Univerzitet u Novom Sadu
Prirodno matematički fakultet
Departman za matematiku i informatiku

Aleksandra Milojević

PRIMENA PANEL REGRESIONE ANALIZE NA DEMOGRAFSKE POKAZATELJE

-MASTER RAD-

Mentor: prof. dr Zagorka Lozanov – Crvenković

Novi Sad, 2020.

SADRŽAJ

UVOD

Demografija se vrlo intezivno razvija poslednjih decenija, i postaje sve značajnija kada se radi o drugim oblastima istraživanja, u globalnim okvirima, posebno kada je reč o promenama i projekcijama naseljenosti planete. Demografski pokazatelji poput broja novorođenih, prirodnog priraštaja, broja umrlih i slično imaju veliki uticaj na ekonomski, tehnološki, naučni i svaki drugi vid razvoja zemlje. Sa druge strane, demografsko projektovanje ima kapitalnu ulogu u planiranju i kreiranju politika u različitim oblastima jedne zemlje poput prosvete, zdravstva, privrede, ali i osiguranja i socijalne zaštite. Kako bi se kvantifikovale određene projekcije neophodno je razviti model čiji će rezultat biti određeno predviđanje. U tom kontekstu, u zavisnosti od datih podataka, u empirijskim analizama, podaci koji se koriste mogu se podeliti u tri grupe i to prema načinu dobijanja i svojim osobinama. To su:

1. **uporedni podaci** (enlg*. Cross selection data*),
2. **podaci vremenskih serija** (engl. *Time series data*) i
3. **panel podaci** (enlg. *Panel data*).

Panel podaci kombiniju podatke vremenskih serija i uporedne podatke. Drugim rečima, panel podaci predstavljaju specijalni tip kombinovanih podataka kod kojih je ista jedinica preseka (na primer, drzava, kompanija, itd.) posmatrana tokom vremena (Gujarati, 2004). Panel podaci mogu biti balansirani i nebalansirani. U slučaju balansiranih panel podataka svaka uporedna jedinica posmatranja ima isti broj opservacija vremenskih serija, odnosno vremenske serije su iste dužine. U ovakvim okolnostima ne postoje problemi uslovljeni nedostatkom podataka. Ukoliko se broj opservacija razlikuje od jednog do drugog subjekta u panelu, onda se radi o nebalansiranom panelu

U disciplinama ekonometrije i statistike, panel podaci se odnose na višedimenzionalne podatke koji uopšteno uključuju merenja tokom određenog vremenskog perioda. Kod panel podataka imamo istu jedinicu preseka, posmatranu tokom vremena. Dakle, imamo izraženu kako prostornu, tako i vremensku komponentu. Kao takvi, panel podaci se sastoje od zapažanja istraživača o brojnim pojavama koje su prikupljene tokom nekoliko vremenskih perioda za istu grupu jedinica ili entiteta. U ovom radu dat je detaljan pregled panel regresionih modela kao i njihova primena u procesu predviđanja promena demografske strukture stanovništva.

Prvo poglavlje rada daje detaljan pregled prednost i ograničenja korišćenja panel podataka u regresionoj analizi

Drugo poglavlje rada sadrži detaljan pregled modela vezanih za panel podatke. U okviru predstavljenih modela dati su OLS model, model fiksnih efekata i model slučajnih efekata. U sekciji modela fiksnih efekata detaljno su razrađene varijacije ovog modela (svi parametri konstantni, promena parametara kroz jedinice posmatranja, promena parametara kroz vreme i promena parametara i kroz jedinice posmatranja i kroz vreme). Dodatno, kroz poglavlje se prožima i ilustrativan primer kako bi se što bolje pojasnili modeli koji su usko vezani za panel podatke.

Treće poglavlje rada uključuje sve potencijalne probleme koji se mogu pojaviti u analizi panel regresije. U okviru navedenih problema detaljno su prikazani problemi heteroskedastičnostii, autokorelacije i multikolinearnosti. Za svaki od navedenih potencijalnih problema detaljno su navedeni testovi kojima se može potvrditi ili odbaciti postojanost istih.

Četvrto poglavlje rada donosi smernice kako bi se odabrao najadekvatniji model. U njemu je prikazan Hausmanov test koji daje odgovor koji je model pogodniji, da li model fiksnih efekata ili model slučajnih (stohastičkih) efekata.

Na samom kraju rada, u ***petom poglavlju*** data je empirijska analiza rađena na stvarnim vrednostima koje su preuzete sa zvaničnog sajta Eurostat (evropske kuće za statistiku). Na samom početku ovog poglavlja date su osnovne informacije o skupu podataka koji će biti korišćen u izradi modela. Nakon toga, centralni deo poglavlja čini konstrukcija modela kako bi se odabrao model sa najboljim performansama, dok sami kraj poglavlja sadrži pregled odabira konkretnog modela.

Pre nego što odpočnu redovi ovog naučnog rada želela bih da se zahvalim svojoj mentorki prof. dr Zagorki Lozanov Crvenković koja mi je u svakom trenutku bila na raspoloaganju, pružala pomoć, savete i razumevanje. Posebno joj se zahvaljujem na uloženom trudu i vremenu, kao i na predloženoj temi koja je bila veoma interesantna, inovativna i na znanjima koja mi je prenela tokom studiranja.

Posebno se zahvaljujem svojim roditeljima i sestri koji su mi pružali beskrajnu podršku i pomoć tokom studiranja.

Novi Sad, 2020. godine *Aleksandra Milojević*

1. PREDNOSTI I OGRANIČENJA KORIŠĆENJA PANEL PODATAKA

Kao što je već rečeno, u empirijskim analizama, pored uporednih podataka i vremenskih serija koriste se i panel podaci. U ovom delu rada opisane su prednosti i ograničenja korišćenja panel podataka. Korišćenje panel serija daje mogućnost vršenja različitih analiza koje se ne bi mogle sprovesti pri samostalnom korišćenju vremenskih serija ili uporednih podataka, jer se panel serijama u ograničenom broju opservacija uzima maksimalna količina informacija. Profesori Cheng Hsiao[1] i Klevmarken[2] su dali pojedine prednosi, ali i ograničenja upotrebe panel podataka.

- **Kontrola individualne heterogenosti** – panel podaci pretpostavljaju da su posmatranja poput države, firme i slično heterogene jedinice. Vremenske serije i uporedni podaci ne mogu kontrolisati heterogenost, a ta nemogućnost kontrole može dovesti do ukidanja nepristrasnosti, odnosno može dovesti do pristrasnosti rezultata istraživanja. Demonstracija ove karakteristike ogleda se kroz sledeći primer: posmatra se potražnja za cigaretama u Sjedinjenim Američkim Državama u periodu od 1963. do 1988. godine. Potrošnja je modelirana kao funkcija zaostale potrošnje, cene i prihoda. Ove varijable variraju od države do države tokom vremena. Međutim, postoji dosta drugih varijabli koje mogu biti invarijantne u odnosu na stanje ili u odnosu na vreme pri čemu mogu uticati na potrošnju. Označimo ih sa Z_i i W_i respektivno. Primeri varijabli označenih sa Z_i su religija i obrazovanje. Za varijablu religija može se pretpostaviti da neće doći do velikih promena u svakoj državi za svaku godinu. Takođe, isto se može pretpostaviti i za procenat populacije koji završi srednju školu ili ima fakultetsko obrazovanje. Primeri varijabli označenih sa W_i uključuju reklame na televiziji ili radiju. Reklamiranja se puštaju širom države i ne variraju. Izostavljanje ovih varijabli dovodi do pristrasnosti u rezultirajućim procenama. Panel podaci u mogućnosti da kontroliču ove vremenski invarijantne varijable pošto vremenske serije i uporedna analiza ne mogu.
- **Veća količina informacija, više varijabilnosti, manje kolinearnosti među varijablama uz veću efikasnosti i više stepeni slobode** – u okviru podataka vremenskih serija često se vezuje problem multikolinearnosti između samih varijabli.

[1] Cheng Hsiao – profesor na Univerzitetu Kalifornija, Departman za ekonomiju, Los Anđeles

[2] Anders Klevmarken – profesor na Upsala Univerzitetu, Institut za tržišnu ekonomiju, Stokholm

Ako se posmatra prethodno navedeni primer potražnje za cigaretama može se uočiti visok nivo multikolinearnosti između cene i prihoda u vremenskoj seriji za SAD. Kako se panel podacima uvodi „treća dimezija” odnosno prostorna dimenzija ta informacija obezbeđuje veću varijabilnost podataka. Dodatno, više „informativnih” podataka mogu proizvesti pouzdanije procene parametara. Naravno, podaci moraju biti poredbeni. Takođe, panel podacima se rešava problem pristrasnosti ocenjenijh parametara koji predstavljati posledicu izostavljanja varijabli koje značajno utiču na zavisnu promenjivu.

- **Bolje istraživanje dinamike prilagođavanja –** različite socio-ekonomske pokazatelje u jednoj državi poput kretanja prihoda, stamblene mobilnosti, perioda nezaposlenosti ili perioda promene poslova je mnogo bolje izučavati korišćenjem panel podataka u odnosu na vremenske serije i uporedne podatke. Dodatno, panel podaci omogućavaju povezivanje iskustvo i ponašanje subjekta u dva različita vremenska perioda.
- **Bolji način za identifikaciju i merenje efekata koji se ne mogu identifikovati upotrebom vremenskih serija ili korišćenjem uporednih podataka –** posmatrajmo primer u kome pretpostavljamo da imamo određene uporedne podatke koji se odnose na žene koje imaju 50% udela u radnoj snazi u toku jedne godine. Jedan od načina interpretiranja ovakve pretpostavke je da svaka žena ima 50% šanse da bude zaposlena u toku godine ili da 50% žena ne radi uopšte a da 50% žena radi sve vreme. Obuhvatanjem oba slučaja može se zaključiti da je u prvom slučaju stepen obrta veliki dok u drugom slučaju stepen obrta ne postoji. Samo se korišćenjem panel podataka mogu sa jedne strane obuhvatiti, i sa druge strane napraviti razlika između ovakvih slučajeva.

Pored navedenih prednosti korišćenja panel podataka, postoje i određena ograničenja u njihovoj upotrebi. Neke od osnovih ograničenja su:

- **Nedostaci i problemi koji se odnose na prikupljanje i izgled podataka –** ovo ograničenje podrazumeva problem nedostajućih podataka tj. nepotpunih podataka o populaciji, nedostatak odgovora (ukoliko se desi da nedostaje saradnja ispitanika ili se desi greška od strane osobe koja sprovodi ispitivanje), razmaka između ispitivanja, učestalosti ispitivanja, opoziva (ispitanik se ne seća tačnog odgovora) itd.
- **Nedostaci i problemi koji se odnose na grešeke u merenju –** do grešaka u merenju najčešće dolazi zbog grešaka u pamćenju, nerazumevanju pitanja, namernim fingiranjem odogovora, neadekvatnosti informacija, uticaja ispitivača itd.
- **Nedostaci i problemi koji se odnose na kratku vremensku seriju –** tipični mikro pane podaci uključuju podatke koji se odnose na godišnje podatke pri čemu pokrivaju kratak vremenski period za svaku jedinicu posmatranja. Sa druge strane, povećanje

broja vremenskih jedinica može da implicira osipanje podataka ili da zakomplikuje kalkulaciju vazanu za panel podatke sa ograničenim zavisnim promenljivima.

- **Nedostaci i problemi koji se odnose na zavisnost uporednih podataka –** paneli regiona ili zemalja sa dužim vremenskim serijama u kojima se ne vodi računa o zavisnosti uporednih podataka često dovode do grešaka u zaključivanju.

Prikupljanje panel podataka je često skupo. Dodatno, uvek postoji nedoumica koliko često ispitivanja (intervjui) treba da se sprovode.

2. TEORIJSKA POSTAVKA MODELA PANEL PODATAKA

Modeli panel podataka pružaju informacije o ponašanju pojedinaca (pojedinih subjekata) kroz same karakteristike subjekta, ali i kroz vreme. Panel podaci i modeli koji se odnose na njih sadrže uporedne podatke, karakteristike i vremenske intervale u okviru kojih se iste posmatraju. U disciplinama ekonometrije i statistike, panel podaci se odnose na višedimenzionalne podatke koji uopšteno uključuju merenja tokom određenog vremenskog perioda. Kod njih imamo istu jedinicu preseka, posmatranu tokom vremena. Dakle, imamo izraženu kako prostornu, tako i vremensku komponentu.

Ukoliko se posmatra zavisna promenljiva Y koja se objašnjava pomoću K nezavisnih promenljivih kao i slučajnom varijacijom koja predstavlja stohastički deo modela, opšti regresioni model koji opisuje panel podatke može se prikazati pomoću sledeće jednačine:

$$Y_{it} = \beta_{1it} \cdot X_{1it} + \beta_{2it} X_{2it} + \cdots + \beta_{Kit} \cdot X_{Kit} + u_{it} = \beta_{1it} + \sum_{k=2}^{K} \beta_{kit} \cdot X_{kit} + u_{it}, \quad ((2.1)$$

$$i = 1, 2, \dots, N$$

$$t = 1, 2, \dots, T$$

$$k = 1, 2, \dots, K.$$

Parametri koji figuriraju u prikazanoj jednačini su sledeći:

- $Y_{it} -$ predstavlja vrednost zavisne promenljive za $i -$ tu jedinicu posmatranja u vremenskom periodu t,
- $X_{kit} -$ predstavlja vrednost $k -$te nezavisne promenljive za $i -$tu jedinicu posmatranja u vremenskom periodu t,
- $X_{1it} = 1$ za svako i i t,
- $\beta_{kit} -$ predstavlja vrednost nepoznatih regresionih parametara koji su promenljivi po $i -$toj jedinici posmatranja i vremenskom periodu t i
- $u_{it} -$ predstavlja slučajnu grešku sa pretpostavkom da je njeno očekivanje jednako nuli $(E(u_{it}) = 0)$ tj. da je njena aritmetička sredina jednaka nuli, i da je njena varijansa konstanta $D(u_{it}) = \sigma^2$, za svako i i t.

Kako se iz prethodne jednačine i opšteg oblika mogu uočiti dva indeksa (i i t), zaključuje se da u modelu figuriraju dve dimenzije: strukturna i vremenska. Opšti regresioni model panel podataka predstavlja najopštiji sadržaj linearnih modela. Drugim rečima, za svaku jedinicu posmatranja postoji različita reakcija zavisne promenljive na varijacije u nezavisnim promenljivim, pri čemu se reakcija takođe razlikuje za svaki vremenski period. Kako bi se različite performanse modela mogle oceniti uvode se određene pretpostavke koje se odnose na:

- osobine nezavisnih promenljivih,
- osobine slučajne greške,
- statističku zavisnost nezavisnih promenljivih i slučajne greške i
- stepen varijabiliteta regresionih parametara.

Izbor modela, o čemu će u nastavku rada diskutovati, zavisi upravo od stepena variabiliteta regresionih parametara. U tom kontekstu, postoji veći broj različitih modela panel podataka, pri čemu se, u najširem smislu, mogu podeliti na tri:

1. **model običnih najmanjih kvadrata** (engl. *Pool OLS*),
2. **model fiksnih efekata** (engl. *Fixed Effects Model*) i
3. **model slučajnih (stohastičkih) efekata** (engl. *Random Effects Model*).

Uzimajući u obzir prethodno pomenute pretpostavke dobijaju se različiti modeli. U tom smislu najjednostavniju grupu pretpostavki čini konstantnost svih regresionih parametara u modelu, što znači:

- $\beta_{kit} = \beta_k, \forall i, t; k = 1, 2, \dots, K$
- $u_{it} \sim N(0, \sigma_u^2)$.

Iz pretpostavki se može zaključiti da model sadrži K regresionih parametara pri čemu se oni mogu oceniti metodom ***običnih najmanjih kvadrata***. U tom slučaju za modeliranje se koristi uzorak od $N \cdot T$ panel podataka.

Naredna grupa pretpostavki polazi od toga da su regresioni parametri uz nezavisne promenljive konstantni pri čemu se slobodni članovi razlikuju po jedinicama posmatranja. Veliki broj faktora, koji nisu eksplicitno uključeni u model, različito deluje na zavisnu promenljivu po jedinicama posmatranja, ali je intezitet tog uticaja tokom posmatranog perioda $t = 1, 2, \dots, T$ isti. Efekti ovih faktora se zovu individualni efekti i mogu biti uključeni u model tako što će razlika između jedinica posmatranja biti obuhvaćena varijabilnim slobodnim članovima β_{1i}, odnosno:

- $\beta_{1it} = \beta_{1i}, \forall\, t = 1, 2, \dots, T$
- $\beta_{kit} = \beta_k, \forall\, i, t;\ k = 1, 2, \dots, K$
- $u_{it} \sim N(0, \sigma_u^2)$.

Iz pretpostavki se može zaključiti da su slobodni članovi β_{1i} varijabilni, ali uzimaju fiksne vrednosti po različitim jedinicama posmatranja.Model koji polazi od navedenih pretpostavki se zove ***model fiksnih efekata***. Opšti oblik ovog modela je:

$$Y_{it} = \beta_{1i} + \sum_{k=2}^{K} \beta_k \cdot X_{kit} + u_{it}.$$

Ako su varijabilni slobodni članovi β_{1i} stohastičke promenljive, tj. nisu fiksni parametri, tada važe sledeće pretpostavke:

- $\beta_{kit} = \beta_k, \forall\, i, t;\ k = 1, 2, \dots, K$
- $\beta_{1i} = \beta_1 + \varepsilon_i$, pa je $\varepsilon_i + u_{it} = w_{it}$
- $u_{it} \sim N(0, \sigma_u^2)$.

U modelu koji se zasniva na prethodnim pretpostavkama tj. da se slučajna greška w_{it} sastoji od dva člana (individualni efekti ε_ii ostatak greške u_{it}) se naziva model ***slučajnih (stohastičkih) efekata***. Opšti oblik ovog modela je:

$$Y_{it} = \beta_1 + \varepsilon_i + \sum_{k=2}^{K} \beta_k \cdot X_{kit} + u_{it} = \beta_1 + \sum_{k=2}^{K} \beta_k \cdot X_{kit} + (\varepsilon_i + u_{it}).$$

U nastavku rada dat je ilustrativan primer panel podataka kao i detaljan pregled prethodno spomenutih modela.

ILUSTRATIVAN PRIMER PANEL PODATAKA

U cilju boljeg razumevanja panel podataka, razmatra se konkretan primer. Posmatraju se podaci dati u Tabeli 1. preuzeti iz studije o teoriji investicija. U primeru se ispituje kako realna bruto investicija (Y) zavisi od realne vrednosti firme (X_2) i kapitala (X_3). Za potrebe ilustracije preuzeti su podaci za četiri kompanije: General Electris (GE), General Motor (GM), U. S. Steel (US) i Westinghouse. Podaci za svaku kompaniju, u kontekstu prethodne tri promenljive, dostupne su za period 1935. – 1954. godine. Shodno tome, postoje četiri uporedne jedinice (jedinice posmatranja) i 20 vremenskih perioda. Dakle, u skupu podataka postoji ukupno 80 opservacija. Iskustveno posmatrano, očekuje se da je Y pozitivno povezan sa X_2 i X_3.

U Tabeli 1. kolona I predstavlja zavisnu promenljivu Y tj. bruto investicije izražene u milionima dolara. Kolona F predstavlja nezavisnu promenljivu X_2 tj. vrednost firme izraženu u milionima dolara, dok kolona C predstavlja kapital.

	GE			US		
Opservacija	**I**	**F**	**C**	**I**	**F**	**C**
1935	33.1	1170.6	97.8	209.9	1362.4	53.8
1936	45	2015.8	104.4	355.3	1807.1	50.5
1937	77.2	2803.3	118	469.9	2673.3	118.1
1938	44.6	2039.7	156.2	262.3	1801.9	260.2
1939	48.1	2256.2	172.6	230.4	1957.3	312.7
1940	74.4	2132.2	186.6	361.6	2202.9	254.2
1941	113	1834.1	220.9	472.8	2380.5	261.4
1942	91.9	1588	287.8	445.6	2168.6	298.7
1943	61.3	1749.4	319.9	361.6	1985.1	301.8
1944	56.8	1687.2	321.3	288.2	1813.9	279.1
1945	93.6	2007.7	319.6	258.7	1850.2	213.8
1946	159.9	2208.3	346	420.3	2067.7	232.6
1947	147.2	1656.7	456.4	420.5	1796.7	264.8
1948	146.3	1604.4	543.4	494.5	1625.8	306.9
1949	98.3	1431.8	618.3	405.1	1667	351.1
1950	93.5	1610.5	647.4	418.8	1677.4	357.8
1951	135.2	1819.4	671.3	588.2	2289.5	341.1
1952	157.3	2079.7	726.1	645.2	2159.4	444.2
1953	179.5	2371.6	800.3	641	2031.3	623.6
1954	189.6	2759.9	888.9	459.3	2115.5	669.7

	GM			WEST		
Opservacija	I	F	C	I	F	C
1935	317.6	3078.5	2.8	12.93	191.5	1.8
1936	391.8	4661.7	52.6	25.9	516	0.8
1937	410.6	5387.1	156.9	35.05	729	7.4
1938	257.7	2792.2	209.2	22.89	560.4	18.1
1939	330.8	4313.2	203.4	18.84	519.9	23.5
1940	461.2	4643.9	207.2	28.57	628.5	26.5
1941	512	4551.2	255.2	48.51	537.1	36.2
1942	448	3244.1	303.7	43.34	561.2	60.8
1943	499.6	4053.7	264.1	37.02	617.2	84.4
1944	547.5	4379.3	201.6	37.81	626.7	91.2
1945	561.2	4840.9	265	39.27	737.2	92.4
1946	688.1	4900	402.2	53.46	760.5	86
1947	568.9	3526.5	761.5	55.56	581.4	111.1
1948	529.2	3245.7	922.4	49.56	662.3	130.6
1949	555.1	3700.2	1020.1	32.04	583.8	141.8
1950	642.9	3755.6	1099	32.24	635.2	136.7
1951	755.9	4833	1207.7	54.38	732.8	129.7
1952	891.2	4924.9	1430.5	71.78	864.1	145.5
1953	1304.4	6241.7	1777.3	90.08	1193.5	174.8
1954	1486.7	5593.6	2226.3	68.6	1188.9	213.5

Tabela 2.1. *Skup podataka korišćen u svrhu ilustrativnog primera*

Kombinovanjem şvih 80 opservacija funkcija ulaganja se može zapisati kao:

$$Y_{it} = \beta_1 + \beta_2 \cdot X_{2it} + \beta_3 \cdot X_{3it} + u_{it} \tag{2.2}$$

$$i = 1, 2, 3, 4$$

$$t = 1, 2, \ldots, 20,$$

gde i važi da je i −ta jedinica posmatranja a t predstavlja t −ti period. Ukoliko se broj jedinica posmatranja i vremenski period povećaju na maksimum, tada se maksimalan broj jedinica posmatranja obeležava sa N, a maksimalan broj vremenskih perioda sa T. Ukoliko svaka jedinica posmatranja ima isti broj opservacija vremenskih serija tada se kaže da su panel podaci ***balansirani***, u suprotnom radi se o ***nebalansiranim*** panel podacima.U datom primeru može se konstatovati da se radi o balansiranim podacima jer svaka od četiri jedinice posmatranja (kompanije) imaju 20 opservacija vremenskih serija.

2.1. MODEL OBIČNIH NAJMANJIH KVADRATA

Panel modeli sa konstantnim regresionim parametrima sadrži $K \cdot N \cdot T$ nepoznatih regresionih parametara, pri čemu se u uzorku nalazi $N \cdot T$ podataka. On predstavlja najjednostavniji panel model. Kako bi se regresioni parametri modela ocenili na osnovu $N \cdot T$opservacija uvode se ograničenja na parametre. Ocena parametara ovog modela dobija se slaganjem podataka o subjektima i u vremenskim periodima t. U prethodnom delu rada opisani su pojedina ograničenja tj. pretpostavke u okviru kojih se jedno od ograničenja zasniva na tome da su svi regresioni parametri konstantni $\beta_{kit} = \beta_k, \forall i, t; k = 1, 2, \dots , K$. Tada bi se za ceo uzorak ocenjivao jedan regresioni parametar (odsečak) $\beta_{1it} = \beta_1$ kao i parametri nagiba β_k. To znači da sve jedinice posmatranja reaguju na isti način. Model običnih najmanjih kvadrata u okviru panel podataka se u literaturi još naziva i model sa konstantnim regresionim parametrima, a opšti oblik ovog modela je:

$$y_{it} = \beta_1 + \sum_{k=2}^{K} \beta_k \cdot x_{kit} + u_{it},$$

$$i = 1, 2, \dots , N$$

$$t = 1, 2, \dots , T.$$

Prethodna jednačina za T opservacija u okviru individue i se može pretstaviti i u matričnom obliku tj:

$$y_i = X_i \beta + u_i$$

$$i = 1, 2, \dots , N$$

gde je:

$$y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \dots \\ y_{iT} \end{bmatrix}, \qquad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_K \end{bmatrix}, \qquad u_i = \begin{bmatrix} u_{i1} \\ u_{i2} \\ \dots \\ u_{iT} \end{bmatrix}$$

$$X_i = \begin{bmatrix} 1 & x_{2i1} & \dots & x_{Ki1} \\ 1 & x_{2i2} & \dots & x_{Ki2} \\ \dots & \dots & \dots & \dots \\ 1 & x_{2iT} & \dots & x_{KiT} \end{bmatrix}$$

pri čemu:

- β predstavlja vektor nepoznatih regresionih parametara dimenzija $K \times 1$,
- u_i predstavlja vektor slučajnih grešaka dimenzija $T \times 1$,
- matrica X_i ima dimenzije $T \times K$ i sadrži opservacije K nezavisnih promenljivih za i −tu jedinicu posmatranja.

Ako se jednačina proširi na svih $N \cdot T$ opservacija, jednačina $y_i = X_i\beta + u_i$ dobija sledeći oblik:

$$y = X \cdot \beta + u,$$

gde parametri koji figuriraju u jednačini imaju sledeći oblik i dimenzije:

$$y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \dots \\ y_{iT} \end{bmatrix}_{NT\times 1}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_K \end{bmatrix}_{K\times 1}, \quad u_i = \begin{bmatrix} u_{i1} \\ u_{i2} \\ \dots \\ u_{iT} \end{bmatrix}_{NT\times 1}$$

pri čemu matrica X ima oblik:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \dots \\ X_N \end{bmatrix}_{NT\times K}.$$

Sublimacija pretpostavki modela najmanjih kvadrata ogleda se u sledećem:

- svi regresioni parametri za sve jedinice posmatranja su konstantni kroz vreme, tj:

$$\beta_{kit} = \beta_k, \forall i, t; k = 1, 2, \dots, K$$

- varijansa slučajne greške nije ista za sve jednice posmatranja po vremenskim periodima, tj:

$$D(u_{it}) = E({u_{it}}^2) - \left(E(u_{it})\right)^2 = E({u_{it}}^2) = \sigma_i^2 (ili\ \sigma_t^2)$$

što drugim rečima predstavlja heteroskedastičnost,

- za isti vremenski period t postoji korelacija između slučajnih grešaka u odnosu na različite jedinice posmatranja $(i \neq j)$, dok za različite vremenske periode t i s korelacija je jednaka nuli, tj.

$$E(u_{it}, u_{jt}) = \sigma_{ij},$$

$$E(u_{it}, u_{js}) = 0,$$

$$i \neq j.$$

Koristeći ocenjene vrednosti regresionih parametara, ocena zavisne promenljive individude i -y_i (u oznaci $\widehat{y_{it}}$) data je sledećom jednačinom:

$$\widehat{y_i} = \widehat{\beta_1} + \sum_{k=2}^{K} \widehat{\beta_k} \cdot x_{ki}$$

Obzirom da je glavni cilj ocena $\beta_1, \beta_2, \beta_3, \dots, \beta_K$ to znači da je neophodno pronaći sledeći skup parametra:

$$\widehat{\beta_1}, \widehat{\beta_2}, \widehat{\beta_3}, \dots, \widehat{\beta_K}.$$

Kako se u modelima greška želi svesti na što manju vrednost, minimiziranjem sume kvadrata iste mogu se dobiti ocene parametara $\widehat{\beta_1}, \widehat{\beta_2}, \widehat{\beta_3}, \dots, \widehat{\beta_K}$. Dakle:

$$\sum_{i=1}^{N} u_i^2 = \sum_{i=1}^{N} (y_i - \widehat{y_i})^2 = \sum_{i=1}^{N} \left(y_i - \left(\widehat{\beta_1} + \sum_{k=2}^{K} \widehat{\beta_k} \cdot x_{ki} \right) \right)^2.$$

Jedan od načina minimiziranja date jednačine jeste putem parcijalnih izvoda po svim parametrima $\widehat{\beta_i}, i = 1, 2, \dots, N$. Izjednačavajući parcijalne izvode sa nulom, a zatim rešavajući sistem od N jednačina dobijaju se ocene parametara $\widehat{\beta_1}, \widehat{\beta_2}, \widehat{\beta_3}, \dots, \widehat{\beta_K}$. Obzirom da se ocene parametara dobijaju minimiziranjem sume kvadrata ovaj metod je proglašen kao metod ***običnih najmanjih kvadrata***.

2.2. MODEL FIKSNIH EFEKATA

Model fiksnih efekata predstavlja model višestruke linearne regresije. Posmatrajući primer korišćen za ilutraciju, može se reći da procena regresionih parametara zavisi od pretpostavki vezanih za odsečak, nagib koeficijenata i slučajne greške.Postoji nekoliko varijacija ovog modela u zavisnosti od pretpostavki i to su:

1. odsečak i nagibi koeficijenata su konstantni kroz vreme i posmatrane jedinice, a slučajna greška obuhvata razlike kroz vreme i pojedinačne posmatrane jedinice;
2. nagibi koeficijenata su konstantni, ali odsečak varira u zavisnosti od pojedinačno posmatranih jedinica;
3. nagibi koeficijenata su konstantni, ali odsečak varira kroz vreme i u zavisnosti od pojedinačno posmatranih jedinica;
4. svi koeficijenti, kako odsečak tako i nagibi koeficijenata, variraju u zavisnosti od posmatranih jedinica;
5. svi koeficijenti, kako odsečak tako i nagibi koeficijenata, variraju kroz vreme i u zavisnosti od posmatranih jedinica.

Kao što se može videti, svaki od slučajeva povećava kompleksnost i procenu parametara modela.

2.2.1. Svi koeficijenti su konstanti kroz vreme i posmatrane jedinice

Ovaj slučaj predstavlja najjednostavniji, „naivni“ pristup koji podrazumeva zanemarivanje vremenske instance ali i prostornih dimenzija i primena obične regresije metodom običnih najmanjih kvadrata o čemu je bilo reči u prethodnom delu rada (pogledati Poglavlje).Ocenjeni regresioni parametri dobijeni pomoću metode običnih najmanjih kvadrata putem Stata softvera su:

Source	SS	df	MS
Model	4849457.34	2	2424728.67
Residual	1560689.71	77	20268.6975
Total	6410147.05	79	81141.1019

Number of obs	80
F(2, 77)	119.63
Prob > F	0
R-squared	0.7565
Adj R-squared	0.7502
Root MSE	142.37

I	Coef.	Std. Err.	t	P>t	[95% Conf. Interval]	
F	0.1100955	0.0137297	8.02	0	0.0827563	0.1374348
C	0.3033932	0.0492957	6.15	0	0.2052328	0.4015535
_cons	-63.30414	29.6142	-2.14	0.036	-122.2735	-4.334735

Tabela 2.2. *Rezultati ilustrativnog primera – koeficijenti konstantni kroz vreme i posmatrane jedinice*

Stoga, dobijena je jednačina oblika:

$$\hat{Y} = -63.30414 + 0.1100955 \cdot X_2 + 0.3033932 \cdot X_3. \tag{2.3}$$

Iz dobijenog sumarnog prikaza regresije može se zaključiti da su obe nezavisne promenljive statistički značajne (p-vrednost <0.05), pri čemu nagibi koeficijenata pozitivni što je bilo i očekivano. Vrednost $R^2 = 0.7565$ je relativno visoka, a Y je pozitivno korelisano sa X_2 i X_3. Predstavljeni procenjeni model pretpostavlja da su vrednost odsečka i koeficijenti nagiba obe promenljiveistiza sve četiri kompanije. Očigledno je da se radi o visoko restrikcionim pretpostavkama te se u narednom delu opisuju fleksibilnije varijacije modela.

2.2.2. Nagibi koeficijenata su konstantni, ali odsečak varira kroz posmatrane jedinice

Posmatrajmo jednačinu (2.1) i ukoliko se dozvoli da odsečak varira kroz posmatrane jedinice($\beta_{1it} = \beta_{1i}, i = 1, 2, \ldots, K$), a restrikcija se nametne nagibima koeficijenata tako da glasi da je svaki nagib konstantan kroz posmatrane jedinice($\beta_{kit} = \beta_k, i = 1, 2, \ldots, K\ i\ t = 1, 2, \ldots, T$) tada se dobija jednačina oblika:

$$Y_{it} = \beta_{1i} + \sum_{k=2}^{K} \beta_k \cdot X_{kit} + u_{it}. \tag{2.4}$$

Kako bi se razmotrila ova varijacija modela fiksnih efekata, posmatra se prethodno dati ilustrativni primer sa kompanijama. Jedan od načina da se uračuna individualnost svake kompanije tj. svake jedinice posmatranja je da se dopusti da odsečak varira za svaku kompaniju pretpostavljajući da su nagibi koeficijenata nepromenjeni. U kontekstu primera, jednačina pod ovakvim pretpostavkama je:

$$Y_{it} = \beta_{1i} + \beta_2 \cdot X_{2it} + \beta_3 \cdot X_{3it} + u_{it}. \tag{2.5}$$

Dakle, indeks i stoji uz odsečak kako bi se sugerisalo da odsečci za četiri firme mogu biti različiti. Razlike mogu nastati zbog specifičnih karakteristika svake kompanije koje zavise od filozofije menadžmenta istih. Takođe, model se naziva ***„model fiksnih efekata“*** zbog činjenice da iako odsečak varira kroz posmatrane jedinice, on se ne menja kroz vreme što znači da je vremenski nepromenljiv. Postavlja se pitanje kako dozvoliti da odsečak varira između kompanija? Uvodeći veštačke promenljive (engl. ***dummy variable***[3]), tj. veštačke diference odsečaka (engl. ***differentialdummy variable***) dozvoljava se da odsečak varira kroz posmatrane jedinice tj. kompanije.Zbog toga jednačina (2.3) se može zapisati kao:

---

[3] Označavanjem određenog obeležja sa 0 (opservacija ne poseduje obeležje) i 1 (opservacija poseduje obeležje) mogu se kvantifikovati atributi. Na primer, 1 može implicira da je osoba ženskog pola, a 0 da je osoba muškog pola. Promenljive koje imaju za pretpostavku binarnu vrednost (0 ili 1) se nazvaju ***veštačke promenljive***.

$$Y_{it} = \alpha_1 + \alpha_2 \cdot D_{2i} + \alpha_3 \cdot D_{3i} + \alpha_4 \cdot D_{4i} + \beta_2 \cdot X_{2it} + \beta_3 \cdot X_{3it} + u_{it}.$$

gde D_{2i}, D_{3i} i D_{4i} predstavljaju veštačke promenljive takve da ako:

- opservacija pripada GM kompaniji tada je $D_{2i} = 1$, u suprotnom je $D_{2i} = 0$,
- opservacija pripada US kompaniji tada je $D_{3i} = 1$, u suprotnom je $D_{3i} = 0$,
- opservacija pripada West kompaniji tada je $D_{4i} = 1$, u suprotnom je $D_{4i} = 0$.

Kako postoje četiri kompanije, u modelu se koriste tri veštačke varijable kako bi se izbegao problem savršene kolinearnosti. Zbog toga za kompaniju GE ne postoji veštačka promenljiva. Drugim rečima,α_1 predstavlja odsečak kompanije GE, a α_2, α_3 i α_4 predstavljaju diference koeficijenata odsečaka, i govore koliko se odsečci kompanija GM, US i West razlikuju od odsečka kompanije GE. U tom smislu, kompanija GE postaje referentna (poredbena), pri čemu se u modelu može odabrati bilo koja druga jedinica posmatranja tj. kompanija koja će imati to svojstvo. S obzirom da se u modelu koriste veštačke promenljive, u literaturi se ovaj model još naziva i ***LSDV model – model najmanjih kvadrata veštačih varijabli*** (engl. ***Least-Squares Dummy Variable Regression Model***). Ocenjeni regresioni parametri iz primera dobijeni pomoću LSDV modela putem Stata softvera su:

Source	SS	df	MS
Model	5990684.12	5	1198136.82
Residual	419462.925	74	5668.4179
Total	6410147.05	79	81141.1019

Number of obs	80
F(5, 74)	211.37
Prob > F	0
R-squared	0.9346
Adj R-squared	0.9301
Root MSE	75.289

I	Coef.	Std. Err.	t	P>t	[95% Conf. Interval]	
D2	161.5722	46.45639	3.48	0.001	69.00581	254.1386
D3	339.6328	23.98633	14.16	0	291.839	387.4266
D4	186.5666	31.50681	5.92	0	123.7879	249.3452
F	0.1079481	0.0175089	6.17	0	0.0730608	0.1428354
C	0.3461617	0.0266645	12.98	0	0.2930315	0.3992918
_cons	-245.7924	35.81112	-6.86	0	-317.1476	-174.4372

Tabela 2.3. *Rezultati ilustrativnog primera – nagibi koeficijenata su konstantni, ali odsečak varira kroz posmatrane jedinice*

Stoga, dobijena je jednačina oblika:

$$\hat{Y} = -245.7924 + 161.5722 \cdot D_2 + 339.6328 \cdot D_3 + 186.5666 \cdot D_4 + 0.1079481 \cdot X_2 + 0.3461617 \cdot X_3 \quad (2.6)$$

Iz dobijenog sumarnog prikaza regresije može se zaključiti da su obe nezavisne kao i tri veštačke promenljive statistički značajne (p-vrednost <0.05). Vrednosti odsečaka za četiri kompanije su: -245.7924 za GE, -84.2202 = -245.7924 + 161.5722 za GM, 93.8404 = -245.7924 + 339.6328 za US i -59.2258 = -245.7924 + 186.5666 za West. Različitosti u odsečku svake kompanije predstavljaju određene karateristike istih koje se mogu ogledati na primer u filozofiji menadžmenta svake od njih.

Ukoliko se poredi model(2.5)sa modelom(2.6), može se reći da model (2.6) bolji i to zbog statističke značajnosti nezavisnih promenljivih kao i povećanja R^2 racija[4] ($R^2 = 0.7565, R^2 = 0.9346$, respektivno). Međutim, povećanje R^2 racija nije iznenađujuće imajući

[4] R^2racio u literaturi poznat kao ***koeficijent determinacije*** pokazuje koji deo varijacija zavisne promenljive je objašnjen modelom, odnosno varijacijama nezavisne promenljive.

u vidu povećanje broja prediktora u drugom modelu. Takođe, valjanost jednog modela u odnosu na drugi se može testirati pomoću F – testa:

$$F: \frac{\frac{(R_2^2 - R_1^2)}{3}}{\frac{(1 - R_2^2)}{74}} = \frac{\frac{(0.9346 - 0.7565)}{3}}{\frac{(1 - 0.9346)}{74}} = 66.998,$$

Gde 3 i 74 predstavljaju stepene slobode u modelima. Kako je $F_{0.05}(3,74) \approx 2.76$ značajno niži od dobijene vrednosti F-testa, može se reći da je prvi, restriktivniji model lošiji.

Ovakav model fiksnih efekata se može uopštiti za K posmatranih jedinica, pa bi u tom slučaju jednačina modela sa uvedenim veštačkim varijablama izgledala kao:

$$Y_{it} = \alpha_1 + \alpha_2 \cdot D_{2i} + \alpha_3 \cdot D_{3i} + \cdots + \alpha_N \cdot D_{Ni} + \beta_2 \cdot X_{2it} + \cdots + \beta_K \cdot X_{Kit} + u_{it},$$

gde su:

$$D_{ji} = \begin{cases} 1, za\ j = i \\ 0, za\ j \neq i \end{cases}, \qquad j = 2, \ldots, N.$$

2.2.3. Nagibi koeficijenata su konstantni, ali odsečak varira kroz vreme i kroz pojedinačno posmatrane jedinice

Ova varijacija modela fiksinh efekata panel podataka predstavlja proširenje u odnosu na prethodnu metodu. Ona se sastoji u tome što je dozvoljeno variranje odsečka kroz jedinice posmatranja, što je prikazano u prethodnom delu, ali i variranje odsečka kroz vreme. Dakle, uvodi se vremenska komponenta odsečka koja predstavlja proširenje LSDV modela.

Obzirom da su svi regresioni parametri uz nezavisne promenljive konstanti, pri čemu su slobodni članovi (odsečci) promenljivi po jedinicama posmatranja i vremenskim periodima, tada se model prikazuje pomoću jednačine:

$$Y_{it} = \beta_{it} + \sum_{k=2}^{K} \beta_k \cdot X_{kit} + u_{it}, \qquad i = 1, 2, \dots, N; t = 1, 2, \dots, T. \tag{2.7}$$

Dakle, indeks i stoji uz odsečak kako bi se sugerisalo da odsečci variraju po jedinicama posmatranja, a indeks t stoji uz odsečak kako bi se sugerisalo da odsečci variraju po vremenskim periodima. Kako bi se omogućilo odsečku da varira:

- po jedinicama posmatranja uvedene su veštačke promenljive (prethodni deo rada govori o tome) u oznaci D_{ji} i
- po vremenskim periodima, uvode se nove veštačke varijable, tako da ih bude za jednu manje od broja vremenskih perioda $(T-1)$ kako bi se izbegla savršena kolinearnost u oznaci A_t,

pa jednačina (2.7) dobija sledeći oblik:

$$\begin{gathered} Y_{it} = \alpha_1 + \alpha_2 \cdot D_{2i} + \alpha_3 \cdot D_{3i} + \cdots + \alpha_N \cdot D_{Ni} \\ +\lambda_0 + \lambda_1 \cdot A_{1t} + \lambda_2 \cdot A_{2t} + \cdots + \lambda_{T-1} \cdot A_{T-1} \\ +\beta_2 \cdot X_{2it} + \cdots + \beta_K \cdot X_{Kit} + u_{it} \\ = \alpha_1 + \lambda_0 + \sum_{j=2}^{N} \alpha_j \cdot D_{ji} + \sum_{l=1}^{T-1} \lambda_{lt} \cdot A_{lt} + \sum_{k=2}^{K} \beta_k \cdot X_{kit} + u_{it}, \end{gathered}$$

pri čemu je:

$$D_{ji} = \begin{cases} 1, za\ j = i \\ 0, za\ j \neq i \end{cases}, \qquad j = 2, 3, \dots, N.$$

$$A_{lt} = \begin{cases} 1, za\ l = t \\ 0, za\ l \neq t \end{cases}, \qquad l = 1, 2, \dots, T-1.$$

U kontekstu ilustrativnog primera jednačina u okviru koje se dozvoljava variranje odsečka po vremenskim periodima i jedinicama posmatranja je:

$$Y_{it} = \alpha_1 + \alpha_2 \cdot D_{GMi} + \alpha_3 \cdot D_{USi} + \alpha_3 \cdot D_{WESTi} + \lambda_0 + \lambda_1 \cdot A_{1935} + \lambda_2 \cdot A_{1936} + \cdots + \lambda_{19} \cdot A_{1953} + \beta_2 \cdot X_{2it} + \beta_3 \cdot X_{3it} + u_{it},$$

gde veštačka promenljivaA_{1935} prima vrednost 1 ukoliko se radi o 1935. godini, a u svim ostalim slučajevima prima vrednost 0 itd., a veštačka promenljiva D_{GMi} prima vrednost 1 ukoluko se radi o kompaniji GM, u suprotnom prima vrednost 0 itd.

Ocenjeni regresioni parametri iz primera dobijeni pomoću LSDV modela putem Stata softvera su:

Source	SS	df	MS
Model	6075088.95	23	264134.302
Residual	335058.101	56	5983.18038
Total	6410147.05	79	81141.1019

Number of obs	80
F(23, 56)	44.15
Prob > F	0
R-squared	0.9477
Adj R-squared	0.9263
Root MSE	77.351

I	Coef.	Std. Err.	t	P>t	[95% Conf. Interval]	
D2i	124.3792	65.71877	1.89	0.064	-7.271326	256.0297
D3i	339.4501	24.83221	13.67	0	289.7052	389.195
D4i	206.8234	39.51051	5.23	0	127.6743	285.9725
A1935	81.26807	54.86044	1.48	0.144	-28.63059	191.1667
A1936	39.34887	51.39371	0.77	0.447	-63.60511	142.3028
A1937	-13.56334	54.2051	-0.25	0.803	-122.1492	95.02252
A1938	-0.6222437	51.30645	-0.01	0.99	-103.4014	102.1569
A1939	-53.4927	49.81506	-1.07	0.288	-153.2843	46.29887
A1940	6.927029	50.1264	0.14	0.891	-93.48822	107.3423
A1941	62.79013	49.69454	1.26	0.212	-36.75999	162.3402
A1942	71.58095	50.27552	1.42	0.16	-29.13302	172.2949

A1944	22.87708	49.48455	0.46	0.646	-76.25239	122.0066
A1945	0.0897458	49.41918	0	0.999	-98.90878	99.08827
A1946	61.56694	49.22212	1.25	0.216	-37.03681	160.1707
A1947	56.35226	49.71731	1.13	0.262	-43.24348	155.948
A1948	49.23402	50.61231	0.97	0.335	-52.15462	150.6227
A1949	-10.44555	50.21818	-0.21	0.836	-111.0447	90.15355
A1950	-4.896163	49.6867	-0.1	0.922	-104.4306	94.63825
A1951	10.6833	47.41816	0.23	0.823	-84.30668	105.6733
A1952	23.15775	47.4353	0.49	0.627	-71.86656	118.1821
A1953	24.84731	48.74272	0.51	0.612	-72.79609	122.4907
F	0.1232068	0.02696	4.57	0	0.0691995	0.1772141
C	0.3489556	0.038509	9.06	0	0.2718129	0.4260984
_cons	-297.9178	62.39597	-4.77	0	-422.9119	-172.9237

Tabela 2.4. *Rezultati ilustrativnog primera – nagibi koeficijenata su konstantni, ali odsečak varira kroz vreme i kroz pojedinačno posmatrane jedinice*

Iz sumarnog prikaza može se uočiti da su individualni efekti posmatranih jedinica statistički značajni, dok se za individualne efekte vremena to ne može tvrditi imajući u vidu statističku značajnost veštačkih varijabli koje se odnose na njih (kod svih važi da je p-vrednost > 0.05). To konkretno znači da je model koji uključuje samo efekte posmatranih jedinica ali ne i efekte vremena bolji.Drugim rečima, investicione funkcije za četiri kompanije se razlikuju u odsečcima.

2.3. MODEL SLUČAJNIH (STOHASTIČKIH) EFEKATA

Iako ih je jednostavno primeniti, modeli fiksnih efekata ili običnih najmanjih kvadrata mogu biti „skupi“ uzimajući u obzir stepene slobode ukoliko u uzorku postoji nekoliko jedinica posmatranja. Pored toga se postavlja pitanje da li je uključivanje pomoćnih promenljivih (kao i posledica tog uvođenja - gubitak stepeni slobode) zaista neophodno.

Model fiksnih efekata je adekvatan kada se razlike između subjekata mogu posmatrati kao pararmetarske smene u regresionom modelu. U tom smislu, rezultati dobijeni modelom fiksnih efekata su primenljivi samo na jedinice koje su odabrane za posmatranje, tj. ne mogu se primeniti na celu populaciju. Sa druge strane, kod modela stohastičkih efekata, smatra se da su jedinice posmatranja koje učestvuju u istraživanju izdvojene iz mnogo veće populacije dostupnih jedinica posmatranja. U tom slučaju interpretacija rezultata dobijenih modelom slučajnih efekata se može primeniti na veću populaciju.

Evaluacijom modela fiksnih efekata, uvedene veštačke promenljive (koje su po svojoj prirodi binarne) se mogu interpretirati i kao nedostatak informacija o modelu. Ovakav nedostatak informacija je moguće izraziti putem slučajne promenljive u_i. Pristalice modela slučajnih efekata predlažu ovakav pristup koji se naziva ***komponenta slučajne greške - ECM*** (enlg. ***error comonents model***) ili ***model slučajnih (stohastičkih) efekata – REM*** (engl. ***random effects model***).

Model slučajnih efekata se može predstaviti kao:

$$Y_{it} = \beta_{1i} + \sum_{k=2}^{K} \beta_k \cdot X_{kit} + u_{it}.$$

Do sada je β_{1i} tretiran kao fiksan parametar. U ovom modelu, umesto fiksnog tretmana parametra β_{1i}, pretpostavlja se da je to stohastička promenljiva, čija je srednja vrednost β_1. Ovako tretiran odsečak za individualne jedinice posmatranja se mogu izraziti na sledeći način:

$$\beta_{1i} = \beta_1 + \varepsilon_i, \qquad i = 1, 2, \dots, N,$$

pri čemu ε_i predstavlja slučajnu komponentu, tj. slučajnu grešku sa srednjom vrednošću nula i varijansom σ^2, odnosno:

$$\varepsilon_i \colon N(0, \sigma_\varepsilon^2).$$

Prethodna konstatacija o srednjoj vrednosti i varijansi greške odsečka svake individualne jedinice posmatranja predstavlja jednu od osnovnih pretpostavki modela slučajnih efekata.

Kako bi se na adekvatan način razumeo model slučajnih efekata, koristiće se ilustrativan primer pomenut na početku poglavlja. Osnovna ideja ovog modela u kontekstu primera jeste jednačina:

$$Y_{it} = \beta_{1i} + \beta_2 X_{2it} + \beta_3 X_{3it} + u_{it}. \tag{2.8}$$

Očigledno, opštu jednačinu ovog modela od n nezavisnih promenljivih smo sveli na jednačinu sa dve nezavisne promenljive koje se menjaju u zavisnosti od $i = 1,2,3,4$ kompanija u vremenu $t = 1935., 1936, \dots, 1953$ godine.

Umesto fiksnog tretmana β_{1i} pretpostavlja se da je ona slučajna promenljiva sa srednjom vrednošću β_1[5], pri čemu je odsečak za svaku kompaniju i (jedinicu posmatranja) se prikazuje kao:

$$\beta_{1i} = \beta_1 + \varepsilon_i, \qquad i = 1, 2, \dots, N,$$

gde β_1 predstavlja fiksni deo, a ε_i predstavlja slučajnu grešku ($\varepsilon_i: N(0, \sigma_\varepsilon^2)$). U primeru sa četiri kompanije, ovakve pretpostavke upućuju na to da su one uzete iz mnogo većeg skupa takvih kompanija i da sve one imaju zajedničku srednju vrednost odsečka (koji je jednak β_1) pri čemu su individualne razlike vrednosti odsečaka za svaku kompaniju reflektovane kroz slučajnu grešku. Zamenjujući $\beta_{1i} = \beta_1 + \varepsilon_i$ u jednačinu modela dobija se:

$$\begin{aligned} Y_{it} &= \beta_{1i} + \beta_2 X_{2it} + \beta_3 X_{3it} + u_{it} \\ &= \beta_1 + \varepsilon_i + \beta_2 X_{2it} + \beta_3 X_{3it} + u_{it} \\ &= \beta_1 + \beta_2 X_{2it} + \beta_3 X_{3it} + w_{it} \end{aligned}$$

Pri čemu je:

$$w_{it} = \varepsilon_i + u_{it}.$$

[5]Koristeći definiciju da je$\beta_{1i} = \beta_1 + \varepsilon_i$i pretpostavku da greška$\varepsilon_i: N(0, \sigma_\varepsilon^2)$dobija se da je srednja (očekivana) vrednost parametraβ_{1i}kao$E(\beta_{1i}) = E(\beta_1) + E(\varepsilon_i) = \beta_1 + 0 = \beta_1$.

Predstavljena greška w_{it} se sastoji od dve komponente:

1. ε_i −uporedna ili individualno specifična komponenta greške i
2. u_{it} −kombinovana(vremensko i individualno specifična) komponenta greške.

Upravo zbog greške w_{it} ovaj model se u literaturi naziva i model komponente slučajne greške.

Pretpostavke koje ovaj model sadrži su sledeće:

1. $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$,
2. $u_{it} \sim N(0, \sigma_u^2)$,
3. $E(\varepsilon_i, u_{it}) = 0,\ E(\varepsilon_i, \varepsilon_j) = 0\ \ (i \neq j)$,
4. $E(u_{it}, u_{is}) = 0,\ E(u_{it}, u_{jt}) = 0, E(u_{it}, u_{js}) = 0\ \ (i \neq j; t \neq s)$.

Drugim rečima, individualno specifične komponente greške nisu međusobno korelisane niti su autokorelisane kroz uporedne jedinice (jedinice posmatranja) niti su autokorelisane kroz jedinice vremenskih serija.

Lako se može uočiti osnovna razlika između modela fiksnih i modela slučajnih efekata. Kod modela fiksnih efekata uporedne jedinice imaju sopstvenu (fiksnu) vrednost odsečka, u svih N opservacija. Sa druge strane, u modelu slučajnih efekata odsečak β_1 predstavlja srednju vrednost svih odsečaka pri čemu greška ε_i predstavlja slučajnu devijaciju individualnih odsečaka od β_1.

Kao rezultat prethodno uvedenih pretpostavki dobija se:

$$E(w_{it}) = 0$$

$$var(w_{it}) = \sigma_\varepsilon^2 + \sigma_u^2.$$

Uočimo, ako je $\sigma_\varepsilon^2 = 0$ modeli (1) i (2) se poklapaju. Dodatno, obzirom da je $var(w_{it}) = \sigma_\varepsilon^2 + \sigma_u^2$ važi da je greška w_{it} homoskedastična. Takođe, može se pokazati da su w_{it} i $w_{is} (t \neq s)$korelisani, odnosno greške datih jedinica posmatrnja u dva različita vremenska momenta su korelisani. Koeficijent korelacije za w_{it} i w_{is} je:

$$corr(w_{it}, w_{is}) = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_u^2}.$$

Nakon teorijske postavke modela, dobijeni su ocenjeni regresioni parametri iz primera pomoću modela slučajnih efekata putem Stata softvera:

```
Random-effects GLS regression                   Number of obs      =        80
Group variable: Firma                           Number of groups   =         4

R-sq:  within  = 0.8068                         Obs per group: min =        20
       between = 0.7303                                        avg =      20.0
       overall = 0.7554                                        max =        20

                                                Wald chi2(2)       =    317.79
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
           I |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           F |   .1076555   .0168169     6.40   0.000     .0746949     .140616
           C |   .3457104   .0265451    13.02   0.000     .2936829    .3977378
       _cons |  -73.03531   83.94957    -0.87   0.384    -237.5734    91.50283
-------------+----------------------------------------------------------------
     sigma_u |  152.15823
     sigma_e |  75.288896
         rho |  .80332023   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Tabela 2.5. *Rezultati ilustrativnog primera – model slučajnih (stohastičkih) efekata*

Iz dobijenog sumarnog prikaza modela slučajnih efekata u odnosu na ilustrativan primer može se zaključiti da su obe nezavisne promenljive statistički značajne (p-vrednost <0.05). Vrednost $R^2 = 0.7554$ je visoka, pri čemu je Y pozitivno korelisano sa X_2 i X_3.

3. POTENCIJALNI PROBLEMI U ANALIZI PANEL PODATAKA

MOTIVACIJA

Različiti matematički modeli se uglavnom bave stohastičkim odnosima. Najjednostavniji oblik stohastičkog odnosa između dve promenljive X i Y naziva se *linearna regresija*. Ovaj model se formalno zapisuje u obliku:

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

gde je:

- Y_i − zavisna promenljvia u i −toj opservaciji (opažanju)
- X_i − nezavisna (objasnidbena) promenljiva u i −toj opservaciji (opažanju),
- ε_i −slučajno odstupanje u i −toj opservaciji (opažanju) i
- α i β − parametri regresije.

Stohastička osobina ovakvog regresionog modela podrazumeva da za svaku vrednost nezavisne pormenljive X postoji raspodela verovatnoća za vrednost zavisne promenljive Y. Drugim rečima, vrednost zavisne promenljive Y se nikada ne može tačno predvideti. Neizvesnost promenljive Y se pojavljuje zbog prisutnosti slučajnog odstupanja ε koje upravo utiče na slučajnost zavisne promenljive Y. U tom smislu osnovne pretpostavke ovakvih tipova regresionog modela jesu:

1. normalnost: ε_i je normalno raspoređeno,
2. sredina je jednaka nuli: $E(\varepsilon_i) = 0$,
3. homoskedastičnost: $var(\varepsilon_i) = \sigma^2$,
4. odsutnost autokorelacije: $cov(\varepsilon_i, \varepsilon_j) = 0$ i
5. nestohastičnost promenljive X i
6. ne postoji jaka linearna veza (ukoliko se radi o višestrukoj regresiji) između nezavisnih promenljvih.

U tom kontekstu, mogu se definisati potencijalni problemi koji se pojavljuju u regresionim modelima. Ovaj deo rada sadrži pojašnjenja za moguće probleme koji se mogu pojaviti u skupu podataka koji se koriste za pravljenje panel modela koji po svojoj suštini pripadaju linearnim regresionim modelima. Shodno tome, mogu se pojaviti sledeći problemi:

1. heteroskedastičnost,
2. autokorelacija i
3. multikolinearnost.

U narednom delu rada dat je detaljan pregled svakog od pomenutih potencijalnih problema.

3.1. HETEROSKEDASTIČNOST

Reč heteroskedastičnost potiče od grčke reči *hetero* što znači *različito* i *skedasis* što znači raspršenost. U regresionim modelima heteroskedastičnost se opisuje kao slučaj u kome se varijansa greške (slučajnog odstupanja) modela menja obzirom na opažanje. Drugim rečima, prema pretpostavci broj 3. datoj u motivaciji ovog poglavlja važi:

$$var(\varepsilon_i) = \sigma^2, \qquad \forall i.$$

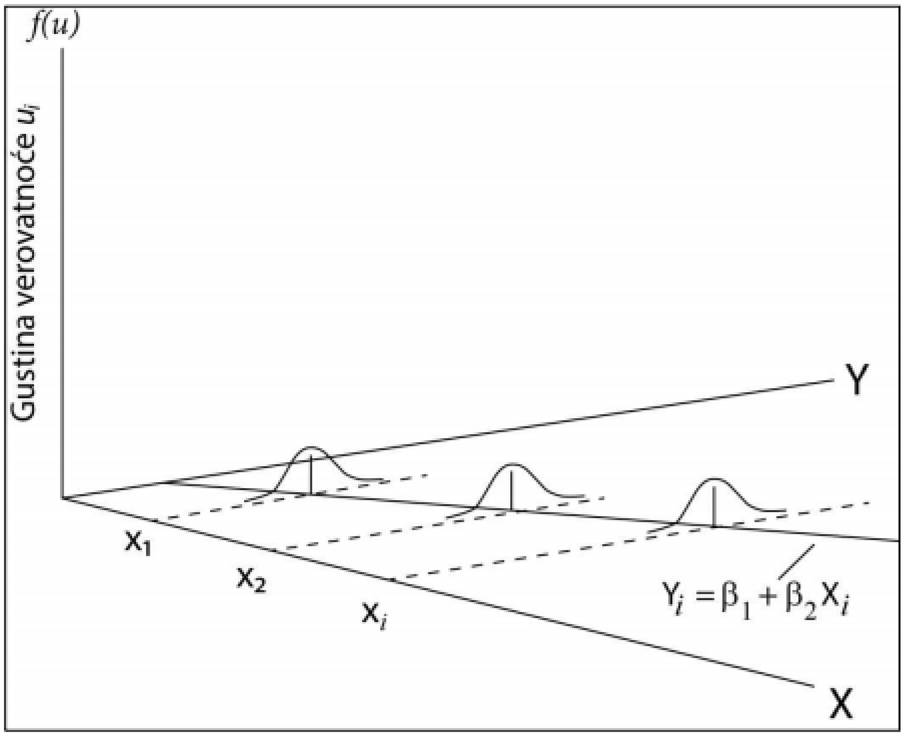
Dodatno, s obzirom da važi je jedna od pretpostavki: sredina je jednaka nuli, tada važi:

$$var(\varepsilon_i) = \sigma^2 = E(\varepsilon_i^2) - \left(E(\varepsilon_i)\right)^2 = E(\varepsilon_i^2) - 0^2 = E(\varepsilon_i^2).$$

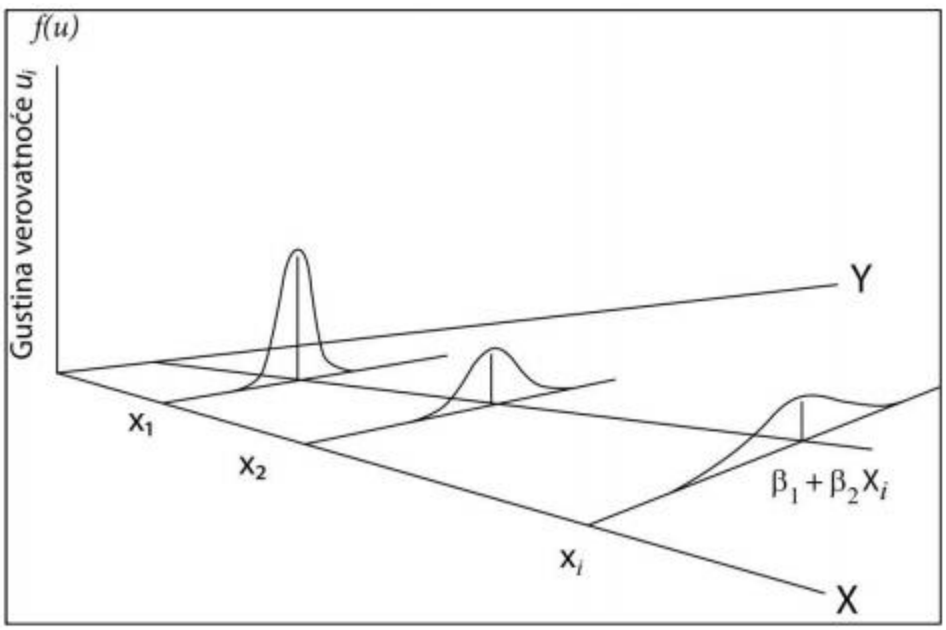
Ovakva osobina slučajnog odstupanja u regresionom modelu se naziva **homoskedastičnost.** Ona kaže da je varijansa slučajnog odstupanja konstantna za sva opažanja.

Definicija 3.1. Greška regresionog modela ε je **homoskedastična** ako $E(\varepsilon_i^2) = \sigma^2$ ne zavisi od opažanja nezavisne promenljvie X (Grafik 1.).

Defincicija 3.2.Greška regresionog modela ε je **heteroskedastična** ako $E(\varepsilon_i^2) = \sigma^2(X)$ zavisi od opažanja nezavisne promenljvie X (Grafik 2.).



Grafik 3.1. *Homoskedastičnost*



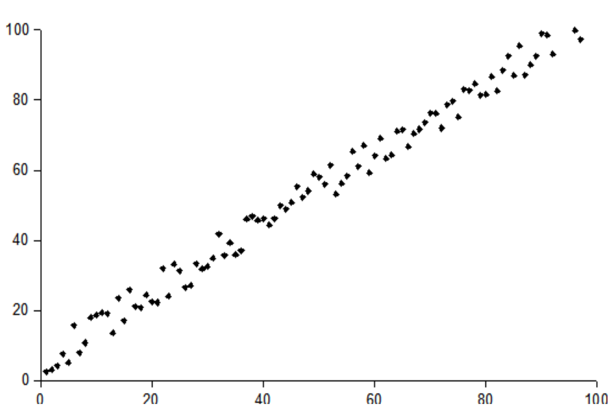
Grafik 3.2. *Heteroskedastičnost*

Ukoliko je slučajna greška heteroskedastična ocenjeni koeficijenti dobijeni regresijom će i dalje biti konzistentni, ali neće biti efikasni. Dodatno, standardna greška ocenjenih vrednosti će biti pristrasna i neobjektivna. Zbog toga neophodno je otkriti da li postoji prisustvo heteroskedastičnosti u panel podacima koji se koriste u analizi.

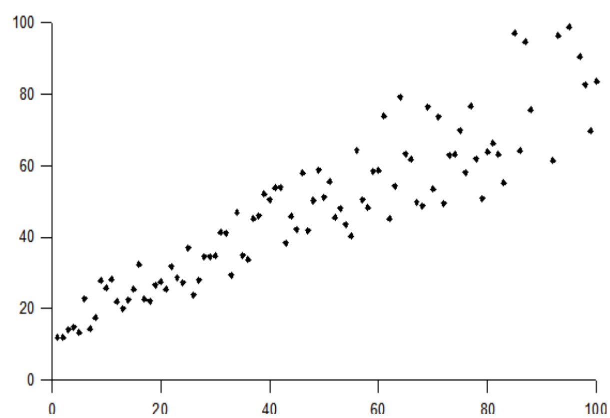
Važno je razumeti da se pojmovi homoskedastičnosti i heteroskedastičnosti odnose na uslovnu varijansu:

$$var(\varepsilon_i) = E(\varepsilon_i^2 \middle| X = x).$$

Osnovi problem jeste detekcija heteroskedastičnosti. Najjednostavniji test za otkrivanje (testiranje pretpostavke homoskedastičnosti) heteroskedastičnosti je test *očne jabučice*. Ovaj test se sastoji u posmatranju dijagrama rasturanja gde su reziduali (greške) prikazani u odnosu na teorijske vrednosti dobijene modelom. Homoskedastičnost grešaka je postignuta ako dijagram rasturanja prikazuje reziduale jednolično za sve teorijske vrednosti Y. Sa druge strane, heteroskedastičnost se detektuje ukoliko dijagram rasturanja prikazuje reziduale „razbacano“ za sve teorijske vrednosti Y. Primeri dijagrama rasturanja u slučaju postojanja homoskedastičnosti i heteroskedastičnosti su dati ispod:



Grafik 3.3. *Primer homoskedastičnosti*



Grafik 3.4. *Primer heteroskedastičnosti*

Od formalnih testova koji se koriste za testiranje homoskedastičnosti su:

- White test i
- Breush-Pagan test

White-ov test predstavlja poseban slučaj Breush-Pagan-ovog testa. Kod ovog testa nulta hipoteza kaže da su sve varijanse grešaka jednake, dok alternativna hipoteza kaže da su varijanse grešaka različite. Dakle:

$$H_0: var(\varepsilon_i) = \sigma^2,$$

$$H_1: var(\varepsilon_i) \neq \sigma^2.$$

Najpoznatiji test homoskedastičnosti je Breush-Pagan test. Ukoliko test pokaže postojanost homoskedastičnosti tada se ocene regresionih parametara dobijene metodom maksimalne verodostojnosti ne bi trebale razlikovati od ocena regresionih parametara dobijenih metodom najmanjih kvadrata. Test polazi od osnovnog modela višestruke linearne regresije:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_n X_{in}, \qquad i = 1,2,\dots,n.$$

Prema Breush –Pagan testu nulta i alternativna hipoteza su:

$$H_0: \sigma_i^2 = \sigma^2$$

$$H_1: \sigma_i^2 = g\left(\theta_0 + \theta_1 X_{i1} + \theta_2 X_{i2} + \cdots + \theta_p X_{ip}\right), \qquad i = 1,2,\dots,p$$

gde je g neprekidna funkcija pri čemu X figuriraju promenljive koje utiču na varijansu greške.

3.2. AUTOKORELACIJA

Kao što je već rečeno, panel podaci predstavljaju kombinaciju uporednih podataka i vremenskih serija. Upravo se pojam autokorelacije vezuje za vremenske serije, a samim tim i za panel analizu. Prisustvo autokorelacije znači narušenost uslova 4. koji je naveden u motivaciji ovog poglavlja – slučajna greška koja se odnosi na jednu opservaciju zavisna je od slučajne greške koja se odnosi na drugu opservaciju. Drugim rečima, prisustvo autokorelacije figurira korelaciju između slučajnih grešaka između perioda. Dakle, pretpostavka 4. iz motivacije ovog poglavlja je:

$$cov(\varepsilon_i, \varepsilon_j) = E\left[\left(\varepsilon_i - E(\varepsilon_i)\right)\left(\varepsilon_j - E(\varepsilon_j)\right)\right] = 0, \qquad \forall i \neq j.$$

Obzirom da je jedna od pretpostavki regresionog modela: *sredine od ε_i i ε_j su jednake nuli,* tada se dobija izraz:

$$cov(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i \varepsilon_j) = 0, \qquad \forall i \neq j.$$

Ukoliko se uzme u obzir pretpostavka normalnosti (slučajna odstupanja imaju normalnu raspodelu: $\varepsilon : N(0, \sigma^2)$) i prethodno razmatrana pretpostavka da je kovarijansa između ε_i i ε_j jednaka nuli, tada se može reći da su slučajna odstupanja ε_i i ε_j nezavisna. Ovakva osobina slučajnih odstupanja se naziva **neautokorelisanost**. Sa druge strane, ukoliko su odstupanja autokorelisana, tada važi:

$$cov(\varepsilon_i, \varepsilon_j) = E(\varepsilon_t \varepsilon_{t-s}) \neq 0, \qquad t > s.$$

Ovako definisan izraz znači da je slučajno odstupanje koje se dogodilo u vremenskom trenutku t povezano sa slučajnim odstupanjem koje se dogodilo u vremenskom trenutku $t - s$.

O svojstavima ocenjivača za α i β koji su dobijeni metodom najmanjih kvadrata, ukoliko se posmatra prosta linearna regresija $Y_t = \alpha + \beta X_t + \varepsilon_t$ u trenutku kada su slučajna odstupanja autoregresiona (svako sledeće odstupanje zavisi od prethodnih odstupanja), govori sledeća teorema.

Teorema 3.1.

1. *Ocenjivači metode najmanjih kvadrata još uvek su linearni i nepristrasni.*
2. *Ocenjivači metode najmanjih kvadrata nisu efikasni u poređenju sa metodama koje uzimaju u obzir autokorelaciju. Drugim rečima, ocenjivači metode najmanjih kvadrata nisu najbolji linearni nepristrasni ocenjivači (BLUE).*
3. *Ocenjene varijanse ocenjivača dobijenih metodom najmanjih kvadrata su pristrasne. U radu sa podacima se dešava da formule za kalkulaciju varijansi i standardnih grešaka ocenjivača dobijenih metodom najmanjih kvadrata daju manje vrednosti od tačnih varijansi i standardnih grešaka.*
4. *Nisu pouzdani uobičajeni t i F test.*
5. *Ocenjena varijansa izračunata pomoću formule*

$$\hat{\sigma}^2 = \frac{SSR}{df}$$

odnosno količnik sume kvadrata reziduala i stepeni slobode predstavlja pristrasan ocenjivač varijanse σ^2.

6. *Dobijeni R^2 izračunat pomoću formule:*

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

predstavlja nepouzdanu meru pravog R^2.

7. *Varijansa i standardna greška koje su izračnate pomoću standardnih formula mogu biti neefikasne.*

S obzirom da je fokus rada na analizi panel podataka, izvođenje dokaza date teoreme neće biti prikazano u ovom radu.

Nakon definisanja autokorelacije, neophodno je utvrditi njenu postojanost u okviru određenog skupa podataka. Najčešći test za detekciju autokorelacije jeste **Durbin – Watson test.** Durbin – Watson test statistika predstavlja statistiku koja se koristi kako bi se identifikovalo prisustvo autokorelacije u slučajnim odstupanjima iz regresione analize. Durbin [6] i Watson[7] su razvili ovu test statistiku koju su primenili na greške najmanjih kvadrata. Dodatno, oni su razvili kritične vrednosti testova za nultu hipotezu: *greške nisu autokorelisane* nasuprot alternativne hipoteze: *greške su autokorelisane i slede autoregresioni proces prvog reda* (greška u momentu t zavisi od greške iz momenta $t-1$). Kako bi se test adekvatno primenio neophodno je izračunati test statistiku d koja je definisana sledećom formulom:

$$d = \frac{\sum_{t=2}^{n}(\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=2}^{n}\varepsilon_t^2},$$

pri čemu ε predstavlja rezidual metode običnih najmanjih kvadrata. Sa druge strane, ukoliko je alternativna hipoteza: *postoji pozitivna autoregresija (drugim rečima jednostrani test$\rho = 1$),* tada važe sledeća pravila:

1. Odbaciti H_0 ako je $d < d_L$,
2. Prihvata se H_0 ako je $d > d_U$,
3. Test je neodređen ako je $d_L < d < d_U$,

gde je:

---

[6]James Durbin (30. jun 1923 – 23. jun 2012) je bio britanski statističar i ekonometrista čiji se rad bazirao na vremenskim serijama i serijskim korelacijama.

[7]Geoffrey Stuart Watson (03. decembar 1921 – 03. januar 1998) je bio australijski statističar koji je statistiku koristio u teoriji kontinentalnog zanosa

- d_L predstavlja donju granicu a
- d_U predstavlja gornju granicu

iz definisane tabele (definiciju dali Durbin i Watson):

Vrednosti $\boldsymbol{\rho}$	**Vrednosti $\boldsymbol{d}$**
$\boldsymbol{\rho = -1}$ **(perfektna negativna korelacija)**	$d = 4$
$\boldsymbol{\rho = 0}$ **(nema korelacije)**	$d = 2$
$\boldsymbol{\rho = 1}$ **(perfektna pozitivna korelacija)**	$d = 0$

Tabela 3.1. *Definisanje vrednosti Durbin-Watson test statistike*

U tabeli ρ predstavlja koeficijent korelacije definisan kao:

$$\rho = \frac{cov(\varepsilon_t, \varepsilon_{t-1})}{\sigma^2}.$$

Ukoliko se pak, alternativna hipoteza odnosi na dvostrani test tj: *postoji autoregresija (drugim rečima $\rho \neq 0$),* tada važe druga pravila:

1. Odbaciti H_0 ako je $d < d_L$ili ako je $d > 4 - d_L$
2. Prihvata se H_0 ako je $d_U < d < 4 - d_U$,
3. Test je neodređen ako je $d_L \leq d \leq d_U$ ili ako je $4 - d_u \leq d \leq 4 - d_L$.

Kako definicije DW test statistike važi:

$$d = \frac{\sum_{t=2}^{n} \varepsilon_t^2}{\sum_{t=1}^{n} \varepsilon_t^2} + \frac{\sum_{t=2}^{n} \varepsilon_{t-1}^2}{\sum_{t=1}^{n} \varepsilon_t^2} - 2\frac{\sum_{t=2}^{n} \varepsilon_t \varepsilon_{t-1}}{\sum_{t=1}^{n} \varepsilon_t^2}$$

tada važi da je $d \approx 2(1 - \hat{\rho})$, pa je:

$$\hat{\rho} = \frac{\sum_{t=2}^{n} \varepsilon_t \varepsilon_{t-1}}{\sum_{t=1}^{n} \varepsilon_t^2}.$$

Dakle, važi da je $0 \leq d \leq 4$ i da je $\hat{\rho} \approx 1 - \frac{d}{2}$.

3.3. MULTIKOLINEARNOST

Kao što je navedeno u motivaciji ovog poglavlja, jedna od osnovnih pretpostavki korišćenja metode najmanjih kvadrata pri ocenjivanju koeficijenata je da nezavisne promenljive nisu međusobno korelisane (pretpostavka 6.). Multikolinearnost predstavlja situaciju u kojoj su dve ili više nezavisnih promenljivih međusobno linearno korelisane. S toga se jedna nezavisna promenljiva tačno može predvideti kao linearna kombinacija drugih nezavisnih promenljvih (npr. bruto društveni proizvod i stopa nezaposlenosti). Takođe, ukoliko se radi o ekstremnom slučaju multikolinearnosti tada se ona naziva perfektna multikolinearnost (npr. jedna nezavisna promenljiva predstavlja starnosno doba osobe izraženo u godinama, dok druga nezavisna promenljvia predstavlja starnosno doba osobe izraženo u danima, te ovakav primer predstavlja perfektnu multikolinearnost jer je prvu promenljivu moguće savršeno predvideti pomoću druge promenljive). Sa druge strane, ukoliko ne postoji linearna veza između nezavisnih promenljvih kaže se da su one ortogonalne jedna na drugu (za detaljnije objašnjenje nepostojanja linearne zavisnosti dve ili više promenljvih i ortogonalnost pogledati *Stojaković Z., Bošnjak I., Linearna algebra*). Dodatno, u slučaju izostanka multikolinearnosti, izostavljanje određene promenljive neće uticati na promenu ocenjenih koeficijenata drugih promenljvih.

Postoje dve vrste multikolinearnosti:

1. strukturna multikolinearnost i
2. multikolinearnost iz podataka.

Strukturna multikolinearnost se javlja kada se u modelu koristi varijabla i njena transformacija (npr. $x_1 = x i x_2 = e^x$– očigledno postoji eksponencijalna povezanost između varijabli x_1 i x_2). Ovakav oblik multikolinearnosti se ređe dešava jer bi trebalo da se vodi računa o input parametrima pri konstrukciji modela. Multikolinearnost iz podataka predstavlja povezanost samih podataka u okviru koji je nemoguća kontrola njihove povezanosti. Treba napomenuti da veza između zavisne promenljive i nezavisne promenljive ne predstavlja multikolinearnost bez obzira na jačinu te linearne veze koja se predstavlja koeficijentom kolinearnosti.

Postoje dva načina identifikacije multikolinernosti. Na osnovu korelacione matrice može se primetiti slaba, srednja ili jaka povezanost između regresora. U nastavku rada je data tabela koja pokazuje jačinu povezanosti između varijabli:

Apsolutna vrednost koeficijenta korelacije	**Jačina povezanosti između promenljvih**
$\boldsymbol{|r| = 1}$	Potpuna (savršena, perfektna) korelacija
$\mathbf{0,8} \leq \boldsymbol{|r|} < 1$	Jaka korelacija
$\mathbf{0,5} \leq \boldsymbol{|r|} < 0{,}8$	Srednje jaka korelacija
$\mathbf{0,2} \leq \boldsymbol{|r|} < 0{,}5$	Relativno slaba korelacija
$\mathbf{0} \leq \boldsymbol{|r|} < 0{,}2$	Neznatna korelacija
$\boldsymbol{|r| = 0}$	Potpuna odsutnost korelacije

Tabela 3.2. *Interpretacija vrednosti koeficijenta korelacije*

Sa druge strane, povezanost dve ili više nezavisnih promenljivih možemo utvrditi kroz **VIF test** (engl. Variance Inflation Factor test) i njegov ekvivalent - **test tolerancije** (engl. Tolerance test). VIF test pokazuje da li je jedna nezavisna promenljiva u snažnoj linearnoj vezi sa ostalim nezavisnim varijablama. Ukoliko je vrednost VIF veća od 10, postoji snažno prisustvo multikolinearnosti. Takođe, tolerancija ispod 0.1 predstavlja problem koji ukazuje na prisustvo multikolinearnosti.

4. MODEL FIKSNIH EFEKATA VS MODEL SLUČAJNIH EFEKATA

Pri sprovođenju empirijskog istraživanja najčešće se postavlja pitanje koji model je pogodniji. Odgovor na ovo pitanje zavisi od različitih pretpostavki koje istraživač uvodi u vezi sa verovatnoćom nastanka korelacije između jedinica posmatranja, ili specifičnosti uporednih podataka, ali i slučajnog odstupanja ε_i.

Ako se pretpostavi da između ε_i i X ne postoji korelacija, model slučajnih efekata je adekvatniji. Sa druge strane ukoliko su ε_i i X korelisani, odgovarajući model je model fiksnih efekata. Sada se postavlja retoričko pitanje zašto očekivati korelaciju između komponente individualne greške i jednog ili više regresora. Posmatrajmo sledeći primer.

Pretpostavimo da imamo uzorak sa velikim brojem individua (jedinice posmatranja) pri čemu se modelira njihova funkcija zarade. Pretpostavimo da je zarada funkcija od obrazovanja, radnog iskustva itd. Dalje, neka ε_i predstavlja urođenu sposobnost individue, pordično okruženje i slično. Tada je, ukoliko se modelira funkcija zarade uključujući i ε_i, vrlo verovatno da će ε_i biti u korelaciji sa obrazovanjem jer često urođena sposobnost i porodično okruženje determinišu obrazovanje određene individue.

Pored ove kratke analize, određena razmatranja i pretpostavke date od strane naučnika Judge, Hill, Griffths, Lutkepolh i Lee upućuju na adekvatan izbor modela.

a.) Ako je T (broj perioda u okviru vremenskih serija datih podataka) veliko i N (jedinica posmatranja u okviru datih podataka) malo, vrlo verovatno je mala razlika u vrednostima parametara dobijenih modelom fiksnih efekata i vrednosti parametara dobijenih modelom slučajnih efekata. Zbog toga je izbor zasnovan na metodi koja je računski pogodnija te se model fiksnih efekata preferira.

b.) Kada je T (broj perioda u okviru vremenskih serija datih podataka) malo i N (jedinica posmatranja u okviru datih podataka) veliko, procene dobijene putem ova dva modela mogu se značajno razlikovati. Kod modela stohastičkih efekata $\beta_{1i} = \beta_1 + \varepsilon_i$ gde je ε_i slučajna komponenta uporednih jedinica, dok model fiksnih efekata posmatra β_{1i} kao fiksan a ne kao slučajan parametar. U slučaju modela fiksnih efekata, statistički zaključak je uslovljen posmatranim uporednim jedinicama u uzorku.Ovakav pristup je adekvatan u slučaju ako se veruje da uporedne jedinice u uzorku nisu slučajno izabrane iz većeg uzorka. U tom slučaju, model fiksnih efekata je odgovarajući. Sa druge strane, ako su uporedne jedninice slučajno odabrane iz većeg uzorka, tada je model slučajnih efekata odgovarajući.

c.) Ako je komponenta individualne greške ε_i u korelaciji sa jednim ili više regresora, tada su ocenjeni parametri modelom slučajnih efekata pristrasni, dok su sa druge strane ocenjeni parametri modelom fiksnih efekata nepristrasni.

d.) Ako je T (broj perioda u okviru vremenskih serija datih podataka) malo i N (jedinica posmatranja u okviru datih podataka) veliko, i ukoliko važe pretpostavke vezane za model slučajnih efekata, ocene dobijene ovim modelom su efikasnije od ocena dobijenih modelom fiksnih efekata.

Pored datih smernica, razvijen je i formalni test od strane Hausmana[8] 1978. godine. Hausmanov test se koristi kako bi se uporedili procenjeni koeficijenti modela sa fiksnim efektom i modela sa slučajnim efektom. Hausman polazi od stajališta da su oba modela konzistentna ako ne postoji korelacija između ε_i i nezavisnih promenljvih X_i. Dakle, kada se utvrdi postojanost individualnih efekata, neophodno je utvrditi prirodu istih (fiksni ili slučajni). Hausmanov test se definiše na sledeći način:

$$H_0: cov(\varepsilon_i, X_{k,it}) = 0, \qquad \forall k = 1, \dots, K$$

$$H_1: cov(\varepsilon_i, X_{k,it}) \neq 0, za\ bar\ jedno\ k$$

Vrednost Hausmanove test statistihe H se računa na osnovu sledeće formule:

$$H = \left(\widehat{\beta_{RE}} - \widehat{\beta_{FE}}\right)^T \left[var\left(\widehat{\beta_{RE}}\right) - var\left(\widehat{\beta_{FE}}\right)\right]^{-1} \left(\widehat{\beta_{RE}} - \widehat{\beta_{FE}}\right).$$

Ako je vrednost H test statistike manja od $X^2(K)$ (K broj procenjenih parametara modela) tada se H_0 ne odbacuje i zaključuje se da je model slučajnih efekata adekvatniji od modela fisknih efekata. Drugim rečima, ukoliko je p −vrednost rezultata ($Prob > chi2$ dobijen kroz Statu) manja od 0,05 tada se koristi model fiksnih efekata, u suprotnom se koristi model slučajnih efekata.

[8] Jerry A. Hausman (rođen 05. maja 1946) je profesor ekonomije na Tehnološkom institutu u Masačusetstu

5. PRIMENA PANEL REGRESIONE ANALIZE – EMPIRIJSKA ANALIZA I RAZULTATI

U ovom poglavlju dat je prikaz praktične primene panel regresione analize na demografske pokazatelje. Poglavlje se sastoji iz dva dela. Prvi deo poglavlja čini motivacija u okviru koje je dato objašnjenje demografije kao pojma, njenih pokazatelja kao i odabira regresora koji će se kasnije koristiti u konstrukciji modela. Nakon motivacije, dat je detaljan pregled različitih panel modela u kojima će se koristiti prethodno odabrani regresori. Na samom kraju poglavlja dat je detaljan pregled odabira finalnog modela, kao i njegovih karakteristika. Svi podaci za nezavisne promenljive kao i zavisnu promenljivu preuzeti su sa zvaničnog sajta Eurostat[9]. Dodatno, za kreiranje modela korišćen je softverski paket STATA.

MOTIVACIJA

„Demografija je matematičko znanje o
stanovništvu, njegovom kretanju,
fizičkom, građanskom, intelektialnom i moralnom stanju“

Guillard, XIX vek

Demografija se vrlo intenzivno razvija poslednjih decenija, i postaje sve značajnija kada se radi o drugim oblastima istraživanja, u globalnim okvirima, posebno kada je reč o promenama i projekcijama naseljenosti planete. Demografski pokazatelji poput broja novorođenih,prirodnog priraštaja, broja umrlih i slično imaju veliki uticaj na ekonomski, tehnološki, naučni i svaki drugi vid razvoja zemlje.

Demografija predstavlja nauku koja proučava stanovništvo. Etimologija reči „*demografija*“ jesu dve grčke reči ***demos*** – narod i ***graphein*** – opisivati. Temelje demografije postavio je Guillard[10] u svom delu „*Elementi humane statistike ili komparativna demografija“.* Usko povezan pojam za demografiju predstavlja demografski razvitak. Demografski razvitak, kao jedan kompleksan proces, čine sa jedne strane prirodno i mehaničko kretanje stanovništva:

[9] Eurostat je kancelarija Evropske Unije zadužena za statistiku čiji je glavni zadatak obrada i objavljivanje uporednih statističkih podataka na nivou cele Evropske Unije.

[10]Achille Gullard (1799 - 1876) je bio poznati francuski statističar, botaničar i demograf.

1. natalitet,
2. ferilitet,
3. smrtnost i
4. migracije.

dok sa druge strane demografski razvitak čine promene u demografskim strukturama:

1. biološkim,
2. socio – ekonomskim i
3. intelektualnim.

Dakle predmet demografije predstavlja stanovništvo koje živi na određenoj teritoriji, a koje se menja u kontekstu različitih obeležja koja su svojstvena za određeno istorijsko razdoblje. Pored izučavanja demografskog razvitka, kao i faktora i uzroka koji taj razvitak uslovljavaju, demografija se bavi projekcijama i prognozama stanovništva. Date projekcije i prognoze su veoma značajne za ekonomsku i svaku drugu nauku koja se bavi planiranjem i predikcijama. Demografsko projektovanje ima kapitalnu ulogu u planiranju i kreiranju politika u različitim oblastima jedne zemlje poput prosvete, zdravstva, privrede, ali i osiguranja i socijalne zaštite. U tom kontekstu, shvatajući važnost demografije, u narednom delu rada biće dat detaljan pregled različitih modela kako bi se projektovala određena promena u populaciji pojedinih zemalja.

5.1. EMPIRIJSKA ANALIZA I RAZULTATI

5.1.1. Osnovne informacije o skupu podataka

Kao što je rečeno u uvodnom delu ovog poglavlja, skup podataka preuzet je sa zvaničnog sajta evropske kuće za statistiku – Eurostat.

Za ***zavisnu promenljivu*** odabrana je sirova ***stopa promene populacije***. Ukoliko stopa ima negativan predznak tada je u toku jedne godine došlo je do pada populacije, dok pozitivna stopa predstavlja rast populacije. Zavisna promenljiva (sirova stopa promene populacije) predstavlja racio promene populacije tokom godine (razlika između veličine populacije na 1. januar dve uzastopne godine) i prosečne populacije u toj godini.

Za ***nezavisne promenljive*** odabrane su:

1. ***stopa razvoda*** **–**stopa razvoda predstavlja odnos broja razvedenih brakova tokom godine i prosečne populacije tokom te godine. Vrednost ove nezavisne promenljive se izražava na 1000 stanovnika.
2. ***stopa brakova*** **–** stopa brakova predstavlja odnos broja sklopljenih brakova tokom godine i prosečne populacije tokom te godine Vrednost ove nezavisne promenljive se izražava na 1000 stanovnika.
3. ***stopa nezaposlenosti*** **–** ovaj pokazatelj se fokusira na starnosno doba od 25 do 64 godine. Stopa nezaposlenosti prikazuje „verovatnoću“ da osoba ostane bez zaposlenja. Dodatno, pokazatelj pruža kvantifikaciju poteškoće na tržištu rada za ljude sa različitim nivoom obrazovanja i nudi prvu ideju o uticaju obrazovanja u smanjenju nezaposlenosti. Nivo obrazovanja je definisan u skladu sa Međunarodnom standardu za klasifikaciju obrazovanja (engl. ISCED).
4. ***stopa rasta bruto društvenog proizvoda*****–**bruto društveni proizvod predstavlja meru ekonomske aktivnosti države, definisan kao vrednost svih dobara ili usluga koje učestvuju u njegovom kreiranju. Obračun godišnje stope rasta obima bruto društvenog proizvoda ima za cilj da omogući poređenje dinamike ekonomskog razvoja kako tokom vremena, tako i između ekonomija različitih veličina.
5. ***stopa smrtnosti*** **–** smrtnost predstavlja trajni nestanak svih dokaza o životu u bilo kom trenutku nakon rođenja (postnatalni prestanak vitalnih funkcija bez mogućnosti reanimacije). Stopa smrtnosti predstavlja racio tj. odnos broja umrlih tokom godine i prosečnog broja stanovnika te godine. Vrednost ove stope se izražava na 1000 stanovnika.

Radi jednostavnost, na dalje će se uvesti sledeće oznake za date promenljive:

Promenljiva	Oznaka	Tip promenljive
Stopa promene populacije	POPUL	Zavisna
Stopa razvoda	RAZV	Nezavisna
Stopa brakova	BRAK	Nezavisna
Stopa nezaposlenosti	NEZAP	Nezavisna
Stopa rasta bruto društvenog proizvoda	BDP	Nezavisna
Stopa smrtnosti	SMR	Nezavisna

Tabela 5.1. *Pregled promenljivih koje učestvuju u regresionoj panel analizi*

Nakon definisanja zavisne i nezavisnih promenljivih, neophodno je definisati jedinice posmatranja (individue) kao i vremensku seriju u okviru koje će se prethodno pomenute promenljive razmatrati.

Za jedinice posmatranja uzete su najjače ekonomije (države) Evropske Unije. Njih čine četiri zemlje i to su:

1. Nemačka
2. Francuska
3. Italija i
4. Španija.

Na dalje, vremenska serija koja je uzeta u razmatranje jeste period od 2008. godine do „najsvežije“ 2019. godine. Razlog izostavljanja Srbije u okviru empirijske analize jeste nedostatak zvaničnih podataka na sajtu evropske kuće za statistiku – Eurostat.

Kada su definisani osnovni subjekti skupa podataka koji će se koristiti u kreiranju modela, neophodno je da se ispitaju osnovne statistike (deskriptivna statistika) pomenutih promenljivih. Da bismo proverili deskriptivnu statistiku neophodno je da se proveri da li se u datom skupu podataka nalaze balansirani ili nebalansirani podaci. Kao rezultat ispitivanja balansiranosti datog skupa podataka javlja se:.

```
xtset Drzava_num Godina, yearly
       panel variable:  Drzava_num (strongly balanced)
        time variable:  Godina, 2008 to 2019
                delta:  1 year
```

Dakle, zaključak ispitivanja balansiranosti podataka je: dati skup podataka je strogo balansiran, što drugim rečima znači da za svaku godinu imamo „punu“ vremensku seriju (vremenska serija gde za svaku godinu postoji podatak tj. nema nedostajućih podataka). Sada se može ispitati deskriptivna statistika. U tabeli ispod dat je pregled osnovnih statistika, kako skupa podatak u celini (engl. *overall*), tako i po državama (engl.*between*) i po godinama (engl. *within*).

Variable		Mean	Std. Dev.	Min	Max	Observations	
POPUL	overall	2.845833	4.163824	-4.6	18.2	N =	48
	between		.6151821	2.225	3.641667	n =	4
	within		4.128868	-4.7375	18.82083	T =	12
RAZV	overall	1.8375	.4465589	.9	2.4	N =	48
	between		.4648028	1.15	2.133333	n =	4
	within		.1842177	1.5875	2.2875	T =	12
BRAK	overall	3.866667	.5846742	3.1	5	N =	48
	between		.6133167	3.433333	4.775	n =	4
	within		.2317831	3.533333	4.633333	T =	12
NEZAP	overall	15.62708	6.491557	7.4	32.7	N =	48
	between		6.201562	12.08333	24.89167	n =	4
	within		3.56119	3.935417	23.43542	T =	12
BDP	overall	.6166667	2.172687	-5.7	4.2	N =	48
	between		.6662499	-.3166667	1.216667	n =	4
	within		2.092981	-6.3	3.783333	T =	12
SMR	overall	9.61875	1.040388	8.2	11.5	N =	48
	between		1.128469	8.616667	10.91667	n =	4
	within		.327114	9.102083	10.20208	T =	12

Tabela 5.2. *Deskriptivna statistika*

Dakle, iz tabele se može uočiti da je deskriptivna statistika prikazana na celom skupu podataka (48 opservacija), po državama ali i po godinama.

5.1.2. Konstrukcija modela

1. OLS model

Osnovni model sa kojim će biti započeta regresiona analiza jeste standardni OLS model. Kao što je rečeno, OLS model predstavlja višestruki regresioni model gde su parametri modela ocenjeni preko modela najmanjih kvadrata. Rezutalti regresione analize dobijeni pomoću ove metode su:

Source	SS	df	MS
Model	226.574859	5	45.3149718
Residual	588.284308	42	14.0067692
Total	814.859167	47	17.3374291

Number of obs	= 48
F(5, 42)	= 3.24
Prob > F	= 0.0147
R-squared	= 0.2781
Adj R-squared	= 0.1921
Root MSE	= 3.7426

POPUL	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
RAZV	-3.871558	2.283412	-1.70	0.097	-8.479669	.7365532
BRAK	4.195715	1.852964	2.26	0.029	.4562825	7.935148
NEZAP	-.1718301	.1251311	-1.37	0.177	-.424355	.0806947
BDP	.1455294	.2763589	0.53	0.601	-.4121854	.7032441
SMR	-2.824892	.8550675	-3.30	0.002	-4.550488	-1.099296
_cons	23.50378	7.505721	3.13	0.003	8.356625	38.65094

Tabela 5.3. *OLS model – prvi model*

Dobijeni rezultati datog modela upućuju da je $R^2 = 0{,}2781$. R^2 predstavlja relativnu meru fitovanja podataka, pri čemu predstavlja proporciju objašnjene varijanse. Obzirom da je dobijeni $R^2 = 0{,}2781$ može se reći da se približno 28% varijanse stope promene populacije može objasniti preko datih nezavisnih varijabli (regresora).

Sa druge strane F test dat u tabeli iznad, testira da li je regresioni model dobar za ove vrednosti. Obzirom da je $F(5{,}42) = 3{,}24, p < 0{,}05$ može se zaključiti da je regresioni model dobar.

Nakon analiziranja modela u globalu, neophodno je izanalizirati model po promenljivima. Iz date tabele se vidi da stopa razvoda, stopa nezaposlenosti i stopa smrtnosti imaju negativan predznak, što je i očekivano. Dakle, porast ili smanjenje ovih pokazatelja utiče na smanjenje ili porast stope populacije, respektivno. Sa druge strane, stopa brakova i stopa rasta bruto društvenog proizvoda imaju pozitivan predznak, te se može reći da sa rastom (padom) ova dva regresora dolazi do rasta(pada) zavisne promenljive. Na dalje, neophodno je utvrditi statističku značajnost nezavisnih promenljivih. Testom se može proveriti koliko je značajna svaka nezavisna promenljiva u prikazanom modelu. Ako je $p < 0{,}05$ zaključuje se da je koeficijent statistički značajno različit od nule, tj. odgovarajuća promenljiva je neophodna u istraživanju. Posmatrajući kolonu $P > |t|$ koja predstavlja rezultate testa, može se zaključiti da su nezavisne promenljive stopa razvoda, stopa nezaposlenosti i stopa rasta bruto društvenog proizvoda nisu statistički značajne promenljive obzirom da je njihova vrednost iz kolone $P > |t|$ veća od 0,05. Isti zaključak se mogao izvesti sa 95% intervalom poverenja. Obzirom da koeficijenti pomenutih promenljivih imaju intervale poverenja koje uključuju nulu, može se reći da isti nisu statistički značajni, te je neophodno iste izbaciti iz modela.

Sada, je neophodno kreirati novi OLS model koji će obuhvatati samo statistički značajne promenljive iz prethodnog modela. Obzirom da je vrednost $P > |t|$ iznosi 0,097 (veoma blizu 0,05 te ćemo pokušati da je uključimo u model), model ćemo kreirati izbacivanjem promenljivih stopa rasta bruto društvenog proizvoda i stope nezaposlenosti. Dobijeni rezultati su:

Source	SS	df	MS
Model	191.511098	3	63.8370327
Residual	623.348068	44	14.1670016
Total	814.859167	47	17.3374291

Number of obs	=	48
F(3, 44)	=	4.51
Prob > F	=	0.0077
R-squared	=	0.2350
Adj R-squared	=	0.1829
Root MSE	=	3.7639

POPUL	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
RAZV	-5.199301	1.729311	-3.01	0.004	-8.684497	-1.714104
BRAK	5.327652	1.540921	3.46	0.001	2.22213	8.433173
SMR	-2.722815	.8256379	-3.30	0.002	-4.386779	-1.058851
_cons	17.98937	6.670512	2.70	0.010	4.545837	31.4329

Tabela 5.4. *OLS model – drugi model*

Iz prikazanih rezultata ($F(3{,}44) = 4{,}51 > 0{,}01$) može se reći da je model dobar, i da smo izbacivanjem dva regresora poboljšali statističku značajnost promenljive stopa razvoda. Takođe, iz date tabele može se zaključiti da je $R^2 = 0{,}2350$ i da je opao u odnosu na inicijalni OLS model (što je i očekivano jer se sa povećanjem promenljivih povećava i R^2) pri čemu je prilagođeni R^2 veoma nizak. Uzimajući u obzir niske performanse ovog modela ali i isključivanje individualnih i vremenskih efekata iz panel serije podataka, ovaj model neće biti razmatran kao kandidat za finalni.

Nakon predstavljanja klasične (višestruke) regresione analize na dalje ćemo prikazati panel regresionu analizu u okviru koje će se posmatrati modeli fiksnih i slučajnih efekata.

2. MODEL FIKSNIH EFEKATA

2.a Nagibi koeficijenata su konstantni ali odsečak varira kroz posmatrane jedinice (države)

Kako bi se razmotrila ova varijacija modela fiksnih efekata, jedan od načina da se uračuna individualnost svake države tj. svake jedinice posmatranja je da se dopusti da odsečak varira za svaku državu pretpostavljajući da su nagibi koeficijenata nepromenjeni. Kako bi se inkorporirao individualni efekat neophodno je uvesti veštačke promenljive. Jednačina koja predstavlja ovu vrstu modela je:

$$POPUL = \alpha_1 + \alpha_2 \cdot D_{2i} + \alpha_3 \cdot D_{3i} + \alpha_4 \cdot D_{4i} + \beta_2 \cdot RAZV + \beta_3 \cdot BRAK + \beta_4 \cdot NEZAP + \beta_5 \cdot BDP + \beta_6 \cdot SMR + u_{it}.$$

Obzirom da postoje četiri države, u modelu se koriste tri veštačke promenljive kako bi se izbegao problem savršene kolinearnosti. Svaka veštačka promenljiva ima sledeći oblik:

$$D_2 = \begin{cases} 1, država = Nemačka \\ 0, ostalo \end{cases},$$

$$D_3 = \begin{cases} 1, država = Italija \\ 0, ostalo \end{cases},$$

$$D_4 = \begin{cases} 1, država = Španija \\ 0, ostalo \end{cases}.$$

Kreiranjem modela sa datim nezavisnim promenljivima uz dodatak tri veštačke promenljive (kako bi se obuhvatio efekat posmatranih jedinica) dobija se sledeći rezultat:

Source	SS	df	MS
Model	351.4892	8	43.93615
Residual	463.369967	39	11.8812812
Total	814.859167	47	17.3374291

Number of obs	=	48
F(8, 39)	=	3.70
Prob > F	=	0.0027
R-squared	=	0.4313
Adj R-squared	=	0.3147
Root MSE	=	3.4469

POPUL	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
RAZV	-5.851115	2.87696	-2.03	0.049	-11.67032	-.0319143
BRAK	-1.283367	3.296446	-0.39	0.699	-7.951058	5.384325
NEZAP	-.6863823	.2133543	-3.22	0.003	-1.117932	-.2548326
BDP	-.0012323	.2670687	-0.00	0.996	-.5414297	.538965
SMR	-1.871311	1.730557	-1.08	0.286	-5.371693	1.629071
D2	4.511316	5.919755	0.76	0.451	-7.462517	16.48515
D3	-4.539385	3.551875	-1.28	0.209	-11.72373	2.644959
D4	7.989825	2.651262	3.01	0.005	2.627143	13.35251
_cons	45.29575	24.41861	1.85	0.071	-4.095553	94.68706

Tabela 5.5. *Model fiksnih efekata– nagibi koeficijenata su konstantni ali odsečak varira kroz posmatrane jedinice (države) – prvi model*

Dobijeni rezultati upućuju na sledeće zaključke:

- Obzirom da je $F(8,39) = 3{,}70$ može se reći da je model dobar,
- $R^2 = 0{,}4313$ pa se može zaključiti da je aproksimativno 43% varijanse stope promene populacije može objasniti preko datih nezavisnih promenljivih (regresora),
- Skok R^2 u ovom modelu je očekivan jer je došlo do povećanja broja ulaznih promenljivih,
- Od datih regresora, jedino promenljive *stopa razvoda* i *stopa nezaposlenosti* imaju koeficijent koji statistički značajno različit od nule (veštačke promenljive ne ulaze u analizu statističke značajnosti)

Na osnovu dobijenih rezultata, naredni korak jeste izbacivanje promenljiviih koje nisu statistički značajne (*stopa sklopljenih brakova, stopa rasta bruto društvenog proizvoda i stopa smrtnosti*). Odstranjivanjem pomenutih promenljivih dobijeni su sledeći rezultati:

Source	SS	df	MS
Model	335.486521	5	67.0973042
Residual	479.372646	42	11.4136344
Total	814.859167	47	17.3374291

Number of obs =	48
F(5, 42) =	5.88
Prob > F =	0.0003
R-squared =	0.4117
Adj R-squared =	0.3417
Root MSE =	3.3784

POPUL	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
RAZV	-5.531125	2.751215	-2.01	0.051	-11.0833	.021053
NEZAP	-.613281	.1423183	-4.31	0.000	-.900491	-.326071
D2	-.9634951	1.458933	-0.66	0.513	-3.907741	1.980751
D3	-6.64915	2.587579	-2.57	0.014	-11.8711	-1.427204
D4	7.403231	2.115061	3.50	0.001	3.134864	11.6716
_cons	22.64542	5.366631	4.22	0.000	11.81512	33.47572

Tabela 5.6. *Model fiksnih efekata– nagibi koeficijenata su konstantni ali odsečak varira kroz posmatrane jedinice (države) –drugi model*

Na osnovu datih rezultata može se konstatovati da je model dobar jer je $F(5,42) = 5.88 > 0{,}05$. Dobijeni $R^2 = 0{,}4117$ se neznatno smanjio što je i očekivano obzirom da se smanjio ulazni broj promenljivih. Sa druge strane, primetimo da se prilagođeni R^2 povećao (prilagođeni R^2 predstavlja meru uspešnosti fitovanja podataka određenog modela, ali na njegovo povećanje ne utiče uključivanje većeg broja promenljivih kao što je to slučaj kod standardnog R^2). Dakle, iz modela su isključene sve promenljive koje nisu statistički značajne i dobijen je model čije su performanse bolje.

2.b Nagibi koeficijenata su konstantni ali odsečak varira kroz vreme i kroz pojedinačno posmatrane jedinice

Ovaj tip modela fiksinh efekata panel podataka predstavlja proširenje u odnosu na prethodni model. Karakteriše ga to što je dozvoljeno variranje odsečka kroz jedinice posmatranja, što je prikazano u prethodnom delu, ali i variranje odsečka kroz vreme. Dakle, uvodi se vremenska komponenta odsečka. Razlika u odnosu na prethodnu metodu je uvođenje dodatnih veštačkih promenljivih koje pokrivaju vremensku komponentu odsečka. Dodatne veštačke promenljive koje se odnose na vreme kreiraju se na sledeći način:

$$T_{2009} = \begin{cases} 1, & godina = 2009 \\ 0, & inače \end{cases}$$

$$T_{2010} = \begin{cases} 1, & godina = 2010 \\ 0, & inače \end{cases}$$

$$\dots$$

$$T_{2019} = \begin{cases} 1, & godina = 2019 \\ 0, & inače \end{cases}$$

Konstrukcija modela u kome su obuhvaćeni kako individualni efekti tako i vremenski efekti dobija se uvrštavanjem prethodno definisanih promenljivih. Rezultati modela sa ovakvim pretpostavkama primenjen na dati skup podataka su:

Source	SS	df	MS		
Model	483.675533	18	26.8708629	Number of obs =	48
Residual	331.183634	29	11.4201253	F(18, 29) =	2.35
				Prob > F =	0.0194
Total	814.859167	47	17.3374291	R-squared =	0.5936
				Adj R-squared =	0.3413
				Root MSE =	3.3794

POPUL	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
RAZV	-2.093553	3.299791	-0.63	0.531	-8.842382	4.655277
BRAK	-1.14538	5.186545	-0.22	0.827	-11.75306	9.462294
NEZAP	-1.001555	.2574578	-3.89	0.001	-1.528116	-.4749951
BDP	-.2941806	.5482806	-0.54	0.596	-1.41554	.8271791
SMR	-.3486592	2.688833	-0.13	0.898	-5.84794	5.150621
D3	-4.573937	6.778139	-0.67	0.505	-18.43679	9.288914
D4	10.97719	3.156674	3.48	0.002	4.521064	17.43331
T2009	-1.506193	3.562295	-0.42	0.676	-8.791904	5.779519
T2010	1.348971	2.911424	0.46	0.647	-4.605561	7.303502
T2011	.5593029	3.485389	0.16	0.874	-6.569119	7.687725
T2012	1.967874	3.424853	0.57	0.570	-5.036737	8.972486
T2013	5.733033	4.020404	1.43	0.165	-2.489616	13.95568
T2014	3.034072	3.587678	0.85	0.405	-4.303553	10.3717
T2015	3.82215	4.427534	0.86	0.395	-5.233174	12.87747
T2016	1.74479	3.631358	0.48	0.634	-5.682171	9.171751
T2017	.4056257	4.744762	0.09	0.932	-9.298502	10.10975
T2018	-.7914265	5.011738	-0.16	0.876	-11.04158	9.458728
T2019	-1.526184	4.789227	-0.32	0.752	-11.32125	8.268885
_cons	27.47453	8.151622	3.37	0.002	10.80259	44.14647

Tabela 5.7. *Model fiksnih efekata– nagibi koeficijenata su konstantni ali odsečak varira kroz vreme i kroz pojedinačno posmatrane jedinice*

Obzirom da je $F(18,29) = 2{,}35$ može se konstatovati da je model dobar. Dodatno, $R^2 = 0{,}5936$ se povećao u odnosu na sve prethodne modele. Međutim, očekivan je skok veličine R^2 obzirom na uvođenje dodatnih 11 veštačkih promenljivih. Sa druge strane, može se uočiti da je prilagođeni R^2 ostao približno isti (neznatno se smanjio u ovom modelu), pa se može zaključiti da se ovaj model neće razmatrati u kontekstu odabira finalnog modela. Dodatno, od nezavisnih promenljivih samo se koeficijent *stope nezaposlenosti* pokazao kao statistički značajno različit od nule.

Posmatrajući sve performanse modela koji se odnose na fiksne efekte, najpogodniji model je prvi u okviru kog su uključeni samo individualni efekti, pri čemu promenljive *stopa razvoda* i *stopa nezaposlenosti* imaju koeficijent koji statistički značajno različit od nule dok preostale nezavisne promenljive nemaju koeficijente sa takvom karakteristikom.

3. MODEL SLUČAJNIH (STOHASTIČKIH) EFEKATA

Pored modela fiksnih efekata (koji mogu biti ponekad i „skupi“ u smislu broja stepeni slobode uvođenjem dodatnih, veštačkih promenljivih) data je analiza i modela stohastičkih efekata. Kao što je u teorijskom delu rada napomenuto, model ima slučajni efekat jer je odsečak individalnih jedinica posmatranja stohastička veličina. Drugim rečima, on se sastoji od determinističke veličine i slučajnog odstupanja koji celo parametar čini stohastičkim (slučajnim).

U ovom slučaju jednačina modela koja se ispituje ima oblik:

$$\begin{aligned} POPUL &= \beta_{1i} + \beta_2 RAZV + \beta_3 BRAK + \beta_3 NEZAP + \beta_3 BDP + \beta_3 SMR + +u_{it} \\ &= \beta_1 + \varepsilon_i + \beta_2 RAZV + \beta_3 BRAK + \beta_3 NEZAP + \beta_3 BDP + \beta_3 SMR + u_{it} \\ &= \beta_1 + \beta_2 RAZV + \beta_3 BRAK + \beta_3 NEZAP + \beta_3 BDP + \beta_3 SMR + w_{it} \end{aligned}$$

Nakon teorijske podloge, puštanjem panel regresionog modela slučajnih efekata dobijeni su sledeći rezultati:

```
Random-effects GLS regression                   Number of obs      =        48
Group variable: Drzava_num                      Number of groups   =         4

R-sq:  within  = 0.3054                         Obs per group: min =        12
       between = 0.4631                                        avg =      12.0
       overall = 0.2781                                        max =        12

                                                Wald chi2(5)       =     16.18
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0064
```

POPUL	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
RAZV	-3.871558	2.283412	-1.70	0.090	-8.346962	.6038465
BRAK	4.195715	1.852964	2.26	0.024	.5639725	7.827458
NEZAP	-.1718301	.1251311	-1.37	0.170	-.4170827	.0734224
BDP	.1455294	.2763589	0.53	0.598	-.396124	.6871828
SMR	-2.824892	.8550675	-3.30	0.001	-4.500794	-1.148991
_cons	23.50378	7.505721	3.13	0.002	8.79284	38.21473
sigma_u	0					
sigma_e	3.4469234					
rho	0	(fraction of variance due to u_i)				

Tabela 5.8. *Model slučajnih (stohastičkih) efekata – prvi model*

Na osnovu prikazanih rezultata mogu se izvesti sledeći zaključci:

- Obzirom da je $Wald\ chi2(5) = 16{,}18$ model je dobar.
- $R^2 = 0{,}2781$ pa se može zaključiti da je aproksimativno 28% varijanse stope promene populacije može objasniti preko datih nezavisnih promenljivih.
- Pored R^2 koji se odnosi na ceo skup podataka, postoje i:
  1. $R^2 between$ – objašnjava varijacije između država,
  2. $R^2 between$ – objašnjava varijacije između između godina.
- Posmatranjem pojedinačnih koeficijenata i njigove statističke značajnosti, promenljive *Stopa razvoda brakova, Stopa nezaposlenosti, stopa rasta bruto društvenog proizvoda* nisu statistički značajne, dok promenljive *Stopa sklopljenih brakova* i *Stopa smrtnosti* jesu statistički značajne (na nivou značajnosti od 95%).

Na osnovu date analize zaključujemo da je neophodno izostaviti promenljive koje nisu statistički značajne. Međutim, obzirom da promenljiva *Stopa razvoda brakova* je vrlo blizu da postane statistički značajna na nivou poverenja od 95%, ovu promenljivu ćemo zadržati u modelu. Sa druuge strane, promenljive *stopa nezaposlenosti, stopa rasta bruto društvenog proizvoda* ćemo izostaviti iz modela.

Novi model slučajnih efekata u kome figuriraju promenljive *stopa razvoda brakova, stopa sklopljenih brakova* i *stopa smrtnosti* ima sledeće performanse:

```
Random-effects GLS regression                   Number of obs      =        48
Group variable: Drzava_num                      Number of groups   =         4

R-sq:  within  = 0.2265                         Obs per group: min =        12
       between = 0.7409                                        avg =      12.0
       overall = 0.2350                                        max =        12

                                                Wald chi2(3)       =     13.52
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0036
```

POPUL	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
RAZV	-5.199301	1.729311	-3.01	0.003	-8.588687	-1.809914
BRAK	5.327652	1.540921	3.46	0.001	2.307503	8.347801
SMR	-2.722815	.8256379	-3.30	0.001	-4.341035	-1.104594
_cons	17.98937	6.670512	2.70	0.007	4.915407	31.06333
sigma_u	0					
sigma_e	3.787614					
rho	0	(fraction of variance due to u_i)				

Tabela 5.9. *Model slučajnih (stohastičkih) efekata – drugi model*

Iz performansi ($Wald\ chi2(3) = 13.52$) se može zaključiti da je model dobar. Dobijeni $R^2 = 0{,}2265$ na nivou celog modela tj. skupa podataka je nešto niži u odnosu na prethodni model. Sa druge strane, primetan je skok R^2 *between* koji se objašnjava varijacije između država. Dodatno, svi koeficijenti nezavisnih promenljivih jesu statistički značajno različiti od nule na nivou poverenja od 95%. Posmatrajući same koeficijente, primetno je da koeficijent promenljive *stopa razvoda brakova* i koeficijent *stopa smrtnosti* imaju negativan predznak (što je i očekivano). Dakle, ukoliko dođe do povećanja jedne od prethodno pomenutih promenljivih, zavisna promenljiva *stopa promene populacije* će se smanjiti, i obrnuto. Promenljiva *stopa sklopljenih brakova* ima pozitivan predznak pri čemu je i ovaj statistički rezultat bio očekivan. Pozitivan predznak ove promenljive implicira da sa povećanjem *stope sklopljenih brakova* dolazi do povećanja *stope promene populacije.*

Poredeći dva modela slučajnih efekata može se zaključiti da je drugi model bolji, jer iako je R^2 veći u prvom modelu (na veći R^2 utiče i veći broj promenljivih koje ulaze u model), drugi model sadrži samo statistički značajne promenljive.

5.1.3. Odabir finalnog modela – model fiksnih vs model slučajnih efekata

Nakon odabira finalnih modela u okviru sekcije fiksnih efekata i u okviru sekcije slučajnih efekata neophodno je doneti konačnu odluku koji model od data dva pristupa je finalni. Za ovu odluku koristi se Hausmanov test (detaljno objašnjen u teorijskom delu rada).

Korišćenjem Hausmanovog testa za poređenje modela dobijeni su sledeći rezultati:

	Coefficients (b) fiksni	(B) slucajni	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
RAZV	-5.531125	-5.199301	-.3318239	2.139783

```
                 b = consistent under Ho and Ha; obtained from regress
      B = inconsistent under Ha, efficient under Ho; obtained from xtreg

   Test:  Ho:  difference in coefficients not systematic

                 chi2(1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                         =        0.02
               Prob>chi2 =      0.8768
```

Tabela 5.10. *Hausmanov test*

Ukoliko je vrednost $Prob > chi2$ manja od 0,05 odluka se donosi u korist modela fiksnih efekata, dok sa druge strane ukoliko je vrednost $Prob > chi2$ veća od 0,05 odluka se donosi u korist modela slučajnih efekata.

Obzirom da je $Prob > chi2 = 0{,}8768$ veće od 0,05, konačni model na osnovu datog skupa podataka jeste model stohastičkih efekata u kome učestvuju promenljive:

- *stopa razvoda,*
- *stopa sklopljenih brakova* i
- *stopa smrtnosti.*

Dobijeni finalni model ima sledeće performanse:

```
Random-effects GLS regression                   Number of obs      =         48
Group variable: Drzava_num                      Number of groups   =          4

R-sq:  within  = 0.2265                         Obs per group: min =         12
       between = 0.7409                                        avg =       12.0
       overall = 0.2350                                        max =         12

                                                Wald chi2(3)       =      13.52
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =     0.0036
```

POPUL	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
RAZV	-5.199301	1.729311	-3.01	0.003	-8.588687	-1.809914
BRAK	5.327652	1.540921	3.46	0.001	2.307503	8.347801
SMR	-2.722815	.8256379	-3.30	0.001	-4.341035	-1.104594
_cons	17.98937	6.670512	2.70	0.007	4.915407	31.06333
sigma_u	0					
sigma_e	3.787614					
rho	0	(fraction of variance due to u_i)				

Tabela 5.11. *Performanse finalnog modela*

Jednačina finalnog modela ima oblik:

$$POPUL \quad = 17{,}98937 - 5{,}199301 \cdot RAZV + 5{,}327652 \cdot BRAK - 2{,}722815 \cdot SMR$$

Dobijena jednačina se može interpretirati na sledeći način:

1. ukoliko se promenljiva *stopa razvoda* poveća za jednu jedinicu, a svi osali parametri ostanu nepromenjeni, *stopa promene populacije* će se smanjiti za 5,199301. Drugim rečima, populacija u državi će se smanjiti za 5,19%;
2. ukoliko se promenljiva *stopa sklopljenih brakova* poveća za jednu jedinicu, a svi osali parametri ostanu nepromenjeni, *stopa promene populacije* će se povećati za 5,327652. Drugim rečima, populacija u državi će se povećati za 5,32%;
3. ukoliko se promenljiva *stopa smrtnosti* poveća za jednu jedinicu, a svi osali parametri ostanu nepromenjeni, *stopa promene populacije* će se smanjiti za 2,722815. Drugim rečima, populacija u državi će se smanjiti za 2,72%.

ZAKLJUČAK

Demografija i demografski razvitak jedne zemlje predstavlja izuzetno kompleksan proces. Njega čine kako prirodno tako i mehaničko kretanje stanovništva. Pored demografskog razvitka, demografija prati faktore i uzroke koji demografski razvitak uslovljavaju. Sa druge strane, veoma važan zadatak demografije kao naučne discipline jeste planiranje i obezbeđivanje određenih predikcija u pogledu nataliteta, mortaliteta, feriliteta ali i migracija. Upravo ove predikcije se obračunavaju kako bi jedna država izgradila strategiju i politike kako u oblasti zdravstva, školstva, privrede tako i u pogledu socijalne zaštite. U ovom radu je dat pregled jednog pristupa rešavanju i obračunavanju predikcija određenih zemalja (koje su vodeće ekonomije Evropske Unije) u pogledu promene stope populacije.

Rad je baziran na pristupu demografskoj strukturnoj promeni kroz panel podatke. Panel podaci, koji su korišćeni u procesu modeliranja, se sastoji od četiri države (Nemačka, Francuska, Italija i Španija). U kombinaciji sa predstavljenim državama korišćena je potpuna vremenska serija, te je za svaku od datih država preuzeta vremenska serija od 2018. godine do 2019. Godine. U tom smislu može se reći da su se za modelovanje koristili strogo balansirani panel podaci.

Rad sadrži teorijsku podlogu u okviru koje je prikazan teorijski pristup različitim modelima. Predstavljeni modeli u prvom delu rada su: OLS model model fiksnih efekata i model slučajnih efekata. Dodatno se u radu daju i smernice odabira najadekvatnijeg modela.

Nakon teorijske podloge koja je bila neophodna kako bi se na pravi način primenila panel regresiona analiza, sledeći deo koji je detaljno objašnjen u radu jeste emiprijska analiza. Za empirijsku analizu korišćeni su podaci preuzeti sa zvaničnog sajta evropske kuće za statistiku – Eurostat. Prvi kreirani model jeste OLS model koji nije ušao kao kandidat za finalni model jer su performanse modela slabije u odnosu na naredna dva. Dodatno, OLS model nije uzeo u obzir niti vremensku niti individualnu komponentu datog panel skupa podataka. S toga analiza je nastavljena u pravcu modela sa fiksnim efektima. Obzirom da se u inicijalnim panel regresionim analizama pojavljuju promenljive koje nisu statistički značajne, fokus je bio na izostavljanju istih, kako bi se napravio što kvalitetniji model. Od predstavljenih varijacija modela sa fiksnim efektom odabran je onaj koji uzima u obzir efekte individualnih jedinica posmatranja (u našem slučaju država) zbog najbolje pokazanih performansi (R^2, prilagođeni R^2 itd.). Nakon kreiranja i odabira predstavnika modela sa fiksnim efektima sprovedena je panel regresiona analiza i za model sa slučajnim efektima. I u ovom modelu se pojavljuju nezavisne promenljive koje se pokazuju da nisu statistički značajne. Kao i u modelu fiksnih efekata, i ovde je vođeno računa da u model ulaze samo one promenljive čiji je koeficijent statistički različit od nule tj. promenljive koje su statistički značajne.

Krajnji deo empirijske analize obuhvata odabir finalnog modela. Neophodno je bilo doneti odluku koji model treba gledati kao finalni. U tom izboru korišćen je Hausmanov test čiji su rezultati uputili na to da za finalni model treba uzeti model slučajnih (stohastičkih) efekata.

LITERATURA

[1.] Crvenković Lozanov Z., *Statistika,* Prirodno-matematički fakultet, Univerzitet u Novom Sadu, Novi Sad, 2012.

[2.] Dragutinović Mitrović R. S., *Analiza panel serija*, Zadužbina Andrejević, Beograd, 2002.

[3.] Gujarati D. N., Dawn C. P., *Basic Econometrics,* New Delhi, 2014.

[4.] Hausman A. J., Taylor E. W., *Panel Data and Unobservable Individual Effects*, Econometrica, 1981.

[5.] Hsiao C., *Analysis of Panel Data*, Cambridge University Press, Cambridge University, United Kingdom, 2003.

[6.] https://ec.europa.eu/eurostat

[7.] https://www.investopedia.com/terms/h/heteroskedasticity.asp

[8.] Knežević A., *Primena panel modela u identifikovanju faktora uspešnosti poslovanja proizvodnih preduzeća,* Doktorska disertacija, Fakultet tehničkih nauka, Univerzitet u Novom Sadu, Novi Sad, 2015.

[9.] Park H. M., *Practical Guides To Panel Data Modeling*: A Step by Step Analysis Using Data, International University of Japan, Niigata, Japan 2011.

[10.] Schmidheniy K., *Panel Data: Fixed and Random Effects,* Short Guides to Microeconometrics, University of Basel, Basel, Switzerland, 2018.

[11.] Vijayamohanan Pillai N., *Panel Data Analysis with Stata Part 1: Fixed Effects and Random Effects Models,* Munich Personal RePEc Archive, Centre for Development Studies, Kerala, India, 2017.

BIOGRAFIJA

Aleksandra Milojević rođena je 22. jula 1993. godine u Prnjavoru, BiH. Završila je osnovnu školu „Nikola Tesla“ u Derventi 2008. godine kao vukovac. Nakon toga upisuje srednju školu Gimnazija „Mihajlo Pupin“, opšti smer koju završava 2012. godine. sa odličnim uspehom.Iste godine upisuje Prirodno – matematički fakultet na Univerzitetu u Novom Sadu, smer primenjena matematika – matematika finansija. Osnovne akademske studije završava 2016. godine kada upisuje master studije primenjene matematike. Sve ispite predviđene planom i programom polaže do 2019. godine čime ostvaruje pravo na odbranu master rada.

Od januara 2019. godine zaposlena je kao IT konsultant u kompaniji „Synechron” u Novom Sadu.

UNIVERZITET U NOVOM SADU

PRIRODNO – MATEMATIČKI FAKULTET

KLJUČNA DOKUMENTACIJSKA INFORMACIJA

Redni broj:	
RBR	
Identifikacioni broj:	
IBR	
Tip dokumentacije:	monografska dokumentacija
TD	
Tip zapisa:	tekstualni štampani materijal
TZ	
Vrsta rada:	master rad
VR	
Autor:	Aleksandra Milojević
AU	
Mentor:	Prof. dr Zagorka Lozanov - Crvenković
MN	
Naslov rada:	Primena panel regresione analize na demografske pokazatelje
NR	
Jezik publikacije:	srpski (latinica)
JP	
Jezik izvoda:	srpski/engleski
JI	
Zemlja publikovanja:	Republika Srbija
ZP	
Uže geografsko područje:	Vojvodina
UGP	
Godina:	2020.
GO	

Izdavač: **IZ**	Autorski reprint
Mesto i adresa: **MA**	Novi Sad, Departman za matematiku i informatiku, Prirodno – matematički fakultet, Univerzitet u Novom Sadu, Trg Dositeja Obradovića 4
Fizički opis rada: **FOR**	(5/65/5/11/4/0) (broj poglavlja/strana/lit. citata/tabela/grafika/priloga)
Naučna oblast: **NO**	Matematika
Naučna disciplina: **ND**	Primenjena matematika
Predmetne odrednice, ključne reči: **PO, UDK**	Panel podaci, modeli fiksnih efekata, modeli slučajnih efekata, metod običnih najmanjih kvadrata
Čuva se: **ČU**	Biblioteka Departmana za matematiku i informatiku Prirodno – matematičkog fakulteta Univerziteta u Novom Sadu
Važna napomena: **VN**	Nema
Izvod: **IZ**	Prvo poglavlje rada daje detaljan pregled prednosti I ograničenja panel podataka u regresionoj analizi. Drugo poglavlje rada sadrži detaljan pregled modela vezanih za panel podatke. Treće poglavlje rada uključuje sve potencijalne probleme koji se mogu pojaviti u analizi panel regresije. U četvrtom poglavlju prikazan je Hausmanov test koji aje odgovor koji je model najpogodniji. U petom poglavlju data je empirijska analiza.
Datum prihvatanja teme od strane NN veća: **DP**	
Datum odbrane: **DO**	
Članovi komisije: **KO**	Predsednik: dr Ljiljana Gajić, redovni profesor, Prirodno-matematički fakultet, Univerzitet u Novom Sadu Mentor: dr Zagorka Lozanov-Crvenković, redovni profesor, Prirodno-matematički fakultet, Univerzitet u Novom Sadu Član: dr Ivana Štajner-Papuga, redovni profesor, Prirodno-matematički fakultet, Univerzitet u Novom Sadu

UNIVERSITY OF NOVI SAD

FACULTY OF SCIENCE

KEY WORDS DOCUMENTATION

Accession number:	
ANO	
Identification number:	
INO	
Document type:	Monograph type
DT	
Type of record:	Printed text
TR	
Contents code:	Master's thesis
CC	
Author:	Aleksandra Milojević
AU	
Menthor:	Zagorka Lozanov – Crvenković PhD
MN	
Title:	Application of panel regression analysis to demographics indicators
TI	
Language of text:	Serbian (Latin)
LT	
Language of abstract:	Serbian/english
LA	
Country of publication:	Republic of Serbia
CP	
Locality of publication:	Vojvodina
LP	
Publication year:	2020.
PY	

Publisher: **PU**	Author's reprint
Publication place: **PP**	Novi Sad, Department of Mathematics and Infomatics, Faculty of Science, University of Novi Sad, Trg Dositeja Obradovića 4
Physical description: **PD**	(5/65/5/11/4/0) (chapters/pages/literature/tables/graphics/add. lists)
Science field: **SF**	Mathematics
Scientific discipline: **SD**	Applied mathematics
Subjectkey word: **SKW**	Panel data, Fixed Effects model, Stohastic Effects model, pooled model
Holding data: **HD**	The library of the Department of Mathematics and Informatics, Faculty of Science, University of Novi Sad
Note: **N**	None

Abstract: **AB**	The first chapter provides a detailed overview of the advantages and limitations of panel data in regression analysis. The second chapter contains a detailed overview of the models related to panel data. The third chapter includes all potenitial problems that appear in the panel regression analysis. In the fourth chapter, the Hausman test is explained, which is the answer to which model is the most suitable. The fifth chapter is an empirical analysis.

Accepted by the Scientific Board on: **AS**	
Defended: **DE**	
Thesis defended board: **DB**	President: Ljiljana Gajić, PhD, full professor, Faculty of Science, University of Novi Sad Mentor: Zagorka Lozanov-Crvenković, PhD, full professor, Faculty of Science, University of Novi Sad Member: Ivana Štajner-Papuga, PhD, full professor, Faculty of Science, University of Novi Sad