

# Socio-ekonomske odrednice transparentnosti državnog proračuna

---

**Andrijević, Marija**

**Master's thesis / Diplomski rad**

**2020**

*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:217:813792>

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

*Download date / Datum preuzimanja:* **2025-03-29**



*Repository / Repozitorij:*

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



# Socio-ekonomske odrednice transparentnosti državnog proračuna

---

**Andrijević, Marija**

**Master's thesis / Diplomski rad**

**2020**

*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:217:813792>

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

*Download date / Datum preuzimanja:* **2024-06-18**



*Repository / Repozitorij:*

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



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

Marija Andrijević

**SOCIO-EKONOMSKE ODREDNICE**  
**TRANSPARENTNOSTI DRŽAVNOG**  
**PRORAČUNA**

Diplomski rad

Voditelj rada:  
prof.dr.sc. Katarina Ott

Zagreb, 2020.

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. \_\_\_\_\_

*Najveće hvala obitelji i prijateljima na neizmjerivoj podršci, posebno majci bez koje ništa od ovoga ne bih uspjela.*

# Sadržaj

<b>Sadržaj</b>	<b>iv</b>
<b>Uvod</b>	<b>2</b>
<b>1 Transparentnost državnog proračuna</b>	<b>3</b>
1.1 Definicija transparentnosti državnog proračuna . . . . .	3
1.2 Indeks otvorenosti proračuna . . . . .	4
1.3 Hipoteze . . . . .	6
<b>2 Podaci</b>	<b>9</b>
2.1 Zavisna varijabla . . . . .	9
2.2 Nezavisne varijable . . . . .	10
2.3 Deskriptivna statistika . . . . .	11
<b>3 Metodologija</b>	<b>14</b>
3.1 Generalizirani linearni modeli . . . . .	14
3.2 Logistička regresija . . . . .	16
3.3 Multinomna logistička regresija . . . . .	17
3.4 Procjena parametara . . . . .	19
3.5 Kvaliteta i odabir modela . . . . .	21
<b>4 Empirijska analiza</b>	<b>24</b>
4.1 Jednostruka ordinalna logistička regresija . . . . .	25
4.2 Višestruka ordinalna logistička regresija . . . . .	28
4.3 Automatizirane procedure u R-u . . . . .	34
<b>5 Zaključak</b>	<b>37</b>
<b>Bibliografija</b>	<b>39</b>

# Uvod

Transparentnost proračuna se kao pojam sve češće pojavljuje među građanima te u medijima te je bitno znati njegovo značenje i važnost za ekonomiju i razvoj svake zemlje, grada, općine... Prema Institutu za javne financije [6] transparentnost proračuna podrazumijeva potpune, točne, pravovremene i razumljive informacije o proračunu i prvi je korak prema pozivanju na odgovornost za prikupljanje i trošenje javnog novca te smanjenje korupcije. Građani trebaju biti uključeni u donošenje odluka o proračunu, suzbijanju korupcije te zahtijevati odgovornost vlasti. Visoka transparentnost vraća povjerenje građana u vladu.

Međunarodni monetarni fond (MMF), Svjetska banka te Organizacija za ekonomsku suradnju i razvoj (OECD) su jedne od mnogih organizacija koje iniciraju povećanje transparentnosti državnih proračuna kroz razne kodekse i smjernice.

Iako je ovo područje istraživanja još u razvoju, postoji brojna literatura koja se bavi transparentnošću pojedinih zemalja, općina, gradova. Međudržavna usporedba transparentnosti moguća je od 2006., od kad International Budget Partnership iz Washingtona objavljuje indeks otvorenosti proračuna (u daljnjem tekstu IOP) dobiven iz opsežnog upitnika kojim se ocjenjuje transparentnost proračuna. Upitnik prikuplja podatke o javnoj dostupnosti i opsežnosti nekoliko proračunskih dokumenata koji se pojavljuju u pojedinim fazama proračunskog procesa. IOP u ovom radu korišten je kao zavisna varijabla te se kao uzorak uzima 117 zemalja obuhvaćenim zadnjim provedenim istraživanjem 2019.[20]. IOP se kreće od 0 do 100 te zemlje svrstava u 5 kategorija, od kategorije gdje se nalaze zemlje koje građanima pružaju opsežne informacije (81-100) do kategorije zemalja koje građanima pružaju neznatne ili nikakve informacije (0-20).

Cilj rada je ispitati socio-ekonomske odrednice transparentnosti državnog proračuna kako bi se moglo doći do zaključka zašto je u nekim zemljama transparentnost visoka dok je u drugima izrazito niska i koji su načini za njeno poboljšanje. Ispituje se utjecaj sljedećih varijabli: stope nezaposlenosti, BDP-a po stanovniku, udjela javnog duga u BDP-u, broja stanovnika, postotka korisnika interneta, indeksa ljudskog razvoja te indeksa percepcije korupcije.

Na temelju prethodnih radova [2, 3, 15, 23, 24, 28, 33] postavljaju se sljedeće hipoteze:

stopa nezaposlenosti te udio javnog duga u BDP-u imaju negativan i statistički značajan utjecaj na IOP, dok preostale varijable imaju pozitivan i statistički značajan utjecaj. Hipoteze se empirijski testiraju na uzorku od 117 zemalja. Promatrajući jednostruke modele, tj. modele s jednom kovarijatom, rezultati su pokazali da stopa nezaposlenosti te broj stanovnika imaju negativan utjecaj na IOP, dok je kod drugih varijabli uočen pozitivan utjecaj. Statistički značajan utjecaj pokazale su varijable BDP po stanovniku, postotak korisnika interneta, indeks ljudskog razvoja te indeks percepcije korupcije. Analizom potpunog modela, tj. modela koji sadrži sve kovarijate redom navedene, negativan utjecaj uočava se kod varijabli: stopa nezaposlenosti, BDP po stanovniku te udjela javnog duga u BDP-u. Kod preostalih varijabli uočena je pozitivna povezanost, dok je koeficijent uz varijablu broj stanovnika približan nuli pa nema značajan utjecaj na IOP. Primjenom ordinalne logističke regresije na potpuni model te promatranjem korelacija među nezavisnim varijablama odabran je najbolji model koji sadrži dvije statistički značajne kovarijate, indeks ljudskog razvoja te indeks percepcije korupcije. Sve se konstruira i provodi u programskom jeziku R.

Rad se sastoji od nekoliko dijelova. Prvi dio rada daje različite definicije transparentnosti državnog proračuna, opis IOP-a te kratku analizu dosadašnjih objavljenih rezultata. Nakon kratkog uvoda postavljaju se navedene hipoteze koju su temeljene na objavljenim radovima te empirijskim istraživanjima. Drugi dio daje kratak opis te deskriptivnu statistiku podataka koji će se koristiti u radu. Nakon opisa podataka treće poglavlje daje opis metodologije korištene u radu. Zbog IOP-a koji je kategorijska varijabla naglasak je na ordinalnoj logističkoj regresiji. Empirijska analiza koja je i glavni dio rada napravljena je u četvrtom poglavlju gdje se promatralo više različitih modela te kriterijima i metodama objašnjenim u trećem poglavlju došlo do najprikladnijeg modela. Na kraju rada dan je zaključak te osvrt na cijeli rad.



# Poglavlje 1

## Transparentnost državnog proračuna

U ovom poglavlju definira se transparentnost državnog proračuna te se navodi njena važnost i razlozi sve većeg zanimanja šire javnosti za ovu tematiku. Zatim se opisuje IOP kao glavna mjera transparentnosti proračuna te se daje kratka analiza prijašnjih rezultata u usporedbi s podacima za 2019. na kojima se temelji rad.

Nakon kratkog opisa postavljaju se hipoteze koje će se u glavnom dijelu rada ispitivati.

### 1.1 Definicija transparentnosti državnog proračuna

Jednu od definicija fiskalne transparentnosti, što je širi pojam od transparentnosti proračuna, dali su Kopits i Craig [23]. Fiskalnu transparentnost definiraju kao otvorenost prema javnosti u pogledu vladine strukture i funkcija, planiranoj fiskalnoj politici, proračunu javnog sektora te budućih projekcija. To uključuje pristup pouzdanim, sveobuhvatnim, pravovremenim, razumljivim i međunarodno usporedivim informacijama o vladinim aktivnostima, tako da biračko tijelo i financijska tržišta mogu točno procijeniti financijski položaj vlade, stvarne troškove te koristi od njezinog upravljanja, uključujući njihove sadašnje i buduće ekonomske i društvene posljedice.

Poterba i von Hagen [29] definiraju transparentan proračunski proces kao onaj koji pruža jasne informacije o svim aspektima državne fiskalne politike. Proračuni koji su lako dostupni javnosti i sudionicima u politici te koji sadrže konsolidirane podatke su transparentni.

Prema Malatestinić [25] transparentan proračunski proces uključuje skup aktivnosti, dokumenata, metoda, intervencija koje izvršni čelnik i/ili predstavničko tijelo provode s ciljem pravovremenog, potpunog, točnog i jasnog informiranja građana o prošlim, sadašnjim i budućim aktivnostima izabrane vlasti.

Organizacija za ekonomsku suradnju i razvoj (OECD-Organization for Economic Cooperation and Development, 2002) definira transparentnost državnog proračuna kao pra-

vovremeno, potpuno i sustavno objavljivanje svih relevantnih fiskalnih informacija te daje smjernice za to područje, s naglaskom na središnju vlast, koje dijeli u tri dijela [26]. U prvom dijelu su navedena glavna proračunska izvješća koja vlade trebaju izraditi i njihov opći sadržaj, u drugom se dijelu opisuju specifični podaci koje bi ta izvješća trebala sadržavati dok treći dio ističe postupke kojima se osigurava kvaliteta i integritet izvješća.

Analogno, Međunarodni monetarni fond, u nastavku MMF, je objavio smjernice za fiskalnu transparentnost podijeljene u četiri dijela: 1. jasnoća uloga i dužnosti državnoga i javnog sektora; 2. otvorenost proračunskog procesa; 3. javnu raspoloživost informacija; 4. jamčenje integriteta izvješća. Izvršni odbor MMF-a usvojio je u kolovozu 2014. novi Kodeks fiskalne transparentnosti (The Fiscal Transparency Code [18]) koji se temelji na tri načela fiskalne transparentnosti: 1. fiskalno izvještavanje; 2. fiskalna predviđanja i proračuni; 3. analiza i upravljanje fiskalnim rizicima. U siječnju 2019. dodano je 4. načelo: upravljanje prihodima od resursa te je Kodeks postao standardom za sve članice MMF-a [19]. Držeći fiskalnu transparentnost jednim od glavnih preduvjeta uspješnog upravljanja državnim financijama, MMF tim Kodeksom želi osigurati javnu dostupnost informacija i povećati odgovornost vlada prema tržištima, zakonodavnim tijelima i građanima.

Uz spomenute te brojne druge organizacije i inicijative diljem svijeta koje se bave ovom tematikom, bitno je spomenuti Institut za javne financije koji kao član Globalnog pokreta za proračunsku transparentnost istražuje i objavljuje rezultate svojih istraživanja u nadi da će tako pridonijeti većoj proračunskoj transparentnosti, odgovornosti vlasti i sudjelovanju građana u Hrvatskoj.

U trenutnoj situaciji u svijetu dok se zdravstvene organizacije bore s virusom Covid 19, vlade se bore s njegovim ekonomskim posljedicama. Vlade moraju odgovarati brzo i odgovorno te će jednu od glavnih uloga u tome imati državni proračuni. Da bi osigurali povjerenje građana od iznimne je važnosti visoka transparentnost državnog proračuna. Veća transparentnost povećava povjerenje stanovništva u vladu te ima pozitivan učinak na cjelokupnu ekonomiju zemlje.

## 1.2 Indeks otvorenosti proračuna

Na temelju smjernica MMF-a, OECD-a te PEFA (Public Expenditure and Financial Accountability) International Budget Partnership (u daljnjem tekstu IBP) iz Washingtona od 2006. objavljuje IOP, koji se računa svake druge godine, a omogućuje usporedbu transparentnosti proračuna zemalja diljem svijeta koje su obuhvaćene istraživanjem. IBP izračunava IOP temeljem upitnika koji se sastoji od 145 pitanja kojeg popunjavaju nezavisni stručnjaci u zemljama obuhvaćenim istraživanjem. Upitnik o otvorenosti proračuna ocjenjuje tri komponente sustava proračunske odgovornosti: dostupnost javnih informacija o proračunu (109 pitanja); mogućnosti javnosti za sudjelovanje u proračunskom pro-

cesu (18 pitanja); i ulogu formalnih nadzornih institucija, uključujući zakonodavno tijelo i državni ured za reviziju (18 pitanja). Upitnik također uključuje dodatna 83 pitanja koja se ne koriste za izračunavanje pojedinačnih bodova, ali pomažu u dovršavanju istraživanja prikupljanjem osnovnih informacija o ključnim proračunskim dokumentima i istraživanjem različitih karakteristika upravljanja javnim financijama u zemlji.

Prema Bronić, Franić [11] otvorenost državnog proračuna je značajna jer je riječ o instrumentu kojim država utječe na živote svih građana. Pravovremena objava kvalitetnih ključnih proračunskih dokumenata omogućuje analiziranje i argumentirano uključivanje građana i stručnjaka u proračunski proces, te usklađivanje njihovih želja i potreba s vladinim ciljevima. IOP mjeri dostupnost online pravovremenih, razumljivih i dovoljno opsežnih informacija o državnom proračun kojima mogu pristupiti obični građani.

U ovom radu gledat će se rezultati istraživanja za 2019. IOP poprima vrijednosti od 0 do 100 te svrstava zemlje u pet kategorija; od onih koje građanima pružaju "opsežne informacije" (indeks od 81 do 100), pa do onih koje "pružaju neznatne ili nikakve informacije" (indeks od 0 do 20). Svi detalji sadržaja Upitnika, provedbe istraživanja kao i svih rezultata dostupni su na stranicama IBP-a [20].

## **Analiza dosadašnjih rezultata**

IOP 2019. rezultat je sedmog kruga IBP-ova istraživanja (prethodni su provedeni 2006., 2008., 2010., 2012., 2015. i 2017.).

Istraživanje pokazuje globalno poboljšanje transparentnosti proračuna, što je u skladu s ukupnim trendom mjerenim istraživanjem u posljednjem desetljeću. Za 115 zemalja koje su sudjelovale u istraživanju 2017. te 2019. prosječni globalni rezultat za IOP - porastao je s 42 na 45 od 100. Prosječna ocjena za ove zemlje i dalje je manja od 61, što se smatra minimalnom razinom transparentnosti proračuna koja omogućava uključivanje javnosti u proračunski proces. Nekoliko regija ima stalni uzlazni trend, posebno Istočna Europa i Srednja Azija, Istočna Azija i Tih ocean te Latinska Amerika i Karibi.

Prosječna vrijednost IOP-a za 2019. te 117 promatranih zemalja iznosi 45. Gotovo 75% anketiranih zemalja je ocijenjeno nedovoljnom razinom (za zemlje s indeksom manjim od 61 smatra se da ne postižu zadovoljavajuću razinu transparentnosti). To znači da je 2019. u analiziranim zemljama u prosjeku bilo dostupno samo 45% ključnih informacija o prihodima i rashodima proračuna. Kao i u prethodnim krugovima, najotvorenije proračune imaju Novi Zeland, Južnoafrička Republika i Švedska, dok su na začelju ljestvice Jemen, Venezuela, Komori i Katar.

Kontinuirano povećanje IOP-a daje nadu za daljni napredak, trenutna razina javno dostupnih informacija o proračunu je i dalje ograničena: samo 31 od 117 analiziranih zemalja ima dovoljne razine transparentnosti proračuna. Vlade često ne objavljuju ključne proračunske

dokumente koji bi trebali jasno objasniti proračunske politike, odluke i javna sredstva. Svih osam ključnih proračunskih dokumenata koji bi se trebali objaviti u svim zemljama nisu dostupni javnosti za trećinu zemalja. Hrvatska je zabilježila značajan napredak te ostvarila najbolji rezultat do sad. IOP 2019. za Hrvatsku iznosi 68, dok je od 2006. do 2017. iznosio redom 42, 59, 57, 61, 53 te 57. Kad su posrijedi susjedne zemlje koje su bile uključene u istraživanje, Slovenija ima jednak IOP kao Hrvatska. IOP za Bosnu i Hercegovinu je 33, Srbiju 40 te Mađarsku 45. Hrvatska je zauzela 21. mjesto na svjetskoj ljestvici i opet je u skupini zemalja koje objavljuju značajnu količinu proračunskih informacija.

### 1.3 Hipoteze

U ovom dijelu postavljaju se hipoteze o socio-ekonomskim odrednicama transparentnosti državnog proračuna temeljene na teorijskim prepostavkama i empirijskim istraživanjima.

To su stopa nezaposlenosti, BDP po stanovniku, udio javnog duga u BDP-u, broj stanovnika, postotak korisnika interneta, indeks ljudskog razvoja te indeks percepcije korupcije.

#### **Hipoteza 1: stopa nezaposlenosti ima negativan i statistički značajan utjecaj na indeks otvorenosti proračuna**

Niže razine gospodarskog razvoja i s njima povezane veće stope nezaposlenosti povezane su s nižom transparentnošću. Tavares i Cruz u svom radu [33], analizirajući 308 općina Portugala, pokazuju negativnu korelaciju između stope nezaposlenosti i transparentnost proračuna.

#### **Hipoteza 2: BDP po stanovniku ima pozitivan i statistički značajan utjecaj na indeks otvorenosti proračuna**

Lowatcharin i Menifield [24] ispituju utjecaj geografskih, socio-ekonomskih te institucionalnih odrednica na transparentnost lokalnih proračuna 12 američkih saveznih država te pokazuju da je varijabla BDP po stanovniku statistički značajna te pozitivno korelirana s transparentnošću proračuna.

Piotrowski i van Ryzin [28] socioekonomski status, mjereno stupnjem obrazovanja i dohotkom, povezuju sa sve većim uključivanjem građana u nadziranje lokalnih vlasti te tvrde da to vodi sve većim zahtjevima za transparentnosti.

### **Hipoteza 3: udio javnog duga u BDP-u ima negativan i statistički značajan utjecaj na indeks otvorenosti proračuna**

Predviđanja da fiskalna transparentnost dovodi do znatno manjeg deficita i akumulacije duga u mnogim su slučajevima dobila čvrstu empirijsku potporu. Mnogi radovi pokazuju negativan odnos između transparentnosti i razine duga. Jedan od njih je rad Alta i Lassena [4] koji su na uzorku od 19 zemalja članica OECD-a to i potvrdili. Njihovi rezultati sugeriraju da je povećana fiskalna transparentnost važan element poboljšanja fiskalne učinkovitosti i da fiskalne institucije zaista utječu na fiskalne rezultate.

### **Hipoteza 4: broj stanovnika ima pozitivan i statistički značajan utjecaj na indeks otvorenosti proračuna**

Jedna od varijabli koja se najčešće koristi u objašnjavanju determinanti transparentnosti proračuna je broj stanovnika pojedine države, regije, općine itd.

U mnogim prijašnjim radovima, primjerice (del Sol, [15] Guillamón, Bastida i Benito, [16]; Lowatcharin i Menifield, [24]) pokazana je pozitivna korelacija broja stanovnika te transparentnosti lokalnog proračuna. Kao objašnjenje takve veze navodi se nekoliko razloga: veće lokalne jedinice imaju više resursa te kapaciteta za veću transparentnost te su zbog većih javnih sredstava podložne većim pritiskom da budu odgovorne i otvorene prema javnosti. Analogno će se pokušati zaključiti u ovom radu, ali na razini države.

### **Hipoteza 5: postotak korisnika interneta ima pozitivan i statistički značajan utjecaj na indeks otvorenosti proračuna.**

Caba Pérez s drugim autorima [27] pokazuje pozitivnu vezu između pristupa internetu te transparentnosti proračuna. Istraživanje je provedeno na španjolskim provincijama te je zaključak da provincije nisu u potpunosti svjesne značaja interneta za transparentnost proračuna. Očekuje se da će upotreba interneta rezultirati time da državne službe postanu izravno i lako dostupne javnosti, što znači da će se informacije lakše i brže širiti te koristiti javnosti. Dostupnost takvih državnih financijskih informacija znači i da javnost može kontinuirano ocjenjivati vladine agencije. Veća internetska transparentnost državnog proračuna može vratiti povjerenje građana u vladu te ih uključiti u proces donošenja odluka.

Internet je pružio dodatni poticaj za objavljivanje proračunskih podataka. Može se reći da je porast interneta povećao transparentnost proračuna omogućavajući brzo i jeftino proaktivno otkrivanje (Darbishire,[14]).

Sve veća dostupnost interneta je dodatni poticaj za objavljivanje proračunskih podataka te se može zaključiti da veći postotak korisnika interneta vodi većoj transparentnosti proračuna.

### **Hipoteza 6: Indeks ljudskog razvoja ima pozitivan i statistički značajan utjecaj na indeks otvorenosti proračuna**

Indeks ljudskog razvoja (Human Development Index-HDI [35]) koji je geometrijska sredina normaliziranih indeksa sljedeće tri stavke: očekivanog životnog vijeka, stupanja obrazovanja te dohotka je pokazatelj stupnja razvijenosti pojedine zemlje. Mnogi preko tog indeksa svrstavaju zemlje u razvijene (zemlje prvog reda), u razvoju (zemlje drugog reda) i treće zemlje svijeta.

Kako se u ovom radu pretpostavlja da razvijenost zemalja pozitivno utječe na transparentnost, tako se pretpostavlja da i kombinacija faktora poput pismenosti, siromaštva, obrazovanja i drugih koje obuhvaća HDI ima pozitivan učinak na transparentnost državnog proračuna.

### **Hipoteza 7: Indeks percepcije korupcije ima pozitivan i statistički značajan utjecaj na indeks otvorenosti proračuna**

Indeks percepcije korupcije (Corruption Perceptions Index–CPI [36]) na godišnjoj razini rangira zemlje po njihovoj percipiranoj razini korupcije u javnom sektoru. Indeks rangira zemlje na skali od 100 (bez korupcije) do 0 (visoka korumpiranost). U ovom radu će se pokušati utvrditi veza između korupcije i transparentnosti državnog proračuna pod pretpostavkom da korupcija negativno utječe na transparentnost. Veća korupcija zemlji dodjeljuje manji indeks percepcije korupcije te se zbog toga u hipotezi pretpostavlja pozitivan utjecaj indeksa percepcije korupcije na indeks otvorenosti proračuna.

# Poglavlje 2

## Podaci

Cilj rada je odrediti socio-ekonomske odrednice koje su u korelaciji i imaju statistički značajan utjecaj na zavisnu varijablu koju promatramo, a to je IOP. IOP koji nam pokazuje razinu transparentnosti državnog proračuna pojedine zemlje modeliramo kao slučajnu varijablu koju nazivamo ovisnom varijablom ili odzivom te sva ostala mjerenja zovemo neovisnim varijablama, predviditeljima ili kovarijatama.

Kao što je spomenuto u Poglavlju 1, na temelju prethodnih istraživanja, promatraju se sljedeće neovisne varijable: stopa nezaposlenosti, BDP po stanovniku, udio javnog duga u BDP-u, broj stanovnika, postotak korisnika interneta, HDI te CPI. Poglavlje 2 daje deskriptivnu statistiku za pojedinu varijablu, njihov opis te izvor podataka.

### 2.1 Zavisna varijabla

Zavisnu varijablu u radu predstavlja IOP (engl OBI-Open Budget Index), kao jedini neovisni i međunarodno usporediv pokazatelj kvalitete i opsežnosti online informacija o proračunu te tako daje uvid u stanje otvorenosti proračuna diljem svijeta. Upitnik o otvorenosti proračuna, na temelju kojeg je dobiven IOP, kao i rezultati provedenog istraživanja javno su dostupni i tako daju mogućnost korištenja svakome tko želi prepoznati trendove, napredak ili pronaći konkretne načine za poboljšanje otvorenosti proračuna.

U ovom radu IOP je promatran za 117 zemalja uključenih u istraživanje te se u obzir uzimaju zadnji dostupni podaci, a to su podaci za 2019.

IOP definiramo kao kategorijalnu varijablu sa sljedećim kategorijama:

- (1) 0-20 neznatne informacije,
- (2) 21-40 vrlo malo informacija,
- (3) 41-60 ograničene informacije,

- (4) 61-80 značajne informacije,
- (5) 81-100 opsežne informacije.

## 2.2 Nezavisne varijable

Sukladno već navedenim socio-ekonomskim varijablama, ovo poglavlje daje njihov kratak opis, osnovne informacije, izvor podataka te kratice koje će biti korištene u radu. Nakon tog prikazuje se deskriptivna statistika pojedine nezavisne varijable.

Stopa nezaposlenosti je broj nezaposlenih radnika podijeljen s ukupnim brojem radno sposobnog stanovništva (radno sposobnim stanovništvom se smatraju osobe između 16 i 65 godina), tj. odnosi se na udio radne snage koja nema posao, ali je dostupna i traži posao. Podaci su uzeti za 2018., izvor je Međunarodna organizacija rada [21].

BDP po stanovniku je ukupan BDP zemlje podijeljen s brojem stanovnika. Podaci su izraženi u američkim dolarima za 2018. te se mogu naći na stranicama Svjetske banke [34].

Javni se dug obično izražava u omjeru prema BDP-u, a taj relativni pokazatelj često služi za usporedbu zaduženosti među državama. Dostupni podaci za udio javnog duga u BDP-u, za sve zemlje obuhvaćene istraživanjem, su iz 2018.

Broj stanovnika je podatak za 2018. te je izražen u milionima. Podaci broja stanovnika te udjela javnog duga u BDP-u su dostupni na stranici Trading Economics [35].

Podaci za pojedinu zemlju te njen postotak korisnika interneta dani su za 2017. godinu, izvor je Međunarodna telekomunikacijska unija te se također mogu naći na stranici Svjetske banke [34, 22].

HDI je stvoren kako bi naglasio da ljudi i njihove sposobnosti te kompetencije trebaju biti jedan od kriterija za ocjenu razvoja zemlje, a ne samo ekonomski rast. HDI je sažeto mjerilo prosječnog postignuća u ključnim dimenzijama ljudskog razvoja: dug i zdrav život, znanje i prihvatljiv životni standard. Zdravstvena dimenzija procjenjuje se očekivanim životnim vijekom od rođenja, a dimenzija obrazovanja mjeri se prosjekom godina školovanja za odrasle u dobi od 25 godina i više i očekivanim godinama školovanja za djecu koja kreću u školu. Dimenzija životnog standarda mjeri se bruto nacionalnim dohotkom po stanovniku. Bodovi za tri dimenzije HDI-a tada se objedinjuju u jedan indeks koristeći geometrijsku sredinu. Svakoj državi daje bodove u rasponu od 0 do 1. Podaci u ovom radu su iz 2018. godine te su preuzeti sa stranice Program Ujedinjenih naroda za razvoj (United Nations Development Programme) [35].

Transparency International od 1995. objavljuje CPI [36] koji na godišnjoj razini rangira zemlje po njihovoj percipiranoj razini korupcije u javnom sektoru, u skladu sa stručnim



procjenama i istraživanjima. Indeks rangira 180 zemalja na skali od 100 (bez korupcije) do 0 (visoka korumpiranost). Na ljestvici od 180 država prema CPI 2018 Hrvatska je na 60. mjestu s 48 bodova. Najtransparentnije države u 2018. su Danska s 88, Novi Zeland s 87, Finska, Singapur, Švedska i Švicarska s 85 bodova. Najkorumpiranije države su Sirija s 13, Južni Sudan s 12, te Somalija sa samo 10 bodova. Hrvatska se i dalje nalazi među korumpiranim državama Europske unije, a osim nas lošije su Rumunjska i Mađarska s 44 boda, te Bugarska s 43 boda.

Zbog širokog raspona podataka nezavisnih varijabli te da bi se olakšala preglednost podataka neki su podaci skalirani. BDP po stanovniku je podijeljen s 1000 te je broj stanovnika izražen u milionima. Stopa nezaposlenosti te udio javnog duga u BDP-u izraženi su u postotku.

Tablica 2.1 daje popis nezavisnih varijabli te ukratko do sad navedeno.

Oznaka varijable	Naziv	Izvor	Razdoblje
stopa_nez	Stopa nezaposlenosti	Međunarodna organizacija rada	2018.
BDP	BDP po stanovniku	Svjetska banka	2018.
dug	Udio javnog duga u BDP-u	Trading Economics	2018.
br_stan	Broj stanovnika	Trading Economics	2018.
net	Postotak korisnika interneta	Međunarodna telekomunikacijska unija	2017.
HDI	Indeks ljudskog razvoja	Program UN-a za razvoj	2018.
CPI	Indeks percepcije korupcije	Transparency International	2018.

Tablica 2.1: Kratak opis nezavisnih varijabli

## 2.3 Deskriptivna statistika

Deskriptivna statistika je dio matematičke statistike koji se koristi za opisivanje i bolje razumijevanje izmjerenog (ili zadanog) skupa podataka te će u ovom poglavlju biti dana za sve korištene varijable u radu.

Broj država po pojedinim kategorijama, tj. frekvencija i relativna frekvenciju varijable IOP prikazana je u tablici 2.2. Tablica 2.3 prikazuje deskriptivnu statistiku varijable IOP te Tablica 2.4 prikazuje deskriptivnu statistiku pojedine nezavisne varijable.

Iz prikazanih podataka uočavamo da samo 31 zemlja ima zadovoljavajuću razinu transparentnosti te zbog toga mnoga istraživanja rade na analizi odrednica transparentnosti ne bi li pridonjeli njenom poboljšanju. Najviše država svojim građanima omogućuje ograničene

Kategorije	Frekvencija	Relativna frekvencija
(1)	20	0.171
(2)	30	0.256
(3)	36	0.308
(4)	25	0.214
(5)	6	0.051

Tablica 2.2: Relativne frekvencije zavisne varijable

informacije tj. pripada kategoriji (3), dok čak 17% država pruža neznatne ili nikakve informacije.

Duljina uzorka	Minimum	Aritmetička sredina	Medijan	Maximum	St. devijacija
117	0	44.64	46	87	22.68

Tablica 2.3: Deskriptivna statistika varijable IOP

Varijabla	Minimum	Aritmetička sredina	Medijan	Maximum	St. devijacija
stopa_nez	0.001	0.069	0.052	0.269	0.053
BDP	0.270	10.749	4.230	81.730	15.824
dug	0.040	0.552	0.490	2.380	0.316
br_stan	0.210	49.230	17.745	1392.270	136.594
net	0.005	0.492	0.510	0.974	0.272
HDI	0.351	0.689	0.712	0.954	0.151
CPI	10.000	39.650	36.000	87.000	16.956

Tablica 2.4: Deskriptivna statistika nezavisnih varijabli

Kako bi se intuitivno mogla prikazati veza između zavisne i nezavisnih varijabli, iduća tablica prikazuje aritmetičke sredine nezavisnih varijabli u odnosu na kategoriju varijable IOP.

Kategorija	stopa_nez	BDP	dug	br.stan	net	HDI	CPI
1	0.071	7.962	0.586	89.180	0.357	0.594	27.100
2	0.075	3.089	0.444	36.121	0.333	0.600	34.200
3	0.069	7.003	0.538	23.750	0.509	0.700	39.028
4	0.054	25.054	0.684	64.002	0.709	0.828	53.200
5	0.112	21.217	0.515	68.702	0.723	0.813	56.000

Tablica 2.5: Aritmetičke sredine nezavisnih varijabli po kategorijama

Odmah se može uočiti da se aritmetička sredina postotka korisnika interneta, HDI-a te CPI-a povećava kako se povećava kategorija IOP-a dok se kod varijable stopa nezaposlenosti uočava pad. BDP po stanovniku je veći u zadnje dvije kategorije, dok se kod preostalih varijabli ne uočava veza.

# Poglavlje 3

## Metodologija

U ovom radu sve kovarijate koje se promatraju poprimaju numeričke vrijednosti dok je OBI, koju modeliramo kao slučajnu varijablu, kategorijalna varijabla s 5 kategorija. Da bi se odredilo koje kovarijate i kako značajno utječu na OBI koristit će se ordinalna logistička regresija kao vrsta multinomne logističke regresije. Multinomni logistički model je specijalan slučaj multivarijatnog generaliziranog linearnog modela zbog čega ovo poglavlje započinje potpoglavljem Generalizirani linearni modeli te nakon tog slijedi detaljni opis multinomne logističke regresije, način procjene parametara modela kao i kvalitete modela. Poglavlje 3 je temeljeno na [1, 7, 8, 12, 13].

### 3.1 Generalizirani linearni modeli

Kod linearne regresije pretpostavlja se da slučajna zavisna varijabla  $Y$  na linearan način ovisi o numeričkim kovarijatama  $X_1, X_2, \dots, X_p$  tj.

$$Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \epsilon_i, \quad i = 1, \dots, N \quad (3.1)$$

gdje  $(\epsilon_i)_i$  čine niz nezavisnih, jednako distribuiranih slučajnih varijabli s očekivanjem 0 i variancom  $\sigma^2$ . Zapisano matrično

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

gdje su  $Y^T = (Y_1, Y_2, \dots, Y_N)$   $N$  međusobno nezavisnih opažanja od  $Y$ ,  $\beta^T = (\beta_0, \beta_1, \dots, \beta_p)$  vektor parametara modela te  $X = [\mathbf{1}, \mathbf{x}_1, \dots, \mathbf{x}_p]$  matrica s vrijednostima kovarijata, tj. podatkovna matrica, pri čemu je  $\mathbf{x}_j$  vektor stupac realizacije  $j$ -te kovarijate.  $\epsilon$  označava vektor slučajnih grešaka.

Osnovna ideja linearnih modela je pretpostavka da postoji linearna veza između očekivanja odziva i kovarijata tj.

$$EY_i = \sum_{j=1}^p \beta_j x_{ij}, \quad (3.2)$$

gdje je  $p$  broj mogućih kovarijata. Generalizirani linearni modeli su proširenje linearnih modela te se kod njih pretpostavlja sljedeće:

$$EY_i = g^{-1} \left( \sum_{j=1}^p \beta_j x_{ij} \right), \quad (3.3)$$

gdje je  $g^{-1}$  inverz tzv. funkcije veze  $g$ ,  $\sum_{j=1}^p \beta_j x_{ij}$  tzv. linearni prediktor te gdje za zadano očekivanje,  $Y_i$  ima unaprijed određenu razdiobu iz tzv. eksponencijalne familije. Iako iz teorijskih razloga, slučajna komponenta odziva mora imati razdiobu iz neke od eksponencijalnih familija, ova restrikcija je u praksi često prihvatljiva jer ove familije uključuju najčešće korištene razdiobe.

### Eksponecijalne familije

Slučajna varijabla  $Y$  pripada nekoj eksponencijalnoj familiji ako joj gustoća (neprekidna ili diskretna) ima sljedeći oblik:

$$f(y; \theta; \varphi) = e^{\frac{y\theta - b(\theta)}{a(\varphi)} + c(y, \varphi)}, \quad (3.4)$$

za neke funkcije  $a, b$  i  $c$ . Familija ima dva parametra:  $\theta$ , tzv. prirodni parametar i  $\varphi$  tzv. parametar disperzije ili skaliranja. Funkcija  $b$  u (3.4) je uvijek dvaput neprekidno diferencijabilna i t.d. je  $b'$  invertibilna. Funkcija  $a$  parametra  $\varphi$  zove se funkcija disperzije, a omogućuje dodatnu fleksibilnost u modelu. Funkciju  $c$  često ignoriramo jer nema utjecaja u procesu procjene parametara GLMa. Za funkciju log-vjerodostojnosti  $l(y; \theta, \varphi) = \log(f(y; \theta, \varphi))$  unutar neke eksponencijalne familije vrijede sljedeća dva rezultata iz statističke teorije:

$$E \left[ \frac{\partial l}{\partial \theta^2} \right] = 0, E \left[ \frac{\partial^2 l}{\partial \theta^2} \right] + E \left[ \left( \frac{\partial l}{\partial \theta} \right)^2 \right] = 0, \quad (3.5)$$

pri čemu je za sve  $\theta$  i  $\varphi$ ,

$$E \left[ \frac{\partial l}{\partial \theta} \right] := E \left[ \frac{\partial}{\partial \theta^2} l(Y; \theta, \varphi) \right]. \quad (3.6)$$

Analogno se definiraju  $E \left[ \frac{\partial^2 l}{\partial \theta^2} \right]$  i  $E \left[ \left( \frac{\partial l}{\partial \theta} \right)^2 \right]$ . Za funkciju log-vjerodostojnosti eksponencijalnih familija vrijedi

$$\frac{\partial l}{\partial \theta} = \frac{y - b'(\theta)}{a(\varphi)}. \quad (3.7)$$

Iz prve jednakosti u (3.5) slijedi

$$\mu = E[Y] = b'(\theta), \quad (3.8)$$

što je ekvivalentno s  $\theta = b'^{-1}(\mu)$ . Kako vrijedi i  $\frac{\partial^2 l}{\partial \theta^2} = \frac{-b''(\theta)}{a(\varphi)}$ , također iz (3.5) imamo

$$\text{Var}(Y) = a(\varphi)b''(\theta) \quad (3.9)$$

Dakle, očekivanje ne ovisi o  $\varphi$ , dok varijanca općenito ovisi o oba parametra. Neke od distribucija koje pripadaju eksponencijalnoj familiji su normalna, Poissonova, binomna. . .

## 3.2 Logistička regresija

Regresija se upotrebljava u svrhu opisivanja i predviđanja zavisne varijable ili varijable odziva na temelju skupa nezavisnih varijabli. Ako je zavisna varijabla kategorijalna i dihotomna, tj. može poprimiti jednu od dvije vrijednosti koristi se logistička regresija. Kako bi se dobio logistički regresijski model izvode se matematičke transformacije zavisne varijable u linearnom regresijskom modelu. Vrijednosti zavisne varijable postaju prirodni logaritmi izgleda (engl. natural logarithm of the odds ratio). Model logističke regresije pripada generaliziranim linearnim modelima. Logistički model se koristi za predviđanje vjerojatnosti događaja pomoću prilagođavanja podataka logističkoj krivulji prepoznatljivijoj po svom S-obliku. Funkcija veze kojom je određen model logističke regresije je logit funkcija:

$$g(\pi_i) = \log \frac{\pi_i}{1 - \pi_i} = \mathbf{X}_i^T \boldsymbol{\beta}, \quad i = 1, \dots, N \quad (3.10)$$

gdje je  $\pi_i$  vjerojatnost realizacije uspjeha pri  $i$ -tom promatranju,  $\mathbf{X}_i$  vektor vrijednosti neovisnih varijabli, a  $\boldsymbol{\beta}$  vektor parametara. Izraz pod algoritmom naziva se i izgled ili sklonost (engl. odds). U ovom slučaju izgled se definira kao omjer vjerojatnosti realizacije uspjeha i vjerojatnosti realizacije neuspjeha.

$\pi_i$  možemo izraziti iz jednakosti (3.10) te se dobiva jednadžba kojom je definiran model logističke regresije:

$$\pi_i = \frac{e^{\mathbf{X}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{X}_i^T \boldsymbol{\beta}}}, \quad i = 1, \dots, N. \quad (3.11)$$

Logit funkcija poprima vrijednosti iz čitavog skupa  $\mathbb{R}$ , a vjerojatnost iz (3.11) poprima vrijednosti iz skupa  $[0,1]$ .

### 3.3 Multinomna logistička regresija

Multinomna logistička regresija koristi se kad je varijabla odziva kategorijalna s barem dvije kategorije, tj.  $J \geq 2$ . Za  $J=2$  naziva se binomna logistička regresija. U ovisnosti o kakvim se kategorijama radi postoje dvije vrste multinomne logističke regresije: nominalna i ordinalna logistička regresija. Kod nominalne logističke regresije ne postoji skala ili redoslijed po kojem je neka kategorija više ili manje vrijedna od druge (npr. plava, zelena, žuta boja), dok kod ordinalne postoji ta skala ili redoslijed kao u kategorijama varijable IOP.

Neka je  $Y$  slučajna varijabla s  $J \geq 2$  kategorija. Neka su  $\pi_j = \mathbf{P}(Y = j)$ ,  $j = 1 \dots J$  odgovarajuće vjerojatnosti tako da vrijedi  $\sum_{j=1}^J \pi_j = 1$ . Promotrimo  $N$  nezavisnih realizacija slučajne varijable  $Y$  tako da s  $y_1$  označimo broj realizacija prve kategorije, s  $y_2$  broj realizacija druge kategorije... Na taj način dolazimo do vektora  $\mathbf{y} = [y_1, y_2, y_3 \dots y_J]^T$ , gdje je  $y_j \in \{0, 1 \dots N\}$  i  $\sum_{j=1}^J y_j = N$ .

**Definicija 3.3.1.** Vektor  $\mathbf{Y}$  ima multinomnu distribuciju s parametrima  $N \in \mathbb{N}$  i  $\pi \in \mathbb{R}^J$  ako ima sljedeću funkciju gustoće:

$$f(\mathbf{y}|N) = \frac{N!}{y_1! y_2! \dots y_J!} \pi_1^{y_1} \pi_2^{y_2} \dots \pi_J^{y_J}.$$

Ukoliko je  $J=2$  multinomna distribucija svodi se na binomnu.

Multinomni logistički model je specijalan slučaj multivarijantnog generaliziranog modela. Neka su  $X_1, X_2, \dots, X_K$  nezavisne varijable, zavisna varijabla  $Y_i$  je  $J$ -dimenzionalni vektor s očekivanjem  $\mu_i = \mathbb{E}(Y_i | \mathbf{X}_i)$ , a s  $\mathbf{X}$  označavamo podatkovnu matricu čiji je  $i$ -ti redak  $\mathbf{X}_i^T$ . Distribucijska pretpostavka je da su  $\mathbf{X}_i^T$  i  $Y_i$  nezavisni i da  $Y_i$  ima distribuciju koja pripada eksponencijalnoj familiji, tj. ima formu kao u izrazu (3.4).

Strukturna pretpostavka je da je očekivanje  $\mu_i$  određeno linearnim prediktorom  $\eta_i = \mathbf{X}_i^T \boldsymbol{\beta}_i$  u obliku  $\mu_i = h(\eta_i)$  za neku injektivnu i glatku funkciju  $h$ , odnosno funkcija veze je definirana kao inverz funkcije  $h$ ,  $g(\mu_i) = \eta_i$ , pri čemu je  $\mathbf{X}$  već navedena podatkovna matrica koja se sastoji od vektora  $\mathbf{1}, \mathbf{X}_1, \dots, \mathbf{X}_N$ , a  $\boldsymbol{\beta}$  matrica nepoznatih parametara koja se sastoji od vektora  $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_K$ .

#### Ordinalna logistička regresija

Postoji više modela ordinalne logističke regresije, kumulativni odnosno proportional odds model je najčešće korišten te će biti korišten u radu. Ostali modeli su modifikacije navedenog modela.

Neka je  $Y_i, i = 1, \dots, N$ , zavisna ordinalna varijabla koja poprima jednu od  $J$  kategorija. Odgovarajuće vjerojatnosti realizacije svake od kategorija pri  $i$ -tom promatranju su

$\pi_{i1} = \mathbb{P}(Y_i = 1), \dots, \pi_{iJ} = \mathbb{P}(Y_i = J)$ . Distribucija slučajne varijable  $Y_i$  je multinomna s parametrom  $\pi_i = (\pi_{i1}, \dots, \pi_{iJ})$ . Za ordinalnu logističku regresiju ključne su kumulativne vjerojatnosti koje se definiraju na sljedeći način:

$$\mathbb{P}(Y_i \leq j) = \pi_{i1} + \pi_{i2} + \dots + \pi_{ij} \quad (3.12)$$

gdje je  $j = 1, \dots, J$  realizirana kategorija zavisne varijable. Kumulativne vjerojatnosti odražavaju redoslijed ordinalnih varijabli:  $\mathbb{P}(Y_i \leq 1) \leq \mathbb{P}(Y_i \leq 2) \leq \dots \leq \mathbb{P}(Y_i \leq J) = 1$ . Kod ovog modela promatra se izgled ili sklonost za svaku pojedinu kategoriju zavisne varijable na način da se gleda omjer sume vjerojatnosti realizacije manjih kategorija i sume vjerojatnosti realizacije većih od njih. (3.10) omogućuje da se svaki izgled zapiše na sljedeći način:

$$\frac{\mathbb{P}(Y_i \leq j)}{\mathbb{P}(Y_i > j)} = \frac{\mathbb{P}(Y_i \leq j)}{1 - \mathbb{P}(Y_i \leq j)} = \frac{\pi_{i1} + \pi_{i2} + \dots + \pi_{ij}}{\pi_{i(j+1)} + \dots + \pi_{iJ}}. \quad (3.13)$$

Time se mjeri koliki je izgled realizacije kategorije koja je manja ili jednaka od  $j$ -te u odnosu na realizaciju kategorije koja je strogo veća od  $j$ -te. Uključe li se nezavisne varijable u model dobiva se konačni kumulativni model logističke regresije:

$$\text{logit}(\mathbb{P}(Y_i \leq j)) := \log\left(\frac{\mathbb{P}(Y_i \leq j)}{1 - \mathbb{P}(Y_i \leq j)}\right) = \beta_{0j} + \beta_{1j}x_{i1} + \beta_{2j}x_{i2} + \dots + \beta_{Kj}x_{iK}, \quad j \in \{1, \dots, J-1\}, \quad (3.14)$$

gdje je  $K$  broj nezavisnih varijabli.

Važna pretpostavka ovog modela je da koeficijenti  $\beta_{1j}, \beta_{2j}, \dots, \beta_{Kj}$  ne ovise o kategoriji  $j$ , nego su jednaki za svaki  $j = 1, \dots, J$ , za razliku od koeficijenata  $\beta_{0j}$  koji ovise o kategoriji  $j$  i variraju za svaku od funkcija. Koeficijenti  $\beta_{0j}$  su kao slobodni član u linearnom regresijskom modelu pa iz tog razloga jednakost (3.12) možemo zapisati tako da slobodni član  $\beta_{0j}$  zamijenimo s  $\alpha_j$ :

$$\log\left(\frac{\mathbb{P}(Y_i \leq j)}{1 - \mathbb{P}(Y_i \leq j)}\right) = \alpha_j + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_K x_{iK}. \quad (3.15)$$

Dakle, vektor  $\alpha$  je oblika  $(\alpha_1, \alpha_2, \dots, \alpha_{J-1})$ .

Iz jednadžbe (3.13) mogu se izraziti kumulativne vjerojatnosti u ovisnosti o poznatim vrijednostima nezavisnih varijabli:

$$\mathbb{P}(Y_i \leq j) = \frac{e^{\alpha_j + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_K x_{iK}}}{1 + e^{\alpha_j + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_K x_{iK}}}. \quad (3.16)$$

Budući da se svaka pojedina vjerojatnost može izraziti pomoću kumulativnih jer vrijedi  $\mathbb{P}(Y_i = j) = \mathbb{P}(Y_i \leq j) - \mathbb{P}(Y_i \leq j-1)$ , lako se pokaže da za slobodne članove vrijedi  $\alpha_j < \alpha_{j+1}$  za svaki  $j \in \{1, \dots, J-1\}$ .



### 3.4 Procjena parametara

Da bi se pojednostavili proračunski detalji procjene, podaci se objedinjuju tako da svaki redak predstavlja jednu različitu kombinaciju vrijednosti nezavisnih varijabli. Ti retci se najčešće nazivaju "populacija".  $N$  je ukupan broj populacija, a  $\mathbf{n}$  vektor stupac s  $n_i$  elemenata tako da vrijedi  $\sum_{i=1}^N n_i = M$ , gdje je  $M$  veličina uzorka.  $\mathbf{y}$  je matrica s  $N$  redaka (za pojedinu populaciju) i  $J - 1$  stupaca. Za svaku populaciju  $y_{ij}$  predstavlja realizaciju kategorije  $j$  pri  $i$ -tom promatranju.

$$\mathbf{y} = \begin{bmatrix} y_{11} & \dots & y_{1(J-1)} \\ \vdots & & \vdots \\ y_{N1} & \dots & y_{N(J-1)} \end{bmatrix}$$

Slično, matrica  $\boldsymbol{\pi}$  je istih dimenzija gdje je svaki  $\pi_{ij}$  vjerojatnost realizacije  $j$ -te vrijednosti u  $i$ -tom promatranju, tj.  $\pi_{ij} = \mathbb{P}(Y_i = j)$  i vrijedi  $\sum_{j=1}^J \pi_{ij} = 1$ .

$$\boldsymbol{\pi} = \begin{bmatrix} \pi_{11} & \dots & \pi_{1(J-1)} \\ \vdots & & \vdots \\ \pi_{N1} & \dots & \pi_{N(J-1)} \end{bmatrix}$$

Matrica nezavisnih varijabli  $\mathbf{X}$  je dimenzije  $N(K + 1)$ , gdje je  $K$  broj neovisnih varijabli. Prvi stupac sadrži samo jedinice,  $x_{i0}=1$ , za svaki  $i \in \{0, \dots, N\}$ , jer se veže uz slobodni član.

$$\mathbf{X} = \begin{bmatrix} x_{10} & \dots & x_{1K} \\ \vdots & & \vdots \\ x_{N0} & \dots & x_{NK} \end{bmatrix}$$

Matrica parametara  $\boldsymbol{\beta}$  je dimenzija  $(K + 1)(J - 1)$ . Parametar  $\beta_{kj}$  veže se uz  $k$ -tu nezavisnu varijablu i  $j$ -tu vrijednost ovisne varijable,  $k \in \{0, \dots, K\}$  i  $j \in \{1, \dots, J - 1\}$ .

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_{01} & \dots & \beta_{0(J-1)} \\ \vdots & & \vdots \\ \beta_{K1} & \dots & \beta_{K(J-1)} \end{bmatrix}.$$

Kod multinomne logističke regresije linearnu komponentu izjednačavamo s logaritmom od sklonosti.  $J$ -ta kategorija se uzima kao bazna. Logaritam od sklonosti prvih  $J - 1$  kategorija

ima sljedeći oblik:

$$\log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) = \log\left(\frac{\pi_{ij}}{1 - \sum_{j=1}^{J-1} \pi_{ij}}\right) = \sum_{k=0}^K x_{ik}\beta_{kj}, \quad i = 1, \dots, N. \quad (3.17)$$

Rješavanjem po  $\pi_{ij}$  dobije se sljedeće:

$$\pi_{ij} = \frac{e^{\sum_{k=0}^K x_{ik}\beta_{kj}}}{1 + \sum_{j=1}^{J-1} e^{\sum_{k=0}^K x_{ik}\beta_{kj}}}, \quad j < J \quad (3.18)$$

odnosno za kategoriju  $J$  vrijedi

$$\pi_{iJ} = 1 - \sum_{j=1}^{J-1} \pi_{ij} = \frac{1}{1 + \sum_{j=1}^{J-1} e^{\sum_{k=0}^K x_{ik}\beta_{kj}}}. \quad (3.19)$$

Procjena parametara multinomnog logističkog modela provodi se metodom maksimalne vjerodostojnosti. Procjenjuje se skup parametara  $\beta_{kj}$ ,  $k = 0, \dots, K$ ,  $j = 1, \dots, J - 1$ . Neka su  $Y_1, Y_2, \dots, Y_N$  međusobno nezavisne varijable koje zadovoljavaju svojstva generaliziranih linearnih modela. Ako je  $f(y_1, \dots, y_N)$  njihova zajednička funkcija gustoće, funkciju maksimalne vjerodostojnosti definiramo kao funkciju parametara  $\boldsymbol{\beta}$  na sljedeći način:

$$L(\boldsymbol{\beta}) = f(y_1, \dots, y_N | \boldsymbol{\beta}). \quad (3.20)$$

Obzirom da su varijable međusobno nezavisne i jednako distribuirane, izraz (3.20) se može pojednostaviti:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^N f(y_i | \boldsymbol{\beta}). \quad (3.21)$$

Prema definiciji multinomne distribucije, zajednička funkcija gustoće, odnosno funkcija maksimalne vjerodostojnosti multinomne distribucije ima sljedeći oblik:

$$f(\mathbf{y} | \boldsymbol{\beta}) = \prod_{i=1}^N \left[ \frac{n_i!}{\prod_{j=1}^J y_{ij}!} \prod_{j=1}^J \pi_{ij}^{y_{ij}} \right], \quad (3.22)$$

gdje je  $N$  broj realizacija, a  $J$  broj kategorija ovisne varijable. Obzirom da je cilj maksimizirati gornju funkciju u terminima parametra  $\boldsymbol{\beta}$  koji ne ovisi o  $n_i$  niti o  $y_{ij}$ , faktorijalni izrazi mogu se tretirati kao konstanta. Zbog tog je dovoljno maksimizirati "jezgru" funkcije maksimalne vjerodostojnosti,

$$f(\mathbf{y} | \boldsymbol{\beta}) = \prod_{i=1}^N \prod_{j=1}^J \pi_{ij}^{y_{ij}}, \quad (3.23)$$

što se može zapisati na sljedeći način:

$$\prod_{i=1}^N \prod_{j=1}^{J-1} \left( \frac{\pi_{ij}}{\pi_{iJ}} \right)^{y_{ij}} \pi_{iJ}^{n_i}. \quad (3.24)$$

Uvrštavanjem jednakosti (3.18) i (3.19) za  $\pi_{ij}$  i  $\pi_{iJ}$  dobiva se sljedeće:

$$\sum_{i=1}^N \sum_{j=1}^{J-1} e^{y_{ij} \sum_{k=0}^K x_{ik} \beta_{kj}} \left( 1 + \sum_{j=1}^{J-1} e^{\sum_{k=0}^K x_{ik} \beta_{kj}} \right)^{-n_i}. \quad (3.25)$$

**Definicija 3.4.1.** Procjenitelj maksimalne vjerodostojnosti parametra  $\beta$  je vrijednost spomenutog parametra koji maksimizira funkciju vjerodostojnosti,

$$\hat{\beta} = \underset{\beta \in \mathbf{R}^{(K+1) \times (J-1)}}{\operatorname{argmax}} L(\beta). \quad (3.26)$$

Umjesto maksimiziranja produkta (3.21) često se maksimizira funkcija log-vjerodostojnosti jer je logaritam rastuća funkcija, stoga je ekvivalentno maksimizirati logaritam funkcije vjerodostojnosti:

$$l(\beta) = \sum_{i=1}^N \log(f(y_i | \beta)). \quad (3.27)$$

Logaritmirajući jednakost (3.25) dobije se logaritam funkcije vjerodostojnosti modela multinomne logističke regresije:

$$l(\beta) = \sum_{i=1}^N \sum_{j=1}^{J-1} \left[ \left( y_{ij} \sum_{k=0}^K x_{ik} \beta_{kj} \right) - n_i \log \left( 1 + \sum_{j=1}^{J-1} e^{\sum_{k=0}^K x_{ik} \beta_{kj}} \right) \right]. \quad (3.28)$$

Da bi se pronašli procjenitelji maksimalne vjerodostojnosti, tj. vrijednosti za parametre  $\beta_{kj}$  koji će maksimizirati funkciju (3.28) koristi se Newton-Raphsonova metoda za koju treba uvesti nekoliko novih pojmova, statistički model morao bi zadovoljavati neke dodatne uvjete te se računaju prva i druga derivacije logaritma funkcije vjerodostojnosti. U ovom radu nećemo se detaljno baviti tom metodom (za više detalja vidi[13]).

### 3.5 Kvaliteta i odabir modela

Nakon procjene parametara modela odabire se najpogodniji model te se ispituje kvaliteta modela. Postoje razni statistički testovi za ocjenu kvalitete modela, u ovom poglavlju će biti navedene neke metode koje su naveli Barnett i Dobson[8].

## Metode odabira modela

Postoje tri metode za odabir modela: *unaprijed* (*engl.forward*), *unatrag* (*engl.backward*) i *korak po korak* (*engl.stepwise*). Svaka od metoda koristi kriterij za dodavanje ili izbacivanje kovarijata u model. Najčešći kriterij za logističke regresijske modele je  $p$ -vrijednost testa značajnosti za određenu kovarijatu. Kovarijate koje imaju  $p$ -vrijednost koja prelazi određenu razinu isključene su iz modela. Jednako tako, kovarijate koje imaju  $p$ -vrijednost manju od neke razine zadržavaju se u modelu. Najčešće korištene razine su 0.05, 0.1 te 0.15.

*Metoda odabira unaprijed* počinje praznim modelom te u prvom koraku dodaje kovarijatu koja je najznačajnija za model. U svakoj sljedećoj fazi razvoja modela svaka kovarijata se testira, zajedno s prethodno prihvaćenim je li za uključivanje u model. Najznačajnija od preostalih kovarijata dodaje se u model, sve dok je  $p$ -vrijednost ispod određene razine. *Metoda odabira unatrag* kreće s potpunim modelom, tj. s modelom koji sadrži sve kovarijate te u svakom sljedećem koraku izbacuje po jednu varijablu koja nije značajna za model. *Metoda korak po korak* je kombinacija prethodne dvije metode. U svakom koraku omogućuje ubacivanje novih ili izbacivanje već uključenih kovarijata. Kod problema korelacije, odnosno u slučaju kada više varijabli ima slično značenje, stepwise regresija eliminira one varijable koje se preklapaju s drugima.

## Kvaliteta modela

U prethodnim metodama  $p$ -vrijednost može se odrediti obzirom na  $z$ ,  $t$  ili *Waldovu* statistiku ili izračunavanjem omjera vjerodostojnosti. Osim korištenja  $p$ -vrijednosti još neki kriterij za uspoređivanje modela su *AIC* (*Akaike infomacijski kriterij*) i *BIC* (*Bayesov informacijski kriterij*). Formula za određivanje AIC indeksa je:

$$AIC = 2p - 2l(\beta), \quad (3.29)$$

gdje je  $p$  broj procijenjenih parametara uključujući slobodni član, a  $l(\beta)$  log-vjerodostojnost modela koji testiramo. Kad se uspoređuju dva modela pomoću Akaikeovog kriterija, bolji je onaj model čiji je AIC indeks manji.

Formula za određivanje BIC indeksa je:

$$BIC = p * \log(N) - 2l(\beta), \quad (3.30)$$

gdje je  $N$  broj promatranja. Za model koji ima BIC indeks manji reći ćemo da je bolji.

ANOVA (analysis of variance) je termin korišten za statističku metodu kojom uspoređujemo očekivanja dvije ili više grupa. Nulta hipoteza ovog testa su jednaka očekivanja promatranih grupa.

Kvaliteta modela može se procijeniti i rezidualima koji su definirani na sljedeći način:

$$r_i = \frac{o_i - e_i}{\sqrt{e_i}}, \quad (3.31)$$

gdje je  $o_i$  promatrana vrijednost, a  $e_i$  očekivana odnosno procijenjena vrijednost za  $i$ -to promatranje. Što su reziduali manji, model je bolji.

Devijanca  $D$  definirana je u terminima maksimalne vrijednosti funkcije vjerodostojnosti  $l(\boldsymbol{\beta})$  i modela s maksimalnim brojem parametara:

$$D = 2[l(\boldsymbol{\beta}_{max}) - l(\boldsymbol{\beta})], \quad (3.32)$$

gdje je  $l(\boldsymbol{\beta}_{max})$  maksimalna vrijednost funkcije vjerodostojnosti za model s najvećim brojem parametara.

## Poglavlje 4

# Empirijska analiza

U ovom poglavlju provodi se empirijska analiza na podacima opisanim u poglavlju 3. Za početak se promatra korelacijska matrica nezavisnih varijabli kako bi se utvrdila međusobna ovisnost varijabli te kako bi sugerirala koje varijable se mogu naći zajedno u modelu. U ovom dijelu pokušat će se doći do modela i kovarijata koje najbolje opisuju model te će se pokušati potvrditi navedene hipoteze. Svi modeli kreirani su naredbom *polr* iz paketa *MASS* koja za rezultat daje parametar  $\beta$  procijenjen metodom maksimalne vjerodostojnosti, standardnu grešku te vrijednost t-statistike. Za početak će se promatrati jednostruka ordinalna logistička regresija, tj. model sa samo jednom kovarijatom te na temelju provedene analize pokušati utvrditi statistički značajne kovarijate. Analiza prediktivnosti provest će se ANOVA testom uspoređujući modele s "null" modelom. Nakon jednostruke regresije, provodi se višestruka regresija počevši od potpunog modela te se na temelju  $p$ -vrijednosti i korelacije među varijablama izbacuju one kovarijate koje nisu statistički značajne za model ili kovarijata koja je u visokoj korelaciji s nekom drugom kovarijatom. Nakon provedene analize dolazi se do najprikladnijeg modela te kovarijata koje imaju statistički značajan utjecaj na IOP. Za provjeru rezultata koriste se automatizirane procedure implementirane u R-u: metoda korak po korak te metoda najboljeg podskupa koje će dati slične rezultate kao i provedena analiza.

### Korelacijska matrica

Korelacijska matrica pokazuje u kojem su odnosu nezavisne varijable te u kojem se smjeru kreću njihove vrijednosti. Odnos je prikazan Spearmanovim koeficijentom korelacije koji se kreće između -1 te 1. Kada je koeficijent blizu 1 varijable su pozitivno korelirane, dok za koeficijent blizu -1 kažemo da su negativno korelirane. Povezanost znači da je vrijednost jedne varijable moguće, s određenom vjerojatnošću, predvidjeti na osnovi saznanja o vrijednosti druge varijable te to sugerira da se neke varijable mogu izbaciti iz modela.

Tablica 4.1 pokazuje matricu korelacije.

	stopa_nez	BDP	dug	br_stan	net	HDI	CPI
stopa_nez	1.000						
BDP	0.091	1.000					
dug	0.199	0.092	1.000				
br_stan	-0.178	0.035	0.096	1.000			
net	0.085	0.884	0.130	0.039	1.000		
HDI	0.092	0.939	0.112	0.047	0.900	1.000	
CPI	0.096	0.664	0.179	-0.085	0.646	0.706	1.000

Tablica 4.1: Spearmanovi koeficijenti korelacije nezavisnih varijabli

Iz tablice je vidljiva visoka povezanost varijable HDI s BDP-om po stanovniku (0.939) kao i postotkom korisnika interneta (0.900). Pozitivna korelacija između HDI-a i BDP-a po stanovniku je bila očekivana jer HDI kao jednu od dimenzija uzima životni standard koji mjeri BDP-om po stanovniku. Pozitivna koreliranost uočava se između varijabli postotak korisnika interneta te BDP-a po stanovniku (0.884). Najmanje korelacije uočavaju se između broja stanovnika te BDP-a po stanovniku (0.035) kao i između broja stanovnika te postotka korisnika interneta (0.039).

## 4.1 Jednostruka ordinalna logistička regresija

Kod jednostruke logističke regresije model sadrži jednu kovarijatu te se za svaki model i pojedinu kovarijatu procjenjuje parametar. Sljedeća jednakost prikazuje model:

$$\text{logit}(\mathbb{P}(Y_i \leq j)) = \beta_{0j} + \beta_1 * \text{nezavisna\_varijabla}.$$

Testiraju se hipoteze  $H_0 : \beta_1 = 0$  i  $H_1 : \beta_1 \neq 0$ . Nezavisna varijabla je statistički značajna ako je  $\beta_1 \neq 0$ . Za razinu značajnosti uzima se  $\alpha=5\%$ . U programskom jeziku R koristi se funkcija *polr* za procjenu parametara ordinalnog logističkog modela. *p*-vrijednost određuje se obzirom na *t*-statistiku koja je jednaka koeficijentu podijeljenom sa standardnom greškom. Ako je  $p \leq \alpha$  onda odbacujemo  $H_0$  u korist  $H_1$  na nivou značajnosti 5%, tj. kovarijata koja ima malu *p*-vrijednost je statistički značajna za model. U slučaju  $p > \alpha$  ne odbacujemo  $H_0$  u korist  $H_1$ . U R-u funkcija *polr* model ordinalne logističke regresije parametrizira kao:  $\text{logit}(\mathbb{P}(Y_i \leq j)) = \beta_{0j} - \eta_1 x_i - \dots - \eta_p x_i$ , gdje je  $\eta_i = -\beta_i$ . Tablica 4.2 daje prikaz rezultata jednostruke ordinalne logističke regresije koja se koristi za određivanje statističke značajnosti nezavisnih varijabli za svaku varijablu posebno.

Iz Tablice 4.2 vidimo da su kovarijate BDP po stanovniku, postotak korisnika interneta, HDI te CPI statistički značajne tj. *p*-vrijednosti su manje od 0.05. Za svaki model u Ta-

Varijabla	$\beta_i$	$e^{\beta_i}$	Standardna greška	t-vrijednost	p-vrijednost
stopa_nez	-1.083	0.339	3.196	-0.339	0.735
BDP	0.060	1.062	0.013	4.641	$3.473 * 10^{-6}$
dug	0.782	2.186	0.524	1.494	0.135
br_stan	$-6.726 * 10^{-5}$	1.000	0.001	-0.069	0.945
net	4.061	58.038	0.710	5.724	$1.041 * 10^{-8}$
HDI	8.474	4787.928	1.322	6.410	$1.457 * 10^{-10}$
CPI	0.073	1.076	0.012	6.137	$8.422 * 10^{-10}$

Tablica 4.2: Rezultati jednostruke ordinalne logističke regresije

Varijabla	$\beta_{01}$	$\beta_{02}$	$\beta_{03}$	$\beta_{04}$
stopa_nez	-1.656	-0.368	0.948	2.847
BDP	-1.184	0.187	1.786	4.177
dug	-1.177	0.119	1.466	3.393
br_stan	-1.583	-0.296	1.016	2.914
net	-0.023	1.524	3.246	5.402
HDI	3.713	5.348	7.191	9.434
CPI	0.953	2.494	4.215	6.690

Tablica 4.3: Koeficijenti  $\beta_{0j}$  za pojedinu varijablu uključenu u model

blici 4.3 prikazani su koeficijenti  $\beta_{0j}$   $j=1, \dots, 4$ . Koeficijenti  $\beta_{0j}$  dani su radi razumijevanja konstrukcije modela, no ne igraju bitnu ulogu u daljnjim analizama te interpretaciji rezultata modela pa se za ostale modele neće navoditi.

Kumulativni logistički model za pojedinu kategoriju uključujući samo jednu kovarijatu, npr. BDP po stanovniku, ima sljedeći oblik:

$$\text{logit}(\mathbb{P}(Y_i \leq 1)) = -1.184 - 0.060 * BDP,$$

$$\text{logit}(\mathbb{P}(Y_i \leq 2)) = 0.187 - 0.060 * BDP,$$

$$\text{logit}(\mathbb{P}(Y_i \leq 3)) = 1.786 - 0.060 * BDP \text{ te}$$

$$\text{logit}(\mathbb{P}(Y_i \leq 4)) = 4.177 - 0.060 * BDP.$$

Koeficijente iz modela je teže interpretirati jer su skalirani u obliku logaritma. Drugi te lakši način interpretacije je pretvaranje koeficijenata u omjere izgleda tako da se za svaki procijenjeni koeficijent  $\beta_i$  gleda njegov eksponent  $e^{\beta_i}$ . Interpretacija parametara je sljedeća, porast BDP-a za jednu jedinicu, točnije za 1000\$ povećava izgleda za prelazak iz niže u višu kategoriju množeći ih s 1.062. Analogna analiza bi se mogla provesti za ostale kovarijate. Uočava se pozitivna veza između IOP-a i sljedećih kovarijata: BDP po stanovniku, javnog duga, postotka korisnika interneta, HDI-a te CPI-a, dok se negativna povezanost



uočava samo kod kovarijate stopa nezaposlenosti. Varijabla broj stanovnika ima koeficijent približan nuli pa se može zaključiti da nema značajan utjecaj na IOP. Uz varijablu HDI dobije se i najveći pozitivan koeficijent te ona ima najveći pozitivan utjecaj na IOP. Pozitivnu vezu između javnog duga i transparentnosti pokazali su neki autori poput Styleasa i Tennysona [32] pa rezultat nije iznenađujuć.

Provodi se i analiza prediktivnosti ANOVA testom, na temelju kojeg ćemo moći izvesti zaključak o korisnosti uključivanja pojedine nezavisne varijable u model. ANOVA analiza provodi se usporedbom dva modela- "null" modela, tj. modela bez ijedne neovisne varijable i modela koji sadrži samo jednu neovisnu varijablu za koju se vrši analiza. Nulta hipoteza su jednaka očekivanja navedena dva modela. Za dobivenu  $p$ -vrijednost manju od 0.05, na razini značajnosti 5% možemo odbaciti nultu hipotezu o jednakim očekivanjima i pretpostaviti da model s neovisnom varijablom bolje opisuje model. Analogno kao prije gledat će se i AIC kriterij među modelima, manja vrijednost AIC kriterija pokazuje bolju prediktivnost modela. Veća razlika reziduala između dva modela, odnosno LR statistika je dodatni pokazatelj bolje prediktivnosti modela s neovisnom varijablom. Rezultati ANOVA analize prikazani su u Tablici 4.4 Usporedbom rezidualne devijance "null" modela te rezidualne devijance promatranog modela s jednom kovarijatom može se zaključiti je li prikladniji model koji ima samo slobodan član ili onaj koji u sebi ima i nezavisnu varijablu.

R.Br.	Model	Resid. df	Resid. Dev	Test	Df	LR stat	Pr(Chi)	AIC
1.	"null" model	113	349.989					357.989
2.	stopa_nez	112	349.874	1 vs 2	1	0.115	0.734	359.874
3.	BDP	112	325.014	1 vs 3	1	24.976	$5.806 * 10^{-7}$	335.014
4.	dug	112	347.817	1 vs 4	1	2.172	0.141	357.817
5.	br_stan	112	349.984	1 vs 5	1	0.005	0.944	359.984
6.	net	112	313.355	1 vs 6	1	36.635	$1.425 * 10^{-9}$	323.355
7.	HDI	112	302.708	1 vs 7	1	47.281	$6.150 * 10^{-12}$	312.708
8.	CPI	112	307.042	1 vs 8	1	42.947	$5.623 * 10^{-11}$	317.042

Tablica 4.4: Test značajnosti nezavisnih varijabli

Usporedbom "null" modela te modela koji sadrži samo varijablu stopa nezaposlenosti ANOVA analizom dobivena  $p$ -vrijednost je 0.734, razlika rezidualnih devijanci je 0.115 te na temelju rezultata ne odbacujemo nultu hipotezu o jednakim očekivanjima i pretpostavljamo da ova varijabla nije prediktivna, tj. ne razlikuje kategorije ovisne varijable i nema je smisla uključivati u model. Analogni zaključci donose se za varijable broj stanovnika te udio javnog duga u BDP-u. Kod ostalih varijabli  $p$ -vrijednost je na razini značajnosti 5% dovoljno mala te odbacujemo nultu hipotezu o jednakim očekivanjima "null" modela i modela s jednom varijablom koji promatramo te ih ima smisla uključiti u model, a to su

BDP po stanovniku, postotak korisnika interneta, HDI te CPI. AIC indeks je manji kod modela koji uključuju navedene varijable u usporedbi s "null" modelom.

## 4.2 Višestruka ordinalna logistička regresija

U ovom dijelu provodi se višestruka ordinalna logistička regresija na nekoliko modela. Kao početni model promatra se potpuni model, tj. model koji uključuje sve zavisne kovarijate. Druge modele određujemo na temelju korelacijske matrice te  $p$ -vrijednosti koja određuje statističku značajnost procijenjenih parametara. Promatrat će se i koeficijenti nezavisnih varijabli te njihov utjecaj na promjenu IOP-a uz uvjet da ostale varijable ostanu nepromijenjene.

### Model 1

Prvi model koji se promatra je potpuni model, tj. model koji se sastoji od svih nezavisnih varijabli promatranih u ovom radu. Kumulativni logistički model zbog pretpostavki modela ima sljedeći prikaz:

$$\text{logit}(\mathbb{P}(Y_i \leq j)) = \beta_{0j} + \beta_1 * \text{stopa\_nez} + \beta_2 * \text{BDP} + \beta_3 * \text{dug} + \beta_4 * \text{br\_stan} + \beta_5 * \text{net} + \beta_6 * \text{HDI} + \beta_7 * \text{CPI} \quad \text{za } j = 1, 2, 3, 4.$$

U Tablici 4.5 su prikazani rezultati za Model 1.

Varijabla	$\beta_k$	$e^{\beta_k}$	Standardna greška	t-vrijednost	p-vrijednost
stopa_nez	-4.688	0.009	3.553	-1.320	0.187
BDP	-0.039	0.962	0.022	-1.781	0.075
dug	-0.060	0.942	0.571	-0.105	0.917
br_stan	0.000	1.000	0.001	-0.341	0.733
net	0.134	1.147	1.521	0.090	0.980
HDI	6.746	850.848	2.910	2.319	0.020
CPI	0.062	1.056	0.020	3.078	0.002

Tablica 4.5: Rezultati ordinalne logističke regresije za Model 1

Koeficijenti regresije  $\beta_k$  ukazuju koliki utjecaj promjena pojedine nezavisne varijable ima na kretanje zavisne varijable. Ukoliko je koeficijent pozitivan, porast nezavisne varijable dovodi i do porasta zavisne varijable, analogno vrijedi i obratno. Koeficijent uz BDP je negativan i iznosi -0.039, njegov eksponent jednak je 0.962 pa se izgledi prelaska iz niže u višu kategoriju IOP-a povećanjem BDP-a za jednu jedinicu, tj. 1000\$ smanjuju za 3.8% (dobiveno kao  $(1-0.962)*100$ ). Analogna analiza se daje za preostale varijable. Kao i

u jednostruko analizi koeficijent uz varijablu broj stanovnika je približan nuli te možemo zaključiti da nema značajan utjecaj na IOP. Tablica 4.5 prikazuje pozitivnu vezu između varijabli postotak korisnika interneta, HDI i CPI. Dok se negativna veza uočava kod varijabli stopa nezaposlenosti, BDP po stanovniku te udio javnog duga u BDP-u. Najveći negativni utjecaj ima varijabla stopa nezaposlenosti dok najveći pozitivan ima varijabla HDI. Varijable koje statistički značajno utječu na IOP su indeksi CPI te HDI. Uz njih manju  $p$ -vrijednost imaju varijable stopa nezaposlenosti te BDP po stanovniku, ostale varijable imaju veliku  $p$ -vrijednost, posebno varijabla udio javnog duga u BDP-u. Za statistički značajne varijable CPI te HDI uočavamo pozitivne koeficijente  $\beta_k$ . Najveći negativni koeficijent ima varijabla stopa nezaposlenosti. Kriterij AIC za Model 1 iznosi 313.684 dok rezidualna devijanca iznosi 291.684.

## Model 2

Na temelju prethodnih analiza te korelacijske matrice prikazane u Tablici 4.1 odabiru su nezavisne varijable za Model 2. Korelacijska matrica prikazuje visoku koreliranost varijabli BDP i net (0.884) te varijabli HDI i net (0.900). Zbog visoke korelacije između navedenih varijabli te velike  $p$ -vrijednosti iz modela se izbacuje kovarijata net. Kumulativni logistički model ima sljedeći prikaz:

$$\text{logit}(\mathbb{P}(Y_i \leq j)) = \beta_{0j} + \beta_1 * \text{stopa\_nez} + \beta_2 * \text{BDP} + \beta_3 * \text{dug} + \beta_4 * \text{br\_stan} + \beta_5 * \text{HDI} + \beta_6 * \text{CPI}$$

za  $j = 1, 2, 3, 4$ .

Rezultati za navedeni model prikazani su u Tablici 4.6.

Varijabla	$\beta_k$	$e^{\beta_k}$	Standardna greška	t-vrijednost	p-vrijednost
stopa_nez	-4.640	0.010	3.516	-1.319	0.187
BDP	-0.039	0.962	0.022	-1.781	0.075
dug	-0.062	0.940	0.571	-0.108	0.914
br_stan	0.000	1.000	0.001	-0.35	0.726
HDI	6.948	1041.519	1.869	3.718	$2.007 * 10^{-4}$
CPI	0.062	1.064	0.020	3.081	$2.064 * 10^{-3}$

Tablica 4.6: Rezultati ordinalne logističke regresije za Model 2

Rezultati pokazuju da su statistički značajne varijable HDI te CPI te imaju pozitivan utjecaj na IOP. Najveći negativni koeficijent ima varijabla stopa\_nez dok najveću  $p$ -vrijednost ima varijabla dug.

Ukoliko je jedan model podskup drugog modela (u ovom slučaju Model 2 je podskup Model 1) može se provesti test omjera vjerodostojnosti te su rezultati prikazani u Tablici 4.7. Test omjera vjerodostojnosti koristi se prilikom usporedbe dva modela. Nulta hipoteza je

da je Model 2 bolji od potpunog Modela 1. Na razini značajnosti 5% ne odbacujemo nultu hipotezu, tj. možemo zaključiti da je Model 2 (bez varijable net) bolji od Modela 1 koji sadrži sve varijable.

Model	Resid. df	Resid. Dev	Test	Df	LR stat	Pr(Chi)	AIC
Model 2	107	291.692					311.692
Model 1	106	291.684	2 vs 1	1	0.008	0.928	313.684

Tablica 4.7: Test omjera vjerodostojnosti za Model 1 te Model 2

### Model 3

Na temelju prethodnih rezultata te Modela 2 odabiremo nezavisne varijable koje će uključivati Model 3. Iz Modela 2 izbacujemo kovarijatu koja ima najveću  $p$ -vrijednost, tj. koja nije statistički značajna za varijablu IOP, a to je dug. Model ima sljedeći prikaz:

$$\text{logit}(\mathbb{P}(Y_i \leq j)) = \beta_{0j} + \beta_1 * \text{stopa\_nez} + \beta_2 * \text{BDP} + \beta_3 * \text{br\_stan} + \beta_4 * \text{HDI} + \beta_5 * \text{CPI}$$

za  $j = 1, 2, 3, 4$ .

Dobiveni rezultati za Model 3 prikazani su u Tablici 4.8.

Varijabla	$\beta_k$	$e^{\beta_k}$	Standardna greška	t-vrijednost	p-vrijednost
stopa_nez	-4.675	0.009	3.499	-1.336	0.181
BDP	-0.039	0.962	0.022	-1.793	0.073
br_stan	0.000	1.000	0.001	-0.365	0.715
HDI	6.950	1042.762	1.869	3.718	$2.005 * 10^{-4}$
CPI	0.062	1.064	0.020	3.087	$2.005 * 10^{-3}$

Tablica 4.8: Rezultati ordinalne logističke regresije za Model 3

Dobiju se slični rezultati kao prije. Statistički su značajne kovarijate HDI te CPI, koje imaju pozitivne koeficijente regresije. Najveću  $p$ -vrijednost ima varijabla broj stanovnika, dok varijable BDP te stopa\_nez imaju negativan koeficijent regresije. Analogno kao prije provodimo test omjera vjerodostojnosti uspoređujući Model 3 s Modelom 2 te s Modelom 1. Nulta hipoteza je da je Model 3 bolji od Modela 2 tj. od Modela 1. Rezultati su prikazani u Tablici 4.9.

Iz dobivenih rezultata te velikih  $p$ -vrijednosti ne odbacujemo nultu hipotezu u oba testa, tj. možemo zaključiti da je Model 3 bolji od Modela 2 te Modela 1 na razini značajnosti 5%. Model 3 koji ne sadrži varijable net te dug bolji je od Modela 2 te Modela 1.

Model	Resid. df	Resid. Dev	Test	Df	LR stat	Pr(Chi)	AIC
Model 3	108	291.704					309.704
Model 2	107	291.692	3 vs 2	1	0.012	0.914	311.692
Model 1	106	291.684	3 vs 1	2	0.020	0.990	313.684

Tablica 4.9: Test omjera vjerodostojnosti za Model 3 i Model 2 te Model 3 i Model 1

## Model 4

Za odabir nezavisnih varijabli Modela 4 gledamo korelacije među varijablama preostalim u Modelu 3, a to su stopa\_nez, BDP, br\_stan, HDI i CPI. Visoka korelacija uočava se između varijabli BDP te HDI (0.939) pa iz modela možemo izbaciti varijablu BDP. Model 4 ima sljedeći prikaz:

$\text{logit}(\mathbb{P}(Y_i \leq j)) = \beta_{0j} + \beta_1 * \text{stopa\_nez} + \beta_2 * \text{br\_stan} + \beta_3 * \text{HDI} + \beta_4 * \text{CPI}$  za  $j = 1, 2, 3, 4$ .  
Dobiveni rezultati za Model 4 prikazani su u Tablici 4.10.

Varijabla	$\beta_k$	$e^{\beta_k}$	Standardna greška	t-vrijednost	p-vrijednost
st_nezap	-3.313	0.036	3.371	-0.983	0.326
br_stan	0.000	1.000	0.001	-0.312	0.754
HDI	5.702	299.505	1.709	3.337	$8.466 * 10^{-4}$
CPI	0.040	1.040	0.016	2.567	$1.026 * 10^{-2}$

Tablica 4.10: Rezultati ordinalne logističke regresije za Model 4

Varijable HDI te CPI su statistički značajne s pozitivnim koeficijentom regresije. Najveću  $p$ -vrijednost ima varijabla broj stanovnika kojoj je koeficijent regresije približno jednak nuli. Provodimo test omjera vjerodostojnosti uspoređujući Model 4 s Modelima 3, 2 i 1. Rezultati su prikazani u Tablici 4.11.

Model	Resid. df	Resid. Dev	Test	Df	LR stat	Pr(Chi)	AIC
Model 4	109	294.923					310.923
Model 3	108	291.704	4 vs 3	1	3.224	0.073	309.704
Model 2	107	291.692	4 vs 2	2	3.236	0.198	311.692
Model 1	106	291.684	4 vs 1	3	3.244	0.355	313.684

Tablica 4.11: Test omjera vjerodostojnosti Modela 4 uspoređujući ga s Modelima 3, 2 i 1

Na temelju danih rezultata te velikih  $p$ -vrijednosti, ne odbacujemo nultu hipotezu, tj. Model 4 smatramo boljim od Modela 3, pa tako i boljim od Modela 2 i 1. Model 4 ima manji AIC indeks od Modela 2 i 1.

## Model 5

Model 4 se pokazao prihvatljivijim od preostalih analiziranih modela te na temelju njega biramo varijable za Model 5. Zbog velike  $p$ -vrijednosti varijable broj stanovnika te koeficijenta regresije blizu nula, možemo zaključiti da na razini značajnosti od 5% varijabla broj stanovnika nije statistički značajna za varijablu IOP te ju možemo zanemariti u modelu. Model 5 se sastoji od sljedećih varijabli: stopa\_nez, HDI i CPI te ima sljedeći prikaz:

$$\text{logit}(\mathbb{P}(Y_i \leq j)) = \beta_{0j} + \beta_1 * \text{stopa\_nez} + \beta_2 * \text{HDI} + \beta_3 * \text{CPI} \quad \text{za } j = 1, 2, 3, 4.$$

Dobiveni rezultati za Model 5 prikazani su u Tablici 4.12.

Varijabla	$\beta_k$	$e^{\beta_k}$	Standardna greška	t-vrijednost	p-vrijednost
stopa_nez	-3.199	0.041	3.353	-0.954	0.340
HDI	5.698	298.321	1.708	3.336	$8.496 * 10^{-4}$
CPI	0.040	1.040	0.015	2.560	$1.046 * 10^{-2}$

Tablica 4.12: Rezultati ordinalne logističke regresije za Model 5

Statistički značajne varijable su CPI te HDI koje kao do sad imaju pozitivne koeficijente regresije. Varijabla stopa\_nez ima negativan koeficijent regresije te  $p$ -vrijednost 0.340 pa za nju možemo zaključiti da nije statistički značajna za varijablu IOP. Model 5 uspoređujemo s prijašnje definiranim modelima u radu te je prikaz rezultata dan u Tablici 4.13.

Model	Resid. df	Resid. Dev	Test	Df	LR stat	Pr(Chi)	AIC
Model 5	110	295.028					309.028
Model 4	109	294.923	5 vs 4	1	0.100	0.752	310.923
Model 3	108	291.704	5 vs 3	2	3.324	0.190	309.704
Model 2	107	291.692	5 vs 2	3	3.336	0.343	311.692
Model 1	106	291.684	5 vs 1	4	3.344	0.502	313.684

Tablica 4.13: Test omjera vjerodostojnosti Modela 5 uspoređujući ga s Modelima 4, 3, 2 i 1

Nulte hipoteze u testovima su da je Model 5 bolji od Modela 4, 3, 2 te 1. U svim testovima dobiju se velike  $p$ -vrijednosti te ne odbacujemo nultu hipotezu. Zaključak je da je Model 5 bolji od preostalih modela promatranih u radu te ujedno ima i najmanji AIC indeks.

## Model 6

Analizom rezultata za Model 5 varijabla stopa\_nez je imala veliku  $p$ -vrijednost te možemo reći da nije statistički značajna za varijablu IOP. U Modelu 6 promatramo samo varijable

HDI te CPI koje su se pokazale kao statistički značajne varijable. Model 6 ima sljedeći prikaz:

$$\text{logit}(\mathbb{P}(Y_i \leq j)) = \beta_{0j} + \beta_1 * HDI + \beta_2 * CPI \quad \text{za } j = 1, 2, 3, 4.$$

Dobiveni rezultati za Model 6 prikazani su u Tablici 4.14.

Varijabla	$\beta_k$	$e^{\beta_k}$	Standardna greška	t-vrijednost	p-vrijednost
HDI	5.582	265.630	1.701	3.281	$1.034 * 10^{-3}$
CPI	0.040	1.041	0.015	2.568	$1.023 * 10^{-2}$

Tablica 4.14: Rezultati ordinalne logističke regresije za Model 6

$p$ -vrijednosti obje nezavisne varijable su male te su obje statistički značajne. Model 6 uspoređujemo s prethodnim modelima. U Tablici 4.15 prikazani rezultati.

Model	Resid. df	Resid. Dev	Test	Df	LR stat	Pr(Chi)	AIC
Model 6	111	295.943					307.943
Model 5	110	295.028	6 vs 5	1	0.916	0.339	309.028
Model 4	109	294.923	6 vs 4	2	1.015	0.602	310.923
Model 3	108	291.704	6 vs 3	3	4.240	0.237	309.704
Model 2	107	291.692	6 vs 2	4	4.251	0.373	311.692
Model 1	106	291.684	6 vs 1	5	4.259	0.513	313.684

Tablica 4.15: Test omjera vjerodostojnosti Modela 6 uspoređujući ga s Modelima 5, 4, 3, 2 i 1

$p$ -vrijednosti testa omjera vjerodostojnosti su velike, te za svaki test ne odbacujemo nultu hipotezu što znači da je Model 6 najprihvatljiviji. Dakle, model koji sadrži dvije od sedam početnih nezavisnih varijabli je najprihvatljiviji. To su HDI te CPI. U jednostrukoj ordinalnoj logističkoj regresiji gdje su se promatrali modeli s po jednom nezavisnom varijablom, varijable BDP po stanovniku, postotak korisnika interneta, HDI te CPI su bile statistički značajne. Zbog visokih korelacija između varijabli BDP-a po stanovniku te HDI-a koji iznosi 0.939, postotka korisnika interneta i HDI-a koji iznosi 0.900 te BDP-a po stanovniku i postotka korisnika interneta koji iznosi 0.884 iz modela su uklonjene varijable postotak korisnika interneta te BDP po stanovniku.

## Model 7

Model 7 se koristi kao provjera mogu li se postići bolji rezultati od Modela 6 izbacivanjem još jedne od preostalih nezavisnih varijabli. U ovom primjeru izbacit ćemo varijablu CPI. Slični rezultati dobiju se i izbacivanjem varijable HDI.

Model 7 ima sljedeći prikaz:

$$\text{logit}(\mathbb{P}(Y_i \leq j)) = \beta_{0j} + \beta_1 * HDI \quad \text{za } j = 1, 2, 3, 4.$$

Dobiveni rezultati za Model 7 prikazani su u Tablici 4.16.

Varijabla	$\beta_k$	$e^{\beta_k}$	Standardna greška	t-vrijednost	p-vrijednost
HDI	8.474	4787.928	1.322	6.410	$1.457 * 10^{-10}$

Tablica 4.16: Rezultati ordinalne logističke regresije za Model 7

Varijabla HDI je statistički značajna za varijablu IOP s pozitivnim koeficijentom.

Model 7 usporedit ćemo s preostalim modelima preko testa omjera vjerodostojnosti te su rezultati prikazani u Tablici 4.17.

Model	Resid. df	Resid. Dev	Test	Df	LR stat	Pr(Chi)	AIC
Model 7	112	302.708					312.708
Model 6	111	295.943	7 vs 6	1	6.765	0.009	307.943
Model 5	110	295.028	7 vs 5	2	7.680	0.021	309.028
Model 4	109	294.923	7 vs 4	3	7.780	0.051	310.923
Model 3	108	291.704	7 vs 3	4	11.005	0.030	309.704
Model 2	107	291.692	7 vs 2	5	11.016	0.051	311.692
Model 1	106	291.684	7 vs 1	6	11.024	0.090	313.684

Tablica 4.17: Test omjera vjerodostojnosti Modela 7 uspoređujući ga s Modelima 6, 5, 4, 3, 2 i 1

Uspoređujući Model 7 s Modelom 6, na razini značajnosti 5% odbacujemo nultu hipotezu da je Model 7 bolji. Dakle, izbacivanjem još jedne od varijabli ne postižu se bolji rezultati. Na temelju provedenih analiza i prikazanih rezultata možemo zaključiti da je Model 6 najprihvatljiviji. U usporedbi s preostalim modelima, Model 7 je bolji od Modela 1, 2 te 4,  $p$ -vrijednosti su veće od 0.05, pa ne odbacujemo nultu hipotezu da je Model 7 bolji.

### 4.3 Automatizirane procedure u R-u

Programski jezik R sadrži automatizirane funkcije koje odabiru najbolji model prema nekom kriteriju. Najčešći su kriterij već navedeni AIC, BIC te drugi. Metode koje se koriste u odabiru varijabli su metode unaprijed, unatrag te korak po korak kratko objašnjene u 3.5. U ovom dijelu provest će se metoda korak po korak te metoda najboljeg podskupa koje će služiti kao provjera prethodno dobivenih rezultata.



## Metoda korak po korak

Metoda korak po korak naizmjenično isključuje najmanje značajnu varijablu te uključuje najznačajniju varijablu koja u tom koraku nije u modelu. U svakom koraku računa se kriterij AIC za dani model. Kod problema korelacije varijabli metoda korak po korak eliminira one varijable koje se preklapaju s drugima i zbog toga imaju zanemariv utjecaj u predviđanju modela. Funkcija koja se koristi u R-u za danu metodu je *stepAIC*. Koraci funkcije *stepAIC* za potpuni model tj. Model 1 s uključenim svim nezavisnim varijablama prikazani su na Slici 4.1.

```

Start: AIC=313.68
factor(Kategorija) ~ Stopa_nez + BDP + Dug + Br_st + Net + CPI +
  HDI

      Df  AIC
- Net    1 313.69
- Dug    1 313.69
- Br_st  1 313.80
- Stopa_nez 1 313.43
<none>   313.68
- BDP    1 314.86
- HDI    1 317.31
- CPI    1 321.61

Step: AIC=311.69
factor(Kategorija) ~ Stopa_nez + BDP + Dug +Br_st
  + CPI + HDI

      Df  AIC
- Dug    1 309.70
- Br_st  1 309.82
- Stopa_nez 1 311.45
<none>   314.16
- BDP    1 312.86
+ Net    1 316.16
- CPI    1 319.63
- HDI    1 324.21

Step: AIC=309.7
factor(Kategorija) ~ Stopa_nez + BDP + Br_st + CPI + HDI

      Df  AIC
- Br_st  1 307.56
- Stopa_nez 1 309.51
<none>   309.70
- BDP    1 310.93
+ Dug    1 311.69
+ Net    1 311.69
- CPI    1 317.66
- HDI    1 322.22

Step: AIC=307.84
factor(Kategorija) ~ Stopa_nez + BDP + CPI + HDI

      Df  AIC
- Stopa_nez 1 307.56
<none>   | 307.84
- BDP    1 309.03
+ Br_st  1 309.70
+ Dug    1 309.82
+ Net    1 309.82
- CPI    1 315.73
- HDI    1 320.33

Step: AIC=307.56
factor(Kategorija) ~ BDP + CPI + HDI

      Df  AIC
<none>   307.56
+ Stopa_nez 1 307.56
- BDP    1 307.94
+ Dug    1 309.49
+ Br_st  1 309.51
+ Net    1 309.55
- CPI    1 314.58
- HDI    1 318.95

```

Slika 4.1: Koraci funkcije *stepAIC*

Slika 4.1 prikazuje kojim je redom algoritam izbacivao varijable te vrijednosti kriterija AIC u svakom koraku. Algoritam je redom izbacivao sljedeće varijable: postotak korisnika interneta, udio javnog duga u BDP-u, broj stanovnika te stopu nezaposlenosti. Model

dobiven metodom korak po korak uključuje tri varijable: BDP po stanovniku, HDI te CPI. Razlika je u varijabli BDP po stanovniku koju Model 6 ne uključuje zbog visokih korelacija među varijablama. Za provjeru su se usporedili Model 6 koji sadrži samo kovarijate HDI te CPI te model dobiven metodom korak po korak koji uključuje dodatnu kovarijatu BDP po stanovniku. Napravljen je analogno kao i prije test omjera vjerodostojnosti gdje je nulta hipoteza da je Model 6 bolji od modela dobivenog metodom korak po korak. Dobivena  $p$ -vrijednost iznosi 0.122 te na razini značajnosti od 5% ne odbacujemo nultu hipotezu, tj. možemo zaključiti da je Model 6 bolji pa se u ovom radu koristi kao najprihvatljiviji iako ima nešto veći kriterij AIC od modela dobivenog metodom korak po korak.

## Metoda najboljeg podskupa

Još jedna od funkcija u R-u kao metoda odabira modela je funkcija *regsubsets*. Početni model je potpuni model te su rezultati metode prikazani na Slici 4.2. Varijable iznad kojih je najviše zvijezdica su najbolje po metodi najboljeg podskupa, a to su varijable HDI i CPI te su tako još jednom potvrđeni rezultati od prije.

```
Subset selection object
Call: regsubsets.formula(Kategorija ~ Stopa_nez + BDP + Dug + Br_st +
  Net + CPI + HDI, data = podaci)
7 Variables (and intercept)
      Forced in Forced out
Stopa_nez  FALSE      FALSE
BDP        FALSE      FALSE
Dug        FALSE      FALSE
Br_st      FALSE      FALSE
Net        FALSE      FALSE
CPI        FALSE      FALSE
HDI        FALSE      FALSE
1 subsets of each size up to 7
Selection Algorithm: exhaustive
      Stopa_nez BDP Dug Br_st Net CPI HDI
1 ( 1 ) " " " " " " " " " " " " " "
2 ( 1 ) " " " " " " " " " " " " " "
3 ( 1 ) " " " * " " " " " " " " " "
4 ( 1 ) " * " " " " " " " " " " " "
5 ( 1 ) " * " " " " * " " " " " " "
6 ( 1 ) " * " " * " * " " " " " " "
7 ( 1 ) " * " " * " * " * " " " " " "
```

Slika 4.2: Rezultati funkcije *regsubsets*

## Poglavlje 5

### Zaključak

Koristeći ordinalnu logističku regresiju u ovom radu pokušalo se odrediti koje socio-ekonomske odrednice imaju statistički značajan utjecaj na indeks otvorenosti proračuna. Sve više radova bavi se ovom tematikom ne bi li se odredile varijable koje mogu poboljšati transparentnost u pojedinoj zemlji. IOP je jedini međunarodno usporedivi pokazatelj transparentnosti državnog proračuna. Prvi put je predstavljen i izračunat 2006. te poprima vrijednosti od 0 do 100 i zemlje svrstava u 5 kategorija. U ovom radu promatrao se indeks za 2019. te 117 zemalja uključenih u ovo istraživanje.

Promatrao se utjecaj sljedećih varijabli: stope nezaposlenosti, BDP-a po stanovniku, udjela javnog duga u BDP-u, broja stanovnika, postotka korisnika interneta, indeksa ljudskog razvoja te indeksa percepcije korupcije.

Hipoteze predstavljene u radu su rezultat prethodnih radova i istraživanja. Pretpostavke su sljedeće: stopa nezaposlenosti i udio javnog duga u BDP-u imaju negativan i statistički značajan utjecaj na indeks otvorenosti proračuna dok preostale varijable imaju pozitivan i statistički značajan utjecaj.

Nakon empirijske analize, rezultati su pokazali da su statistički značajne četiri od sedam navedenih varijabli, a to su: BDP po stanovniku, postotak korisnika interneta, indeks ljudskog razvoja te indeks percepcije korupcije. Sve četiri varijable pokazuju pozitivan utjecaj na indeks otvorenosti proračuna. Nakon promatranja korelacijske matrice varijabli te analize rezultata za nekoliko modela, uspoređujući ih po različitim kriterijima, najprikladnijim se pokazao model koji sadrži dvije varijable: indeks ljudskog razvoja te indeks percepcije korupcije.

Može se zaključiti da visoka korumpiranost značajno smanjuje transparentnost u pojedinoj zemlji kao i da zemlje s obrazovanim stanovništvom te višim životnim standardom imaju veću transparentnost. Obrazovanje stanovništvo se više aktivira te zahtijeva transpa-

rentnost vladajućih. Države s većim postotkom korisnika interneta također pokazuju veću transparentnost, a kao razlog može se navesti da su građani bolje, više i lakše informirani.

U svrhu ovog i sličnih radova mogla se provesti i panel analiza na podacima dostupnim od 2006. te su se mogle promatrati i druge odrednice koje bi mogle imati statistički značajan utjecaj na indeks otvorenosti proračuna. Više podataka dalo bi više informacija iako je 2006. u istraživanju bilo uključeno tek 59 zemalja pa se u ovom radu ipak odlučilo za ordinalnu logističku regresiju.

Zbog trenutne krize izazvane pandemijom koronavirusa, vlade su primorane donositi brze te velike odluke koje utječu na proračun te cjelokupnu ekonomiju zemlje. Kako vlada ne bi izgubila povjerenje građana transparentnost ima bitnu ulogu i značaj. Vlade trebaju povećati transparentnost koja predstavlja dobro upravljanje te povećava odgovornost vlasti. Posebno je potrebno raditi na sprječavanju korupcije, koja se može izrazito povećati u ovakvim krizama. Povećanjem transparentnosti moguće je bar malo ublažiti posljedice ove krize.

# Bibliografija

- [1] A. Agresti, *Categorical Data Analysis*, University of Florida, Gainesville, Florida, 2002.
- [2] D. Albalade, *The Institutional, Economic and Social Determinants of Local Government Transparency*, Journal of Economic Policy Reform, 2012., dostupno na [http://diposit.ub.edu/dspace/bitstream/2445/58408/1/IR12-010\\_Albalade.pdf](http://diposit.ub.edu/dspace/bitstream/2445/58408/1/IR12-010_Albalade.pdf) (srpanj 2020.)
- [3] J. Alt, D. Lassen, S. Rose, *The Causes of Fiscal Transparency: Evidence from the American States*, Sixth Jacques Polak Annual Research Conference, 2005. dostupno na <https://www.imf.org/external/np/res/seminars/2005/arc/pdf/alt.pdf> (srpanj 2020.).
- [4] J. Alt, D. Lassen, *Fiscal transparency, political parties, and debt in OECD countries*, Economic Policy Research Unit Working Paper No. 03-02, 2003. dostupno na <http://rwj.harvard.edu/papers/alt.pdf> (srpanj, 2020.).
- [5] N. Andreula A. Chong, J. Guillén, *Institutional Quality and Fiscal Transparency*, IDB Working Paper Series No.IDB-WP-125, 2015.
- [6] M. Bađun, *Transparentnost proračuna*, Institut za javne financije, dostupno na <http://www.ijf.hr/FTP/2009/4/badjun.pdf> (srpanj, 2020.).
- [7] B. Basrak, H. Planinić *Generalizirani linearni modeli*, dostupno na [https://web.math.pmf.unizg.hr/nastava/finprakt/Materijali1920/GLM/GLM\\_Notes2.pdf](https://web.math.pmf.unizg.hr/nastava/finprakt/Materijali1920/GLM/GLM_Notes2.pdf) (srpanj, 2020.).
- [8] A. G. Barnett, A. J. Dobson, *An introduction to generalized linear models*, CRC Press, Boca Raton, 2008.
- [9] L. Birskyte *Determinants of Budget Transparency in Lithuanian Municipalities*, Public Performance and Management Review, 2018.

- [10] M. Bronić *Kodeks fiskalne transparentnosti*, dostupno na <http://www.ijf.hr/newsletter/53.pdf> (srpanj, 2020.).
- [11] M. Bronić, J. Franić *Otvorenost proračuna središnje države*, dostupno na <http://www.ijf.hr/upload/files/1141.pdf> (srpanj, 2020.).
- [12] B. Christensen *Cumulative Link Models for Ordinal Regression with the R Package ordinal*, Journal of Statistical Software, dostupno na [https://cran.r-project.org/web/packages/ordinal/vignettes/clm\\_article.pdf](https://cran.r-project.org/web/packages/ordinal/vignettes/clm_article.pdf) (srpanj, 2020.).
- [13] S. A. Czepiel *Maximum Likelihood Estimation of Logistic Regression Models: Theory and Implementation*, dostupno na <https://czep.net/stat/mlelr.pdf> (srpanj, 2020.).
- [14] H. Darbishire *Proactive Transparency: The future of the right to information*, World Bank Institute Working Paper, No. 60., 2010. dostupno na <https://doi.org/10.1596/25031> (srpanj, 2020.).
- [15] D. A. del Sol, *The institutional, economic and social determinants of local government transparency*, Journal of Economic Policy Reform, dostupno na <https://www.tandfonline.com/doi/abs/10.1080/17487870.2012.759422> (srpanj, 2020.).
- [16] D. Guillamón, F. Bastida, B. Benito *The Determinants of Local Government's Financial Transparency*, Local Government Studies, 2011.
- [17] United Nations Development Programme *Human Development Index*, dostupno na <http://hdr.undp.org/en/content/human-development-index-hdi> (kolovoz, 2020.).
- [18] International Monetary Fund *Fiscal transparency code*, 2014., dostupno na <https://blog-pfm.imf.org/files/ft-code.pdf> (srpanj, 2020.).
- [19] International Monetary Fund *Fiscal transparency code*, 2019., dostupno na <https://www.imf.org/external/np/fad/trans/Code2019.pdf> (srpanj, 2020.).
- [20] International Budget Partnership *Open Budget Survey*, dostupno na <https://www.internationalbudget.org/open-budget-survey/open-budget-survey-2019> (srpanj, 2020.).
- [21] International Labour Organization *Unemployment rate*, dostupno na <https://ilostat.ilo.org/data/> (kolovoz, 2020.).

- [22] International Telecommunication Union *Individuals using the Internet (% of population)*, dostupno na <https://www.itu.int/en/ITU-D/Statistics/Pages/default.aspx> (kolovoz, 2020.).
- [23] G. Kopits i J. Craig *Transparency in Government Operations*, International Monetary Fund, Washington DC, 1998. dostupno na <https://www.imf.org/external/pubs/ft/op/158/op158.pdf> (srpanj, 2020.).
- [24] G. Lowatcharin, C. E. Menifield, *Determinants of Internet-enabled Transparency at the Local Level: A Study of Midwestern County Web Sites*, State and Local Government Review, 2015.
- [25] I. Malatestinić *Lokalni proračun i uključivanje javnosti u proces njegovog donošenja*, Udruga gradova u Republici Hrvatskoj, 2011. dostupno na [https://www.udruga-gradova.hr/wordpress/wp-content/uploads/2013/09/Lokalni-proracun\\_gotovo.pdf](https://www.udruga-gradova.hr/wordpress/wp-content/uploads/2013/09/Lokalni-proracun_gotovo.pdf) (srpanj, 2020.).
- [26] OECD, 2002. *OECD Best Practices for Budget Transparency*, Paris, dostupno na <https://www.oecd.org/gov/budgeting/Best%20Practices%20Budget%20Transparency%20-%20complete%20with%20cover%20page.pdf> (srpanj, 2020.).
- [27] C. Pérez R. Bolívar, L. Hernández, *e-Government process and incentives for online public financial information*, Emerald Insight, 2008.
- [28] S. J. Piotrowski i G. Van Ryzin, *Citizen Attitudes Toward Transparency in Local Government*, The American Review of Public Administration, 2007. dostupno na <https://journals.sagepub.com/doi/abs/10.1177/0275074006296777> (srpanj, 2020.).
- [29] J. M. Poterba, J. von Hagen *Fiscal Institutions and Fiscal Performance*, University of Chicago, 1999.
- [30] B. Stanić, *Determinants of subnational budget/fiscal transparency: a review of empirical evidence*, Public Sector Economics, 2018. dostupno na <http://www.pse-journal.hr/upload/files/pse/2018/4/4.pdf> (srpanj, 2020.).
- [31] B. Stanić, *Globalna inicijativa za fiskalnu transparentnost*, dostupno na <https://www.ijf.hr/upload/files/110.pdf> (srpanj, 2020.).
- [32] A. K. Styles, M. Tennyson, *The Accessibility of Financial Reporting of U.S. Municipalities on the Internet*

- [33] A. F. Tavares, N. F. da Cruz, *Explaining the transparency of local government websites through a political market framework*, LSE Research Online, 2017.
- [34] The World Bank, *GDP per capita and Individuals using the Internet (% of population)* dostupno na <https://data.worldbank.org/indicator> (kolovoz, 2020.).
- [35] Trading Economics, *Debt to GDP and Population* dostupno na <https://tradingeconomics.com/country-list> (kolovoz, 2020.).
- [36] Transparency International, *Corruption Perceptions Index*, dostupno na <https://www.transparency.org/en/cpi> (kolovoz, 2020.).
- [37] J. Wehner, P. Renzio, J. Braams, D. Carlisle i Ch. Rowly, *Citizens, Legislators, and Executive Disclosure: The Political Determinants of Fiscal Transparency*, Elsevier, 2013.
- [38] J. Wehner, P. Renzio *Citizens, legislators, and executive disclosure: The political determinants of fiscal transparency*, International Budget Partnership, 2011.



# Sažetak

Transparentnost proračuna može povećati odgovornosti vlasti, uključuje građane u donošenje odluka te vraća povjerenje u vlast. Cilj ovog rada bio je utvrditi koje socio-ekonomske varijable mogu utjecati na indeks otvorenosti proračuna te na koji način se transparentnost u pojedinoj zemlji može poboljšati. Promatraju se rezultati za 117 zemalja za koje su bili dostupni zadnji podaci o transparentnosti proračuna. Varijable koje su se promatrale su: stopa nezaposlenosti te udio javnog duga u BDP-u s pretpostavkom negativnog utjecaja na transparentnost, dok se preostale varijable: BDP po stanovniku, broj stanovnika, postotak korisnika interneta, indeks ljudskog razvoja te indeks percepcije korupcije promatraju pod pretpostavkom pozitivnog utjecaja. Četiri varijable su pokazale statistički značajan i pozitivan utjecaj, a to su: postotak korisnika interneta, BDP po stanovniku, indeks ljudskog razvoja te indeks percepcije korupcije. Koristeći ordinalnu logističku regresiju dobilo se nekoliko modela, od kojih je najprihvatljiviji sadržavao dvije varijable, a to su indeks ljudskog razvoja te indeks percepcije korupcije.

# Summary

Budget transparency might increase government accountability, involve citizens in decision-making, and restore citizens' trust in government. The aim of this paper was to determine which socio-economic variables might affect the OBI (Open Budget Index) and how transparency in a particular country can be improved. Results are observed for 117 countries for which the latest budget transparency data were available. The following variables were observed: unemployment rate and the public debt to GDP ratio with the assumption of a negative impact on transparency, while the remaining variables: GDP per capita, population, percentage of Internet users, human development index and corruption perceptions index are observed under assuming a positive impact. Four variables showed a statistically significant and positive impact: the percentage of Internet users, GDP per capita, the human development index and the corruption perception index. Using ordinal logistic regression, several models were obtained, the most acceptable containing two variables, the human development index and the corruption perception index.

# Životopis

Marija Andrijević je rođena 9.12.1994. u Požegi. Nakon završene osnovne škole "Antun Kanižlić" pohađa Opću gimnaziju u Požegi. Završetkom srednjoškolskog obrazovanja upisuje preddiplomski studij Matematike na Prirodoslovno matematičkom fakultetu u Zagrebu. Po završetku preddiplomskog studija, na istom fakultetu upisuje diplomski studij, smjer Financijska i poslovna matematika.