

# Institucionalne odrednice transparentnosti državnih proračuna

---

**Kovačević, Marija Pia**

**Master's thesis / Diplomski rad**

**2022**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:217:243085>

*Rights / Prava:* [In copyright](#)/[Zaštićeno autorskim pravom.](#)

*Download date / Datum preuzimanja:* **2024-10-09**



*Repository / Repozitorij:*

[Repository of the Faculty of Science - University of Zagreb](#)



**SVEUČILIŠTE U ZAGREBU**  
**PRIRODOSLOVNO–MATEMATIČKI FAKULTET**  
**MATEMATIČKI ODSJEK**

Marija Pia Kovačević

**INSTITUCIONALNE ODREDNICE**  
**TRANSPARENTNOSTI DRŽAVNIH**  
**PRORAČUNA**

Diplomski rad

Voditeljica rada:  
prof. dr. sc. Katarina Ott

Zagreb, 2022.

Ovaj diplomski rad obranjen je dana \_\_\_\_\_ pred ispitnim povjerenstvom u sastavu:

1. \_\_\_\_\_, predsjednik
2. \_\_\_\_\_, član
3. \_\_\_\_\_, član

Povjerenstvo je rad ocijenilo ocjenom \_\_\_\_\_.

Potpisi članova povjerenstva:

1. \_\_\_\_\_
2. \_\_\_\_\_
3. \_\_\_\_\_

*Ovaj diplomski rad želim posvetiti svojoj obitelji, posebno roditeljima Mariji i Juraju bez čije žrtve i ljubavi ovo ne bi bilo moguće. Zahvaljujem mentorici prof. dr. sc. Katarini Ott na strpljenju, pomoći te vodstvu pri izradi rada. Hvala i kolegama s faksa zbog čije potpore i nesebičnosti sam stigla do završetka te koji su postali prijatelji na koje ću se uvijek moći osloniti.*

# Sadržaj

<b>Sadržaj</b>	<b>iv</b>
<b>Uvod</b>	<b>2</b>
<b>1 Definicije i motivacija</b>	<b>3</b>
1.1 Definicija transparentnosti državnog proračuna . . . . .	3
1.2 Definicija i mjerenje institucionalnosti . . . . .	4
1.3 Indeks otvorenosti poraćuna . . . . .	7
<b>2 Podaci i hipoteze</b>	<b>10</b>
2.1 Podaci . . . . .	10
2.2 Hipoteze . . . . .	13
2.3 Deskriptivna statistika . . . . .	15
<b>3 Opis metodologije</b>	<b>19</b>
3.1 Linearna regresija . . . . .	19
3.2 Jednostavna linearna regresija . . . . .	20
3.3 Višestruka linearna regresija . . . . .	24
3.4 Određivanje prikladnog modela . . . . .	27
<b>4 Empirijska analiza</b>	<b>32</b>
4.1 Testiranje hipoteza . . . . .	32
4.2 Interpretacija rezultata . . . . .	54
<b>5 Zaključak</b>	<b>56</b>
<b>Bibliografija</b>	<b>57</b>
<b>A Usporedba WGI i Avg_WGI</b>	<b>64</b>

# Uvod

Proračunski sustav i javni proračunski proces ogledalo su djelovanja vlasti svake zemlje. Transparentnost proračuna doprinosi političkom procesu i učinkovitosti vlade te time predstavlja dobru mjeru uspješnog provođenja vlasti. S ciljem mjerenja transparentnosti, International Budget Partnership 2006. počinje provođenje ankete kojom, među ostalim, mjeri i transparentnost. Uspješno provođenje vlasti ne utječe samo na ekonomski razvoj zemlje i njeno gospodarsko poboljšanje, već i na razvoj njenog društva, veće povjerenje javnosti te iskorjenjivanje korupcije. Vladanje uvelike utječe na društvo te za dobrobit svih trebamo težiti što boljem upravljanju. Kao indikatore uspješnosti upravljanja i vladanja promatramo šest svjetskih pokazatelja upravljanja koji opisuju glas i odgovornost, političku stabilnost i odsutnost nasilja/terorizma, učinkovitost vlade, kvalitetu regulatornog okvira, vladavinu prava te kontrolu korupcije. U ovom radu promatramo vezu između transparentnosti državnog proračuna i svjetskih pokazatelja upravljanja.

U prvom poglavlju detaljnije opisujemo način mjerenja transparentnosti državnog proračuna i institucionalnosti. Kao mjeru transparentnosti gledamo indeks otvorenosti proračuna (OBI, eng. *Open Budget Index*), a institucionalnost mjerimo svjetskim pokazateljima upravljanja (WGI, eng. *Worldwide Governance Indicators*). Kao uzorak uzimamo 117 zemalja kojima je 2019. izračunat OBI te za iste zemlje imamo dostupne WGI pokazatelje sve od 2017. do 2019. Spominjemo i posljednje objavljen OBI iz 2021. te promatramo koliko se razlikuje od prethodnog rezultata iz 2019.

Hipoteze navedene u drugom poglavlju postavljene su na temelju ostalih istraživačkih radova kao u [7], [32] i [33]. Provjeravamo jesu li svi svjetski pokazatelji upravljanja statistički značajni za indeks otvorenosti proračuna te utječu li pozitivno na njega. Umjesto promatranja šest zasebnih pokazatelja uvodimo pojam agregiranog pokazatelja upravljanja kao sažetog indikatora kvalitete institucije. Linearnom regresijom ćemo provjeriti odnos između transparentnosti proračuna i svjetskih pokazatelja upravljanja.

U trećem poglavlju opisujemo metodologiju korištenu prilikom testiranja hipoteza. Opisujemo metodu jednostavne i višestruke linearne regresije te način procjene parametara danih modela.

Na kraju navodimo zaključke donesene na temelju rezultata prezentiranih u četvrtom poglavlju. Zaključujemo da svi WGI pokazatelji statistički značajno i pozitivno utječu na

OBI. Također, svi WGI pokazatelji su međusobno jako korelirani. Agregirani svjetski pokazatelj upravljanja prati isti rezultat zasebnih pokazatelja, to jest pozitivno utječe na OBI. Dakle, ovim radom pokazuje da upravljanje zaista utječe na transparentnost državnog proračuna.

# Poglavlje 1

## Definicije i motivacija

### 1.1 Definicija transparentnosti državnog proračuna

Državni proračun ukazuje na djelotvornost te uspješnost rada i kvalitete institucija, samog vodstva države te njen ekonomski rast. Preciznije, za određivanje učinkovitosti vlade je bitna transparentnost državnog proračuna. Organizacija za ekonomsku suradnju i razvoj (OECD, engl. *Organisation for Economic Cooperation and Development*) transparentnost definira kao potpuno objavljivanje svih relevantnih fiskalnih informacija u pravodobnom i sistematičnom obliku [35]. Transparentnost omogućuje građanima uvid u potpune, točne, pravovremene i razumljive informacije o proračunu kako bi oni mogli vlastitim angažmanom utjecati na efikasnost prikupljanja javnih sredstava i ponude javnih dobara i usluga, na povećanje odgovornosti regionalnih/lokalnih vlasti i smanjenje mogućnosti za korupciju [21].

Iako je gore navedena definicija najčešće korištena u praksi, postoje još neke definicije otvorenosti državnih proračuna i procesa. Primjerice, Premchad [36] definira transparentnost ili otvorenost kao dostupnost informacija javnosti o transakcijama vlada i transparentnost procesa donošenja odluka.

Bitno je spomenuti i fiskalnu transparentnost, pojam koji je nešto širi od proračunske transparentnosti. Prema Međunarodnom monetarnom fondu (MMF, eng. *International Monetary Fund*), fiskalna transparentnost odnosi se na informacije dostupne javnosti o vladinom procesu donošenja fiskalne politike. Odnosi se na jasnoću, pouzdanost, učestalost, pravodobnost i relevantnost javnog fiskalnog izvješćivanja i otvorenost takvih informacija [23]. Transparentnost pruža građanima uvid u vladine proračune i pomaže im da ih drže odgovornima. Pomaže u izgradnji povjerenja i održivosti tržišta. Fiskalnu transparentnost spominju i Alt i Lassen [8] analizirajući efekte fiskalne transparentnosti na akumuliranost državnog duga te kako institucionalni faktori utječu na ovaj tip transparentnosti. OECD i MMF su osmislili smjernice za obje transparentnosti (proračunsku i fiskalnu redom).



OECD je na temelju iskustava zemalja članica sastavio smjernice za središnju vlast te su one podijeljene na tri dijela. Prvi dio navodi osnovne proračunske izvještaje i njihov općeniti sadržaj kao upute vladama. Naglasak je na samom proračunu, prijedlogu proračuna, mjesečnim i polugodišnjim izvješćima te na godišnjem obračunu proračuna. Drugi dio opisuje specifične podatke koje bi ova izvješća trebala sadržavati. Primjerice, detaljni podaci o javnom dugu, financijskoj i nefinancijskoj imovini, mirovinskim doprinosima i sl. U trećem dijelu su istaknuti postupci za osiguranje kvalitete i integriteta izvješća. Posljednji dio se odnosi na računovodstvene standarde, internu financijsku kontrolu, reviziju i kontrolu koju provodi parlament.[11]

MMF-ove smjernice pokrivaju četiri područja: (1) jasnoću uloga i dužnosti državnoga i javnog sektora; (2) otvorenost proračunskog procesa; (3) javnu raspoloživost informacija; (4) jamčenje integriteta izvješća. [11]. Smjernice su osmišljene kao referentni alat za zemlje članice i nečlanice kako bi povećale stupanj transparentnosti proračuna u svojim zemljama. Organizirane su oko specifičnih izvješća samo kao primjer te se podrazumijeva da će različite zemlje imati različite načine izvješćivanja. Samim time mogu imati i drugo područje kojim bi istakle razinu transparentnosti. Treba naglasiti da smjernice ne predstavljaju službeni "standard" za proračunsku transparentnost [35].

Državni proračun je korišten kao bitan element upravljanja zemljom i kao uspješan pokazatelj o njenom stanju. Transparentan proračun ukazuje na uključivanje javnosti u proces vladanja što produbljuje povjerenje javnosti i promiče pravedniju politiku vladanja.

## 1.2 Definicija i mjerenje institucionalnosti

### *Good Governance*

Prema radovima Acoste [30], Ellisa i Fendera [20], javni proračun je motor koji pokreće gospodarstvo i rad vlade. Struktura proračuna, proračunski proces i način na koji se izvršava igraju značajnu ulogu u gospodarskom rastu i održivom razvoju. Samim time, transparentnost proračuna je povezana s dobrim vladanjem i kvalitetom institucija.

Kao i za transparentnost proračuna, ne postoji univerzalna definicija dobrog upravljanja i vladanja. Prema Neumayeru [34], upravljanje (eng. *governance*) je način na koji su politički protagonisti i akteri ovlašteni donositi odluke, način na koji se političke odluke formuliraju i provode te stupanj u kojem je vladi u nekom obliku inetrvcije dopušteno zadirati u prava građana. Program Ujedinjenih naroda za razvoj (UNDP, eng. *United Nations Development Programme*) definira upravljanje (ili vladanje) kao vršenje ekonomske, političke i administrativne ovlasti za upravljanje poslovima zemlje na svim razinama. Sastoji se od mehanizama, procesa i institucija kroz koje građani i skupine artikuliraju svoje

interese, ostvaruju svoja zakonska prava, ispunjavaju svoje obveze i posreduju u njihovim razlikama [38].

Intuitivno, možemo pretpostaviti da za sam ekonomski rast i razvoj ne možemo imati bilo kakvo upravljanje, ono bi trebalo biti donekle dobro i uspješno. Sam pojam dobrog upravljanja (eng. *good governance*) spominje se kao ključna stavka za gospodarski razvoj i smanjenje siromaštva u mnogim radovima. Neke međunarodne financijske institucije kao MMF i Svjetska banka te zemlje donatori (npr. Sjedinjene Države i Ujedinjeno Kraljevstvo), koriste dobro upravljanje kao standard za ocjenjivanje poslova i sustava ostalih zemalja.

Dobro upravljanje omogućuje bolje usluge za građane, političku stabilnost i učinkovitost vlade. Nadalje, povezano je s borbom protiv korupcije i pozivanjem birokrata i političara na odgovornost. Slijedom toga, dobre prakse upravljanja preduvjet su financijske i nefinancijske pomoći donatora zemljama kojima je potrebna pomoć[7].

Osim toga, međunarodne organizacije smatraju da je dobro upravljanje uvjet za gospodarski razvoj i napore u borbi protiv korupcije. Ujedinjeni narodi su uveli osam glavnih karakteristika dobrog upravljanja koje definiraju i artikuliraju prakse dobrog upravljanja: „[dobro upravljanje je] participativno, orijentirano na konsenzus, odgovorno, transparentno, reaktivno (eng. *responsive*), uspješno i učinkovito, pravedno i uključivo, te slijedi zakonske propise” [39]. Također, dobro upravljanje opisuju Kosack i Fung [29] te Denhardt i Denhardt [19] kao ono koje poštuje ljudska prava i usvaja demokratska načela (npr. sudjelovanje građana i transparentnost) u svojim procesima donošenja odluka.

U nastavku rada ćemo kao vodeću definiciju upravljanja koristiti onu koju su dali Kaufmann, Kraay i Zoido-Lobaton u svom radu [27]. Oni definiraju upravljanje kao tradicije i institucije pomoću kojih se provodi vlast u državi. To uključuje (1) procese kojima se vlada izabire, nadzire i mijenja, (2) kapacitet vlade da učinkovito formulira i primijeni zdrave politike te (3) poštovanje građana i države prema institucijama koje upravljaju ekonomskim i društvenim odnosima među njima samima.

## **Svjetski pokazatelji upravljanja**

Iako postoje razni pokazatelji upravljanja, u nastavku rada koristit ćemo svjetske pokazatelje upravljanja (WGI, eng. *Worldwide Governance Indicators*) koje su razvili Kaufmann i Kraay 1999.. Svojim radom su željeli prikupiti skup pokazatelja koji mjere subjektivne percepcije o kvaliteti upravljanja u različitim zemljama [27]. Za velike doprinose, razvoj i poboljšanje samih pokazatelja zaslužni su i Pablo Zoido te Massimo Mastruzzi. U radu u kojem opovrgavaju kritike usmjerene prema WGI-jevima autori spominju da njihove pokazatelje često koriste i političari i znanstvenici [24].

Njihove tvrdnje potvrđuju i Andreula, Chong i Guillen u svom radu gdje za pokazatelje kažu da su pružili novi način kombiniranja usporedivih indeksa uz različite dimenzije

upravljanja. Također ukazuju na to da uključuju više zemalja i podataka nego što je to prije bilo dostupno. Kod kreiranja pokazatelja su i jasno naznačili temeljne izvore podataka te kako su podaci konstruirani, eksplicitno izvještavajući o rasponima pogrešaka za sve procjene. Ovo je bila velika inovacija i dalje ostaje kao bitna stavka [9].

Arndt i Oman kažu da, iako savršeni pokazatelj upravljanja vjerojatno nikada neće postojati, neki pokazatelji imaju veću valjanost i vjerodostojnost od drugih. Broj izvora korištenih za strukturiranje indeksa, sveobuhvatnost kojom je pokriven proces upravljanja i točnost rezultata su čimbenici koji među korisnicima čine jedan indeks vjerodostojnijim od drugih. Iako za njih imaju i kritike, u svom radu navode da svjetski pokazatelji upravljanja pokrivaju najvažnije aspekte samog procesa koji proučavaju [10].

Kako su pokazatelji korišteni i u brojnim drugim radovima te sadrže informacije još od 1996., logična su polazna točka i indikator dobrog ili lošeg upravljanja.

WGI su složeni pokazatelji upravljanja temeljeni na više od 30 temeljnih izvora podataka, koji se mijenjaju i kombiniraju kako bi se stvorilo šest agregatnih pokazatelja. Sama činjenica da se računaju za više od 200 država i teritorija ukazuje na njihovu korisnost i rasprostranjenost. Spomenuti agregatni pokazatelji su:

1. glas i odgovornost (eng. *Voice and Accountability*),
2. politička stabilnost i odsutnost nasilja/terorizma (eng. *Political Stability and Absence of Violence/Terrorism*),
3. učinkovitost vlade (eng. *Government Effectiveness*),
4. kvaliteta regulatornog okvira (eng. *Regulatory Quality*),
5. vladavina prava (eng. *Rule of Law*),
6. kontrola korupcije (eng. *Control of Corruption*).

Pokazatelje ćemo koristiti u hipotezama u nastavku ovog rada. Agregatni pokazatelji se dobiju korištenjem statističke metodologije poznate kao model neopaženih komponenti. Ključna značajka metodologije je da generira margine pogreške za svaku procjenu upravljanja, što je bitna informacija za bilo kakav tip procjene. Sami agregatni pokazatelji su kreirani na način da im je srednja vrijednost jednaka 0, standardna devijacija jednaka 1, a raspon vrijednosti između -2,5 i 2,5 gdje veći broj ukazuje na bolji rezultat.

Šest agregatnih pokazatelja korisni su kao alat za široke usporedbe među zemljama i za ocjenjivanje širokih trendova tijekom vremena. WGI su komplementarni velikom broju drugih nastojanja da se stvore detaljnije mjere upravljanja, često samo za jednu zemlju.

Dobro vladanje je ključno za gospodarski razvoj zemlje, povezano je s borbom protiv

korupcije, pravedno je i uključivo te nam govori kako se provodi vlast u državi. Svjetski pokazatelji upravljanja su konstruirani s namjerom mjerenja i boljeg shvaćanja samog procesa upravljanja.

### 1.3 Indeks otvorenosti proračuna

#### Open Budget Survey

S istim ciljem s kojim su stvorene i objavljene smjernice OECD-a i MMF-a, a to je što ispravnije mjerenje i prakticiranje transparentnog državnog proračuna, IBP (eng. *International Budget Partnership*) 2006. kreće s anketom pod nazivom *Open Budget Survey* (OBS). Namjera IBP-a jest pružiti svima priliku i stvoriti mjesto na kojem se javnost može uključiti u proračunske procese na smislen i značajan način. Cilj je smanjenje zlouporaba izvršne vlasti, pružanje većeg broja boljih informacija o proračunu te institucionaliziranje reformi koje propagiraju i podržavaju odgovornost. Prema IBP-u, OBS je jedini neovisni, usporedivi i činjenični istraživački instrument koji se temelji na činjenicama za mjerenje nekih bitnih aspekata upravljanja i odgovornosti kao što su : (1) sudjelovanje - postoje li formalne i značajne mogućnosti za javnost da se uključi u proces državnog proračuna; (2) nadzor - jesu li nadzorne institucije uspostavljene i osposobljene za ispravno funkcioniranje; (3) transparentnost – jesu li sveobuhvatne informacije o proračunu središnje vlade dostupne javnosti u korisnom vremenskom okviru?[6]

Rezultati ankete iz 2019. i 2021. temelje se na 228 pitanja koja su ista za svaku zemlju. Anketu provode istraživači koji se obično nalaze u dotičnoj zemlji. Gotovo svi istraživači dolaze iz organizacija civilnog društva (od kojih je većina značajno usmjerena na proračunska pitanja) ili akademskih institucija. Od 228 pitanja, njih 145 se bode u uključuje 109 pitanja koja ocjenjuju koliko su informacije o proračunu dostupne javnosti, 18 pitanja koja ocjenjuju mogućnosti sudjelovanja javnosti u proračunskom procesu i 18 pitanja koja ocjenjuju ulogu zakonodavnog i vrhovnog tijela institucija.

83 pitanja koja se ne ocjenjuju pomažu u ispunjavanju OBS ankete prikupljanjem osnovnih informacija o ključnim proračunskim dokumentima i različitim karakteristikama upravljanja javnim financijama u zemlji.

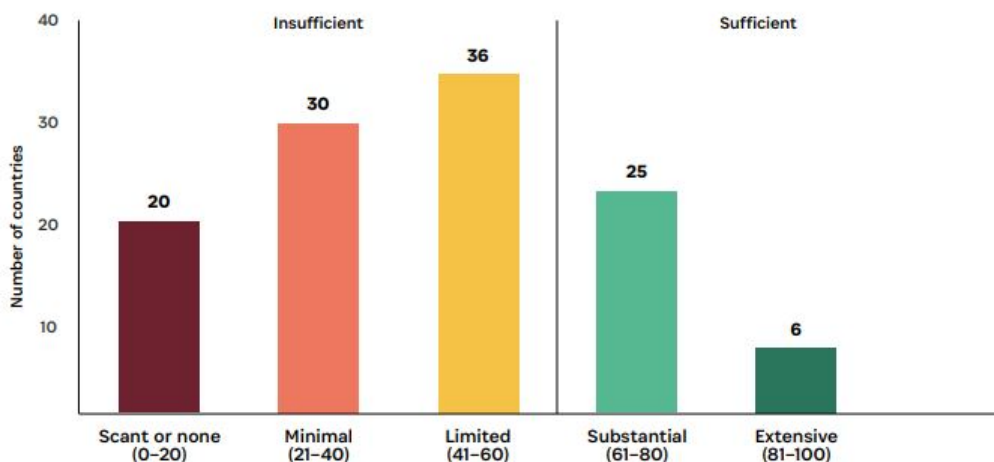
Popunjen nacrt upitnika svake države također provjeravaju anonimni neovisni stručnjaci kao i predstavnici vlade zemlje. Primjerice, u Republici Hrvatskoj OBS upitnik popunjavaju stručnjaci Instituta za javne financije. Provjeru provodi anonimni neovisni stručnjak te predstavnik Ministarstva financija RH.

Iako OBS procjenjuje sudjelovanje, nadzor i transparentnost proračuna kako bi ispravno izmjerio bitne aspekte upravljanja i odgovornosti države, u ovom radu ćemo se fokusirati na procjenu transparentnosti proračuna.

## Mjerenje transparentnosti proračuna

Razina transparentnosti proračuna poznata i kao indeks otvorenosti proračuna (OBI, eng. *Open Budget Index*). OBI analizira dostupnost osam ključnih proračunskih dokumenata koji zajedno pružaju potpuni uvid u to kako su javna sredstva prikupljena, planirana i utrošena tijekom proračunske godine. Kako bi se smatrali "javno dostupnima", dokumenti moraju biti objavljeni na internetu, u roku koji je u skladu s dobrom praksom, te moraju sadržavati informacije koje su sveobuhvatne i korisne. Prvi indeks otvorenosti proračuna objavljen je 2006., kada je i provedena prva anketa IBP-a (OBS). Indeks se većinom računa svake dvije godine otada. Provodi se za više od 100 zemalja (2019. izračunat OBI za čak 117 zemalja [1]), a poprima vrijednosti od 0 do 100.

OBS anketa iz 2019. otkriva skromno globalno poboljšanje transparentnosti proračuna, što je u skladu s općim trendom mjerenim istraživanjem u posljednjih desetak godina. Iako ovi rezultati pružaju nadu za poboljšanje, razine javno dostupnih proračunskih informacija i dalje su ograničene: prosječna ocjena transparentnosti u svijetu u ovom krugu istraživanja je 45 od 100 [1]. Posljednje izračunat OBI od 2021. kaže da ocjena od 61 ili više ukazuje na to da će zemlja vjerojatno objaviti dovoljno materijala za potporu informiranoj javnoj raspravi o proračunu [6]. RH u 2021. ima OBI od 64 pa možemo reći da ima znatno transparentan proračun.



Izvor: The Open Budget Survey 2019. Dostupno na <https://www.internationalbudget.org/open-budget-survey/open-budget-survey-2019>

Slika 1.1: Rezultati proračunske transparentnosti

Iako je moguća vrijednost OBI-ja između 0 i 100, IBP unutar OBS izvještaja često prikazuje rezultate grupirane u 5 skupina, opisujući transparentnost proračuna kao: (1) malu ili

nikakvu (0 – 20); (2) minimalnu (21 – 40); (3) ograničenu (41 - 60), (4) znatnu (61 – 80) i (5) opsežnu (81 – 100).

Iako je posljednje dostupan OBI iz 2021., u nastavku rada koristit ćemo OBI objavljen za 2019. Razlog tome je dostupnost svjetskih pokazatelja upravljanja te činjenica da su posljednji rezultati njihovih izračuna objavljeni za 2019. Također, kod ispunjavanja OBS ankete, pa posljedično i izračuna OBI-ja, bitni su proračunski izvještaji godina koje prethode toj godini.

Ulogu OBI-ja kao primjerenog alata za mjerenje transparentnosti proračuna spominju i De Renzio i Masud [18] te tvrde da su brojne zemlje poboljšale količinu i pokrivenost proračunskih informacija koje objavljuju. Neki od slučajeva poboljšanja su proizašli iz pritiska civilnog društva na temelju nalaza OBI-ja. U svom radu spominju da OBI podaci pomažu identificirati jednostavne korake koje bi vlade i drugi akteri mogli poduzeti za daljnje poboljšanje proračunske transparentnosti. Primjer unaprijeđenja i zahtjeva za većom transparentnosti proračuna vidimo i u radovima Bađun, Urban i Bronić [12], [13], [14] u kojima pri kraju rada navode svoje preporuke o objavljivanju sredinom godine ili pregled sa svim potrebnim informacijama Ministarstvu financija te da bi ono moglo značajno poboljšati ocjenu, odnosno OBI.

Seifert, Carlitz i Mondo [37] kažu da OBI doprinosi transparentnosti na isključiv i neovisan način, pružajući objektivne podatke o transparentnosti proračunskih postupaka brojnih zemalja tijekom dužeg perioda. Sama neovisnost procesa jamči pouzdanost prezentiranih podataka i dobivenih rezultata.

Indeks otvorenosti proračuna jedan je od mnogih pokazatelja transparentnosti proračuna te najčešće korišten. Proizvod je inicijative kojoj je cilj pronaći jednostavne korake koje bi vlade mogle poduzeti kako bi dodatno poboljšali transparentnost proračuna.

Proračunski sustav i proces državnog proračuna očito su povezani s načinom na koji vlada djeluje. Osnova financijskih sustava u zemljama počinje s razvojem državnog proračuna. Zaključili smo i da državni proračun utječe na povjerenje javnosti, gospodarski razvoj zemlje i borbu protiv korupcije. S druge strane, dobro upravljanje predstavljeno je kao odraz boljih usluga za javnost, političku stabilnost i učinkovitost vlade. Zato ćemo promatrati indeks transparentnosti proračuna i svjetske pokazatelje upravljanja kao indikatore transparentnosti proračuna i upravljanja koje vlada provodi te vidjeti u kakvom su oni odnosu.

# Poglavlje 2

## Podaci i hipoteze

U ovom poglavlju ćemo detaljnije opisati svjetske pokazatelje upravljanja te način na koji se računaju. U nastavku rada ćemo ih koristiti kao nezavisne varijable. Također, objasniti ćemo i način mjerenja OBI-ja i koji sve izvještaji ulaze u obzir pri njegovom izračunu. OBI će biti zavisna varijabla u svim hipotezama. Dakle, promatrat ćemo koliko su svjetski pokazatelji upravljanja dobri pokazatelji same transparentnosti proračuna.

### 2.1 Podaci

#### Zavisna varijabla

Indeks transparentnosti proračuna se tipično računa svake dvije godine. Počevši od 2006., kada je započeo Open Budget Survey, OBI se izračunao za 2006., 2008., 2010., 2012., 2015., 2017., 2019. te najnoviji za 2021. Za svako pitanje u anketi, mogući izbori se kreću od "a" do "d" ili su binarni: "a" ili "b". Svaki izbor odgovara brojčanoj ocjeni, u rasponu od 0 do 100, s time da veći brojevi ukazuju na veću transparentnost. Za pitanja čiji odgovori nisu binarni, izbori odgovora "a", "b", "c" i "d" odgovaraju brojčanim rezultatima 100, 67, 33 i 0, redom. Za pitanja s binarnim opcijama odgovora, izbor odgovora "a" odgovara brojčanoj ocjeni 100, a "b" ocjeni 0. Obje vrste pitanja imaju i posljednju opciju ("e", odnosno "c") koja odgovara odgovoru "nije primjenjivo". Takvi odgovori nisu uključeni u zbirni rezultat.

Ključni dokumenti koji se ocjenjuju u OBS anketi su prijedlog proračuna (eng. *Executive Budget Proposal*), proračunske smjernice (eng. *Pre-Budget Statement*), proračunski vodič za građane (CB, eng. *Citizens Budget*), usvojeni proračun (eng. *Enacted Budget*), mjesečna izvješća (eng. *In-Year Reports*), polugodišnje izvješće o izvršenju proračuna (eng. *Mid-Year Review*), godišnje izvješće o izvršenju proračuna (eng. *Year-End Report*) i izvješće o obavljenoj reviziji (eng. *Audit Report*). Svaki put kad jedan od osam ključnih proračunskih

dokumenata nije izrađen ili nije dostupan javnosti, sva pitanja vezana uz taj dokument dobiju ocjenu nula. OBI se izračuna kao običan prosjek dobivenih rezultata. U tablici 2.1 vidimo broj pitanja koji se odnosi na svaki proračunski dokument te time i udio koji dokument donosi u konačnom rezultatu OBI-ja za 2019.

Proračunski dokument ocijenjen u OBS-u	Broj pitanja uključenih u izračun OBI-ja koja se odnose na dokument	Udio u konačnom OBI rezultatu (%)
Prijedlog proračuna	54	49,5
Proračunske smjernice	6	5,5
Proračunski vodič za građane	4	3,7
Usvojeni proračun	6	5,5
Mjesečna izvješća	9	8,3
Polugodišnje izvješće o izvršenju proračuna	9	8,3
Godišnje izvješće o izvršenju proračuna	14	12,8
Izvješće o obavljenoj reviziji	7	6,4
Ukupno	109	100

Tablica 2.1: Udio vrijednosti priložen svakom dokumentu korištenom u izračunu OBI-ja

Tipična poboljšanja u transparentnosti proračuna (prema OBI-ju) mjerena u kasnijim istraživanjima odgovaraju promjenama kao što je uključivanje proračuna iz prethodnih godina u nacrt proračuna, stavljanje postojećih dokumenata na raspolaganje javnosti ili otvaranje relevantnih parlamentarnih procedura. (Seifert, 2013. [37])

Po napatku IBP-a, De Renzio i Castro [17] su istražili alternativne načine ocjenjivanja pitanja iz ankete., proučivši tri metode vaganja (tj. ponderiranja). Prvi način je bio model "jednakih težina" (eng. *Equal Weight*) u kojem rezultat svakog proračunskog dokumenta nosi istu težinu. Utvrdili su da je koeficijent korelacije s tada važećim OBI rezultatom bio prilično visok (0,88) te iako su postojale značajne razlike kod par zemalja, model "jednakih težina" su smatrali manje obranjivim od jednostavnog prosjeka jer se određeni proračunski dokumenti mogu opravdano smatrati važnijima (time nose i veću težinu). Kao drugu metodu kombiniraju jednostavan prosjek i model jednake težine. Koeficijent korelacije je također visok za ovaj model (0,97). Njihova treća metoda je model analize glavnih komponenti (PCA, eng. *principal component analysis*), uobičajena statistička metoda za generiranje sažete mjere iz skupa opažanja. Njime se postiže koeficijent korelacije od 0,99 s originalnim OBI rezultatima.



Iz njihovog rada možemo zaključiti da zamjena OBI metode ponderiranja s nekom od navedenih metoda ne mijenja značajno rezultate zemalja i još manje utječe na njihove relativne rangove.

U ovom radu koristit ćemo OBI podatke izračunate za 2019. te usporediti koliko se razlikuju od onih za 2021. Podaci su dostupni na stranicama International Budget Partnerships. [6]

## Nezavisne varijable

Kao nezavisne varijable u nastavku rada ćemo koristiti svih šest svjetskih pokazatelja upravljanja. Također ćemo kreirati "agregirani" pokazatelj upravljanja kao prosjek šest osnovnih pokazatelja. Agregirani pokazatelj upravljanja koriste Andreula, Chong i Guillen u svom radu [9] kao generalni indikator kvalitete institucije te istražuju vezu između kreiranog pokazatelja i fiskalne transparentnosti.

Kaufmann, Kraaya i Zoido-Lobatón daju detaljan opis svakog od ovih pokazatelja i njihovih izvora te korištenjem modela neopaženih komponenti (UCM, eng. *unobserved components model*) konstruiraju pokazatelje koji odgovaraju osnovnim konceptima upravljanja. [28]. Izvore podataka o upravljanju klasificiraju u dvije dimenzije, ovisno o njihovoj:

1. prirodi, to jest tehnike kojima prakticiraju ili mjere upravljanje
2. pokrivenosti, to jest u kojoj mjeri je skup zemalja obuhvaćenih svakim pokazateljem reprezentativan globalno.

Tijekom kreiranja pokazatelja, podaci su grupirani s namjerom da svaki indikator stvoren iz svake skupine (eng. *cluster*) mjeri sličan (osnovni) koncept upravljanja. Navedene skupine su kombinirane u kumulativni pokazatelj upravljanja koristeći UCM koji izražava promatrane podatke u svakom klasteru kao linearnu funkciju nepromatrane zajedničke komponente upravljanja, plus izraz poremećaja koji bilježi pogreške opažanja ili varijacije uzorkovanja u svakom indikatoru. (Kauffman, 1999. [27])

Gore opisan proces se provodi kroz tri koraka. U prvom koraku se provodi grupacija podataka. U drugom koraku se reskaliraju podaci tako da im vrijednost bude između 0 i 1, gdje veća vrijednost odgovara boljem ishodu. Treći korak korištenjem UCM modela konstruira ponderirani prosjek pojedinačnih pokazatelja za svaki izvor. Dobivene procjene upravljanja sadrže pondere koji odražavaju uzorak korelacije između izvora podataka. Dostupni su na stranicama Svjetske banke [5]. (Kaufmann, 2005. [25])

UCM pridaje veću težinu izvorima podataka koji su međusobno snažnije povezani. Iako ovo ponderiranje poboljšava statističku preciznost agregatnih pokazatelja, obično ne utječe mnogo na rangiranje zemalja s agregatnim pokazateljima. Kompozitne mjere upravljanja koje generira UCM su u jedinicama standardne normalne distribucije, s očekivanjem nula, standardnom devijacijom od jedan i krećući se od približno -2,5 do 2,5, s višim vrijednostima koje odgovaraju boljem upravljanju. Podaci se izvještavaju i u terminima za postotak

ranga, u rasponu od 0 (najniži rang) do 100 (najviši rang). Standardne greške pretvorene su u 90-postotne intervale pouzdanosti. (Kaufman, 1999. [26])

Korišteni su posljednje dostupni podaci koji sadrže informacije od 1996. do 2020., dostupni na stranicama Svjetske banke [5]

## 2.2 Hipoteze

### **Hipoteza 1: WGI pokazatelj glas i odgovornost ima pozitivan utjecaj na OBI**

WGI pokazatelj glas i odgovornost (oznaka *VA*) odražava u kojoj mjeri građani neke zemlje mogu sudjelovati u odabiru vlasti, kao i slobodu izražavanja, slobodu udruživanja i slobodu medija. Autori izvješća za OBI 2008. izračunali su korelaciju OBI te pokazatelja glasa i odgovornosti (0,737), globalnog indeksa integriteta (0,681) i indeksa demokracije koji objavljuje Freedom House (0,691). [17]. Ovi pozitivni koeficijenti ukazuju da WGI glas i odgovornost ima pozitivan utjecaj na OBI.

### **Hipoteza 2: WGI pokazatelj politička stabilnost i odsutnost nasilja/terorizma ima negativan utjecaj na OBI**

WGI pokazatelj politička stabilnost i odsutnost nasilja/terorizma (oznaka *PS*) mjeri percepciju vjerojatnosti političke nestabilnosti i/ili politički motiviranog nasilja, uključujući terorizam. Cimpoeru [16] i Albassam [7] analiziraju odnose WGI-ja i OBI-ja kroz raspon od nekoliko godina. Suprotno od ostalih WGI pokazatelja, dolaze do zaključka da je pokazatelj političke stabilnosti i odsutnosti nasilja/terorizma negativno koreliran s OBI-jem, no da nema statistički značajan utjecaj.

### **Hipoteza 3: WGI pokazatelj učinkovitost vlade ima pozitivan utjecaj na OBI**

WGI pokazatelj učinkovitost vlade (oznaka *GE*) odražava percepciju kvalitete javnih usluga, kvalitete državne službe i stupnja njezine neovisnosti od političkih pritisaka, kvalitetu oblikovanja i provedbe politike te vjerodostojnost vladine predanosti takvim politikama. Montes, Bastos i de Oliveira [32] analiziraju ulažu li zemlje napore da poboljšaju fiskalnu transparentnost i utječe li fiskalna transparentnost na učinkovitost vlade i učinkovitost državne potrošnje te dolaze do zaključka da fiskalna transparentnost ima statistički značajan utjecaj na učinkovitost vlade.

#### **Hipoteza 4: WGI pokazatelj kvaliteta regulatornog okvira ima pozitivan utjecaj na OBI**

WGI pokazatelj kvaliteta regulatornog okvira (oznaka *RQ*) odražava percepciju sposobnosti vlade da formulira i provodi zdrave politike i propise koji dopuštaju i promiču razvoj privatnog sektora. Albassam [7] proučava utjecaj proračunske transparentnosti na kvalitetu upravljanja, čiji je jedan pokazatelj kvaliteta regulatornog okvira. Zaključuje da ima pozitivan utjecaj te potvrđuje značajan utjecaj ljudskog razvoja na odnos transparentnosti proračuna s jedne strane i regulatorne kvalitete i učinkovitost vlade s druge strane. Možemo to protumačiti u smislu da ljudi posvećuju više pažnje na čimbenike koji imaju izravan utjecaj na njihov svakodnevni život (npr. kvalitetu propisa i učinkovitost vlade) nego na ostale pokazatelje upravljanja.

#### **Hipoteza 5: WGI pokazatelj vladavina prava ima pozitivan utjecaj na OBI**

WGI pokazatelj vladavina prava (oznaka *RL*) odražava percepciju u kojoj mjeri agenti imaju povjerenja i pridržavaju se društvenih pravila, a posebno kvalitete izvršenja ugovora, imovinskih prava, policije i sudova, kao i vjerojatnosti kriminala i nasilja. Mensah i Qi [31] proučavaju utjecaj vodstva države na kvalitetu regulatornog okvira i transparentnost proračuna i pokazuju da je stil vodstva orijentiran na učinak (eng. *Performance - Oriented*) pozitivno povezan s vladavinom prava, dok je zaštitnički (eng. *Self/Group-Protective*) pozitivno povezan s većom transparentnošću proračuna.

#### **Hipoteza 6: WGI pokazatelj kontrola korupcije ima pozitivan utjecaj na OBI**

WGI pokazatelj kontrola korupcije (oznaka *CC*) odražava percepciju u kojoj se mjeri koliko javna vlast koristi svoju poziciju radi privatne dobiti, uključujući i sitne i velike oblike korupcije, kao i "zarobljavanje" države od strane elita i privatnih interesa. Više radova istražuje odnos korupcije u državi i transparentnosti proračuna. Haque i Neandis [33] pronalaze dokaze da povećanje fiskalne transparentnosti može smanjiti korupciju. Istom metodom Chen i Neshkova [15] smatraju da je fiskalna transparentnost najvažnija u završnim fazama proračunskog procesa te da proračunski vodič za građane može poslužiti kao snažan antikorupcijski alat. Montes i Luna [33] dolaze do zaključka da bi društva trebala zahtijevati višu razinu transparentnosti od svojih vlada kako bi povećala kontrolu nad korupcijskim djelima koje mogu prakticirati javni dužnosnici i političari.

### **Hipoteza 7: WGI pokazatelj glas i odgovornost ima najveći utjecaj na OBI od svih WGI pokazatelja**

U radu Albassama [7] vidimo da je pokazatelj glas i odgovornost za 2006., 2008., 2010. i 2012. pozitivno povezan s OBI-jem te je njegov koeficijent najveći od svih pokazatelja.

### **Hipoteza 8: agregatni WGI pokazatelj ima pozitivan utjecaj na OBI**

Andreula, Chong i Guillen [9] koriste WGI pokazatelje, no umjesto korištenja individualnih pokazatelja, kreiraju novi agregirani pokazatelj kojom mjere kvalitetu institucije. Definiiraju ga kao običan prosjek svih 6 WGI-ja. U radu koriste i makroekonomske te političke varijable za analiziranje transparentnosti proračuna.

S postavljenim hipotezama želimo provjeriti na koji način nezavisne varijable WGI utječu na indeks transparentnosti proračuna, ako utječu. Pretpostavljamo da svi WGI pokazatelji imaju statistički značajan i pozitivan utjecaj na OBI osim pokazatelja političke stabilnosti i odsutnosti nasilja/terorizma.

## **2.3 Deskriptivna statistika**

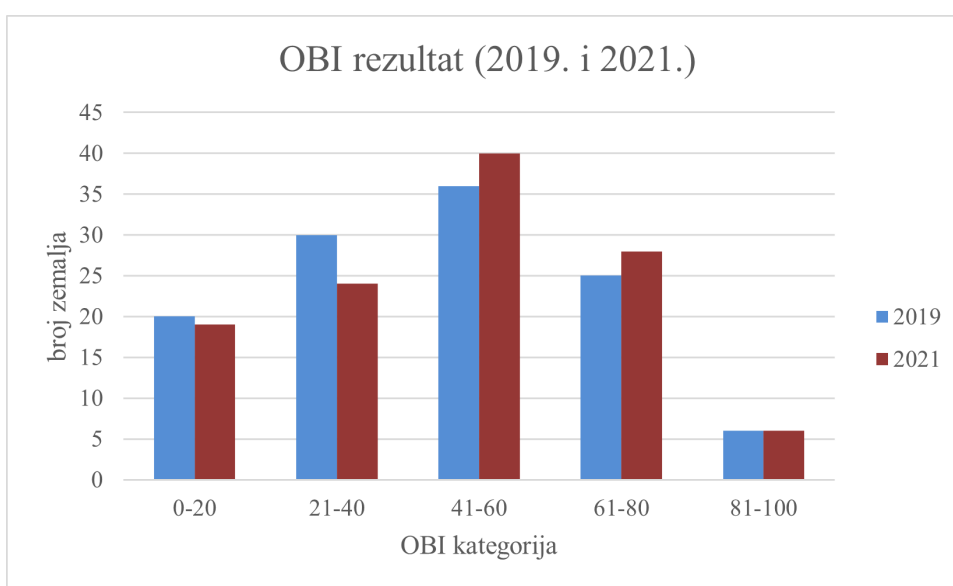
U ovom dijelu rada ćemo deskriptivnom statistikom opisati podatke koje koristimo kod testiranja gore navedenih hipoteza. Deskriptivna ili opisna statistika bavi se organizacijom sakupljenih podataka te njihovim sažetim opisom s pomoću numeričkih i grafičkih prikaza. Vodeći se primjerom IBP-a, OBI ćemo koristiti kao kategorijalnu varijablu iako je svakoj zemlji dodijeljena ocjena između 0 i 100 (kategorije navedene u potpoglavlju Mjerenje transparentnosti proračuna). Uzorak je 117 zemalja koje su za 2019. i 2021. priložili potrebne proračunske dokumente te im je izračunat OBI.

Godina	Minimum	Maksimum	Srednja vrijednost	Medijan	Mod	Standardna devijacija	Varijanca
2019	0	87	44,64	46	38	22,78	518,94
2021	0	87	45,54	46	50	22,31	497,84

Tablica 2.2: Deskriptivna statistika OBI-ja za 2019. i 2021.

Iz tablice 2.2 vidimo da se prosječni OBI rezultat nije puno promijenio iz 2019. u 2021., povećao se za 1 bod, što znači da je većina zemalja uspjela održati transparentost proračuna te da pandemija nije poništila ostvaren napredak u transparentnosti proračunskih procesa.[2]

Histogrami na slici 2.1 prikazuju frekvencije zavisne varijable za 2019. i 2021. Možemo uočiti da najmanji broj zemalja ima najbolje ocijenjenu transparentnost proračuna, dok je za njih 20 (odnosno 19), transparentnost proračuna mala ili nikakva. Također vidimo da se povećao broj zemalja čiji OBI leži u kategoriji 3 (srednjoj kategoriji kojoj je transparentnost proračuna ograničena). Primjerice, iako je Hrvatska 2019. imala bolji OBI nego 2021. (OBI joj je pao sa 68 na 64), ostala je u istoj kategoriji u obje godine.



Slika 2.1: Histogram frekvencija OBI-ja za 2019. i 2021.

Kako bismo prikazali i udio svake kategorije kod oba OBI-ja, promatrat ćemo relativne frekvencije njihovih kategorija u tablici 2.3.

Godina	Kategorija 1	Kategorija 2	Kategorija 3	Kategorija 4	Kategorija 5
2019.	17,09	25,64	30,77	21,37	5,13
2021.	16,24	20,51	34,19	23,93	5,13

Tablica 2.3: Relativne frekvencije OBI-ja za 2019. i 2021. (u %)

Na temelju gornjih usporedbi možemo pretpostaviti da se OBI rezultati ne razlikuju značajno između 2019. i 2021. Pošto su za WGI indikatore (nezavisne varijable) posljednje dostupni podaci iz 2020., promatrat ćemo samo podatke za OBI iz 2019. jer su unutar OBS ankete najbitniji procijenjeni proračunski dokumenti iz trenutne godine.

Iako je većina proračunskih dokumenata iz godine za koju se računa OBI, priloženi u OBS anketi mogu biti izvještaji iz prethodnih godina. Primjerice, za 2019. su neke države priložile izvještaje i iz 2017. i 2018. Zbog tog razloga ćemo provesti deskriptivnu statistiku na WGI pokazateljima za 2019., no isto ćemo provesti i na običnom prosjeku pokazatelja kroz 2017. - 2019. Prosjek pokazatelja ćemo označavati s *Avg\_WGI*. Oznake pojedinačnih pokazatelja specificirane su u dijelu potpoglavlja 2.2.

Pokazatelj <sup>1</sup>	Minimum	Maksimum	Srednja vrijednost	Medijan	Standardna devijacija	Varijanca
VA	-1,9888	1,6553	-0,1824	-0,0885	0,8892	0,7906
PS	-2,7708	1,4113	-0,3531	-0,2677	0,9002	0,8103
GE	-2,4519	1,8619	-0,1970	-0,1954	0,9199	0,8462
RQ	-2,3593	1,8811	-0,1308	-0,2191	0,8976	0,8056
RL	-2,3509	1,9853	-0,2575	-0,4111	0,9072	0,8230
CC	-1,7735	2,1702	-0,2727	-0,4160	0,8899	0,7920

Tablica 2.4: Deskriptivna statistika 6 WGI pokazatelja za 2019. godinu

Iako uz podatke preuzete sa stranice Svjetske banke stoji napomena da su vrijednosti pokazatelja upravljanja otprilike između -2,5 (koji predstavlja slab rezultat) i 2,5 (koji predstavlja jak rezultat), u tablici 2.4 vidimo na primjeru PS pokazatelja da mogu i izaći iz navedenog raspona. Pokazatelj PS je jedini pokazatelj čija vrijednost izlazi iz raspona, s minimumom -2,7708.

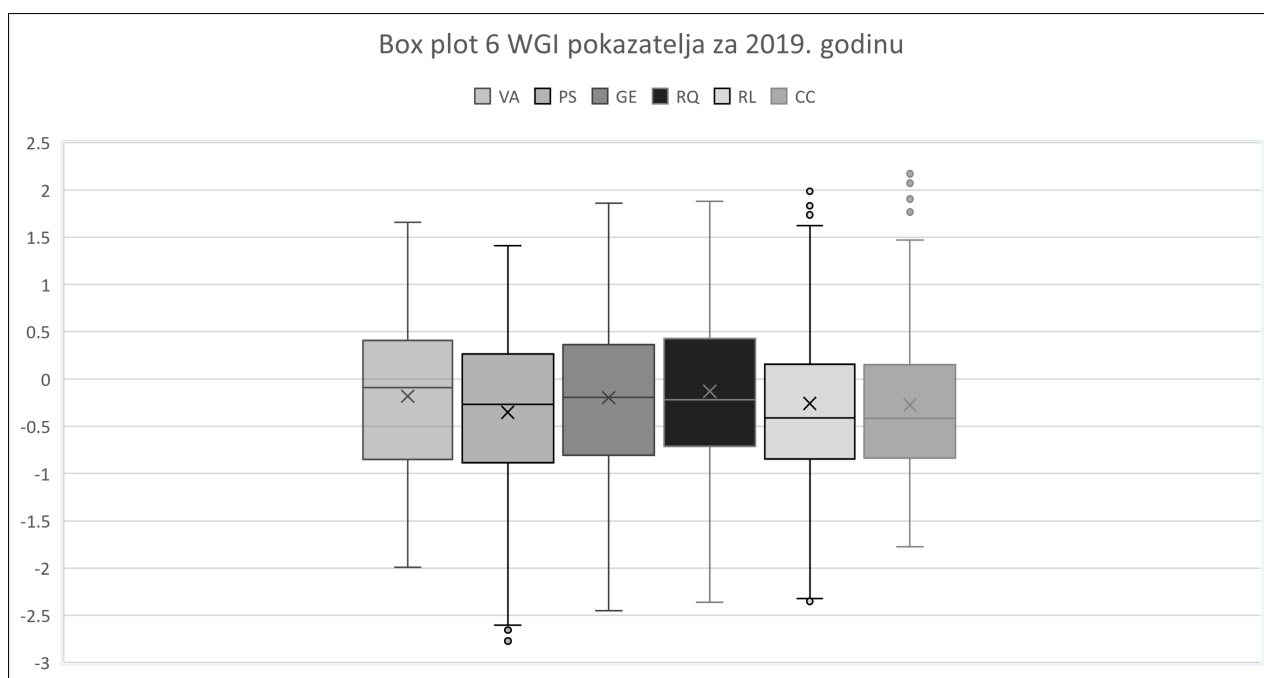
Pokazatelj <sup>1</sup>	Minimum	Maksimum	Srednja vrijednost	Medijan	Standardna devijacija	Varijanca
Avg_VA	-1,9188	1,6777	-0,1625	-0,0617	0,8875	0,7877
Avg_PS	-2,8785	1,5149	-0,3441	-0,2693	0,9028	0,8151
Avg_GE	-2,4273	1,9008	-0,1869	-0,2609	0,9057	0,8203
Avg_RQ	-2,2530	2,0045	-0,1300	-0,2369	0,8875	0,7877
Avg_RL	-2,3314	2,0033	-0,2510	-0,3660	0,9079	0,8243
Avg_CC	-1,7622	2,2184	-0,2725	-0,4594	0,8903	0,7927

Tablica 2.5: Deskriptivna statistika *Avg\_WGI* (razdoblje 2017. - 2019.)

<sup>1</sup>VA = glas i odgovornost, PS = politička stabilnost i odsutnost nasilja, GE = učinkovitost vlade, RQ = kvaliteta regulatornog okvira, RL = vladavina prava, CC = kontrola korupcije

U tablici 2.5 vidimo deskriptivnu statistiku prosječnog pokazatelja *Avg\_WGI*. Vrijednosti kreiranog pokazatelja koji u sebi sadrži tri promatrane godine (2017., 2018. i 2019.) ne odskakuju puno od službenog pokazatelja za 2019. Najveće razlike vidimo u vrijednostima minimuma pokazatelja političke stabilnosti i odsutnosti nasilja (PS).

Na slici 2.2 vidimo i grafički prikaz deskriptivne statistike za WGI pokazatelje iz 2019. pomoću kutijastih dijagrama. Svaki kutijasti dijagram ili box-plot predstavlja jedan WGI pokazatelj.



Slika 2.2: Box-plot WGI pokazatelja za 2019.

Iz deskriptivne statistike zavisne varijable OBI vidimo da većina zemalja i dalje ima nedovoljnu transparentnost državnog proračuna te da se prosječan OBI neznatno promijenio u zadnje dvije godine što bi ukazalo na to da je proces unaprijeđenja državnog proračuna spor postupak. Za WGI-jeve, koji predstavljaju nezavisne varijable, vidimo da većinom poprimaju vrijednosti u rasponu -2,5 do 2,5. Iznimka je pokazatelj političke stabilnosti i odsutnosti nasilja/terorizma. Uvodimo i novu varijablu *Avg\_WGI* koja je dobivena kao prosjek pojednog pokazatelja u razdoblju od 2017. do 2019., za svaku državu iz uzorka.

# Poglavlje 3

## Opis metodologije

U ovom poglavlju opisat ćemo jednu od najbitnijih grana statistike, regresijsku analizu, moćnu statističku metodu koja nam omogućuje ispitivanje odnosa između dvije ili više varijabli od interesa. Provođenje regresije omogućuje nam da pouzdano odredimo koji su čimbenici najvažniji, koji se mogu zanemariti i kako utječu jedni na druge. Regresijske tehnike omogućuju nam da zavisnost, odnosno korelaciju, kvantitativno izrazimo i dobiveni model koristimo ili za predviđanje nekih podataka za koje nemamo mjerenja ili da dođemo do nekih konstanti koje nam tu zavisnost opisuju. Model odabran za obradu podataka ovisi o informacijama koje znamo o podacima i onome što predstavljaju. Ovo poglavlje temelji se na [22], [4], [3] i [40]

### 3.1 Linearna regresija

Veličinu od interesa modeliramo kao slučajnu varijablu koju nazivamo zavisnom varijablom ili odzivom (eng. *response*), a sva ostala mjerenja zovemo nezavisnim varijablama, predviđateljima ili kovarijatama (eng. *predictors*).

Podatke koje želimo opisati predstavljamo kao niz parova

$$(y_i, \mathbf{x}_i), i = 1, \dots, n, \quad (3.1)$$

gdje je  $y_i$  realizacija slučajne varijable  $Y_i$  čija razdioba ovisi o kovarijatama  $\mathbf{x}_i$ . Kovarijate mogu poprimati numeričke i kategorijalne vrijednosti te mogu biti proizvoljno velike dimenzije. Kategorijalne vrijednosti zovemo i faktori.



## 3.2 Jednostavna linearna regresija

U jednostavnoj linearnoj regresiji, zavisna varijabla  $Y$  ovisi o samo jednoj nezavisnoj varijabli (kovarijati)  $X$ . Dakle, vezu između dvije varijable možemo pisati kao:

$$Y = \beta_0 + \beta_1 X \quad (3.2)$$

No u stvarnom životu nezavisne varijable nikada nisu savršeni prediktori zavisnih varijabli. Zbog toga u izrazu 3.2 dodajemo izraz slučajne pogreške. Slučajna pogreška obuhvaća sve ostale čimbenike koji, osim nezavisne varijable  $X$ , utječu na ovisnu varijablu  $Y$ .

Dakle, jednostavni linearni model sada glasi:

$$Y = \beta_0 + \beta_1 X + \epsilon, \quad (3.3)$$

gdje je:

- $\beta_0$  - nepoznati regresijski parametar koji trebamo procijeniti, a predstavlja presjek pravca s  $y$ -osi; nazivamo ga odsječak
- $\beta_1$  - nepoznati regresijski parametar koji trebamo procijeniti, a predstavlja nagib (ili koeficijent smjera) pravca
- $\epsilon$  - slučajna pogreška

Neka je  $n$  veličina uzorka kojeg promatramo. Tada za svaki podatak  $i = 1, \dots, n$  i opažene vrijednosti uzorka  $(y_i, x_i)$  model glasi

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad (3.4)$$

### Pretpostavke modela

Prije korištenja linearne regresije za modeliranje odnosa između dviju varijabli, moramo provjeriti nekoliko pretpostavki modela. Uvjeti koji trebaju biti ispunjeni prije nego što upotrijebimo model i izvedemo iz njega zaključke su:

$$(1) \mathbb{E}[\epsilon_i] = 0, \forall i = 1, \dots, n \quad (1)$$

$$(2) \text{Var}[\epsilon_i] = \sigma^2 > 0, \forall i = 1, \dots, n \quad (2)$$

$$(3) \text{Cov}[\epsilon_i \epsilon_j] = 0, \forall i, j \text{ t.d. } i \neq j \quad (3)$$

Pretpostavka koja pokriva uvjete (1), (2) i (3) je da  $(\epsilon_i)_i$  mora činiti niz nezavisnih jednako distribuiranih normalnih slučajnih varijabli s očekivanjem nula i varijancom  $\sigma^2$ , tj.  $\epsilon_i \sim N(0, \sigma^2)$ , t.d.  $\epsilon_i, \epsilon_j$  nezavisne  $\forall i \neq j$ . Navedene uvjete zovemo Gauss-Markovljevi uvjeti te nam koriste kod procjene parametara modela. Vidimo da gornja je gornja pretpostavka ekvivalentna i tvrdnji  $Y \sim N(\mu, \sigma^2)$ , gdje je

$$\mu = \mathbb{E}[Y] = \mathbb{E}[\beta_0 + \beta_1 X + \epsilon] = \beta_0 + \beta_1 X. \quad (3.5)$$

### Procjena parametara - metoda najmanjih kvadrata

Problem određivanja jednostavnog linearnog regresijskog modela svodi se na procjenu nepoznatih parametara  $\beta_0$  i  $\beta_1$ . U ovom potpoglavlju obradit ćemo jednu od najpoznatijih metoda za procjenu parametara u regresijskim modelima.

Uz pretpostavku postojanja greške modela, opažena vrijednost  $y_i$  i očekivana vrijednost  $y'_i$  nisu u potpunosti iste. To nam govori da točke  $(x_i, y_i)$  ne leže u potpunosti na pravcu  $Y = \beta_0 + \beta_1 X$ . Ideja je dobiti procijenjene parametre  $\hat{\beta}_0$  i  $\hat{\beta}_1$  tako da razlika  $y_i - y'_i$ , gdje je  $y'_i$  teorijska vrijednost koja odgovara vrijednosti  $\beta_0 + \beta_1 x_i$ , bude što manja. Metoda najmanjih kvadrata (OLS, eng. *ordinary least squares*) to postiže minimizacijom sume kvadrata odstupanja teorijskih od opaženih vrijednosti.

Definirajmo funkciju

$$L(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - y'_i)^2 = \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2. \quad (3.6)$$

Dakle, cilj je odrediti  $\hat{\beta}_0$  i  $\hat{\beta}_1$  tako da minimiziraju funkciju  $L(\beta_0, \beta_1)$ . Znamo da je  $(\hat{\beta}_0, \hat{\beta}_1)$  stacionarna točka za funkciju  $L$  3.6 ako vrijedi

$$\frac{\partial L}{\partial \beta_0}(\hat{\beta}_0, \hat{\beta}_1) = 0, \quad \frac{\partial L}{\partial \beta_1}(\hat{\beta}_0, \hat{\beta}_1) = 0. \quad (3.7)$$

Problem se svodi na rješavanje sljedećeg sustava jednačbi:

$$\begin{cases} \frac{\partial L}{\partial \beta_0}(\beta_0, \beta_1) = \sum_{i=1}^n -2(y_i - (\beta_0 + \beta_1 x_i)) = 0 \\ \frac{\partial L}{\partial \beta_1}(\beta_0, \beta_1) = \sum_{i=1}^n -2x_i(y_i - (\beta_0 + \beta_1 x_i)) = 0, \end{cases}$$

što je ekvivalentno sustavu

$$\begin{cases} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i)) = 0 \\ \sum_{i=1}^n x_i (y_i - (\beta_0 + \beta_1 x_i)) = 0. \end{cases} \quad (3.8)$$

Iz prve jednadžbe možemo izraziti  $\beta_0$  kao

$$-n\beta_0 + \sum_{i=1}^n (y_i - \beta_1 x_i) = 0 \Rightarrow \sum_{i=1}^n (y_i - \beta_1 x_i) = n\beta_0 \Rightarrow \frac{\sum_{i=1}^n (y_i - \beta_1 x_i)}{n} = \beta_0. \quad (3.9)$$

Ubacivanjem izraženog  $\beta_0$  iz jednadžbe 3.9 u drugu jednadžbu sustava 3.8, dobijemo

$$\begin{aligned} \sum_{i=1}^n x_i \left( y_i - \frac{\sum_{i=1}^n (y_i - \beta_1 x_i)}{n} - \beta_1 x_i \right) &= 0 \\ \Rightarrow \sum_{i=1}^n x_i \left( y_i - \frac{1}{n} \sum_{i=1}^n y_i \right) - \beta_1 \sum_{i=1}^n x_i \left( x_i - \frac{1}{n} \sum_{i=1}^n x_i \right) &= 0 \\ \Rightarrow \beta_1 &= \frac{\sum_{i=1}^n x_i \left( y_i - \frac{1}{n} \sum_{i=1}^n y_i \right)}{\sum_{i=1}^n x_i \left( x_i - \frac{1}{n} \sum_{i=1}^n x_i \right)} \end{aligned} \quad (3.10)$$

Označimo s  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  i  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  aritmetičku sredinu podataka  $y_1, \dots, y_n$  i  $x_1, \dots, x_n$ , redom. Tada iz 3.9 i 3.10 dobijemo izraze za procjenu  $\hat{\beta}_0$  i  $\hat{\beta}_1$ :

$$\hat{\beta}_0 = \bar{y} - \beta_1 \bar{x}_i \quad (3.11)$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (3.12)$$

Potrebno je provjeriti da se u točkama određenim 3.11 i 3.12 postiže minimum funkcije  $L$ . No to lako možemo provjeriti jer minimum funkcije mora zadovoljavati sljedeće uvjete:

$$\frac{\partial^2 L}{\partial \beta_0^2}(\hat{\beta}_0, \hat{\beta}_1) \frac{\partial^2 L}{\partial \beta_1^2}(\hat{\beta}_0, \hat{\beta}_1) - \frac{\partial^2 L}{\partial \beta_1 \partial \beta_0}(\hat{\beta}_0, \hat{\beta}_1) > 0 \quad (3.13)$$

i

$$\frac{\partial^2 L}{\partial \beta_0^2}(\hat{\beta}_0, \hat{\beta}_1) = 0. \quad (3.14)$$

Ponovnim deriviranjem jednadžbi iz sustava 3.8 dobijemo

$$2n \cdot \left( 2n \sum_{i=1}^n x_i^2 \right) - 0,$$

što očito zadovoljava uvjet 3.13 jer  $x_i^2 > 0, \forall i$ .

Nadalje,

$$\frac{\partial^2 L}{\partial \beta_0^2}(\hat{\beta}_0, \hat{\beta}_1) = 2n,$$

što je strogo veće od nula jer je  $n \geq 1$ .

Dakle,  $\hat{\beta}_0$  i  $\hat{\beta}_1$  su t.d. je razlika  $y_i - y'_i$  minimalna.

Za dobivene procijenjene parametre možemo definirati i procijenjenu vrijednost mjerenja  $y_i$  kao

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i, \forall i = 1, \dots, n. \quad (3.15)$$

**Teorem 3.2.1** (Gauss-Markovljev teorem). *Ukoliko su zadovoljeni uvjeti (1), (2) i (3), onda su  $\hat{\beta}_0$  i  $\hat{\beta}_1$  najbolji linearni nepristrani procjenitelji za  $\beta_0$  i  $\beta_1$  (BLUE, eng. Best Linear Unbiased Estimator).*

Nepriistranost znači  $\mathbb{E}[\hat{\beta}_j] = \beta_j, \forall j$ , a linearnost znači da se procjena može prikazati kao linearna funkcija vrijednosti ovisne varijable, odnosno,  $\hat{\beta}_j = \sum_{i=1}^n \omega_{ij} y_i$ , gdje je  $\omega_{ij}$  težinska funkcija uzoračkih vrijednosti neovisne varijable.

Preciznije,  $\omega_{ij} = \frac{x_{ij} - \bar{x}_j}{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}, j = 1, \dots, k$ .

Najbolji procjenitelji znači da imaju najmanju varijancu (najmanje kvadratno odstupanje od točne vrijednosti). Primjenom BLUE načela mogu se naći i varijance procjenitelja.

Zbog linearosti procjenitelja mora vrijediti:

$$\hat{\beta}_1 = \sum_{i=1}^n a_i y_i, i = 1, \dots, n,$$

gdje su  $a_i, i = 1, \dots, n$  konstante koje treba odrediti.

Iz nepristranosti procjenitelja slijedi:

$$\mathbb{E}[\hat{\beta}_1] = \mathbb{E}\left[\sum_{i=1}^n a_i y_i\right] = \sum_{i=1}^n a_i \mathbb{E}[y_i] = \sum_{i=1}^n a_i \mathbb{E}[\beta_0 + \beta_1 x_i] = \beta_0 \sum_{i=1}^n a_i + \beta_1 \sum_{i=1}^n a_i x_i, \quad (3.16)$$

Dakle, da bi procjenitelj  $\hat{\beta}_1$  bio nepristran, mora vrijediti  $\sum_{i=1}^n a_i = 0$  i  $\sum_{i=1}^n a_i x_i = 0$ . Varijanca procjenitelja  $\hat{\beta}_1$  glasi:

$$\text{Var}[\hat{\beta}_1] = \text{Var}\left[\sum_{i=1}^n a_i y_i\right] = \sum_{i=1}^n a_i^2 \text{Var}[y_i] = \sum_{i=1}^n a_i^2 \sigma^2 = \sigma^2 \sum_{i=1}^n a_i^2, \quad (3.17)$$

gdje su  $y_i$  nezavisni s varijancom  $\sigma^2$ , te  $a_i$  nestohastički.

Dakle, određivanje varijance se svodi na problem minimizacije  $\sigma^2 \sum_{i=1}^n a_i^2$  uz uvjet da je  $\hat{\beta}_1 = \sum_{i=1}^n a_i y_i$ .

Problem minimizacije uz ograničenja riješimo pomoću metode Lagrangeovih multiplikatora, iz koje slijedi da  $a_i$  glasi:

$$a_i = \frac{-\sum_{i=1}^n x_i + n x_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}, i = 1, \dots, n. \quad (3.18)$$

Uvrštavanjem 3.18 u izraz 3.17 i korištenjem uvjeta  $\sum_{i=1}^n a_i = 0$  i  $\sum_{i=1}^n a_i x_i$ , slijedi

$$\text{Var}[\hat{\beta}_1] = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}. \quad (3.19)$$

Analogno se izvede varijanca za  $\hat{\beta}_0$  te glasi

$$\text{Var}[\hat{\beta}_0] = \sigma^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right). \quad (3.20)$$

Gauss-Markovljev teorem nam govori da je OLS metoda najbolja metoda procjene parameta ukoliko su zadovoljeni Gauss-Markovljevi uvjeti.

Jednostavna linearna regresija koristi se kod proučavanja veze između dviju varijabli  $X$  i  $Y$ . Predstavlja linearnu vezu koju možemo predočiti pravcem kojemu je koeficijent smjera parametar  $\beta_1$ . Parametri regresije se mogu procijeniti metodom najmanjih kvadrata koja minimizira zbroj kvadrata svih vertikalnih odstupanja točaka od pravca.

### 3.3 Višestruka linearna regresija

Višestruka linearna regresija sadrži jednu zavisnu i dvije ili više nezavisnih varijabli. Višestruku regresiju koristimo kada želimo znati koliko je jaka veza između dvije ili više nezavisnih varijabli i jedne zavisne varijable ili vrijednost zavisne varijable za određene vrijednosti nezavisnih varijabli. Model je dan s

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \epsilon,$$

gdje je  $Y$  i dalje nezavisna varijabla, dok  $p$  označava broj nezavisnih varijabli  $X_1, \dots, X_p$ , te su:

- $\beta_0$  - nepoznati regresijski parametar koji trebamo procijeniti, nazivamo ga odsječak,
- $\beta_k, k = 1, \dots, p$  - nepoznati regresijski parametri koje trebamo procijeniti, predstavlja nagib (ili koeficijent smjera) pravca svake nezavisne varijable  $X_k$ ,
- $\epsilon$  - slučajna pogreška.

Za razliku od jednostavne linearne regresije, u slučaju višestruke regresije, procjenitelji se nalaze u  $p$ -dimenzionalnom prostoru te je teže prikazati odnos između podataka grafički. Možemo reći da  $Y$  ovisi o  $X$ , gdje je  $X = (X_1, \dots, X_p)$  U uzorku veličine  $n$ , opažanje za  $i$ -ti podatak označavamo kao:

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}, i = 1, \dots, n,$$

gdje su  $y_i, x_{i1}, \dots, x_{ip}$  opažene vrijednosti uzorka.  
Model možemo prikazati i u matricnom zapisu kao:

$$y = X\beta + \epsilon,$$

gdje je

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \epsilon = \begin{bmatrix} \epsilon_0 \\ \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

U ovom slučaju,  $y$  se naziva vektorom odgovora,  $X$  matricom dizajna,  $\beta$  vektorom parametara te  $\epsilon$  vektorom greške ili odstupanja.

Kao i kod jednostavne linearne regresije, postaviti ćemo pretpostavke modela te opisati procjenu parametara regresije.

### Pretpostavke modela

Pretpostavke višestruke linearne regresije su jednake kao i kod jednostavne linearne regresije.

Kod višestruke linearne regresije pretpostavljamo linearnost u parametrima. Također, podaci kojima se koristimo za procjenu regresijskog modela dolaze iz nezavisnog slučajnog uzorka s  $n$  mjerenja (slučajni uzorak) te nijedna varijabla nije konstanta, niti postoji linearna veza među nezavisnim varijablama (multikolinearnost).

Varijanca greške uz dane nezavisne varijable je konstanta (homoskedastičnost) te je očekivanje greške modela jednako nula.

Iz prethodno navedenih pretpostavki vidimo da i u višestrukom modelu vrijede Gauss-Markovljevi uvjeti. Razlika je u tome da je  $\epsilon$  sada vektor grešaka:

$$(1) \mathbb{E}[\epsilon_i] = 0, \forall i = 1, \dots, n \quad (1)$$

$$(2) \text{Var}[\epsilon_i] = \sigma^2 > 0, \forall i = 1, \dots, n \quad (2)$$

$$(3) \text{Cov}[\epsilon_i \epsilon_j] = 0, \forall i, j, i \neq j \quad (3)$$

### Procjena parametara

Cilj ovog potpoglavlja je pronaći nepoznate parametre  $\beta_j$  vektora  $\beta = (1, \beta_0, \beta_1, \dots, \beta_p)^T$  tako da što bolje opisuju utjecaj nezavisne varijable  $X_j$  na  $Y$ . Kako vrijede Gauss-Markovljevi uvjeti, po teoremu 3.2.1 znamo da ćemo metodom najmanjih kvadrata doći do najboljih linearnih nepristranih procjenitelja parametara. Dakle, želimo pronaći  $\beta_j$  minimizirajući sumu kvadrata razlika između opaženih i procijenjenih vrijednosti.

Neka je sada  $L$  funkcija  $p + 1$  varijabli:

$$L(\beta_0, \beta_1, \dots, \beta_p) = \sum_{i=1}^n (y_i - y'_i)^2 = \sum_{i=1}^n \left( y_i - \left( \beta_0 + \sum_{j=1}^p \beta_j x_{ij} \right) \right)^2, i = 1, \dots, n. \quad (3.21)$$

Kako bi  $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$  minimizirao funkciju  $L$ , znamo da mora zadovoljiti 3.7, za svaki  $\beta_j$ . No, tada bismo morali riješiti sustav od  $p + 1$  jednažbi.

Zbog olakšanog prikaza rješenja, promatrat ćemo problem minimizacije u matričnom zapisu. Tada funkcija  $L$  glasi:

$$L(\beta) = (y - X\beta)^T (y - X\beta) \quad (3.22)$$

$\beta^T X^T y$  je skalar pa vrijedi

$$L(\beta) = y^T y - \beta^T X^T y - y^T X\beta + \beta^T X^T X\beta = y^T y - 2\beta^T X^T y + \beta^T X^T X\beta,$$

jer  $(\beta^T X^T y)^T = y^T X\beta$ . Sada iz uvjeta 3.7 i 3.2 slijedi

$$\frac{\partial L}{\partial \beta} = 0 \Rightarrow -2X^T y + 2X^T X\beta = 0 \Rightarrow X^T X\beta = X^T y.$$

Množenjem zadnjeg izraza s obje strane s  $(X^T X)^{-1}$  dobijemo procjenitelj

$$\hat{\beta} = (X^T X)^{-1} X^T y, \quad (3.23)$$

Napomenimo da je množenje inverzom matrice  $(X^T X)$  u prethodnom koraku moguće zbog linearne nezavisnosti varijabli  $X_j$ , iz čega slijedi da nijedan stupac matrice  $X$  nije linearna kombinacija ostalih stupaca. Nužan uvjet regularnosti matrice  $(X^T X)$  je da je  $p < n$ , to jest da broj parametara koje tražimo regresijom bude manji od broja opažanja iz uzorka. Ne samo to, kada su varijable  $X_j$  linearno nezavisne, znači da je matrica  $X$  regularna. To nam osigurava i jedinstvenost rješenja 3.23.

Procijenjene vrijednosti  $\hat{y}$  tada možemo zapisati kao

$$\hat{y} = X\hat{\beta} = (X^T X)^{-1} X^T y,$$

gdje su  $y$  opažene vrijednosti varijable  $Y$ .

Pokažimo i neka statistička svojstva procjenitelja  $\hat{\beta}$ .

Nepriistranost slijedi iz  $\mathbb{E}(\epsilon) = 0$  i  $(X^T X)^{-1} X^T X = I$ :

$$\begin{aligned} \Rightarrow \mathbb{E}(\hat{\beta}) &= \mathbb{E}[(X^T X)^{-1} X^T y] \\ &= \mathbb{E}[(X^T X)^{-1} X^T (X\beta + \epsilon)] \\ &= \mathbb{E}[(X^T X)^{-1} X^T X\beta + (X^T X)^{-1} \epsilon] \\ &= \beta \end{aligned} \quad (3.24)$$

Varijacijsko-kovarijacijsku matricu procjenitelja  $\hat{\beta}$  dobijemo kao:

$$\begin{aligned} \Rightarrow \text{Var}(\hat{\beta}) &= \text{Var}[(X^T X)^{-1} X^T y] \\ &= (X^T X)^{-1} X^T \text{Var}(y) [(X^T X)^{-1} X]^T \\ &= \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1} \\ &= \sigma^2 (X^T X)^{-1} \end{aligned} \quad (3.25)$$

Druga jednakost slijedi iz činjenice da je matrica  $(X^T X)^{-1}$  konstanta pa je njena varijanca jednaka 0. Iz  $\text{Var}(y) = \sigma^2 I$  slijedi treća jednakost. Varijancu  $\sigma^2$  procjenjujemo s

$$\hat{\sigma}^2 = \frac{1}{n - p - 1} \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

Konačno, uz zadovoljene Gauss-Markovljeve uvjete i pretpostavku da je  $\epsilon \sim N(0, \sigma^2)$  n.j.d., slijedi da je procjenitelj parametara  $\hat{\beta}$  normalno distribuiran,

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1}). \quad (3.26)$$

Kao i jednostavna linearna regresija, višestruka linearna regresija se koristi za ispitavanje veze između nezavisnih i zavisne varijable. Za razliku od jednostavne, u ovom modelu možemo proučavati utjecaj više nezavisnih varijabli. Metoda procjene parametara je ista kao i kod jednostavne linearne regresije (OLS).

### 3.4 Određivanje prikladnog modela

Uz procjenu parametara modela, zanima nas i koliko dobro tako dobiveni regresijski model opisuje opažene podatke. Takva mjerenja se mogu koristiti za testiranje statističkih hipoteza ili provjeru distribucije frekvencija ishoda. Pojam dobrote pristajanja (eng. *goodness of fit*) govori koliko uspješno model opisuje zavisnost (tj. korelaciju) između zavisne i nezavisne varijable.

U regresijskoj analizi se kao pokazatelji pristajanja modela koriste koeficijent determinacije, nedostatak zbroja kvadrata, reducirana  $\chi^2$  statistika te validacija regresije.



## Koeficijent determinacije

Korelacijski koeficijent  $r^2$  nam govori o korelaciji i smjeru linearne povezanosti između dvije varijable, gdje  $r = 1$  ukazuje na pozitivnu korelaciju,  $r = -1$  negativnu korelaciju te  $r = 0$  znači da korelacije nema. Moramo imati na umu da  $r^2$  daje samo djelomičnu informaciju o uspješnosti regresije jer objašnjava jedino korelaciju između zavisne i nezavisne varijable.

Koeficijent determinacije predstavlja omjer modelom opisanih varijacija u odnosu na ukupne varijacije podataka. Koeficijent determinacije je dan s:

$$R^2 = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{SS_{err}}{SS_{tot}},$$

gdje je:

- $e_i$  - reziduali definirani kao  $e_i = y_i - \hat{y}_i, i = 1, \dots, n$ ,
- $SS_{err}$  - ukupna varijacija u vrijednostima  $y$ ,
- $SS_{tot}$  - udio varijacija u vrijednostima  $y$  koji nije objašnjen modelom.

Koeficijent  $R^2$  poprima vrijednosti između 0 i 1, gdje  $R^2 = 1$  ako su sva odstupanja  $u_i = 0$ , to jest model savršeno opisuje podatke. Ako je  $R^2 = 0$ , tada model ne objašnjava kretanje  $Y$  ništa bolje od sredine uzorka  $\bar{y}$ . Primijetimo i da je koeficijent  $R^2$  osjetljiv u odnosu na varijablu  $Y$  s obzirom da mjeri njenu objašnjenu varijaciju.

Iako je koeficijent determinacije često korišten pokazatelj, prate ga tri problema.

1. Ne vrijedi za regresijske modele koji su izračunati metodom najmanjih kvadrata bez konstantnog parametra  $\beta_0$ .

U tom slučaju koristimo necentrirani koeficijent determinacije koji glasi:

$$R^2 = 1 - \frac{\sum_{i=1}^n u_i^2}{\sum_{i=1}^n y_i^2}.$$

2. Ne vrijedi za regresijske modele koji nisu izračunati pomoću metode najmanjih kvadrata.

U tom slučaju koeficijent  $R^2$  možemo računati kao  $R^2 = (\text{corr}(y_i, \hat{y}_i))^2$  gdje vidimo povezanost koeficijenta determinacije s korelacijskim koeficijentom (kao kvadriran koeficijent korelacije).

3.  $R^2$  nikada ne opada s povećanjem broja nezavisnih varijabli niti onda kada dodatne varijable ne objašnjavaju kretanje zavisne varijable.

U tom slučaju koristimo korigirani koeficijent determinacije koji u svom izračunu uključuje stupnjeve slobode te glasi:

$$\bar{R}^2 = 1 - \frac{n-1}{n-p-1}(1-R^2).$$

### Testiranje hipoteza

Pri statističkom testiranju hipoteza moramo paziti na dvije vrste grešaka. Greška I. tipa se odnosi na odbacivanje istinite hipoteze, dok je neodbacivanje pogrešne hipoteze greška II. tipa. Vjerojatnost nastanka greške I. tipa kontroliramo odabirom razine značajnosti  $\alpha$ , koja uobičajeno iznosi 5%.

Ekonometrijski paketi koji se često koriste pri testiranju hipoteza imaju ugrađen izračun  $p$ -vrijednosti ili točne razine značajnosti testa. Ukoliko je  $p$ -vrijednost manja od razine značajnosti  $\alpha$ , možemo odbaciti nultu hipotezu  $H_0$ . Za rezultate kažemo da su statistički značajni kada možemo odbaciti nultu hipotezu  $H_0$ , inače nisu statistički značajni.

Za testiranje hipoteza o parametrima modela mogu se koristiti testovi koje ćemo opisati u nastavku,  $t$ -test i  $F$ -test.

#### $t$ - test

Neka je  $z$  varijabla dana s

$$Z = \frac{\hat{\beta} - \beta}{\sqrt{\sigma^2(X^T X)^{-1}}}.$$

Tada iz 3.26 možemo zaključiti da  $Z$  ima standardnu normalnu distribuciju, s očekivanjem nula i varijancom 1. Kako su reziduali uzorka  $e_i$  normalno distribuirani, suma njihovih kvadrata ima  $\chi^2$  distribuciju. Ukoliko  $\sigma^2$  zamijenimo procijenjenom nepristanom varijancom, dobijemo da tako definirana varijabla ima Studentovu  $t$  distribuciju.

Iz ovoga definiramo testnu statistiku  $T$  kao

$$T = \frac{\hat{\beta}_k - b_k}{\hat{\sigma} \sqrt{(X^T X)^{-1}_{jj}}} \sim t(n-p-1), \quad (3.27)$$

pri čemu  $b_k$  predstavlja vrijednost parametra  $\beta_k$  kojeg želimo provjeriti hipotezom. To jest, slijedi  $t$ -distribuciju s  $(n-p-1)$  stupnjeva slobode. Testnom statistikom  $t$  možemo testirati dvije vrste hipoteza.

Prva glasi

$$\begin{aligned} H_0 : \beta_k &= b_k \\ H_1 : \beta_k &\neq b_k \end{aligned} \quad (3.28)$$

Budući da alternativna hipoteza  $H_1$  dozvoljava vrijednosti koje su veće ili manje od  $b_k$ , test zovemo dvostrani  $t$ -test. Ovaj test je koristan kada želimo provjeriti ima li varijabla  $X_k$  ikakav utjecaj na zavisnu varijablu  $Y$ . Tada hipotezu testiramo za  $b_k = 0$ . Dakle, ako možemo odbaciti nultu hipotezu  $H_0 : \beta_k = 0$ , tada kažemo da je utjecaj varijable  $X_k$  na  $Y$  statistički značajan.

Testnom statistikom  $T$  možemo testirati i hipoteze koje imaju jednostranu alternativnu hipotezu, primjerice:

$$\begin{aligned} H_0 : \beta_k &\leq b_k \\ H_1 : \beta_k &> b_k \end{aligned} \quad (3.29)$$

Takav test zovemo jednostranim  $t$ -testom.

### **$F$ - test**

$F$ -test koistimo ukoliko želimo testirati zajedničke hipoteze nad parametrima. Tada definiramo ograničeni model kao  $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$ , gdje je  $k < p$ . Prošireni model s  $p$  nezavisnih varijabli mora sadržavati sve zavisne varijable ograničenog modela.

Ovim testom želimo testirati hipotezu

$$\begin{aligned} H_0 : \beta_{k+1} &= \beta_{k+2} = \dots = \beta_p = 0 \\ H_1 : \exists i \in \{k+1, \dots, p\} &t.d. \beta_i \neq 0 \end{aligned} \quad (3.30)$$

Cilj je testirati utječu li nezavisne varijable  $X_{k+1}, \dots, X_p$  na  $Y$  kako bismo dobili optimalan model te eventualno izbacili one varijable koje nemaju utjecaj na zavisnu varijablu.

Testna statistika glasi

$$F = \frac{SSE_1 - SSE_2}{SSE_2} \cdot \frac{n - p - 1}{p - k} \sim F(p - k, n - p - 1),$$

gdje je

- $SSE_1$  - suma kvadrata reziduala malog modela,
- $SSE_2$  - suma kvadrata reziduala proširenog modela.

Testna statistika  $F$  ima  $F(p-k, n-p-1)$  distribuciju u slučaju neodbacivanja nulte hipoteze. ANOVA metoda na temelju  $F$ -testa računa rezultate koji se koriste pri analiziranju utjecaja jedne ili više nezavisnih kategorijalnih varijabli na jednu zavisnu neprekidnu varijablu.

### Normalnost uzorka

Normalnost uzorka testiramo normalnim vjerojatnosnim grafom te Lillieforsovom inačicom Kolmogorov-Smirnovljeva testa. Općeniti test prilagodbe uzorka  $X_1, X_2, \dots, X_n$  s neprekidnom distribucijom  $F$  nekoj teoretskoj distribuciji  $F_0$  možemo testirati Kolmogorov-Smirnovljevim testom. Hipoteza tada glasi

$$\begin{aligned} H_0 : F &= F_0 \\ H_1 : F &\neq F_0 \end{aligned} \tag{3.31}$$

Njena testna statistika je

$$D_n = \sup_{x \in \mathbb{R}} |\hat{F}_n - F_0(x)| = \max\left\{\max_{1 \leq i \leq n} \left| \frac{i-1}{n} - F_0(x_i) \right|, \left| \frac{i}{n} - F_0(x_i) \right|\right\},$$

gdje je  $(x_1, \dots, x_n)$  uređena realizacija slučajnog uzorka.

Vidimo da ovisno o postavljenoj hipotezi, odabiremo prikladan model. Također se pri provođenju testa moraju provjeriti uvjeti na kojima se zasniva teorija kako bi rezultati bili opravdani. Moramo imati na umu da ne postoji savršen model kod testiranja, već moramo pronaći metodu koja najviše odgovara postavljenoj hipotezi.

Jednostavnu i višestruku linearnu regresiju koristit ćemo pri testiranju hipoteza iz ovog rada. Vidimo da ove metode opisuju veze između dviju ili više varijabli, a nas točno zanima utjecaj svjetskih pokazatelja upravljanja na OBI. Kod testiranja utjecaja svakog pokazatelja posebno ćemo koristiti jednostavnu linearnu regresiju. Provest ćemo i višestruku linearnu regresiju te vidjeti možemo li donijeti snažne zaključke te u kakvom odnosu su nezavinske varijable.

# Poglavlje 4

## Empirijska analiza

U ovom poglavlju ćemo testirati hipoteze postavljene u potpoglavlju 2.2 te interpretirati dobivene rezultate.

### 4.1 Testiranje hipoteza

Pri svakom testiranju hipoteze, potrebno je provjeriti jesu li zadovoljeni Gauss-Markovljevi uvjeti kako bismo mogli provesti metode opisane u poglavlju 3.

Dakle, provjeravamo normalnost reziduala svake regresije i to grafičkim prikazom reziduala, ili qqplot() funkcijom u programu R, no možemo isto provjeriti i funkcijom lillie.test() koja provodi Lillieforsovu inačicu Kolmogorov-Smirnovljevog testa normalnosti uzorka.

```
> lillie.test(reziduali)
```

```
Lilliefors (Kolmogorov-Smirnov) normality test
```

```
data: reziduali  
D = 0.069086, p-value = 0.1859
```

Lillieforsov test normalnosti nam kaže da ne možemo odbaciti nultu hipotezu ( $H_0$  : podaci su normalno distribuirani) zbog toga što je  $p$ -vrijednost = 0,1859 veća od razine značajnosti 5%. Dakle, zaključujemo da su podaci, odnosno reziduali, normalno distribuirani. Sada možemo prijeći na testiranje hipoteza.

Pri testiranju prve hipoteze ćemo provesti linearnu regresiju nad običnim WGI pokazateljem i prosječnim Avg\_% pokazateljem. Uočiti ćemo da oba pokazatelja daju jako sličan rezultat pa je dovoljno promatrati samo jedan od njih. U ostalim hipotezama ćemo koristiti samo pokazatelj za 2019. jer ćemo vidjeti da za njih imamo malo bolji koeficijent determinacije  $\bar{R}^2$ . Rezultati linearnih regresija za pokazatelje koji promatraju raspon 2017. -

2019. bit će navedeni u dodatku A jer su dokumenti korišteni kod kreiranja OBI-ja često iz ranijih godina. Linearnu regresiju provodimo ugrađenom funkcijom `lm()` u programu R.

## Hipoteza 1

Želimo testirati ima li WGI pokazatelj „glas i odgovornost“ pozitivan utjecaj na OBI. Provest ćemo linearnu regresiju za model

$$OBI_{2019} = \beta_0 + \beta_1 \cdot VA .$$

```
> model = lm(OBI_2019 ~ VA)
> summary(model)
```

Call:

```
lm(formula = OBI_2019 ~ VA)
```

Residuals:

Min	1Q	Median	3Q	Max
-39.641	-10.368	-0.386	10.625	45.983

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	47.884	1.555	30.78	<2e-16 ***
VA	17.773	1.721	10.33	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

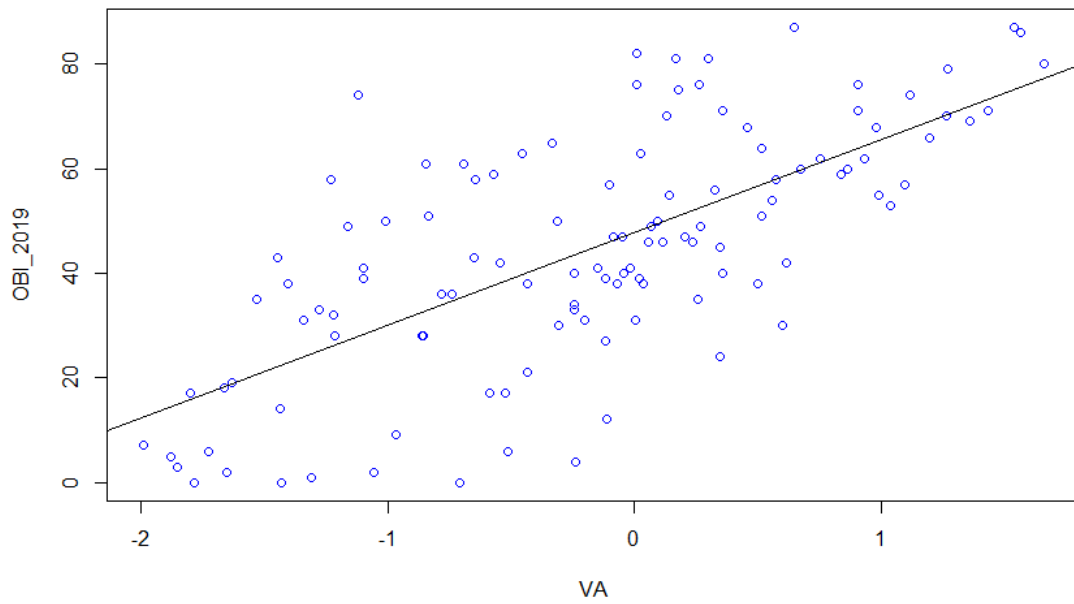
Residual standard error: 16.48 on 115 degrees of freedom

Multiple R-squared: 0.4812, Adjusted R-squared: 0.4767

F-statistic: 106.7 on 1 and 115 DF, p-value: < 2.2e-16

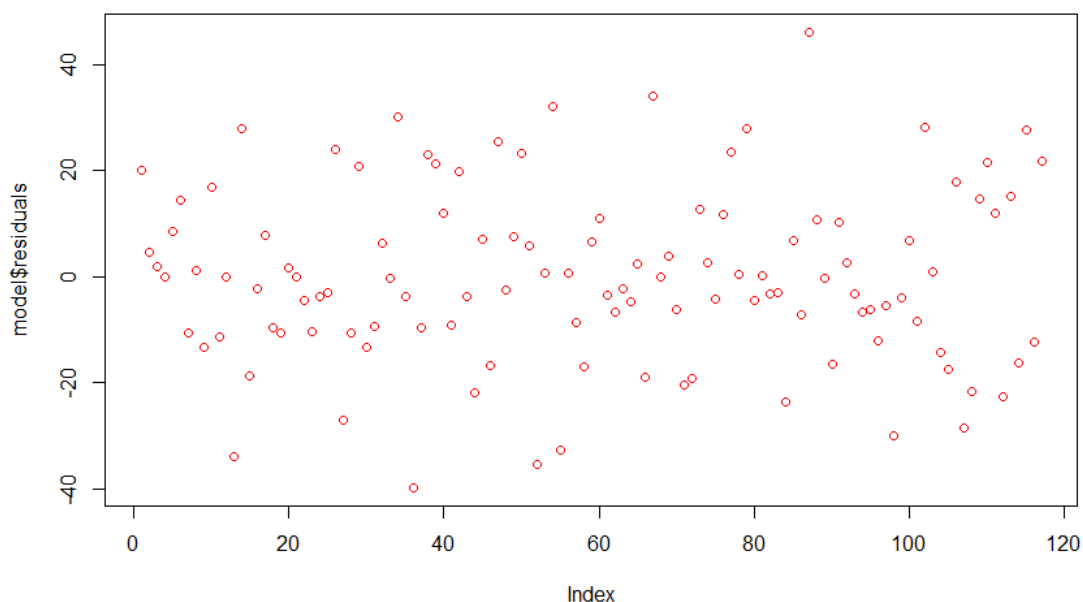
Iz rezultata vidimo da je  $p$ -vrijednost dovoljno mala ( $< 0,05$ ) pa možemo zaključiti da pokazatelj VA iz 2019. ima statistički značajan utjecaj na OBI. Iz procijenjenih parametara modela vidimo i da ima pozitivan utjecaj na OBI ( $\beta_1 = 17,773$ , uz slobodni član  $\beta_0 = 47,884$ ).

Pogledajmo i grafički prikaz podataka te pravac dobiven linearnom regresijom. Vidimo da za ovaj model koeficijent determinacije  $R^2$  iznosi 0,4812, dok prilagođeni koeficijent determinacije  $\bar{R}^2$  iznosi 0,4767. Dakle, model dobro opisuje 48,12% podataka.



Slika 4.1: Podaci (VA, OBI\_2019) i pravac dobiven modelom

Provjerimo zadovoljavaju li reziduali pretpostavke normalne distribuiranosti. Ovo ćemo provjeriti vizualnom reprezentacijom reziduala:



Slika 4.2: Raspršenost reziduala modela

Vidimo da su reziduali nasumično raspršeni što navodi na normalnu distribuiranost reziduala. Provjerimo to još i Lillieforsovom inačicom Kolmogorov-Smirnovljeva testa.

```
> lillie.test(model$residuals)
```

Lilliefors (Kolmogorov-Smirnov) normality test

```
data: model$residuals
D = 0.075272, p-value = 0.1067
```

Lillieforsov test normalnosti nam kaže da ne možemo odbaciti nultu hipotezu ( $H_0$  : podaci su normalno distribuirani) zbog toga što je  $p$ -vrijednost = 0,1067 veća od razine značajnosti 5%. Dakle, zaključujemo da su podaci, odnosno reziduali, normalno distribuirani.

Pogledajmo sada rezultate za Avg\_VA: Promatramo model

$$OBI\_2019 = \beta_0 + \beta_1 \cdot Avg\_VA .$$

```
> model = lm(OBI_2019 ~ Avg_VA)
```



```
> summary(model)
```

```
Call:
```

```
lm(formula = OBI_2019 ~ Avg_VA)
```

```
Residuals:
```

```
      Min       1Q   Median       3Q      Max
-39.364 -10.312  -1.449   10.545  46.107
```

```
Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   47.534      1.549   30.68  <2e-16 ***
Avg_VA        17.800      1.724   10.32  <2e-16 ***
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

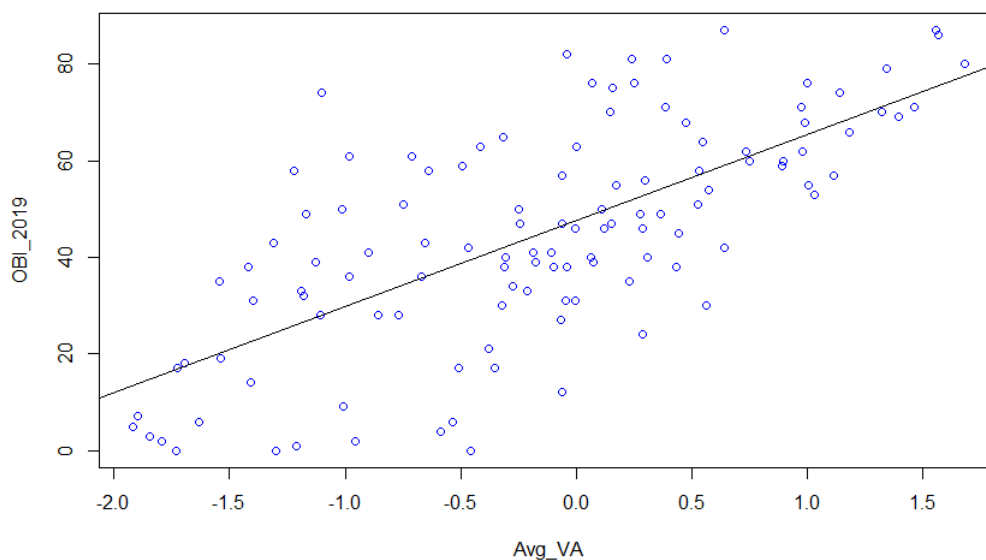
```
Residual standard error: 16.48 on 115 degrees of freedom
```

```
Multiple R-squared:  0.4809, Adjusted R-squared:  0.4764
```

```
F-statistic: 106.5 on 1 and 115 DF, p-value: < 2.2e-16
```

Linearnom regresijom dobijemo da parametri modela glase  $\beta_0 = 47,534$  i  $\beta_1 = 17,800$ , što je veoma slično rezultatima dobivenim kod pokazatelja VA iz 2019. Vidimo da je  $p$ -vrijednost ista kao u prethodnom modelu ( $< 0,05$ ) te zaključujemo da i Avg\_VA ima statistički značajan i to pozitivan utjecaj na OBI. Vizualnu reprezentaciju podataka i regresijskog pravca vidimo na slici 4.3.

Koeficijent determinacije  $R^2$  sada iznosi 0,4809, a prilagođeni koeficijent determinacije  $\bar{R}^2$  0,4764. Model dobro opisuje 48,09% podataka. Iz koeficijenata determinacije i procijenjenih parametara možemo zaključiti da oba parametra (obični i prosječni) jako slično opisuju OBI te je dovoljno promatrati samo jedan.



Slika 4.3: Podaci (Avg\_VA, OBI\_2019) i pravac dobiven modelom

Lillieforsovom inačicom Kolmogorov-Smirnovljeva testa provjerimo jesu li reziduali i dalje normalno distribuirani.

```
> lillie.test(model$residuals)
```

Lilliefors (Kolmogorov-Smirnov) normality test

```
data: model$residuals
```

```
D = 0.06633, p-value = 0.2346
```

Iz dobivene  $p$ -vrijednosti = 0,2346 > 0,05 zaključujemo da ne možemo odbaciti nultu hipotezu da su podaci normalno distribuirani.

## Hipoteza 2

Želimo testirati ima li WGI pokazatelj „politička stabilnost i odsutnost nasilja/terorizma“ pozitivan utjecaj na OBI. Provest ćemo linearnu regresiju za model

$$OBI_{2019} = \beta_0 + \beta_1 \cdot PS .$$

```
> #Political Stability
```

```
> model = lm(OBI_2019 ~ PS)
> summary(model)
```

```
Call:
lm(formula = OBI_2019 ~ PS)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-57.27 -12.23  -1.14   12.88  43.62
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   49.209      1.954   25.185 < 2e-16 ***
PS             12.937      2.028    6.378 3.87e-09 ***
```

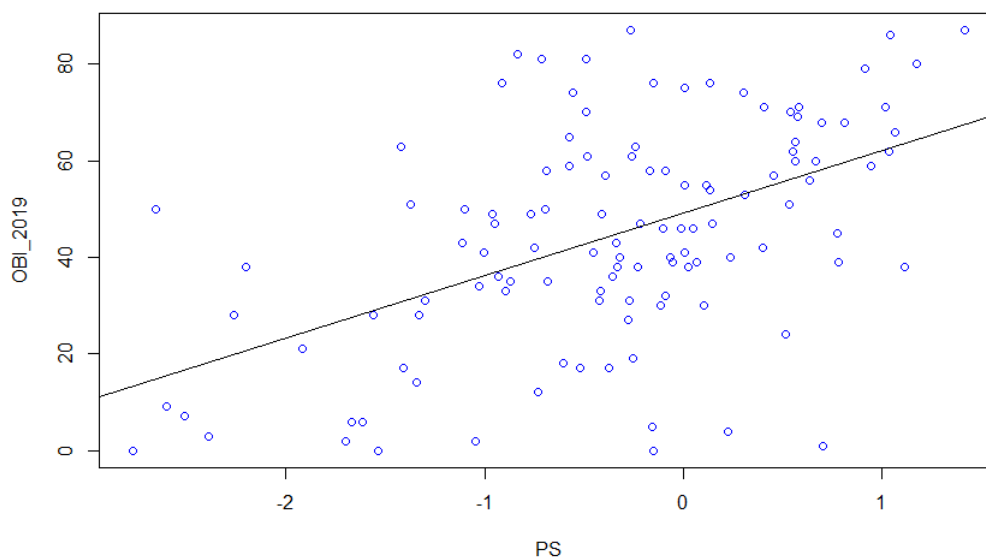
```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 19.66 on 115 degrees of freedom
Multiple R-squared:  0.2613, Adjusted R-squared:  0.2549
F-statistic: 40.68 on 1 and 115 DF,  p-value: 3.873e-09
```

Iz rezultata vidimo da je  $p$ -vrijednost dovoljno mala ( $< 0,05$ ) pa možemo zaključiti da pokazatelj PS iz 2019. ima statistički značajan utjecaj na OBI. Iz procijenjenih parametara modela vidimo i da ima pozitivan utjecaj na OBI ( $\beta_1 = 12,937$ , uz slobodni član  $\beta_0 = 49,209$ ).

Pogledajmo i grafički prikaz podataka te pravac dobiven linearnom regresijom na slici 4.4.

Vidimo da za ovaj model koeficijent determinacije  $R^2$  iznosi 0,2613, dok prilagođeni koeficijent determinacije  $\bar{R}^2$  iznosi 0,2549. Dakle, model dobro opisuje samo 25,49% podataka, što znači da ne opisuje dobro dovoljno podataka.



Slika 4.4: Podaci (PS, OBI\_2019) i pravac dobiven modelom

Provjerimo zadovoljavaju li reziduali pretpostavke normalne distribuiranosti. Ovo ćemo provjeriti vizualnom reprezentacijom reziduala:

Vidimo da su reziduali nasumično raspršeni što navodi na normalnu distribuiranost reziduala. Provjerimo to još i Lillieforsovom inačicom Kolmogorov-Smirnovljeva testa.

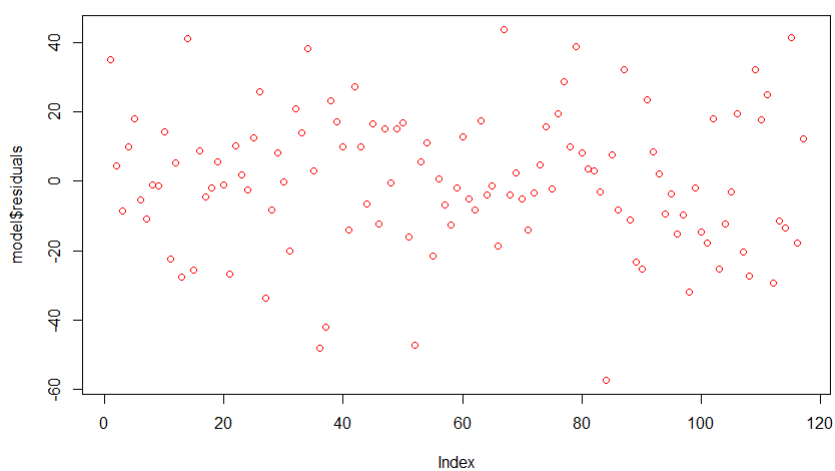
```
> lillie.test(model$residuals)
```

Lilliefors (Kolmogorov-Smirnov) normality test

```
data: model$residuals  
D = 0.042173, p-value = 0.8771
```

Lillieforsov test normalnosti nam kaže da ne možemo odbaciti nultu hipotezu ( $H_0$  : podaci su normalno distribuirani) zbog toga što je  $p$ -vrijednost = 0,8771 veća od razine značajnosti 5%. Dakle, zaključujemo da su podaci, odnosno reziduali, normalno distribuirani.

Linearna regresija za prosječni pokazatelj Avg\_PS nalazi se u dodatku A.



Slika 4.5: Raspršenost reziduala modela

### Hipoteza 3

Želimo testirati ima li WGI pokazatelj „učinkovitost vlade“ pozitivan utjecaj na OBI. Provest ćemo linearnu regresiju za model

$$OBI_{2019} = \beta_0 + \beta_1 \cdot GE .$$

```
> #Government Effectiveness
> model = lm(OBI_2019 ~ GE)
> summary(model)
```

Call:

```
lm(formula = OBI_2019 ~ GE)
```

Residuals:

Min	1Q	Median	3Q	Max
-59.015	-10.610	0.262	10.196	36.673

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	47.999	1.569	30.59	<2e-16 ***
GE	17.046	1.675	10.18	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

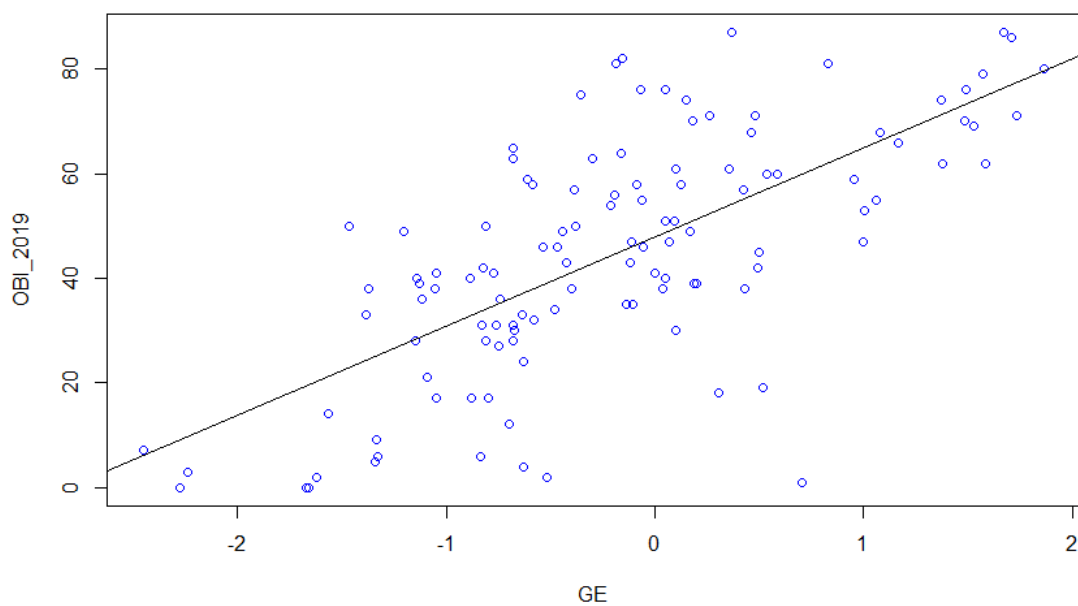
Residual standard error: 16.6 on 115 degrees of freedom

Multiple R-squared: 0.4738, Adjusted R-squared: 0.4693

F-statistic: 103.6 on 1 and 115 DF, p-value: &lt; 2.2e-16

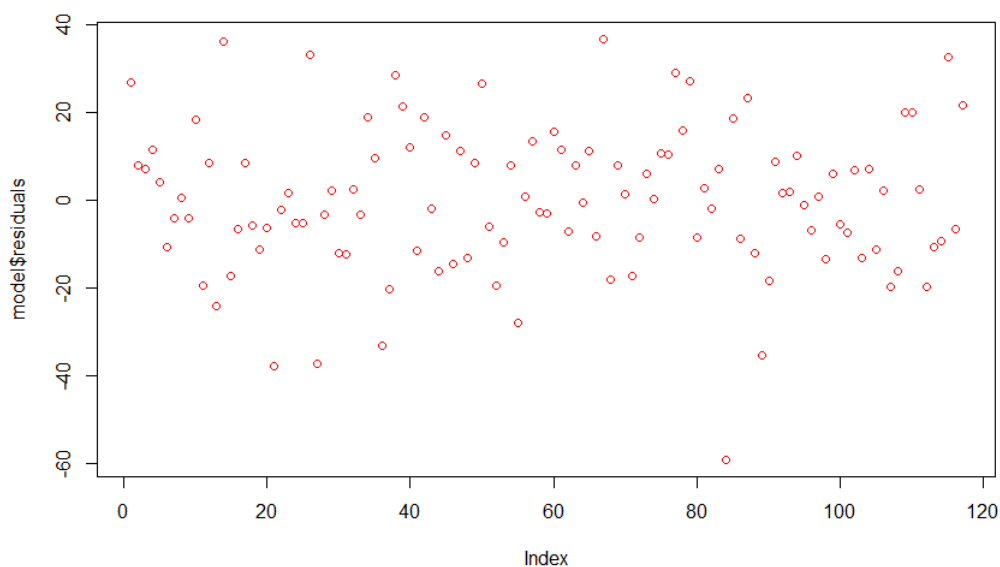
Iz rezultata vidimo da je  $p$ -vrijednost dovoljno mala ( $< 0,05$ ) pa možemo zaključiti da pokazatelj GE iz 2019. ima statistički značajan utjecaj na OBI. Iz procijenjenih parametara modela vidimo i da ima pozitivan utjecaj na OBI ( $\beta_1 = 17,046$ , uz slobodni član  $\beta_0 = 47,999$ ).

Pogledajmo i grafički prikaz podataka te pravac dobiven linearnom regresijom.



Slika 4.6: Podaci (GE, OBI\_2019) i pravac dobiven modelom

Vidimo da za ovaj model koeficijent determinacije  $R^2$  iznosi 0,4738, dok prilagođeni koeficijent determinacije  $\bar{R}^2$  iznosi 0,4693. Dakle, model dobro opisuje samo 47,38% podataka. Provjerimo zadovoljavaju li reziduali pretpostavke normalne distribuiranosti. Ovo ćemo provjeriti vizualnom reprezentacijom reziduala na slici 4.7.



Slika 4.7: Raspršenost reziduala modela

Vidimo da su reziduali nasumično raspršeni što navodi na normalnu distribuiranost reziduala. Provjerimo to još i Lillieforsovom inačicom Kolmogorov-Smirnovljeva testa.

```
> lillie.test(model$residuals)
```

Lilliefors (Kolmogorov-Smirnov) normality test

```
data: model$residuals  
D = 0.052051, p-value = 0.6097
```

Lillieforsov test normalnosti nam kaže da ne možemo odbaciti nultu hipotezu ( $H_0$  : podaci su normalno distribuirani) zbog toga što je  $p$ -vrijednost = 0,6097 veća od razine značajnosti 5%. Dakle, zaključujemo da su podaci, odnosno reziduali, normalno distribuirani.

Linearna regresija za prosječni pokazatelj Avg\_GE nalazi se u dodatku A.

## Hipoteza 4

Želimo testirati ima li WGI pokazatelj „kvaliteta regulatornog okvira“ pozitivan utjecaj na OBI.

Provest ćemo linearnu regresiju za model

$$OBI_{2019} = \beta_0 + \beta_1 \cdot RQ .$$

```
> #Regulatory Quality
> model = lm(OBI_2019 ~ RQ)
> summary(model)
```

Call:

```
lm(formula = OBI_2019 ~ RQ)
```

Residuals:

Min	1Q	Median	3Q	Max
-58.718	-9.534	-0.649	8.409	37.246

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	47.075	1.454	32.38	<2e-16 ***
RQ	18.609	1.609	11.56	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.56 on 115 degrees of freedom

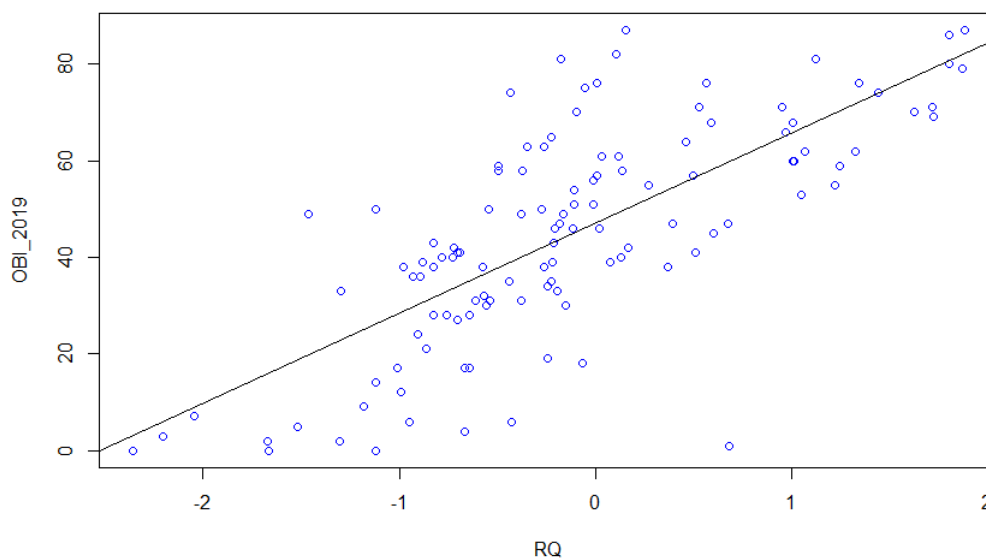
Multiple R-squared: 0.5376, Adjusted R-squared: 0.5336

F-statistic: 133.7 on 1 and 115 DF, p-value: < 2.2e-16

Iz rezultata vidimo da je  $p$ -vrijednost dovoljno mala ( $< 0,05$ ) pa možemo zaključiti da pokazatelj RQ iz 2019. ima statistički značajan utjecaj na OBI. Iz procijenjenih parametara modela vidimo i da ima pozitivan utjecaj na OBI ( $\beta_1 = 18,609$ , uz slobodni član  $\beta_0 = 47,075$ ).

Pogledajmo i grafički prikaz podataka te pravac dobiven linearnom regresijom.





Slika 4.8: Podaci (RQ, OBI\_2019) i pravac dobiven modelom

Vidimo da za ovaj model koeficijent determinacije  $R^2$  iznosi 0,5376, dok prilagođeni koeficijent determinacije  $\bar{R}^2$  iznosi 0,5336. Dakle, model dobro opisuje 53,76% podataka.

Provjerimo zadovoljavaju li reziduali pretpostavke, tj. jesu li normalno distribuirani. Ovo ćemo provjeriti vizualnom reprezentacijom reziduala na slici 4.9.

Vidimo da su reziduali nasumično raspršeni što navodi na normalnu distribuiranost reziduala. Provjerimo to još i Lillieforsovom inačicom Kolmogorov-Smirnovljeva testa.

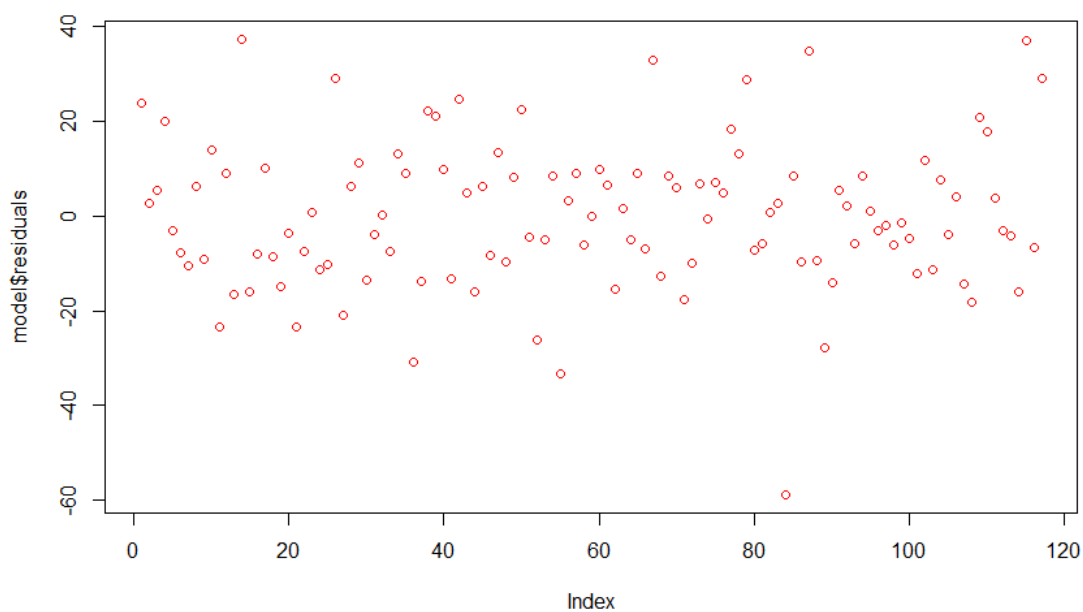
```
> lillie.test(model$residuals)
```

Lilliefors (Kolmogorov-Smirnov) normality test

```
data: model$residuals
D = 0.069086, p-value = 0.1859
```

Lillieforsov test normalnosti nam kaže da ne možemo odbaciti nultu hipotezu ( $H_0$  : podaci su normalno distribuirani) zbog toga što je  $p$ -vrijednost = 0,1859 veća od razine značajnosti 5%. Dakle, zaključujemo da su podaci, odnosno reziduali, normalno distribuirani.

Linearna regresija za prosječni pokazatelj Avg\_RQ nalazi se u dodatku A.



Slika 4.9: Raspršenost reziduala modela

### Hipoteza 5

Želimo testirati ima li WGI pokazatelj „vladavina prava“ pozitivan utjecaj na OBI. Provest ćemo linearnu regresiju za model

$$OBI_{2019} = \beta_0 + \beta_1 \cdot RL .$$

```
> #Rule of Law
> model = lm(OBI_2019 ~ RL)
> summary(model)
```

```
Call:
lm(formula = OBI_2019 ~ RL)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-59.197 -10.995  -0.534   9.321  43.641
```

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	48.679	1.718	28.339	< 2e-16	***
RL	15.683	1.829	8.576	5.29e-14	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

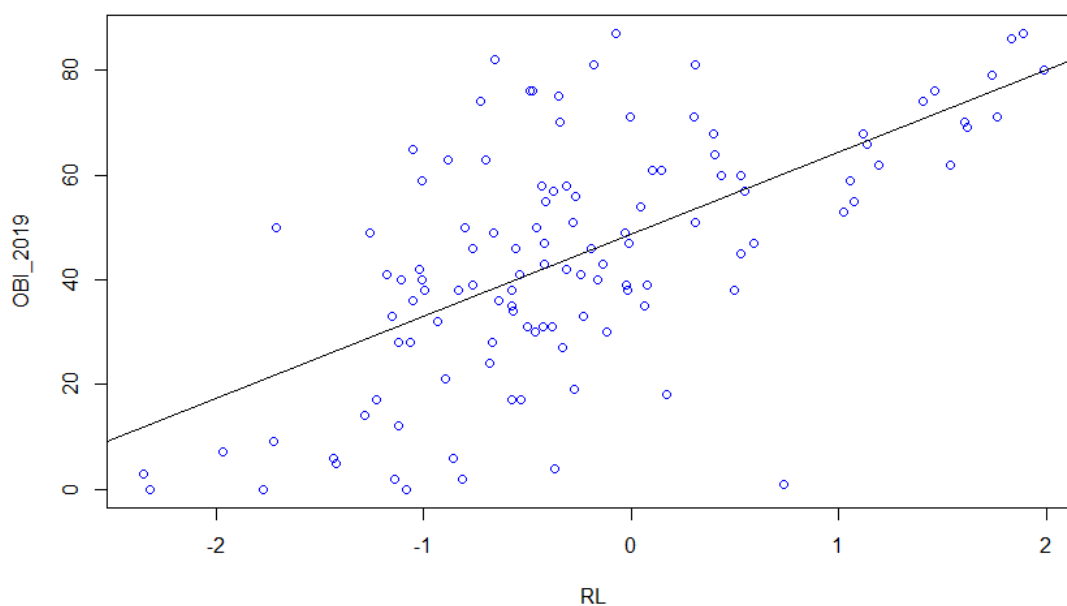
Residual standard error: 17.87 on 115 degrees of freedom

Multiple R-squared: 0.3901, Adjusted R-squared: 0.3848

F-statistic: 73.54 on 1 and 115 DF, p-value: 5.292e-14

Iz rezultata vidimo da je  $p$ -vrijednost dovoljno mala ( $< 0,05$ ) pa možemo zaključiti da pokazatelj RL iz 2019. ima statistički značajan utjecaj na OBI. Iz procijenjenih parametara modela vidimo i da ima pozitivan utjecaj na OBI ( $\beta_1 = 15,683$ , uz slobodni član  $\beta_0 = 478,679$ ).

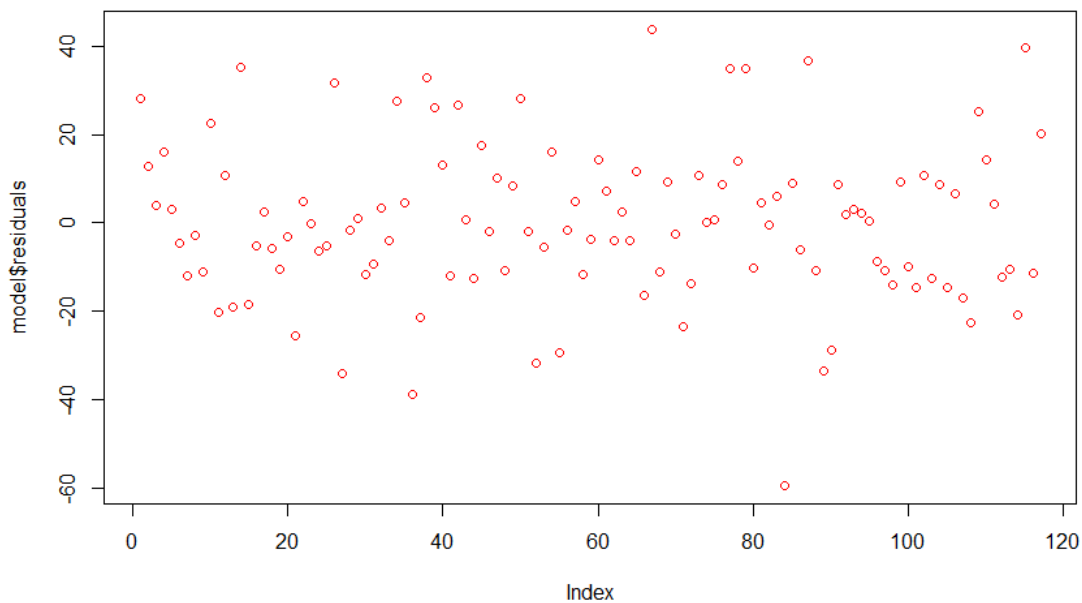
Pogledajmo i grafički prikaz podataka te pravac dobiven linearnom regresijom.



Slika 4.10: Podaci (RL, OBI\_2019) i pravac dobiven modelom

Vidimo da za ovaj model koeficijent determinacije  $R^2$  iznosi 0,5376, dok prilagođeni koeficijent determinacije  $\bar{R}^2$  iznosi 0,5336. Dakle, model dobro opisuje 53,76% podataka.

Provjerimo zadovoljavaju li reziduali pretpostavke normalne distribuiranosti. Ovo ćemo provjeriti vizualnom reprezentacijom reziduala na slici 4.11.



Slika 4.11: Raspršenost reziduala modela

Vidimo da su reziduali nasumično raspršeni što navodi na normalnu distribuiranost reziduala. Provjerimo to još i Lillieforsovom inačicom Kolmogorov-Smirnovljeva testa.

```
> lillie.test(model$residuals)
```

Lilliefors (Kolmogorov-Smirnov) normality test

```
data: model$residuals  
D = 0.05969, p-value = 0.3878
```

Lillieforsov test normalnosti nam kaže da ne možemo odbaciti nultu hipotezu ( $H_0$ : podaci su normalno distribuirani) zbog toga što je  $p$ -vrijednost = 0,3878 veća od razine značajnosti 5%. Dakle, zaključujemo da su podaci, odnosno reziduali, normalno distribuirani.

Linearna regresija za prosječni pokazatelj Avg\_RL nalazi se u dodatku A.

## Hipoteza 6

Želimo testirati ima li WGI pokazatelj „kontrola korupcije“ pozitivan utjecaj na OBI. Provest ćemo linearnu regresiju za model

$$OBI_{2019} = \beta_0 + \beta_1 \cdot CC .$$

```
> #Control of Corruption
> model = lm(OBI_2019 ~ CC)
> summary(model)
```

```
Call:
lm(formula = OBI_2019 ~ CC)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-60.174 -11.431   0.788   9.584  45.526
```

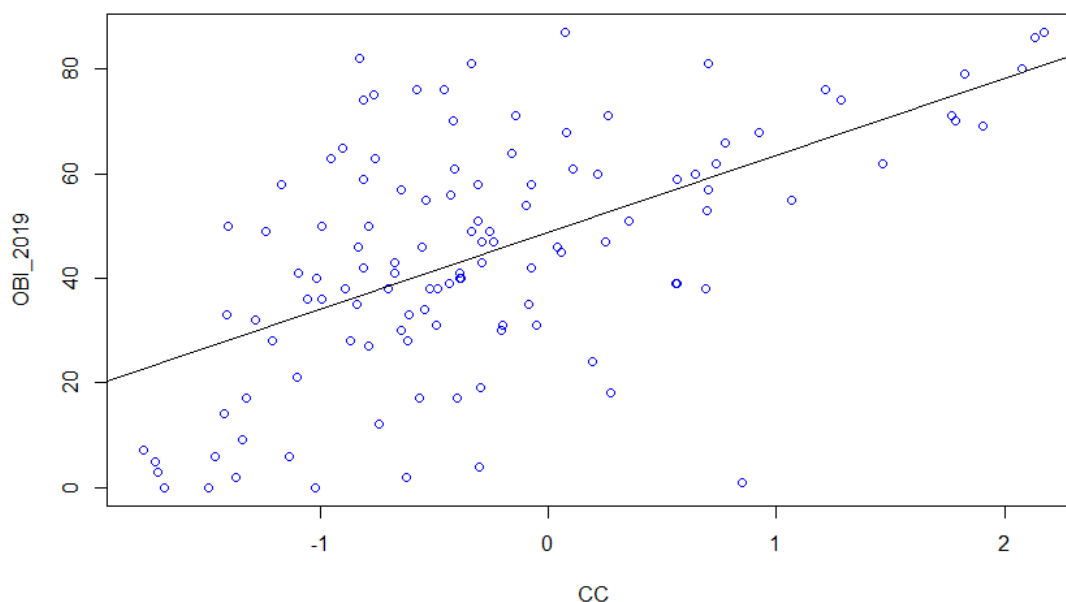
```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  48.659      1.810  26.889 < 2e-16 ***
CC           14.735      1.952   7.549 1.13e-11 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 18.71 on 115 degrees of freedom
Multiple R-squared:  0.3313, Adjusted R-squared:  0.3255
F-statistic: 56.99 on 1 and 115 DF, p-value: 1.129e-11
```

Iz rezultata vidimo da je  $p$ -vrijednost dovoljno mala ( $< 0,05$ ) pa možemo zaključiti da pokazatelj CC iz 2019. ima statistički značajan utjecaj na OBI. Iz procijenjenih parametara modela vidimo i da ima pozitivan utjecaj na OBI ( $\beta_1 = 14,735$ , uz slobodni član  $\beta_0 = 48,659$ ).

Pogledajmo i grafički prikaz podataka te pravac dobiven linearnom regresijom.



Slika 4.12: Podaci (CC, OBI.2019) i pravac dobiven modelom

Vidimo da za ovaj model koeficijent determinacije  $R^2$  iznosi 0,3313, dok prilagođeni koeficijent determinacije  $\bar{R}^2$  iznosi 0,3255. Dakle, model dobro opisuje 33,13% podataka. Provjerimo zadovoljavaju li reziduali pretpostavke normalne distribuiranosti. Ovo ćemo provjeriti vizualnom reprezentacijom reziduala na slici 4.13.

Vidimo da su reziduali nasumično rasršeni što navodi na normalnu distribuiranost reziduala. Provjerimo to još i Lillieforsovom inačicom Kolmogorov-Smirnovljeva testa.

```
> lillie.test(model$residuals)
```

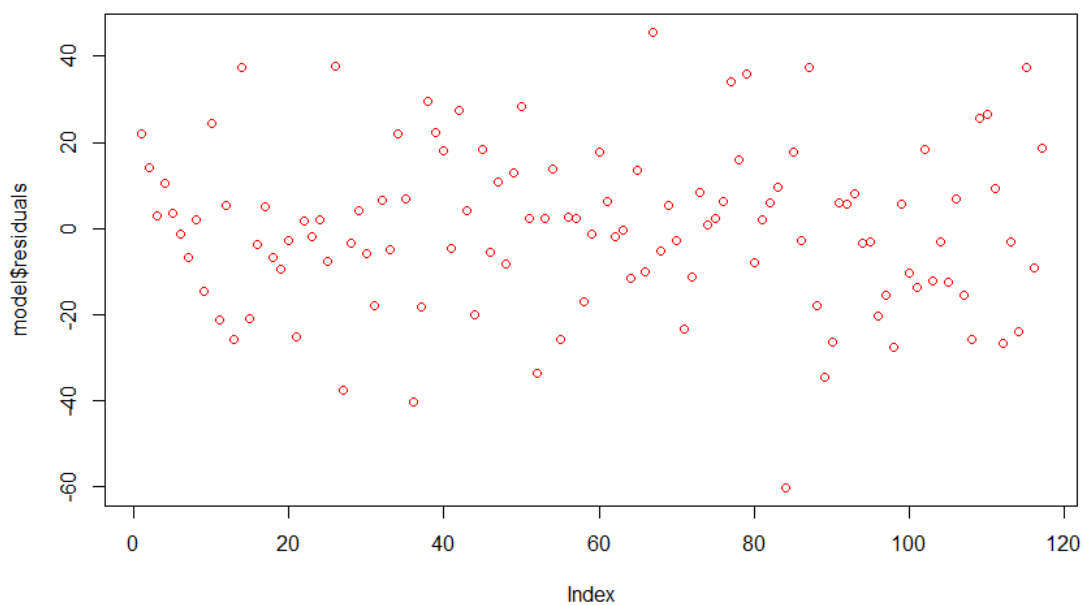
```
Lilliefors (Kolmogorov-Smirnov) normality test
```

```
data: model$residuals
```

```
D = 0.073813, p-value = 0.122
```

Lillieforsov test normalnosti nam kaže da ne možemo odbaciti nultu hipotezu ( $H_0$  : podaci su normalno distribuirani) zbog toga što je  $p$ -vrijednost = 0,122 veća od razine značajnosti 5%. Dakle, zaključujemo da su podaci, odnosno reziduali, normalno distribuirani.

Linearna regresija za prosječni pokazatelj Avg\_CC nalazi se u dodatku A.



Slika 4.13: Raspršenost reziduala modela

### Hipoteza 7

WGI pokazatelj „glas i odgovornost“ ima najveći utjecaj na OBI od svih WGI pokazatelja. Promatrat ćemo model koji se sastoji od svih nezavisnih varijabli. Tada model glasi

$$OBI = \beta_0 + \beta_1 \cdot VA + \beta_2 \cdot RQ + \beta_3 \cdot GE + \beta_4 \cdot PS + \beta_5 \cdot RL + \beta_6 \cdot CC .$$

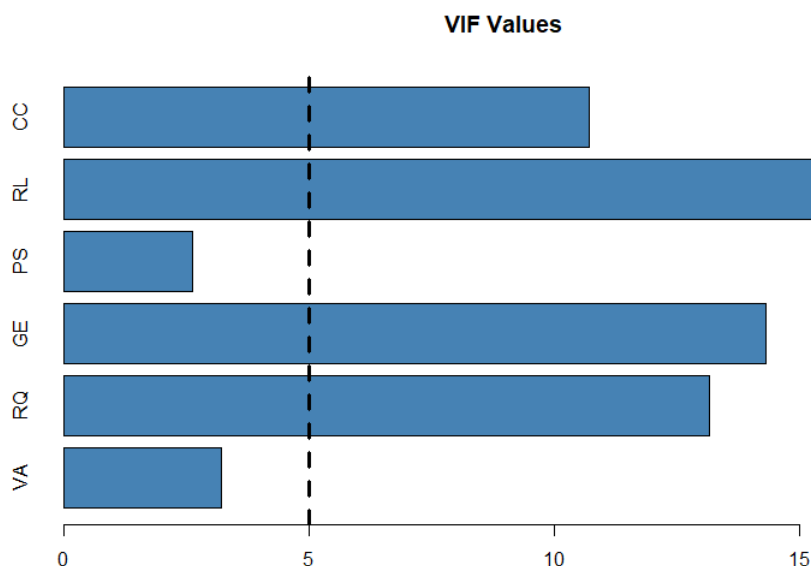
Prvo provjerimo matricu koreliranosti za nezavisne varijable.

	VA	PS	GE	RQ	RL	CC
VA	1,0000					
PS	0,6921	1,0000				
GE	0,7367	0,7482	1,0000			
RQ	0,7912	0,7317	0,9459	1,0000		
RL	0,7712	0,7674	0,9459	0,9371	1,0000	
CC	0,7822	0,7490	0,9129	0,8936	0,9462	1,0000

Tablica 4.1: Matrica korelacijskih koeficijenata nezavisnih varijabli

Iz tablice 4.1 vidimo da su sve varijable pozitivno korelirane. Najmanja korelacija s ostalim nezavisnim varijablama se vidi kod varijable VA ("glas i odgovornost"). Najveći koeficijent korelacije imaju CC ("kontrola korupcije") i RL ("vladavina prava") koji iznosi 0,9462. Možemo primijetiti i da varijabla GE ("učinkovitost vlade") ima visoke koeficijente korelacije s varijablama RQ ("kvaliteta regulatornog okvira"), RL i CC ("kontrola korupcije").

Međusobnu koreliranost nezavisnih varijabli možemo i vizualno predočiti grafom faktora inflacije varijance (VIF, eng. *Variation Inflation Factor*). VIF mjeri koreliranost i jačinu korelacije među nezavisnim varijablama modela. Ukoliko poprimi vrijednost između 1 i 5, znači da postoji korelacija s ostalim nezavisnim varijablama, no zanemariva je. Ako pak prelazi 5, moguća je jaka korelacija s ostalim varijablama. U ovom slučaju su procijenjeni parametri i  $p$ -vrijednosti nepouzdana. Vidimo da VIF graf odgovara matrici korelacije:



Slika 4.14: Graf faktora inflacije varijance

Zbog visoke koreliranosti varijabli GE, RQ, RL i CC (koeficijent korelacije  $\geq 90\%$ ), nema smisla promatrati potpuni model. Budući da za niti jednu od varijabli ne možemo reći da je bolja od drugih, možemo promatrati njihov prosjek kao novu varijablu.

Model tada glasi

$$OBI = \beta_0 + \beta_1 \cdot VA + \beta_2 \cdot PS + \beta_3 \cdot Z,$$

gdje varijabla  $Z$  predstavlja prosjek varijabli GE, RQ, RL i CC.

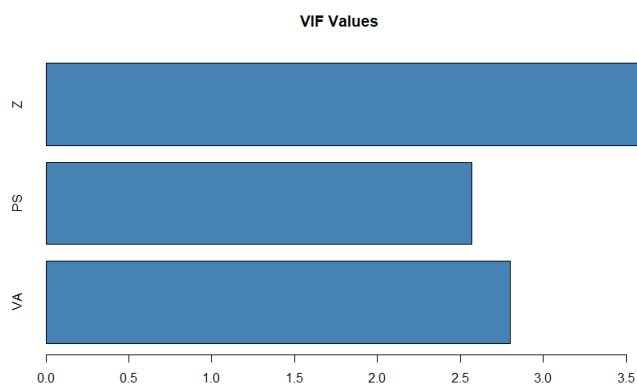
Promotrimo sada matricu korelacije za nezavisne varijable novog modela u tablici 4.2.



	VA	PS	Z
VA	1,0000		
PS	0,6921	1,0000	
Z	0,7910	0,79652	1,0000

Tablica 4.2: Matrica korelacijskih koeficijenata nezavisnih varijabli novog modela

Iz nove matrice korelacije vidimo da su varijable i dalje pozitivno korelirane, što je i očekivano, ali su koeficijenti korelacije sada manji (ili jednaki) 80%. Za varijablu Z, koja je prosjek varijabli s visokim koeficijentima korelacije, se i dalje postižu najveći korelacijski koeficijenti. No, iz VIF grafa sada vidimo da je za sve varijable vrijednost strogo manja od pet.



Slika 4.15: Graf faktora inflacije varijance novog modela

Provođenjem linearne regresije na ovakvom modelu dobivamo niže prikazane rezultate:

```
> model = lm(OBI_2019 ~ VA + PS + Z)
> summary(model)
```

```
Call:
lm(formula = OBI_2019 ~ VA + PS + Z)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-37.471  -9.199  -0.370   8.335  42.154
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	47.995	1.584	30.306	< 2e-16 ***
VA	11.646	2.772	4.202	5.31e-05 ***
PS	-2.791	2.622	-1.064	0.28943
Z	10.327	3.166	3.262	0.00146 **

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.86 on 113 degrees of freedom

Multiple R-squared: 0.5276, Adjusted R-squared: 0.515

F-statistic: 42.06 on 3 and 113 DF, p-value: &lt; 2.2e-16

Iz rezultata vidimo da  $p$ -vrijednost varijable PS nije dovoljno mala ( $0,28943 > 0,05$ ) te zaključujemo da ona nije statistički značajna u modelu. Izbacivanjem varijable PS i ponovnim provođenjem i linearne regresije dobijemo sljedeći model

$$OBI = 48,502 + 11,016 \cdot VA + 8,633 \cdot Z,$$

u kojem su obje varijable statistički značajne. Prilagođeni koeficijent determinacije iznosi 0,5145. Vidimo da se OBI za 1 bod poboljšanja mjerenog glasa i odgovornosti poveća za čak 11,016, dok se u slučaju povećanja varijable Z poveća za 8,633, manje nego za VA.

## Hipoteza 8

Spomenuli smo da su Andreula, Chong i Guillen [9] u svom radu kreirali novi agregirani pokazatelj kojim mjere kvalitetu institucije tako da su gledali prosjek svih WGI pokazatelja. Slijedom njihovog rada, promatrat ćemo model

$$OBI = \beta_0 + \beta_1 \cdot WGI_{aggregate}.$$

```
> model = lm(OBI_2019 ~ WGI_aggregate)
> summary(model)
```

Call:

```
lm(formula = OBI_2019 ~ WGI_aggregate)
```

Residuals:

Min	1Q	Median	3Q	Max
-55.471	-10.256	-1.224	9.417	40.426

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	49.033	1.589	30.86	<2e-16 ***
WGI_aggregate	18.909	1.847	10.24	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.55 on 115 degrees of freedom

Multiple R-squared: 0.4769, Adjusted R-squared: 0.4723

F-statistic: 104.8 on 1 and 115 DF, p-value: &lt; 2.2e-16

Iz rezultata vidimo da agregatni WGI ima statistički značajan utjecaj na OBI ( $p$ -vrijednost < 0,05) te je  $R^2 = 0,4769$ . Iz koeficijenta smjera pravca  $\beta_1 = 18,909$  vidimo da ima pozitivan utjecaj na OBI. Dakle, ako agregirani pokazatelj kvalitete institucije poraste (to jest kvaliteta institucije se poboljša), povećat će se i OBI. Dakle, transparentnost proračuna će biti bolja.

## 4.2 Interpretacija rezultata

Iz provedenih regresija i dobivenih rezultata vidimo da svi svjetski pokazatelji upravljanja imaju statistički značajan i to pozitivan utjecaj na OBI. Dakle, ukoliko zemlji poraste neki od svjetskih pokazatelja upravljanja, to jest određeni aspekt upravljanja zemljom se poboljša, možemo zaključiti da će se povećati i OBI. To nam govori da bolje upravljanje zemljom vodi do veće transparentnosti državnog proračuna. Ovaj rezultat je u skladu s prethodno spomenutim analizama. Iako su u ostalim radovima autori proučavali i utjecaj transparentnosti proračuna na kvalitetu upravljanja, iz naših rezultata vidimo da upravljanje utječe na transparentnost. Kao iznimku navodimo pokazatelj političke stabilnosti i odsutnosti nasilja/terorizma (PS). Za političku stabilnost i odsutnost nasilja/terorizma Cimpoera [16] i Albassam [7] zaključuju da je negativno koreliran s OBI-jem te da nema statistički značajan utjecaj, što se ne slaže s rezultatima dobivenim u ovom poglavlju. Njihove analize provedene su na podacima iz ranijih godina (promatrali su podatke za 2006., 2008., 2010. i 2012.). Preciznije, za 2010. i 2012. zaključuju da nema statistički značajan utjecaj. Logističkom regresijom su promatrali utjecaj OBI-ja na pokazatelje te došli do zaključka. Svoje rezultate možemo smatrati opravdanim jer intuitivno možemo zaključiti da bolja politička stabilnost zemlje dovodi do uređenijeg upravljanja i pružanja više informacija javnosti u smislu transparentnijeg proračuna. Također uočavamo da gledajući WGI pokazatelj samo za jednu godinu ili kao prosjek godina koje prethode promatranom OBI-ju

dobijemo jako slične rezultate. Dakle, WGI pokazatelji se ne mijenjaju drastično kroz vrijeme.

Kod testiranja sedme hipoteze smo vidjeli da ne možemo jednostavno zaključiti koji svjetski pokazatelj najviše utječe na OBI. Zaključili smo da su svi pokazatelji jako korelirani. Zato ne možemo reći da primjerice pokazatelj kontrole korupcije najznačajnije utječe na OBI budući se njegovom promjenom mijenja i pokazatelj vladavine prava. Vidimo da je pokazatelj glas i odgovornost najmanje koreliran s ostalim pokazateljima te da pokazatelj političke stabilnosti u višestrukom linearnom modelu nije statistički značajan.

Zbog velike koreliranosti svjetskih pokazatelja ima smisla gledati njihov prosjek kao indikator kvalitete institucije. Indikator kao takav koriste Andreula, Chong i Guillen [9]. Zaključuju da agregirani svjetski pokazatelj pozitivno utječe na fiskalnu transparentnost. Kao što smo spomenuli kod definicije transparentnosti proračuna, fiskalna transparentnost je širi pojam od proračunske transparentnosti. No iz naših rezultata vidimo da tako kreiran pokazatelj kvalitete institucije pozitivno utječe i na proračunsku transparentnost. Možemo zaključiti da bolje upravljanje i provođenje vlasti vodi do transparentnijeg proračuna.

## Poglavlje 5

### Zaključak

Promatranjem odnosa između svjetskih pokazatelja upravljanja (WGI) i indeksa transparentnosti proračuna (OBI) vidimo da svi WGI-jevi imaju statistički značajan pozitivan utjecaj na OBI. Ovaj zaključak donosimo provođenjem linearne regresije na podacima. Dakle, poboljšanjem nekog aspekta upravljanja, državni proračun je transparentniji. Možemo vidjeti i da su svi WGI pokazatelji međusobno snažno korelirani, s najmanjim koeficijentom korelacije između pokazatelja glasa i odgovornosti (VA) te političke stabilnosti i odustnosti nasilja/terorizma (PS). Iz provedene višestruke linearne regresije ne možemo zaključiti koji model najbolje predstavlja podatke zbog snažne koreliranosti među nezavisnim varijablama, WGI-jevima. Agregirani pokazatelj upravljanja kojeg dobijemo kao prosjek svih šest WGI-jeva također statistički značajno i pozitivno utječe na OBI. Zato možemo uzeti agregirani pokazatelj kao sažet indikator kvalitete institucije. Iz rezultata navedenih u dodatku vidimo da promatranjem WGI-jeva za 2019. ili njihovih prosjeka u razdoblju 2017. - 2019. dobijemo jako slične procijenjene parametre i  $p$ -vrijednosti. Na kraju, WGI-jevi su dobri pokazatelji transparentnosti proračuna te je dovoljno promatrati samo dobivene pokazatelje za posljednju godinu.

# Bibliografija

- [1] *Open Budget Survey 2019, howpublished* = <https://www.internationalbudget.org/open-budget-survey/open-budget-survey-2019>, note = Accessed: 2022-05-20.
- [2] *Open Budget Survey 2021.* <https://internationalbudget.org/open-budget-survey/open-budget-survey-2021>. Accessed: 2022-06-09.
- [3] *Regresijska analiza.* [https://www.pmf.unizg.hr/\\_download/repository/PREDAVANJE11.pdf](https://www.pmf.unizg.hr/_download/repository/PREDAVANJE11.pdf). Accessed: 2022-06-08.
- [4] *Statistički praktikum 2 - materijali.* <https://web.math.pmf.unizg.hr/nastava/statpr2/materijali.html>. Accessed: 2022-06-08.
- [5] *WGI Aggregation Methodology.* <http://info.worldbank.org/governance/wgi/Home/Documents>. Accessed: 2022-06-08.
- [6] *Why an Open Budget Survey?* <https://www.internationalbudget.org/open-budget-survey>. Accessed: 2022-05-20.
- [7] A., AlbassamBassam: *The Influence of Budget Transparency on Quality of Governance.* Journal of Law and Governance, 10(3), Dec. 2015. <https://journals.vu.edu.au/index.php/jbsge/article/view/864>.
- [8] Alt, James E i David Dreyer Lassen: *Fiscal transparency, political parties, and debt in OECD countries.* European Economic Review, 50(6):1403–1439, 2006.
- [9] Andreula, Nicolás, Alberto Chong i Jorge B Guillen: *Institutional quality and fiscal transparency.* 2009.
- [10] Arndt, Christiane i Charles Oman: *Uses and Abuses of Governance Indicators.* Paris: Organization for Economic Cooperation and Development. Development Centre Series, 2006.

- [11] Bađun, Marijana: *Transparentnost proračuna*. Financijska teorija i praksa, 33(4):495–497, 2009.
- [12] Bađun, Marijana i Ivica Urban: *Open Budget Index 2008-Small Advances Made in the Accessibility of Information about Croatian Public Finance*. Press releases, 2(8):1–4, 2009.
- [13] Bađun, Marijana i Ivica Urban: *Croatia's Score on Open Budget Index 2010*. Press releases, 3(26):1–2, 2010.
- [14] Bronić, Mihaela i Ivica Urban: *Croatia's Open Budget Index 2012 score-a slight improvement in the quality and comprehensiveness of budget information*. Press releases, 6(47):1–4, 2013.
- [15] Chen, Can i Milena I Neshkova: *The effect of fiscal transparency on corruption: A panel cross-country analysis*. Public Administration, 98(1):226–243, 2020.
- [16] Cimpoeru, Violeta Maria i Valentin Cimpoeru: *Influence of Government Quality on Budgetary Transparency*.
- [17] De Renzio, P i M Castro: *Is the open budget index an accurate measure of budget transparency? a research note investigating alternative aggregation methods*. Unpublished manuscript, 2011.
- [18] De Renzio, Paolo i Harika Masud: *Measuring and promoting budget transparency: The open budget index as a research and advocacy tool*. Governance, 24(3):607–616, 2011.
- [19] Denhardt, Janet V i Robert B Denhardt: *The New public service: Serving, not Steering*. Language, 13(222p):24cm, 2007.
- [20] Ellis, Christopher James i John Fender: *Corruption and transparency in a growth model*. International Tax and Public Finance, 13(2):115–149, 2006.
- [21] financije, Institut za javne: *Transparentnost regionalnih/lokalnih proračuna (studeni 2020. - travanj 2021.)*. listopad 2021. [http://www.ijf.hr/transparentnost/?params\\_1=transparentnost](http://www.ijf.hr/transparentnost/?params_1=transparentnost).
- [22] Frees, Edward W: *Regression modeling with actuarial and financial applications*. Cambridge University Press, 2009.
- [23] Handbook, Fiscal Transparency i Tax Policy Div: *International Monetary Fund*. Fiscal Affairs Dept, 2018. <https://doi.org/10.5089/9781484331859.069>.

- [24] Kaufmann, Daniel, Aart Kraay i Massimo Mastruzzi: *The Worldwide Governance Indicators Project: Answering the Critics*.
- [25] Kaufmann, Daniel, Aart Kraay i Massimo Mastruzzi: *Governance matters IV: governance indicators for 1996-2004*. World bank policy research working paper series, (3630), 2005.
- [26] Kaufmann, Daniel, Aart Kraay i Massimo Mastruzzi: *The worldwide governance indicators: Methodology and analytical issues*. Hague journal on the rule of law, 3(2):220–246, 2011.
- [27] Kaufmann, Daniel, Aart Kraay i Pablo Zoido: *Governance matters*. Available at SSRN 188568, 1999.
- [28] Kaufmann, Daniel, Aart Kraay i Pablo Zoido-Lobatón: *Aggregating governance indicators*, svezak 2195. world Bank publications, 1999.
- [29] Kosack, Stephen i Archon Fung: *Does transparency improve governance?* Annual review of political science, 17, 2014.
- [30] Mejía Acosta, Andrés: *The impact and effectiveness of accountability and transparency initiatives: The governance of natural resources*. Development Policy Review, 31:s89–s105, 2013.
- [31] Mensah, Yaw M i Yanxia Qi: *Culturally Preferred Leadership Styles, Government Budget Transparency, and the Perceived Rule of Law: An International Analysis*. Government Budget Transparency, and the Perceived Rule of Law: An International Analysis (October 12, 2012), 2012.
- [32] Montes, Gabriel Caldas, Júlio Cesar Albuquerque Bastos i Ana Jordânia de Oliveira: *Fiscal transparency, government effectiveness and government spending efficiency: Some international evidence based on panel data approach*. Economic Modelling, 79:211–225, 2019.
- [33] Montes, Gabriel Caldas i Paulo Henrique Luna: *Fiscal transparency, legal system and perception of the control on corruption: empirical evidence from panel data*. Empirical economics, 60(4):2005–2037, 2021.
- [34] Neumayer, Eric: *The pattern of aid giving: The impact of good governance on development assistance*. Routledge, 2003.
- [35] OECD: *OECD Best Practices for Budget Transparency*. OECD Journal on Budgeting, vol. 1/3:7–14, listopad 2002. <https://doi.org/10.1787/budget-v1-art14-en>.



- [36] Premchand, Arigapudi: *Public expenditure management*. International Monetary Fund, 1993.
- [37] Seifert, Jan, Ruth Carlitz i Elena Mondo: *The Open Budget Index (OBI) as a comparative statistical tool*. Journal of Comparative Policy Analysis: Research and Practice, 15(1), 2013.
- [38] UNDP: *Public Sector Management, Governance, and Sustainable Human Development*, 1995.
- [39] United Nations Economic and Social Commission for Asia and the Pacific (UNESCAP): *What is Good Governance?*, 10. srpnja, 2009 [Online]. <https://www.unescap.org/resources/what-good-governance#>.
- [40] Weisberg, Sanford: *Applied linear regression*, svezak 528. John Wiley & Sons, 2005.

## Sažetak

U ovom radu promatramo kako svjetski pokazatelji upravljanja (WGI) utječu na transparentnost proračuna. Kao razinu transparentnosti proračuna uzimamo indeks otvorenosti proračuna (OBI) te ga promatramo kao zavisnu varijablu. Nezavisne varijable predstavljaju WGI-jevi glas i odgovornost (VA), politička stabilnost i odsutnost nasilja/terorizma (PS), učinkovitost vlade (GE), kvaliteta regulatornog okvira (RQ), vladavina prava (RL) te kontrola korupcije (CC). Za uzorak uzimamo 117 zemalja za koje je izračunat OBI u 2019. WGI-jeve promatramo za istih 117 zemalja, no uz 2019., promatramo i prosjek svakog pokazatelja u razdoblju 2017. - 2019. Metodom linearne regresije dolazimo do zaključka da su svi WGI statistički značajni i to pozitivno utječu na OBI. Uvodimo pojam agregiranog WGI pokazatelja kao prosjek svih šest svjetskih pokazatelja upravljanja. Zaključujemo da je agregirani WGI također statistički značajan za OBI i pozitivno utječe na njega. Također vidimo da su WGI pokazatelji međusobno snažno korelirani. Na kraju zaključujemo da bolje upravljanje ukazuje na veću transparentnost državnog proračuna.

# Summary

In this paper, we consider how Worldwide Governance Indicators (WGIs) affect budget transparency. The Open Budget Index (OBI) is taken as the measure of budget transparency and viewed as a dependent variable. WGIs are used as independent variables, 6 in total: Voice and Accountability (VA), Political Stability and Absence of Violence / Terrorism (PS), Government Effectiveness (GE), Regulatory Quality (RQ), Rule of Law (RL) and Control of Corruption (CC). We take as a sample 117 countries for which OBI was calculated in 2019. Regarding WGIs, we observe them for the same 117 countries, but except for 2019, we also observe the average of each indicator in the period 2017. - 2019. Using the linear regression method, we determined a positive and statistically significant effect of each WGI on OBI. We introduce the concept of an aggregate WGI indicator as an average of all six Worldwide Governance Indicators. Aggregated WGI is also statistically significant for OBI and positively affects it. Strong correlation between all WGIs is also observed. Finally, we conclude that better governance indicates greater budget transparency for a country.

# Životopis

Marija Pia Kovačević rođena je 4. veljače 1997. u Splitu. Pohađala je Osnovnu školu Hvar u Hvaru, te upisala opću gimnaziju u Srednjoj školi Hvar. Svoje obrazovanje nastavlja u Zagrebu upisom preddiplomskog studija Matematika na Prirodoslovno matematičkom fakultetu u Zagrebu. Na istom fakultetu upisuje i diplomski smjer Financijska i poslovna matematika.

# Dodatak A

## Usporedba WGI i Avg\_WGI

### Avg\_PS

Pogledajmo sada rezultate za Avg\_PS: Promatramo model

$$OBI_{2019} = \beta_0 + \beta_1 \cdot Avg\_PS$$

```
> model = lm(OBI_2019 ~ Avg_PS)
> summary(model)
```

Call:

```
lm(formula = OBI_2019 ~ Avg_PS)
```

Residuals:

Min	1Q	Median	3Q	Max
-57.211	-11.549	-0.601	12.498	42.203

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	48.983	1.961	24.976	< 2e-16 ***
Avg_PS	12.619	2.037	6.194	9.39e-09 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

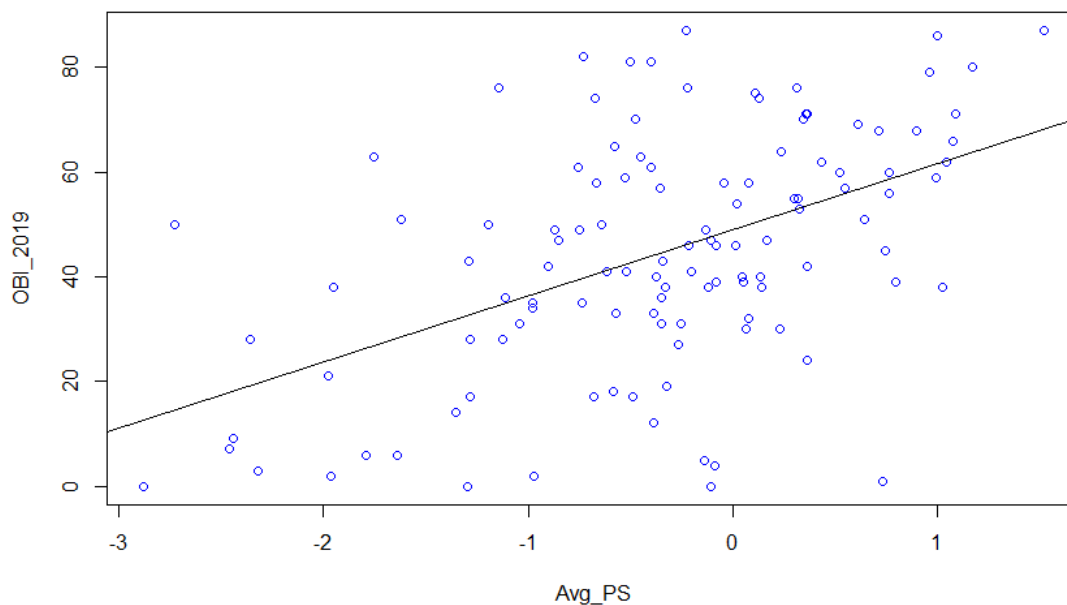
Residual standard error: 19.81 on 115 degrees of freedom

Multiple R-squared: 0.2501, Adjusted R-squared: 0.2436

F-statistic: 38.36 on 1 and 115 DF, p-value: 9.391e-09

Linearnom regresijom dobijemo da parametri modela glase  $\beta_0 = 48,983$  i  $\beta_1 = 12,619$ , što je veoma slično rezultatima dobivenim kod pokazatelja PS iz 2019. Vidimo da je  $p$ -

vrijednost dovoljno mala ( $< 0,05$ ) te zaključujemo da i Avg\_PS ima statistički značajan pozitivan utjecaj na OBI. Vizualnu reprezentaciju podataka i regresijskog pravca vidimo na slici A.1



Slika A.1: Podaci (Avg\_PS, OBI\_2019) i pravac dobiven modelom

Koeficijent determinacije  $R^2$  sada iznosi 0,2501, a prilagođeni koeficijent determinacije  $\bar{R}^2$  0,2436. Model dobro opisuje 25,01% podataka. Iz koeficijenata determinacije i procijenjenih parametara možemo zaključiti da oba parametra (obični i prosječni) jako slično opisuju OBI te je dovoljno promatrati samo jedan.

Lillieforsovom inačicom Kolmogorov-Smirnovljeva testa provjerimo jesu li reziduali i dalje normalno distribuirani.

```
> lillie.test(model$residuals)
```

Lilliefors (Kolmogorov-Smirnov) normality test

```
data: model$residuals
```

```
D = 0.045022, p-value = 0.8117
```

Iz dobivene  $p$ -vrijednosti = 0,8117  $>$  0,05 zaključujemo da ne možemo odbaciti nultu hipotezu da su podaci normalno distribuirani.

## Avg\_GE

Pogledajmo sada rezultate za Avg\_GE: Promatramo model

$$OBI_{2019} = \beta_0 + \beta_1 \cdot Avg\_GE$$

```
> model = lm(OBI_2019 ~ Avg_GE)
> summary(model)
```

Call:

```
lm(formula = OBI_2019 ~ Avg_GE)
```

Residuals:

Min	1Q	Median	3Q	Max
-58.785	-10.112	0.488	9.605	37.619

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	47.821	1.591	30.066	<2e-16 ***
Avg_GE	17.018	1.727	9.854	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.85 on 115 degrees of freedom

Multiple R-squared: 0.4578, Adjusted R-squared: 0.4531

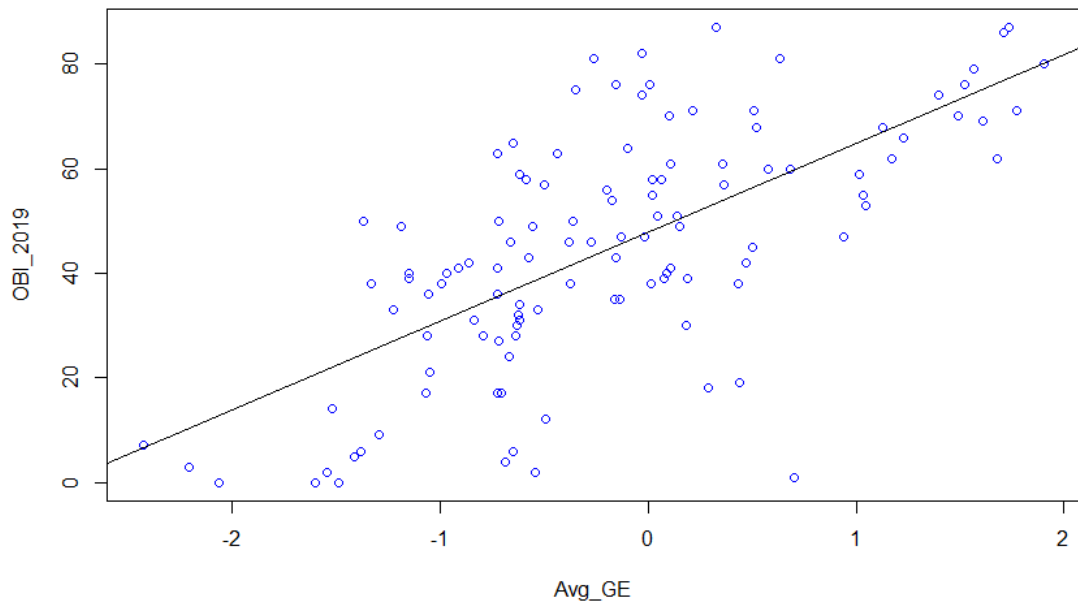
F-statistic: 97.1 on 1 and 115 DF, p-value: < 2.2e-16

Linearnom regresijom dobijemo da parametri modela glase  $\beta_0 = 47,821$  i  $\beta_1 = 17,018$ , što je veoma slično rezultatima dobivenim kod pokazatelja GE iz 2019. Vidimo da je  $p$ -vrijednost dovoljno mala ( $< 0,05$ ) te zaključujemo da i Avg\_GE ima statistički značajan pozitivan utjecaj na OBI. Vizualnu reprezentaciju podataka i regresijskog pravca vidimo na slici A.2

Koeficijent determinacije  $R^2$  sada iznosi 0,4578, a prilagođeni koeficijent determinacije  $\bar{R}^2$  0,4531. Model dobro opisuje 45,31% podataka. Iz koeficijenata determinacije i procijenjenih parametara možemo zaključiti da oba parametra (obični i prosječni) jako slično opisuju OBI te je dovoljno promatrati samo jedan.

Lillieforsovom inačicom Kolmogorov-Smirnovljeva testa provjerimo jesu li reziduali i dalje normalno distribuirani.

```
> lillie.test(model$residuals)
```



Slika A.2: Podaci (Avg\_GE, OBI\_2019) i pravac dobiven modelom

Lilliefors (Kolmogorov-Smirnov) normality test

```
data: model$residuals
D = 0.054068, p-value = 0.5486
```

Iz dobivene  $p$ -vrijednosti = 0,5486 > 0,05 zaključujemo da ne možemo odbaciti nultu hipotezu da su podaci normalno distribuirani.

## Avg\_RQ

Pogledajmo sada rezultate za Avg\_RQ: Promatramo model

$$OBI_{2019} = \beta_0 + \beta_1 \cdot Avg\_RQ$$

```
> model = lm(OBI_2019 ~ Avg_RQ)
> summary(model)
```

```
Call:
lm(formula = OBI_2019 ~ Avg_RQ)
```



Residuals:

Min	1Q	Median	3Q	Max
-57.113	-9.970	0.088	9.420	37.944

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	47.046	1.483	31.73	<2e-16 ***
Avg_RQ	18.492	1.660	11.14	<2e-16 ***

---

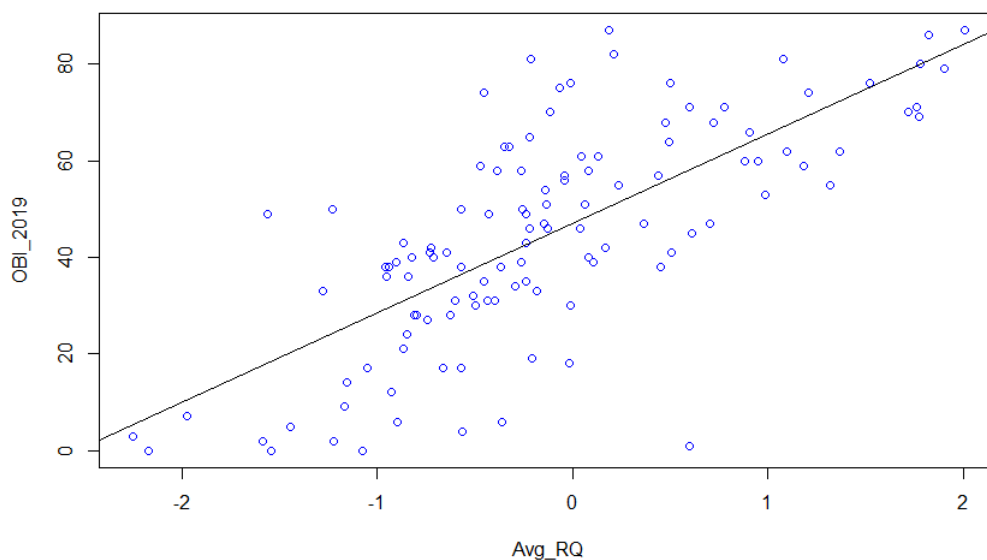
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.87 on 115 degrees of freedom

Multiple R-squared: 0.519, Adjusted R-squared: 0.5148

F-statistic: 124.1 on 1 and 115 DF, p-value: < 2.2e-16

Linearnom regresijom dobijemo da parametri modela glase  $\beta_0 = 47,046$  i  $\beta_1 = 18,492$ , što je veoma slično rezultatima dobivenim kod pokazatelja RQ iz 2019. Vidimo da je  $p$ -vrijednost dovoljno mala ( $< 0,05$ ) te zaključujemo da i Avg\_RQ ima statistički značajan pozitivan utjecaj na OBI. Vizualnu reprezentaciju podataka i regresijskog pravca vidimo na slici A.3



Slika A.3: Podaci (Avg\_RQ, OBI\_2019) i pravac dobiven modelom

Koeficijent determinacije  $R^2$  sada iznosi 0,5190, a prilagođeni koeficijent determinacije  $\bar{R}^2$  0,5148. Model dobro opisuje 51,90% podataka. Iz koeficijenata determinacije i procijenjenih parametara možemo zaključiti da oba parametra (obični i prosječni) jako slično opisuju OBI te je dovoljno promatrati samo jedan.

Lillieforsovom inačicom Kolmogorov-Smirnovljeva testa provjerimo jesu li reziduali i dalje normalno distribuirani.

```
> lillie.test(model$residuals)
```

Lilliefors (Kolmogorov-Smirnov) normality test

```
data: model$residuals
```

```
D = 0.060819, p-value = 0.3585
```

Iz dobivene  $p$ -vrijednosti = 0,3585 > 0,05 zaključujemo da ne možemo odbaciti nultu hipotezu da su podaci normalno distribuirani.

## Avg\_RL

Pogledajmo sada rezultate za Avg\_RL: Promatramo model

$$OBI_{2019} = \beta_0 + \beta_1 \cdot Avg\_RL$$

```
> model = lm(OBI_2019 ~ Avg_RL)
> summary(model)
```

Call:

```
lm(formula = OBI_2019 ~ Avg_RL)
```

Residuals:

Min	1Q	Median	3Q	Max
-59.353	-11.702	-0.113	9.345	42.888

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	48.591	1.710	28.419	< 2e-16 ***
Avg_RL	15.736	1.822	8.635	3.87e-14 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

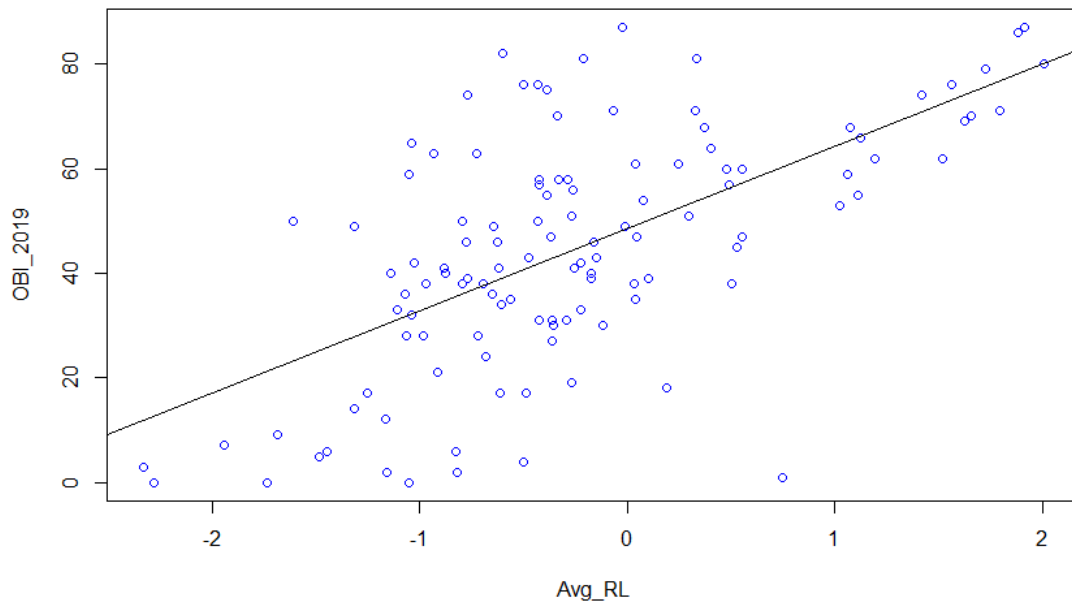
Residual standard error: 17.82 on 115 degrees of freedom

Multiple R-squared: 0.3933, Adjusted R-squared: 0.3881

F-statistic: 74.56 on 1 and 115 DF, p-value: 3.867e-14

Linearnom regresijom dobijemo da parametri modela glase  $\beta_0 = 48,591$  i  $\beta_1 = 15,736$ , što je veoma slično rezultatima dobivenim kod pokazatelja RL iz 2019. Vidimo da je  $p$ -vrijednost dovoljno mala ( $< 0,05$ ) te zaključujemo da i Avg\_RL ima statistički značajan pozitivan utjecaj na OBI. Vizualnu reprezentaciju podataka i regresijskog pravca vidimo na slici A.4

Koeficijent determinacije  $R^2$  sada iznosi 0,3933, a prilagođeni koeficijent determinacije  $\bar{R}^2$  0,3881. Model dobro opisuje 39,33% podataka. Iz koeficijenata determinacije i procijenjenih parametara možemo zaključiti da oba parametra (obični i prosječni) jako slično opisuju OBI te je dovoljno promatrati samo jedan.



Slika A.4: Podaci (Avg\_RL, OBI\_2019) i pravac dobiven modelom

Lillieforsovom inačicom Kolmogorov-Smirnovljeva testa provjerimo jesu li reziduali i dalje normalno distribuirani.

```
> lillie.test(model$residuals)
```

Lilliefors (Kolmogorov-Smirnov) normality test

```
data: model$residuals
D = 0.066349, p-value = 0.2343
```

Iz dobivene  $p$ -vrijednosti = 0,2343 > 0,05 zaključujemo da ne možemo odbaciti nultu hipotezu da su podaci normalno distribuirani.

## Avg\_CC

Pogledajmo sada rezultate za Avg\_CC: Promatramo model

$$OBI_{2019} = \beta_0 + \beta_1 \cdot Avg\_CC$$

```
> model = lm(OBI_2019 ~ Avg_CC)
> summary(model)
```

```
Call:
lm(formula = OBI_2019 ~ Avg_CC)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-59.445 -11.895  -0.239   8.667  45.257
```

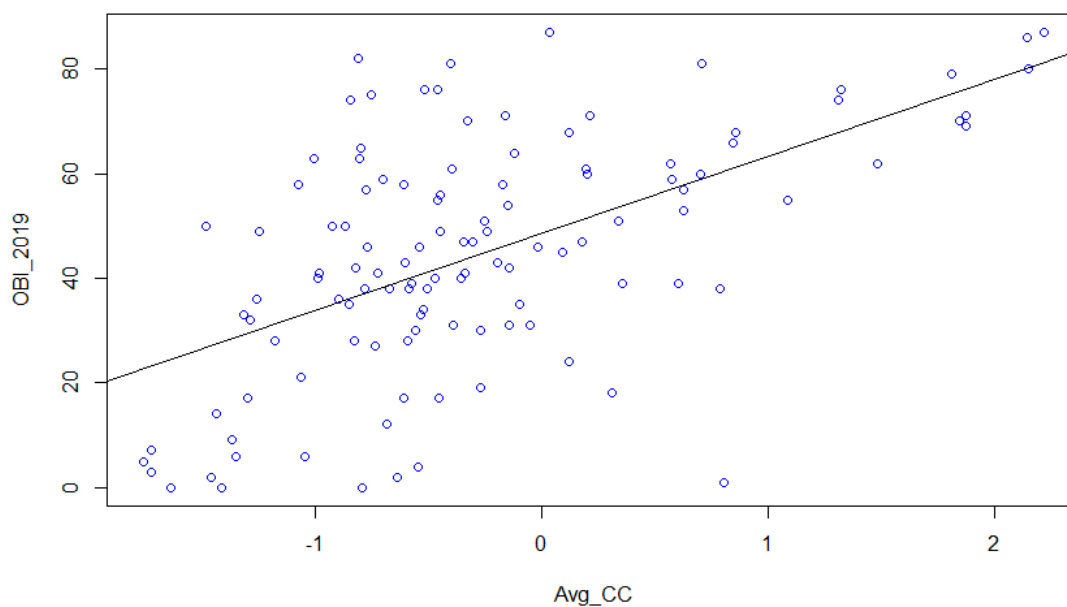
```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   48.645      1.812   26.853 < 2e-16 ***
Avg_CC        14.694      1.953    7.523 1.29e-11 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 18.73 on 115 degrees of freedom
Multiple R-squared:  0.3298, Adjusted R-squared:  0.324
F-statistic: 56.59 on 1 and 115 DF,  p-value: 1.29e-11
```

Linearnom regresijom dobijemo da parametri modela glase  $\beta_0 = 48,645$  i  $\beta_1 = 14,694$ , što je veoma slično rezultatima dobivenim kod pokazatelja CC iz 2019. Vidimo da je  $p$ -vrijednost dovoljno mala ( $< 0,05$ ) te zaključujemo da i Avg\_CC ima statistički značajan pozitivan utjecaj na OBI. Vizualnu reprezentaciju podataka i regresijskog pravca vidimo na slici A.5

Koeficijent determinacije  $R^2$  sada iznosi 0,3298, a prilagođeni koeficijent determinacije  $\bar{R}^2$  0,3240. Model dobro opisuje 32,98% podataka. Iz koeficijenata determinacije i procijenjenih parametara možemo zaključiti da oba parametra (obični i prosječni) jako slično opisuju OBI te je dovoljno promatrati samo jedan.



Slika A.5: Podaci (Avg\_CC, OBI\_2019) i pravac dobiven modelom

Lillieforsovom inačicom Kolmogorov-Smirnovljeva testa provjerimo jesu li reziduali i dalje normalno distribuirani.

```
> lillie.test(model$residuals)
```

Lilliefors (Kolmogorov-Smirnov) normality test

```
data: model$residuals
D = 0.07319, p-value = 0.1291
```

Iz dobivene  $p$ -vrijednosti = 0,1291 > 0,05 zaključujemo da ne možemo odbaciti nultu hipotezu da su podaci normalno distribuirani.

### Hipoteza 7 - stepAIC()

Ugrađenom funkcijom stepAIC() (eng. Akaike Information Criteria) u R-u metodom "korak po korak" dolazimo do najboljeg linearnog regresijskog modela. U slučaju više modela, smatramo boljim onim koji ima manju AIC vrijednost. Dakle, želimo odabrati one nezavisne varijable koje će nam u modelu najbolje opisivati zavisnu varijablu.

```

> model_AIC <- stepAIC(model, direction = "both")
Start:  AIC=618.77
OBI_2019 ~ VA + RQ + GE + PS + RL + CC

```

	Df	Sum of Sq	RSS	AIC
- PS	1	95.1	20655	617.31
<none>			20560	618.77
- RL	1	584.5	21145	620.05
- CC	1	1532.3	22093	625.18
- GE	1	1572.0	22132	625.39
- RQ	1	2122.3	22683	628.26
- VA	1	4310.2	24871	639.03

```

Step:  AIC=617.31
OBI_2019 ~ VA + RQ + GE + RL + CC

```

	Df	Sum of Sq	RSS	AIC
<none>			20655	617.31
+ PS	1	95.1	20560	618.77
- RL	1	680.0	21335	619.10
- GE	1	1499.2	22155	623.50
- CC	1	1550.3	22206	623.77
- RQ	1	2219.5	22875	627.25
- VA	1	4256.7	24912	637.23

Algoritam staje već nakon drugog koraka te iz njega vidimo da model koji u sebi ne sadrži varijablu PS ("politička stabilnost"), po metodi "korak po korak" najbolje opisuje OBI. Prilagođeni koeficijent determinacije ovog model iznosi 0,6414, što znači da model dobro opisuje 64,14% podataka. No, ove rezultate moramo promatrati oprezno zbog velike koreliranosti nezavisnih varijabli.