

УНИВЕРЗИТЕТ У БЕОГРАДУ
МАТЕМАТИЧКИ ФАКУЛТЕТ



Стефан Малбашић

УПРАВЉАЊЕ КРЕДИТНИМ РИЗИКОМ У
БАНКАРСКОМ СЕКТОРУ РЕПУБЛИКЕ
СРБИЈЕ ПРИМЈЕНОМ МЕТОДА
СТАТИСТИЧКОГ УЧЕЊА

мастер рад

Београд, 2023.

Ментор:

др Бојана МИЛОШЕВИЋ, ванредни професор
Универзитет у Београду, Математички факултет

Чланови комисије:

др Милош БОЖОВИЋ, редовни професор
Универзитет у Београду, Економски факултет

др Милан ЈОВАНОВИЋ, доцент
Универзитет у Београду, Математички факултет

Датум одбране: _____

ЗАХВАЛНИЦА

Прије свега, неизмјерну захвалношћу дућујем менторки проф. др Бојани Милошевић, ванредном професору Математичког факултета, на уложеном времену и труду да своје математичко знање и искуство подијели са мном. Захваљујем јој се на подршци, помоћи и сирљењу приликом писања овог рада.

Захваљујем се и члановима Комисије за преглед и оцјену мастер рада проф. др Милошу Божовићу и др Милану Јовановићу на пажљивом читању рукописа, као и корисним коментарима и сугестијама који су допринијели квалитету овог рада.

Посебну захвалношћу дућујем Одбору за управљање ризицима и капиталом Удружења банака Србије на усћућању временске серије аритраних података о стопама неизмирења обавеза на нивоу банкарског сектора Републике Србије која је послужила као основ за специфичне циљеве овог рада.

Највећу захвалношћу дућујем родитељима који су кроз бескрајну љубав и вјеру у мене увијек били моја највећа подршка и охрабрење да истрајем на свом путу.

Београд, септембар 2023.

Стефан Малбашић

Наслов мастер рада: Управљање кредитним ризиком у банкарском сектору Републике Србије примјеном метода статистичког учења

Апстракт: Банке се у свом пословању неминовно сусрећу с различитим врстама ризика, из којих могу проистећи негативни ефекти на само пословање. Нарочито могу бити изложене кредитном ризику који представља могућност настанка негативних ефеката на финансијски резултат и капитал банке усљед неизвршавања обавеза дужника према банци. Управљање ризиком добија посебну важност у савременим условима сталних промјена у банкарском окружењу и у организацији пословања који доносе веома изражену ризичност. Предмет овог рада је управљање кредитним ризиком у банкарском сектору Републике Србије и квантификовање степена изложености ризику примјеном метода статистичког учења. Интерни модели које банке користе најчешће дају оцјене вјероватноће неизвршења обавеза које обухватају читав пословни циклус. Међутим, за потребе имплементације МСФИ 9 неопходне су оцјене вјероватноће за конкретан временски тренутак. У раду су приказани модели статистичког учења (случајне шуме, градијентно појачавање и рекурентне неуронске мреже) који у обзир узимају текућа макроекономска кретања и анализирају постојање везе између остварених стопа неизвршења обавеза и макроекономских показатеља. Експериментална анализа је извршена на основу временске серије агреgirаних података о стопама неизвршења обавеза на нивоу банкарског сектора Републике Србије. Методе статистичког учења су постигле значајне резултате у побољшању тачности оцјена у односу на регресионе моделе, а као најзначајнији предиктори издвајају се девизни курс динара према еврџу, монетарни агрегат М3 и учешће проблематичних кредита у банкарском сектору. Добијени резултати могу представљати важну информацију за институције и регулаторна тијела надлежна за јачање стабилности финансијског система, јер приказане модерне методе статистичког учења имају значајан потенцијал за анализу и управљање кредитним ризиком.

Кључне ријечи: статистичко учење, кредитни ризик, стопа неизвршења обавеза, макроекономски показатељи, банкарски сектор

Садржај

1	Кредитни ризик	1
1.1	Увод	1
1.2	Кредитни ризик и кредитни рејтинг	2
1.3	Параметри кредитног ризика	4
2	Макроекономска позиција Републике Србије	7
2.1	Преглед макроекономских кретања	7
2.2	Карактеристике банкарског сектора	11
3	Имплементација регулаторних оквира у банкарском сектору	17
3.1	Базелски комитет за супервизију банака	17
3.2	Међународни стандарди финансијског извјештавања	21
4	Методе статистичког учења	25
4.1	Стабла одлучивања	25
4.2	Случајне шуме	34
4.3	Градијентно појачавање	36
4.4	Неуронске мреже	39
4.5	Рекурентне неуронске мреже	41
5	Стопа неизвршења обавеза и макроекономски предиктори: случај банкарског сектора Републике Србије	43
6	Закључак	61
	Библиографија	62

Глава 1

Кредитни ризик

1.1 Увод

У деценијама након глобалне економске и финансијске кризе, банке све више воде рачуна о потреби управљања ризиком. Развијају се софистицирани системи за контролу финансијског ризика, који укључује кредитни ризик, односно вјероватноћу да ће корисници отплатити своје кредите и тржишни ризик, односно вјероватноћу да ће вриједност хартије од вриједности варирати и ризик ликвидности и способности банке да испуни своје обавезе према депонентима и уговореним странама.

Током вишевијековне праксе конвенционалног банкарства, развијене су бројне технике и инструменти за ублажавање кредитног ризика и процес њихове еволуције, због сталних промјена у банкарском пословању, и даље траје. Кредитни ризик је опасност да друга уговорена страна неће испунити своје обавезе у складу са договореним условима. Сматра се најзначајнијим ризиком у банкарском пословању и кључним узроком стечајева банака, а то нису промијениле ни бројне финансијске иновације које су се појавиле у последњих неколико деценија. Ефикасно управљање портфолиом изложености кредитном ризику и кредитном функцијом је основа сигурности и здравог пословања банака. Експертиза банака за управљање ризиком потиче из њихове способности да мјере ризик свог пословања, као и ризик пословања својих клијената, па је исправна процјена ризика основна за адекватно управљање ризиком.

Комплексност управљачких система наметнула се као неопходност у условима када се финансијско тржиште развија великом брзином, када се креи-

рају нови инструменти и банкарски производи. Циљ је да се на адекватан начин контролише изложеност ризику и прате сва евентуална погоршања као и да се предузму превентивне мјере како би се неповољне ситуације спријечиле. Интерни модели које су банке развиле за потребе оцјене способности својих депонената најчешће дају оцјене вјероватноће неизмирења које обухватају читав пословни циклус. За потребе имплементације МСФИ 9¹ у банкарски сектор, неопходне су оцјене вјероватноће неизмирења за конкретан временски тренутак, као и укључивања различитих макроекономских сценарија.

1.2 Кредитни ризик и кредитни рејтинг

Х. Леви² и М. Сарнат³ наводе да је у средњем вијеку употребљавана ријеч *risicum*, која је имала значење данашњег појма игре на срећу, односно означавала је могућност доброг и лошег исхода [26]. У савременој финансијској анализи појам ризика се користи управо у том значењу, као могућност појаве различитих исхода, како позитивних, тако и негативних.

Настајање ризика повезано је са економским промјенама које су углавном резултат развоја фактора који детерминишу исход пословања. У статичким условима нема ризика, јер нема промјена, све је познато или може да буде познато. Са друге стране, у динамичким условима, који имају алеаторни карактер, настају економске промјене које није могуће са сигурношћу предвидјети због недостатка потпуног знања о тим промјенама у будућности.

Савремено банкарско пословање, усљед развоја финансијског тржишта у нашој земљи, претпоставља преузимање високих ризика, а примарно кредитног ризика, те се адекватна процјена изложености истом мора сматрати емпиријском потребом финансијских институција. Базелска комисија за супервизију банака [1] дефинише кредитни ризик као ризик да супротна страна у уговору не изврши обавезу у пуној вриједности или у тренутку доспијећа или у било којем тренутку након доспјећа. Погоршање кредитног ризика

¹ International Financial Reporting Standard 9: Међународни стандард финансијског извјештавања 9 - финансијски инструменти. Циљ овог стандарда је установљавање принципа за финансијско извјештавање о финансијским средствима и финансијским обавезама којим се презентују релевантне и корисне информације корисницима финансијских извјештаја на основу којих врше процјене износа, временског распореда и неизвјесности будућих токова готовине ентитета.

² Haim Levy (1978-), израелски економиста

³ Marshall Sarnat (1927-2006), израелски економиста

представља опадање кредитног квалитета једног или више депонената, такав да је вјероватно да банка неће бити у могућности да наплати, или више не постоји разумно осигурање да ће банка моћи да наплати укупну суму потраживања у складу са одредбама уговора који садржи кредитни ризик.

Губици у кредитном пословању могу настати у ситуацији када дужник не извршава отплату добијеног кредита. Кредитни ризик настаје као резултат кредитног процеса који почиње дефинисањем стратешке одређености банке према одређеном сегменту потенцијалних клијената. На основу потребне документације и добијених информација о платној способности клијената банка доноси одлуку о ризичности клијената и додјељује му оцјену кредитног квалитета (кредитног рејтинга). Висок кредитни рејтинг представља велику вјероватноћу да дужник врати кредит у потпуности без икаквих потешкоћа у отплати, док низак кредитни рејтинг сугерише да је дужник у прошлости имао проблема у отплати кредита и изгледно је да ће у будућности бити суочен са истим. Стога, кредитни рејтинг представља мишљење о кредитном ризику у наредном периоду и подразумејева процјену способности и спремност дужника да правовремено и у потпуности измирује своје обавезе према кредиторима. Уобичајено је давати словне ознаке за кредитну способност неке земље. Словне оцјене у свакој од категорија (страна и домаћа валута) варирају од ААА, што се сматра највишим рејтингом, до СД оцјене која означава да се дуг сервисира селективно (енг. Selective Default). Република Србија сарађује са рејтинг агенцијама S&P, Fitch и Moody's које оцјењују њен кредитни рејтинг за дугорочно и краткорочно задуживање, у иностраној и домаћој валути. С обзиром да агенције Moody's Investors Service и Standard & Poor's користе различите ознаке за рејтинг, у табели 1.1 дата је класификација и упоредни преглед тих ознака. Агенције дијеле све ознаке рејтинга у неколико категорија према критеријумима кредитне способности и кредитног ризика а које се означавају словима од А до Ц, односно Д. У оквиру сваке категорије постоје нивои рејтинга у зависности од тога како је процијењена позиција државе, као јака, просјечна или слаба. При томе, агенција Moody's Investors Service за означавање тих нивоа користи нумеричке ознаке, а Standard & Poor's ознаку плус (+) или минус (-). Агенције дају мишљење за изгледе или очекивања у виду описне оцјене у ком смјеру би се рејтинг могао кретати у будућности. Изгледи могу бити позитивни, стабилни, негативни и промјенљиви у зависности од развоја догађаја.

Moody's	S&P	Кратак опис категорије рејтинга
Инвестициони ниво		
Aaa	AAA	Најбоља кредитна способност, минималан кредитни ризик
Aa1	AA+	Висока кредитна способност, веома низак кредитни ризик
Aa2	AA	
Aa3	AA-	
A1	A+	Средње висока кредитна способност, низак кредитни ризик
A2	A	
A3	A-	
Baa1	BBB+	Средња кредитна способност, умјерен кредитни ризик
Baa2	BBB	
Baa3	BBB-	
Неинвестициони (спекулативни) ниво		
Ba1	BB+	Кредитна способност са спекулативним елементима, знатан кредитни ризик
Ba2	BB	
Ba3	BB-	
B1	B+	Спекулативна кредитна способност, висок кредитни ризик
B2	B	
B3	B-	
Цaa1	ЦЦЦ+	Лоша кредитна способност, веома висок кредитни ризик
Цaa2	ЦЦЦ	
Цaa3	ЦЦЦ-	
Цa	ЦЦ	Високоспекулативна кредитна способност, извјесно неизвршавање обавеза, селективни банкрот, банкрот
Ц	Ц	
	СД Д	

Табела 1.1: Упоредни преглед ознака рејтинга⁴

1.3 Параметри кредитног ризика

Како банкарско пословање на тржиштима у развоју, као што је наше, претпоставља преузимање ризика, то се адекватна процјена његовој изложениости сматра основном потребом финансијских институција. Параметри кредитног ризика имају значајну улогу у процесу управљања кредитним ризиком у финансијским институцијама и веома су важни у контексту регулаторних захтјева у погледу третмана ризика у банкарском пословању и подстицаја за примјену квантитативних модела. Један од специфичних циљева овог рада

⁴Извор: Bloomberg, Moody's Investors Service, Standard & Poor's.

јесте креирање модела за показивање везе између макроекономских показатеља и остварених стопа неизмирења на подацима банака које послују у Србији. С тим у вези дефинисаћемо различите параметре које помажу у одређивању изложености кредитном ризику: вјероватноћа неизмирења обавеза (PD), изложеност банке у вријеме неизмирења обавеза дужника (EAD) и губитак усљед наступања статуса неизмирења обавезе (LGD). Због широке примјене у поступку мјерења кредитног ризика постоје бројне методе за оцјену параметара кредитног ризика [14].

Вјероватноћа неизмирења обавеза

Вјероватноћа неизмирења обавеза (енг. probability of default - PD) представља вјероватноћу да дужник неће бити у могућности да испуни своју платежну обавезу у одређеном временском периоду. С обзиром на захтјеве или ограничења, вјероватноћа неизмирења обавеза може се рачунати за једног дужника или групу дужника са сличним карактеристикама кредитног ризика. Оцјена вјероватноћа неизмирења обавеза зависи од два фактора: макроекономских (незапосленост, стопа раста бруто домаћег производа, каматна стопа, инфлација и сл) и специфичности дужника (демографске информације, финансијски показатељи и сл).

Изложеност банке у вријеме неизмирења обавеза дужника

Кључни параметар кредитног ризика је изложеност усљед неизмирења обавеза (енг. exposure at default – EAD) и представља бруто износ који дужник није вратио у тренутку догађаја неиспуњења обавеза. Генерално, изложеност се посматра као процјена у којој мјери би банка могла бити изложена уговорној страни у случају неизвршења обавеза партнера. Према Банци за међународна поравнања (BIS) ⁵ тај износ не смије да буде нижи од књиговодствене вриједности потраживања из биланса стања и мора се рачунати без разматрања резерви.

⁵Банка за међународна поравнања (енг. Bank for International Settlements, BIS) јесте међународна финансијска институција у власништву централних банака која поспјешује међународну монетарну и финансијску сарадњу и служи као банка централним банкама.

Губитак усљед наступања статуса неизмирења обавезе

Банка биљежи губитак када компанија са којом је склопила уговор не измирује обавезе плаћања. Према Банци за међународна поравнања (BIS), до неизмирења обавеза долази када наступи један од сљедећих догађаја: прошло је више од 90 дана од било које кредитне обавезе дужника или је дужник поднио захтјев за стечај или сличну заштиту од повјерилаца.

Губитак усљед наступања статуса неизмирења обавезе (енг. loss given default – LGD) дефинише се као стопа губитка насталог усљед изложености банке у тренутку када је наступио статус неизмирења обавеза, при чему се под губитком подразумијева економски губитак одређен узимањем у обзир временске вриједности новца, као и директни и индиректни трошкови у вези са наплатом потраживања.

Глава 2

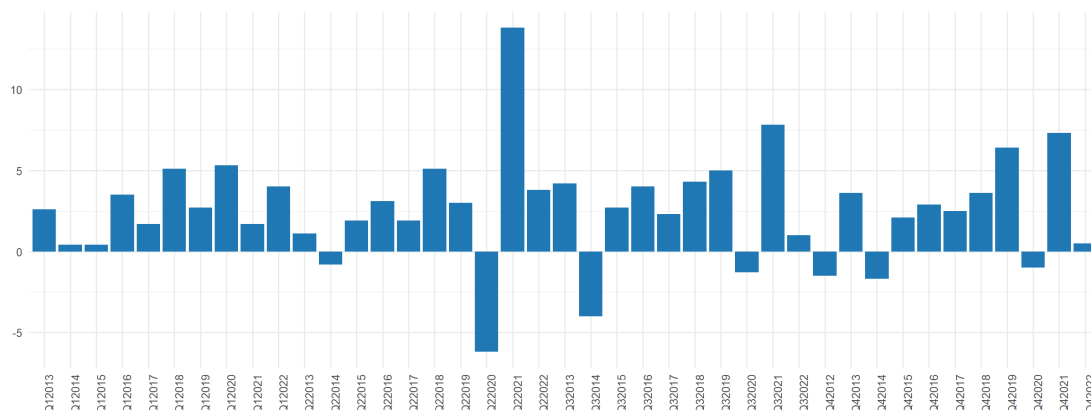
Макроекономска позиција Републике Србије

Изложеност кредитном ризику зависи од многих фактора међу којима се издвајају: усклађеност пословања са важећим регулативама, одабир адекватних кредитних стратегија и политика, степен развијености привреде и макроекономска кретања на тржишту на којем банке послују. Емпиријска истраживања спроведена у банкарском сектору препознају макроекономска кретања као једну од битних одредница у објашњавању степена изложености кредитном ризику у Републици Србији. У овој глави су приказане основне макроекономске перформансе, актуелни трендови и кретање економске активности путем текстуалног и графичког приказа основних статистичких, монетарних и фискалних показатеља.

2.1 Преглед макроекономских кретања

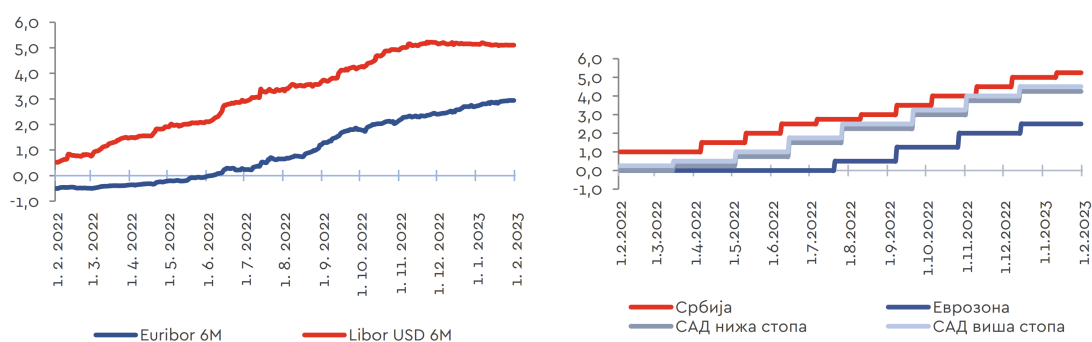
На почетку 2023. године глобална економија се и даље налази у тренутку високе неизвијесности након кумулативних негативних ефеката пандемије COVID-19, украјинске кризе, прекида глобалних ланаца снабдијевања и многих других фактора. Привредна активност Србије, мјерена стопом реалног раста бруто домаћег производа (БДП), остварила је раст од 2.3% у 2022. години вођен растом индустрије и услужних сектора, што се може окарактерисати као релативно добар економски резултат, имајући у виду све економске и геополитичке изазове током 2022. године.

ГЛАВА 2. МАКРОЕКОНОМСКА ПОЗИЦИЈА РЕПУБЛИКЕ СРБИЈЕ



Слика 2.1: Реални раст БДП-а за период од Т4 2012. до Т4 2022. године

Током 2022. године Европска централна банка је започела циклус повећања каматних стопа. Референтна каматна стопа је од фебруара 2022. године повећавана четрнаест пута и тренутно износи 6.25% (јул 2023). У великом броју економија широм свијета, независно од степена развијености, на снази је тренд раста каматних стопа као ефикасан инструмент за сузбијање високе инфлације у ситуацији повећаних економских ризика.

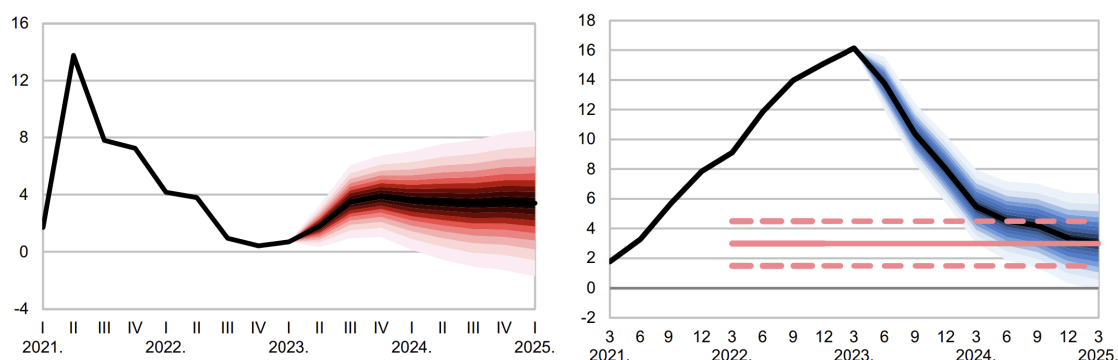


Слика 2.2: Кретање каматних стопа у последњих годину дана у Србији (лијево) и кретање референтних каматних стопа у Србији, еврозони и САД у последњих годину дана (десно)¹

¹Извор: Макроекономски осврт за друго тромјесечје 2023. године, Привредна комора Србије

ГЛАВА 2. МАКРОЕКОНОМСКА ПОЗИЦИЈА РЕПУБЛИКЕ СРБИЈЕ

Просјечна годишња инфлација у 2022. износила је 11.9%, док је међугодишња инфлација у децембру износила 15.1%. И у условима комбинованог дјеловања већег броја фактора, прије свега последица енергетске кризе и суше која је погодила велики дио Европе, инфлација у Србији била је мања од инфлације у упоредивим земљама у окружењу. Инфлација је у периоду јануар–јун 2023. године износила 15.3% што је знатно изнад циљане границе Народне банке Србије ($3.0\% \pm 1.5$ п.п.). У јуну 2023. године у поређењу са истим мјесецем 2022. године, инфлација је износила 13.7%, што показује да је врхунац инфлације на мјесечном нивоу прошао и слиједи процес њеног успоравања који ће бити веома дуг и спор. Према пројекцији Народне банке Србије, инфлација је достигла врхунац у првом кварталу године и од априла се налази у паду. Пад инфлације ће се интензивирати у другој половини године, а њен повратак у границе циља очекује се средином 2024. године.



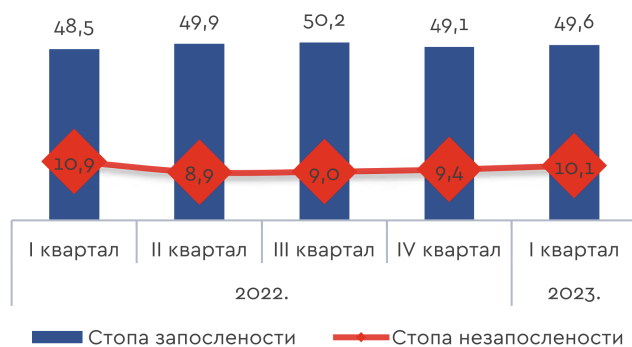
Слика 2.3: Пројекција реалног раста БДП-а (лијево) и пројекција инфлације (десно), мг. стопе, (у %) ²

Од марта 2021. године, агенција Moody's није оцјењивала кредитни рејтинг Србије (Ba2), док су агенције Fitch Ratings и Standard and Poor's задржале кредитни рејтинг на ББ+. S&P је у априлу 2023. године потврдио стабилне изгледе за даље унапређење кредитног рејтинга, чиме су потврђени макроекономска стабилност и повољни изгледи раста, као и адекватност економске политике током кризе. У августу 2023. године, Fitch Ratings је потврдио кредитни рејтинг Србије на нивоу од ББ+ са стабилним изгледима за даље унапређење, у условима појачаних геополитичких тензија, што додатно потврђује адекватност економских политика у Србији.

²Извор: Годишњи извјештај о стабилности финансијског система у 2022. години, Народна банка Србије

ГЛАВА 2. МАКРОЕКОНОМСКА ПОЗИЦИЈА РЕПУБЛИКЕ СРБИЈЕ

Стопа запослености популације старости 15 и више година у другом тромјесечју 2023. године износила је 49.6%, а стопа незапослености 10.1%. Ситуација на тржишту рада се може окарактерисати као стабилна, али треба имати у виду да су високи инфлаторни притисци и даље на снази и да могу имати одређене посљедице по нашу привреду.



Слика 2.4: Кретање стопе запослености/незапослености становништва старости 15 и више година, (у %) ³

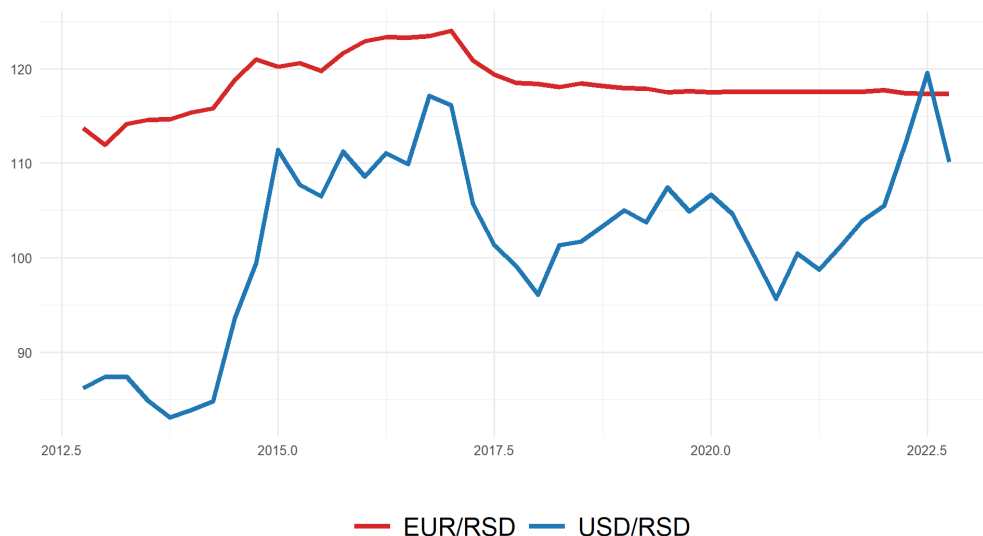
Просјечна бруто зарада обрачуната за период јануар–април 2023. године износила је 114.939 динара (980 евра), док је просјечна нето зарада износила 83.360 динара (711 евра). У односу на исти период претходне године, зарада је реално мања за 0.4%. Уједно, медијална нето зарада за април 2023. године износила је 63.954 динара (545 евра), што значи да је 50% запослених остварило зараду мању од наведеног износа.



Слика 2.5: Просјечне нето и бруто зараде (у динарима)

³Извор: Анкета о радној снази, Републички завод за статистику

Просјечан девизни курс у периоду јануар– јун 2023. године износио је 117.3067 динара за један евро, као и 108.5272 динара за један амерички долар. На слици 2.6 примјетно је да је динар показао завидан степен стабилности према водећој валути еврозоне, упркос бројним економским и геополитичким искушењима.



Слика 2.6: Кретање девизних курсева EUR/RSD и USD/RSD у периоду од Т4 2012. до Т4 2022. године

2.2 Карактеристике банкарског сектора

Према Извјештају банкарског сектора Народне банке Србије за IV тромјесечје 2019. године, 26 банака активно послује на територији Републике Србије. У власништву страних правних лица је 19 банака, а 7 банака је у власништву домаћих правних лица. Детаљан преглед одабраних параметара банкарског сектора Србије (у млрд РСД, у %) дат је у табели 2.1.

Укупна нето билансна актива банкарског сектора Србије на крају децембра 2019. износила је 4084.1 милијарду динара, а укупан билансни капитал 705.7 милијарди динара. На основу упоредног приказа власничке структуре банкарског система Србије датог на слици 2.7 примјетна је контракција банкарског тржишта у форми смањења укупног броја банака, тј. 8 банака је мање

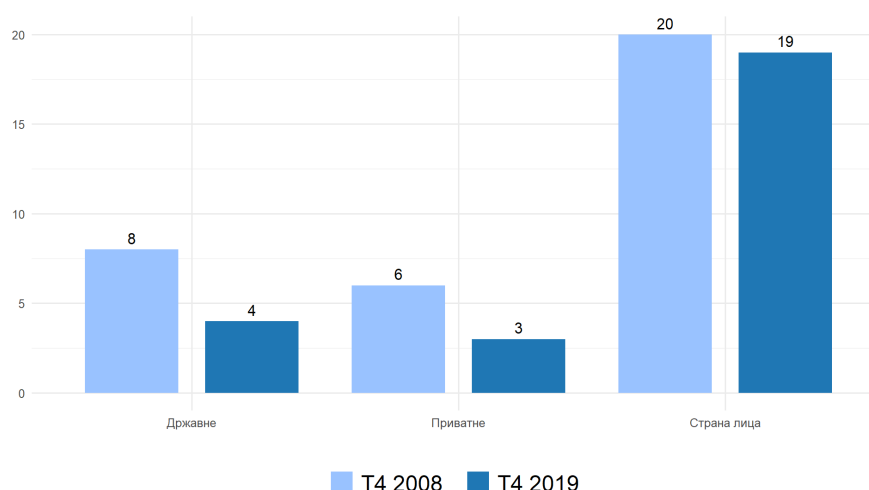
³Сви подаци у Извјештају банкарског сектора Србије дати су на основу извјештаја које су банке дужне да достављају Народној банци Србије.

ГЛАВА 2. МАКРОЕКОНОМСКА ПОЗИЦИЈА РЕПУБЛИКЕ СРБИЈЕ

	Број банака	Актива		Капитал	
		Износ	Учешће	Износ	Учешће
Банке у власништву домаћих лица	7	991	24.3%	181	25.6%
Државне	4	686	16.8%	109	15.4%
Приватне	3	305	7.5%	72	10.2%
Банке у власништву страних лица	19	3093	75.7%	525	74.4%
Италија	2	1094	26.8%	186	26.3%
Аустрија	2	550	13.5%	90	12.7%
Француска	1	119	2.9%	12	1.8%
Мађарска	2	551	13.5%	82	11.6%
Остали	12	779	19.1%	155	22.0%
Укупно	26	4084	100.00%	706	100.00%

Табела 2.1: Преглед одабраних параметара банкарског сектора Србије

на банкарском тржишту, као и повећање удјела банака у власништву страних лица са 59% на 74.4%, али и смањења удјела банака у власништву домаћих лица, како државних тако и приватних. Доминантно учешће у банкарском сектору Србије имају банке поријеклом из Италије, Мађарске и Аустрије (укупно шест банака) са удјелом у укупној билансној суми од 53.7%.



Слика 2.7: Власничка структура банкарског система Србије у четвртом тромјесечју 2019. године у односу на четврто тромјесечје 2008. године

ГЛАВА 2. МАКРОЕКОНОМСКА ПОЗИЦИЈА РЕПУБЛИКЕ СРБИЈЕ

Првих пет банака (Banca Intesa А.Д.- Београд, Unicredit Банка Србија А.Д.- Београд, Комерцијална Банка А.Д.- Београд, ОТП Банка Србија А.Д. Београд, Raiffeisen Банка А.Д.- Београд) посматрано у односу на билансну суму, бруто кредите и депозите, чини више од половине банкарског сектора Србије у наведеним сегментима, с обзиром на то да са 53.4% учествују у нето билансној активи, са 52.0% у бруто кредитима и са 54.9% у депозитима. На основу наведеног, може се закључити да банкарско тржиште Србије одликује задовољавајући степен конкуренције и ниска концентрација активности. Наиме, вриједност Херфиндал-Хиршмановог индекса ⁴ дата у табели 2.2 указује на одсуство концентрације у свим посматраним категоријама. Дужи временски период највише вриједности Херфиндал-Хиршмановог индекса забиљежене су код депозита становништва и прихода од накнада, док је најнижа вриједност забиљежена код прихода од камата и укупних кредита.

	Првих 5 банака	Првих 10 банака	ХХИ
Актива	53.4	79.7	800
Кредити (укупни)	52.0	78.8	789
Кредити становништву	53.3	83.0	838
Кредити привредним друштвима	54.5	83.2	840
Депозити (укупни)	54.9	81.2	840
Депозити становништва	58.5	82.9	968
Приходи (укупни)	54.7	80.5	818
Приходи од камата	50.8	78.1	755
Приходи од накнада	58.8	82.3	930

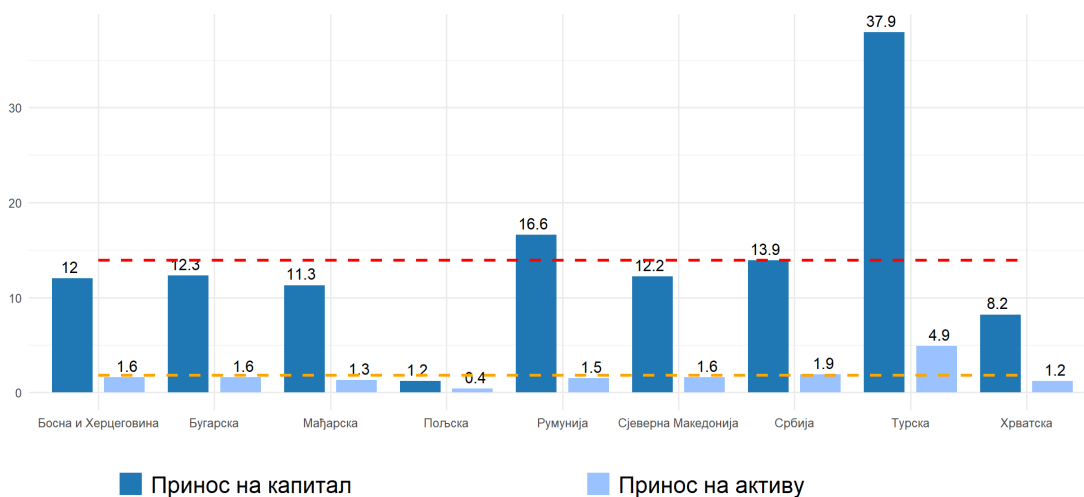
Табела 2.2: Показатељи концентрације и конкуренције банкарског сектора Србије за четврто тромјесечје 2019. године

Банкарски сектор Србије остварио је позитиван финансијски резултат у 2022. години. Остварена профитабилност банкарског сектора на крају године резултирала је приносом на активу овог сектора у износу од 1.9% и приносом на капитал од 13.9%. Принос на активу био је благо изнад просјека региона, а принос на капитал је, и поред високе капитализованости домаћег банкарског сектора, био на нивоу просјека региона.

⁴Херфиндал–Хиршманов индекс концентрације израчунава се као сума квадраних учешћа појединачних банака у укупној категорији која се посматра (актива, кредити, депозити итд.). Вриједност показатеља до 1.000 указује на одсуство концентрације, вриједност између 1.000 и 1.800 на постојање умјерене концентрације, а вриједност изнад 1.800 на постојање изражене концентрације.



Слика 2.8: Показатељи профитабилности (у %)



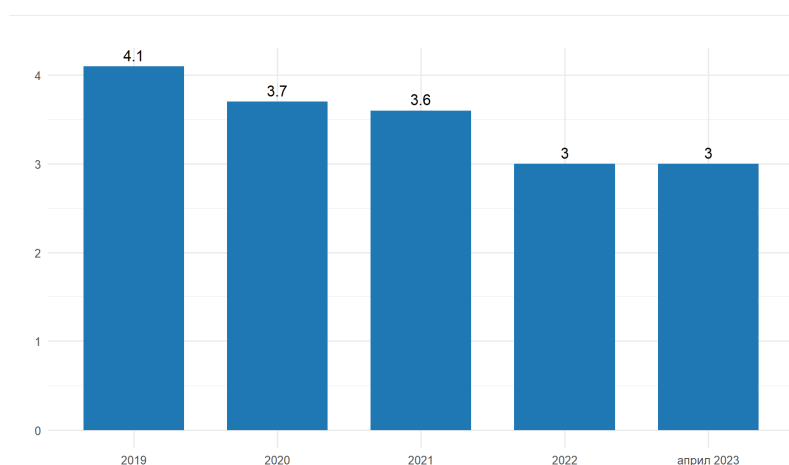
Слика 2.9: Принос на капитал и принос на активу земаља региона за 2022. годину (у %)

ГЛАВА 2. МАКРОЕКОНОМСКА ПОЗИЦИЈА РЕПУБЛИКЕ СРБИЈЕ

Емпиријски показатељ који се најчешће користи за утврђивање степена изложености банака кредитном ризику је удио високо ризичних кредита у укупним кредитима (квалитет активе). Праћење нивоа и тренда проблематичних кредита (eng. non performing loans - NPL) од велике је важности за препознавање могућих проблема у наплати потраживања и праћење кредитног ризика, будући да ови кредити и с њима повезани показатељи представљају показатеље погоршања квалитета кредитног портфолија банкарског сектора.

У складу с методологијом коју примјељује Народна банка Србије, под NPL се подразумијева стање укупног преосталог дуга сваког појединачног кредита (укључујући и износ доцње):

- по основу кога дужник касни са отплатом главнице или камате 90 и више дана од иницијалног рока доспјећа;
- по коме је камата у висини тромјесечног износа (и виша) приписана дугу, капитализована, рефинансирана или је њено плаћање одложено;
- по основу ког дужник касни мање од 90 дана, али је банка процијенила да је способност дужника да отплати дуг погоршана и да је отплата дуга у пуном износу доведена у питање.



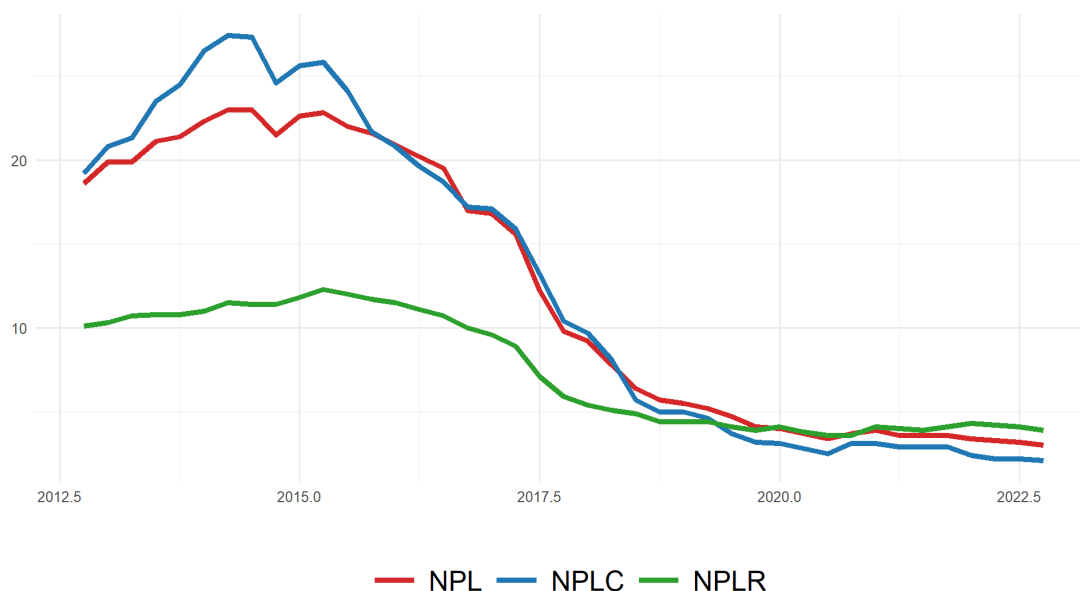
Слика 2.10: Удио проблематичних кредита у укупним кредитима (у %)

Након свјетске економске кризе удио NPL у укупним кредитима биљежили су тренд раста, те је реакција банака која је услиједила као одговор на кризу, довела до појаве смањења обима пласмана, раста ликвидности и смањења

ГЛАВА 2. МАКРОЕКОНОМСКА ПОЗИЦИЈА РЕПУБЛИКЕ СРБИЈЕ

активних и пасивних каматних стопа [27]. Све то је резултирало да укупни NPL банкарског сектора почевши од трећег тромјесечја 2015. године биљеже тренд опадања. Константан пад бруто NPL уз раст укупних кредита условио је даљи пад показатеља учешћа NPL у укупним бруто кредитима на ниво од 3.00% у првом тромјесечју 2023. године, што је њихов нови историјски минимум од прописивања једнообразне дефиниције и обавезе извјештавања 2008. године.

На слици 2.11 приказано је кретање удјела проблематичних кредита у кредитном портфолију и то - банкарски сектор (NPL), кредити одобрени привреди (NPLC) и кредити одобрени становништву (NPLR) у периоду од Т4 2012. до Т4 2022. године.



Слика 2.11: Удио проблематичних кредита у укупним кредитима у периоду од Т4 2012. до Т4 2022. године (у %)

Глава 3

Имплементација регулаторних оквира у банкарском сектору

3.1 Базелски комитет за супервизију банака

Базелски комитет за супервизију банака је у јуну 2004. године издао документ под називом „International Convergence of Capital Measurement and Capital Standards” [2] који представљају тзв. Базел II стандарди, чија је имплементација започела 2006. године. У децембру 2009. године, Базелска комисија издала је консултативни документ под називом „Strengthening the Resilience of the Banking Sector” [3] као дио референтног пакета како би се посветила пажња искуствима из економске кризе. Базел II садржи по први пут интегрисани приступ који покушава да обједини кредитни, тржишни и оперативни ризик, а у циљу обрачуна свеобухватне изложености банке ризику и обрачуна потребног капитала у складу са специфичним портфолиом ризика банке.

Два основна принципа споразума су да осигура адекватан ниво капитала у међународном банкарству и да изгради такав систем да банке не могу да повећавају обим пласмана без повећања капитала. Основни циљеви Базел II стандарда су повећање покрића изложености ризицима и смањење процикличности са тежњом промовисања отпорнијег банкарског сектора у будућности.

Процјена кредитног ризика за потребе утврђивања минимално потребног капитала раније се базирала искључиво на регулаторно прописаним нормама. Међутим, Базел II се заснива на флексибилнијим и софистициранијим при-

ступима мјерења кредитног и оперативног ризика, те подразумијева иновације које резултирају већом осјетљивошћу на ризике и то [19]:

- стандарди су допуњени са два додатна стуба који се баве надзорном ревизијом и тржишном дисциплином;
- већим банкама је дозвољено да користе сопствене приступе за процјену кредитног ризика засноване на интерним рејтинзима (енг. Internal rating based approach – IRB);
- банкама је дозвољено да користе рејтинге које обезбјеђују екстерне институције за процјену кредитног ризика (енг. External credit rating institutions - ECAI) за класификацију изложености државама и привредним друштвима.

Нови Базел II стандарди настоје побољшати оквире за адекватност капитала и стављају главни фокус на управљање ризицима и подстицање побољшања у способности банака да процјењују ризике. Први пут се уводи нови концепт три стуба који тежи усклађивању прописаних захтјева са економским принципима управљања ризицима.

Стуб I: Минимални капитални захтјеви

Стуб I уводи новине у капитални оквир усклађујући минималне капиталне захтјеве банке са обимом и врстом њеног пословања и природом ризика који из њега произилазе. Задржава се постојећа дефиниција капитала из Базел I и минимални ниво коефицијента адекватности (енг. Capital adequacy ratio - CAR) од 8%. ¹ Побољшања у оквиру Базел II фокусирана су на унапређење мјерења ризика користећи укупну активу пондерисану ризиком, односно

$$CAR = \frac{\text{Укупан капитал}}{\text{Кредитни ризик} + \text{Тржишни ризик} + \text{Оперативни ризик}}$$

Стуб I стандарда Базела II примјењује се у Србији по основу Одлуке о адекватности капитала банке („Службени гласник РС”, бр. 46/2011).

¹У литератури се обично третира коефицијент адекватности преко 10% као добро капитализована банка, између 8% и 10% као адекватно капитализована банка, испод 8% као подкапитализована банка, а испод 2% као значајно подкапитализована банка.

Стуб II: Надзор над адекватношћу капитала

Стуб II у фокус ставља процес надзорне ревизије који се имплементира на националном нивоу, чији је акценат на анализи процеса интерне процјене адекватности капитала банке. Супервизор треба да анализира и процјени адекватност тог интерног процеса како би одредио да ли руководство банке примјењује разбориту процјену свих ризика и да ли је банка обезбиједила адекватан износ капитала у односу на обим и врсту својих пословних активности. Процес супервизије заснива се на сљедећа четири принципа чији је циљ унапређење интерних система управљања ризицима и контроле:

- процес интерне процене адекватности капитала (ICAAP): банке треба да врше континуирану интерну процјену адекватности капитала у складу са својим портфолиом ризика;
- процес супервизије (енг. Supervisory Review and Evaluation Process – SREP): супервизори треба да процјењују ICAAP и стратегију одржавања потребног нивоа капитала банака;
- капитал изнад минималног нивоа: очекивање супервизора да банке посједују капитал изнад минималног нивоа дефинисаног стубом I;
- интервенције супервизора: супервизори треба да интервенишу у раном стадијуму како би се спријечило да капитал падне испод минималног нивоа потребног да подржи ризике.

Стуб III: Тржишна дисциплина

Стуб III поспјешује јачање тржишне дисциплине путем захтјева за објављивање релевантних информација о пословању банака. Ефикасно објављивање података је основа осигурања да други учесници на финансијском тржишту имају прилику да изврше сопствену анализу пословања банке, да разумију профил ризика банке, као и адекватност њихових капиталних позиција. Развој тржишне дисциплине учесника на финансијском тржишту, у складу са стубом III, доприноси сигурности и стабилности банкарског сектора и финансијског система. Стуб III стандарда Базела II примјењује се у Србији по основу Одлуке о објављивању података и информација банке („Службени гласник РС”, бр. 45/2011).

Увођење Базела II у Србији завршено је пакетом прописа које је Извршни одбор Народне банке Србије донио у јуну 2011. године. Овај пакет прописа, којима се правила за израчунавање адекватности капитала банке и управљање ризицима усклађују с Базелом II, односно којима се повећава транспарентност у пословању банака – чини шест одлука објављених у „Службеном гласнику РС”, бр. 45/2011 и 46/2011, и то: Одлука о адекватности капитала банке (с прилозима), Одлука о извјештавању о адекватности капитала банке, Одлуке о управљању ризицима банке, Одлука о објављивању података и информација банке, Одлука о контроли банкарске групе на консолидованој основи и Одлука о извјештавању банака.

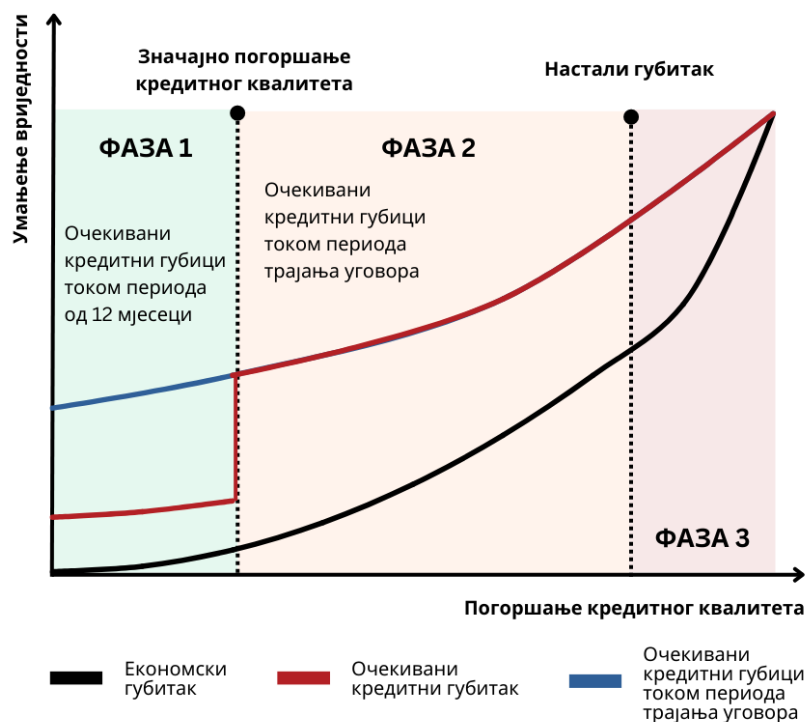
С друге стране, имајући у виду да је увођење Базела II у Србији спроведено у вријеме кад је Базелски комитет за супервизију банака већ био објавио Базел III, неки елементи Базела III већ су уведени у правни систем Србије и то Одлуком о адекватности капитала банке. Најзначајније новине предвиђене Базел III стандардима објављене у [28] односе се на:

- унапређење квалитета капитала и јачање капиталних захтјева;
- увођење минималних стандарда за управљање ризиком ликвидности, као и минималних показатеља ликвидности;
- побољшање процеса управљања ризицима у циљу одговарајућег покрића свих ризика капиталом;
- увођење левериџ рација као додатног показатеља пословања банака који није заснован на степену ризичности активе;
- увођење низа мјера у циљу формирања што адекватнијих резерви капитала које би биле расположиве у периодима кризе.

3.2 Међународни стандарди финансијског извјештавања

Банке у Републици Србији имају обавезу, утврђену одредбама Закона о банкама („Службени гласник РС”, бр. 107/2005, 91/2010 и 14/2015), да при састављању годишњих финансијских извјештаја примјењују Међународне стандарде финансијског извјештавања, у даљем тексту: МСФИ, (енг. International Financial Reporting Standard - IFRS) од дана који је надлежно међународно тијело одредило као дан почетка њихове примјене. Одбор за међународне рачуноводствене стандарде објавио је да се документ МСФИ 9 – финансијски инструменти [4] примјењује од 1. јануара 2018. године а његов значај огледа се у промјени начина класификације, вредновања и обрачуна умањења вриједности финансијских инструмената. Поред тога, МСФИ 9 за банке представља важан заокрет у погледу признавања и исказивања кредитних губитака, будући да умјесто концепта насталих кредитних губитака уводи концепт очекиваних кредитних губитака који, између осталог, укључује и макроекономске моделе приликом вршења њихове процјене и обрачуна.

МСФИ 9 у себи инкорпорира нови модел умањења вриједности финансијске имовине заснован на очекиваним губицима, а овај модел ће директно утицати на знатно раније признавање кредитних губитака од стране банака што ће довести до раста исправки вриједности са једне стране, а са друге стране до реалнијег и квалитетнијег портфолија кредита. Главни фокус МСФИ 9 није на томе да ли постоје објективни индикатори умањења вриједности, већ се кроз модел очекиваних губитака процјењује да ли ће до губитака доћи или не. Овакав приступ резултира идентификацијом губитака значајно прије него што се исти догоди, а то повећава квалитет финансијских информација у финансијским извјештајима, али истовремено доприноси побољшању финансијске стабилности у земљи. Кључна новина у погледу исправке вриједности по финансијским инструментима коју уводи МСФИ 9 јесте модел очекиваних кредитних губитака (енг. Expected Credit Losses - ECL) умјесто претходног модела насталих губитака. Базелски комитет за супервизију банака дао је подршку примјени модела очекиваних кредитних губитака у циљу ранијег признавања кредитних губитака и поштовања здравих пракси управљања кредитним ризиком [20].



Слика 3.1: Модел очекиваног губитка унутар МСФИ 9

Као што можемо видјети на слици 3.1, модел очекиваних губитака унутар МСФИ 9 подразумијева три фазе квалитета финансијске имовине [8]. У прву фазу убрајамо све финансијске инструменте чији се кредитни ризик није значајно повећао након почетног признавања или који имају низак кредитни ризик на датум извјештавања. У случају да дође до значајног повећања кредитног квалитета, већ сада долази до раста исправке вриједности по основу кредита, на прелазу из фазе 1 у фазу 2. Према МСФИ 9 умањење вриједности у овој фази једнако је очекиваним кредитним губицима током периода трајања уговора, за разлику од прве фазе када су кредитни губици били једнаки очекиваним кредитним губицима током периода од 12 мјесеци. Коначно, у случају утврђивања објективних индикатора умањења вриједности кредита, исти прелази у фазу 3. На основу ове анализе можемо закључити да МСФИ 9 много раније уочава промјене кредитног квалитета, те сходно томе утиче на раст исправке вриједности по кредитима.

Очекивани губици за прелазак из фазе 1 у фазу 2 рачунају се на основу обрасца [13]:

$$ECL = \sum_{t=1}^T MPD_t \cdot LGD \cdot EAD_t \cdot D_t + MPD_{T+\tau} \cdot LGD \cdot EAD_{T+\tau} \cdot D_{T+\tau},$$

гдје је

- MPD_t - маргинална вјероватноћа неизмирења обавезе,
- $MPD_{T+\tau}$ - маргинална вјероватноћа неизмирења обавезе у посљедњем (непотпуном) периоду,
- LGD - губитак усљед наступања статуса неизмирења обавезе,
- EAD - изложеност усљед неизмирења обавеза,
- T - за фазу 2: цијели дио преосталог животног вијека финансијског инструмента (у годинама) од датума извјештавања; за фазу 1: $T = 1$ ако је очекивани животно вијек финансијског инструмента бар 1 година, $T = 0$ иначе;
- τ - за фазу 2: разломљени дио преосталог животног вијека финансијског инструмента (у годинама) од датума извјештавања; за фазу 1: $\tau = 0$ ако је очекивани животно вијек финансијског инструмента мањи од 1 године;
- $D_t = \frac{1}{(1+r)^{t-0,5}}$
 $D_{T+\tau} = \frac{1}{(1+r)^{T+\tau/2}}$ } - дисконтни фактор за посматрани период;
- r - ефективна каматна стопа.

Маргинална вјероватноћа неизмирења обавезе може се добити као

$$MPD_t = PD_t \cdot (1 - CPD_{t-1}) = CPD_t - CPD_{t-1},$$

гдје је CPD_t кумулативна вјероватноћа неизмирења

$$CPD_t = \begin{cases} CPD_{t-1} + (1 - CPD_{t-1}) \cdot PD_t, & t > 0 \\ 0, & t = 0. \end{cases}$$

Са друге стране, очекивани губици у фази 3 рачунају се по обрасцу:

$$ECL = LGD_{t_{def}} \cdot EAD_{act}$$

гдје је

- EAD_{act} - стварна вриједност кредитних обавеза које дужник није вратио на датум извјештавања;
- $LGD_{t_{def}}$ - очекивани ниво губитака усљед наступања статуса неизмирења обавеза дефинисан у зависности од броја година које су прошле од наступања статуса неизмирења;
- t - број година неизмирења обавеза.

Новим моделом обезбиједиће се знатно веће исправке вриједности, првенствено кредита у портфолију, а у коначници финансијска стабилност банке, а кроз банку и укупног финансијског система у једној земљи. На овај начин, финансијски извјештаји ће бити састављени на реалнијој основи него што су били до сада, а то значи да ће се смањити ризик од негативних дејстава тржишних шокова као што је то био случај током свјетске финансијске кризе.

Глава 4

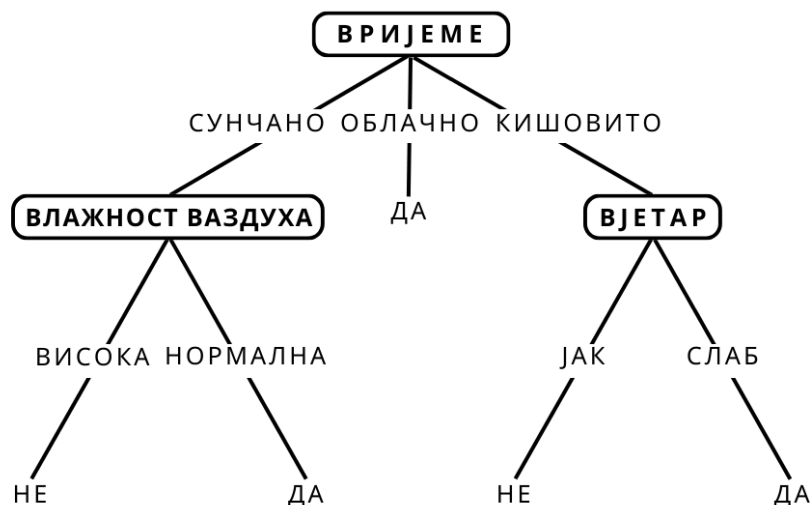
Методе статистичког учења

Статистичко учење је област вјештачке интелигенције која се бави изградњом рачунарских система који уче из искуства методама вјероватноће и статистике, без експлицитног програмирања. У посљедњој деценији, развојем дубоког учења (енг. deep learning), долази до многих пробоја у области вјештачке интелигенције, резултирајући изненађујућим перформансама рачунарских система. Фокус проучавања се помјера са дедукције на индукцију и тиме формализује уопштавање од ограниченог броја узорака ка универзалним закључцима. Теоријски оквири ове главе приказани су по узору на [7].

4.1 Стабла одлучивања

Стабла одлучивања представљају једноставан и интерпретабилан модел статистичког учења, примјенљив и на регресију и на класификацију. У сваком чвору стабла, осим у листовима, налази се по један тест који има више од једног исхода. Сваком исходу одговара једна грана стабла која води до сљедећег чвора. Листови су означени вриједностима које представљају предвиђање стабла. Стабло одлучивања се може посматрати и као списак ако-онда правила, гдје свака грана представља конјункцију услова над предикторима, а цијело стабло дисјункцију. Примјер стабла одлучивања [9] дат је на слици 4.1, који представља израз:

$$\begin{aligned} & (\text{вријеме} = \text{сунчано}) \wedge (\text{влажност ваздуха} = \text{нормална}) \\ & \vee (\text{вријеме} = \text{облачно}) \\ & \vee (\text{вријеме} = \text{кишовито} \wedge \text{вјетар} = \text{слаб}). \end{aligned}$$



Слика 4.1: Примјер стабла одлучивања

Из примјера примјећујемо да су инстанце представљене листама парова предиктор-вриједност, циљана функција је Булова, али се иста може уопштити на дискретну или чак реалну функцију.

Приликом дизајна алгоритма за учење стабла одлучивања, циљ је изабрати најкориснији предиктор који ће бити тестиран у коријену стабла за дати скуп инстанци. Основна идеја је изабрати предиктор који најбоље раздваја инстанце, тако да добијене подгрупе буду што хомогеније у односу на класе. Стога, потребно је дефинисати мјеру квалитета предиктора у односу на разврставање које нуди, односно мјеру чистоће (енг. purity) или хомогености добијених подгрупа. Неки од познатих приступа квантификавању хомогености скупа инстанци су ентропија и Ђинијев индекс. Оба приступа почивају на удјелима инстанци различитих класа у укупном скупу. Статистичка особина информационе добити (енг. information gain) дефинише се ентропијом скупа S као

$$H(S) = -p_{\oplus} \log p_{\oplus} - p_{\ominus} \log p_{\ominus},$$

гдје је p_{\oplus} пропорционалан однос позитивних инстанци у S , а p_{\ominus} однос негативних (узима вриједност нула ако сви примјери припадају истој класи). У општем случају, ако предиктор има C дискретних вриједности онда се ентропија дефинише као

$$H(S; p_1, \dots, p_C) = - \sum_{i=1}^C p_i \log p_i.$$

Лако се уочава да ентропија представља мјеру нехомогености. Заиста, претходни израз достиже максималну вриједност ако за свако i важи $p_i = \frac{1}{C}$, а вриједност 0, ако је $p_i = 1$ за неко i . Ђинијев индекс дефинише се као

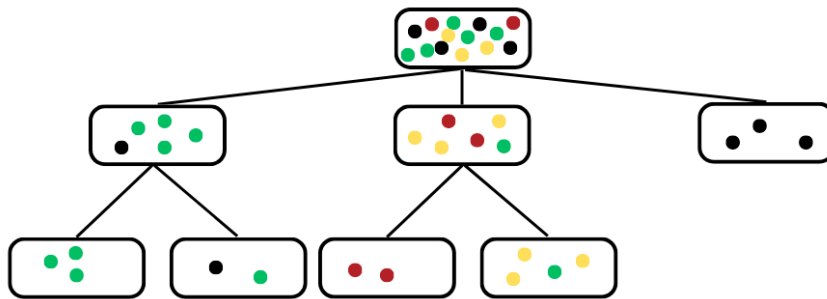
$$G(S; p_1, \dots, p_C) = 1 - \sum_{i=1}^C p_i^2.$$

Примијетимо да је вриједност Ђинијевог индекса једнака 0 у случају да све инстанце припадају једној класи, док ако је расподјела класа униформна, достиже се максимална вриједност.

Ако са $V(A)$ означимо скуп вриједности предиктора A и са $S_v = \{s \in S : A(s) = v\}$ скуп инстанци за које је вриједност предиктора A једнака v , онда се предиктор обично евалуира тако што се израчуна смањење нехомогености након партиционисања скупа инстанци по вриједностима тог предиктора, односно добит предиктора дата је изразом

$$\text{Gain}(S; A) = H(S) - \sum_{v \in V(A)} \frac{|S_v|}{|S|} H(S_v),$$

гдје је $|\cdot|$ ознака за кардиналност скупа. Добит предиктора очигледно изражава разлику између нехомогености полазног скупа и просјечне нехомогености његових подскупова добијених партиционисањем по вриједностима датог предиктора. Смањење нехомогености дуж путање од коријена до листа стабла илустровано је на слици 4.2.



Слика 4.2: Илустрација хомогености инстанци у чворовима стабла

Постоји експоненцијално велики број стабала одлучивања која се могу конструисати из датог скупа података [11]. У наставку наводимо верзију ID3 алгоритма за Булове функције. Као критеријум гранања стабла користи се информациона добит. Главна предност ID3 алгоритма јесте његова

једноставност, те се често користи као база алгоритама за индукцију стабла одлучивања.

ID3(S,C,prediktori)

1. креирај чвор Коријен стабла
2. ако су сви примјери у S позитивни, врати Коријен са ознаком = +
3. ако су сви примјери у S негативни, врати Коријен са ознаком = -
4. ако је листа предиктора prediktori које треба тестирати празна, врати Коријен са ознаком = најчешћа вриједност у примјерима
5. иначе
 - a) $A \leftarrow$ из листе предиктора prediktori који најбоље класификује према Gain (S,A), Коријен $\leftarrow A$
 - b) за сваку вриједност v_i предиктора A:
 - i. додај нову грану испод Коријен за услов $A = v_i$
 - ii. ако је S[v_i] празан онда
 - испод додај лист са ознаком = најчешћа вриједност у примјерима
 - инаће додај испод поддрво ID3(S[v_i],C,prediktori-A)
6. врати Коријен

Овакав ID3 алгоритам се лако може уопштити за функцију дискретних вриједности и користи се обично као прозор који чини подскуп примјера задате величине над којима се примијени овакав алгоритам, а онда се у прозор унесу сви изузеци из скупа примјера који не одговарају класификацији и ако их има понови се поступак. На тај начин превазилази се ограничење величине скупа примјера.

Дубока стабла су обично склона преприлагођавању, пошто примјеном великог броја тестова могу да опишу небитне специфичности података на којима су обучавана. Са друге стране, плитка стабла обично имају проблем потприлагођавања. Дакле, примијетимо да дубина стабла представља регуларизациони хиперпараметар, али упркос могућности регуларизације, у пракси испољавају велику дисперзију.

Основна предност стабла одлучивања је њихова интерпретабилност и јасна важност појединих предиктора за конкретан проблем. Такође, стабла одлучивања природно комбинују предикторе различитих врста (категоричке и непрекидне), неосјетљива су на монотоне трансформације предиктора и врло су ефикасна у вријеме предвиђања.

Модел стабла одлучивања који се заснива на регресији

Претпоставимо да се скуп података \mathcal{D} састоји од p улазних вриједности и одговора за сваку од укупно N инстанци, тј. имамо парове опсервација (x_i, y_i) за $i = 1, 2, \dots, N$, гдје је $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$. Алгоритам треба да идентификује тачке подјеле, као и да одреди топологију коју стабло одлучивања треба да има. Најприје, примијетимо да сваком чвору стабла m одговара једна област простора предиктора R_m и вриједност c_m која је тој области придружена. Тада се стабло може записати као

$$f(x) = \sum_{m=1}^M c_m I\{x \in R_m\}. \quad (4.1)$$

Ако као критеријум користимо минимизовану суму квадрата $\sum (y_i - f(x_i))^2$, лако добијамо да ће оцјена за \hat{c}_m бити средња вриједност одговора y_i који припадају области R_m :

$$\hat{c}_m = \frac{1}{\#\{x_i \in R_m\}} \sum_{(x_i, y_i) \in \mathcal{D}, x_i \in R_m} y_i.$$

Проналажење најбоље бинарне партиције у смислу минимизације суме квадрата обично је рачунски неизводљиво. Стога настављамо са похлепним алгоритмом. Уочимо подиону варијаблу j и подиону тачку s , те дефинишимо полуравни

$$R_1(j, s) = \{X \mid X_j \leq s\} \quad \text{и} \quad R_2(j, s) = \{X \mid X_j > s\}.$$

Тада подиону варијаблу j и подиону тачку s добијамо као рјешење минимизационог проблема

$$\min_{j,s} \left[\min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2 \right].$$

За сваки избор j и s , добија се да је

$$\hat{c}_1 = \frac{1}{\#\{x_i \in R_1(j, s)\}} \sum_{(x_i, y_i) \in \mathcal{D}, x_i \in R_1(j, s)} y_i,$$

$$\hat{c}_2 = \frac{1}{\#\{x_i \in R_2(j, s)\}} \sum_{(x_i, y_i) \in \mathcal{D}, x_i \in R_2(j, s)} y_i.$$

За сваку подиону варијаблу, одређивање подионе тачке се може обавити веома брзо и стога пролазимо кроз све улазне податке ради избора најбољег пара (j, s) . Након што пронађемо најбоље дијелење, партиционисемо податке у двије области и понављамо описани поступак дијелења на сваку од тих области. Затим исти поступак понављамо за све резултујуће области.

Када смо креирали потпуно стабло одлучивања са свим могућим подионим областима, постоји могућност да је креирано стабло исувише комплексно и да дубина стабла није одговарајућа. Често нису сви крајњи чворови стабла хомогени, што оставља одређени степен грешке. Метода скраћивања стабла (енг. pruning) рјешава проблем преклапања (нехомогености) података. Подрезана верзија је мање комплексна и лакша је за разумијевање.

Претпоставимо да смо на основу улазних података креирали стабло одлучивања T_0 са свим предикторима (потпуни модел), а затим дато стабло скраћујемо методом заснованом на мјери сложености скраћивања (енг. cost-complexity pruning) стабла, коју описујемо у наставку. Нека је подстабло $T \subset T_0$ било које подстабло које се може добити скраћивањем стабла T_0 , односно спајањем одређеног броја његових унутрашњих чворова. Индексирајмо крајње чворове (листовете) стабла са m , при чему m одговара области R_m и нека је $|T|$ број крајњих чворова стабла. Уведимо ознаке:

$$N_m = \#\{x_i \in R_m\}, \quad \hat{c}_m = \frac{1}{N_m} \sum_{x_i \in R_m} y_i$$

$$Q_m(T) = \frac{1}{N_m} \sum_{x_i \in R_m} (y_i - \hat{c}_m)^2.$$

Теорема 4.1. Сложеност скраћивања стабла је мјера дата формулом

$$C_\alpha(T) = \sum_{m=1}^{|T|} N_m \cdot Q_m(T) + \alpha \cdot |T|, \quad (4.2)$$

гдје је $\alpha \in \mathbb{R}^+$ ненегативан реалан параметар који представља сложеност додавања новог предиктора моделу (енг. complexity parameter).

Ако означимо са $T_\alpha \subset T_0$ подстабло потпуног модела које има најмању вриједност сложености, тада је α регуларизациони хиперпараметар који минимизује вриједност $C_\alpha(T)$. Веће вриједности параметра α индуковаће стабла T_α мање дубине, и обрнуто. Може се показати да за свако α постоји јединствено најмање подстабло T_α које минимизује вриједност сложености $C_\alpha(T)$. Да бисмо нашли такво стабло T_α , користимо методу скраћивања најслабијих карика, односно постепено спајамо унутрашње чворове који резултују најмање повећање вриједности $\sum_m N_m Q_m(T)$ по сваком чвору, те поступак настављамо док не дођемо до коријена стабла. Овај поступак резултира коначним низом подстабала и може се показати да овај низ мора садржати подстабло T_α са описаним особинама [5].

Теорема 4.2. Ако су T_1 и T_2 подстабла стабла T , при чему важи $C_\alpha(T_1) = C_\alpha(T_2)$, тада важи да је стабло T_1 подстабло стабла T_2 или је стабло T_2 подстабло стабла T_1 , па је $|T_1| < |T_2|$ или $|T_2| < |T_1|$.

Теорема 4.3. Ако је $\alpha > \beta$ тада је $T_\alpha = T_\beta$ или је T_α подстабло стабла T_β .

Теорема 4.4. За дате вриједности параметара комплексности $\alpha_1, \alpha_2, \dots, \alpha_m$, $T_{\alpha_1}, T_{\alpha_2}, \dots, T_{\alpha_m}$ и $C_{\alpha_1}(T), C_{\alpha_2}(T), \dots, C_{\alpha_m}(T)$ се могу ефикасно израчунати.

Докази претходних тврђења као и додатне информације о овој теми могу се пронаћи у [5].

Посљедица 4.1. Приметијетимо да је скраћено стабло T_α најмање подстабло стабла T за које је вриједност сложености $C_\alpha(T)$ минимална, односно за које је сложеност креирања стабла минимална

$$\min_{T_\alpha \subset T} C_\alpha(T).$$

Модел стабла одлучивања који се заснива на класификацији

Ако је циљ стабла одлучивања алгоритам који ће вршити класификацију по категоријам $1, 2, \dots, K$, једине промјене у односу на модел који се заснивао на регресији биће критеријум за формирање чворова и скраћивање стабла. Наиме, код модела са регресијом користили смо $Q_m(T)$ као мјеру нехомогености, а у наставку дефинишемо нове за класификацију.

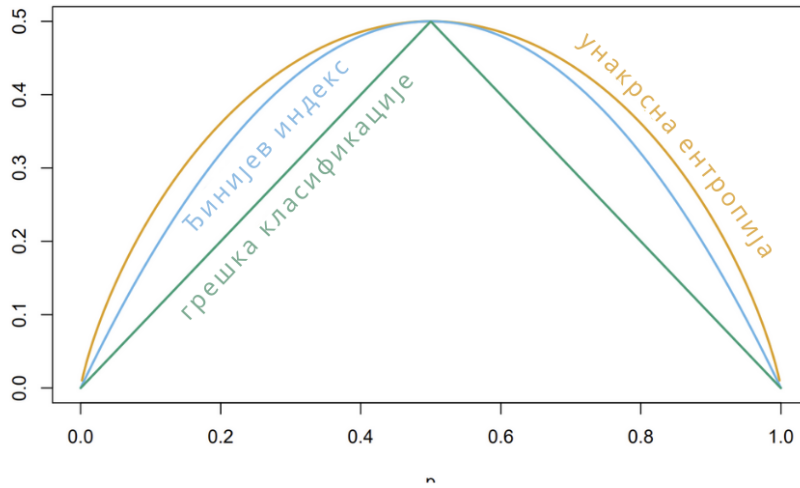
Ако чвор m који одговара области R_m има N_m инстанци, тада је

$$\hat{p}_{mk} = \frac{1}{N_m} \sum_{x_i \in R_m} I\{y_i = k\}, \quad (4.3)$$

удио инстанци класе k у чвору m . Инстанцама из чвора m додјелујемо класу $k(m) = \arg \max_k \hat{p}_{mk}$, најбројнију класу у чвору m . Дефинишимо сљедеће мјере нехомогености чворова:

- грешка класификације: $\frac{1}{N_m} \sum_{i \in R_m} I\{y_i \neq k(m)\} = 1 - \hat{p}_{mk(m)}$;
- Ђинијев индекс: $\sum_{k \neq k'} \hat{p}_{mk} \hat{p}_{mk'} = \sum_{k=1}^K \hat{p}_{mk} (1 - \hat{p}_{mk})$;
- унакрсна ентропија: $-\sum_{k=1}^K \hat{p}_{mk} \log \hat{p}_{mk}$.

На слици 4.3 приказане су мјере нехомогености инстанци у чворовима стабла за класификацију у двије категорије, као функције у зависности од удјела p инстанци одређене класе у датом чвору датог изразом 4.3.



Слика 4.3: Мјере нехомогености инстанци у чворовима стабла

Примјер 4.1. Претпоставимо да вршимо метод класификације у двије класе. Ако је p удио инстанци у другој класи, тада су претходно дефинисане

мјере нехомогености инстанци у чворовима редом: $1 - \max\{p, 1 - p\}$, $2p(1 - p)$ и $-p \log p - (1 - p) \log(1 - p)$. Са претходног графика можемо примијетити да су унакрсна ентропија и Ђинијев индекс диференцијабилне функције и као такве погодније су за даљу нумеричку оптимизацију.

Примјер 4.2. Унакрсна ентропија и Ђинијев индекс су осјетљивији на промјене у вјероватноћама чворова у односу на грешку класификације. Заиста, ако разматрамо проблем класификације у двије класе са 400 инстанци у свакој (означимо то са [400,100]) и претпоставимо једно гранање резултује подјелом на чворове [300,100] и [100,300], док друго гранање формира чворове [200,400] и [200,0]. Оба описана гранања имају грешку класификације 0,25, али другим гранањем добијамо хомоген чвор који је вјероватно пожељнији. Такође, и Ђинијев индекс и унакрсна ентропија имају мање вриједности за друго гранање, те су стога чешће у употреби за формирање стабала одлучивања. Са друге стране, за процес заснован на сложености скраћивања стабла могу да се користе све три описане мјере нехомогености, али је најчешће у употреби мјера грешке класификације.

Ђинијев индекс можемо интерпретирати на два начина. Наиме, умјесто да класификујемо инстанце у најбројнију класу у чвору, можемо их класификовати у класу k са вјероватноћом \hat{p}_{mk} . Тада је очекивана грешка једнака $\sum_{k \neq k'} \hat{p}_{mk} \hat{p}_{mk'}$ Ђинијевом индексу. Слично, ако кодирамо сваку опсервацију са 1 ако припада класи k и са 0 иначе, тада је дисперзија у 0-1 чворовима једнака $\hat{p}_{mk}(1 - \hat{p}_{mk})$. Сумирањем претходног израза по класама k добијамо Ђинијев индекс.

Избор оптималног параметра комплексности

Најбољи избор скраћеног стабла T_α , односно вриједности параметра комплексности α , могуће је извршити методом унакрсне валидације (енг. cross-validation), кроз сљедеће кораке [12]:

1. Груписати све могуће вриједности параметра комплексности α у m дисјунктних интервала, гдје је $m < |T|$. Интервали су облика $I_1 = [0, \alpha_1]$, $I_2 = (\alpha_1, \alpha_2]$, \dots , $I_m = (\alpha_{m-1}, \infty]$, гдје све вриједности параметра $\alpha \in I_i$ дијеле исто минимизовано подстабло. Овакво груписање је могуће као последица теореме 4.3.

2. Оцијенити комплетан модел са датим подацима и израчунати вриједности параметара $\beta_1 = 0, \beta_2 = \sqrt{\alpha_1 \alpha_2}, \dots, \beta_{m-1} = \sqrt{\alpha_{m-2} \alpha_{m-1}}, \beta_m = \infty$, гдје је свако β_i вриједност која одговара интервалу I_i .
3. Подијелити податке у s група G_1, G_2, \dots, G_s гдје је свака величине $\frac{s}{n}$, а затим за сваку групу појединачно понављати сљедећи поступак:
 - оцијенити потпун модел на основу свих вриједности осим оних из групе G_i и креирати стабла $T_{\beta_1}, T_{\beta_2}, \dots, T_{\beta_m}$ за овај редуковани скуп података;
 - одредити оцијењену класу за свако G_i на основу сваког модела T_{β_j} , $1 \leq j \leq m$;
 - израчунати мјеру сложености креирања модела за свако од креираних стабала.
4. Сумирати добијене вриједности по групи G_i у циљу добијања оцијењеног ризика за свако β_j . Стабло T_β је најбоље одабрано (скраћено) подстабло за одговарајућу вриједност β за коју је израчунати параметар комплексности најмањи.

4.2 Случајне шуме

Један од најуспјешнијих приступа статистичком учењу су ансамбли (енг. ensemble), односно скупови већег броја модела који заједнички доносе одлуке. Главна претпоставка на којој почива овај приступ јесте да већи број модела, конструисан на адекватан начин, може дати значајно бољу прецизност од само једног модела. Постоје различити приступи конструкцији ансамбла, а један од најкоришћенијих је проста агрегација.

Проста агрегација (енг. bootstrap aggregation, bagging) подразумијева обучавање већег броја модела који појединачно не морају имати високу прецизност, али чије су грешке независне. Приликом предвиђања, сви модели нуде своја предвиђања, која се затим агрегирају како би се добило предвиђање ансамбла. Уколико се ради о регресионим моделима, агрегација се обично врши упросјечавањем, а у случају класификације гласањем. Снага овог модела почива на идеји да се упросјечавањем независне грешке поништавају, што значајно поправља предвиђања.

Претпоставимо да разматрамо регресиони модел и да имамо тренинг скуп података $\mathbf{Z} = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ за који је потребно одредити предикцију $\hat{f}(x)$ за улазне податке x . Проста агрегација налази тражену предикцију упросјечавањем бутстреп узорака. За сваки бутстреп узорак \mathbf{Z}^{*b} , $b = 1, 2, \dots, B$, оцјењујемо модел налазећи вриједност предикције $\hat{f}^{*b}(x)$. Оцјена добијена простром агрегацијом дата је са

$$\hat{f}_{bag}(x) = \frac{1}{B} \sum_{b=1}^B \hat{f}^{*b}(x). \quad (4.4)$$

Означимо са $\hat{\mathcal{P}}$ емпиријску расподелу која свакој инстанци (x_i, y_i) додјељује једнаку вјероватноћу $\frac{1}{N}$. Заправо, стварна вриједност оцјене методом прости агрегације дефинише се као $E_{\hat{\mathcal{P}}} \hat{f}^*(x)$, гдје је $\mathbf{Z}^* = \{(x_1^*, y_1^*), (x_2^*, y_2^*), \dots, (x_N^*, y_N^*)\}$ и важи $(x_i^*, y_i^*) \sim \hat{\mathcal{P}}$. Израз 4.4 представља Монте Карло оцјену стварне вриједности, која тежи тој вриједности када $B \rightarrow \infty$.

Један од најуспјешнијих и практично најчешће коришћених примјера прости агрегације и ансамбла уопште су случајне шуме. Метод случајних шума (енг. random forest) заснива се на простој агрегацији стабала одлучивања. Ансамбл се састоји од m стабала тренираних на различитим подскуповима скупа за обучавање, како би њихове грешке биле што слабије корелисане, што оставља простор за поправку агрегацијом. У наставку наводимо алгоритам генерисања случајних шума.

1. За $b=1, \dots, B$:
 - a) Изабрати бутстреп узорак \mathbf{Z}^* обима N из тренинг података.
 - b) Формирати стабло T_b методом случајне шуме, рекурзивно понављајући наредне кораке за сваки крајњи чвор стабла, док се не достигне минимални број чворова.
 - Одабрати m предиктора од датих p .
 - Наћи најбољу подиону тачку међу одабраним предикторима.
 - Подијелити чвор на два потомка.
2. Излаз је ансамбл стабала $\{T_1, T_2, \dots, T_B\}$.

На основу претходно описаног алгоритма, предвиђање за нову тачку x добијамо на сљедећи начин:

- за моделе регресије: $\hat{f}_{rf}^B(x) = \frac{1}{B} \sum_{b=1}^B T_b(x)$;
- за моделе класификације: Нека је $\hat{C}_b(x)$ предвиђање класе за b -то стабло у ансамблу. Тада је $\hat{C}_{rf}^B(x) = \text{највише гласова } \{\hat{C}_b(x)\}_1^B$.

У пракси се препоручује сљедеће:

- У моделима за класификацију, подразумијевана вриједност за m је $\lfloor \sqrt{p} \rfloor$, а минимална величина чвора је један.
- У моделима за регресију, подразумијевана вриједност за m је $\lfloor \frac{p}{3} \rfloor$, а минимална величина чвора је пет.

На овај начин смо обезбиједили да, уколико у скупу имамо један или више предиктора који имају значајно већу предиктивну моћ од осталих, избјегнемо ситуацију гдје имамо доста сличних стабала одлучивања у ансамблу. Наравно, када су у питању реални подаци, сваком проблему треба приступити посебно и ове параметре третирати као регуларизационе хиперпараметре.

4.3 Градијентно појачавање

Приликом прости агрегације, модели се конструишу потпуно независно. Ипак, то не значи да није могуће унаприједити приступ конструкцији ансамбла, ако би модели некако узимали у обзир понашање других модела. Основна идеја појачавања (енг. boosting) је да се ансамбли граде додајући модел по модел, при чему се сваки од модела обучава тако да што боље надомјести слабости текућег модела, односно да га појача. Алгоритми појачавања најчешће почињу изградњу ансамбла од модела $F_0(x) = 0$ и претпостављајући да је доступан ансамбл F_{m-1} граде ансамбл F_m додавањем још једног модела на сљедећи начин:

$$(\beta_m, f_m) = \arg \min_{\beta, f} \sum_{i=1}^N L(y_i, F_{m-1}(x_i) + \beta f(x_i)),$$

$$F_m = F_{m-1}(x) + \beta_m f_m(x),$$

гдје је L функција грешке. Број модела од којих се састоји ансамбл је хипер-параметар. Примијетимо да се у случају регресије, ако је $L(u, v) = (u - v)^2$, добија врло интуитивно рјешење

$$L(y_i, F_{m-1}(x_i) + \beta f(x_i)) = ((y_i - F_{m-1}(x_i)) - \beta f(x_i))^2,$$

односно, нови модел βf покушава да што боље предвиди грешку предвиђања постојећег ансамбла F_{m-1} .

По питању прецизности, методе градијентног појачавања (енг. gradient boosting) су међу најбољим методама појачавања и статистичког учења уопште. Основна идеја почива на градијентним методама оптимизације, у којима се текуће рјешење оптимизационог проблема поправља додавањем вектора пропорционалног негативној вриједности градијента функције која се минимизује. Наиме, негативна вриједност градијента показује смјер опадања функције, а у контексту појачавања, ова идеја се реализује тако што се модел f_m којим се допуњује ансамбл F_{m-1} одреди тако да апроксимира градијент грешке по функцији F_{m-1} .

У наставку демонстрирамо алгоритам градијентног појачавања помоћу стабала дефинисаних изразом 4.1, који се у пракси најчешће користи. Први модел f_0 одређује се стандардним одлучивањем, Модел f_m којим се допуњује ансамбл F_{m-1} одређује се тако што за сваку инстанцу (x_i, y_i) израчуна вриједност

$$r_{im} = - \left[\frac{\partial L(y, F(x))}{\partial F(x)} \right]_{F=F_{m-1}, (x,y)=(x_i,y_i)},$$

а потом се обучи регресионо стабло над свим паровима (x_i, r_{im}) . Ово стабло дефинише области R_i који су од значаја, међутим, вриједности c_i не морају бити релевантне. Наиме, стабло је регресионо, али ансамбл може бити и класификациони. Стога, ове вриједности се рачунају поново рјешавањем проблема

$$c_i = \arg \min_c \sum_{(x,y) \in \mathcal{D}, x \in R_i} L(y, F_{m-1}(x) + c_i).$$

Након тога, ансамбл се ажурира на стандардан начин:

$$F_m(x) = F_{m-1}(x) + \sum_{i=1}^M c_i I\{x \in R_i\}.$$

У табели 4.1 приказани су градијенти најчешће коришћених функција губитака. У случају квадратне грешке негативни градијент представља резидуале

модела $-r_{im} = y_i - F(x_i)$, а ако је апсолутна грешка негативни градијент је знак резидуала. Хуберова функција грешке дата је са

$$L(y, f(x)) = \begin{cases} [y - f(x)]^2, & \text{за } |y - f(x)| \leq \delta, \\ 2\delta|y - f(x)| - \delta^2, & \text{иначе.} \end{cases}$$

У нашем случају је $\delta_m = \alpha$ -квартил $\{|y_i - F(x_i)|\}$, те Хуберова грешка представља компромис квадратне и апсолутне грешке.

Модел	Функција губитка	$-\partial L(y_i, F(x_i))/\partial F(x_i)$
Регресија	$\frac{1}{2}[y_i - F(x_i)]^2$	$y_i - F(x_i)$
Регресија	$ y_i - F(x_i) $	$\text{sgn}[y_i - F(x_i)]$
Регресија	Хубер	$y_i - F(x_i), \quad y_i - F(x_i) \leq \delta_m$ $\delta_m \text{sgn}[y_i - F(x_i)], \quad y_i - F(x_i) > \delta_m$
Класификација	Мултиномијална	к-та компонента: $I\{y_i = \mathcal{G}_k\} - p_k(x_i)$

Табела 4.1: Градијенти најчешћих функција губитака

За модел класификације у K класа, претпоставимо да се вриједности зависне промјенљиве Y налазе у неуређеном скупу $\mathcal{G} = \{\mathcal{G}_1, \dots, \mathcal{G}_k\}$. Мултиномијална функција губитка дефинише се као

$$\begin{aligned} L(y, p(x)) &= - \sum_{k=1}^K I\{y = \mathcal{G}_k\} \log p_k(x) \\ &= - \sum_{k=1}^K I\{y = \mathcal{G}_k\} f_k(x) + \log \left(\sum_{l=1}^K e^{f_l(x)} \right), \end{aligned}$$

гдје је за свако $k = 1, 2, \dots, K$

$$p_k(x) = P\{Y = \mathcal{G}_k|x\} = \frac{e^{f_k(x)}}{\sum_{l=1}^K e^{f_l(x)}}.$$

За свако стабло T_{km} одговарајући негативни градијент је

$$\begin{aligned} -r_{ikm} &= \left[\frac{\partial L(y_i, F_1(x_i), \dots, F_K(x_i))}{\partial F_K(x_i)} \right]_{F(x_i)=F_{m-1}(x_i)} \\ &= I\{y_i = \mathcal{G}_k\} - p_k(x_i). \end{aligned}$$

4.4 Неуронске мреже

Неуронске мреже (енг. neural networks) представљају параметризовану презентацију која може послужити за апроксимацију других функција и као такве су једна од најпопуларнијих и најпримјењенијих метода статистичког учења. Могу се користити као модели засновани на регресији или класификацији. Основни модел неуронске мреже дат је са

$$\begin{aligned} Z_m &= \sigma(\alpha_{0m} + \alpha_m^T X), \quad m = 1, 2, \dots, M, \\ T_k &= \beta_{0k} + \beta_k^T Z, \quad k = 1, 2, \dots, K, \\ f_k(X) &= g_k(T), \quad k = 1, 2, \dots, K, \end{aligned} \tag{4.5}$$

гдје је $T = (T_1, T_2, \dots, T_K)$, $Z = (Z_1, Z_2, \dots, Z_M)$ вектор линеарних комбинација предиктора и активациона функција $\sigma(v)$. Могуће је изабрати различите активационе функције, али се у пракси најчешће користи сигмоидна функција $\sigma(v) = 1/(1 + e^{-v})$. Функција излаза g_k омогућава финалну трансформацију излазног вектора T . За модел регресије најчешћи избор је функција $g_k(T) = T_k$, а за модел класификације у K класа функција меког максимума (енг. softmax)

$$g_k(T) = \frac{e^{T_k}}{\sum_{l=1}^K e^{T_l}}.$$

Проблем обучавања неуронске мреже у општем случају заснива се на мјерама квалитета, односно минимизовању средње грешке. За моделе регресије, у пракси се користи сума квадрата резидуала

$$R(\theta) = \sum_{k=1}^K \sum_{i=1}^N (y_{ik} - f_k(x_i))^2,$$

а за моделе класификације средње квадратна грешка или унакрсна ентропија

$$R(\theta) = - \sum_{k=1}^K \sum_{i=1}^N y_{ik} \log f_k(x_i),$$

и одговарајући класификатор $G(x) = \arg \max_k f_k(x)$. У специјалном случају, ако је активациона функција меког максимума и мјера квалитета унакрсна ентропија, тада је модел неуронске мреже заправо линеарна логистичка регресија у скривеним слојевима мреже, а оцјењивање параметара се врши методом максималне вјеродостојности.

Најчешће, није нам потребна оцјена глобалног минимума функције $R(\theta)$ како не би дошло до преприлагођавања подацима, те се користе регуларизациони приступи засновани на методи градијентне оптимизације. Врши се алгоритмом пропације уназад (енг. back-propagation) који описујемо у наставку за квадратну грешку. Нека је $z_{mi} = \sigma(\alpha_{0m} + \alpha_m^T x_i)$ дефинисано у изразу 4.5 и $z_i = (z_{1i}, z_{2i}, \dots, z_{Mi})$. Имамо да је грешка дата са

$$\begin{aligned} R(\theta) &= \sum_{i=1}^N R_i \\ &= \sum_{k=1}^K \sum_{i=1}^N (y_{ik} - f_k(x_i))^2, \end{aligned}$$

и одговарајући парцијални изводи

$$\begin{aligned} \frac{\partial R_i}{\partial \beta_{km}} &= -2(y_{ik} - f_k(x_i))g'_k(\beta_k^T z_i)z_{mi}, \\ \frac{\partial R_i}{\partial \alpha_{ml}} &= -\sum_{k=1}^K 2(y_{ik} - f_k(x_i))g'_k(\beta_k^T z_i)\beta_{km}\sigma'(\alpha_m^T x_i)x_{il}. \end{aligned}$$

На основу претходног добијамо да је градијент у $(r + 1)$ -ој итерацији дат са

$$\begin{aligned} \beta_{km}^{(r+1)} &= \beta_{km}^{(r)} - \gamma_r \sum_{i=1}^N \frac{\partial R_i}{\partial \beta_{km}^{(r)}}, \\ \alpha_{ml}^{(r+1)} &= \alpha_{ml}^{(r)} - \gamma_r \sum_{i=1}^N \frac{\partial R_i}{\partial \alpha_{ml}^{(r)}}, \end{aligned}$$

гдје је γ_r стопа учења (енг. learning rate). Ако су δ_{ki} и s_{mi} вриједности грешака тренутног модела на излазним и скривеним јединицама, редом, тада оне задовољавају једнакост

$$s_{mi} = \sigma'(\alpha_m^T x_i) \sum_{k=1}^K \beta_{km} \delta_{ki},$$

која се још назива једначина пропације уназад.

Алгоритам пропације уназад није лако примјенљив на дубоке неуронске мреже, због великог броја узастопних множења. Конкретно, често долази до тога да вриједности парцијалних извода који се рачунају буду практично нула или да дође до прекорачења или просто нестабилности оптимизационих метода. Овај феномен се назива проблемом нестајућих и експлодирајућих градијената (енг. exploding and vanishing gradients).

4.5 Рекурентне неуронске мреже

Постоје различите врсте неуронских мрежа, а у наставку ће бити дискутоване рекурентне неуронске мреже које се најчешће користе за обраду података налик низовима промјенљиве дужине. Дијелови текста приказани су по узору на [22].

Рекурентне неуронске мреже (енг. recurrent neural networks) представљају архитектуру мрежа специјализовану за обраду секвенцијалних података, попут реченица природног језика и временских серија. Конструисане су са идејом да се моделује зависност међу инстанцама. Елементи улазне секвенце се обрађују у корацима, мрежа има скривено стање које акумулира информацију о елементима секвенце обрађеним у претходним корацима, а параметри одређују на који начин се то стање мијења из корака у корак на основу претходног стања и текућих улаза и како се генерише излаз мреже у зависности од текућег стања.

Основни модел рекурентне неуронске мреже са једним скривеним слојем дат је на сљедећи начин:

$$\begin{aligned} h^{(0)} &= 0, \\ h^{(t)} &= \sigma(b_h + Wh^{(t-1)} + Ux^{(t)}), \\ o^{(t)} &= \sigma(b_o + Vh^{(t)}), \end{aligned}$$

за $t = 1, 2, \dots, T$, гдје је T дужина секвенце и може се разликовати од секвенце до секвенце. Вектор $x^{(t)}$ представља улаз у тренутку t , $h^{(t)}$ вектор вриједности скривеног слоја, σ је нелинеарна активациона функција, $o^{(t)}$ је вектор излаза мреже који даје предвиђање праве вриједности $y^{(t)}$. Параметри су матрица трансформације стања у стање W , улаза у стање U , стања у излаз V и вектори слободних чланова b_h и b_o . Сви параметри заједно биће означавани са w . На слици 4.4 дата је графичка репрезентација рекурентне мреже, која сакрива број јединица по слојевима, а наглашава временску димензију.

Проблем обучавања рекурентне неуронске мреже може бити различито постављен у зависности од специфичности архитектуре, али као и до сада, минимизује се средња грешка. У општем случају грешка се рачуна на основу секвенце вриједности y и секвенце предвиђања, те је функција грешке дата формулом:

$$L([y^{(1)}, \dots, y^{(T)}], [f_w^{(1)}, \dots, f_w^{(T)}]),$$

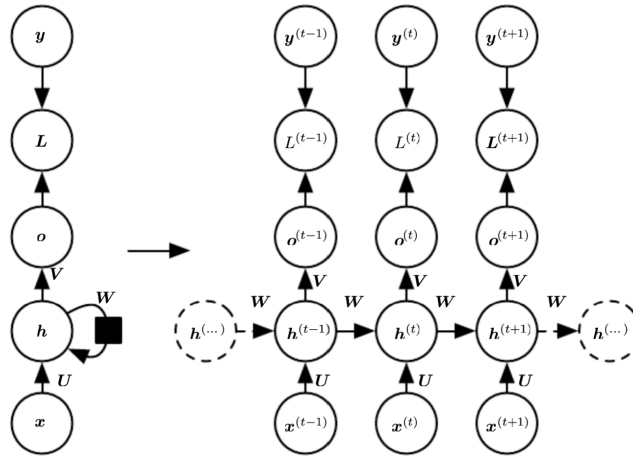
гдје $f_w^{(t)}$ представља предвиђање модела у кораку t и може се записати као

$$f_w^{(t)} = f_w([x^{(1)}, \dots, x^{(t)}]).$$

Најчешће, функција грешке се може разложити на сљедећи начин:

$$L([y^{(1)}, \dots, y^{(T)}], [f_w^{(1)}, \dots, f_w^{(T)}]) = \sum_{t=1}^T L(y^{(t)}, f_w^{(t)}).$$

Граф израчунавања је важан технички појам за разумијевање рекурентних неуронских мрежа и представља основу имплементација различитих библиотека за рад са произвољним неуронским мрежама. Граф израчунавања је граф чији су чворови промјенљиве које могу представљати или улазе или међурезултате израчунавања. Овакав граф се не користи само за представљање модела, већ свих израчунавања која су укључена у поставку оптимизационог проблема (регуларизација, функција грешке и сл). Приказ рекурентне неуронске мреже на слици 4.4 управо представља граф израчунавања (слика лијево) и алтернативни приказ графа добијен размотавањем (слика десно).



Слика 4.4: Рекурентна неуронска мрежа

Захваљујући томе што моделују прелазе из стања у стање, умјесто директне зависности излаза од цјелокупног улаза, рекурентне мреже су у стању да представе врло дубоке структуре зависности, да обрађују секвенце промјенљиве и моделују дугорочне зависности. Са друге стране, познато је да је њихово обучавање теже хардверски убрзати због високог обима меморијског протока.

Глава 5

Стопа неизвршења обавеза и макроекономски предиктори: случај банкарског сектора Републике Србије

У овој глави представићемо истраживање утицаја системских фактора на стопу неизвршења обавеза у случају банкарског сектора Републике Србије. Користићемо двије групе података: (1) подаци о стопама неизвршења обавеза и (2) макроекономски подаци. Временске серије обухватају период од четвртог тромјесечја 2012. године до четвртог тромјесечја 2022. године.

Подаци о стопама неизвршења обавеза

Удружење банака Србије формирало је крајем 2019. године националну базу података стопе неизвршења обавеза по основу кредита. База обухвата податке на агрегатном нивоу и по појединачним сегментима, почев од првог тромјесечја 2012. године. У оквиру наведене базе података, извршене су двије сегментације годишње стопе неизвршења обавеза по основу кредита: према типу кредита (десет основних категорија) и према системском значају банке (три групе банака). Рачунају се двије стопе неизвршења обавеза:

- i на основу броја клијената који уђу у статус неизвршења обавеза;
- ii на основу износа изложености који је ушао у статус неизвршења обавеза.

ГЛАВА 5. СТОПА НЕИЗВРШЕЊА ОБАВЕЗА И МАКРОЕКОНОМСКИ ПРЕДИКТОРИ: СЛУЧАЈ БАНКАРСКОГ СЕКТОРА РЕПУБЛИКЕ СРБИЈЕ

У оба случаја, статус неизвршења обавеза представља кашњење у измирењу обавеза према банци дуже од 90 дана. Годишња стопа неизвршења обавеза по основу кредита (енг. default rate) одређује се по обрасцу

$$DR = \frac{N_T^D}{N_{T-12m}},$$

гдје је

- DR - стопа неизвршења обавеза по основу кредита;
- N_T^D - број дужника (односно износ изложености) од датума $T - 12m$ који су у статусу неизвршења обавеза на датум T ;
- N_{T-12m} - број дужника (односно износ изложености) који нису у статусу неизвршења обавеза на датум $T - 12m$.

За стопе неизвршења обавеза користићемо базу података Удружења банака Србије о броју пласмана у статусу неизвршења обавеза, а не на основу изложености (кредита) који је у статусу неизвршења обавеза. Приступ према броју клијената је у складу са дефиницијом неизвршења обавеза по основу кредита, гдје вјероватноћа неизвршења представља вјероватноћу да ће одређени број клијената, у оквиру једне рејтинг категорије, ући у статус неизвршења обавеза у наредних 12 мјесеци. Са друге стране, стопа неизвршења обавеза према изложености може бити пристрасна, посебно када су велики клијенти у питању, јер улазак кредита велике вриједности у статус неизвршења обавеза може бити резултат дејства специфичних фактора (нпр. лоше управљање предузећем, а не дејство рецесије).

Због потребе да се унификују подаци који се прикупљају неопходна је јединствена сегментација клијената који су правна лица. Из тог разлога коришћена је сегментација правних лица у складу са Законом о рачуноводству: Large Corporate - велика правна лица, SME - мала и средња правна лица и Small Business - микро правна лица. Даље, да би се груписали подаци на основу сегмената кредитног портфолија, извршено је здруживање података на начин приказан у табели 5.1 са истакнутим ознакама сегмената.

Дескриптивне статистике овдје посматраних зависних промјенљивих приказане су у табели 5.2. Примјећујемо да највећи варијабилитет имају стопа неизвршења обавеза по основу кредита у сегменту микро правних лица и великих корпоративних клијената. Агрегатна стопа неизвршења обавеза има

ГЛАВА 5. СТОПА НЕИЗВРШЕЊА ОБАВЕЗА И МАКРОЕКОНОМСКИ ПРЕДИКТОРИ: СЛУЧАЈ БАНКАРСКОГ СЕКТОРА РЕПУБЛИКЕ СРБИЈЕ

својства приближна стопа неизвршења обавеза у сегменту физичких лица, будући да у току цијелог посматраног периода њихово учешће је најдоминантније у укупном броју клијената и укупном броју клијената који уђу у статус неизвршења обавеза.

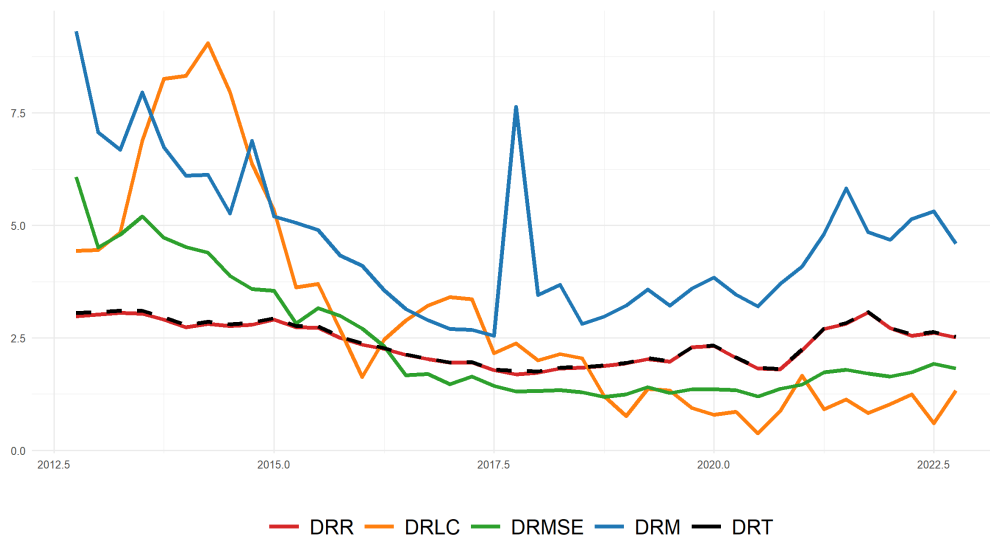
Назив промјенљиве	Ознака	Детаљан опис сегмента
Стопа неизвршења обавеза у банкарском сектору Републике Србије	DRT	
Стопа неизвршења обавеза по основу кредита одобрених великим корпоративним клијентима	DRLC	<ul style="list-style-type: none"> ○ кредити одобрени великим правним лицима (LC) ○ кредити одобрени малим и средњим предузећима (SME)
Стопа неизвршења обавеза по основу кредита одобрених малим и средњим предузећима	DRSME	<ul style="list-style-type: none"> ○ кредити одобрени локалним самоуправама (LG)
Стопа неизвршења обавеза по основу кредита одобрених микро правним лицима	DRM	<ul style="list-style-type: none"> ○ кредити одобрени микро правним лицима (SB) ○ прекорачење по текућем рачуну (OVD) ○ кредитне картице (CC) ○ готовински и потрошачки кредити (CCL) ○ стамбени кредити (MTG) ○ регистрована пољопривредна газдинства (RPG)
Стопа неизвршења обавеза по основу кредита одобрених становништву	DRR	<ul style="list-style-type: none"> ○ предузетници (ENT)

Табела 5.1: Стопа неизвршења обавеза по основу кредита

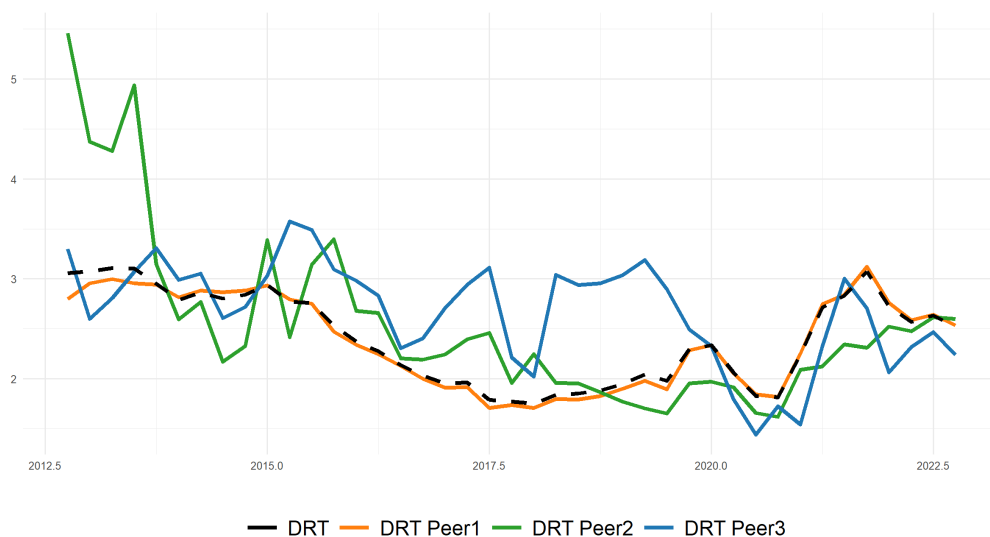
	DRT	DRLC	DRSME	DRM	DRR
средња вриједност	2.411%	2.948%	2.389%	4.658%	2.385%
ст. девијација	0.459%	2.411%	1.363%	1.641%	0.448%
мин.	1.751%	0.380%	1.189%	2.550%	1.691%
макс.	3.106%	9.050%	6.075%	9.320%	3.068%
медијана	2.376%	2.140%	1.710%	4.330%	2.348%
мјера асиметрије	0.029	1.175	1.119	0.866	-0.010
мјера спљоштености	1.529	3.312	2.971	3.124	1.545

Табела 5.2: Дескриптивне статистике агрегатне стопе неизвршења обавеза и стопа неизвршења по сегментима кредитног портфолија

ГЛАВА 5. СТОПА НЕИЗВРШЕЊА ОБАВЕЗА И МАКРОЕКОНОМСКИ ПРЕДИКТОРИ: СЛУЧАЈ БАНКАРСКОГ СЕКТОРА РЕПУБЛИКЕ СРБИЈЕ



Слика 5.1: Стопа неизвршења обавеза у периоду од Т4 2012. до Т4 2022. године у банкарском сектору

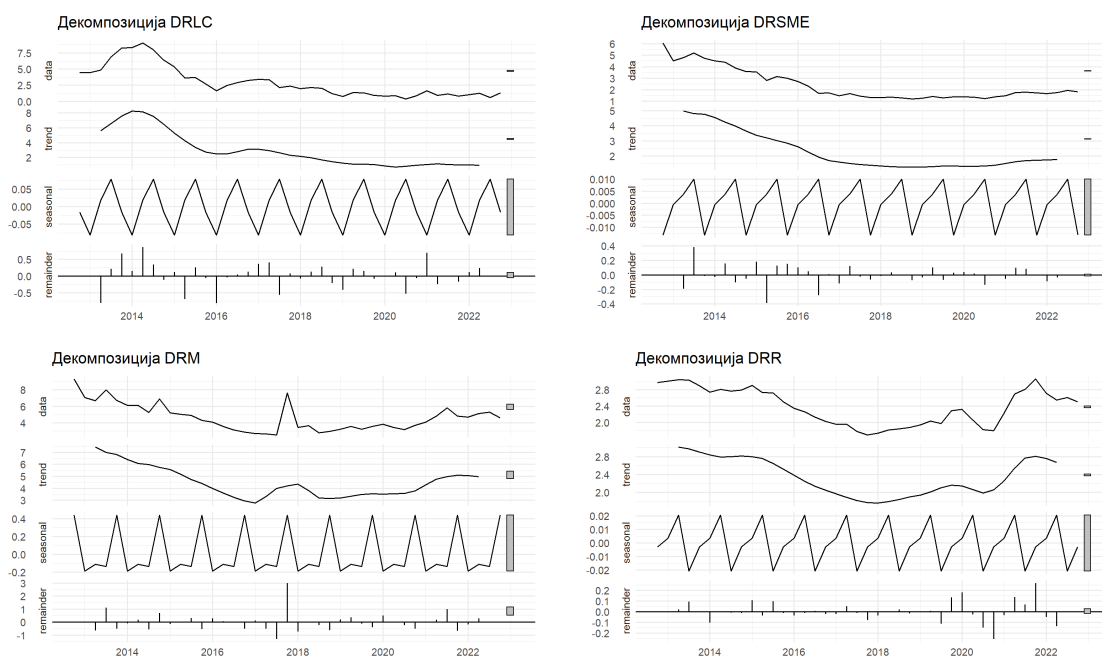


Слика 5.2: Стопа неизвршења обавеза у периоду од Т4 2012. до Т4 2022. године по групама упоредивих банака

На слици 5.1 приказано је кретање агрегатне стопе неизвршења обавеза (DRT) и по сегментима кредитног портфолија (DRLC, DRMSE, DRM и DRR) у периоду од четвртог тромјесечја 2012. до четвртог тромјесечја 2022. године,

ГЛАВА 5. СТОПА НЕИЗВРШЕЊА ОБАВЕЗА И МАКРОЕКОНОМСКИ ПРЕДИКТОРИ: СЛУЧАЈ БАНКАРСКОГ СЕКТОРА РЕПУБЛИКЕ СРБИЈЕ

док је на слици 5.2 приказана еволуција стопе неизмирења по групама упоредивих банака. Упадљиво се издваја да прва група најутицајнијих банака прати тренд стопе неизмирења обавеза у читавом банкарском сектору.



Слика 5.3: Декомпозиција стопе неизмирења обавеза по сегментима кредитног портфолија

У циљу испитивања стационарности временских серија применићемо три теста јединичног коријена: Дики-Фулеров тест (ADF), KPSS тест и Филипс-Перонов тест. У табели 5.3 приказани су резултати за сваки од наведених тестова, одакле слиједи нестационарност посматраних серија.

	Дики-Фулер (ADF)		KPSS		Филипс-Перон		Бр. диф.
	Тест стат.	р	Тест стат.	р	Тест стат.	р	
DRT	-1.3461	0.8307	0.4038	0.0755	-4.0118	0.8796	I(1)
DRLC	-3.2993	0.0858	0.8270	<0.01	-10.945	0.4404	I(1)
DRSME	-1.2455	0.8703	0.8498	<0.01	-4.0083	0.8798	I(2)
DRM	-1.8060	0.6499	0.5158	0.0381	-16.843	0.0904	I(1)
DRR	-1.3626	0.8242	0.3876	0.0825	-4.1805	0.8689	I(1)

Табела 5.3: Тестови јединичног коријена агрегатне стопе неизвршења обавеза и стопа неизвршења по сегментима кредитног портфолија

Макроекономски подаци

У раду ће бити коришћени макроекономски показатељи преузети од Народне банке Србије и Републичког завода за статистику. Временске серије обухватају период од четвртог тромјесечја 2012. до четвртог тромјесечја 2022. године. Преглед показатеља који су укључени у истраживање дат је у табели 5.4, а основне дескриптивне статистике истих у табели 5.5.

Ознака	Макроекономски показатељи
NPL	Учешће проблематичних кредита у банкарском сектору Србије (у %)
NPLC	Учешће проблематичних кредита одобрених привредним субјектима
NPLR	Учешће проблематичних кредита одобрених физичким лицима
GDP	Стопа раста бруто домаћег производа (у %)
EUR	Курс динара према евр
USD	Курс динара према америчком долару
KIR	Референтна каматна стопа Народне банке Србије
IPC	Годишња стопа потрошачких цијена (у %)
IRE	Реални индекс зарада
M3	Монетарни агрегат М3
CAR	Показатељ адекватности капитала
ROE	Принос на капитал
SP	Берзански индекс S&P500
B	BELIBOR 6М

Табела 5.4: Преглед макроекономских показатеља

Одлуком о класификацији билансне активе и ванбилансних ставки банке дата је прецизна дефиниција проблематичних кредита, у складу са којом се исти користе као мјера изложености кредитном ризику. Основна макроекономска детерминанта проблематичних кредита јесте бруто домаћи производ, те због тога испитујемо утицај ове макроекономске промјенљиве на ниво кредитног ризика. Референтна каматна стопа је један од главних канала трансмисије монетарне политике у свим земљама са развијеном тржишном економијом. Када је привреда оптерећена структурним проблемима и високим степеном евроизације, онда овај канал преузима девизни курс. Референтна каматна стопа, као кључна каматна стопа у монетарном систему једне привреде, позитивно је корелисана са нивоом проблематичних кредита ($\rho = 0.7861$). Поред тога, и Белибор (енг. Belgrade Interbank Offered Rate) као референтна каматна стопа утврђена на Панелу банака, на српском међународном тржишту, позитивно је корелисана са нивоом проблематичних

ГЛАВА 5. СТОПА НЕИЗВРШЕЊА ОБАВЕЗА И МАКРОЕКОНОМСКИ ПРЕДИКТОРИ: СЛУЧАЈ БАНКАРСКОГ СЕКТОРА РЕПУБЛИКЕ СРБИЈЕ

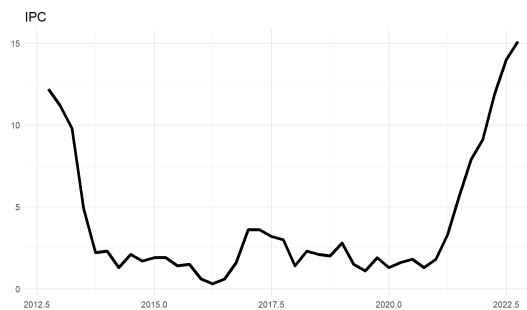
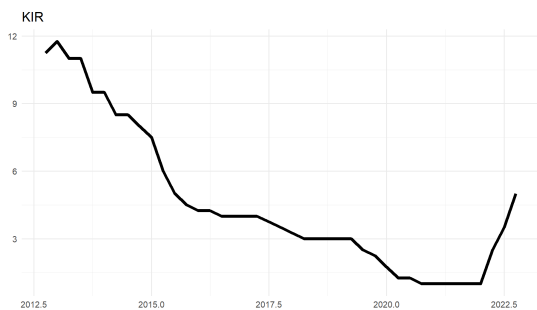
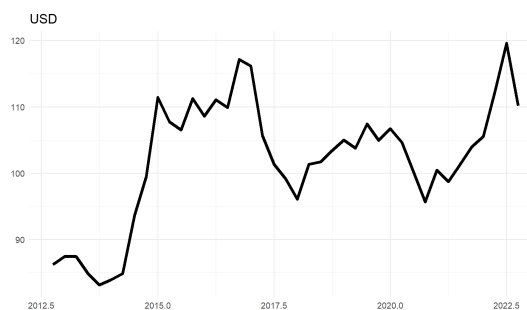
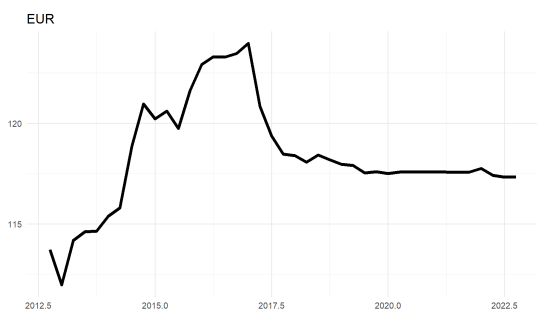
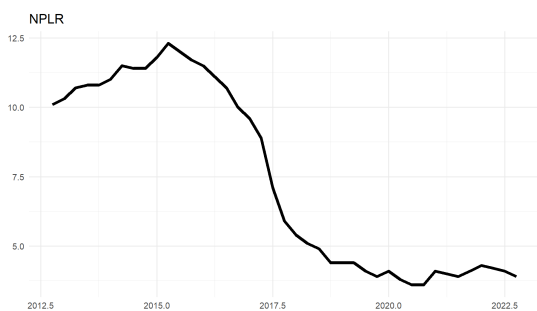
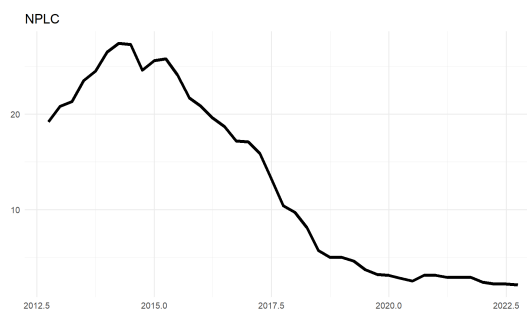
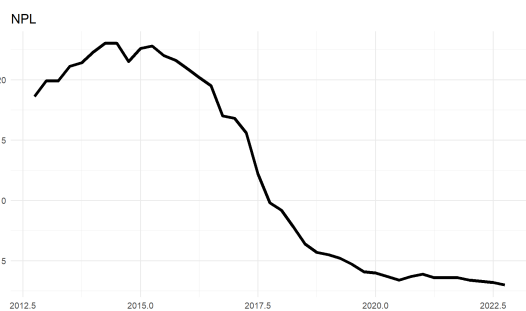
кредита ($\rho = 0.7930$). Индекс реалне нето зараде је један од три макроекономске детерминанте проблематичних кредита у моделу Народне банке Србије и користи се за макропруденцијалне стрес-тестове отпорности банкарског сектора на повећање кредитног ризика усљед погоршања макроекономске позиције једне земље [21]. Раст индекса реалне нето зараде утиче на раст отплатног капацитета физичких лица и на тај начин смањује ниво проблематичних кредита ($\rho = -0.7305$). Ниво солвентности банкарског сектора, мјерен показатељем адекватности капитала, негативно је корелисан са нивоом проблематичних кредита ($\rho = -0.4487$), а ова тврдња се заснива на феномену моралног хазарда. Показатељ профитабилности банке, мјерен приносом на капитал, негативно је корелисан са нивоом кредита коме је изложена банка ($\rho = -0.6830$).

	сред. вр.	ст. дев.	мин.	макс.	мед.	м. асим.	м. спљошт.
NPL	12.2610	8.1035	3.000	23.000	9.800	0.1141	1.2190
NPLC	12.7415	9.4772	2.100	27.400	10.400	0.2014	1.3721
NPLR	7.4366	3.3875	3.600	12.300	5.900	0.1509	1.1926
GDP	2.5561	3.3410	-6.200	13.800	2.700	0.3692	5.4656
EUR	118.3493	2.7215	111.960	123.970	117.750	0.1918	3.0426
USD	101.9341	9.4321	83.130	119.550	103.770	-0.4860	2.5809
KIR	4.5854	3.2407	1.000	11.750	3.750	0.8820	2.6144
IPC	3.9220	3.9999	0.300	15.100	2.100	1.5499	4.0703
IRE	102.5951	3.8935	94.600	110.400	102.700	0.0577	2.1512
M3 (10^{12})	2.4972	0.7394	1.623	40.372	2.255	0.6309	2.0652
CAR	21.5317	1.2479	19.400	23.700	21.600	-0.1206	1.8391
ROE	7.6195	3.3309	-0.400	13.900	7.900	-0.5802	2.6451
SP	187.0468	60.9048	98.280	328.440	173.610	0.6929	2.4874
B	4.6195	3.2522	1.000	11.800	3.800	0.8689	2.5609

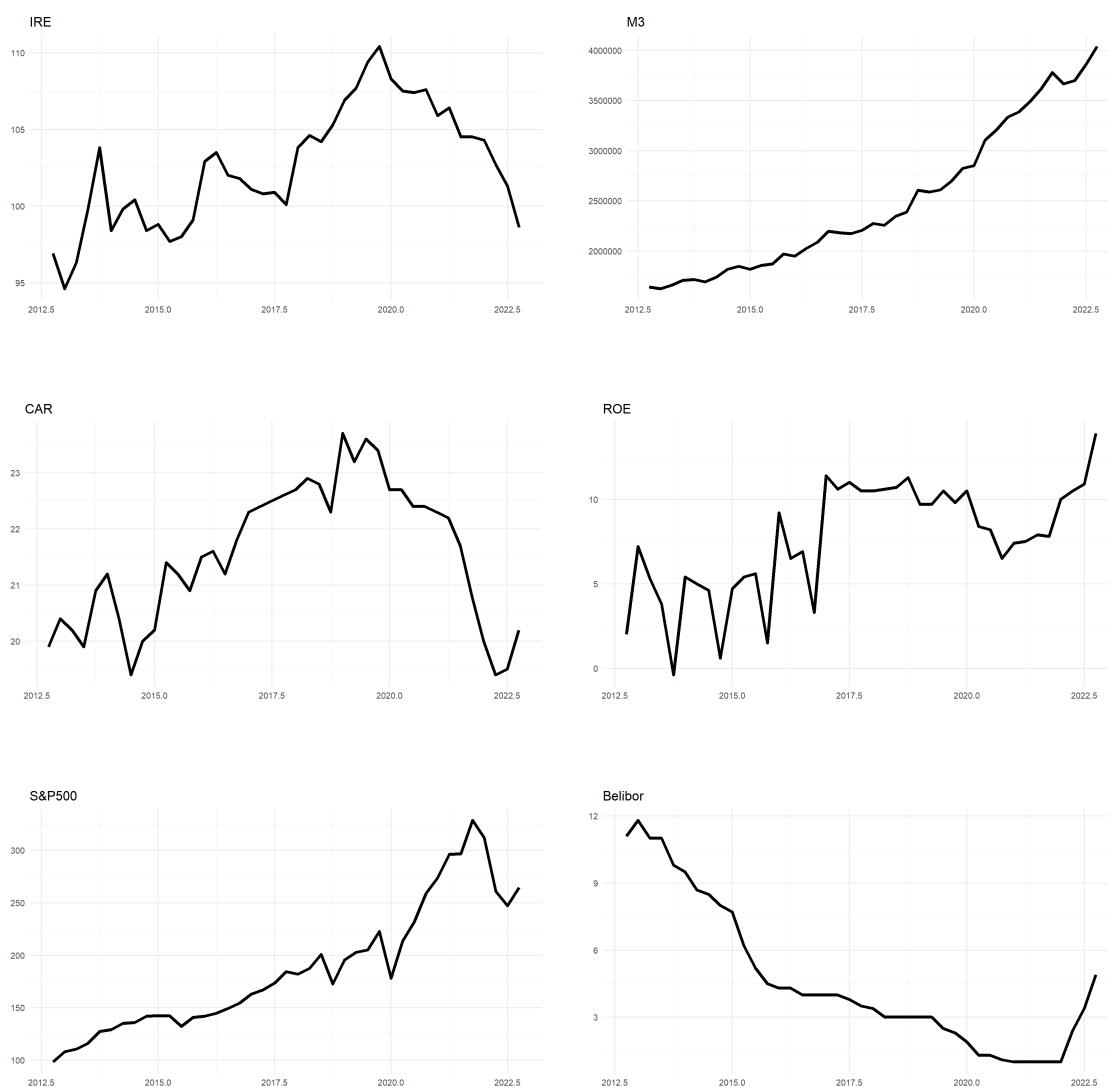
Табела 5.5: Дескриптивне статистике макроекономских показатеља

Из табеле 5.5 запажамо изразиту асиметрију временских серија референтне каматне стопе и Белибора удесно, те закључујемо да је десни реп ових расподјела тежи од репа нормалне расподјеле, одакле слиједи постојање структурног лома који је касније и приказан на слици 5.4. Средња вриједност девизног курса у односу на евро је благо асиметрична улијево, а индекс реалних нето зарада јако асиметричан удесно. Од јако спљоштених временских серија издвајају се стопа раста БДП-а и индекс реалних нето зарада.

ГЛАВА 5. СТОПА НЕИЗВРШЕЊА ОБАВЕЗА И МАКРОЕКОНОМСКИ ПРЕДИКТОРИ: СЛУЧАЈ БАНКАРСКОГ СЕКТОРА РЕПУБЛИКЕ СРБИЈЕ



ГЛАВА 5. СТОПА НЕИЗВРШЕЊА ОБАВЕЗА И МАКРОЕКОНОМСКИ ПРЕДИКТОРИ: СЛУЧАЈ БАНКАРСКОГ СЕКТОРА РЕПУБЛИКЕ СРБИЈЕ



Слика 5.4: Макроекономски показатељи (национални рачуни и банкарски сектор) у периоду од Т4 2012. до Т4 2022. године

ГЛАВА 5. СТОПА НЕИЗВРШЕЊА ОБАВЕЗА И МАКРОЕКОНОМСКИ ПРЕДИКТОРИ: СЛУЧАЈ БАНКАРСКОГ СЕКТОРА РЕПУБЛИКЕ СРБИЈЕ

Стационарност временских серија макроекономских показатеља испитујемо тестовима јединичног коријена: Дики-Фулеров тест (ADF), KPSS тест и Филипс-Перонов тест. У табели 5.6 приказани су резултати тест статистика и р-вриједности за сваки од наведених тестова, као и број потребних диференцирања.

	Дики-Фулер (ADF)		KPSS		Филипс-Перон		Бр. диф.
	Тест стат.	р	Тест стат.	р	Тест стат.	р	
NPL	-2.1305	0.5222	1.0147	<0.01	-5.8920	0.7605	I(1)
NPLC	-1.6864	0.6969	1.0137	<0.01	-7.0946	0.6843	I(1)
NPLR	-2.6918	0.3015	0.9680	<0.01	-5.1116	0.8100	I(1)
GDP	-5.2479	<0.01	0.2683	>0.10	-23.739	0.0134	I(0)
EUR	-2.4449	0.3986	0.1944	>0.10	-4.4653	0.8509	I(0)
USD	-2.6641	0.3123	0.3867	0.083	-8.4388	0.5992	I(1)
KIR	0.2950	>0.99	0.9185	<0.01	3.2342	>0.99	I(2)
IPC	-0.0401	>0.99	0.25	>0.10	-2.8532	0.9372	I(0)
IRE	-0.4359	0.9799	0.6983	0.014	-3.6925	0.8998	I(1)
M3	-0.9088	0.9398	1.0732	<0.01	-1.4395	0.9774	I(1)
CAR	-0.1986	>0.99	0.3616	0.094	-2.7385	0.9422	I(1)
ROE	-1.1378	0.9049	0.7674	<0.01	-26.506	<0.01	I(1)
SP	-2.7082	0.2950	1.0116	<0.01	-13.292	0.2918	I(1)
B	0.3833	>0.99	0.9247	<0.01	3.3325	>0.99	I(2)

Табела 5.6: Тестови јединичног коријена макроекономских показатеља

Када је временска серија стационарна са ауторегресионим параметром који је близак вриједности 1, тада се коришћењем ADF теста најчешће добија закључак да је серија нестационарна. Овај проблем се обично рјешава тако што се нултом хипотезом претпоставља стационарност временске серије. Управо на овај начин је дефинисан KPSS тест јединичног коријена.

Предвиђање стопе неизвршења

Како успјешност пословања банке непосредно зависи од способности предвиђања и квалификације ризика, то је од интереса исправно процијенити кредитну способност клијената. Интерни модели које банке користе по правилу дају оцјене вјероватноће неизвршења обавеза које обухватају читав пословни циклус. Међутим, за потребе примјене МСФИ 9 неопходне су оцјене вјероватноће неизвршења за конкретан временски тренутак, као и укључивање различитих макроекономских сценарија. У наставку анализирамо постојање везе између макроекономских показатеља и остварених стопа неизвршења у банкарском сектору Србије. Користимо неколико различитих модела: линеарна регресија (LR), метод случајних шума (RandomForest), метод градијентног појачавања (XGBoost) и рекурентне неуронске мреже (RNN). Мјере квалитета које смо користили су средња апсолутна процентуална грешка (енг. Mean Absolute Percentage Error - MAPE) као и њену симетричну верзију (енг. Symmetric Mean Absolute Percentage Error - SMAPE), средња апсолутна скалирана грешка (енг. Mean Absolute Scaled Error - MASE) и геометријска средња релативна апсолутна грешка (енг. Geometric Mean Relative Absolute Error - GMRAE). Подаци су подијељени тако да је однос података на тренинг и тест скупу једнак 70:30%, а подјела је извршена случајним одабиром. Кодови су писани у програмском језику R и Python.

Модел 1: Линеарна регресија

Модел 1.1

У складу са резултатима из табеле 5.3 и 5.6, промјенљиве диференцирамо назначени број пута. Тестови јединичног коријена над диференцама стопе неизвршења и макроекономских показатеља одбацују хипотезу нестационарности на свим релевантним нивоима значајности за овако формиране временске серије. Стога посматрамо модел линеарне регресије дат формулом:

$$\Delta DR_t = \beta_0 + \sum_{k: X_{k,t} \in \mathcal{D}_0} \beta_k X_{k,t} + \sum_{j: X_{j,t} \in \mathcal{D}_1} \beta_j \Delta X_{j,t} + \sum_{s: X_{s,t} \in \mathcal{D}_2} \beta_s \Delta^2 X_{s,t} + \varepsilon_t,$$

гдје су $X_{k,t}$ различити макроекономски показатељи, а $\mathcal{D}_0 = \{\text{GDP, EUR, IPC}\}$, $\mathcal{D}_1 = \{\text{NPL, NPLC, NPLR, USD, IRE, M3, CAR, ROE, SP}\}$ и $\mathcal{D}_2 = \{\text{KIR, B}\}$, тј. \mathcal{D}_i је скуп макроекономских показатеља које диференцирамо i пута.

Модел 1.2

Како је стопа неизвршења ограничена вриједност између 0 и 1, алтернатива која је предложена у [17] јесте употреба логистичке трансформације облика:

$$y_t = \ln \left(\frac{DR_t}{1 - DR_t} \right).$$

Користећи ову трансформацију, могуће је замијенити претходни модел линеарном регресијом облика:

$$\Delta y_t = \beta_0 + \sum_{k: X_{k,t} \in \mathcal{D}_0} \beta_k X_{k,t} + \sum_{j: X_{j,t} \in \mathcal{D}_1} \beta_j \Delta X_{j,t} + \sum_{s: X_{s,t} \in \mathcal{D}_2} \beta_s \Delta^2 X_{s,t} + \varepsilon_t,$$

Модел 1.3

Комбиноваћемо претходно описане моделе уз додатне трансформације макроекономских показатеља. Модел формирамо одбацивањем предиктора који нису значајни и који су високо корелисани са преосталим. Посматрамо модел линеарне регресије облика:

$$\begin{aligned} \Delta \log DR_t = & \beta_0 + \beta_1 \Delta \log EUR_t + \beta_2 \Delta NPLR_t \\ & + \beta_3 \Delta \log IRE_t + \beta_4 \Delta \log M3_t + \varepsilon_t \end{aligned}$$

У табели 5.7 приказане су вриједности мјера квалитета за сваки од претходно описаних модела, на основу којих можемо закључити да највећу предиктивну моћ има модел 1.3.

	LR 1.1	LR 1.2	LR 1.3
MAPE	1.935020	1.865317	2.071217
SMAPE	1.230320	1.234320	1.068239
MASE	0.724361	0.733335	0.671445
GMRAE	0.937366	1.022032	0.891093

Табела 5.7: Резултати регресионих модела

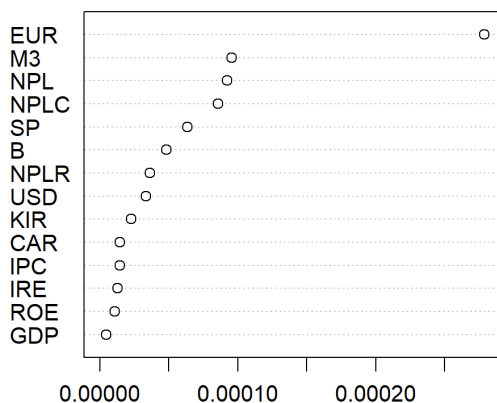
Модел 2: Случајна шума

Користећи метод стабала одлучивања који се заснива на регресији формирамо случајну шуму. Монте Карло симулацијама за $N = 10.000$ бирамо оптималан број стабала. У свакој симулацији на основу средње апсолутне скалиране грешке (MASE) бирамо оптималан број стабала поредећи грешке модела са бројем стабал од 1 до 200, редом.

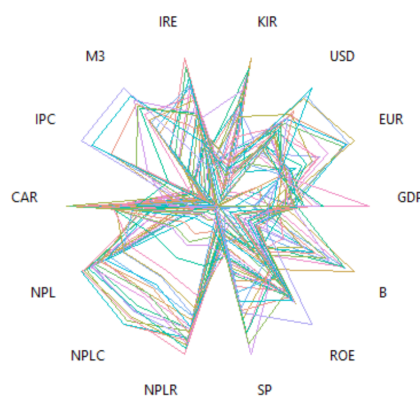
У табели 5.8 приказане су вриједности мјера квалитета на тест скупу за најоптималнији ($k_1 = 36$) и најфреквентнији ($k_2 = 8$) број стабала добијен симулацијама. На слици 5.5 је приказана значајност сваког од макроекономских показатеља, а на слици 5.6 радијални координатни дијаграм предиктора.

	RF 2.1	RF 2.2
MAPE	0.0563855	0.0508326
SMAPE	0.0558130	0.0518419
MASE	0.3890906	0.3743140
GMRAE	0.0337626	0.0370701

Табела 5.8: Резултати модела случајних шума на тест скупу



Слика 5.5: Значајност предиктора према Тинијевом индексу



Слика 5.6: Радијални координатни дијаграм предиктора

Модел 3: Градијентно појачавање

Слично као за методу случајних шума, симулацијама оцјењујемо вриједности параметара за градијентно појачавање: број стабала, максимална дубина сваког стабла и стопа учења.

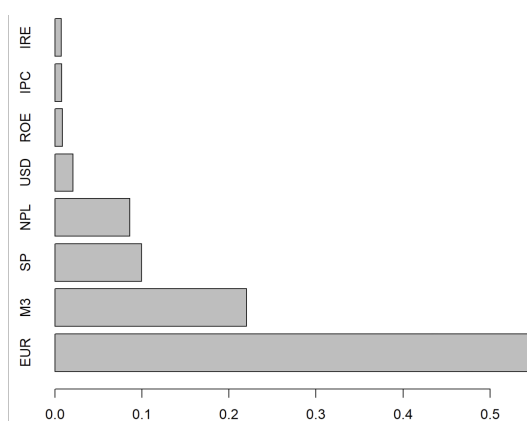
У табели 5.9 приказани су резултати на тест скупу за моделе:

- XGB 3.1: број стабала 91, максимална дубина 4, стопа учења 0.08,
- XGB 3.2: број стабала 103, максимална дубина 3, стопа учења 0.07,

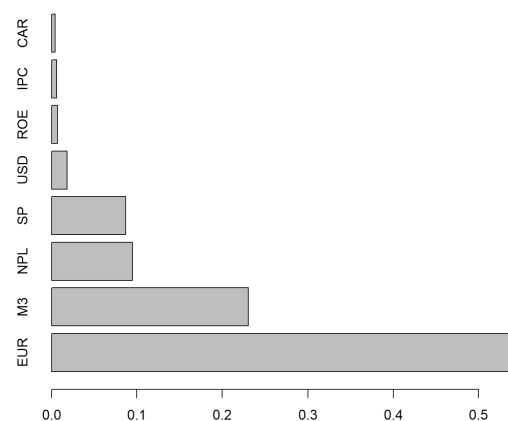
гдје су одабране најоптималније вриједности параметара за број стабала из интервала [1,200], максималну дубину сваког стабла [1, 10] и стопу учења [0.01, 0.15] тако да средња апсолутна скалирана грешка (MASE) буде најмања. На сликама 5.7 и 5.8 приказана је значајност предиктора у оба одабрана модела према Ђинијевом индексу.

	XGB 3.1	XGB 3.2
MAPE	0.0343937	0.0344287
SMAPE	0.0344795	0.0344941
MASE	0.3365120	0.3344056
GMRAE	0.0084191	0.0165691

Табела 5.9: Резултати модела градијентног појачавања на тест скупу



Слика 5.7: Значајност предиктора у моделу XGB 3.1



Слика 5.8: Значајност предиктора у моделу XGB 3.2

Модел 4: Рекурентне неуронске мреже

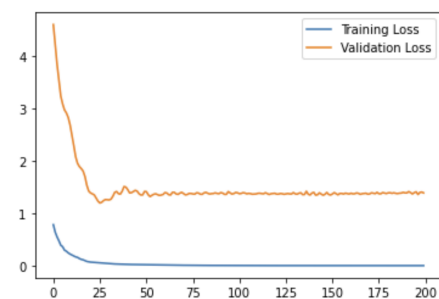
Разматрамо два модела рекурентних неуронских мрежа и то:

- RNN 4.1: рекурентна неуронска мрежа са два скривена слоја. У првом слоју има 32 неурона, а у другом 16 неурона. Мрежа је тренирана у 200 епоха, а за оптимизацију је коришћен оптимизатор Adam са активационом функцијом $\sigma(v) = \tanh(v)$,
- RNN 4.2: рекурентна неуронска мрежа са два скривена слоја. У првом слоју има 8 неурона, а у другом 4 неурона. Мрежа је тренирана у 150 епоха, а за оптимизацију је коришћен оптимизатор Adam са активационом функцијом $\sigma(v) = \tanh(v)$.

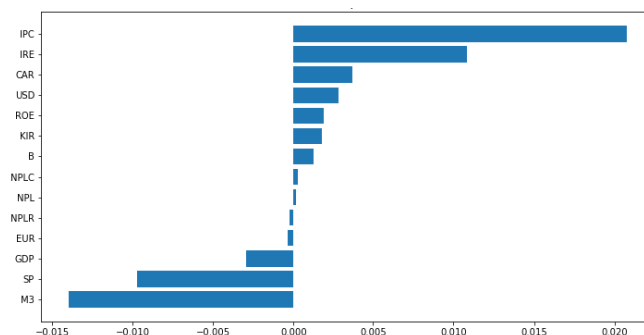
У табели 5.10 приказане су вриједности мјера квалитета за претходно описане моделе. По форми добијених вриједности функција губитака приказаних на слици 5.9 примјећујемо да мрежа учи иако би за ефикасније обучавање био неопходан дужи временски период серија, а на слици 5.10 су издвојени најзначајнији предиктори према Ђинијевом индексу за модел RNN 4.1.

	RNN 4.1	RNN 4.2
MAPE	0.0910836	0.1049363
SMAPE	0.0959418	0.1117815
MASE	0.6427240	0.6616280
GMRAE	0.0392599	0.0508291

Табела 5.10: Резултати модела рекурентне неуронске мреже на тест скупу



Слика 5.9: Функција губитака на тренинг и валидационом скупу у моделу RNN 4.1



Слика 5.10: Значајност предиктора према важности пермутације у моделу RNN 4.1

Зкључна разматрања о моделима

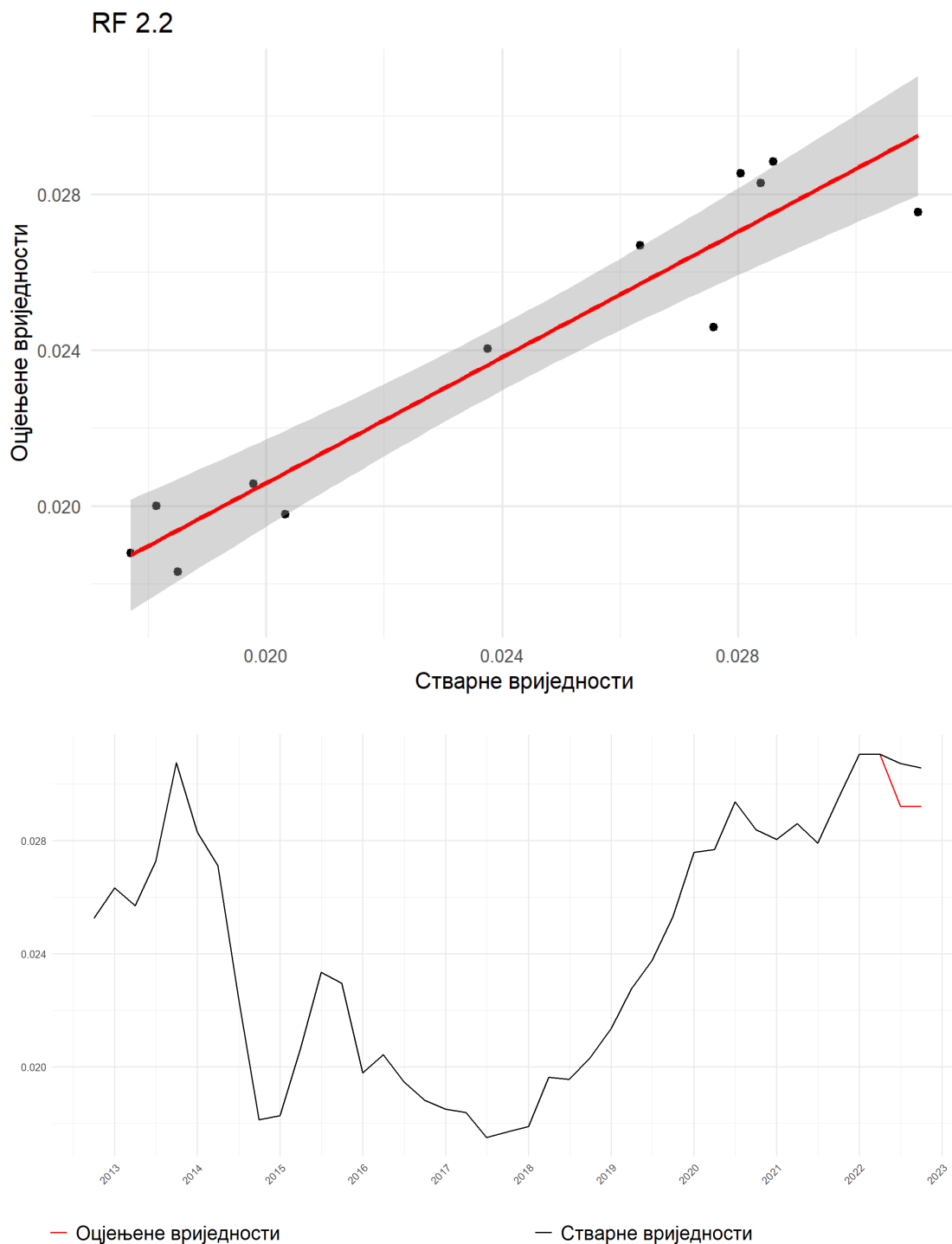
У табели 5.11 приказани су резултати добијени у свим претходно описаним моделима. Можемо примијетити да модел градијентног појачавања даје најбоље резултате, као и да су сви модели статистичког учења ефикаснији у односу на постојеће интерне моделе засноване на линеарној регресији. Од најзначајнијих макроекономских показатеља издвајају се девизни курс динара према еврџу, монетарни агрегат МЗ и учешће проблематичних кредита у банкарском сектору. Сви наведени фактори су у складу са економском интуицијом и макроекономским кретањима.

	LR 1.3	RF 2.1	RF 2.2	XGB 3.1	XGB 3.2	RNN 4.1	RNN 4.2
MAPE	2.07122	0.05639	0.05083	0.03439	0.03443	0.09108	0.10494
SMAPE	1.06824	0.05581	0.05184	0.03448	0.03449	0.09594	0.11178
MASE	0.67145	0.38909	0.37431	0.33651	0.33441	0.64272	0.66163
GMRAE	0.89109	0.03376	0.03707	0.00842	0.01657	0.03926	0.05083

Табела 5.11: Резултати модела статистичког учења

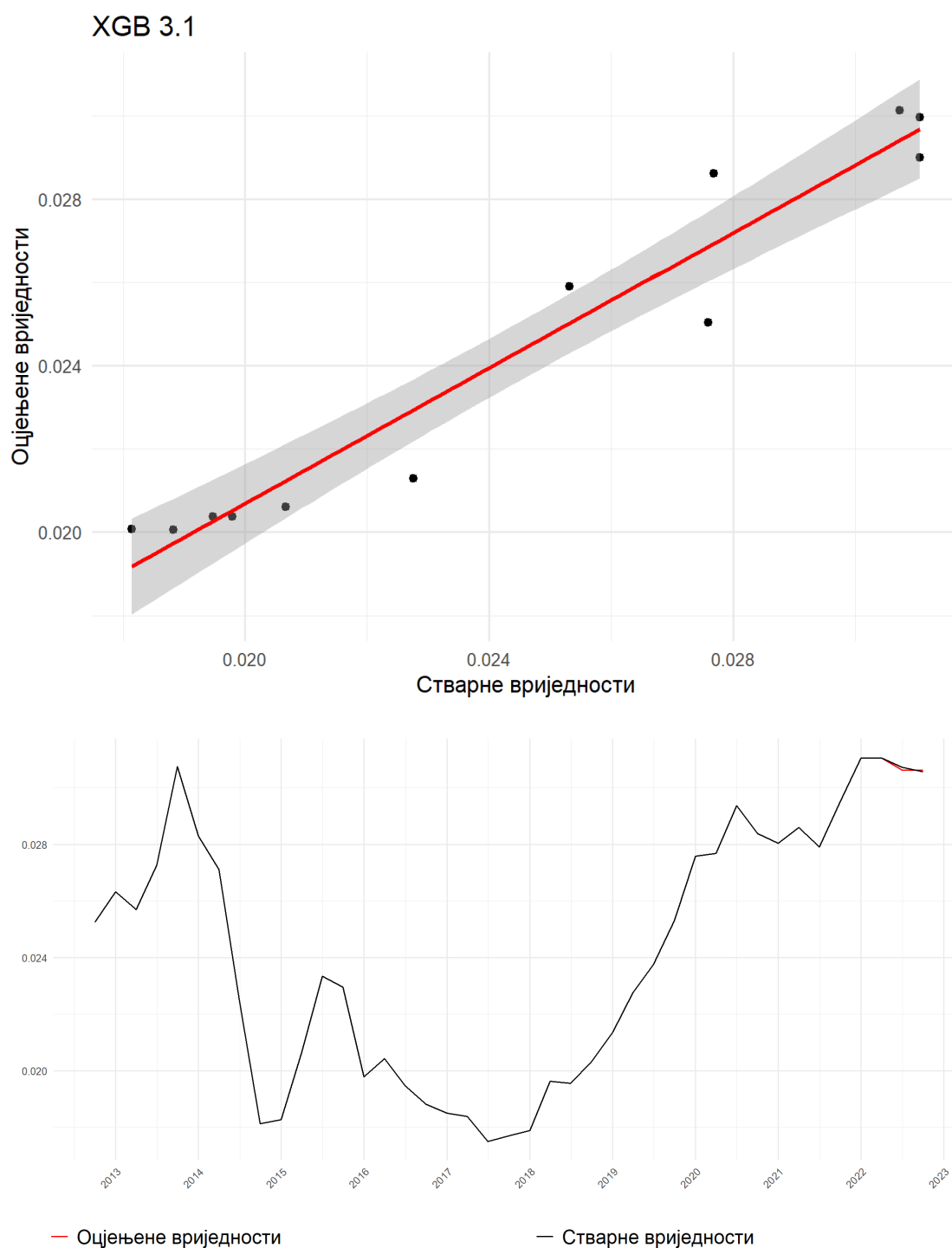
С обзиром да је привреда Републике Србије оптерећена високим степеном евроизације, то је девизни курс динара према еврџу један од главних канала трансмисије монетарне политике и као такав је статистички значајна детерминанта кредитног ризика. Индекс реалних нето зарада повећава агрегатну тражњу, што утиче на раст инвестиционе и привредне активности и коначно има резултат на смањење проблематичних кредита. Увећање монетарног агрегата МЗ повећава расположивост извора финансирања, што са друге стране доводи до алоцирања тих средстава на нове пласмане, потенцијално на ефикаснији начин него што је то био случај са претходним.

На сликама 5.11 и 5.12 приказано је предвиђање кретања стопе неизвршења обавеза за два тромјесечја примјеном модела случајних шума RF 2.2 и градијентног појачавања XGB 3.1, редом. Како ова два модела не узимају у обзир временску компоненту посматраних серија, за предвиђање будућих вриједности стопе неизвршења неопходно је имати вриједности макроекономских показатеља за период предвиђања, уколико то није случај исти се могу оцијенити неким познатим методама, а након тога извршити предвиђање стопе неизвршења.



Слика 5.11: Предвиђање кретања стопе неизвршења обавеза за два тромјесечја примјеном модела случајних шума RF 2.2

ГЛАВА 5. СТОПА НЕИЗВРШЕЊА ОБАВЕЗА И МАКРОЕКОНОМСКИ ПРЕДИКТОРИ: СЛУЧАЈ БАНКАРСКОГ СЕКТОРА РЕПУБЛИКЕ СРБИЈЕ



Слика 5.12: Предвиђање кретања стопе неизвршења обавеза за два тромјесечја примјеном модела градијентног појачавања XGB 3.1

Глава 6

Закључак

У овом мастер раду анализирали смо да ли постоји веза између стопа неизвршења обавеза у банкарском сектору Републике Србије и макроекономских показатеља користећи агрегиране податке о неизвршењима Удружења банака Србије. Излагање је било хронолошки, од увођења потребне теоријске основе до презентовања резултата са циљем да представимо значај управљања кредитним ризиком и дођемо до нових начина оцјењивања вјероватноћа неизвршења обавеза, заснованих на текућим макроекономским кретањима, а који не зависе од пословног циклуса банака. Примењено је неколико различитих приступа за провјеру везе између стопа неизвршења и макроекономских показатеља: линеарна регресија, случајне шуме, градијентно појачавање и рекурентне неуронске мреже. Показали смо да методе статистичког учења постижу значајне резултате у побољшању тачности оцјена у односу на моделе линеарне регресије. Штавише, сви уочени макроекономски показатељи у моделима у складу су са економском интуицијом и макроекономским кретањима. Важно је напоменути да добијени резултати могу представљати значајну информацију за институције и регулаторна тијела надлежна за јачање стабилности финансијског система, будући да модели представљени у раду дају статистички значајно боље резултате за процјену кредитног ризика у односу на стандардне регресионе моделе. Предмет даљег истраживања могле би бити друге софистицираније методе статистичког учења са додатним нумеричким оптимизацијама, као и анализа кредитног ризика и макроекономских кретања на нивоу више упоредивих земаља у развоју.

Библиографија

- [1] Bank for International Settlement Basel Committee on Banking Supervision. Sound Practices for Loan Accounting and Disclosure. *Basel Committee on Banking Supervision Publications*, 1996.
- [2] Bank for International Settlement Basel Committee on Banking Supervision. International Convergence on Capital Measurement and Capital Standards. *Basel Committee on Banking Supervision Publications*, 2004.
- [3] Bank for International Settlement Basel Committee on Banking Supervision. Strengthening the Resilience of the Banking Sector. *Basel Committee on Banking Supervision Publications*, 2009.
- [4] International Accounting Standards Board. IFRS 9 Financial Instruments. 2014.
- [5] Leo Breiman, Jerome Friedman, Charles J. Stone, and R.A. Olshen. Classification and Regression Trees. 1984.
- [6] Greg M. Gupton, Christopher C. Finger, and Mickey Bhatia. CreditMetrics - Technical Document. 1997.
- [7] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2009.
- [8] Lars Meyer and Thomas Pöhlmann. Challenges of IFRS 9 Impairment: Requirement to prepare early for the new impairment approach. 2014.
- [9] Tom M. Mitchell. Machine Learning. 1997.
- [10] Anthony Saunders and Linda Allen. Credit Risk Measurement-New Approaches to Value at Risk and Other Paradigms. *John Wiley & Sons*, 2002.

- [11] Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, and Vipin Kumar. Introduction To Data Mining. 2005.
- [12] Terry M. Therneau, Elizabeth J. Atkinson, and Mayo Foundation. An Introduction to Recursive Partitioning Using the RPART Routines. 2022.
- [13] Alfiya Vasilyeva and Elvina Frolova. Methods of Calculation of Expected Credit Losses Under Requirements of IFRS 9. *Journal of Corporate Finance Research*, 2019.
- [14] Милица Јанковић. Параметри кредитног ризика. *Универзитет у Нишу, Природно-математички факултет*, 2018.
- [15] Жељко М. Јовић. Анализа детерминанти кредитног ризика у условима изражене информационе асиметрије: пример банкарског сектора Републике Србије. *Универзитет у Београду, Економски факултет*, 2017.
- [16] Жељко Рачић и Лидија Барјактаровић. Анализа емпиријских детерминанти кредитног ризика у банкарском сектору Републике Србије. *Банкарство, Удружење банака Србије*, 2016.
- [17] Милош Божовић. Постоје ли макроекономски предиктори за point-in-time PD? Резултати на основу базе података стопа неизмирења Удружења банака Србије. *Банкарство, Удружење банака Србије*, 2019.
- [18] Милан Брковић. Имплементација IFRS 9 у банкама и макроекономски сценарији - неки методолошки аспекти. *Банкарство, Удружење банака Србије*, 2017.
- [19] Милош Б. Вујновић. Валидација модела кредитног ризика. *Универзитет у Београду, Факултет организационих наука*, 2016.
- [20] Татјана Драшковић. Процес конвергенције регулаторних захтева и рачуноводствених стандарда. *Банкар, Удружење банака Црне Горе*, 2017.
- [21] Светлана Ј. Дрљача. Анализа системске компоненте кредитног ризика у банкарском сектору Републике Србије. *Универзитет у Београду, Економски факултет*, 2022.
- [22] Младен Николић и Анђелка Зечевић. Машинско учење. *Универзитет у Београду, Математички факултет*, 2019.

- [23] Дарко Будиша и Марко Тешић и Марко Крајишник. Главне одредбе примјене МСФИ 9 и утицај на ниво кредитних губитака банака. *Зборник радова ЕконБиз*, 2018.
- [24] Слободанка Јанковић и Бојана Милошевић. Елементи финансијске математике. *Универзитет у Београду, Математички факултет*, 2017.
- [25] Реамаре Новко. Кредитни ризик. *Свеучилиште у Загребу, Природословно-математички факултет*, 2017.
- [26] Аднан Ровчанин. Управљање финансијама. 2010.
- [27] Народна банка Србије. Извештај о резултатима анкете о кредитној активности - прво тромесечје 2015.
- [28] Народна банка Србије. Стратегија за увођење стандарда Базел III у Србији. 2013.

Биографија аутора

Стефан Малбашић рођен је 25. децембра 1998. године. Основну школу и Гимназију завршио је у Источном Сарајеву као ученик генерације и носилац дипломе “Вук Караџић” за постигнути успјех. Дипломирао је на Математичком факултету Универзитета у Београду 2022. године на смјеру Статистика, актуарска и финансијска математика. Исте године уписао је мастер академске студије и изабран је за сарадника у настави на Катедри за вероватноћу и статистику. Држао је вјежбе из сљедећих предмета: Елементи актуарске математике, Статистички софтвер 3, Вероватноћа и статистика А, Вероватноћа и статистика Б, Временске серије и примене у финансијама и Увод у теорију узорака. Године 2022. изабран је за студента продекана Математичког факултета, потпредседника за наставу Студентског парламента Универзитета у Београду и члана скупштине Студентске конференције универзитета Србије „СКОНУС”. Члан је Сената Универзитета у Београду и један од оснивача Студентске уније математичара “СУМА”. Стипендиста је Фонда за младе таленте - Доситеја у оквиру Министарства науке, технолошког развоја и иновација Републике Србије за најбоље студенте за академску 2021/22. и 2022/23. годину.