanja, v katerem se število zmag s časom povečuje. Števni proces je torej strogo monoton nepopolni martingal. Ker je povprečje martingala enako 0, je njegova varianca enaka $E(M^2)$, ki je tudi nepopolni martingal z zgodovino \mathcal{F}_t enako zgodovini martingala M. Tudi M^2 lahko razčlenimo na vsoto kompenzatorja in martingala, pri čemer kompenzator imenujemo napovedljiv variacijski proces (ang. *predictable variation process*) in ga običajno zapišemo v obliki $\langle M, M \rangle(t)$ oziroma $\langle M \rangle(t)$. Predstavlja limito v verjetnosti izraza

$$\sum_i E[(M(t_{i+1}) - M(t_i))^2 | \mathcal{F}_{t_i}] = \sum_i Var[(M(t_{i+1}) - M(t_i)) | \mathcal{F}_{t_i}],$$

na intervalu $[0, t]$, $0 = t_0 < t_1 < \dots < t_n = t$. Z oznakami diferencialov zapišemo

$$d\langle M \rangle(t) = \text{var}(dM(t)|\mathcal{F}_{t-})$$

in za martingal števnega procesa

$$d\langle M \rangle(t) = \text{var}(dN(t) - \lambda(t)dt|\mathcal{F}_{t-}).$$

Ob dani filtraciji \mathcal{F}_{t-} je $dN(t)$ Poissonova slučajna spremenljivka s povprečjem in varianco enako $\lambda(t)dt$, kar pomeni, da je napovedljiv variacijski proces martingala števnega procesa enak kompenzatorju števnega procesa:

$$\langle M \rangle(t) = \int_0^t \lambda(s)ds = E(t).$$

V teoriji martingalov poleg napovedljivega variacijskega procesa obstaja tudi tako imenovan opcijski variacijski (optional variation) proces $[M]$, ki je prvemu soroden, vendar enostavnejši. Definiran je kot limita v verjetnosti za $\sum [M(t_{i+1}) - M(t)]^2$ na intervalu $[0, t]$, $0 = t_0 < t_1 < \dots < t_n = t$. Je desno zvezni proces za katerega velja, da je $M^2 - [M]$ martingal, kar pomeni, da je $[M]$ nepristranska cenilka variance martingala s povprečjem 0. V primeru, ko je M zvezen, velja $[M] = \langle M \rangle$, v primeru ko M ni zvezen pa je

$$[M] = \sum_{s \leq t} \Delta M(s)^2,$$

kjer je $\Delta M(s) = M(s) - M(s-)$. To pomeni, da je za martingal števnega procesa opcijski variacijski proces enak $[M] = N$, saj se spremeni samo ob spremembi števnega procesa, ta sprememba pa je po definiciji enaka 1. V tem primeru $[M]$ predstavlja nepopolni martingal, $\langle M \rangle$ pa je njegov kompenzator. Pravzaprav $[M]$ predstavlja opazovano, $\langle M \rangle$ pa pričakovano informacijo. V primeru martingala s povprečjem 0 velja

$$\text{Var}(M(t)) = E(\langle M \rangle(t)) = E([M](t)) = E(M^2(t)).$$

Za n martingalov z enako filtracijo $(\mathcal{F}_t, t \geq 0)$ je tudi $\bar{M}(t) = \sum_i M_i(t)$ martingal z $(\mathcal{F}_t, t \geq 0)$. Proces variance \bar{M} -ja pa je odvisen od kovariacije med martingali $M_i(t)$. Ker produkt $M_i M_j$ lahko zapišemo kot razliko dveh nepopolnih martingalov

$$M_i M_j = 1/4[(M_i + M_j)^2 - (M_i - M_j)^2],$$

je kompenzator produkta $M_i M_j$ napovedljiv variacijski proces $\langle M_i, M_j \rangle = \text{plim} \sum \text{cov}[\{(M_i(t_{i+1}) - M_i(t_i))\}]$. V primeru, ko za vsak t velja $\langle M_i, M_j \rangle = 0$, sta M_i in M_j tako imenovana ortogonalna martingala. To pomeni, da je tudi produkt $M_i M_j$ martingal. Za martingale števnih procesov velja $E(\langle M_i, M_j \rangle) = \text{cov}(M_i, M_j)$ zato so ortogonalni martingali števni procesov nekolerirani. Opcijski kovariacijski proces je enak $[M_i, M_j] = \text{plim} \sum \{[M_i(t_{i+1}) - M_i(t_i)][M_j(t_{i+1}) - M_j(t_i)]\}$. V primeru zveznih martingalov velja $[M_i, M_j] = \langle M_i, M_j \rangle$, sicer pa je

$$[M_i, M_j](t) = \sum_{s \leq t} \Delta M_i(s) \Delta M_j(s).$$

$\langle M_i, M_j \rangle$ je kompenzator procesa $[M_i, M_j]$. Tako pridemo do pomembnega rezultata, ki pravi, da sta martingala števnih procesov ortogonalna, če imata enako filtracijo in njune spremembe ne nastopajo sočasno. To pomeni, da je $[M_i, M_j] = 0$, s čimer je tudi $\langle M_i, M_j \rangle = 0$. To pomeni, da imajo lahko tudi močno odvisni števni procesi nekorelirane martingale.

Napovedljiv in opcijijski variacijski proces za \bar{M} je tako odvisen od napovedljivih in opcijskih kovariacijskih procesov posameznih martingalov:

$$\begin{aligned} \langle \bar{M} \rangle(t) &= \sum_i \sum_j \int_0^t d \langle M_i, M_j \rangle(s) \\ [\bar{M}](t) &= \sum_i \sum_j \int_0^t d [M_i, M_j](s), \end{aligned}$$

oziroma v primeru ortogonalnih martingalov števnih procesov

$$\begin{aligned} \langle \bar{M} \rangle(t) &= \sum_i \langle M_i \rangle(t) \\ [\bar{M}](t) &= \sum_i [M_i](t). \end{aligned}$$

Zgornji formuli v obliki martingalov predstavljata v statistiki znano dejstvo, da je varianca vsote slučajnih spremenljivk enaka vsoti kovarianc vseh parov slučajnih spremenljivk oziroma, da je varianca vsote neodvisnih slučajnih spremenljivk enaka vsoti njihovih varianc.

Vzemimo, da je glede na filtracijo \mathcal{F}_t , M martingal s povprečjem 0 in H nek napovedljiv proces z enako filtracijo. Potem je transformacija $U(t) = \int_0^t H(s)dM(s)$ ravno tako martingal s povprečjem 0. Na podlagi (9.2) je mogoče pokazati

$$\begin{aligned} E(dU(t)|\mathcal{F}_{t-}) &= E(H(t)dM(t)|\mathcal{F}_{t-}) \\ &= H(t)E(dM(t)|\mathcal{F}_{t-}) \\ &= H(t) = 0. \end{aligned}$$

Zaradi napovedljivosti je H del σ algebре, definirane z $\mathcal{F}_{\sqcup -}$. Napovedljivi in opcijski variacijski proces za U dobimo s preprostimi transformacijami obeh procesov za M :

$$\begin{aligned} \langle U \rangle(t) &= \int_0^t H^2(s)d\langle M \rangle(s) \\ [U](t) &= \int_0^t H^2(s)d[M](s). \end{aligned}$$

V primeru števnih procesov je martingal enak $M(t) = N(t) - \int_0^t \lambda(s)ds$, zgornji enačbi pa torej zapišemo kot:

$$\begin{aligned} \langle U \rangle(t) &= \int_0^t H^2(s)\lambda(s)ds \\ [U](t) &= \int_0^t H^2(s)dN(s). \end{aligned}$$

Za n martingalov $M_i, i = 1, \dots, n$ z enako filtracijo $\mathcal{F}_t, t \geq 0$ in n napovedljivih procesov H_i s filtracijo $\mathcal{F}_t, t \geq 0$ je tudi proces $\sum_i \int H_i dM_i$ martingal z enako filtracijo. V primeru krnjnih podatkov lahko skoraj vse cenilke in testne statistike zapišemo v obliki napovedljivega procesa H glede na martingal, ki ga dobimo na podlagi opazovanih podatkov, pri katerih se je dogodek zgodil $dN(t)$, in kompenzatorja, ki ga dobimo na osnovi modeliranja stopnje ogroženosti.

V praksi se najpogosteje srečujemo z multivariatnim števnim procesom $\mathbf{N}(t) = (N_1(t), \dots, N_k(t))'$ za katerega predpostavljamo, da spremembu katerihkoli dveh procesov ne more nastopiti sočasno ($P[\Delta N_i(t) = \Delta N_j(t) = 1] = 0$). Nekatere pomembnejše zveze martingalov števnih procesov v tem delu prikazujemo še v matrični obliki.

Proces ali funkcijo ogroženosti multivariatnega števnega procesa lahko zapišemo kot vektor funkcij ogroženosti posameznih števnih procesov:

$$\lambda(t) = (\lambda_1(t), \dots, \lambda_k(t))^T.$$

Naj bo $\int_0^t \lambda(u)du$ k razsežni vektor z elementi $\int_0^t \lambda_j(u)du; j = 1, \dots, k$. Martingal multivariatnega števnega procesa v matrični obliki je tedaj enak

$$\mathbf{M}(t) = \mathbf{N}(t) - \int_0^t \lambda(u)du.$$

Napovedljiv ($\langle \mathbf{M} \rangle(t)$) in opcijski variacijski proces ($[\mathbf{M}](t)$) martingala v zgornji obliki je $k \times k$ razsežna matrika, v kateri je (h, j) -ti element enak $\langle \mathbf{M}_h, \mathbf{M}_j \rangle(t)$ oziroma $[\mathbf{M}_h, \mathbf{M}_j]$. Zaradi ortogonalnosti martingalov števnih procesov in $[M](t) = N(t)$ lahko oba variacijska procesa multivariatnega števnega procesa enostavneje zapišemo kot

$$\begin{aligned}\langle \mathbf{M} \rangle(t) &= \text{diag} \int_0^t \lambda(u)du \\ [\mathbf{M}](t) &= \text{diag} \mathbf{N}(t),\end{aligned}$$

kjer je za k razsežni vektor \mathbf{v} $\text{diag} \mathbf{v}$ $k \times k$ razsežna diagonalna matrika z elementi vektorja \mathbf{v} na diagonali.

Zaradi predostavke multivariatnega števnega procesa $P[\Delta N_i(t) = \Delta N_j(t) = 1] = 0$ so tudi spremembe agregiranega števnega procesa $N_{\bullet}(t) = \sum_{j=1}^k N_j(t)$ enake 1 ($\Delta N_{\bullet}(t) = 1$). Funkcija ogroženosti za tako definiran števni proces je potem enaka $\lambda_{\bullet}(t) = \sum_{j=1}^k \lambda_j(t)$, pripadajoči martingal pa $M_{\bullet}(t) = \sum_{j=1}^k M_j(t)$.

Vzemimo $p \times k$ razsežno matriko $\mathbf{H}(t)$, pri kateri (i, j) -te elemente predstavljajo napovedljivi procesi $H_{ij}(t); i = 1, \dots, p, j = 1, \dots, k$. $\int_0^t \mathbf{H}(u)d\mathbf{M}(u)$ je torej p razsežni vektor transformacij, v katerem je i -ti element enak vsoti posameznih transformacij $\sum_{j=1}^k \int_0^t H_{ij}(u)dM_j(u)$. Velja

$$\begin{aligned}\langle \int \mathbf{H} d\mathbf{M} \rangle(t) &= \int_0^t \mathbf{H}(u) \text{diag}[\lambda(u)du] \mathbf{H}(u)^T \\ [\int \mathbf{H} d\mathbf{M}](t) &= \int_0^t \mathbf{H}(u) \text{diag}[d\mathbf{N}(u)] \mathbf{H}(u)^T,\end{aligned}$$

zapisano v obliki matričnih komponent

$$\begin{aligned}\left\langle \sum_{j=1}^k \int_0^t H_{ij}(u)dM_j(u), \sum_{j=1}^k \int_0^t H_{lj}(u)dM_j(u) \right\rangle(t) &= \sum_{j=1}^k \int_0^t H_{ij}(u)H_{lj}(u)\lambda_j(u)du \\ \left[\sum_{j=1}^k \int_0^t H_{ij}(u)dM_j(u), \sum_{j=1}^k \int_0^t H_{lj}(u)dM_j(u) \right](t) &= \sum_{j=1}^k \int_0^t H_{ij}(u)H_{lj}(u)dN_j(u).\end{aligned}$$