Modeliranje rasta prihoda od prodaje iz panel podataka

Vašarević, Filip

Sveučilište J.J. Strosmayera u Osijeku
Odjel za matematiku

Filip Vašarević

Modeliranje rasta prihoda od prodaje iz panel podataka

Diplomski rad

Osijek, 2017.

Sveučilište J.J. Strosmayera u Osijeku
Odjel za matematiku

Filip Vašarević

Modeliranje rasta prihoda od prodaje iz panel podataka

Diplomski rad

Mentor: prof.dr.sc. Mirta Benšić
Komentor: prof.dr.sc. Nataša Šarlija

Osijek, 2017.

Sadržaj

Uvod

U ovom diplomskom radu ćemo raditi sa specifičnim tipom podataka, to su takozvani panel podaci. U ovom trenu ih nećemo definirati jer ćemo u nastavku puno govoriti o njima. Nama je cilj u ovom radu proučiti linearne modele koji se najčešće koriste u analizi panel podataka. Gledat/ ćemo teorijsku pozadinu svakog takvog modela, te pokazati primjenu u analizi rasta prihoda od prodaje u poduzećima iz prerađivačke industrije na osnovu panel podataka kojima raspolažemo.

U početku ovog diplomskog rada ćemo navesti osnovne pojmove i definicije linearne regresije i vremenskih nizova jer su nam ta dva koncepta vrlo bitna kako bi što bolje razumjeli ono što radimo u nastavku. U 2. i 3. poglavlju ćemo iskazati osnovne definicije o linearnoj regresiji te primjer linearne regresije, a zatim i osnovne definicije vremenskih nizova i njihove primjere.

U četvrtom poglavlju ćemo detaljno govoriti o statičkim linearnim panel modelima. Obraditi ćemo tri statička modela, te ćemo reći koje su pretpostavke tih modela, što nam pojedini model govori, te na koji način se koristi. Unutar istog poglavlja ćemo nešto reći i o procjeniteljima koji se koriste pri određivanju koeficijenata u svakom modelu. Na kraju poglavlja ćemo pogledati statističke testove pomoću kojih utvrđujemo opravdanost korištenja pojedinih modela. Primjenu tih testova moći ćemo vidjet kasnije u praktičnom djelu rada.

Poglavlje 5 nam govori o prihodima od prodaje kao ekonomskoj veličini s kojom se mjeri rast nekog poduzeća, u našem slučaju poduzeća u prerađivačkoj industriji. Također ćemo reći koje varijable utječu na rast poduzeća, te koji su najznačajniji financijski omjeri kada gledamo rast poduzeća, odnosno rast prihoda od prodaje. Zatim ćemo ukratko nešto reći i o onim varijablama koje smo mi koristili u modelu, te opisati sve financijske omjere koje smo koristili.

Kroz zadnje poglavlje ćemo opisati panel podatke kojima smo raspolagali te ćemo prikazati kako smo pristupili modeliranju, te kako smo došli do našeg konačnog modela. Objasniti ćemo rezultate i predstaviti detaljnije konačan model koji ima najbolje statističke karakteristike i ekonomsku pozadinu.

Linearna regresija

Kao što smo rekli u uvodu, ovaj će rad govoriti o panel podacima i modeliranju panel podataka. No prije nego što definiramo panel podatke i modele na panel podacima moramo nešto reći o linearnoj regresiji i vremenskim nizovima. To radimo iz razloga što su modeli kojima se bavimo usko povezani s linearnom regresijom. Isto tako je i sa vremenskim nizovima, no njih ćemo definirati u sljedećem poglavlju.

2.1 Linearna regresija

U statistici linearna regresija je metoda u kojoj jednu varijablu opisujemo s jednom ili više drugih varijabli. Linearna regresija se često koristi u praksi i vrlo je korisna i značajna metoda statističke analize. U linearnoj regresiji ispitujemo ovisnost zavisne varijable o jednoj ili više nezavisnih varijabli.

Definicija 2.1. *Statističkim modelom linearne regresije pretpostavljamo da su vrijednosti zavisne varijable Y linearno povezane s nezavisnom varijablom X. Veza se može zapisati na sljedeći način*

$$Y = \beta_0 + \beta_1 X + \epsilon \tag{2.1}$$

pri čemu su β_0, β_1 nepoznati parametri, a ϵ je slučajna varijabla koja predstavlja grešku modela.

Osnovna pretpostavka linearne regresije je da vrijedi

$$E[Y|X = x] = \beta_0 + \beta_1 x, \quad \forall x \in R(X)$$

Model linearne regresije možemo prikazati i bez korištenja uvjetnih očekivanja. To se događa u slučaju kada je X deterministička varijabla, a ne slučajna. Pretpostavimo da je Y_i slučajna varijabla koja opisuje zavisnu varijablu u slučaju kada nezavisna varijabla ima vrijednost x_i, $i = 1, .., n$. Tada pretpostavljamo

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

gdje je ϵ_i slučajna varijabla greške modela.
Cilj analize podataka je procijeniti β_0 i β_1. Pretpostavimo da imamo neki niz podataka $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$.

Procjenu parametara možemo napraviti koristeći metodu najmanjih kvadrata (Ordinary Least Square, kasnije OLS). Metoda najmanjih kvadrata znači da moramo minimizirati funkciju

$$F(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2,$$

po nepoznatim parametrima β_0 i β_1. Primjenom metode OLS dobivamo procjenitelje $\hat{\beta}_0, \hat{\beta}_1$ i oni iznose

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

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x},$$

pri čemu je

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i,$$

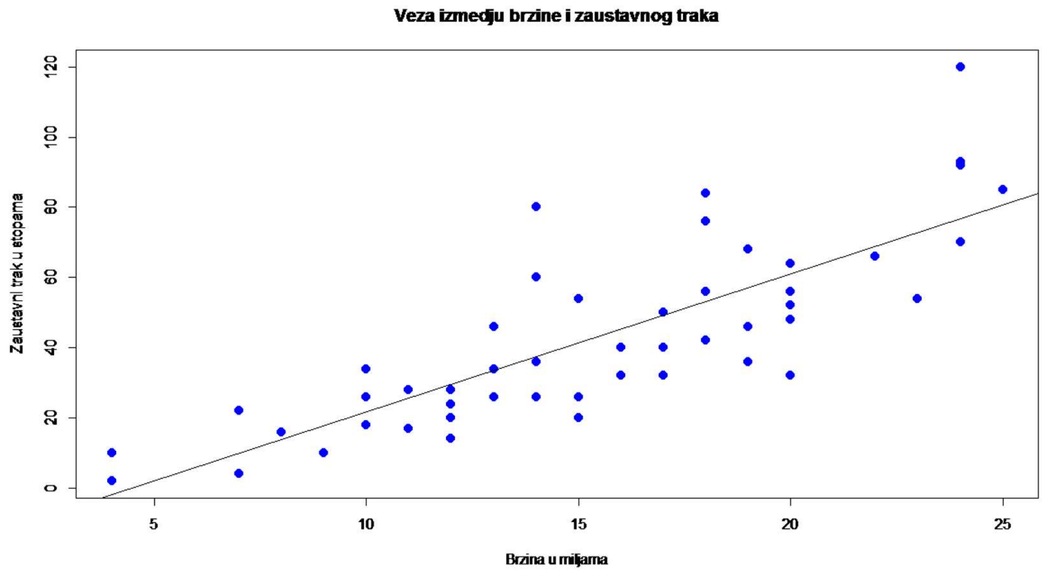
$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

Pretpostavke koje slučajna varijabla greške ϵ_i za svaki i = 1,...,n mora zadovoljiti u analizi ovakvog modela su:

- $E[\epsilon_i] = 0$
- $Var[\epsilon_i] = \sigma^2$
- $Cov[\epsilon_i, \epsilon_j] = 0, i \neq j$
- $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$

Četvrtu točku pretpostavke možemo ispustiti iz razloga što se ne narušava bitno metodologija zaključivanja pri ovakvom modeliranju kod slučaja kad imamo veliki broj podataka [13]

Primjer 2.1. *Na slici 2.1 vidimo vezu između zaustavnog traka vozila izraženog u stopama i brzine vozila u miljama na sat. Plavim točkama su označeni stvarni podaci. Vidimo da bi njihovu vezu mogli opisati pravcem, pa zato koristimo pravac procijenjen iz svih podataka metodom najmanjih kvadrata i on je na grafu označen sa crnom bojom.*



Slika 2.1: Model linearne regresije za brzinu u miljama na sat i zaustavni trak u stopama

Također možemo vidjet da ovaj pravac relativno dobro opisuje zadane podatke, no na temelju grafičkog prikaza ne možemo ništa zaključiti. Da bi vidjeli koliko je model kvalitetan moramo koristiti statističke testove. Više o tome se može vidjeti u [14].

2.2 Višestruka linearna regresija

Višestruka linearna regresija je generalizacija jednostavne linearne regresije s više nezavisnih varijabli.
Osnovna pretpostavka višestruke linearne regreije je da vrijedi

$$E[Y|\boldsymbol{X}] = \beta_0 + \beta_1 X_1 + \cdots + \beta_K X_K$$

Definicija 2.2 (*Višestruka linearna regresija*)**.** *Neka je Y zavisna varijabla koja je linearno povezana s K nezavisnih varijabli $X_1, \ldots, X_K$ preko parametara $\beta_0, \beta_1, \ldots, \beta_K$. Veza se može zapisati na sljedeći način*

$$Y = \beta_0 + X_1\beta_1 + X_2\beta_2 + \cdots + X_K\beta_K + \boldsymbol{\epsilon} \tag{2.2}$$

pri čemu su $\beta_0, \beta_1, \ldots, \beta_K$ regresijski koeficijenti povezani s $X_1, \ldots, X_K$ respektivno, a $\boldsymbol{\epsilon}$ predstavlja grešku modela.

Imamo N promatranja za Y i vektor $(X_1, \dots, X_K)$ koje možemo zapisati na sljedeći način:

$$(Y, \boldsymbol{X}) = \begin{bmatrix} Y_1 & x_{11} & . & . & . & x_{K1} \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ Y_N & x_{1N} & . & . & . & x_{KN} \end{bmatrix} = \begin{bmatrix} Y_1, X_1^T \\ \vdots \\ Y_N, X_K^T \end{bmatrix}$$

pri čemu je $Y = (Y_1, \dots, Y_N)^T$ vektor dimenzija N, $\boldsymbol{X}_i = (x_{1i}, \dots, x_{Ki})^T$ vektor dimenzija K. Napomenimo da su prva i treća matrica iz gore navedenoga izraza particijske matrice. U takvom slučaju imamo N jednadžbi oblika kao u definiciji 2.2 i njih možemo napisati na sljedeći način:

$$Y_i = \boldsymbol{X}_i^T \boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad i = 1, \dots, N$$

gdje je $\boldsymbol{\beta}^T = (\beta_0, \dots, \beta_K)$.
Višestruki linearni regresijski model možemo zapisati i matrično

$$Y = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{2.3}$$

gdje je $\boldsymbol{X}$ matrica dimenzija $N \times K$ napravljena za N opažanja i svih K nezavinsih varijabli i $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_N)^T$. Također i kod višestruke regresije bitno je odrediti procjenitelje za nepoznati parametar β. To ćemo napraviti tako da minimiziramo funkciju

$$\boldsymbol{S}(\boldsymbol{\beta}) = \sum_{i=1}^{n} (\beta_i x_{1i} + ... + \beta_k x_{Ki} - Y_i)^2 \tag{2.4}$$

Tako da procjena za β iznosi

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T Y, \tag{2.5}$$

pri čemu $\boldsymbol{X}^T \boldsymbol{X}$ mora biti punog ranga. Ovi procjenitelji se još mogu i teorijski proučiti i analizirati, no to nije tema ovoga rada. Detaljnije možete pogledati u [14]. Kao i kod modela linearne regresije i ovdje vrijede iste pretpostavke za slučajne greške:

- $E[\epsilon_i] = 0$
- $Var[\epsilon_i] = \sigma^2$
- $Cov[\epsilon_i, \epsilon_j] = 0, i \neq j$
- $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$
- X nije stohastička matrica

U ovom slučaju procjenitelji i model se moraju ispitati i moramo provjeriti pretpostavke modela kako bi se utvrdila ispravnost istog prije donošenja konačnog zaključka.

Vremenski nizovi

Drugi bitan dio panel podataka su vremenski nizovi. U ekonomiji i ekonomskim istraživanjima često se podaci prikupljaju u nekim vremenskim trenucima (sat,dan,mjesec,...) te se na taj način pokušava predvidjet ponašanje budućih pojava. Takvu analizu podataka nazivamo analiza vremenskih nizova. U nastavku ćemo vidjeti neke osnovne definicije i primjere vremenskih nizova.

3.1 Definicije i primjeri vremenskih nizova

Definicija 3.1. ***Vremenski niz*** *je uređen niz vrijednosti*

$$(x_t, t \in T)$$

pri čemu je T skup indeksa koji predstavljaju vremenske trenutke.

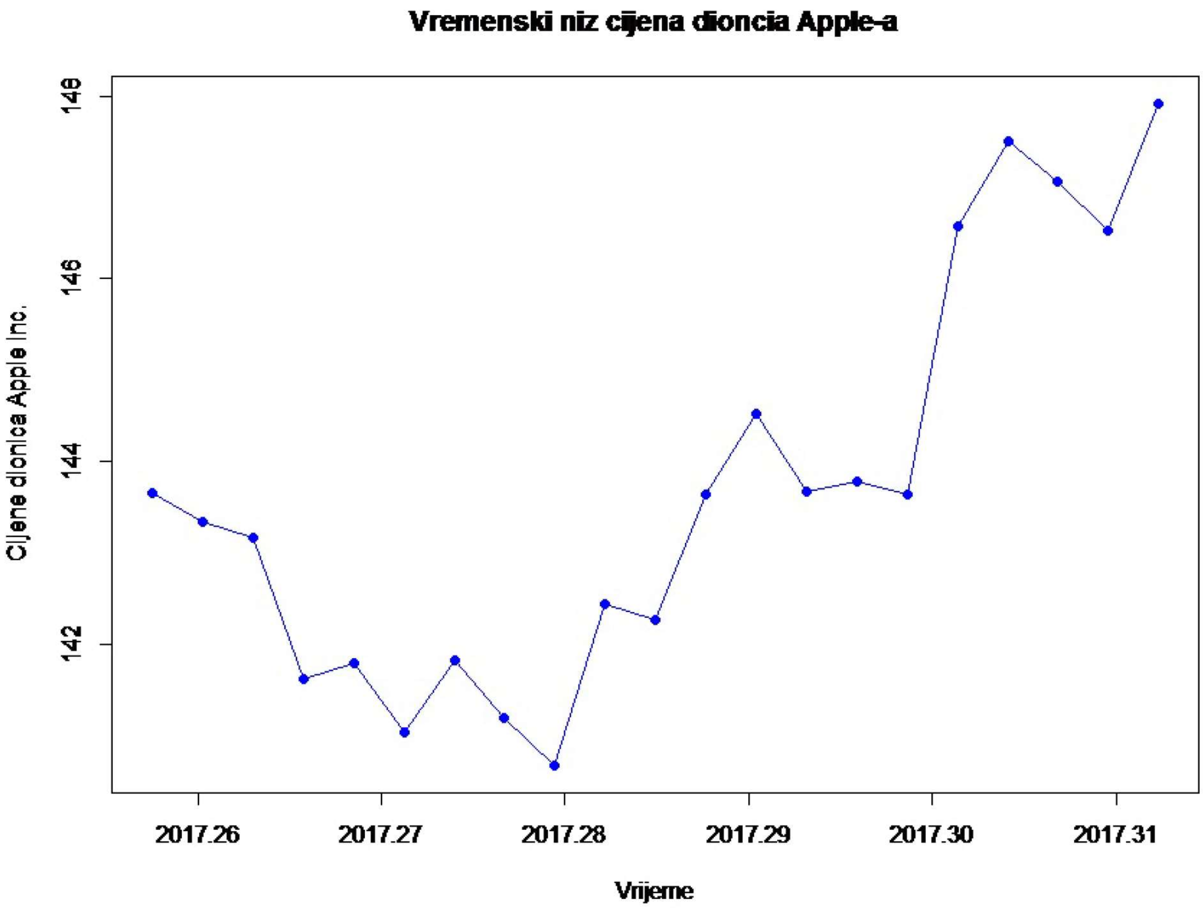
Kod vremenskih nizova je cilj pronaći prikladan model kojim bismo što bolje opisali promatrane podatke u vremenu. Možemo pretpostaviti da svaka vrijednost x_t zapravo jedna realizacija slučajne varijable X_t. To znači da za vremenski niz $(x_t, t \in T)$ kao model koristimo slučajni proces $(X_t, t \in T)$. Vrijednosti vremenskog niza predstavljaju realizacije slučajnog procesa. Cilj analize vremenskog niza je pronaći slučajni proces kojim se on može opisati. Više o slučajnim procesima možete vidjeti u [4]

Primjer 3.1. *Pogledajmo cijene dionica u Tablici 3.1. Radi se o multinacionalnoj tvrtki Apple inc. Cijene tih dionica predstavljaju realizacije u trenucima od 6.4.2017. do 5.5.2017. Podaci koje gledamo u tablici čine jedan vremenski niz. Na slici 3.1. možemo vidjeti i grafički prikaz tog vremenskog niza.*

Datum	Cijena	Datum	Cijena
6.4.2017	143,66	24.4.2017	143,64
7.4.2017	143,34	25.4.2017	144,53
10.4.2017	143,17	26.4.2017	143,68
11.4.2017	141,63	27.4.2017	143,79
12.4.2017	141,80	28.4.2017	143,65
13.4.2017	141,05	1.5.2017	146,58
17.4.2017	141,83	2.5.2017	147,51
18.4.2017	141,20	3.5.2017	147,06
19.4.2017	140,68	4.5.2017	146,53
20.4.2017	142,44	5.5.2017	147,92
21.4.2017	142,27		

Tablica 3.1: Cijene dionica Apple-a od 6.4.2017 do 5.5.2017.

Da bi vremenske nizove što bolje shvatili postoje različiti testovi i metode kojima se pokušava pronaći poznati slučajni proces koji najbolje odgovara vremenskom nizu. Znamo da je slučajni proces potpuno određen ukoliko znamo sve njegove konačnodimenzionalne distribucije. Jedna od najbitnijih stvari kod analize vremenskih nizova je pojam stacionarnosti. No da bi njega mogli definirati potrebna je definicija funkcije autokovarijanci.



Slika 3.1: Prikaz vremenskog niza za cijene dionica iz Tablice 3.1.

Definicija 3.2. *Neka je* $\{X_t, t \in T\}$ *slučajni proces za koji je* $Var(X_t) < \infty$, $\forall t \in T$ *Tada funkciju* $\gamma : T \times T \to \mathbb{R}$ *danu izrazom*

$$\gamma(s,t) = Cov(X_s, X_t) = E[(X_s - E[X_s])(X_t - E[X_t])], \forall s,t \in T \tag{3.1}$$

nazivamo ***funkcija autokovarijanci*** *slučajnog procesa* $\{X_t, t \in T\}$.

Definicija 3.3. *Za slučajni proces* $\{X_t, t \in \mathbb{Z}\}$ *kažemo da je* ***stacionaran*** *ako za njega vrijede sljedeći uvjeti:*

- $E[X_t^2] < \infty$, $\forall t \in \mathbb{Z}$
- $E[X_t] = c, \quad c \in \mathbb{R}, \quad \forall t \in T$
- $\gamma(s,t) = \gamma(s+r, t+r)$, $\forall s,t,r \in \mathbb{Z}$

Ako kod analize podataka možemo pretpostaviti stacionarnost tada često u modeliranju možemo primijeniti slučajne procese specifičnih struktura kao što su proces pomičnih prosjeka ili MA proces, autoregresivni proces ili AR proces i ARMA proces.

Ukoliko imamo nestacionarnost uobičajeno je da prvo provedemo određene modifikacije podataka zbog kojih ćemo sa velikom vjerojatnošću moći tvrditi da podaci nisu nestacionarni. Najčešća takva modifikacija ili metoda je diferenciranje. Također, tu u obzir dolazi i logaritmiranje, potenciranje, itd. Najpoznatiji procesi pomoću kojih se modeliraju nestacionarni vremenski nizovi su ARIMA, ARCH, GARCH, itd. Trebamo još napomenuti da vremenski nizovi ne moraju nužno pratiti samo jednu varijablu. Stoga možemo razlikovati jednodimenzionalne i višedimenzionalne vremenske nizove. Više o tome se može vidjeti u [1]

Modeliranje panel podataka

U prošla dva poglavlja smo vidjeli osnovne definicije i pojmove linearne regresije i vremenskih nizova. Oni su nam bili potrebni da bi mogli definirati panel podatke i panel modele. Panel podaci kao što smo ranije spomenuli su spoj vremenskih nizova i linearne regresije. I oni su vrlo bitni za ekonomska i druga istraživanja koja svoje teorije temelje na kreiranju modela za podatke s kojima raspolažu. Kroz ovo poglavlje, ali i kroz sam diplomski rad ćemo se baviti sa statičkim panel modelima. U ovom poglavlju još ćemo obraditi vrste statičkih modela, njihove procjenitelje, te statističke testove za panel modele.

4.1 Panel podaci

U ovom poglavlju ćemo na primjeru intuitivno objasniti kako izgledaju panel podaci. Koristiti ćemo izvadak iz baze na kojoj ćemo raditi kasnije. **Panel podaci** su nova struktura podataka u kojoj imamo više primjeraka za koje gledamo neke njihove odrednice kroz više vremenskih razdoblja, te ih onda analiziramo. U sljedećoj tablici možemo vidjet kako izgledaju panel podaci u praksi. U drugom stupcu je varijabla "ID_SUBJEKTA" koja govori o kojem se poduzeću radi: Prvi stupac je varijabla "GODINA". Ona pokazuje vremensko razdoblje kroz koje promatramo poduzeća, u ovom slučaju ćemo ih promatrati u razdoblju od 5 godina. Ostale varijable su financijski omjeri koji su od značaja za poduzeća koja promatramo.

GODINA	ID_SUBJEKTA	prodaja	velicina	NKD	djelatnost	kdjel	NematImovina	kImovkObv	pPrihZapos	Zkz	Zdk
2010	24	156214	1	3250	C	B	0	0,0827		3,2705	-1,4404
2011	24	172577	1	3250	C	B	0	0,159	172577	3,6123	-1,3827
2012	24	150359	1	3250	C	B	0	0,0938		7,4312	-1,1554
2013	24	124555	1	3250	C	B	0	0,137		6,3306	-1,1875
2014	24	109899	1	3250	C	B	0	0,1263	111076	7,0156	-1,1662
2010	29	11847020	1	2361	C	B	0	3,5681	749081,8	0,2024	0,2537
2011	29	9409754	1	2361	C	B	0	4,8687	860278,9	0,1515	0,1786
2012	29	7131351	1	2361	C	B	0	3,4537	650871,8	0,2067	0,2607
2013	29	4710421	1	2361	C	B	0	2,3027	523438,2	0,308	0,4451
2014	29	5084242	1	2361	C	B	0	1,76	635980,3	0,4075	0,6878
2010	50	1151678	1	1392	C	B	0	7,4015	191946,3	0,1297	0,149
2011	50	1411284	1	1392	C	B	0	4,3972	282256,8	0,1904	0,2352
2012	50	1732974	1	1392	C	B	0,001	10,0093	346594,8	0,0747	0,0807
2013	50	1714955	1	1392	C	B	0,0004	6,7014	342991	0,1188	0,1348
2014	50	1579462	1	1392	C	B	0	6,4493	315892,4	0,1254	0,1434

Tablica 4.1: Primjer panel podataka za 3 poduzeća u razdoblju od 2010. do 2014

U nastavku ćemo proučiti najbitnije statičke linearne panel modele za ovakav tip podataka, gdje ćemo veći fokus postaviti na onaj model koji smo koristili u modeliranju naše baze panel podataka.

4.2 Statički panel modeli

Statički panel modeli se izgrađuju na fiksnim godinama, tj. ne oslanjaju se na regresiju s prethodnim godinama, pa zbog toga kažemo da su takvi modeli **statički**. U nastavku ćemo definirati panel model te proučiti najbitnije statičke modele.

Neka imamo panel s N jedinki promatranih kroz T vremenskih trenutaka. Najopćenitiji oblik linearnog statičkog panel modela možemo prikazati sljedećom formulom

$$Y_{it} = \alpha_{it} + X_{it}^T \beta_{it} + \epsilon_{it}, \quad i = 1, ..., N, \quad t = 1, ..., T \tag{4.1}$$

gdje β_{it} mjeri učinke za X_{it} u vremenu t za jedinku i. Ovaj model je previše generalan da bi bio koristan i koeficijent β_{it} mora biti bolje strukturiran i iskorišten. Y_{it} predstavlja zavisnu slučajnu varijablu koju modeliramo za jedinku i u vremenskom trenutku t, dok je X_{it} vektor nezavisnih slučajnih varijabli za i-tu jedinku u vremenskom trenutku t. α_{it} je slobodni član također za i-tu jedinku u vremenskom trenutku t, a ϵ_{it} predstavlja grešku modela za koju se u većini slučajeva pretpostavlja da ima očekivanje 0 i varijancu σ_ϵ^2.

Kao što smo rekli, ovaj model je previše generalan. Da bi došli do nama korisnih modela koji se koriste u praksi, moramo uvesti restrikcije na generalni model. Uvođenjem restrikcija dobivamo različite modele koje onda možemo primjenjivati na panel podatke. Objasniti ćemo tri tipa takvih modela. To su **Model konstantnih koeficijenata(Pooled model), Model fiksnih efekata(Fixed effects model) i Model slučajnih efekata(Random effects model)**.

4.2.1 Model konstantnih koeficijenata (Pooled model)

Pretpostavimo da je panel za N jedinki promatram kroz T vremenskih trenutaka. Tada ga možemo prikazati formulom:

$$Y_{it} = \alpha + \boldsymbol{X}_{it}^T \boldsymbol{\beta} + u_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{4.2}$$

gdje i označava jedinke koje promatramo, a t vremenske trenutke.

Kao i u općenitom modelu Y_{it} predstavlja zavisnu slučajnu varijablu za i-tu jedinku u vremenu t. $\boldsymbol{X}_{it}$ je vektor nezavisnih slučajnih varijabli dimenzija K. β je vektor parametara, dok je α slobodni član. u_{it} je slučajna varijabla s očekivanjem 0 i varijancom σ^2 i ona predstavlja grešku modela.
Možemo uočiti da je u odnosu na najopćenitiji oblik linearnog panel modela (4.1) ovaj model (4.2) ima fiksiran slobodan član i vektor koeficijenata. To znači da oni ne variraju kroz jedinke i vrijeme nego su jednaki za sva promatranja. To je snažna restrikcija u odnosu na početni model.Pretpostavke za primjenu modela konstantnih koeficijenata su:

- Sve pretpostavke karakteristične za primjenu metode najmanjih kvadrata. Vidi [13]
- Fiksni slobodni član je konstantan po svim jedinkama i
- Efekt za bilo koji X na Y je konstantan kroz sva promatranja

Možemo vidjeti da fiksni slobodni član govori da ovakav model ima smisla primijeniti u slučaju individualnih učinaka svake promatrane jedinke. U slučaju da gore navedene pretpostavke ne možemo pretpostaviti, onda ne bismo trebali koristiti model konstantnih koeficijenata. Vidi [5].

Teorijske pretpostavke modela konstantnih koeficijenata su sljedeće:

1. $E[\boldsymbol{X}_{it}u_{it}] = 0, \forall i = 1, \ldots, N, \ \forall t = 1, \ldots, T$
2. $r(\sum_{i=1}^{T} E[\boldsymbol{X}_{it}\boldsymbol{X}_{it}^T]) = K \ \forall i = 1, ..., N$ pri čemu smo sa r označili rang matrice.
3. (a) $E[u_{it}^2\boldsymbol{X}_{it}\boldsymbol{X}_{it}^T] = \sigma^2 E[X_{it}X_{it}^T], \quad \forall i, t$ pri čemu je $\sigma^2 = E[u_{it}^2]$.
   (b) $E[u_{it}u_{is}\boldsymbol{X}_{it}\boldsymbol{X}_{is}^T] = 0, \quad i \neq s, \quad \forall i = 1, ..., N, \quad t, s = 1, ..., T.$

Iz pretpostavke 1. možemo vidjet da je ovaj model dosta sličan linearnoj regresiji o kojoj smo pričali u 1. poglavlju. Ta pretpostavka govori da je greška modela nekorelirana s nezavisnim varijablama, ali samo onima iz istog vremenskog trenutka. Ona ništa ne govori o odnosu $\boldsymbol{X}_{it}$ i u_{is}, za $t \neq s$.

Što se tiče druge pretpostavke, ona govori o rangu matrice koji dobijemo od vektora nezavisnih varijabli. Kod definicije 4.2 smo rekli da svaki vektor $\boldsymbol{X}_{it}$ ima dimenziju K, pa onda možemo pretpostaviti da je matrica iz pretpostavke 2. punog ranga.

Treća pretpostavka ovog modela govori da varijanca greške ne korelira s nezavisnim varijablama u bilo kojem vremenskom trenutku. To svojstvo se naziva **stroga homoskedastičnost**. Pretpostavka pod b) zahtjeva da kovarijance grešaka budu jednake nula kroz različite vremenske trenutke. Također je potrebno da greške budu međusobno nekorelirane da bi ta pretpostavka vrijedila, tj. mora vrijediti da je $E[u_{it}u_{is}] = 0$, za $t \neq s$.

4.2.2 Model fiksnih efekata (Fixed effects model)

Model fiksnih efekata uključuje slobodni član iz jednadžbe (4.1), dok vektor β ostaje fiksan. Možemo ga zapisati na sljedeći način

$$Y_{it} = \alpha_i + \gamma_t + \boldsymbol{X}_{it}^T \boldsymbol{\beta} + u_{it}. \tag{4.3}$$

Ovaj model je koristan iz razloga što ćemo u praksi gledati podatke koji imaju puno jedinaka, a promatraju se kroz kraći vremenski period. Zbog toga se vremenski učinci γ_t mogu puno kvalitetnije procijeniti jer imamo više promatranja, dok sa slobodnim članovima za individue α_i imamo problem.

Model iz defincije 4.3 se u praksi preoblikuje na način da se koristi oblik u kojem varira samo slobodni član za jedinke. Takav model izgleda ovako

$$Y_{it} = \alpha_i + \boldsymbol{X}_{it}^T \boldsymbol{\beta} + \epsilon_{it} \tag{4.4}$$

Kao i u prethodnom potpoglavlju 4.2.1 neke od ovih varijabli imaju jednako značenje pa ih nećemo ponavljati. Ostaje varijabla α_i koja je slučajna varijabla koja opisuje neopaženi individualni efekt i-te jedinke. ϵ_{it} je slučajna varijabla s očekivanjem 0 i varijancom σ^2 i on predstavlja grešku modela. Varijabla γ_t nemamo u modelu (4.4), ali ona se može uvesti ako ima smisla promatrati vremenske efekte. Ona je također slučajna varijabla.
Teorijske pretpostavke modela fiksnih efekata su sljedeće:

1. $E[\epsilon_{it}|X_{i1}, ..., X_{iT}, \alpha_i] = 0, \quad t = 1, ..., T$ **(Svojstvo egzogenosti)**
2. $r(\sum_{i=1}^{T} E[\ddot{\boldsymbol{X}}_{\boldsymbol{it}} \ddot{\boldsymbol{X}}_{\boldsymbol{it}}^{\boldsymbol{T}}] = r(E[\ddot{\boldsymbol{X}}_{\boldsymbol{i}}{}^T \ddot{\boldsymbol{X}}_{\boldsymbol{i}}]) = K$, pri čemu je $\ddot{\boldsymbol{X}}_{\boldsymbol{it}} = \boldsymbol{X}_{\boldsymbol{it}} - \bar{\boldsymbol{X}}_{\boldsymbol{i}}$, a $\bar{\boldsymbol{X}}_{\boldsymbol{i}} = T^{-1} \sum_{i=1}^{T} \boldsymbol{X}_{it}$
3. $E[\epsilon_{it}^2|X_{i1}, ..., X_{iT}, \alpha_i] = \sigma_\epsilon^2$

Prva pretpostavka, tj. svojstvo egzogenosti govori da je slučajna varijabla koja predstavlja grešku modela, uvjetno na prošlost, sadašnjost i budućnost vektora nezavisnih varijabli, te na slučajnu varijablu α_i ima očekivanje 0.

Kao i kod modela konstantnih koeficijenata druga pretpostavka govori o rangu matrice. Vidimo da je matrica dimenzija $K \times K$, pa možemo pretpostaviti da je punog ranga.

Treća pretpostavka govori o uvjetnoj varijanci grešaka, tj. ako znamo da vrijedi pretpostavka 1., onda ova pretpostavka kaže da je varijanca greške uvjetovana na prošlost, sadašnjost i budućnosti nezavisne varijable, te na α_i konstantna i jednaka σ_ϵ^2.

Izazov za procjenu je prisutnost N individualnih efekata koji se povećavaju kako $N \to \infty$. Dodatan problem je što možemo imati i mali broj vremenskih trenutaka, pa je procjena za koeficijente α_i prilično nepouzdana. Iz razloga što takva slučajna varijabla sakuplja individualne efekte koji su ostali neopaženi u modelu konstantnih koeficijenata, ovaj model zovemo model fiksnih efekata. Procijeniti β možemo s metodom najmanjih kvadrata, ali problem je taj da u modelu možemo imati jako puno regresora što može biti numerički neatraktivno. Srećom imamo jednostavniji način za procjenu parametra β. Može se pokazati da potpuno isti procjenitelj za β je dobiven u regresiji koja se provodi u odstupanjima od pojedinačnih prosjeka. O procjeniteljima ćemo reći nešto više u posebnom poglavlju.

4.2.3 Model slučajnih efekata (Random effects model)

Uobičajeno je za pretpostaviti u regresijskoj analizi da svi faktori koji utječu na zavisnu varijablu, ali nisu uključeni kao regresori, mogu biti propisno sumirani kao slučajna greška. U našem slučaju to vodi do pretpostavke da je α_i slučajni faktor koji je nezavisan i jednako distribuiran preko svih jedinki. Tako da model slučajnih efekata izgleda ovako:

$$Y_{it} = \boldsymbol{X}_{it}^T \boldsymbol{\beta} + \alpha_i + \epsilon_{it} \tag{4.5}$$

gdje su $\epsilon_{it} \sim IID(0, \sigma_\epsilon^2), \quad \alpha_i \sim IID(0, \sigma_\alpha^2)$.

Kao i u prethodna dva modela parametri ove formule su slični. Dakle, i označava jedinke koje promatramo, a indeks t vremenske trenutke. Y_{it} je zavisna slučajna varijabla, dok za $\boldsymbol{X}_{it}$ kažemo da vektor nezavisnih slučajnih varijabli dimenzija K. β je vektor parametara dimenzija K. Sada dolazimo do parametara koji se razlikuju od ostala dva modela. α_i su slučajne varijable koje kao što smo na početku poglavlja naveli opisuju neopažene efekte, dok je ϵ_{it} slučajna varijabla kako smo je naveli u definiciji 4.4 i ona predstavlja grešku modela. Moramo još navesti da $\alpha_i + \epsilon_{it}$ možemo tretirati kao grešku modela koja se sastoji od dvije komponente, individualne određene komponente, na koju ne utječe vrijeme i komponenta za koju pretpostavljamo da je nekorelirana kroz vrijeme. Mi zapravo pokušavamo pronaći zajedničku specifičnost koju nije moguće opisati nezavisnim varijablama, pa ju tražimo slučajnom varijablom. Grešku modela možemo onda definirati kao:

$$u_{it} = \alpha_i + \epsilon_{it} \tag{4.6}$$

Više o definiranju ove greške možete vidjeti u [13]. Sada model (4.5) možemo napisati u ovom obliku kroz sve vremenske trenutke T:

$$\boldsymbol{Y}_i = \boldsymbol{X}_i^T \boldsymbol{\beta} + \boldsymbol{u}_i$$

gdje $\boldsymbol{u}_i$ može biti zapisana kao $\boldsymbol{u}_i = \alpha_i \boldsymbol{j}_T + \boldsymbol{\epsilon}_i$, gjd $\boldsymbol{j}_T$ je vektor jedinica dimenzije T. Sada možemo definirati i matricu kovarijanci za $\boldsymbol{u}_i$ na sljedeći način:

$$\boldsymbol{\Omega} = E[\boldsymbol{u}_i \boldsymbol{u}_i^T]$$

To je matrica dimenzija $T \times T$ i za nju pretpostavljamo da je pozitivno definitna. Napomenimo još da je ova matrica nužno ista za svaki i. Više o elementima ove matrice i kako se dobivaju ćemo reći u nastavku kroz pretpostavke ovog modela. Kao i kod modela fiksnih efekata i ovdje imamo neke pretpostavke. One su:

1. $E[\epsilon_{it} | X_{i1}, ..., X_{iT}, \alpha_i] = 0, \ \ \forall i = 1, ..., N$ **(Svojstvo stroge egzogenosti)**
2. $E[\alpha_i | X_{i1}, ..., X_{iT}] = E[\alpha_i] = 0, \ \ \forall i = 1, ..., N$
3. $r(E[\boldsymbol{X}_i^T \boldsymbol{\Omega}^{-1} \boldsymbol{X}_i)] = K, \forall i = 1, \ldots, N,$
4. $E[\epsilon_{it}^2 | X_{i1}, ..., X_{iT}, \alpha_i] = \sigma_\epsilon^2, \ \ \forall i = 1, ..., N, \ \ \forall t = 1, \ldots, T$
5. $E[\alpha_i^2 | X_{i1}, ..., X_{iT}] = \sigma_\alpha^2, \ \ \forall i = 1, ..., N$

Vidimo da se neke pretpostavke identične onima iz pretpostavki o modelu fiksnih efekata. Prva pretpostavka je jedna od njih, tvz. svojstvo stroge egzogenosti i ona govori da je očekivanje greške modela uvjetno na nezavisne varijable u svim vremenskim trenucima i slučajnu varijablu α_i jednako nuli.

Druga pretpostavka daje razliku između modela fiksnih efekata i modela slučajnih efekata. Kod modela fiksnih efekata ne znamo ništa o očekivanju slučajnih varijabli α_i uvjetovanih na nezavisne varijable, dok kod modela slučajnih efekata takvo je očekivanje jednako nuli. Jednakost pretpostavke 2. govori da je α_i nezavisna od vektora nezavisnih varijabli u prošlosti, sadašnjosti i budućnosti. Iz tog razloga u modelu slučajnih efekata imamo nekoreliranost grešaka i nezavisnih varijabli, dok kod modela fiksnih efekata dopuštamo da greška bude korelirana sa slučajnom varijablom koja opisuje individualni neopaženi efekt.

Kod treće pretpostavke govorimo o rangu matrice. Prva pretpostavka koju ćemo koristiti je da ϵ_{it} ima konstantnu varijancu za sve t, tj. $E[\epsilon_{it}^2] = \sigma_\epsilon, \forall t = 1, \dots T$. Druga pretpostavaka koju navodi [13] kaže da možemo pretpostaviti da je greška modela serijalno nekorelirana, tj. $E[\epsilon_{it}\epsilon_{is}] = 0, \quad \forall t \neq s$. Koristeći te dvije pretpostavke možemo izvesti varijance i kovarijance elemenata od $\boldsymbol{u}_i$. Pod pretpostavkom 1. znamo da vrijedi da je $E[\epsilon_{it}\alpha_i] = 0, \quad \forall t = 1, \dots, T$. (Vidi [13]) Iz toga slijedi:

$$E[u_{it}^2] = E[\alpha_i^2] + 2E[\alpha_i\epsilon_{it}] + E[\epsilon_{it}^2] = \sigma_\alpha^2 + \sigma_\epsilon^2$$

gdje je $\sigma_\alpha^2 = E[\alpha_i^2]$. Također vrijedi $\forall t \neq s$,

$$E[u_{it}u_{is}] = E[(\alpha_i + \epsilon_{it})(\alpha_i + \epsilon_{is})] = E[\alpha_i^2] = \sigma_\alpha^2$$

Ako to napišemo preko kovarijance, onda to izgleda ovako:

$$Cov[(\alpha_i + \epsilon_{it})(\alpha_i + \epsilon_{is})] = \begin{cases} \sigma_\alpha^2, & t \neq s \\ \sigma_\alpha^2 + \sigma_\epsilon^2, & t = s \end{cases}$$

Ako znamo da vrijede sve pretpostavke koje smo dosad rekli, onda matrica $\boldsymbol{\Omega}$ dobiva poseban oblik

$$\Omega = \begin{bmatrix} \sigma_\alpha^2 + \sigma_\epsilon^2 & \dots & \sigma_\alpha^2 & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_\epsilon^2 & \dots & \sigma_\alpha^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \dots & \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_\epsilon^2 \end{bmatrix}$$

Sada odmah možemo i odrediti korelaciju koja na temelju kovarijance koja je definirana matricom Ω. Nju možemo zapisati na sljedeći način

$$Corr[u_{it}, u_{is}] = \begin{cases} \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\epsilon^2}, & t \neq s \\ 1, & t = s \end{cases}$$

Matrica koju dobijemo kao produkt tri matrica iz pretpostavke 3. će biti dimenzija $K \times K$ pa je i u ovom slučaju pretpostavka je da je matrica definirana u pretpostavci 3 punog ranga. (Vidi [13]).

Kod četvrte pretpostavke imamo isti slučaj kao i kod modela fiksnih efekata. Ona govori da su varijance greške uvjetno na nezavisne slučajne varijable i slučajnu varijablu α_i, konstantne.
Peta pretpostavka govori o uvjetnoj varijanci od α_i, za koju se pretpostavlja da je konstantna uvjetno na nezavisne slučajne varijable u svim vremenskim trenucima.

Kroz ova tri poglavlja smo vidjeli najbitnije modele za analizu panel podataka. U svakom slučaju nije jednostavno odlučiti koji je model prikladniji za uporabu. Možemo reći da se u praksi model konstantnih koeficijenata gotovo nikad ne koristi. Uvijek se bira između modela fiksnih efekata i modela slučajnih efekata [11]. Koji ćemo koristiti i koji je prikladniji ovisi o tome kakvim panel podacima raspolažemo. Model fiksnih efekata opisuje svaku jedinku i on je dobar ako želimo informacije o svakoj takvoj jedinki jer je ona specifična sama za sebe. No ukoliko svaku jedinku možemo na neki način smatrati slučajno odabranim reprezentantom cijele populacije i sama po sebi se ne razlikuje puno od ostalih jedinki po svim svojim osobinama, onda bi prikladniji bio model slučajnih efekata. No ovakva pravila su prilično generalizirana i ne možemo sa sigurnošću tvrditi da će ta pravila uvijek vrijediti. Kad bi gledali što se više koristi u praksi, ekonomisti će reći da oni više koriste model slučajnih efekata. To je iz razloga što se pojedinačni utjecaji mogu staviti pod jednakodistribuiranu nezavisnu varijablu, dok kod modela fiksnih efekata imamo previše koeficijenata, pogotovo ako se radi o ekonometrijskoj analizi gdje imamo veliki broj financijskih koeficijenata. Kao što vidimo iz navedenoga, nije lako odlučiti se koji model koristiti. Zato će nam u tom slučaju najviše pomoći Hausmanov test, o kojem ćemo detaljnije u poglavlju o statističkim testovima za panel modeliranje.

4.3 Procjenitelji parametara statičkih panel modela

U ovom poglavlju ćemo pogledati koji se sve procjenitelji parametara statičkih panel modela koriste i analizirati ćemo za koji su model oni adekvatni.

4.3.1 OLS procjenitelj parametara za model konstantnih koeficijenata

Formula za model konstantnih koeficijenata je

$$Y_{it} = \alpha + \boldsymbol{X}_{it}^T \boldsymbol{\beta} + u_{it}, \quad i = 1, ..., N, \quad t = 1, ..., T$$

U statistici se taj model naziva i populacijsko-prosječni model. Ovim načinom procjene se promatraju panel podaci kroz sve godine i individue. Na taj način imamo $N \times T$ promatranja, što za slučaj kad imamo veliku bazu panel podataka nije praktično, stoga se ta procjena uvodi kada imamo malo vremenskih trenutaka, tj. kada promatramo recimo poduzeća ili ljude u nekom kratkom vremenskom razdoblju.

Praktično nam je da uvedemo matričnu notaciju koju navodi [13]. Vrijede i pretpostavke za OLS procjenitelje o kojima više možete vidjeti u [13], te sve pretpostavke koje smo naveli u 4.2.1. Definiramo prvo općeniti OLS procjenitelj za parametar $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = (N^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i^T \boldsymbol{X}_i)^{-1} (N^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i^T \boldsymbol{Y}_i)$$

Za računanje procjenitelja, ponekad nam je bolje koristiti matričnu notaciju, pa ovaj izraz možemo zapisati na sljedeći način:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{Y}$$

gdje je $\boldsymbol{X} = (\boldsymbol{X}_1^T, \ldots, \boldsymbol{X}_N^T)^T$ matrica dimenzije $NT \times K$ i $\boldsymbol{Y} = (Y_1^T, \ldots, Y_N^T)^T$ je vektor dimenzije NT.
Sada možemo vidjeti kako izgleda OLS procjenitelj paramatara za model konstantnih koeficijenata.Pogledajmo sljedeće izraze:

$$\sum_{i=1}^{N} \boldsymbol{X}_i^T \boldsymbol{X}_i = \sum_{i=1}^{N} \sum_{t=1}^{T} \boldsymbol{X}_{it}^T \boldsymbol{X}_{it}, \quad \sum_{i=1}^{N} \boldsymbol{X}_i^T \boldsymbol{Y}_i = \sum_{1=1}^{N} \sum_{t=1}^{T} \boldsymbol{X}_{it}^T Y_{it}$$

Sada možemo zapisati OLS procjenitelj na sljedeći način:

$$\hat{\boldsymbol{\beta}} = (\sum_{i=1}^{N} \sum_{t=1}^{T} \boldsymbol{X}_{it}^T \boldsymbol{X}_{it})^{-1} (\sum_{1=1}^{N} \sum_{t=1}^{T} \boldsymbol{X}_{it}^T Y_{it}) \tag{4.7}$$

Procjenitelj iz (4.7) koristi se za iste jedinke uzrokovane kroz različite vremenske trenutke. [13] navodi da je OLS procjenitelj konzistentan ako vrijedi uvjet da je $E[\boldsymbol{X}_{it}\epsilon_{it}] = 0, t = 1, \ldots T$ i uz pretpostavku 2. iz poglavlja 4.2.1. OLS procjenitelj pod uvjetima navedenim u [5] je konzistentan i asimptotski normalan. Međutim, u panel okruženju malo je vjerojatno da je $\boldsymbol{\Omega} = \sigma^2 \boldsymbol{I}_{NT}$ tako da OLS ne mora nužno biti efikasan osim možda u nekim specijalnim slučajevima kad su regresori invarijantni na vrijeme. Ako je $Cov[u_{it}, x_{it}] = 0$ onda je za konzistentnost procjenitelja dovoljno da vrijedi da $NT \to \infty$. Odmah vidimo da model konstantnih koeficijenata zadovoljava taj uvjet jer imamo pretpostavku da su nezavisne varijable nekorelirane s greškom. Također i u slučaju modela slučajnih efekata će ovaj procjenitelj biti konzistentan jer smo rekli da je taj model jedna varijanta modela konstantnih koeficijenata.(Vidi [13])

No ako pogledamo model fiksnih efekata vidjet ćemo da će ovaj procjenitelj biti nekozništentan. Ako ga zapišemo kao

$$Y_{it} = \alpha + \boldsymbol{X}_{it}^T \boldsymbol{\beta} + (\alpha_i - \alpha + \epsilon_{it})$$

OLS procjenitelj može biti nekonzistentan jer imamo slučaj da α_i može potencijalno biti koreliran s vektorom nezavisnih varijabli kako smo u definiciji tog modela i rekli. Iz toga slijedi da izraz u zagradi koji predstavlja mješovitu grešku modela je potencijalno koreliran s nezavisnim varijablama, pa iz toga odmah slijedi da parametar β neće biti konzistentan. (Vidi [5])

4.3.2 Procjenitelji parametara modela fiksnih efekata

Formule za model fiksnih efekata iznosi

$$Y_{it} = \alpha_i + \boldsymbol{X}_{it}^T \boldsymbol{\beta} + \epsilon_{it}$$

Izazov za procjenu se javlja zbog prisutnosti N individualno specifičnih efekata koji se povećavaju kako N ide prema beskonačnosti. Za ovaj model postoji nekoliko načina da se konzistentno procijeniti $\boldsymbol{\beta}$. Jedan od njih naveli smo u nastavku poglavlja.
Procjenitelj parametara fiksnih efekata je dobiven oduzimanjem vremenski prosječnog modela $\bar{Y}_i = \alpha_i + \bar{\boldsymbol{X}}_i^T \boldsymbol{\beta} + \bar{\epsilon}_i$ od originalnoga modela. Dobivamo izraz

$$Y_{it} - \bar{Y}_i = (\boldsymbol{X}_{it} - \bar{\boldsymbol{X}}_i)^T \boldsymbol{\beta} + (\epsilon_{it} - \bar{\epsilon}_i), \tag{4.8}$$

Kao rezultat oduzimanja slobodni član α_i eliminiran. Koristeći metodu najmanjih kvadrata dobivamo procjenitelj parametara za model fiksnih efekata, gdje oznaka FE označava da se radi o modelu fiksnih efekata, izražen sljedećom formulom

$$\hat{\beta}_{FE} = \left[\sum_{i=1}^{N} \sum_{t=1}^{T} (\boldsymbol{X}_{it} - \bar{\boldsymbol{X}}_i)(\boldsymbol{X}_{it} - \bar{\boldsymbol{X}}_i)^T \right]^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (\boldsymbol{X}_{it} - \bar{\boldsymbol{X}}_i)(Y_{it} - \bar{Y}_i) \tag{4.9}$$

Tradicionalan pristup, koji detaljnije možemo vidjeti u [13], je da α_i bude gledan kao parametar, te da se procjenjuje uz β. Jednu od mogućnosti procjene α_i kao parametra možete vidjeti u [13] i ona iznosi:

$$\hat{\alpha}_i = \bar{Y}_i - \bar{\boldsymbol{X}}_i^T \hat{\beta}_{FE}, \quad i = 1, ..., N$$

Možemo uočiti da je $\hat{\alpha}_i$ dobiven kao prosjek kroz vrijeme, pa iz te činjenice možemo zaključiti kako će takav procjenitelj biti konzistentan ako imamo veliki panel gdje $T \to \infty$. No ako gledamo kratke panele, onda taj procjenitelj nije konzistentan, ali u definiciji modela fiksnih efekata smo rekli da ti individualni efekti nisu bitni. Oni su zapravo pomoćni koeficijenti koji smetaju u procjeni pravog učinka nezavisnih varijabli, pa iz tog razloga njihova procjena ne igra značajnu ulogu. Najbitnije je ipak procijeniti vektor β koji zapravo govori o vezi nezavisnih varijabli sa zavisnom.

Kao i kod procjenitelja konstantnih efekata i ovdje je konzistentnost od β zadovoljena u smislu konvergencije $NT \to \infty$. Dodatno mora vrijediti

$$E[\epsilon_{it} - \bar{\epsilon i}|\boldsymbol{X}_{it} - \bar{\boldsymbol{X}}_i] = 0, \tag{4.10}$$

a dovoljan uvjet da vrijedi (4.9) koji navodi [13] je da vrijedi

$$E[\epsilon_{it}|X_{i1}, \ldots, X_{iT}] = 0, \tag{4.11}$$

što smo i pretpostavili kada smo definirali model fiksnih efekata. Po pretpostavkama koje smo naveli u prethodnom poglavlju ovaj procjenitelj je konzistentan i asimptotski normalan kada N teži u beskonačnost bez obzira koliki je broj vremenskih trenutaka T. Pa možemo zapisati da vrijedi sljedeće

$$\hat{\beta}_{FE} \overset{as}{\sim} \mathcal{N}(\boldsymbol{\beta}, V(\hat{\beta}_{FE})), \quad N \to \infty$$

pri čemu je $V(\hat{\beta}_{FE}) = \frac{\sigma_\epsilon^2}{N}\left(\sum_{i=1}^{T} E[\ddot{\boldsymbol{X}}_{\boldsymbol{it}}\ddot{\boldsymbol{X}}_{\boldsymbol{it}}^{\boldsymbol{T}}]\right)^{-1}$ matrica kovarijanci za procjenitelja $\hat{\beta}_{FE}$, gdje je $\hat{\beta}_{FE}$ oznaka za procjenitelja modela fiksnih efekata.

4.3.3 Procjenitelj parametara modela slučajnih efekata

Formula za model slučajnih efekata glasi:

$$Y_{it} = \alpha_i + \boldsymbol{X}_{it}^T\boldsymbol{\beta} + \epsilon_{it}$$

Model se može zapisati i na sljedeći način:

$$Y_{it} = \mu + \alpha_i + \boldsymbol{X}_{it}^T\boldsymbol{\beta} + \epsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T$$

gdje je neslučajan skalar μ dodan kako bi slučajni efekti mogli biti normalizirani tako da imaju očekivanje nula. Procjenitelj slučajnih efekata za μ i β je dopustivi GLS (Generalized Least Squares) procjenitelj od modela (4.12). Detaljnije o dopustivom GLS pogledajte [5]. Iako je OLS procjenitelj konstantnih koeficijenata učinkovit procjenitelj za slučajne efekte, sa dopustivim GLS procjeniteljem imamo bolju učinkovitost. On se dobije iz transformiranog modela. Vidi [5].

$$Y_{it} - \hat{\lambda}\bar{Y}_i = (1 - \hat{\lambda})\mu + (\boldsymbol{X}_{it} - \hat{\lambda}\bar{\boldsymbol{X}}_i)^T\boldsymbol{\beta} + v_{it} \tag{4.12}$$

Pri čemu je

$$v_{it} = (1 - \hat{\lambda})\alpha_i + (\epsilon_{it} - \hat{\lambda}\bar{\epsilon}_i)$$

Prema [5] $\hat{\lambda}$ je konzistentan za

$$\lambda = 1 - \frac{\sigma_\epsilon}{\sqrt{\sigma_\epsilon^2 + T\sigma_\alpha^2}}$$

Možemo dodatno uočiti da za $\hat{\lambda} = 0$ imamo procjenitelj fiksnih efekata, dok za $\hat{\lambda} = 1$ imamo OLS procjenjivanje kao kod modela konstantnih koeficijenata. Još možemo napomenuti i kada $T \to \infty$, onda $\lambda \to 1$.

U nastavku radi lakšeg snalaženja i da bi bili u skladu s literaturom [13] uvodimo oznaku indeksa RE što nam označava model slučajnih efekata(Random Effects).Ako koristimo formulu (4.12) onda GLS procjenitelj parametara modela slučajnih efekata možemo zapisati sljedećom formulom

$$\hat{\boldsymbol{\beta}}_{RE} = \left[\sum_{i=1}^{N} \sum_{t=1}^{T} (\boldsymbol{X}_{it} - \hat{\lambda}\bar{\boldsymbol{X}}_i)(\boldsymbol{X}_{it} - \hat{\lambda}\bar{\boldsymbol{X}}_i)^T \right]^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (\boldsymbol{X}_{it} - \hat{\lambda}\bar{\boldsymbol{X}}_i)(Y_{it} - \hat{\lambda}\bar{Y}_i) \tag{4.13}$$

Konzistentnost u smislu konvergencije zahtjeva da imamo $NT \to \infty$. Prema [13] ćemo pretpostaviti da imamo konzistentne procjenitelje za σ_ϵ^2 i σ_α^2. Sada možemo definirati procjenu za $\boldsymbol{\Omega}$:

$$\hat{\boldsymbol{\Omega}} = \hat{\boldsymbol{\sigma}}^{\mathbf{2}} \boldsymbol{I}_T + \hat{\sigma}_\epsilon^2 \boldsymbol{j}_T \boldsymbol{j}_T^T$$

Gdje su $\boldsymbol{I}_T$ i $\boldsymbol{j}_T$ definirani u 4.2.3. To je matrica dimenzija $T \times T$ i za nju pretpostavljamo da je pozitivno definitina. Procjenitelja parametara za model slučajnih efekata možemo zapisati način

$$\hat{\boldsymbol{\beta}}_{RE} = (\sum_{i=1}^{N} \boldsymbol{X}_i^T \hat{\boldsymbol{\Omega}}^{-1} \boldsymbol{X}_i)^{-1} (\sum_{i=1}^{N} \boldsymbol{X}_i^T \hat{\boldsymbol{\Omega}}^{-1} \boldsymbol{Y}) \tag{4.14}$$

Gdje su $\boldsymbol{X}_i$ i $\boldsymbol{\Omega}$ matrice definirane u 4.2.3. i 4.3.1. Procjenitelj je konzistentan ako ima asimptotski normalnu distribuciju, pa iz toga slijedi

$$\hat{\boldsymbol{\beta}}_{RE} \overset{as}{\sim} \mathcal{N}\left(\boldsymbol{\beta}, V\left(\hat{\boldsymbol{\beta}}_{RE}\right)\right)$$

pri čemu je matrica kovarijanci $V\left(\hat{\boldsymbol{\beta}}_{RE}\right)$ dana izrazom

$$V\left(\hat{\boldsymbol{\beta}}_{RE}\right) = \frac{\sigma_\epsilon^2}{N} \left(E[\boldsymbol{X}_i^T \Omega^{-1} \boldsymbol{X}_i] \right)^{-1}$$

4.3.4 Procjenitelj prvim diferencijama

Model prvih diferencija povezan je sa modelom fiksnih efekata i on se dobiva kada od početnog modela fiksnih efekata oduzmemo vrijednosti istog tog modela pomaknutog za jedan vremenski trenutak unatrag. To znači da od početnog modela

$$Y_{it} = \alpha_i + \boldsymbol{X}_{it}^T \boldsymbol{\beta} + \epsilon_{it}$$

oduzmemo model pomaknut za jedan vremenski trenutak unazad

$$Y_{i,t-1} = \alpha_i + \boldsymbol{X}_{i,t-1}^T \boldsymbol{\beta} + \epsilon_{i,t-1}.$$

Rezultat tog oduzimanja je

$$Y_{it} - Y_{i,t-1} = (\boldsymbol{X}_{it} - \boldsymbol{X}_{i,t-1})^T \boldsymbol{\beta} + (\epsilon_{it} - \epsilon_{i,t-1}).$$

Možemo odmah uočiti da smo tim oduzimanjem eliminirali individualne komponente α_i, a metodom najmanjih kvadrata dobivamo procjenitelj za parametar β koji je dan formulom

$$\hat{\beta}_{FD} = \left[\sum_{i=1}^{N} \sum_{t=2}^{T} (\boldsymbol{X}_{it} - \boldsymbol{X}_{i,t-1})(\boldsymbol{X}_{it} - \boldsymbol{X}_{i,t-1})^T \right]^{-1} \sum_{i=1}^{N} \sum_{t=2}^{T} (\boldsymbol{X}_{it} - \boldsymbol{X}_{i,t-1})(Y_{it} - Y_{i,t-1}) \tag{4.15}$$

gdje FD je oznaka za prvu diferenciju (First Difference). Sa ovakvim procjeniteljem imamo $N \times (T-1)$ promatranja. Prema [13] procjenitelj prvog diferenciranja će biti konzistentan zbog $E[(\boldsymbol{X}_{it} - \boldsymbol{X}_{i,t-1})(\epsilon_{it} - \epsilon_{i,t-1})] = 0$. Ustvari, stroga egzogenost je prisutna u jednadžbi $E[\epsilon_{it} - \epsilon_{i,t-1} | \boldsymbol{X}_{it} - \boldsymbol{X}_{i,t-1}] = 0$ što znači da je procjenitelj prvim diferenciranjem uvjetno nepristran na X. To je jači uvjet nego $E[\epsilon_{it} | \boldsymbol{X}_{it}]$, ali je slabiji od uvjeta kod procjenitelja fiksnih efekata (4.11).

4.4 Statistički testovi za panel modele

U ovom poglavlju ćemo definirati statističke testove za panel podatke. Oni pomažu u detekciji koji model je bolje koristiti, te koliko su kvalitetni modeli koje koristimo. Definirat ćemo nekoliko statističkih testova, te ćemo objasniti njihove glavne značajke. Testove o kojima ćemo govoriti u nastavku su **Hausmanov test**, te **testovi individualnih i vremenskih komponenti**.

4.4.1 Hausmanov test

Kada smo govorili o statičkim panel modelima rekli smo da će se u praksi uglavnom voditi bitka koristiti li model fiksnih efekata ili model slučajnih efekata. Ako pogledamo formulu, vidjet ćemo da su to isti modeli, no stvar je u tome kako tretiramo individualni koeficijent α_i. Teško je odgovoriti na pitanje tretirati li ga kao fiksnog ili slučajnog. Tako da možda u danom trenutku više odgovara da imamo koeficijente za svaku jedinku i na taj način da maksimalno smanjimo utjecaj nebitnih čimbenika u procijenjenim koeficijentima uz nezavisne varijable. Tada više odgovara model fiksnih efekata. No, ako se dogodi slučaj da imamo previše jedinki i ne želimo procjene za svaku jedinku, nego želimo sve pojedinačne učinke staviti pod jednu nezavisnu i jednako distribuiranu slučajnu varijablu, tada će bolje odgovarati model slučajnih efekata. U statističkoj analizi upravo zbog takvih problema, ne možemo lako izabrati model. Zato je potreban test koji bi nas riješio te dileme, tj. koji bi nam rekao koji je model prikladniji za pojedinačne efekte - model fiksnih efekata ili model slučajnih efekata. Test koji se može primijeniti u tu svrhu zove se **Hausmanov test**.

Hausmanov test uspoređuje dva modela - model fiksnih efekata i model slučajnih efekata. Hipoteze ovog testa su sljedeće:

$$H_0 : \text{Slučajna varijabla } \alpha_i \text{ je nekorelirana s } \boldsymbol{X}_{it}.$$
$$H_1 : \text{Slučajna varijabla } \alpha_i \text{ je korelirana s } \boldsymbol{X}_{it}.$$

Pod pretpostavkama za slučajne efekte definirane u 4.2.3. uzimamo slučaj gdje $\boldsymbol{X}_{it}$ sadrži samo vremenski promjenjive komponente kad su to jedini koeficijenti koje možemo procijeniti koristeći fiksne efekte. Tada je $V(\boldsymbol{\hat\beta_{FE}}) = \sigma_u^2[E[\boldsymbol{\ddot X_i^T \ddot X_i}]]^{-1}/N$ i $V(\boldsymbol{\hat\beta_{RE}}) = \sigma_u^2[E[\boldsymbol{\check X_i^T \check X_i}]]^{-1}/N$, gdje je t-ti red od $\boldsymbol{\ddot X_i} = \boldsymbol{X}_{it} - \boldsymbol{\bar X}_i$ i t-ti red od $\boldsymbol{\check X}_i = \boldsymbol{X}_{it} - \lambda\boldsymbol{\bar X}_i$, a V je oznaka za matricu kovarijanci. Sada prema [13] slijedi da:

$$E[\boldsymbol{\check X_i^T \check X_i}]] - E[\boldsymbol{\ddot X_i^T \ddot X_i}]] = (1-\lambda)TE[\boldsymbol{\bar X}_i^T \boldsymbol{\bar X}_i] \ , \qquad \lambda < 1$$

iz čega slijedi da je

$$[V\{\hat\beta_{FE}\}]^{-1} - [V\{\hat\beta_{RE}\}]^{-1}$$

pozitivno definitno, pa je onda i $V\{\hat\beta_{FE}\} - V\{\hat\beta_{RE}\}$ pozitivno definitno. Kako $\lambda \to 1$ kada $T \to \infty$ ovi izrazi nam pokazuju kako asimptotska varijanca od procjenitelja slučajnih efekata teži u varijance procjenitelja fiksnih efekata kako T postaje veći.

Test statistika Hausmanovog testa dana je sljedećim izrazom

$$\xi_H = (\hat\beta_{FE} - \hat\beta_{RE})^T[\hat V\{\hat\beta_{FE}\} - \hat V\{\hat\beta_{RE}\}]^{-1}(\hat\beta_{FE} - \hat\beta_{RE}).$$

$\hat{\beta}_{FE}$ i $\hat{\beta}_{RE}$ predstavljaju procijenjene vrijednosti za parametar β za slučajne i fiksne efekte, dok $\hat{V}$ predstavlja procijenjene matrice kovarijanci. Ako je nulta hipoteza istinita Hausmanova test statistika je asimptotski distribuirana kao χ^2_M pod pretpostavkama koje smo definirali za slučajne efekte u 4.2.3. Vidi [13]

Kod nulte hipoteze vrijedi $\hat{\beta}_{FE} - \hat{\beta}_{RE} \to 0$. To znači da pretpostavljamo da ne postoji značajna razlika u procijenjenim koeficijentima s dva različita procjenitelja. Odmah vidimo da je koeficijent slučajnih efekata bolji i efikasniji. Također, u tom slučaju dobivamo i konzistentnost oba procjenitelja. Ako pogledamo alternativnu hipotezu, vidjet ćemo da nemamo konzistentnost kod oba procjenitelja, tj. procjenitelj slučajnih efekata je nekonzistentan. Tako da odbacivanjem nulte hipoteze dolazimo do zaključka da ćemo odabrati model fiksnih efekata, dok pri odabiru pretpostavke nulte hipoteze naš model postaje model slučajnih efekata.

Hausmanov test je jedan od jako bitnih alata pri određivanju koji panel model nam bolje koristi, tj. kod kojeg panel modela imamo efikasniji rezultat. Još možemo i napomenuti da ako je nulta hipoteza istinita test statistika Hausmanovog testa ima asimptotsku hi-kvadrat distribuciju s K stupnjeva slobode.

4.4.2 Testovi individualnih i vremenskih komponenti

U ovom potpoglavlju ćemo vidjet ima li potrebe za modeliranjem više kompleksnih modela od modela konstantnih koeficijenata. Pokazati ćemo kako se provode testovi o postojanju dodatnih efekata koji nisu uključeni u takav model. Tu prvenstveno mislimo na nepromatrane efekte koje sakuplja slučajna varijabla kod modela slučajnih efekata ili na nepromatrane efekte individualnih ili vremenskih komponenti kod modela fiksnih efekata. Navodimo tri vrste testova, Wooldrigeov test za nepromatrane efekte, F test i Lagrangeov multiplikacijski test.

Wooldrigeov test za nepromatrane efekte

Ako u modelu slučajnih efekata vrijede sve pretpostavke, ali model zapravo ne sadrži nepromatrane efekte onda je model konstantnih koeficijenata efikasan i može se koristiti. Ovaj test će reći postoje li ti nepromatrani efekti. Ako takvih efekata nema onda je model konstantnih koeficijenata prikladan za panel podatke.

Wooldrigeov test za nepromatrane efekte. Nultu i alternativnu hipotezu možemo zapisati sljedećim izrazom

$$\begin{aligned} H_0 &: \sigma_\alpha = 0 \\ H_1 &: \sigma_\alpha \neq 0. \end{aligned}$$

Kako je definiran σ_α možemo vidjeti u [13]. Procjenitelj za σ_α skaliran sa $\sqrt{N}\ \hat{\sigma}_\alpha$ dan je sljedećim izrazom:

$$N^{-1/2} \sum_{i=1}^{N} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \hat{v}_{it} \hat{v}_{is}$$

gdje $\hat{v}_{it}$ predstavlja reziduale modela i oni su procjena za v_{it} što smatramo mješovitom greškom modela. Nultu hipotezu testiramo pomoću test statistike koju definiramo sljedećim izrazom:

$$T_W = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \hat{v}_{it} \hat{v}_{is}}{\left[\sum_{i=1}^{N} \left(\sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \hat{v}_{it} \hat{v}_{is} \right)^2 \right]^{1/2}}$$

Prema [13] za bilo koju distribuciju od v_{it}, $N^{-1/2} \sum_{i=1}^{N} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} v_{it} v_{is}$ ima asimptotski normalnu distribuciju(pod nultom hipotezom vrijedi da je v_{it} serijski nekorelirana). S tom test statistikom možemo odbaciti H_0 za negativne procjene od σ_α^2 iako su negativne procjene rijetke u praksi. Ova statistika može detektirati mnogo vrsta serijalnih korelacija u mješovitoj greški v_{it}, no odbacivanje nulte hipoteze se ne bi trebalo interpretirati tako da se implicira da struktura greške slučajnih efekata mora biti istinita. Što znači da kad gledamo hipoteze vidimo da odbacivanje nulte hipoteze ne govori da je model slučajnih efekata najbolji, nego samo pokazuje činjenicu da postoje nekakvi individualni efekti koje modelom konstantnih koeficijenata zanemarujemo. To je posljedica što za takvu test statistiku pretpostavljamo da je standardna normalna u asimptotskom smislu, a za mješovitu grešku pretpostavljamo serijsku nekoreliranost. Odbacivanjem nulte hipoteze ne govori da je model slučajnih efekata najprikladniji, nego samo otkriva da postoji učinak i nepromatrani efekti.

F test

Ovaj test će nam pomoći da ispitamo utjecaje individualnih, vremenskih ili oba utjecaja u podacima za model. U ovom testu uspoređujemo dva modela. To su model konstantnih koeficijenata dan izrazom (4.2) i model fiksnih efekata dan izrazom (4.3).Model fiksnih efekata je "veći" model. Oba utjecaja u podacima na model je najlakše testirati pomoću F statistike. Neka SSR_p i SSR_f budu kvadratne rezidualne sume iz modela konstantnih koeficijenata i modela fiksnih efekata. Tada test statistiku možemo zapisati na sljedećom formulom

$$F_n = \frac{(SSR_p - SSR_f)}{SSR_f} \frac{N(T-1) - K}{N-1} \tag{4.16}$$

Nulta hipoteza kod F testa izgleda ovako

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_{N-1} = 0$$

Iz nje možemo vidjet da su svi individualni koeficijenti jednaki nuli osim jednog, pa ćemo u tom slučaju koristiti model konstantnih koeficijenata. Ako odbacimo nultu hipotezu možemo zaključiti da je prikladnije koristiti model koji sadrži individualne efekte. To znači da odbacivanjem nulte hipoteze zaključujemo da nam je prikladnije koristiti model fiksnih efekata u odnosu na model konstantnih koeficijenata.

Lagrangeov multiplikacijski test

Ovim testom također testiramo prisutnost individualnih, vremenskih ili oba efekta. Kod ovog testa se koriste i reziduali iz modela konstantnih koeficijenata koje ćemo označiti sa e_{it}. Prvi korak ovog testiranja je da počnemo sa modelom konstantnih koeficijenata, te njega ugnijezdimo sa kompliciranijim modelom. Tada popularni model testiramo protiv generalnog modela da bi utvrdili jesmo li dobro odredili originalni model. Ne želimo procjenjivat kompliciraniji model ako nemamo značajne dokaze protiv restringiranog oblika modela. U praksi ovaj test daje primarnu važnost restringiranom obliku tog modela. Mi možda uopće nemamo namjeru procjenjivati generalni model ako je nulta hipoteza odbačena. Test statistiku koju koristimo za ovaj test definiramo izrazom:

$$LM = \frac{NT}{2(T-1)}\left[\frac{\sum_{i=1}^{N}\left(\sum_{t=1}^{T} e_{it}\right)^2}{\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}^2} - 1\right]^2$$

Hipoteze ovog testa ćemo definirati na sljedeći način

Lagrangeov multiplikacijski test sa test statistikom LM ima sljedeće hipoteze:

H_0 : Ne postoje individualni efekti.
H_1 : Prisutni su individualni efekti.

Napomenimo prvo da kada pogledamo test statistiku vidimo omjer sume kvadrata grešaka po jedinkama sa sumom kvadrata svih grešaka. Ako su te vrijednosti približno jednake onda će vrijednost test statistike bit približno nuli. Ako je test statistika blizu nuli nećemo odbaciti nultu hipotezu i tada nemamo potrebu uvoditi nikakve individualne komponente u model konstantnih koeficijenata. U slučaju odbacivanja nulte hipoteze, vidimo da su prisutni individualni efekti ili vremenski efekti. No, postoji dosta različitih Lagrangeovih testova za ovakvu analizu. Jedni od najpoznatijih su napravljeni od strane Breushca i Pagana, Kinga i Wua. Detaljnija analiza ovog testa se može pronaći i pročitati u [13].

Predviđanje rasta prodaje prerađivačkih poduzeća

U nastavku ovog diplomskog rada ćemo se baviti panel podacima koji se odnose na rast poduzeća prerađivačke industrije u Hrvatskoj. Mi želimo statistički analizirati rast prihoda od prodaje poduzeća i vidjet koje varijable su ključne kako bi uspješno procijenili rast nekog poduzeća. U ovom radu ćemo se fokusirati na rast prihoda od prodaje poduzeća iz prerađivačke industrije, tj. provest ćemo analizu i pokušati napraviti panel model za poduzeća iz navedene industrije. Moramo napomenuti da će zavisna varijabla za mjerenje rasta poduzeća biti njihova prodaja.

Rast poduzeća je tema koja je zanimljiva na mnogim istraživačkim područjima iz razloga što samo razumijevanje faktora koji izazivaju rast poduzeća su bitni za stratešku dimenziju nekog poduzeća. Sam pogled na rast poduzeća kroz godine nam može dati na uvid kako to poduzeće funkcionira. Radi li se o stabilnom poduzeću sa stalnim rastom ili je to poduzeće u gubitku. S tim saznanjima možemo procijeniti koliko je poslovanje tog poduzeća sigurno, te što bi se moglo popraviti ili poboljšati da bi to poduzeće ostvarilo veći rast.

Mnogi ekonomisti se slažu sa činjenicom da je najbitnije odrediti varijable koje utječu na rast. U dosadašnjim istraživanjima u ovom području koristili su se različiti financijski i ekonomski pokazatelji, te je za neke od njih utvrđeno da imaju značajan utjecaj na rast poduzeća.

Ako pogledamo istraživanja koja su radili ekonomisti i statističari na panel podacima vidimo da su uočili uzorak u bitnosti varijabli koje utječu na rast poduzeća. Neke od bitnih varijabli čiji se rast u poduzeću prati su imovina, broj zaposlenika, udio dionica, profit i prodaja [9]. Također [9] navodi da u se u većini slučajeva najčešće koristi prodaja kao pokazatelj rasta poduzeća. Istraživanje ukazuje da je nezavisna varijabla koja predstavlja starost firme dosta bitan faktor koji utječe na rast prodaje poduzeća[7], no mi u našim podacima nemamo uključenu starost firme nego ćemo se u ovom radu fokusirati na financijske omjere kao nezavisne varijable.

Nadalje, istraživanje iz članka [2] govori o još nekim nezavisnim varijablama koje su bitne za rast poduzeća. Istraživanja su pokazala da poduzeća u različitim industrijama različito rastu. Iako se javljaju mnogi problemi kod prikupljanja podataka od velikih poduzeća u industrijskom sektoru, [2] ipak navodi da je industrijski sektor bitan pokazatelj rasta poduzeća. Ovaj članak je rađen na uzorku od 11 tisuća poduzeća iz Švedske. Njihovi rezultati također spominju starost poduzeća kao bitan faktor za rast (starije firme rastu sporije). Veličina poduzeća je još jedan bitan faktor u rezultatima, kao i industrijski sektor koji smo već spomenuli, te lokacija poduzeća [3]. U našem radu ćemo se više fokusirati na financijske pokazatelje poduzeća i vidjeti na koji način oni utječu na prihod od prodaje, tj. rast poduzeća.

U članku [12] govori se o nekim problemima koji mogu nastati kada mjerimo rasta poduzeća. Kako vrijeme prolazi, tako se mijenja i dinamika istraživanja, pa se može dogoditi da ekonomisti dođu do različitih formula i saznanja iako svi istražuju rast poduzeća temeljen na prodaji. Na primjer, kada gledamo prediktore rasta, možemo očekivati da će oni biti različiti ako gledamo novo nastala poduzeća u odnosu ako gledamo poduzeća iz Fortune 500. Tako da moramo uzeti u obzir da se naši rezultati možda neće slagati sa rezultatima ovih istraživanja. Još možemo napomenuti da se naša baza sastoji od poduzeća koji su iz prerađivačke industrije, no o tome ćemo detaljnije u sljedećem poglavlju.

Spomenuli smo da se rast poduzeća proučava na način da gledamo utjecaj varijabli na sam rast. U nastavku ćemo navesti sve varijable s kojima raspolažemo, tj. sve financijske omjere, te ćemo ih ukratko objasniti.

Prodaja

Kao što smo u prethodnim odlomcima naveli, varijabla prodaja je jedna od zavisnih varijabli i ona služi za mjerenje rasta poduzeća. To je i logično, jer ako poduzeće ostvaruje sve veću prodaju, tj. prodaje sve više svojih zaliha ili usluga, dolazi do velikog kapitala koji mu omogućuje daljnji rast. Naravno, takva poduzeća onda imaju visoku likvidnost i neto obrtni kapital o kojima ćemo reći više u nastavku.

Financijski pokazatelji

U nastavku možemo vidjet tablicu sa svim financijskim omjerima s kojima ćemo raspolagati. Ukratko ćemo ih objasnit i reći zašto su oni bitni, te koji bi potencijalno mogli biti važni za naš model. Financijski omjeri su odnosi dviju ili više ekonomskih veličina. Oni se računaju radi sinteze i jednostavnijeg pregleda informacija o poslovanju poduzeća. Da bi mogli izračunati financijske omjere potreban je uvid u financijske izvještaje poduzeća. Financijske omjere najčešće koriste revizori, banke, te ulagači na financijskim tržištima, a i svi koji žele imati uvid u poslovanje određenog poduzeća. Oni se računaju na različite načine i možemo reći da ima puno smislenih financijskih pomjera, no možemo reći da se većina tih pokazatelja računa na isti način. Dodatan problem je usporedivost tih pokazatelja u odnosu na različite sektore gospodarstva. Uz to, ne možemo uvijek pretpostaviti da su financijski podaci i omjeri uvijek točni jer na podatke može utjecati ljudski faktor i uvijek je moguće da se dođe do pogreške. U tablici u nastavku možemo vidjeti sve financijske omjere koje smo koristili. Oni su preuzeti iz financijskih izvještaja poduzeća u obliku računa dobiti i gubitka, te bilance.

Oznaka financijskog omjera	Značenje
NematImovina	Nematerijalna imovina/Ukupna imovina
kImovkObv	Kratkoročna imovina / Kratkoročne obaveze
pPrihZapos	Poslovni prihodi/Broj zaposlenih
Zkz	Ukupne obveze/Imovina
Zdk	Ukupne obveze/Kapital
Blta	Dugoročne i kratkoročne obveze prema bankama i drugim financijskim institucijama/Ukupna imovina
Zkvf	Kapital/Imovina
ROSdg	Dobit nakon poreza/Prihod od prodaje (brojnik ili pozitivan ili negativan)
Lclnw	Tekuća pasiva/Kapital
Lubrl	Koeficijent ubrzane likvidnosti= (kratkotrajna imovina – zalihe)/ Kratkoročne obveze
Lkiui	Kratkotrajna imovina/Ukupna imovina
Zlongdca	Dugoročne obveze/Tekuća imovina
Aukupni	Koeficijenti obrtaja ukupne imovine = prihod/ukupna imovina
Adug	Koeficijent obrtaja dugotrajne imovine = ukupni prihod/dugotrajna imovina
Akrat	Koeficijent obrtaja kratkotrajne imovine =ukupni prihod/ kratkotrajna imovina
Anap1	Trajanje naplate potraživanja 1 =broj dana razdoblja /koeficijent obrta potraživanja
Akrdob	Kreditiranje dobavljača =broj dana razdoblja/(materijalni troškovi/obveze prema dobavljačima)
Azal1	Koeficijent obrta zaliha prihodi od prodaje/zalihe
Pnmpdg	Neto marža profita= (neto dobit ili gubitak/ ukupni prihodi)*100
Pnroadg	neto dobit ili gubitak/ukupna imovina) *100
Pnroedg	(neto dobit ili gubitak/vlastiti kapital)*100
Prearnta	Zadržana dobit/ukupna imovina
Ctrenl	Koeficijent trenutne likvidnosti=novac/kratkoročne obveze
ImpLATA	Uvoz u razdoblju/imovina
ExpdnTA	Izdaci za razvoj/imovina
CPLTA	Koncesije, patenti, licencije, robne i uslužne marke, softver i ostala prava/imovina
GWTA	Goodwill/imovina

Tablica 5.1: Tablica financijskih omjera

Omjer nematerijalne imovine i ukupne imovine

Omjer nematerijalne i ukupne imovine govori o tome koliko udio nematerijalne imovine je sadržan u ukupnoj imovini, tj. aktivi.

$$NematImovina = \frac{Nematerijalna\ imovina}{Ukupna\ imovina}$$

Prosječna vrijednost ovog omjera za prerađivačku industriju je za naše podatke 0.01. To znači da je vrlo mali udio nematerijalne imovine u ukupnoj imovini u prerađivačkoj industriji.

Koeficijent tekuće likvidnosti

Koeficijent tekuće likvidnosti (eng. current ratio) omjer je likvidnosti najvišeg stupnja jer u odnos dovodi pokriće i potrebe za kapitalom u roku od godine dana.

$$kImovkObv = \frac{Kratkorocna\ imovina}{Kratkorocne\ obaveze}$$

Ovaj omjer mjeri sposobnost poduzeća da podmiri svoje kratkoročne obveze.Smatra se kako koeficijent tekuće likvidnosti manji od 1,5 implicira mogućnost da poduzeće ostane bez sredstava za podmirenje kratkoročnih obveza. U našim podacima prosjek je 3.2.

Poslovni prihodi po broju zaposlenih

Ovaj omjer govori o tome koliko prihodi rastu u odnosu na broj zaposlenih, tj. koji je njihov omjer. Logično je za pretpostaviti da što je taj omjer veći to je bolje za poduzeće. A njegov omjer glasi :

$$pPrihZapos = \frac{Poslovni\ prihodi}{Broj\ zaposlenih}$$

Koeficijent zaduženosti

Koeficijent zaduženosti (eng. debt ratio, leverage ratio, indebtedness coefficient) pokazuje do koje mjere poduzeće koristi zaduživanje kao oblik financiranja, odnosno koji je postotak imovine nabavljen zaduživanjem.

$$Zkz = \frac{Ukupne\ obveze}{Ukupna\ imovina}$$

U pravilu bi vrijednost koeficijenta zaduženosti trebala biti 50% ili manja.

Odnos duga i kapitala

Gornja granica odnosa duga i glavnice je najčešće 2:1, s udjelom dugoročnog duga ne većim od jedne trećine.

$$Zdk = \frac{Ukupne\ obveze}{Kapital}$$

Visoka vrijednost ovog omjera ukazuje na moguće poteškoće pri vraćanju posuđenih sredstava i plaćanju kamata.

Omjer obveza prema financijskim institucijama i ukupne imovine

Ovaj koeficijent je sličan koeficijentu zaduženosti. Razlika je u tome što se ovaj koeficijent fokusira samo na zaduženja u financijskim institucijama.

$$Blta = \frac{Dugorocne\ i\ kratkorocne\ obveze\ prema\ financijskim\ institucijama}{Ukupna\ imovina}$$

Koeficijent vlastitog financiranja

Koeficijent vlastitog financiranja govori koliko je imovine financirano iz vlastitog kapitala (glavnice) a njegova vrijednost bi trebala biti veća od 50%.

$$Zkvf = \frac{Kapital}{Ukupna\ imovina}$$

Omjer neto dobiti i prihoda od prodaje

Ovaj koeficijent je vrlo sličan neto profitnoj marži. Razlika je u tome što ovdje gledamo samo prihod od prodaje dok u neto profitnoj marži gledamo ukupan prihod.

$$ROSdg = \frac{Neto\ dobit}{Prihod\ od\ prodaje}$$

Omjer kratkotrajnih obveza i kapitala

Ovaj omjer je vrlo sličan odnosu duga, tj. svih obaveza i kapitala koje smo već naveli.

$$Lclnw = \frac{Tekuce\ obveze}{Kapital}$$

Koeficijent ubrzane likvidnosti

Ovaj omjer govori o tome ima li poduzeće dovoljno kratkoročnih sredstva da podmiri dospjele obveze, a bez prodaje zaliha.

$$Lubrl = \frac{(Kratkotrajna\ imovina - Zalihe)}{Kratkorocne\ obveze}$$

Ukoliko industrijski prosjek nije poznat, tada je poželjna vrijednost koeficijenta ubrzane likvidnosti približno jednaka omjeru 1:1.

Omjer kratkotrajne imovine i ukupne imovine

Ovaj omjer samo govori koji je udio kratkotrajne imovine u ukupnoj imovini.

$$Lkiui = \frac{Kratkotrajna\ imovina}{Ukupna\ imovina}$$

Omjer dugoročnih obveza i kratkoročne imovine

Kod ovog omjera prosjek u našim podacima je 0.48. Ovaj omjer je izražen sljedećom formulom

$$Zlongdca = \frac{Dugorocne\ obveze}{Tekuca\ imovina}$$

Koeficijent obrtaja ukupne imovine

Koeficijent obrta ukupne imovine govori koliko puta se ukupna imovina tvrtke obrne u tijeku jedne godine, odnosno koliko tvrtka uspješno koristi imovinu s ciljem stvaranja prihoda.

$$Aukupni = \frac{Ukupni\ prihod}{Ukupna\ imovina}$$

Ovo je jedan od najčešće korištenih omjera aktivnosti koji izražava veličinu imovine potrebne za obavljanje stanovite razine prodaje ili obrnuto, kune iz prodaje koje donosi svaka kuna imovine.

Koeficijent obrtaja dugotrajne imovine

Obrt dugotrajne imovine predstavlja omjer ukupnih prihoda i fiksne imovine, a pokazuje koliko uspješno tvrtka koristi dugotrajnu imovinu s ciljem stvaranja prihoda.

$$Adug = \frac{Ukupni\ prihod}{Dugotrajna\ imovina}$$

Opadajući omjer može biti indikator preinvestiranja u postrojenja, opremu ili neku drugu stalnu imovinu.

Koeficijent obrtaja kratkotrajne imovine

Koeficijent obrta kratkotrajne imovine govori koliko puta se kratkotrajna imovina tvrtke obrne u tijeku jedne godine, odnosno on mjeri relativnu efikasnost kojom poduzeće rabi kratkotrajnu imovinu za stvaranje prihoda.

$$Akrat = \frac{Ukupni\ prihod}{Kratkotrajna\ imovina}$$

Trajanje naplate potraživanja

Ovaj omjer mjeri dužinu vremena potrebnog za pretvaranja prosječne prodaje u novac. Obično se računa na kvartalnoj ili godišnjoj razini. Niski koeficijenti ovog omjera ukazuju na to da poduzeće uspješno naplaćuje svoja potraživanja.

$$Anap1 = \frac{Broj\ dana\ razdoblja}{Koeficijent\ obrta\ potrazivanja}$$

Kreditiranje dobavljača

Kreditiranje dobavljača je mjeri dužinu vremena potrebnog za vraćanje kredita dobavljačima. I u ovom slučaju je bolje da ovaj omjer ima niske koeficijente.

$$Akrdob = \frac{Broj\ dana\ razdoblja}{(Materijalni\ troskovi/Obveze\ prema\ dobavljacima)}$$

Koeficijent obrta zaliha

Koeficijent obrta zaliha pokazuje koliko se puta zalihe obrnu u tijeku jedne godine. Obzirom da su zalihe najnelikvidniji oblik imovine, poželjna je veća vrijednost koeficijenta.

$$Azal1 = \frac{Prihodi\ od\ prodaje}{Zalihe}$$

Neto marža profita

Neto profitna marža ukazuje na sposobnost managementa u vođenju poduzeća. Odnos neto dobiti prema prodaji u biti izražava razinu odnosa trošak/cijena u poslovanju. Ona pokazuje koliko se neto dobiti ostvaruje na ukupnim prihodima.

$$Pnmpdg = \frac{Neto\ dobit}{Ukupni\ prihod} * 100$$

Stopa povrata imovine

Stopa povrata imovine (eng. return on assets) je indikator uspješnosti korištenja imovine u stvaranju dobiti. Ovaj omjer pokazuje kolika je intenzivnost imovine poduzeća.

$$Pnroadg = \frac{Neto\ dobit}{Ukupna\ imovina} * 100$$

Stopa povrata glavnice

Stopa povrata glavnice (eng. return on equity) predstavlja možda i najznačajniji omjer profitabilnosti. Pokazuje koliko novčanih jedinica dobiti poduzeće ostvaruje na jednu jedinicu vlastitog kapitala.

$$Pnroedg = \frac{Neto\ dobit}{Vlastiti\ kapital} * 100$$

Omjer zadržane dobiti i ukupne imovine

Ovaj omjer govori koji je omjer zadržane dobiti i ukupne imovine. Zadržana dobit služi za povećanje investicija, pa tako i povećanje ukupne imovine. Formula ovog omjera je sljedeća

$$Prearnta = \frac{Zadrzana\ dobit}{Ukupna\ imovina}$$

Koeficijent trenutne likvidnosti

Koeficijent trenutne likvidnosti se računa prema formuli:

$$Ctrenl = \frac{Novac}{Kratkorocne\ obveze}$$

Omjer uvoza u razdoblju i ukupne imovine

Ovaj omjer se računa po sljedećoj formuli:

$$ImpLATA = \frac{Uvoz\ u\ razdoblju}{Ukupna\ imovina}$$

Omjer izdataka za razvoj i ukupne imovine

Izdaci za razvoj su dio nematerijalne imovine, te nam ovaj omjer govori koliki je udio tih izdataka u ukupnoj imovini.

$$ExpdnTA = \frac{Izadci\ za\ razvoj}{Ukupna\ imovina}$$

Omjer koncesija, patenata,ostalih prava i ukupne imovine

Kao i kod prethodnog omjera i ovdje se radio o djelu nematerijalne imovine i gledanju njegovog udjela u ukupnoj imovini.

$$CPLTA = \frac{Koncesije, patenti\ i\ ostalo}{Ukupna\ imovina}$$

Omjer Goodwill-a i ukupne imovine

Goodwill je razlika između vrijednosti neke tvrtke na tržištu dionica u određenom vremenu i njezine računovodstveno utvrđene neto imovine. I on je također dio nematerijalne imovine i ovim omjerom gledamo njegov udio u ukupnoj imovini.

$$GWTA = \frac{Goodwill}{Ukupna\ imovina}$$

Svi ovi omjeri daju vrlo važne informacije za analizu poslovanja poduzeća, a i nama su bitni za razvijanje našeg modela. Jedna od bitnih varijabli koje su značajne za rast poduzeća su koeficijenti ukupne likvidnosti, te koeficijent trenutne likvidnosti. Za koeficijent trenutne likvidnosti [8] kaže da ima znatan utjecaj na brzo rastuća poduzeća, tj. ima pozitivnu korelaciju sa brzo rastućim poduzećima.

Nadalje, u istraživanju iz članka [10] autori rade tri vrste regresije i gledaju koji financijski omjeri najviše utječu na rast firme. Kod sve tri regresije omjeri profitabilnosti se pokazuju vrlo značajni. Stopa povrata imovine, te neto marža profita pokazuju negativnu odnosno pozitivnu jaku vezu sa zavisnom varijablom rasta. Također koeficijenti trajanja naplate potraživanja i koeficijent obrta zaliha se pokazuju bitnima u sve tri regresije. Što se tiče likvidnosti, autori navode da koeficijenti likvidnosti imaju pozitivan i statistički značajan utjecaj na rast, dok koeficijenti povezani sa zaduženosti pokazuju negativnu ali statistički značajnu vezu sa rastom u njihovom modelu. Detaljnije možete pročitati u [10].

Istraživanje rasta prodaje prerađivačkih poduzeća u Hrvatskoj

U ovom poglavlju ćemo se malo bolje upoznati s podacima koje smo koristili u diplomskom radu. Istraživanje za ovaj rad je provedeno na uzorku od 3839 poduzeća koji se nalaze u prerađivačkoj industriji. Poduzeća su u uzorak birana na način da su u obzir uzeta samo ona poduzeća koja imaju zadovoljene sve komponente. To znači da smo uzeli u obzir poduzeća koja imaju podatke za cijelo razdoblje za koje je bila rađena analiza i regresija. Vremensko razdoblje u kojem smo promatrali poduzeća je od 2010. - 2014. godine. Dakle, kroz svih 5 godina smo imali jednak broj poduzeća, tako da možemo reći da se ovdje radi o balansiranim panel podacima. Shodno Pravilniku o razvrstavanju poslovnih subjekata prema Nacionalnoj klasifikaciji djelatnosti iz 2007. odredili smo koja poduzeća pripadaju prerađivačkoj industriji. Podaci za ovaj diplomski rad su dobiveni kroz projekt Hrvatske zaklade za znanost "RAZVOJ I PRIMJENA MODELA PREDIKCIJE RASTA ZA MALA I SREDNJA PODUZEĆA U HRVATSKOJ " voditeljice prof.dr.sc. Nataša Šarlija.

Za poduzeća u uzorku na raspolaganju su nam bili godišnji financijski izvještaji u obliku financijskih omjera koje smo koristili kroz svih 5 godina. Svi financijski omjeri su navedeni u prethodnom poglavlju, a mi ćemo u ovom poglavlju pogledati kako izgleda deskriptivna statistika onih varijabli koje smo koristili kasnije u modelu i koje bi nam mogle biti zanimljive za promatranje.

6.1 Opis panel podataka i varijabli u istraživanju

Da bi se bolje snašli u tumačenju deskriptivne statistike nezavisnih varijabli u nastavku pogledajmo koju varijablu svaka oznaka u tablici predstavlja:

- $log(pPrihZapos + 1)$ - logaritam omjera poslovnih prihoda i broja zaposlenih
- $NematImovina$ - omjer nematerijalne imovine i ukupne imovine
- $log(Zkz + 1)$ - logaritam koeficijenta zaduženosti
- $korijen(Zdk)$ - treći korijen odnosa duga i kapitala
- $log(blta + 1)$ - logaritam omjera obveza prema financijskim institucijama i ukupne imovine

- $korijen(Lclnw)$ - treći korijen od omjera kratkotrajnih obveza i kapitala
- $log(Lubrl + 1)$ - logaritam koeficijenta ubrzane likvidnosti
- $log(Aukupni + 1)$ - logaritam koeficijenta obrtaja ukupne imovine
- $log(Anap1 + 1)$ - logaritam trajanja naplate potraživanja
- $log(Akrdob + 1)$ - logaritam kreditiranja dobavljača
- $log(Azal1 + 1)$ - logaritam koeficijenta obrta zaliha
- $korijen(Pnmpdg)$ - treći korijen od neto marže profita
- $korijen(Prearnta)$ - treći korijen stope povrata glavnice
- $log(GWTA + 1)$ - logaritam omjera goodwill-a i ukupne imovine

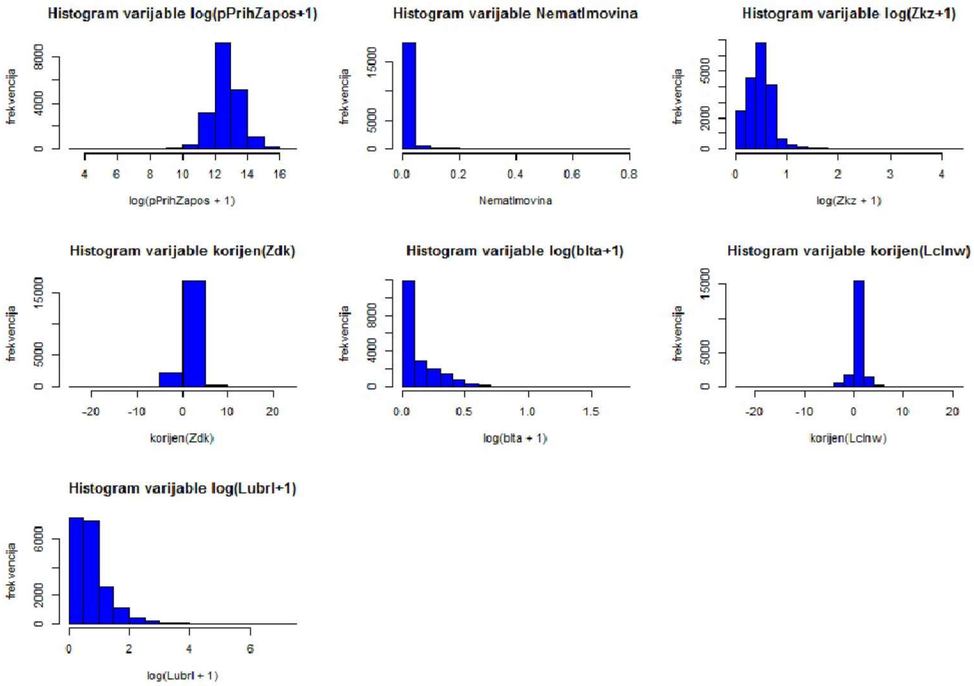
Za većinu varijabli koje ćemo koristiti u modelu koristili logaritamsku transformaciju ili transformaciju u treći korijen. To smo napravili iz razloga što smo kod tih varijabli uočili velike vrijednosti u podacima, pa da bi imali bolju kontrolu nad podacima napravili smo logaritamsku transformaciju za varijable koje imaju samo pozitivne vrijednosti, te transformaciju u treći korijen za varijable koje sadrže i pozitivne i negativne vrijednosti.

U tablici 6.1 možete vidjet deskriptivnu statistiku svih varijabli koje ćemo koristiti u modelu.

	Min	1st Qu.	Median	Mean	3rd Qu.	Max
log(pPrihZapos+1)	3,8290	12,1600	12,6700	12,7000	13,2000	16,9400
NematImovina	0,0000	0,0000	0,0000	0,0105	0,0014	0,7875
log(Zkz+1)	0,0004	0,3196	0,4837	0,4854	0,6166	4,3670
korijen(Zdk)	-22,8400	0,6837	1,0350	0,9323	1,4740	22,8300
log(blta+1)	0,0000	0,0000	0,0357	0,1105	0,1881	1,7570
korijen(Lclnw)	-22,8400	0,6079	0,9076	0,8307	1,3100	21,1500
log(Lubrl+1)	0,0000	0,3533	0,6132	0,7464	0,9641	7,4420
log(Aukupni+1)	0,0001	0,5225	0,7535	0,7943	1,0110	3,9080
log(Anap1+1)	0,0000	3,4970	4,2490	4,0120	4,7990	14,7500
log(Akrdob+1)	0,0000	3,6800	4,4240	4,3210	5,0540	12,7100
log(Azal1+1)	0,0001	1,3460	1,9790	2,2140	2,8300	9,8900
korijen(Pnmpdg)	-33,6000	0,5066	1,1130	0,6175	1,7150	4,5530
korijen(Prearnta)	-3,4320	0,0000	0,4523	0,2692	0,6520	1,2700
log(GWTA+1)	0,0000	0,0000	0,0000	0,0003391	0,0000	0,5709

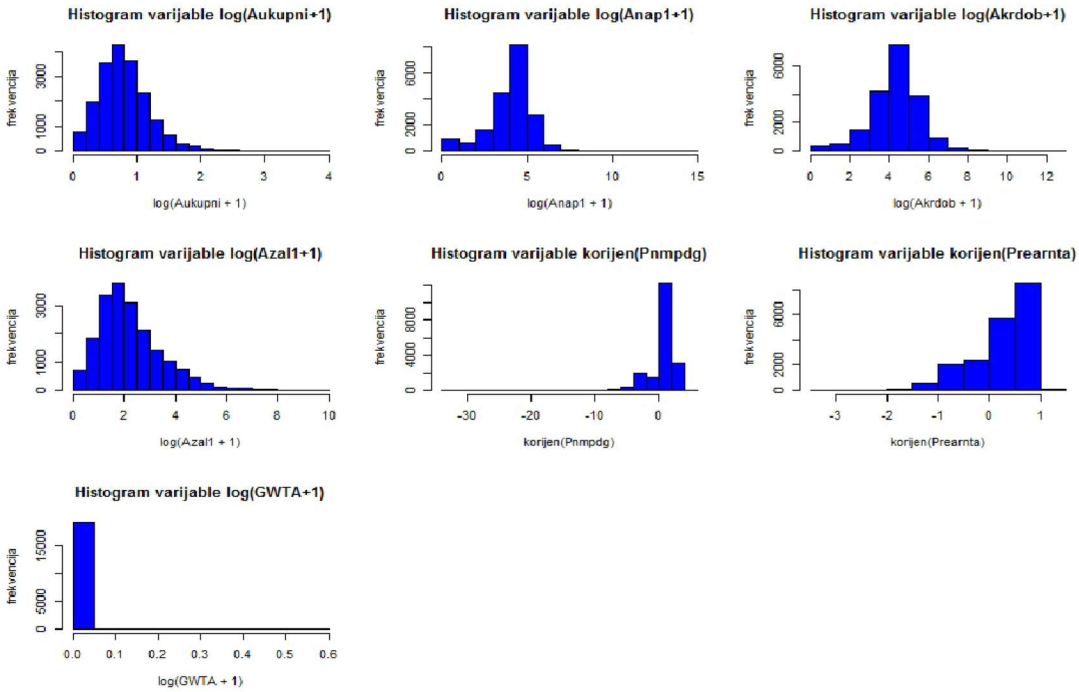
Tablica 6.1: Deskriptivna statistika nezavisnih varijabli

Dok u nastavku možemo vidjet histograme i kutijaste dijagrame ovih nezavisnih varijabli.



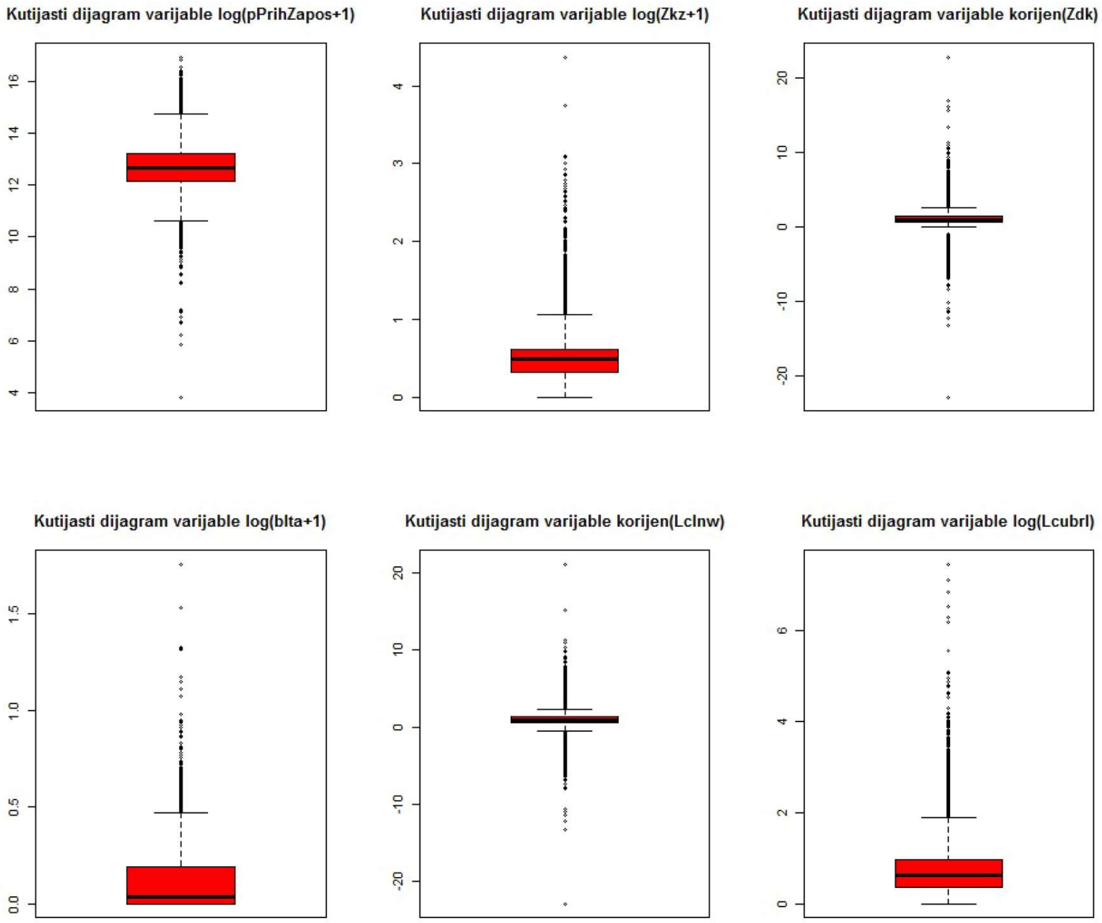
Slika 6.1: Histogrami nezavisnih varijabli

Na slici 6.1 možemo vidjet prvih 7 varijabli koje smo koristili u našem modelu. Jedino kod varijabli $log(pPrihZapos+1)$ i $korijen(Lclnw)$ vidimo naznaku zvonolikog oblika pa možda možemo naslutiti normalnost kod njih.



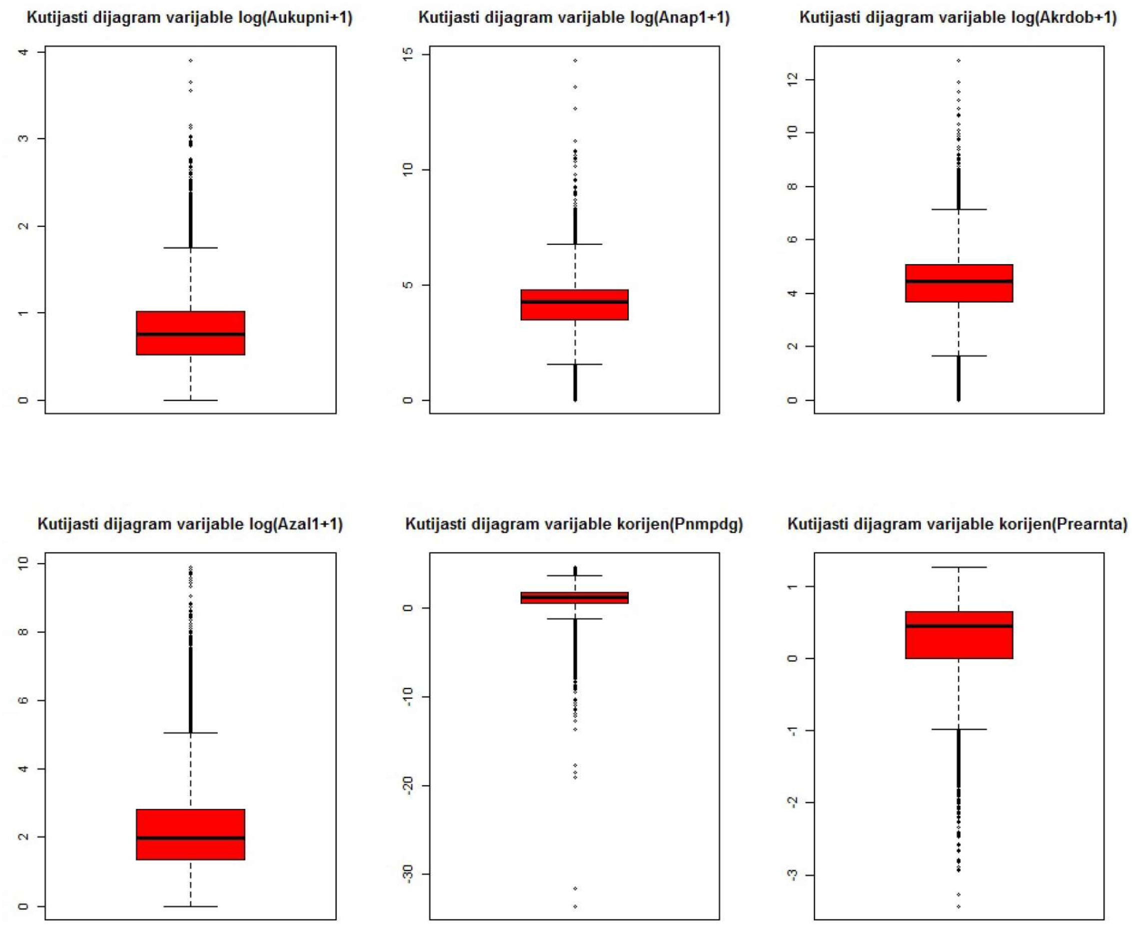
Slika 6.2: Histogrami nezavisnih varijabli-nastavak

Isto tako i na slici 6.2 možemo uočiti da neke varijable imaju zvonolik oblik, pa možemo naslutiti normalnost kod njih. Da bi još malo bolje shvatili i opisali naše nezavisne varijable potrebno je pogledamo i kutijaste dijagrame.



Slika 6.3: Kutijasti dijagrami nezavisnih varijabli

Na kutijastim dijagramima nezavisnih varijabli modela možemo uočiti i postojanje stršećih vrijednosti. Zbog takvih vrijednosti gubimo na kvaliteti modela, no nije isključeno da bismo napravili pogrešku njihovom eliminacijom iz panela jer ekonomski gledano oni mogu imati utemeljenost i logičnost, iako to otežava posao statističarima pri modeliranju modela. Dodatno možemo uočiti da imamo prilično raznovrsne kutijaste dijagrame, no nekako zajednička odlika im je što su komparativno kratki, što znači da puno poduzeća, izuzev stršećih vrijednosti, ima slične vrijednosti financijskih koeficijenata.



Slika 6.4: Kutijasti dijagrami nezavisnih varijabli - nastavak

Kada pogledamo i ostale kutijaste dijagrame nezavisnih varijabli možemo uočiti slične stvari. I kod njih imamo puno stršećih vrijednosti. Naizgled nisu toliko komparativno kratki, ali i kod njih možemo zaključiti da poduzeća imaju slične vrijednosti financijskih koeficijenata.

Naš panel se sastoji od 3839 poduzeća, što znači da imamo 19195 promatranja. Za sva poduzeća imamo popunjene sve podatke o financijskim koeficijentima, pa možemo reći da imamo slučaj balansiranih panel podataka. To nam je u svakom slučaju poželjno i model će zbog toga biti kvalitetniji, te ćemo moći koristiti više testova iz razloga što su neki statistički testovi isključivo za balansirane podatke.

6.2 Rezultati modeliranja prihoda poduzeća

Kroz ovaj diplomski rad nam je bio cilj vidjeti na koji način možemo modelirati panel podatke te pokušati napraviti reprezentativan model u kojem je varijabla PRODAJA, koja predstavlja prihod od prodaje, zavisna varijabla, dok su svi ostali financijski omjeri nezavisne varijable. Radimo na balansiranim podacima, tj. imamo jednak broj promatranja za sva poduzeća, pa će nam samim time model biti kvalitetniji.

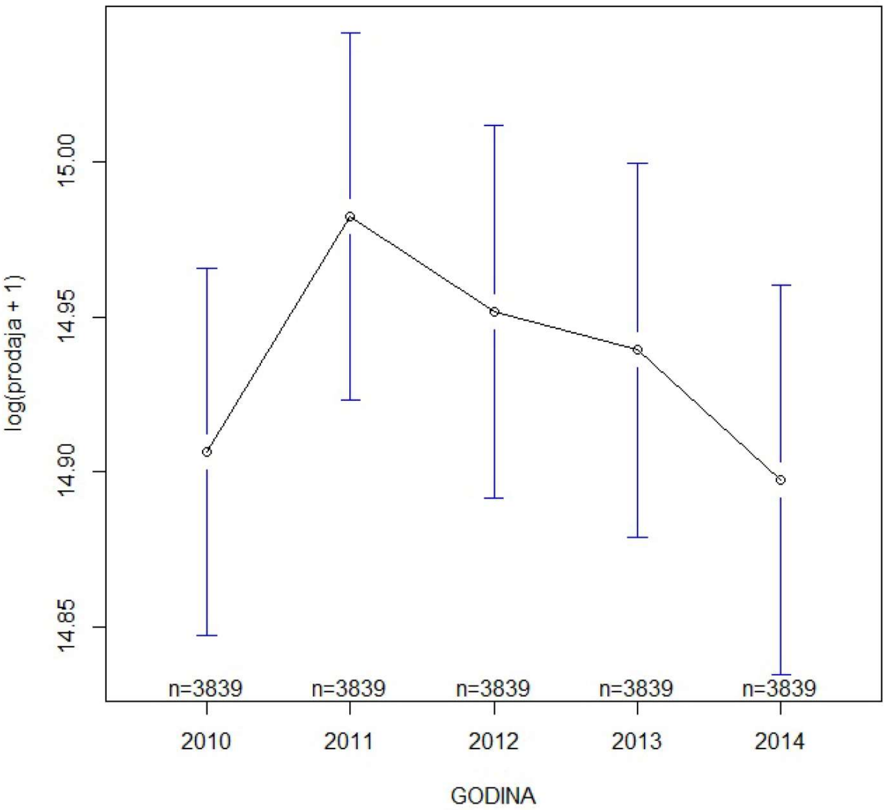
Kod samog modeliranja smo isprobavali različite varijante modela, u smislu da smo kombinirali različite nezavisne varijable. Isprobavali smo model sa uključenim svim varijablama, ali isto tako smo isprobavali modele sa manjim brojem varijabli. Neke varijable su se pokazale vrlo sličnima ili ih nije imalo smisla uvrstiti u naš model. Naravno, smanjivanjem broja varijabli smo izgubili na statističkoj značajnosti dobivenih koeficijenata. Na kraju smo odabrali 12 varijabli koje su pokazale da imaju najbolju statistički značajnu vezu sa zavisnom varijablom.

Za odabranih tih 12 varijabli napravili smo model konstantnih koeficijenata, model fiksnih efekata i model slučajnih efekata. Za početak smo proveli statističke testove koji će nam ukazati na opravdanost korištenje nekog od tih modela.

Prvo smo napravili model konstantnih koeficijenata nad podacima i dobili koeficijente s dosta dobrim statističkim značajnostima. Sve varijable osim slobodnog koeficijenta i varijable $NematImovina$ su pokazale da postoji statistička značajnost.

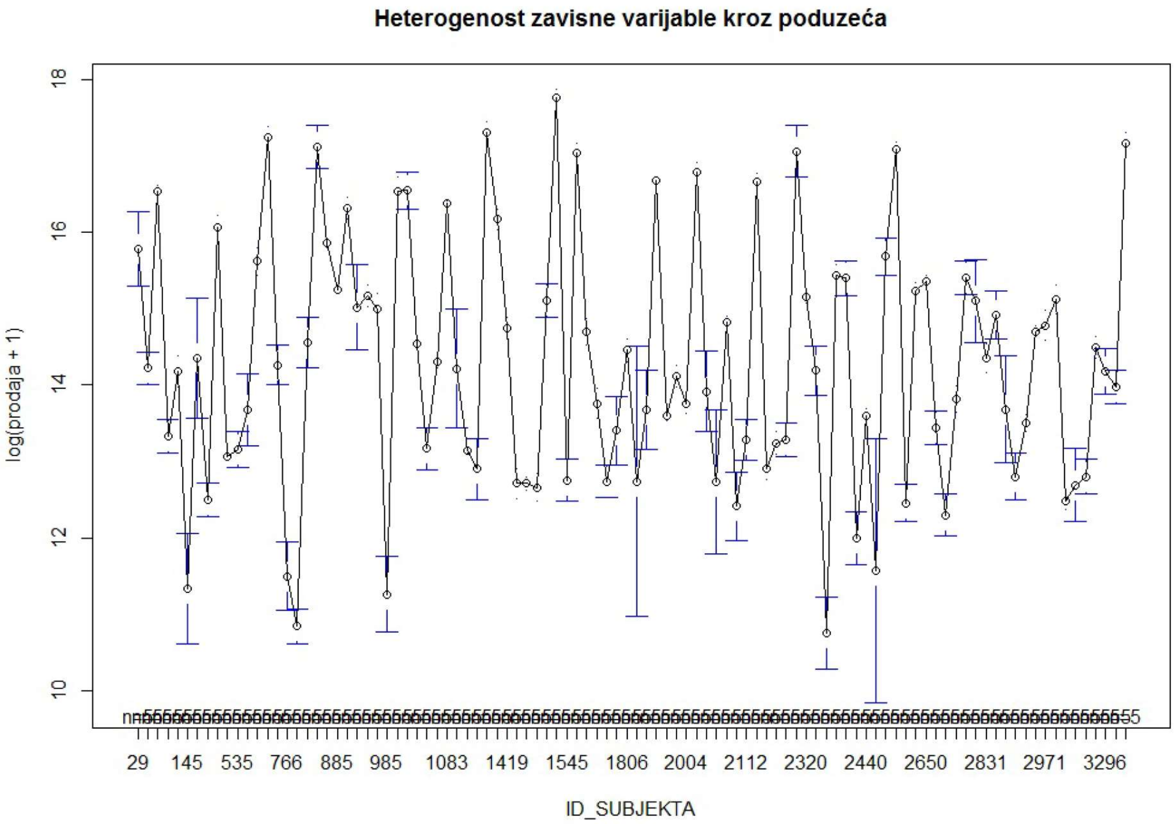
Nadalje, htjeli smo se uvjeriti postoji li bolji model od ovoga navedenoga, pa smo proveli dva F-test individualnih efekata koji uspoređuje model konstantnih koeficijenata sa modelom fiksnih efekata, odnosno model konstantnih koeficijenata sa modelom slučajnih efekata. Njime smo zapravo ispitivali postoji li značajan učinak pojedinačnih slobodnih koeficijenata koji bi pripadali pojedinom poduzeću ili pojedinom vremenskom razdoblju. Nulta hipoteza ovog testa je da ne postoje značajni individualni učinci i u tom slučaju je model konstantnih koeficijenata bolji od preostala navedena dva, a alternativna hipoteza govori da takvi efekti postoje, odnosno da model konstantnih koeficijenata nije dovoljno dobar model. Kod obadva testa smo dobili $p - vrijednost =< 2.2e^{-16}$ pa smo iz tog razloga odbacili nultu hipotezu i zaključili kako model konstantnih koeficijenata ipak nije dovoljno dobar model za nas.

Dodatno smo također proveli i Wooldrigeov test, koji smo opisali u poglavlju o testovima, te smo dobili $p - vrijednost = 0.00325$ što znači da odbacujemo nultu hipotezu koja kaže da nema nepromatranih efekata. Na taj način smo dobili potvrdu da model konstantnih koeficijenata nije najbolji odabir za nas jer postoje neki nepromatrani efekti koji nisu sakupljeni nezavisnim varijablama modela.



Slika 6.5: Heterogenost zavisne varijable kroz godine

Pogledamo li sliku 6.5 možemo uočiti izraženu heterogenost kroz vrijeme. To se slaže s rezidualima testova koje smo proveli. Na slici 6.7. također vidimo izraženu heterogenost, čak i izraženiju nego na prethodnoj slici pa zaključujemo da postoje individualni efekti u takvom modelu, te nam je sljedeći korak da pogledamo model slučajnih efekata, te model fiksnih efekata.



Slika 6.6: Heterogenost zavisne varijable kroz poduzeća

Napravili smo dva modela slučajnih efekata sa 15 i sa 13 varijabli. Za obadva modela smo proveli Hausmanov test te smo utvrdili da nije preporučljiv za korištenje umjesto modela fiksnih efekata. Nultu hipotezu smo odbacili jer nam je $p - vrijednost =< 2.2e^{-16}$. Također i kada smo proveli Hausmanov test na punom modelu slučajnih efekata sa svim varijablama, također smo utvrdili da nije preporučljiv za korištenje umjesto modela fiksnih efekata. Proveli smo još par Hausmanovih testova na različitim restrikcijama varijabli modela slučajnih efekata i uvijek smo odbacili nultu hipotezu iz čeka možemo zaključiti da je prikladan model za korištenje model fiksnih efekata.

Sljedeći korak nam je bio da odredimo ima li model fiksnih efekata nekakve individualne ili vremenske efekte. Način na koji smo to provjeravali je da smo u naš zadnji model uvrstili vrijeme, tj. godine kao prediktorsku varijablu, te smo s F-testom individualnih efekata testirali postoji li značajan učinak pojedinačnih slobodnih koeficijenata. Dobili smo $p - vrijednost =< 2.2e^{-16}$, te smo došli do zaključka kako postoje vremenski efekti u našem modelu, te da bi dobili kvalitetniji model ako uvrstimo vrijeme kao prediktorsku varijablu. Tu tvrdnju je potvrdio i Lagrange-ov multiplikacijiski test.

Kao što smo ranije pokazali da imamo heteroskedastičnost podataka, morali smo napraviti nešto da bi kontrolirali heteroskedastičnost i dobili bolju procjenu koeficijenata u modelu, te vidjeti kakve su statističke značajnosti. To smo napravili na način da smo koristili procjenu robusnim matricama kovarijanci. [6] navodi da prvo moramo provjeriti postoji li serijalna korelacija u našem modelu. Koristili smo Wooldrige-ov test za serijalne korelacije u panel podacima i dobili smo $p - vrijednost = 0.00003723$ pa na razini značajnosti od 0.05 možemo odbaciti nultu hipotezu koja govori da nemamo razloga sumnjati u postojanje serijalnih korelacija. Ako za model fiksnih efekata imamo heteroskedastičnost i serijalne korelacije, preporuča se korištenje "arellano" metode za procjenu. Vidi [6]

U tablic 6.7 prikazana su 4 glavna modela na kojima smo provodili modeliranje. U tablici smo ih nazvali redom: Pooled OLS koji predstavlja model konstantnih koeficijenata, Random Effects koji predstavlja model slučajnih efekata, Fixed Effect koji predstavlja model fiksnih efekata i Fixed Effects 2 koji predstavlja model fiksnih efekata u kojem smo izbacili varijable koje nisu pokazivale statistički značajnu vezu sa zavisnom varijablom, te u kojem smo ubacili faktor vremena kao prediktorsku varijablu . Ako pogledamo sve modele, vidimo da je za svaki model p-vrijednost bila reda veličine e^{-12}, što govori da su modeli bili valjani, tj. to je zapravo vrijednost F-testa koji ispituje jesu li svi koeficijenti jednaki nuli, a kako u našem slučaju je $p - vrijednost =< 2.2e^{-12}$, onda odbacujemo nultu hipotezu i prihvaćamo da su koeficijenti različiti od nule.

U tablici možemo uočiti da su R^2 vrijednosti između Fixed effects i Fixed effect 2 približno jednake, ali smo odabrali ovaj drugi model upravo zbog faktora godina koji je bitan u ovom modelu. Možemo vidjet i da su R^2 vrijednosti ostalih modela manje od našeg odabranoga. Moramo napomenuti da R^2 nije dovoljno velik, pa samim time ovaj model nije primjenjiv za predikciju. U tablici još možemo vidjet i totalne sume kvadrata(TSS) i rezidualne sume kvadrata(RSS).

	Pooled OLS		Random Effects		Fixed Effects		Fixed Effects 2	
Variable	**Estimate**	**Pr(>|t|)**	**Estimate**	**Pr(>|t|)**	**Estimate**	**Pr(>|t|)**	**Estimate**	**Pr(>|t|)**
Intercept	-0,2340015	0.175666	7,3421229	< 2.2e-16 ***	-	-	-	-
log(pPrihZapos + 1)	1,24086120	< 2.2e-16 ***	0,5878288	< 2.2e-16 ***	0,51177612	< 2.2e-16 ***	0,5090846	< 2.2e-16 ***
NematImovina	-0,1908556	0.419936	-0,345965	0.01298 *	-0,39442490	0.004147 **	-0,4197856	0.002255 **
log(Zkz + 1)	-2,11406350	< 2.2e-16 ***	-0,191727	1.089e-10 ***	-0,06591425	0.023145 *	-0,0757279	0.008650 **
3.korijen(Zkd)	0,2191651	1.351e-07 ***	0,028591	0.03195 *	0,01649214	0.198817	-	-
log(blta + 1)	3,06623070	< 2.2e-16 ***	0,630260	< 2.2e-16 ***	0,46841688	< 2.2e-16 ***	0,4522427	< 2.2e-16 ***
3.korijen(Lclnw)	-0,2218927	1.261e-06 ***	-0,013445	0.36158	-0,00059115	0.966770	-0,0680669	1.318e-10 ***
log(Lubrl + 1)	-0,1386244	3.157e-08 ***	-0,072880	3.948e-11 ***	-0,06972455	8.589e-11 ***	0,5140454	< 2.2e-16 ***
log(Aukupni + 1)	0,11692	0.001133 **	0,448132	< 2.2e-16 ***	0,52144659	< 2.2e-16 ***	-0,093995	< 2.2e-16 ***
log(Anap1 + 1)	-0,0925113	< 2.2e-16 ***	-0,087258	< 2.2e-16 ***	-0,09323914	< 2.2e-16 ***	-	-
log(Akrdob + 1)	0,1062724	< 2.2e-16 ***	-0,016064	8.387e-05 ***	-0,02178988	3.636e-08 ***	-0,0213299	7.119e-08 ***
log(Azal1 + 1)	0,04902	2.919e-07 ***	0,068538	< 2.2e-16 ***	0,0765538	< 2.2e-16 ***	0,0781874	< 2.2e-16 ***
korijen(Pnmpdg)	-0,0275735	2.105e-05 ***	0,0353056	< 2.2e-16 ***	0,04279742	< 2.2e-16 ***	0,0443608	< 2.2e-16 ***
korijen(Prearnta)	-0,360096	< 2.2e-16 ***	0,2986802	< 2.2e-16 ***	0,34094627	< 2.2e-16 ***	0,3449236	< 2.2e-16 ***
log(GWTA + 1)	7,7711282	3.600e-13 ***	3,1771230	8.376e-08 ***	2,7472415	2.472e-06 ***	2,7796166	1.828e-06 ***
Factor(GODINA)2011	-	-	-	-	-	-	0,0464253	8.642e-11 ***
Factor(GODINA)2012	-	-	-	-	-	-	0,0368376	2.895e-07 ***
Factor(GODINA)2013	-	-	-	-	-	-	0,0175738	0.014372 *
Factor(GODINA)2014	-	-	-	-	-	-	-0,0067469	0,348228
R^2	0,44862		0,47168		0,52424		0,52552	
R^2 ajd.	0,44822		0,47129		0,40479		0,40631	
p-value	< 2.22e-16		< 2.22e-16		< 2.22e-16		< 2.22e-16	
TSS	69421		3861,9		3146,6		3146,1	
RSS	38277		2040,3		1496,2		1492,7	

Tablica 6.7: Pregled svih modela

Kod procjene koeficijenata robusnim matricama kovarijanci mijenjaju se p-vrijednosti koje ukazuju na statistički značajnu vezu sa zavisnom varijablom, dok koeficijenti ostaju isti. U tablici 6.9 možemo vidjeti usporedbu dva modela Fixed effects 2 i Fixed effects 2 (Robust). Fixed effects 2 (Robust) je naš konačan model, čije smo koeficijente procijenili robusnim matricama kovarijanci te utvrdili statističke značajnosti. U tablici 6.9 vidimo koje varijable pokazuju statistički značajnu vezu sa zavisnom varijablom.

	Fixed Effects 2		Fixed Effects 2 (Robust)	
Variable	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)
log(pPrihZapos + 1)	0,5090846	< 2.2e-16 ***	0,5090846	< 2.2e-16 ***
NematImovina	-0,4197856	0.002255 **	-0,4197856	0,1515546
log(Zkz + 1)	-0,0757279	0.008650 **	-0,0757279	0,4561586
log(blta + 1)	0,4522427	< 2.2e-16 ***	0,4522427	< 2.2e-16 ***
log(Lubrl + 1)	-0,0680669	1.318e-10 ***	-0,0680669	1.318e-10 ***
log(Aukupni + 1)	0,5140454	< 2.2e-16 ***	0,5140454	< 2.2e-16 ***
log(Anap1 + 1)	-0,093995	< 2.2e-16 ***	-0,093995	< 2.2e-16 ***
log(Akrdob + 1)	-0,0213299	7.119e-08 ***	-0,0213299	0.0445315 *
log(Azal1 + 1)	0,0781874	< 2.2e-16 ***	0,0781874	< 2.2e-16 ***
korijen(Pnmpdg)	0,0443608	< 2.2e-16 ***	0,0443608	< 2.2e-16 ***
korijen(Prearnta)	0,3449236	< 2.2e-16 ***	0,3449236	< 2.2e-16 ***
log(GWTA + 1)	2,7796166	1.828e-06 ***	2,7796166	1.828e-06 ***
Factor(GODINA)2011	0,0464253	8.642e-11 ***	0,0464253	8.642e-11 ***
Factor(GODINA)2012	0,0368376	2.895e-07 ***	0,0368376	2.895e-07 ***
Factor(GODINA)2013	0,0175738	0.014372 *	0,0175738	0.0373344 *
Factor(GODINA)2014	-0,0067469	0,348228	-0,0067469	0,4945509

Tablica 6.8: Usporedba konačnog modela s Fixed Effect 2 (Robust) modelom

Odabrani model je:

$$\begin{aligned} log(prodaja+1) = 0,5090846 * log(pPrihZapos+1) + 0,4522427 * log(blta+1) \\ +(-0,0680669) * log(Lubrl+1) + 0,5140454 * log(Aukupni+1) \\ +(-0,093995) * log(Anap1+1) + (-0,0213299) * log(Akrdob+1) \\ +0,0781874 * log(Azal1+1) + 0,0443608 * korijen(Pnmpdg) \\ +0,3449236 * korijen(Prearnta) + 2,7796166 * log(GWTA+1) \\ +0,0464253 * \mathbb{1}_{t=2011} + 0,0368376 * \mathbb{1}_{t=2012} + 0,0175738 * \mathbb{1}_{t=2013} \end{aligned} \tag{6.1}$$

U nastavku ćemo usporediti dobivene rezultate s pretpostavkama o učinku nekih od varijabli rasta prodaje.

Ako usporedimo naš model sa pretpostavkama vidimo da imamo jedan financijski koeficijent koji je povezan sa likvidnosti. Logaritam koeficijenta ubrzane likvidnosti ima negativan učinak sa logaritmom prodaje kako je pretpostavljeno u ranije navedenim istraživanjima. Ako pogledamo koeficijente zaduženosti jedino omjer obveza prema financijskim institucijama i imovine pokazuje statistički značajnu vezu s prodajom, te u našem slučaju pokazuje pozitivan učinak, dok je u pretpostavkama rečeno da zaduženost ima negativan učinak. Međutim u [10] možemo naći objašnjenje da za poduzeća koja rastu korištenje zaduživanja kao oblik financiranja može biti ograničen, pa u tome možemo naći objašnjenje našeg pozitivnog učinka. Što se tiče pokazatelja profitabilnosti neto marža profita ima pozitivan učinak isto kako je navedeno i u pretpostavkama.Logaritam koeficijenta obrtaja zaliha te trajanje naplate potraživanja imaju pozitivan, odnosno negativan učinak, dok [10] govori samo da ta dva koeficijenta imaju značajan utjecaj na rast prodaje. Ostalo je još reći da prihod po broju zaposlenih pokazuje pozitivan učinak, a isti slučaj je i sa varijablom $log(GWTA+1)$ koja predstavlja omjer Goodwilla i ukupne imovine. Što se tiče godina kao prediktora, vidimo da 2011,2012, 2013 imaju pozitivan učinak u odnosu na baznu godinu 2010. To znači da u tim godinama dolazi do povećanja prihoda od prodaje u odnosu na baznu 2010. godinu, tj. bilježi se rast. Na kraju možemo zaključiti da naš model osim statističkog opravdanja ima i ekonomsko opravdanje te smo odabrali dobar model koji modelira prihod od prodaje poduzeća iz prerađivačke industrije na osnovu svih varijabli koje smo uvrstili u modeliranje.

Bibliografija

[1] J.D. Cryer and K-S Chan. *Time Series Analysis with applications in R.* Springer-Verlag, 2008.

[2] P. Davidsson, B. Kirchhoff, A. Hatemi–J, and H. Gustavsson. Empirical analysis of business growth factors using swedish data. *Journal of Small Business Management*, 40(4), 2002.

[3] F. Delmar, P Davidson, and W. Gartner. Arriving at the high growth firm. *Journal of Business Venturing*, 18(2), 2003.

[4] R. Durrett. *Essentials of Stochastic Processes.* Springer, 1999.

[5] Cheng Hsiao. *Analysis of Panel Data.* Cambridge University Press, 2003.

[6] Christian Kleiber and Achim Zeileis. *Applied Econometrics with R.* Springer, 2008.

[7] Miroslav Mateev and Anastasov Yanko. Determinants of small and medium sized fast growing enterprises in central and eastern europe: A panel data analysis. *Financial Theory and Practice*, 34(3), 2010.

[8] A. M. Moreno and J.C. Casillas. High-growth smes versus non-highgrowth smes: a discriminant analysis. *Entrepreneurship & Regional Development: An International Journal*, 19(1), 2007.

[9] Gabriele Sampagnaro. Predicting rapid growth smes through a reversal of credit scoring principles: a note.

[10] Gabriele Sampagnaro and Giuseppe Lubrano Lavadera. Identifying high growth smes through balance sheet ratios, 2013.

[11] Marno Verbeek. *A Guide to Modern Econometrics.* John Wiley Sons, 2008.

[12] L. G. Weinzimmer, P. C. Nystrom, and S. J. Freeman. Measuring organizational growth: Issues, consequences and guidelines. *Journal of Management*, 24(2), 1998.

[13] Jerey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data.* The MIT Press, 2010.

[14] Željko Pauša. *Uvod u matematičku statistiku.* Školska knjiga, 1993.

Sažetak

U radu smo predstavili posebnu strukturu podataka koju nazivamo panel podaci. Upoznali smo se s modelima koji se primjenjuju na takve podatke, njihovim pretpostavkama te procijeniteljima uz pomoć kojih dobivamo koeficijente svakog modela. Obradili smo teorijsku pozadinu najvažnijih statističkih testova koji služe za usmjeravanje pri modeliranju i ispitivanju ispravnosti pojedinog modela. Objasnili smo rast poduzeća temeljen na prihodu od prodaje što je vrlo zanimljivo za ekonomska istraživanja te smo se upoznali s osnovnim varijablama koje pokazuju vezu s rastom prihoda od prodaje. Detaljnije smo opisali sve financijske omjere koji su bile korištenje u modeliranju. Modelirali smo na primjeru od 3839 poduzeća iz prerađivačke industrije u razdoblju od 2010. do 2014. godine. Opisali smo koje financijske omjere sadrže te smo pogledali deskriptivnu statistiku. Nakon toga smo modelirali prihod od prodaje pomoću nezavisnih varijabli iz panela. U konačnom modelu, kojeg smo dobili primjenom postupaka napravljenih u teorijskom djelu rada izdvojeno je 13 nezavisnih varijabli, te smo taj model diskutirali u ekonomskom smislu.

Title and summary

Panel data model for sales growth
In this paper panel data are explained and used for modeling revenue growth from sales. Different models for panel data modeling are discussed as well as statistical tests which can be used for model selection. Revenue growth from sales and its variables are also explained, which is very interesting for economic research. The panel data for 3839 Croatian firms in manufacturing industry in the period form 2010. to 2014. was used for revenue growth modeling. The modeling was done with the program R by using estimators and statistical tests which are explained in the theoretical part of this paper.

Životopis

Rođen sam 9. srpnja 1991. godine u Virovitici. Osnovnu školu sam završio 2006. godine u Virovitici te potom upisao Prirodoslovno-matematičku gimnaziju u Gimnaziji Petra Preradovića u Virovitici. Za vrijeme osnovnoškolskog i srednjoškolskog obrazovanja sudjelovao sam na općinskih i županijskim natjecanjima iz matematike i kemije, te na sportskim natjecanjima iz rukometa i tenisa. 2010. godine upisujem Preddiplomski studij matematike na Prirodoslovno matematičkom fakultetu u Zagrebu. te ga završavam 2014. godine. Iste godine upisujem Diplomski studij matematike, smjer Financijska i poslovna matematika pri Odjelu za matematiku u Osijeku.

Dodatak - R kod

```
#paketi koje sam koristio
library(foreign)
library(plm)
library(mice)
library(VIM)
library(caret)
library(psych)
library(Hmisc)
library(nlme)
library(geoR)
library(car)
library(gplots)
library(tseries)
library(lmtest)

#Funkcija 3.korijen, jer sam neke podatke koji imaju velike negativne
vrijednosti transformirao s tom funkcijom
korijen <- function(x) {
    sign(x) * abs(x)^(1/3)
}


#Ucitavanje podataka i panela

podaci<-read.csv2("Preradjivacka_5.csv", header=T)
attach(podaci)
panel<-plm.data(podaci, index=c("ID_SUBJEKTA","GODINA"))

#Analiza varijabli, korelacije , histogrami, boxplot

summary(panel)

#histogrami, kod nekih histograma sam gledao outliere, csto da vidim
koliko ih ima i dali bi se mogli izbaciti.
par(mfrow=c(3,3))
hist(log(prodaja+1))
hist(NematImovina,col="blue1", main="Histogram varijable NematImovina", ylab="frekvencija")
hist(log(kImovkObv+1))
hist(log(pPrihZapos+1), col="blue1", main="Histogram varijable log(pPrihZapos+1)", ylab="frekvencija")
hist(log(Zkz+1),col="blue1", main="Histogram varijable log(Zkz+1)", ylab="frekvencija")
hist(korijen(Zdk),col="blue1", main="Histogram varijable korijen(Zdk)", ylab="frekvencija")
hist(log(blta+1),col="blue1", main="Histogram varijable log(blta+1)", ylab="frekvencija")
hist(korijen(Zkvf))
hist(korijen(ROSdg))
hist(korijen(Lclnw),col="blue1", main="Histogram varijable korijen(Lclnw)", ylab="frekvencija")
hist(log(Lubrl+1),col="blue1", main="Histogram varijable log(Lubrl+1)", ylab="frekvencija")
hist(Lkiui)
hist(log(Zlongdca))

hist(log(Aukupni+1),col="blue1", main="Histogram varijable log(Aukupni+1)", ylab="frekvencija")

hist(log(Adug+1))

hist(log(Akrat+1))

hist(log(Anap1+1),col="blue1", main="Histogram varijable log(Anap1+1)", ylab="frekvencija")

hist(log(Akrdob+1),col="blue1", main="Histogram varijable log(Akrdob+1)", ylab="frekvencija")

hist(log(Azal1+1),col="blue1", main="Histogram varijable log(Azal1+1)", ylab="frekvencija")

hist(korijen(Pnmpdg),col="blue1", main="Histogram varijable korijen(Pnmpdg)", ylab="frekvencija")

hist(korijen(Pnroadg))

hist(korijen(Prearnta),col="blue1", main="Histogram varijable korijen(Prearnta)", ylab="frekvencija")

hist(log(Ctrenl+1))

hist(log(ImpTA+1))

hist(log(ExpdnTA+1))

hist(log(CPLTA+1))

hist(log(GWTA+1),col="blue1", main="Histogram varijable log(GWTA+1)", ylab="frekvencija")

#boxplots
```

```
par(mfrow=c(2,3))
boxplot(log(prodaja+1))
boxplot(NematImovina,col="red", main="Kutijasti dijagram varijable NematImovina")

boxplot(log(kImovkObv+1)))

boxplot(log(pPrihZapos+1),col="red", main="Kutijasti dijagram varijable log(pPrihZapos+1)")

boxplot(log(Zkz+1),col="red", main="Kutijasti dijagram varijable log(Zkz+1)")

boxplot(korijen(Zdk),col="red", main="Kutijasti dijagram varijable korijen(Zdk)")

boxplot(log(blta+1),col="red", main="Kutijasti dijagram varijable log(blta+1)")

boxplot(korijen(Zkvf))

boxplot(korijen(ROSdg))
boxplot(korijen(Lclnw),col="red", main="Kutijasti dijagram varijable korijen(Lclnw)")
boxplot(log(Lubrl+1),col="red", main="Kutijasti dijagram varijable log(Lcubrl)")
boxplot(Lkiui)
boxplot(log(Zlongdca))
boxplot(log(Aukupni+1),col="red", main="Kutijasti dijagram varijable log(Aukupni+1)")
boxplot(log(Adug+1))
boxplot(log(Akrat+1))
boxplot(log(Anap1+1),col="red", main="Kutijasti dijagram varijable log(Anap1+1)")
boxplot(log(Akrdob+1), col="red", main="Kutijasti dijagram varijable log(Akrdob+1)")
boxplot(log(Azal1+1), col="red", main="Kutijasti dijagram varijable log(Azal1+1)")
boxplot(korijen(Pnmpdg),col="red", main="Kutijasti dijagram varijable korijen(Pnmpdg)")
boxplot(korijen(Prearnta),col="red", main="Kutijasti dijagram varijable korijen(Prearnta)")
boxplot(log(GWTA+1),col="red", main="Kutijasti dijagram varijable log(GWTA+1)")
boxplot(log(Ctrenl+1))


#model random i fixed effects

#RANDOM

#random model
random.model <- plm(log(prodaja+1)~log(pPrihZapos+1)+NematImovina+log(kImovkObv+1)+log(Zkz+1)+korijen(Zdk)+
log(blta+1)+korijen(Zkvf)+korijen(ROSdg)+korijen(Lclnw)+log(Lubrl+1)+log(Aukupni+1)+log(Akrat+1)+log(Anap1+1)
+log(Akrdob+1)+log(Azal1+1)+korijen(Pnmpdg)+korijen(Prearnta)+log(Ctrenl+1)+log(GWTA+1) ,index=c("ID_SUBJEKTA", "GODINA"
 data=panel, na.action=na.omit, model="random")
summary(random.model)

#model kod kojeg su izbacene:kimovkobv, rosdg, akrat, ctrenl

random.modelnovi <- plm(log(prodaja+1)~log(pPrihZapos+1)+NematImovina+log(Zkz+1)+korijen(Zdk)+
log(blta+1)+korijen(Zkvf)+korijen(Lclnw)+log(Lubrl+1)+log(Aukupni+1)+log(Anap1+1)+log(Akrdob+1)
+log(Azal1+1)+korijen(Pnmpdg)+korijen(Prearnta)+log(GWTA+1) ,index=c("ID_SUBJEKTA", "GODINA"),
 data=panel, na.action=na.omit, model="random")
summary(random.modelnovi)

#model gdje sam izbacio varijable za koje ne postoji znacajna veza
random.modelnovi2 <- plm(log(prodaja+1)~log(pPrihZapos+1)+NematImovina+log(Zkz+1)+log(blta+1)+korijen(Zkvf)+
log(Lubrl+1)+log(Aukupni+1)+log(Anap1+1)+log(Akrdob+1)+log(Azal1+1)+korijen(Pnmpdg)+korijen(Prearnta)+log(GWTA+1) ,
index=c("ID_SUBJEKTA", "GODINA"), data=panel, na.action=na.omit, model="random")
summary(random.modelnovi2)

#FIXED

#fixed effects model
within.model <- plm(log(prodaja+1)~log(pPrihZapos+1)+NematImovina+log(kImovkObv+1)+log(Zkz+1)+korijen(Zdk)+
log(blta+1)+korijen(Zkvf)+korijen(ROSdg)+korijen(Lclnw)+log(Lubrl+1)+log(Aukupni+1)+log(Akrat+1)+
log(Anap1+1)+log(Akrdob+1)+log(Azal1+1)+korijen(Pnmpdg)+korijen(Prearnta)+log(Ctrenl+1)+log(GWTA+1),
index=c("ID_SUBJEKTA", "GODINA"), data=panel, na.action=na.omit, model="within")
summary(within.model)

#model kod kojeg su izbacene:kimovkobv, rosdg, akrat, ctrenl, zkvf
within.modelnovi <- plm(log(prodaja+1)~log(pPrihZapos+1)+NematImovina+log(Zkz+1)+korijen(Zdk)+log(blta+1)+
korijen(Lclnw)+log(Lubrl+1)+log(Aukupni+1)+log(Anap1+1)+log(Akrdob+1)+log(Azal1+1)+korijen(Pnmpdg)+korijen(Prearnta)+
log(GWTA+1)+factor(GODINA) ,index=c("ID_SUBJEKTA", "GODINA"), data=panel, na.action=na.omit, model="within")
summary(within.modelnovi)


#model gdje sam izbacio varijable za koje ne postoji znacajna veza
within.modelnovi2 <- plm(log(prodaja+1)~log(pPrihZapos+1)+NematImovina+log(Zkz+1)+log(blta+1)+log(Lubrl+1)+
log(Aukupni+1)+log(Anap1+1)+log(Akrdob+1)+log(Azal1+1)+korijen(Pnmpdg)+korijen(Prearnta)+log(GWTA+1) ,
index=c("ID_SUBJEKTA", "GODINA"), data=panel, na.action=na.omit, model="within")
summary(within.modelnovi2)

#test koji nam govori koji je model bolji za koristenje
phtest(random.model,within.model)

phtest(random.modelnovi, within.modelnovi)

phtest(random.modelnovi2, within.modelnovi2)



#fixed effects model sa ukljucenim faktorom vremena na kraju, da vidimo imamo li ukljucene vremenske efekte


within.model.t <- plm(log(prodaja+1)~log(pPrihZapos+1)+NematImovina+log(kImovkObv+1)+log(Zkz+1)+korijen(Zdk)+
log(blta+1)+korijen(Zkvf)+korijen(ROSdg)+korijen(Lclnw)+log(Lubrl+1)+log(Aukupni+1)+log(Akrat+1)+log(Anap1+1)+
log(Akrdob+1)+log(Azal1+1)+korijen(Pnmpdg)+korijen(Prearnta)+log(Ctrenl+1)+log(GWTA+1)+factor(GODINA),
index=c("ID_SUBJEKTA", "GODINA"), data=panel, na.action=na.omit, model="within")
summary(within.model.t)
```

```
within.modelnovi.t <- plm(log(prodaja+1)~log(pPrihZapos+1)+NematImovina+log(Zkz+1)+korijen(Zdk)+log(blta+1)+
korijen(Lclnw)+log(Lubrl+1)+log(Aukupni+1)+log(Anap1+1)+log(Akrdob+1)+log(Azal1+1)+korijen(Pnmpdg)+
korijen(Prearnta)+log(GWTA+1)+factor(GODINA) ,index=c("ID_SUBJEKTA", "GODINA"), data=panel, na.action=na.omit, model="w
summary(within.modelnovi.t)


within.modelnovi2.t <- plm(log(prodaja+1)~log(pPrihZapos+1)+NematImovina+log(Zkz+1)+log(blta+1)+log(Lubrl+1)+
log(Aukupni+1)+log(Anap1+1)+log(Akrdob+1)+log(Azal1+1)+korijen(Pnmpdg)+korijen(Prearnta)+log(GWTA+1)+factor(GODINA),
index=c("ID_SUBJEKTA", "GODINA"), data=panel, na.action=na.omit, model="within")
summary(within.modelnovi2.t)


#Test koji nam govori imamo li vremenske efekte ukljucene
pFtest(within.model.t , within.model)

pFtest(within.modelnovi.t, within.modelnovi)

pFtest(within.modelnovi2.t, within.modelnovi2)

#lagrange multiplikacijski test

plmtest(within.modelnovi, effect="time", type="bp")
plmtest(within.modelnovi2, effect="time", type="bp")

#cross-sectional ovisnost dva testa

pcdtest(within.model, test=c("cd"))
pcdtest(within.model, test=c("lm"))


pcdtest(within.modelnovi, test=c("cd"))
pcdtest(within.modelnovi, test=c("lm"))


pcdtest(within.modelnovi2, test=c("cd"))
pcdtest(within.modelnovi2, test=c("lm"))



#test za serijalnu korelaciju

pbgtest(within.model)

pbgtest(within.modelnovi)

pbgtest(within.modelnovi2)



#Vrijeme kao predikcija

# fixed sa godinama kao prediktorima
within.modeltime <- plm(log(prodaja+1)~log(pPrihZapos+1)+NematImovina+log(Zkz+1)+korijen(Zdk)+log(blta+1)+
korijen(Lclnw)+log(Lubrl+1)+log(Aukupni+1)+log(Anap1+1)+log(Akrdob+1)+log(Azal1+1)+korijen(Pnmpdg)+korijen(Prearnta)+
log(GWTA+1) + factor(GODINA) ,index=c("ID_SUBJEKTA", "GODINA"), data=panel, na.action=na.omit, model="within")
summary(within.modeltime)


within.modelvrijeme2 <- plm(log(prodaja+1)~log(pPrihZapos+1)+log(Zkz+1)+korijen(Zdk)+log(blta+1)+korijen(Lclnw)+
log(Lubrl+1)+log(Aukupni+1)+log(Anap1+1)+log(Akrdob+1)+log(Azal1+1)+korijen(Pnmpdg)+korijen(Prearnta)+log(GWTA+1) ,
index=c("ID_SUBJEKTA", "GODINA"), data=panel, na.action=na.omit, model="within", effect="time")
summary(within.modelvrijeme2)

within.modeltime2 <- plm(log(prodaja+1)~log(pPrihZapos+1)+NematImovina+log(Zkz+1)+korijen(Zdk)+log(blta+1)+
korijen(Lclnw)+log(Lubrl+1)+log(Aukupni+1)+log(Anap1+1)+log(Akrdob+1)+log(Azal1+1)+korijen(Pnmpdg)+korijen(Prearnta)+
log(GWTA+1) + factor(GODINA) +log(Anap1+1)*factor(GODINA) ,index=c("ID_SUBJEKTA", "GODINA"), data=panel,
 na.action=na.omit, model="within")
summary(within.modeltime2)

#cross-sectional
pcdtest(within.modeltime, test=c("cd"))
pcdtest(within.modeltime, test=c("lm"))


pcdtest(within.modelvrijeme2, test=c("cd"))
pcdtest(within.modelvrijeme2, test=c("lm"))

#serijalna korelacija
pbgtest(within.modeltime)

pbgtest(within.modelvrijeme2)


#heteroskedasticnost

bptest(log(prodaja+1)~log(pPrihZapos+1)+NematImovina+log(Zkz+1)+korijen(Zdk)+log(blta+1)+korijen(Lclnw)+log(Lubrl+1)+
log(Aukupni+1)+log(Anap1+1)+log(Akrdob+1)+log(Azal1+1)+korijen(Pnmpdg)+korijen(Prearnta)+log(GWTA+1)+factor(GODINA) ,
 data=panel, studentize=F)

coeftest(within.modelvrijeme2, vcovHC(within.modelvrijeme2, method = "arellano"))

t(sapply(c("HC0", "HC1", "HC2", "HC3", "HC4"), function(x) sqrt(diag(vcovHC(within.modelvrijeme, type = x)))))

plotmeans(log(prodaja+1)~ID_SUBJEKTA, data=panel)
```