Sveučilište J.J. Strossmayera u Osijeku
Fakultet primijenjene matematike i informatike
Sveučilišni diplomski studij matematike
Financijska matematika i statistika

Nikolina Lozina

Primjena jackknife metode na linearni regresijski model za procjenu rasta poduzeća

Diplomski rad

Osijek, 2023.

Sveučilište J.J. Strossmayera u Osijeku
Fakultet primijenjene matematike i informatike
Sveučilišni diplomski studij matematike
Financijska matematika i statistika

Nikolina Lozina

Primjena jackknife metode na linearni regresijski model za procjenu rasta poduzeća

Diplomski rad

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

Osijek, 2023.

Sadržaj

1 Uvod

Osnovna ideja metoda ponovljenog uzorkovanja je generirati uzorke iz postojećeg uzorka i na temelju generiranih uzoraka procijeniti određene parametre. Ova metoda je prihvaćena kao alternativa asimptotskim metodama i široko je primjenjiva, a često i preciznija od egzaktnih ili asimptotskih metoda. Jedan od nedostataka metode je taj što zahtijeva veću računalnu snagu, što za posljedicu ima da većina znanstvenika koristi asimptotske metode za potrebne izračune, a metode ponovljenog uzorkovanja koriste za konačno izvještavanje.

Tri su vrste metoda: jackknife, bootstrap i poduzorkovanje. *Jackknife metoda* temelji se na distibuciji dobivenoj iz n "izbaci-jedan" uzoraka i najčešće se koristi za procjenu varijance. Iz uzorka veličine n definiraju se novi uzorci na način da se za svaki $i = 1, \ldots, n$ iz početnog uzorka izbaci i-ti podatak. Na temelju svakog od n novih uzoraka veličine $n - 1$ računaju se određene statistike. *Bootstrap metoda* koristi uzorkovanje s vraćanjem, tj. iz empirijske distribucije podataka generiraju se novi uzorci koji su jednake veličine kao početni uzorak i sadrže samo vrijednosti iz početnog uzorka. Na temelju tako generiranih uzoraka izračunavaju se određene statističke veličine, a najčešće se koristi za procjenu varijance, konstrukciju pouzdanih intervala i testiranje hipoteza. *Poduzorkovanje* je metoda koja za procjenu koristi uzorkovanje bez vraćanja, tj. iz empirijske distribucije podataka generira nove uzorke koji su manji od početnog uzorka i zatim se izračunavaju određene statističke veličine. Može se koristiti u iste svrhe kao i bootstrap metoda.

U ovom radu predstaviti ćemo jackknife metodu od koje je sve počelo i primijeniti je na linearni regresijski model za procjenu rasta poduzeća. U 2. poglavlju kratko ćemo predstaviti metodu linearne regresije i navesti osnovne pretpostavke linearne regresije, zatim ćemo u 3. poglavlju definirati jackknife metodu i navesti njenu primjenu. Na kraju, u empirijskom dijelu rada izgraditi ćemo linearni regresijski model za procjenu rasta malih i srednjih poduzeća u Hrvatskoj na koji ćemo primjeniti jackknife metodu za izračun standardne greške procjenitelja i pouzdanih intervala i usporediti dobivene rezultate.

2 Linearna regresija

S obzirom da ćemo u empirijskom dijelu rada izgraditi linearni regresijski model za procjenu rasta malih i srednjih poduzeća u Hrvatskoj u ovom poglavlju kratko ćemo predstaviti osnovne pojmove i pretpostavke linearne regresije. U nastavku rada nećemo razlikovati u oznakama slučajne varijable i podatke koji su njihova realizacija, već ćemo smatrati da se to iz konteksta podrazumijeva.

Linearna regresija je metoda statističke analize kojom se ispituje veza između varijabli, tj. jedna varijabla opisuje se pomoću jedne ili više drugih varijabli.
Varijabla čije se promjene objašnjavaju pomoću drugih varijabli naziva se *ovisna varijabla* (regresand ili output), a varijable kojima se objašnjavaju promjene ovisne varijable zovu se *neovisne varijable* (regresori ili inputi). Regresijski modeli mogu biti linearni ili nelinearni u parametrima, te linearni ili nelinearni u varijablama. Pod pojmom linearni regresijski model podrazumijeva se model linearan u parametrima.

Neka je y ovisna varijabla koja je linearno povezana s k neovisnih varijabli $x_1, x_2, \dots, x_k$. Model višestruke linearne regresije pretpostavlja da vrijedi:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon, \tag{1}$$

gdje su $\beta_0, \beta_1, \dots, \beta_k$ koeficijenti modela, a ϵ greška modela.

Pretpostavimo da imamo n promatranja $\{(y_i, \boldsymbol{x}_i) : i = 1, \dots, n\}, \boldsymbol{x}_i = (x_{i1}, \dots, x_{ik})$ koji su nezavisne realizacije slučajnog vektora $(y, \boldsymbol{x}), \boldsymbol{x} = (x_1, \dots, x_k)$. Tada se model višestruke linearne regresije (1) može u matričnoj notaciji zapisati kao

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{2}$$

pri čemu su

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1^T \\ \boldsymbol{x}_2^T \\ \vdots \\ \boldsymbol{x}_n^T \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

U izrazu (2) $\boldsymbol{y}$ predstavlja n-dimenzionalni vektor stupac opaženih vrijednosti ovisne varijable y, $\boldsymbol{X}$ predstavlja $(n \times (k+1))$ matricu čiji prvi stupac sadrži jedinice, a ostali opažene vrijednosti neovisnih varijabli $x_1, x_2, \dots, x_k$. $\boldsymbol{\beta}$ je $(k+1)$-dimenzionalni vektor stupac nepoznatih parametara $\beta_i, i = 0, 1, \dots, k$, a $\boldsymbol{\epsilon}$ predstavlja n-dimenzionalan slučajan vektor grešaka modela.
U takvom slučaju imamo n jednadžbi oblika kao u (1) i možemo ih zapisati kao:

$$y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \epsilon_i, i = 1, \dots, n.$$

Ključna pretpostavka višestruke linearne regresije je da vrijedi

$$\mathbb{E}[y|\boldsymbol{x} = \boldsymbol{x}_i] = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}.$$

Model jednostavne linearne regresije je model u kojem je $k = 1$.

Jedan od zadataka linearne regresije je procijeniti nepoznate parametre $\beta_0, \ldots, \beta_k$. Za procjenu parametara najčešće se koristi metoda najmanjih kvadrata (engl. Ordinary Least Square, OLS). Metoda najmanjih kvadrata podrazumijeva minimiziranje srednje kvadratne greške $SSE(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$ po $\boldsymbol{\beta}$, odakle proizlazi procjenitelj za $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} \tag{3}$$

Pretpostavka je da je matrica $\boldsymbol{X}$ punog ranga, pa je matrica $\boldsymbol{X}^T\boldsymbol{X}$ invertibilna.

O kvaliteti modela zaključivati ćemo na temelju analize grešaka modela. S obzirom da su greške $\epsilon_i, i = 1, \ldots, n$ nemjerljive veličine potrebno ih je procijeniti. Procjenom koeficijenata modela dobit ćemo teorijske vrijednosti, na temelju kojih onda dobivamo jednog od procjenitelja grešaka modela, reziduale. Reziduali predstavljaju razliku između stvarnih vrijednosti i procijenjenih vrijednosti ovisne varijable, tj.

$$\hat{\epsilon}_i = y_i - \hat{y}_i = y_i - \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}, i = 1, \ldots, n.$$

Za procjenu se može koristiti i greška predikcije. Osnovna ideja greške predikcije je da u procjeni $\boldsymbol{\beta}$ ne koristi i-ti podatak, tj. $(y_i, \boldsymbol{x}_i)$. Tada imamo "izbaci-jedan" procjenitelja

$$\hat{\boldsymbol{\beta}}_{(-i)} = (\boldsymbol{x}_{(-i)}^T\boldsymbol{x}_{(-i)})^{-1}\boldsymbol{x}_{(-i)}^T\boldsymbol{y}_{(-i)}, i = 1, \ldots, n$$

i pri tome je predikcija na temelju $\boldsymbol{x}_i$ jednaka $\tilde{y}_i = \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}_{(-i)}$. Greška predikcije je razlika između stvarne vrijednosti i prediktirane vrijednosti varijable y, tj.

$$\tilde{\epsilon}_i = y_i - \tilde{y}_i = y_i - \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}_{(-i)}, i = 1, \ldots, n.$$

U analizi modela linearne regresije pretpostavljamo da slučajna varijabla greške ϵ_i, za svaki $i = 1, \ldots, n$ ispunjava sljedeće pretpostavke:

1) $\mathbb{E}[\epsilon_i] = 0$
2) $Cov(\epsilon_i, \epsilon_j) = \mathbb{E}[\epsilon_i\epsilon_j] = 0, i \neq j$
3) $Var[\epsilon_i] = \sigma^2, i = 1, \ldots, n$
4) $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$

Pretpostavka 1) posljedica je ključne pretpostavke linarne regresije i primjene teorema o dvostrukom očekivanju, tj. $\mathbb{E}[\epsilon_i] = \mathbb{E}[\mathbb{E}[\epsilon_i|\boldsymbol{x}]] = \mathbb{E}[\mathbb{E}[\epsilon_i|\boldsymbol{x}_i]] = 0$, a 2) je posljedica pretpostavke o nezavisnim mjerenjima i ključne pretpostavke linearne regresije. Pretpostavka 3) naziva se homoskedastičnost modela i specijalan je slučaj heteroskedastičnog modela koji pretpostavlja $Var[\epsilon_i] = \sigma_i^2, i = 1, \ldots, n$. U nastavku poglavlja pretpostavlja se heteroskedastičnost modela, te će se navesti rezultati za specijalan slučaj. Pretpostavka 4) koja govori o normalnosti grešaka modela može se ispustiti kada imamo velik broj podataka, jer se metodologija zaključivanja neće bitno narušiti.

Kako bi mogli izračunati varijancu procjenitelja $\hat{\boldsymbol{\beta}}$ u nastavku ćemo kratko prikazati načine izračuna matrice kovarijanci procjenitelja $\hat{\boldsymbol{\beta}}$,

$$\begin{aligned}
\boldsymbol{V}_{\hat{\boldsymbol{\beta}}} &= E[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T] \\
&= E[((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\epsilon})((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\epsilon})^T] \\
&= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T)\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1} \\
&= (\boldsymbol{X}^T\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{D}\boldsymbol{X})(\boldsymbol{X}^T\boldsymbol{X})^{-1},
\end{aligned}$$

gdje je $\boldsymbol{D} = diag[\sigma_1^2, \dots, \sigma_n^2]$. Dijagonalna matrica $\boldsymbol{D}$ je nemjerljiva veličina jer su greške modela $\epsilon_i, i = 1, \dots, n$ nemjerljive veličine pa je potrebno koristiti procjenitelje.

Za $\boldsymbol{V}_{\hat{\boldsymbol{\beta}}}$ u upotrebi su sljedeći procjenitelji, a razlikuju se u načinu procjene matrice $\boldsymbol{D}$:

$$1)\ \hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}}^{HCO} = \big(\boldsymbol{X}^T\boldsymbol{X}\big)^{-1}\big(\sum_{i=1}^{n}\boldsymbol{x}_i^T\boldsymbol{x}_i\hat{\epsilon}_i\big)\big(\boldsymbol{X}^T\boldsymbol{X}\big)^{-1}$$

$$2)\ \hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}}^{HC1} = \big(\frac{n}{n-k}\big)\big(\boldsymbol{X}^T\boldsymbol{X}\big)^{-1}\big(\sum_{i=1}^{n}\boldsymbol{x}_i^T\boldsymbol{x}_i\hat{\epsilon}_i\big)\big(\boldsymbol{X}^T\boldsymbol{X}\big)^{-1}$$

$$3)\ \hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}}^{HC2} = \big(\boldsymbol{X}^T\boldsymbol{X}\big)^{-1}\big(\sum_{i=1}^{n}\boldsymbol{x}_i^T\boldsymbol{x}_i\bar{\epsilon}_i\big)\big(\boldsymbol{X}^T\boldsymbol{X}\big)^{-1}, \bar{\epsilon}_i^2 = \frac{1}{1-h_{ii}}\hat{\epsilon}_i^2$$

$$4)\ \hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}}^{HC3} = \big(\boldsymbol{X}^T\boldsymbol{X}\big)^{-1}\big(\sum_{i=1}^{n}\boldsymbol{x}_i^T\boldsymbol{x}_i\tilde{\epsilon}_i\big)\big(\boldsymbol{X}^T\boldsymbol{X}\big)^{-1}, \tilde{\epsilon}_i^2 = \frac{1}{(1-h_{ii})^2}\hat{\epsilon}_i^2$$

pri čemu su h_{ii}[1] dijagonalni elementi hat matrice $\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$ za koje vrijedi $0 \leq h_{ii} \leq 1, i = 1, \dots, n$.

Kako je $\frac{1}{(1-h_{ii})^2} > \frac{1}{1-h_{ii}} > 1$, za procjenitelje matrice kovarijanci vrijedi:

$$\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}}^{HCO} < \hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}}^{HC2} < \hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}}^{HC3}.$$

Nejednakost matrica $\boldsymbol{A} < \boldsymbol{B}$ označava da je matrica $\boldsymbol{B} - \boldsymbol{A}$ pozitivno definitna, odnosno da su joj sve svojstvene vrijednosti veće od nule.
Kada linearni regresijski model zadovoljava svojstvo homoskedastičnosti koristi se nepristrani procjenitelj za varijancu greške modela

$$s^2 = \frac{1}{n-k}\sum_{i=1}^{n}\hat{\epsilon}_i^2,$$

pa je u tom slučaju procjenitelj za $\boldsymbol{V}_{\hat{\boldsymbol{\beta}}}$ jednak:

$$\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}s^2.$$

Nakon određivanja procjenitelja i modela linearne regresije potrebno je provjeriti pretpostavke modela kako bi se utvrdila njegova ispravnost prije donošenja konačnih zaključaka. Detaljnije o linearnoj regresiji može se pogledati u [4].

[1] Bruce E. Hansen, Econometrics, str. 80

3 Jackknife metoda

Jackknife metoda jedna je od metoda ponovljenog uzorkovanja čija je osnovna ideja generirati uzorak iz već postojećeg i na temelju generiranih uzoraka procijeniti određene statističke veličine. Povijesno je sve počelo 1949. kada je Maurice H. Quenouille predstavio jackknife procjenitelja pristranosti. Metodu je kasnije proširio John W. Tukey (1958.) na jackknife procjenitelja varijance. Bradley Efron (1979.) je prvi predložio korištenje jackknife metode za procjenu distribucije uzorka procjenitelja parametara u linearnom regresijskom modelu, a koju su kasnije razvili David A. Freedman (1981.) i Chien-Fu J. Wu (1986.).

Jackknife metoda koristi se za procjenu standardnih grešaka, pristranosti, intervala pouzdanosti i grešaka predikcije, a generirana je uzastopnim brisanjem jednog opažanja iz početnog uzorka stvarajući tako n zasebnih uzoraka veličine $n-1$.

Neka je $\hat{\boldsymbol{\theta}}$ procjenitelj za vektor parametara $\boldsymbol{\theta}$ koji je funkcija slučajnog uzorka veličine n. Neka je $\boldsymbol{V}_{\hat{\boldsymbol{\theta}}} = Var(\hat{\boldsymbol{\theta}})$ matrica kovarijanci od $\hat{\boldsymbol{\theta}}$. "Izbaci-jedan" procjenitelj $\hat{\boldsymbol{\theta}}_{(-i)}$ izračunava se koristeći formulu za $\hat{\boldsymbol{\theta}}$ tako da se izbriše i-to opažanje. John W. Tukey je definirao jackknife procjenitelja za $\boldsymbol{V}_{\hat{\boldsymbol{\theta}}}$ kao skaliranu varijancu uzorka svih "izbaci-jedan" procjenitelja:

$$\hat{\boldsymbol{V}}^{jack}_{\hat{\boldsymbol{\theta}}} = \frac{n-1}{n} \sum_{i=1}^{n} (\hat{\boldsymbol{\theta}}_{(-i)} - \bar{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_{(-i)} - \bar{\boldsymbol{\theta}})^T, \tag{4}$$

gdje je $\bar{\boldsymbol{\theta}}$ prosjek "izbaci-jedan" procjenitelja, tj.

$$\bar{\boldsymbol{\theta}} = \frac{1}{n} \sum_{i=1}^{n} \hat{\boldsymbol{\theta}}_{(-i)}.$$

Za procjenitelja $\hat{\boldsymbol{\theta}}$ jackknife standardna greška je kvadratni korijen elemenata dijagonale matrice kovarijanci $\hat{\boldsymbol{V}}^{jack}_{\hat{\boldsymbol{\theta}}}$, tj. standardna greška pojedinog procjenitelja $\hat{\theta}_j$ definirana je formulom:

$$s^{jack}_{\hat{\theta}_j} = \sqrt{[\hat{\boldsymbol{V}}^{jack}_{\hat{\boldsymbol{\theta}}}]_{jj}}$$

Zgodna značajka jackknife procjenitelja $\hat{\boldsymbol{V}}^{jack}_{\hat{\theta}}$ je ta da je formula (4) prilično generalna i ne zahtijeva analitičke izračune iz distribucije od $\hat{\boldsymbol{\theta}}$. Nedostatak jackknife metode je što zahtjev od n zasebnih procjenitelja u nekim slučajevima može biti računalno zahtjevan.

U većini slučajeva $\hat{\boldsymbol{V}}^{jack}_{\hat{\boldsymbol{\theta}}}$ će biti sličan robusnoj asimptotskoj varijanci procjenitelja. Najznačajnija prednost jackknife procjenitelja je ta da se može koristiti kada eksplicitna asimptotska formula za varijancu nije poznata i što se može koristiti kao procjena realnosti asimptotske formule. Formula (4) nije odmah intuitivna, stoga ju je potrebno detaljnije objasniti.

U tu svrhu počinjemo s ispitivanjem uzoračkog očekivanja $\bar{\boldsymbol{y}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{y}_i, \boldsymbol{y} \in \mathcal{R}^m$ kao procjenitelja za populacijsko očekivanje. "Izbaci-jedan" procjenitelj je

$$\bar{\boldsymbol{y}}_{(-i)} = \frac{1}{n-1} \sum_{j \neq i} \boldsymbol{y}_j = \frac{n}{n-1}\bar{\boldsymbol{y}} - \frac{1}{n-1}\boldsymbol{y}_i \tag{5}$$

Prosjek "izbaci-jedan" procjenitelja jednak je

$$\frac{1}{n} \sum_{i=1}^{n} \bar{\boldsymbol{y}}_{(-i)} = \frac{n}{n-1}\bar{\boldsymbol{y}} - \frac{1}{n-1}\bar{\boldsymbol{y}} = \bar{\boldsymbol{y}},$$

tj. odgovara uzoračkom očekivanju.
Razlika između "izbaci-jedan" uzoračkog očekivanja i njegovog prosjeka jednaka je

$$\bar{\boldsymbol{y}}_{(-i)} - \bar{\boldsymbol{y}} = \frac{n}{n-1}\bar{\boldsymbol{y}} - \frac{1}{n-1}\boldsymbol{y}_i - \bar{\boldsymbol{y}} = \frac{1}{n-1}(\bar{\boldsymbol{y}} - \boldsymbol{y}_i).$$

Jackknife procjenitelj varijance (4) je tada

$$\begin{aligned} \hat{\boldsymbol{V}}_{\bar{\boldsymbol{y}}}^{jack} &= \frac{n-1}{n}\sum_{i=1}^{n}\left(\frac{1}{n-1}\right)^2(\bar{\boldsymbol{y}} - \boldsymbol{y}_i)(\bar{\boldsymbol{y}} - \boldsymbol{y}_i)^T \\ &= \frac{1}{n}\left(\frac{1}{n-1}\right)\sum_{i=1}^{n}(\bar{\boldsymbol{y}} - \boldsymbol{y}_i)(\bar{\boldsymbol{y}} - \boldsymbol{y}_i)^T. \end{aligned} \tag{6}$$

Ovaj procjenitelj varijance jednak je uobičajenom procjenitelju za varijancu od $\bar{\boldsymbol{y}}$. John W. Tukey je predložio skaliranje s $(n-1)/n$ u (4) tako da je $\hat{\boldsymbol{V}}_{\bar{\boldsymbol{y}}}^{jack}$ upravo jednaka tom procjenitelju. To pokazuje da je za uzoračko očekivanje jackknife procjenitelj varijance identičan uobičajenom procjenitelju varijance.

U nastavku slijedi analiza koeficijenata dobivenih metodom najmanjih kvadrata, a s ciljem izračuna jackknife procjenitelja matrice kovarijanci za $\hat{\boldsymbol{\beta}}$. U prethodnom poglavlju definiran je "izbaci-jedan" OLS procjenitelj:

$$\hat{\boldsymbol{\beta}}_{(-i)} = (\boldsymbol{x}_{(-i)}^T\boldsymbol{x}_{(-i)})^{-1}\boldsymbol{x}_{(-i)}^T\boldsymbol{y}_{(-i)}, i = 1, \ldots, n$$

i tako definiran zahtijeva računanje n zasebnih regresija što može biti dosta zahtjevno. Sljedeći teorem daje formulu za jednostavniji izračun "izbaci-jedan" procjenitelja, pa je time i primjena Jackknife metode kod linearne regresije bitno pojednostavljena.

Teorem 3.1. *Za "izbaci-jedan" OLS procjenitelja vrijedi:*

$$\hat{\boldsymbol{\beta}}_{(-i)} = \hat{\boldsymbol{\beta}} - (1 - h_{ii})^{-1}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i\hat{\epsilon}_i,$$

gdje je $h_{ii} = \boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i$.

U dokazu teorema koristi se Sherman-Morrison formula (vidi [9]) koja daje formulu za izračun inverza sume nesingularne matrice $\boldsymbol{A} \in R^{nxn}$ i produkta vektora $\boldsymbol{u}, \boldsymbol{v} \in R^n$:

$$(\boldsymbol{A} - \boldsymbol{u}\boldsymbol{v}^T)^{-1} = \boldsymbol{A}^{-1} + \frac{\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^T\boldsymbol{A}^{-1}}{1 - \boldsymbol{v}^T\boldsymbol{A}^{-1}\boldsymbol{u}}.$$

Dokaz.

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{(-i)} &= (\boldsymbol{x}_{(-i)}^T \boldsymbol{x}_{(-i)})^{-1} \boldsymbol{x}_{(-i)}^T \boldsymbol{y}_{(-i)} \\
&= (\boldsymbol{X}^T\boldsymbol{X} - \boldsymbol{x}_i\boldsymbol{x}_i^T)^{-1}(\boldsymbol{X}^T\boldsymbol{y} - \boldsymbol{x}_i y_i) \\
&= \left[(\boldsymbol{X}^T\boldsymbol{X})^{-1} + \frac{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}}{1 - \boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i}\right](\boldsymbol{X}^T\boldsymbol{y} - \boldsymbol{x}_i y_i) \\
&= \left[(\boldsymbol{X}^T\boldsymbol{X})^{-1} + (1-h_{ii})^{-1}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\right](\boldsymbol{X}^T\boldsymbol{y} - \boldsymbol{x}_i y_i) \\
&= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} + (1-h_{ii})^{-1}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} - \\
&- (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i y_i - (1-h_{ii})^{-1}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i\boldsymbol{y}_i \\
&= \hat{\boldsymbol{\beta}} - (1-h_{ii})^{-1}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i\left[-\boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} + (1-hii)y_i + h_{ii}y_i\right] \\
&= \hat{\boldsymbol{\beta}} - (1-h_{ii})^{-1}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i\big(y_i - \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}\big) \\
&= \hat{\boldsymbol{\beta}} - (1-h_{ii})^{-1}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i\hat{e}_i
\end{aligned}$$

□

Prosjek svih "izbaci-jedan" procjenitelja je tada

$$\begin{aligned}
\bar{\boldsymbol{\beta}} = \frac{1}{n}\sum_{i=1}^{n}\hat{\boldsymbol{\beta}}_{(-i)} &= \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\boldsymbol{\beta}} - (1-h_{ii})^{-1}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i\hat{e}_i\right) \\
&= \hat{\boldsymbol{\beta}} - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\frac{1}{n}\sum_{i=1}^{n}(1-h_{ii})^{-1}\boldsymbol{x}_i\hat{e}_i \\
&= \hat{\boldsymbol{\beta}} - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\tilde{e}_i \\
&= \hat{\boldsymbol{\beta}} - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\tilde{\boldsymbol{\mu}}
\end{aligned}$$

gdje je $\tilde{\boldsymbol{\mu}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\tilde{e}_i$. Prema tome je

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{(-i)} - \bar{\boldsymbol{\beta}} &= \hat{\boldsymbol{\beta}} - (1-h_{ii})^{-1}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i\hat{e}_i - \hat{\boldsymbol{\beta}} + (\boldsymbol{X}^T\boldsymbol{X})^{-1}\tilde{\boldsymbol{\mu}} \\
&= -(\boldsymbol{X}^T\boldsymbol{X})^{-1}\left((1-h_{ii})^{-1}\boldsymbol{x}_i\hat{e}_i - \tilde{\boldsymbol{\mu}}\right) \\
&= -(\boldsymbol{X}^T\boldsymbol{X})^{-1}(\boldsymbol{x}_i\tilde{e}_i - \tilde{\boldsymbol{\mu}}).
\end{aligned}$$

Jackknife procjenitelj varijance za $\hat{\boldsymbol{\beta}}$ je

$$\begin{aligned}
\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}}^{jack} &= \frac{n-1}{n}\sum_{i=1}^{n}\left(\hat{\boldsymbol{\beta}}_{(-i)} - \bar{\boldsymbol{\beta}}\right)\left(\hat{\boldsymbol{\beta}}_{(-i)} - \bar{\boldsymbol{\beta}}\right)^T \\
&= \frac{n-1}{n}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\left(\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^T\tilde{e}_i^2 - n\tilde{\boldsymbol{\mu}}\tilde{\boldsymbol{\mu}}^T\right)(\boldsymbol{X}^T\boldsymbol{X})^{-1} \qquad (7) \\
&= \frac{n-1}{n}\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}}^{HC3} - (n-1)(\boldsymbol{X}^T\boldsymbol{X})^{-1}\tilde{\boldsymbol{\mu}}\tilde{\boldsymbol{\mu}}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1},
\end{aligned}$$

gdje je $\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}}^{HC3}$ HC3 procjenitelj matrice kovarijanci temeljen na greškama predikcije. Drugi izraz u (7) je prilično mali dok je $\tilde{\boldsymbol{\mu}}$ tipično male magnitude, pa je prema tome $\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}}^{jack} \simeq \hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}}^{HC3}$.

Ova analiza pokazuje da je za regresijske koeficijente jackknife procjenitelj varijance sličan asimptotskom procjenitelju varijance.

U većini istraživanja područje interesa u regresijskom modelu je transformacija od $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)$ i u tom slučaju parametar od interesa $\boldsymbol{\theta}$ može se definirati kao funkcija vektora koeficijenata, tj. $\boldsymbol{\theta} = r(\boldsymbol{\beta})$, za funkciju $r : \mathcal{R}^k \to \mathcal{R}^q$. Procjenitelj od $\boldsymbol{\theta}$ je tada $\hat{\boldsymbol{\theta}} = r(\hat{\boldsymbol{\beta}})$. Ako je r neprekidna funkcija onda je $\hat{\boldsymbol{\theta}}$ konzistentan procjenitelj za $\boldsymbol{\theta}$[2]. Štoviše, pod pretpostavkama:
1) vektori $(y_i, x_i), i = 1, \ldots, n$ su nezavisni i jednako distribuirani
2) $E[y^4] < \infty$
3) $E||\boldsymbol{x}_i^4|| < \infty$
4) $E[\boldsymbol{X}\boldsymbol{X}^T]$ je pozitivno definitna
5) r je neprekidno diferencijabilna u pravoj vrijednosti $\boldsymbol{\beta}$ i $\boldsymbol{R} = \left[\frac{\partial}{\partial \boldsymbol{\beta}} r(\boldsymbol{\beta})\right]^T$ je matrica punog ranga, tj. ima rang q,
vrijedi:

Teorem 3.2. *Za $n \to \infty$ vrijedi:*
$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \underset{d}{\to} \mathcal{N}(\boldsymbol{0}, \boldsymbol{V}_{\boldsymbol{\theta}})$,
gdje je $\boldsymbol{V}_{\boldsymbol{\theta}} = \boldsymbol{R}^T \boldsymbol{V}_{\boldsymbol{\beta}} \boldsymbol{R}$.

Dokaz teorema slijedi iz primjene delta metode[3].
Pod uvjetima ovog teorema, $\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\theta}}} = \hat{\boldsymbol{R}}^T \hat{\boldsymbol{V}}_{\boldsymbol{\beta}} \hat{\boldsymbol{R}}$ je konzistentan procjenitelj za $\boldsymbol{V}_{\boldsymbol{\theta}}$.

Definiranjem $r(\boldsymbol{\beta}) = \boldsymbol{x}^T \boldsymbol{\beta}$ procjenitelj od $\boldsymbol{\theta}$ je $\hat{\boldsymbol{\theta}} = \boldsymbol{x}^T \hat{\boldsymbol{\beta}}$ i onda je $\boldsymbol{R} = \boldsymbol{x}$. Standardna greška procjenitelja $\hat{\boldsymbol{\theta}}$ je jednaka $s(\hat{\boldsymbol{\theta}}) = \sqrt{\hat{\boldsymbol{R}}^T \hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}} \hat{\boldsymbol{R}}} = \sqrt{\boldsymbol{x}^T \hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}} \boldsymbol{x}}$.
Dakle, formula za 95% - interval pouzdanosti glasi $\left[\boldsymbol{x}^T \boldsymbol{\beta} \pm 1.96 \sqrt{\boldsymbol{x}^T \hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}} \boldsymbol{x}}\right]$

U nastavku slijedi definicija jackknife procjenitelja za funkciju $\hat{\boldsymbol{\theta}} = r(\hat{\boldsymbol{\beta}})$ procjenitelja metodom najmanjih kvadrata. "Izbaci-jedan" procjenitelj od $\boldsymbol{\theta}$ je

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_{(-i)} &= r(\hat{\boldsymbol{\beta}}_{(-i)}) \\
&= r(\hat{\boldsymbol{\beta}} - (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{x}_i \tilde{e}_i) \\
&\approx \hat{\boldsymbol{\theta}} - \hat{\boldsymbol{R}}^T (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{x}_i \tilde{e}_i.
\end{aligned}$$

Posljednja aproksimacija dobije se koristeći $r(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\theta}}$ i definiranjem $\hat{\boldsymbol{R}} = (\partial / \partial \boldsymbol{\beta}) r(\hat{\boldsymbol{\beta}})^T$.

Jackknife procjenitelj varijance za $\hat{\boldsymbol{\theta}}$ je tada jednak

$$\begin{aligned}
\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\theta}}}^{jack} &= \frac{n-1}{n} \sum_{i=1}^{n} (\hat{\boldsymbol{\theta}}_{(-i)} - \bar{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_{(-i)} - \bar{\boldsymbol{\theta}})^T \\
&\approx \frac{n-1}{n} \hat{\boldsymbol{R}}^T (\boldsymbol{X}^T \boldsymbol{X})^{-1} (\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^T \tilde{e}_i^2 - n \tilde{\boldsymbol{\mu}} \tilde{\boldsymbol{\mu}}^T) (\boldsymbol{X}^T \boldsymbol{X})^{-1} \hat{\boldsymbol{R}} \\
&= \hat{\boldsymbol{R}}^T \hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}}^{jack} \hat{\boldsymbol{R}} \\
&\approx \hat{\boldsymbol{R}}^T \hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}}^{HC3} \hat{\boldsymbol{R}}.
\end{aligned}$$

[2]Econometrics, B.Hansen, Teorem 6.6, str. 156
[3]Econometrics, B.Hansen, Teorem 6.8, str. 157

Zadnji redak u gornjem izrazu jednak je procjenitelju delta metodom za varijancu od $\hat{\boldsymbol{\theta}}$ konstruiranog s HC3 procjeniteljem varijance. Ovim je potvrđeno da jackknife daje procjenitelja koji je asimptotski sličan procjenitelju dobivenog korištenjem asimptotskih metoda, unatoč činjenici da se izračunava bez referenca na asimptotsku teoriju. Ovaj se argument direktno proširuje na bilo kojeg procjenitelja "glatke" funkcije.

4 Model za predikciju rasta malih i srednjih poduzeća u Hrvatskoj

Rastuća poduzeća imaju značajan utjecaj na nacionalna gospodarstva jer se stvaranjem i širenjem tvrtki generiraju nova zapošljavanja i mogućnosti, omogućujući ljudima prosperitetniji život. Prepoznajući važnost rasta tvrtki, ekonomisti, političari i razvojne agencije ulažu značajna sredstva za stvaranje i provedbu programa koji pomažu rastu poduzeća. Kako bi ti programi davali odgovarajuće rezultate nužno je razumjeti što ograničava ili poboljšava rast poduzeća.

Dosadašnja istraživanja predikcije rasta malih i srednjih poduzeća uključivala su razne faktore vezane uz osobine vlasnika poduzeća, poduzeće i okoline poduzeća, no još uvijek ne postoji opće prihvaćeni model odrednica rasta poduzeća. Jedan od razloga tome je činjenica da su istraživanja temeljena na podacima pojedine države i to najčešće dobro razvijenih država. Malo je istraživanja usredotočeno na čimbenike koji utječu na rast poduzeća u gospodarstvima u razvoju, a upravo u takvim gospodarstvima je ključan gospodarski rast kako bi se smanjile razlike s dobro razvijenim gospodarstvima i potaknula dobrobit njihovog stanovništva. U dosadašnjim istraživanjima pokazalo se da su najvažniji sudionici u razvoju segmenata rastućih poduzeća vlasnik i poduzeće, dok je država također važan sudionik kroz stvaranja poduzetničke okoline koja podržava rast poduzeća.

Rast poduzeća može se mjeriti kvalitativno u smislu kvalitete proizvoda ili usluga koje nude ili tržišne pozicije, te kvantitativno kroz generiranje fizičkog outputa ili ekspanzije volumena poslovanja. Najčešće kvantitativne mjere za definiranje rasta su rast prodaje, prihoda, zaposlenih i imovine (P. Davidsson, P. Steffens i J. Fitzsimmons, vidi [1]). Prilikom analize prethodnih istraživanja nužno je voditi računa o korištenoj definiciji rasta poduzeća.

4.1 Prethodna istraživanja

J. Wiklund, H. Potzelt i D. A. Shepherd (vidi [8]) proučavali su podatke 413 malih švedskih poduzeća. Pokazali su postojanje pozitivne veze između rasta poduzeća i stava prema rastu, poduzetničke orijentacije te dinamičke i heterogene okoline, dok je neprijateljska okolina u negativnoj vezi s rastom malih poduzeća.

F. Diaz Hermelo i R. Vassolo (vidi [2]) proučavali su podatke prikupljene od 34 poduzeća iz pokrajine Tucman, Argentina iz vremenskog razdoblja od 1994. do 1996. godine. Korištena definicija rasta poduzeća je stopa rasta prodaje u trogodišnjem periodu. Došli su do zaključka kako postoji pozitivna veza između rasta poduzeća i dostupnosti financijskih izvora, ulaganja u noviju tehnologiju i orijentacije na međunarodna tržišta.

G. Trovato i L. Becchetti (vidi [7]) proučavali su podatke prikupljene od oko 4.000 poduzeća u Italiji u periodu od 1989. do 1997. Analizom podataka utvrdili su postojanje pozitivne veze između rasta poduzeća i dostupnosti financijskih izvora, te ukupnog udjela vlasništva u poduzeću. Pokazali su i da mlađe poduzeće ima veću priliku za rastom, te da veću priliku za rastom imaju tvrtke kojima je ograničenost proračuna ublažena državnim subvencijama.

N. Šarlija i A. Bilandžić (vidi [6]) proučavale su podatke prikupljene od 265 malih i srednjih poduzeća u Hrvatskoj u periodu od 2010. do 2015. Definicija rasta poduzeća koju su koristile u istraživanju je godišnji rast 20% ili više u trogodišnjem razdoblju. Istraživanje je pokazalo da su dobre vještine umrežavanja i manje godina radnog iskustva poduzetnika najvažnije karakteristike poduzetnika. Mlađa poduzeća, novija tehnologija i spremnost na rast ključni su pokretači uspjeha, dok je odnos zadržane dobiti i ukupne imovine te prodaja po zaposlenom u negativnoj vezi s rastom poduzeća. Što se tiče okoline poduzeća, ključni pokazatelj je održavanje razine konkurentnosti, odnosno što je viša razina konkurentnosti, poduzeće ima veću priliku za rastom.

4.2 Podaci

Baza podataka korištena u ovom radu nastala je u okviru projekta *"Razvoj i testiranje modela za procjenu potencijala za rast malih i srednjih poduzeća u Hrvatskoj"*[4], Ekonomskog fakulteta u Osijeku. Podaci na temelju kojih je napravljeno istraživanje dijele se na primarne i sekundarne podatke.

Primarni podaci prikupljeni su pomoću ankete koja se sastoji od 3 dijela, a kojima su obuhvaćene karakteristike poduzetnika, poduzeća i poduzetničke okoline. Anketirano je 265 poduzetnika u 2015. godini.
Sekundarne podatke ustupila je financijska agencija FINA, a koji obuhvaćaju podatke iz financijskih izvještaja za 191 malih i srednjih poduzeća u razdoblju 2012. – 2015.
Reduciranjem baze podataka i isključivanjem poduzeća za koja su nedostajali neki od podataka, dobivena je baza podataka od 156 poduzeća.

S obzirom na dostupnost baze podataka u ovom radu za ovisnu varijablu odabrana je varijabla *omjer poslovnih prihoda*, a koja predstavlja odnos poslovnih prihoda iz 2015. i 2014. godine, tj.

$$Omjer\ poslovnih\ prihoda = \frac{poslovni\ prihodi\ 2015.}{poslovni\ prihodi\ 2014.}$$

Prilikom izrade modela za predikciju rasta poduzeća, u prvom koraku primjenom jednostavne linerane regresije analizirana je veza između ovisne varijable *omjer poslovnih prihoda* i pojedine neovisne varijable. Temeljem navedene analize, u početni model uključene su slijedeće 24 varijable koje su se pokazale značajnima:

∗ **Nematerijalna imovina** - udio nematerijalne imovine u ukupnoj imovini poduzeća.
Analizom je utvrđena pozitivna statistička veza, tj. s porastom udjela nematerijane imovine u ukupnoj imovini poduzeće ostvaruje veći rast poslovnih prihoda.

∗ **Godina rođenja** - godina rođenja poduzetnika.
Analizom je utvrđena pozitivna statistička veza, odnosno poduzeća čiji su poduzetnici mlađi ostvaruju veći rast poslovnih prihoda.

∗ **Prethodno iskustvo** - ima li poduzetnik iskustva u poduzetništvu iz vremena prije nego je počeo raditi za trenutno poduzeće? (0-ne, 1-da)
Pokazalo se da poduzeća čiji poduzetnici imaju prethodno iskustvo u poduzetništvu ostvaruju veći rast poslovnih prihoda od poduzeća čiji poduzetnici nemaju prethodno iskustvo.

[4]Projekt financiran od Hrvatske zaklade za znanost, voditeljica projekta: prof. dr. sc. Nataša Šarlija, 2014.-2018. godina

∗ **Broj propalih poduzeća** - broj propalih poduzeća koja je poduzetnik imao prije trenutnog poduzeća.
Analizom je utvrđena negativna statistička veza, tj. što je veći broj propalih poduzeća to je manji rast poslovnih prihoda.

∗ **Strateško razmišljanje** - ocjena poduzetnikove vještine strateškog razmišljanja (1 - najlošije, 5 - najbolje).
Analizom je utvrđena pozitivna statistička veza, tj. što je bolja ocjena poduzetnikove vještine strateškog razmišljanja poduzeće ostvaruje veći rast poslovnih prihoda.

∗ **Sposobnosti učenja** - ocjena poduzetnikove vještine sposobnosti učenja (1 - najlošije, 5 - najbolje).
Analizom je utvrđena pozitivna statistička veza, tj. što je bolja ocjena poduzetnikove vještine sposobnosti učenja poduzeće ostvaruje veći rast poslovnih prihoda.

∗ **Strateško planiranje** - postotak radnog vremena poduzetnika koji odlazi na strateško razmišljanje.
Analizom je utvrđena pozitivna statistička veza, tj. što je veći postotak radnog vremena koji odlazi na strateško razmišljanje poduzeće ostvaruje veći rast poslovnih prihoda.

∗ **Orijentacija rastu** - u viziji i misiji poduzeća je jasno izražena orijentacija prema rastu (0-ne, 1-da).
Analizom je utvrđeno da poduzeća u kojima je jasno izražena orijentacija rastu ostvaruju veći rast poslovnih prihoda od poduzeća u kojima nije jasno izražena orijentacija rastu.

∗ **Obuka** - broj sati obuke svih zaposlenika u posljednje dvije godine.
Analizom je utvrđena pozitivna statistička veza, tj. s porastom broja sati obuke svih zaposlenika raste i poslovni prihod poduzeća.

∗ **Starost tehnologije** - starost tehnologije koja se koristi pri razvoju proizvoda/usluga u poduzeću.
Analizom je utvrđena negativna statistička veza, odnosno što je starija tehnologija to poduzeće ostvaruje manji rast poslovnih prihoda.

∗ **Inozemni kupci** - postotak kupaca izvan Hrvatske.
Analizom je utvrđena pozitivna statistička veza, tj. što je veći postotak inozemnih kupaca poduzeće ostvaruje veći rast poslovnih prihoda.

∗ **Snaga za rastom** - u poduzeću su svim snagama orijentirani prema rastu poduzeća (1 - *uopće se ne slažem*, 5 - *u potpunosti se slažem*).
Analizom je utvrđena pozitivna statistička veza, tj. što je snažnija orijentacija rastu u poduzeću, veći je rast poslovnih prihoda.

∗ **Utjecaj na događaje** - poduzetnik često ima osjećaj kako ne može utjecati na stvari koje mu se događaju (1 - *uopće se ne slažem*, 5 - *u potpunosti se slažem*).
Analizom je utvrđena negativna statistička veza, tj. što češće poduzetnik ne može utjecati na događaje poduzeće ostvaruje manji rast poslovnih prihoda.

∗ **Odluke** - koliko često je zaposlenicima dopušteno donošenje odluka? (1 - *gotovo nikada*, 5 - *gotovo uvijek*)
Analizom je utvrđena pozitivna statistička veza, tj. što češće zaposlenici mogu donositi odluke poduzeće ostvaruje veći rast poslovnih prihoda.

∗ **Planiranje** - koliko često se planiraju sve poslovne aktivnosti u poduzeću? (1 - *gotovo nikada*, 5 - *gotovo uvijek*)
Analizom je utvrđena pozitivna statistička veza, tj. što se češće planiraju sve poslovne aktivnosti poduzeće ostvaruje veći rast poslovnih prihoda.

∗ **Ciljevi** - koliko često su u poduzeću orijentirani postizanju ciljeva koje su si postavili? (1 - *gotovo nikada*, 5 - *gotovo uvijek*)
Analizom je utvrđena pozitivna statistička veza, odnosno što je veća orijentacija postizanju ciljeva poduzeće ostvaruje veći rast poslovnih prihoda.

∗ **Konkurencija** - u kojoj mjeri je povećana konkurencija barijera rastu poduzeća? (1 - *uopće nije barijera*, 5 - *vrlo visoka barijera*)
Analizom je utvrđena negativna statistička veza, odnosno što je povećana konkurencija veća barijera rastu manji je rast poslovnih prihoda.

∗ **Zakon o radu** - u kojoj mjeri je zakon o radu barijera rastu poduzeća? (1 - *uopće nije barijera*, 5 - *vrlo visoka barijera*)
Analizom je utvrđena negativna statistička veza, osnosno što je zakon o radu veća barijera rastu manji je rast poslovnih prihoda.

∗ **Korupcija** - u kojoj mjeri je korupcija barijera rastu poduzeća? (1 - *uopće nije barijera*, 5 - *vrlo visoka barijera*)
Analizom je utvrđena negativna statistička veza, osnosno što je korupcija veća barijera rastu manji je rast poslovnih prihoda.

Kako je ranije navedeno, primarni podaci prikupljeni su na temelju anketiranja poduzetnika. Za dio pitanja u anketi korištena je Likertova skala $1-5$ koja odražava u kojoj mjeri se ispitanik slaže s pojedinom tvrdnjom. Pritom ocjena 1 najčešće označava tvrdnju *uopće se ne slažem, gotovo nikada i sl.*, a ocjena 5 označava tvrdnju *u potpunosti se slažem, gotovo uvijek i sl.* Temeljem toga definiraju se teorijski konstrukti[5] kao suma ili prosjek odgovora na skup tvrdnji koje se odnose na isti objekt. U ovom radu istraživani su i uključeni u modeliranje sljedeći teorijski konstrukti:

Postignuće je varijabla dobivena kao prosječna vrijednost odgovora na slijedeće izjave:
∗ *Čak i nakon što ostvarim uspjeh, nastojim biti još bolji.*
∗ *Rado se uspoređujem s drugima.*
∗ *Činim sve kako bi postigao cilj.*

Rizik je varijabla dobivena kao prosječna vrijednost odgovora na slijedeće izjave:
∗ *Volim kockanje.*
∗ *Poduzimam aktivnosti, čak i ako podrazumijevaju rizik.*

[5]Više o konstruktima u članku *Determinants and dimensions of firm growth*, https://ondernemerschap.panteia.nl/pdf-ez/200903.pdf

∗ *Spremam sam preuzeti rizik.*

Samoefikasnost je varijabla dobivena kao prosječna vrijednost odgovora na slijedeće izjave:

∗ *Mogu donijeti dobre strateške odluke.*
∗ *U raspravama dajem značajan doprinos.*
∗ *Otvoren sam za nove i netradicionalne ideje.*
∗ *Obično ja vodim implementaciju novih ideja, proizvoda i usluga.*
∗ *Postavljam pitanja kojih se nitko drugi nije sjetio.*
∗ *Postavljam si ciljeve i radim u skladu s tim ciljevima.*
∗ *Orijentiran sam ciljevima.*

Iskustvo je varijabla dobivena kao suma odgovora na slijedeća pitanja:

∗ *Koliko godina imate u industriji u kojoj posluje Vaše poduzeće?*
∗ *Koliko ukupno godina radnog iskustva imate?*
∗ *Koliko godina radite u ovom poduzeću?*
∗ *Koliko godina ste u poduzetništvu?*

Tržišna orijentacija je varijabla dobivena kao prosječna vrijednost odgovora na slijedeće izjave:

∗ *Zadovoljstvo kupaca mjerimo strukturno i periodički.*
∗ *Raspravljamo o tome što konkurencija radi.*
∗ *Interno dijelimo informacije o željama klijenata.*
∗ *Naše interne procedure i pravila su fokusirani na zadovoljavanje potreba na tržištu.*
∗ *Razmišljamo i radimo u skladu s potrebama kupaca koje će se tek pojaviti na tržištu.*
∗ *Fokusirani smo na privlačenje novih kupaca s novim potrebama.*

Poduzetnička orijentacija je varijabla dobivena kao prosječna vrijednost odgovora na slijedeće izjave:

∗ *Aktivno tražimo inovativne koncepte proizvoda/usluga i proizvodnih procesa.*
∗ *U odnosu na ostala poduzeća, mi preuzimamo veći rizik.*
∗ *Na aktivnosti konkurencije reagiramo snažno i agresivno.*

Spremnost je varijabla dobivena kao prosječna vrijednost odgovora na slijedeće izjave:

∗ *Spremni smo na snažan rast našeg poduzeća.*
∗ *Unutar našeg poduzeća, svi znaju da želimo snažno rasti.*
∗ *S trenutnom organizacijskom strukturom i resursima, možemo rasti stopom od 20%.*

Organizacijsko učenje je varijabla dobivena kao prosječna vrijednost odgovora na slijedeće izjave :

∗ *Unutar poduzeća svi se slažu s ciljevima poduzeća.*
∗ *Imamo snažan timski duh.*
∗ *Edukacija zaposlenika je investicija, a ne trošak.*
∗ *Smatramo da je učenje ključno ukoliko želimo poboljšati poslovanje.*
∗ *Odvajamo dovoljno vremena da naučimo iz vlastitih pogrešaka.*
∗ *Proučavamo uspješne i neuspješne poslovne aktivnosti i međusobno ih raspravljamo.*

Intezitet konkurencije je varijabla dobivena kao prosječna vrijednost odgovora na slijedeće izjave:

∗ *Naš tržišni udio je pod utjecajem jake konkurencije.*
∗ *Naše tržište je karakterizirano snažnom konkurencijom.*

Tržišna dinamika je varijabla dobivena kao prosječna vrijednost odgovora na slijedeće izjave:

∗ *Kupci konstantno traže nove proizvode/usluge.*
∗ *Proizvodi i usluge na našem tržištu brzo zastarjevaju.*

Potencijal je varijabla dobivena kao prosječna vrijednost odgovora na slijedeće izjave:

∗ *Na našem tržištu postoji neotkriveni tržišni potencijal.*
∗ *Naše najznačajnije tržište brzo raste.*

Regulatorne barijere rastu je varijabla dobivena kao prosječna vrijednost odgovora u kojoj su mjeri slijedeći faktori barijera rastu poduzeća:

∗ *Administrativne barijere (dozvole, procedure, ...)*
∗ *Nedostatak podrške države*
∗ *Politike gospodarstva koje provodi vlada*
∗ *Zakon o radu*
∗ *Porezna politika*
∗ *Korupcija*

Od definiranih teorijskih konstrukata značajnima su se pokazali i u početni model uključeni su: *iskustvo*, *potencijal*, *poduzetnička orijentacija*, *spremnost* i *organizacijsko učenje*.

Nakon odabira varijabli za koje je uočeno postojanje veze s ovisnom varijablom, u drugom koraku konstruirano je više modela, s različitim kombinacijama varijabli. Model koji se pokaže najboljim bit će prikazan u radu.

4.3 Model predikcije rasta malih i srednjih poduzeća u Hrvatskoj

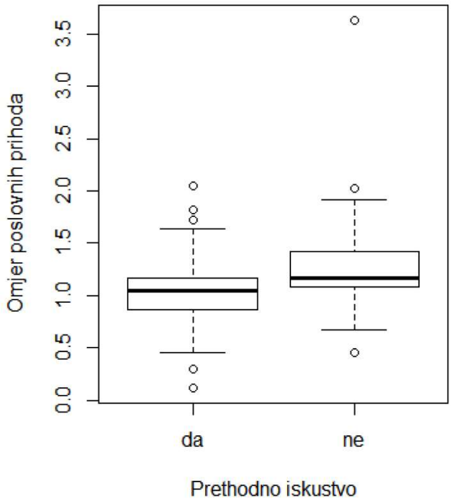
U ovom poglavlju, za izgradnju modela za procjenu rasta malih i srednjih poduzeća korištena je višestruka linearna regresija uz primjenu forward selekcijske procedure, a rezultati su dobiveni korištenjem programskog jezika *R*.

Primjenom forward stepwise selekcije[6] na početni model, analizom utjecajnih mjerenja i odbacivanjem varijabli koje se nisu pokazale statistički značajne, dobiven je linearni regresijski model koji uključuje 5 neovisnih varijabli koje su kratko predstavljene u nastavku rada.

Prethodno iskustvo je kategorijalna varijabla koja predstavlja prethodno iskustvo vlasnika poduzeća (0−ne, 1−da). U bazi podataka su 63 (40,38%) poduzetnika koji nemaju iskustvo i 93 (59,62%) poduzetnika koji imaju iskustvo u poduzetništvu iz vremena prije nego je počeo raditi za poduzeće.

[6] Postupak regresije korak po korak pri čemu se u početni model bez varijabli sukcesivno uvode nove neovisne varijable na svakom koraku regresije, do dobivanja najboljeg regresijskog modela. Najznačajnija varijabla, tj. varijabla s najmanjom *p*-vrijednosti se uključuje u model, a ostale varijable se uključuju u model jedna po jedna. Ovaj pristup odbacuje varijable koje ne doprinose ili neznatno doprinose predviđanju modela. Primjenom selekcijske procedure odabire se model s najmanjom AIC vrijednosti.

Na slijedećoj slici koja predstavlja usporedni boxplot varijable *prethodno iskustvo* može se uočiti postojanje razlika u očekivanom iznosu *omjera poslovnih prihoda* između poduzeća čiji poduzetnik ima prethodno iskustvo i poduzeća čiji je poduzetnik bez prethodnog iskustva.



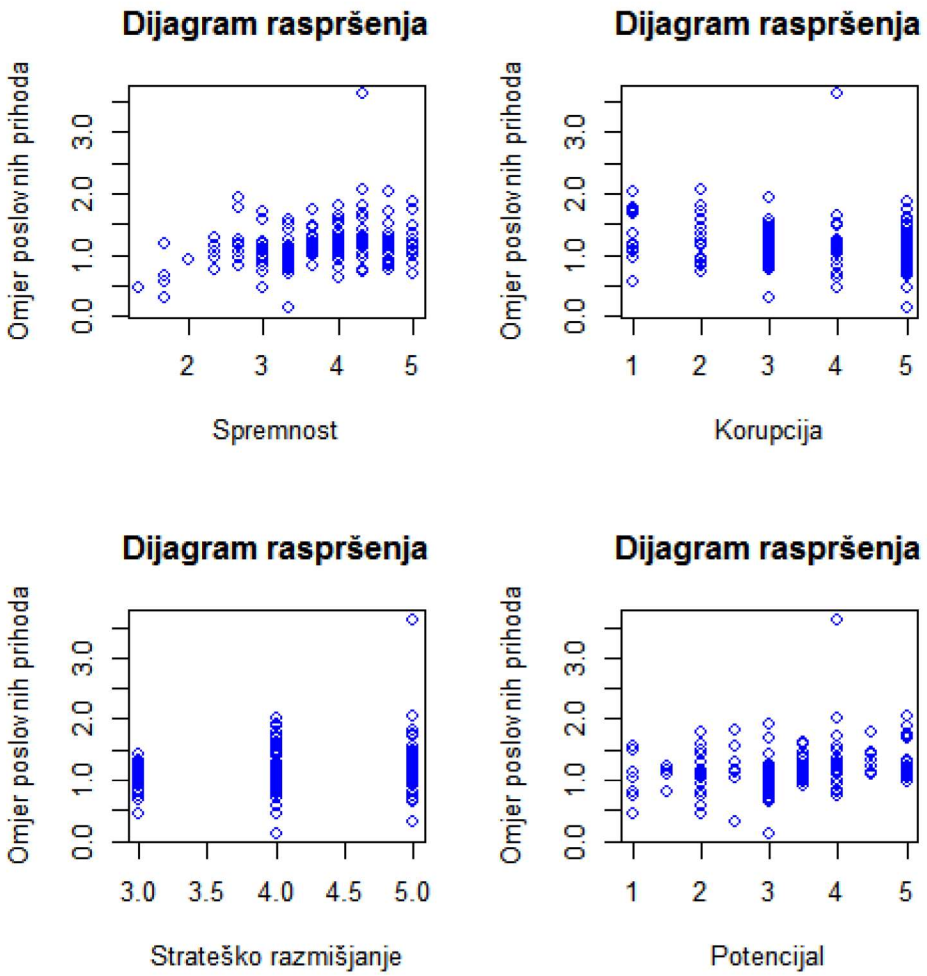
Slika 1: Usporedni boxplot varijable *prethodno iskustvo* u odnosu na *omjer poslovnih prihoda*

Statistički značajna razlika u očekivanom iznosu *omjera poslovnih prihoda* potvrđena je Welchovim t-testom o jednakosti očekivanja (p-vrijednost = 0.0003034). Može se zaključiti da poduzeća čiji poduzetnici imaju prethodno iskustvo u poduzetništvu, u prosjeku imaju veći *omjer poslovnih prihoda* od poduzeća čiji poduzetnik nema prethodnog iskustva u poduzetništvu. Za poduzeća čiji poduzetnik nema prethodnog iskustava on u prosjeku iznosi 1.039549, a za poduzeća čiji poduzetnik ima prethodno iskustvo 1.248828, tj. poduzeća čiji poduzetnik ima prethodno iskustvo u prosjeku ostvaruju veći rast poslovnih prihoda od poduzeća čiji poduzetnik nema prethodnog iskustva.

U nastavku se mogu vidjeti osnovna svojstva numeričkih varijabli *spremnost*, *korupcija*, *strateško razmišljanje* i *potencijal* te njihovi dijagrami raspršenja.

Ime varijable	**Min.**	**Median**	**Mean**	**Max.**	**St. Dev.**
Spremnost	1.333	3.667	3.776	5.000	0.8537412
Korupcija	1.000	4.000	3.716	5.000	1.327799
Strateško razmišljanje	3.000	4.000	4.219	5.000	0.6957288
Potencijal	1.000	3.000	3.287	5.000	1.052129

Tablica 1: Deskriptivna statistika varijabli



Slika 2: Dijagrami raspršenja numeričkih varijabli

Postojanje linearne veze pojedine neovisne varijable s ovisnom varijablom *omjer poslovnih prihoda* provodi se Pearsonovim testom korelacije čiji su rezultati prikazani u slijedećoj tablici.

Ime varijable	Koef. korelacije	p-vrijednost
Spremnost	0.260023	0.001045
Korupcija	-0.1953612	0.01452
Strateško razmišljanje	0.1680967	0.03594
Potencijal	0.2500607	0.001642

Tablica 2: Pearsonov test korelacije

Pearsonovim testom korelacije potvrđeno je postojanje pozitivne linearne veze između ovisne varijable *omjer poslovnih prihoda* i pojedine neovisne varijable *spremnost*, *potencijal* i *strateško razmišljanje*, dok je između ovisne varijable *omjer poslovnih prihoda* i neovisne varijable *korupcija* potvrđeno postojanje negativne linearne veze.

4.4 Ispitivanje pretpostavki modela

S obzirom da postoje situacije kada točnost procjene koeficijenata i njihova interpretacija može biti upitna, potrebno je izvršiti određene statističke testove kako bi odbacili mogućnost postojanja problema.

Značajan problem koji se pri modeliranju može pojaviti je multikolinearnost. Problem multikolinearnosti je prisutan u slučaju kada se neki od regresora može jako dobro opisati linearnom kombinacijom preostalih regresora. Faktor inflacije varijance (VIF) daje informaciju o jačini linearne veze pojedinog regresora s preostalim regresorima. Za svaku pojedinu varijablu $X_j, j = 1, \dots, k$ definira se kao:

$$VIF_j = \frac{1}{1 - R_j^2},$$

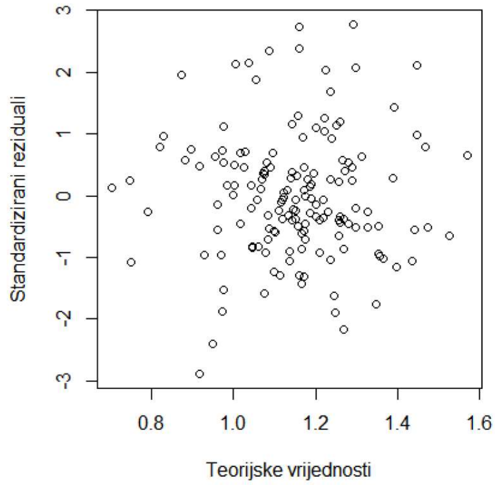
gdje je R_j^2 koeficijent determinacije u linearnom regresijskom modelu za X_j na temelju preostalih $(k-1)$ regresora. Ukoliko je $R_j^2 > 0.8$, odnosno $VIF_j > 5$ u modelu je prisutan ozbiljan problem multikolinearnosti. Drugi pokazatelj multikolinearnosti je kondicijski broj (CN) koji ukazuje na problem multikolinearnosti s numeričkog aspekta. U linearnom regresijskom modelu nije prisutan problem multikolinearnosti ako je kondicijski broj manji od 1000. U sljedećoj tablici prikazene su VIF vrijednosti regresora:

Ime varijable	**VIF**
Prethodno iskustvo	1.037259
Spremnost	1.430343
Korupcija	1.022260
Strateško razmišljanje	1.080515
Potencijal	1.363722

Tablica 3: VIF vrijednosti regresora

Budući da su sve vrijednosti faktora inflacije varijance manje od 5 i kondicijski broj iznosi 61.58064, može se zaključiti kako u modelu nije prisutan problem multikolinearnosti.

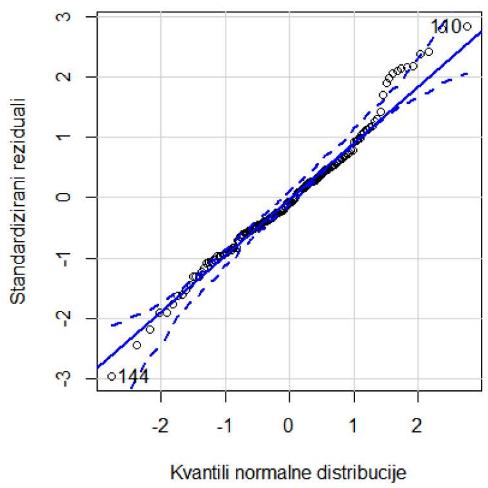
Homoskedastičnost, odnosno heteroskedastičnost grešaka modela je sljedeće na što treba obratiti pažnju pri modeliranju. S obzirom da su stvarne greške modela nepoznate, za zaključivanje o homoskedastičnosti koriste se standardizirani reziduali kao procjena za greške. Oni bi trebali biti "jednoliko raspršeni" kako bi mogli vjerovati u homogenost varijance modela.



Slika 3: Dijagrami raspršenja standardiziranih reziduala

Raspršenost standardiziranih reziduala izgleda jednoliko, no kao potvrdu provodimo Breush-Pagan test koji u nul-hipotezi sadrži homogenost varijance. U tu svrhu u programskom jeziku *R* korištena je naredba *ncvTest* paketa *car*. Dobivena p-vrijednost iznosi 0.92126 što znači da nema razloga sumnjati u homoskedastičnost grešaka modela.

Preostalo je ispitati normalnost grešaka i njihovu nezavisnost od regresora. Iz grafičkog prikaza može se uočiti odstupanje standardiziranih reziduala od normalne distribucije, što je dodatno potvrđeno Shapiro - Wilk testom (p-vrijednost=0.04534).



Slika 4: qqPlot standardiziranih reziduala

4.5 Interpretacija modela

Nakon ispitanih pretpostavki modela, slijedi prikaz i interpretacija koeficijenata modela.

	Ime varijable	Procjena	p-vrijednost
β_1	Prethodno iskustvo	0.18091	0.000155
β_2	Spremnost	0.06666	0.036497
β_3	Korupcija	-0.05367	0.002132
β_4	Strateško razmišljanje	0.05909	0.081465
β_5	Potencijal	0.04420	0.079449

Tablica 4: Tablica koeficijenata modela

Jednadžba konačnog modela glasi:

$$\begin{aligned}\mathbb{E}[\textit{Omjer poslovnih prihoda}] &= 0.59413 + 0.18091 * \textit{prethodno iskustvo}\\ &+ 0.06666 * \textit{spremnost} - 0.05367 * korupcija + 0.04420 * potencijal\\ &+ 0.05909 * stratesko\ razmisljanje\end{aligned}$$

Poduzeća u vlasništvu poduzetnika koji imaju prethodno iskustvo u poduzetništvu u prosjeku imaju veći omjer poslovnih prihoda za 0.18091 od poduzeća čiji poduzetnik nema prethodno iskustvo (95% interval pouzdanosti iznosi $[0.0888245, 0.272997]$). Poduzeća čiji poduzetnik ima prethodno iskustvo u istim ili sličnim poslovima ostvaruje veći rast poslovnih prihoda u odnosu na poduzeća čiji poduzetnik nema prethodnog iskustva u poduzetništvu.

Jedinično povećanje varijable spremnost u prosjeku se reflektira povećanjem omjera poslovnih prihoda za 0.06666 (95% interval pouzdanosti iznosi $[0.0042437, 0.1207226]$), uz sve ostale varijable fiksne. Što je veća spremnost poduzeća na snažan rast to je veći rast poslovnih prihoda koje poduzeće ostvaruje.

Jedinično povećanje korupcije kao barijere rastu u prosjeku se reflektira smanjenjem omjera poslovnih prihoda za 0.05367 (95% interval pouzdanosti iznosi $[-0.0875953, -0.0197424]$), uz sve ostale varijable fiksne. Što je korupcija veća barijera rastu to je manji rast poslovnih prihoda koje poduzeće ostvaruje.

Jedinično povećanje varijable strateško razmišljanje u prosjeku se reflektira povećanjem omjera poslovnih prihoda za 0.05909 (95% interval pouzdanosti iznosi $[-0.0074749, 0.1256612]$), uz sve ostale varijable fiksne. Što je veća sposobnost strateškog razmišljanja poduzetnika to je veći rast poslovnih prihoda koje poduzeće ostvaruje.

Jedinično povećanje varijable potencijal tržišta u prosjeku se reflektira povećanjem omjera poslovnih prihoda za 0.04420 (95% interval pouzdanosti iznosi $[-0,0052566, 0.0936475]$), uz sve ostale varijable fiksne. Što je veći tržišni potencijal to je veći rast poslovnih prihoda koje poduzeće ostvaruje.

Dobivene vrijednosti koeficijenata su logične i u skladu s dosadašnjim istraživanjima. Klepper (2001.) te Lee i Tsang (2001.) i N. Šarlija pokazali su da je poduzetničko iskustvo u pozitivnoj vezi s rastom poduzeća. Pozitivnu vezu između rasta poduzeća i varijable spremnost, a koja opisuje spremnost poduzeća na rast potvrdili su u svojim istraživanjima F. Diaz Hermelo i R. Vassolo (2007.), J. Wiklund, H. Potzelt i D. A. Shepherd (2009.) te N. Šarlija i A. Bilandžić (2018.). Z. Griliches i T. Klette (2000.), A. Del Monte i E. Papagni (2003.), te T. Yasuda (2005.) potvrdili su pozitivnu vezu između rasta poduzeća i strateškog razmišljanja. M. Ayyagari (2008.) u svojem istraživanju navodi kako je između ostalih korupcija mjera koja usporava rast poduzeća.

Vrijednost korigiranog R^2 iznosi $21,63\%$ što znači da je tek $21,63\%$ ukupne varijabilnosti podataka objašnjeno modelom. Ovaj podatak nije iznenađujući s obzirom da su i dosadašnja istraživanja dobila sličan rezultat (vidi [8]). Jedan od razloga tomu je činjenica da je jako teško obuhvatiti sve čimbenike koji utječu na rast poduzeća.

5 Primjena jackknife metode na linearni regresijski model za procjenu rasta poduzeća

U ovom dijelu rada primijeniti ćemo jackknife metodu na linearni regresijski model. S obzirom da je linearni regresijski model izgrađen na temelju 155 opažanja, primjenom jackknife metode dobit ćemo 155 skupova veličine 154 i na temelju svakog od njih ćemo izračunati standardne greške i intervale pouzdanosti za procjenitelje $\beta_i, i = 0, \ldots, 5$ i usporediti ih s rezultatima linearne regresije.

U programskom jeziku *R* korištena je funkcija *jackknife* za čije je pozivanje potrebno uključiti paket *bootstrap*.
Iz sljedećih tablica mogu se iščitati dobiveni podaci:

	Ime varijable	**Donja granica**	**Gornja granica**
β_0	Slobodni član	0.2633656	0.9250122
β_1	Prethodno iskustvo	0.0873925	0.2744154
β_2	Spremnost	-0.0028948	0.1361855
β_3	Korupcija	-0.0914434	-0.0159006
β_4	Strateško razmišljanje	0.0023695	0.1158106
β_5	Potencijal	-0.0085464	0.0969494

Tablica 5: Intervali pouzdanosti primjenom jackknife metode

	β_0	β_1	β_2	β_3	β_4	β_5
Homosk.	0.17727	0.04660	0.03159	0.01717	0.03369	0.02503
HCO	0.159914	0.045628	0.033387	0.018320	0.027463	0.025380
HC1	0.163102	0.046537	0.034053	0.018685	0.028011	0.025886
HC2	0.163896	0.046546	0.034334	0.018745	0.028124	0.026072
HC3	0.168008	0.047489	0.035316	0.019182	0.028805	0.026788
Jackknife	0.167464	0.047336	0.035202	0.019120	0.028712	0.026701

Tablica 6: Standardne greške procjenitelja

Uspoređujući dobivene podatke s linearnom regresijom može se uočiti da su 95% intervali pouzdanosti za varijable *prethodno iskustvo*, *spremnost*, *korupcija* i *potencijal* širi kod jackknife procjene. Uspoređujući intervale pouzdanosti različitih koeficijenata, ukoliko je jedan pouzdani interval dosta uži od drugog pouzdanog intervala sugerira da je odgovarajući koeficijent procijenjen s većom preciznošću. Na temelju navedenog može se zaključiti da je koeficijent uz varijablu *korupcija* procijenjen s većom preciznošću u odnosu na ostale koeficijente u modelu.

Uspoređujući vrijednosti standardnih grešaka jackknife metodom može se uočiti kako su vrijednosti najbliže vrijednostima dobivenim koristeći procjenitelj $HC3$. Procjenitelj $HC3$ je najkonzervativniji, u smislu da je veći od svih ostalih procjenitelja koji se koriste u slučaju heteroskedastičnog modela. Preporuka je koristiti $HC3$ procjenitelj kada je narušena pretpostavka homoskedastičnosti modela.

Literatura

[1] P. Davidsson, P. Steffens and J. Fitzsimmons, *Growing profitable or growing from profits: Putting the horse in front of the cart?*, Journal of Business Venturing 24(2009.), no. 4, 388-406.

[2] F. Diaz Hermelo, R. Vassolo, *The determinants of firms' growth: An empirical examination*, Abante 10(2007.), 3-20.
http://www.abante.cl/files/ABT/Contenidos/Vol-10-N1/Diaz.pdf

[3] P. A. Esemokumo, R. Bekesuoyeibo,I. Okenwe, *Jackknife Algorithm on Linear Regression Estimation*, International Journal for Research in Mathematics and Statistics 1(2015.), 34-40.
https://gnpublication.org/index.php/ms/article/view/910/718

[4] B. E. Hansen, *Econometrics*, University of Wisconsin, 2000.

[5] R. Johns, *Likert items and scales*, University of Strathclyde, 2010.
https://dam.ukdataservice.ac.uk/media/262829/discover_likertfactsheet.pdf

[6] N. Šarlija, A. Bilandžić, *Modeling and Predicting the Growth of SME*, Proceedings of the 6th International Conference on Innovation and Entrepreneurship University of the District of Columbia Washington DC, USA 5-6 March 2018., Washington: Academic Conferences and Publishing International Limited Reading(2018.), 391-398.

[7] G. Trovato, L. Becchetti, *The Determinants of Growth for Small and Medium Sized Firms. The Role of the Availability of External Finance*, Small Business Economics 10(2002.), 291-306.
https://www.researchgate.net/publication/5158138_The_Determinants_of_Growth_for_Small_and_Medium_Sized_Firms_The_Role_of_the_Availability_of_External_Finance

[8] J. Wiklund, H. Patzelt, D. A. Shepher, *Building an Integrative Model of Small Business Growth*, Small Business Economics 32(2009.), 351-374.
https://www.researchgate.net/publication/42242049_Building_an_Integrative_Model&_of_Small_Business_Growth

[9] Sherman - Morrison formula, Wikipedia
https://en.wikipedia.org/wiki/Sherman%E2%80%93Morrison_formula

Sažetak

U radu je predstavljena jackknife metoda ponovljenog uzorkovanja i njena primjena na linearni regresijski model. U prvom dijelu rada kratko su predstavljeni osnovni pojmovi i pretpostavke linearne regresije. Zatim je predstavljena jacknnife metoda i njena primjena u statistici. Drugi dio rada obuhvaća izgradnju linearnog regresijskog modela za procjenu rasta malih i srednjih poduzeća na temelju podataka prikupljenih od 156 hrvatskih poduzeća u razdoblju 2012. – 2015. U dobivenom regresijskom modelu nije prisutan problem multikolinearnosti, kao ni heteroskedastičnosti, te je logičan i u skladu s prethodnim istraživanjima. Jackknife metoda za izračun standardne greške procjenitelja i pouzdanih intervala primjenjena je na regresijski model i uspoređeni su dobiveni rezultati.

Ključne riječi

Metode ponovljenog uzorkovanja, jackknife metoda, linearna regresija, metoda najmanjih kvadrata

Jaccknife method application to a linear regression model of company growth

Summary

In this paper we introduced jackknife method and its application to linear regression model. The basic terms and assumptions of linear regression are briefly presented in the first part of the paper as well as the jackknife method and its application in statistics. The second part of the paper includes the construction of linear regression model for estimating the growth of small and medium-sized enterprises based on data collected from 156 Croatian enterprises in the period 2012. – 2015. The obtained regression model has no problem of multicollinearity, nor heteroscedasticity, and it is logical and in accordance with previous research. The jackknife method for calculating the standard error of the estimator and confidence intervals was applied to the regression model and the given results were compared.

Keywords

Resampling methods, jackknife method, linear regression, least square estimator

Životopis

Rođena sam 26. rujna 1993. godine u Makarskoj. 2000. godine započinjem obrazovanje u Osnovnoj školi oca Petra Perice, a 2008. godine se upisujem u SŠ fra A.K. Miošića, ekonomski smjer u Makarskoj. Nakon srednjoškolskog obrazovanja, 2012. godine upisujem preddiplomski studij matematike na Odjelu za matematiku u Rijeci. Preddiplomski studij završavam 2016. godine s temom završnog rada *Modeliranje privatnosti u sustavima suradnje* pod mentorstvom dr. sc. Tajane Ban Kirigin. Iste godine upisujem diplomski studij Financijske matematike i statistike na Odjelu za matematiku u Osijeku. Tijekom diplomskog studija odradila sam stručnu praksu u Raiffeisen bank d.d. u Zagrebu, u odjelu Upravljanje kreditnim rizicima retail klijenata, tim za strateški razvoj.