Ekonometrija – I deo

Doktorske studije

Predavač: Aleksandra Nojković

Beograd, školska 2023/24

Struktura predavanja

- Klasični višestruki linearni regresioni model
  - metod ONK
  - matrična notacija
  - pretpostavke višestrukog KLRM

- Pokazatelji kvaliteta ocenjenog modela. Korelacija u višestrukom KLRM. Običan i korigovani R^2

- Testovi linearnih ograničenja na parametre

Dvostruki linearni regresioni model

- Ako pretpostavimo model sa dve objašnjavajuće promenljive: $Y=\beta_0+\beta_1 X_1+\beta_2 X_2+\varepsilon$
- Populaciona regresiona jednačina (za $E(\varepsilon)=0$) je:

$$E(Y)=\beta_0+\beta_1 X_1+\beta_2 X_2$$

- Parametri β_0, β_1 i β_2 su populacioni parametri ili regresioni koeficijenti.
- Populaciona reg. **jednačina** ne opisuje pravu, nego **ravan.**
- Parametar β_0 je odsečak (presek ravni sa y-osom), a parametri β_1 i β_2 su **parcijalni koeficijenti nagiba.**

Dvostruki linearni regresioni model (nastavak)

- Uključivanjem konkretnih podataka model postaje:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i, \quad \text{za } i = 1,2,\ldots,n.$$

- Uzoračka regresiona funkcija je:

$$\hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i},$$

gde su b_0, b_1 i b_2 ocene parametara.

- Ocene nepoznatih parametara primenom metoda ONK, koji se sastoji u minimiziranju sume kvadrata reziduala.

Ocene metodom ONK

- Eksplicitni izrazi za ocene metodom ONK:

$$b_0 = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2,$$

$$b_1 = \frac{\sum_{i=1}^{n} x_{1i} y_i \sum_{i=1}^{n} x_{2i}^2 - \sum_{i=1}^{n} x_{1i} x_{2i} \sum_{i=1}^{n} x_{2i} y_i}{\sum_{1=1}^{n} x_{1i}^2 \sum_{1=1}^{n} x_{2i}^2 - \left(\sum_{i=1}^{n} x_{1i} x_{2i}\right)^2},$$

$$b_2 = \frac{\sum_{i=1}^{n} x_{2i} y_i \sum_{i=1}^{n} x_{1i}^2 - \sum_{i=1}^{n} x_{1i} x_{2i} \sum_{i=1}^{n} x_{1i} y_i}{\sum_{1=1}^{n} x_{1i}^2 \sum_{1=1}^{n} x_{2i}^2 - \left(\sum_{i=1}^{n} x_{1i} x_{2i}\right)^2}.$$

- Pokazati...
- Važe relacije:

$$\sum_{i=1}^{n} e_i = 0; \ \sum_{i=1}^{n} X_{1i} e_i = 0; \ \sum_{i=1}^{n} X_{2i} e_i = 0.$$

Pretpostavke KLRM (višestrukog)

1. *$E(\varepsilon_i)=0$, za svako i.*
2. *$Var\ (\varepsilon_i)=E(\varepsilon_i^2)=\sigma^2$, za svako i.*
3. *$Cov(\varepsilon_i,\varepsilon_j)=E(\varepsilon_i\varepsilon_j)=0$, za svako i,j, tako da $i\neq j$.*
4. *$E(\varepsilon_i X_i)=0$, za svako i.*
5. *$\varepsilon_i \sim N(0,\ \sigma^2)$.*
6. Ne postoji tačna linearna zavisnost izmedu objašnjavajućih promenljivih ($r_{x1x2}\neq 1$, tj. jedna objašnjavajuća promenljiva nije linearna funkcija druge).

Klasicni višestruki linearni regresioni model

- Analiticki oblik višestrukog linearnog regresionog modela:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_{k-1} X_{k-1i} + \varepsilon_i, \quad \text{za } i = 1,2,\ldots,n.$$

- Parametri β_1, $\beta_{2,\ldots,}$ β_{k-1} su parcijalni koeficijenti nagiba.

Npr: *ako se* X_{1i} poveća za jednu jedinicu, očekivana promena Y_i je β_1 jedinica, pod pretpostavkom da se ne menja uticaj ostalih objašnjavajućih promenljivih $X_2, X_{3,\ldots,} X_{k-1}$.

Višestruki linearni regresioni model (matrična notacija)

- U matričnoj notaciji (k=3):

$$y = XB + e,$$

gde je: y (nx1) vektor kolona; X (nxk) matrica; B (kx1) vektor kolona; e (nx1) vektor kolona.

- Matrica X: svaki red predstavlja vrednost svih eksplanatornih prom. koje odgovaraju jednoj opservaciji, a svaka kolona predstavlja sve vrednosti jedne eksplanatorne prom. u uzorku.

Dvostruki linearni regresioni model (k=3): matrična notacija

- Model: $Y = XB + e$,

gde je:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, X = \begin{bmatrix} 1 & X_{11} & X_{21} \\ 1 & X_{12} & X_{22} \\ \dots & \dots & \dots \\ 1 & X_{1n} & X_{2n} \end{bmatrix}, \mathrm{B} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}, e = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Alternativna forma polaznog modela

- Matrična forma matričnog modela u centriranim podacima:

$$y = xb + e,$$

pri čemu za n –opservacija dolazimo do ocena ONK sadržanih u vektoru b:

$$b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_{k-1} \end{pmatrix} = \left(x' x\right)^{-1} x'y.$$

Određivanje standardnih grešaka ocena u višestrukom modelu

- U višestrukom modelu ocene varijanse σ^2 je:

$$\hat{\sigma}^2 = s^2 = \frac{e'e}{n-k}$$

- U višestrukom modelu ocene varijansi vektora ocena je:

$$\widehat{\mathrm{var}}(b) = s_b^2 = s^2\left(x'x\right)^{-1}$$

Testiranje pojedinačne značajnosti parametara

Pretpostavimo da je hipoteza od interesa:

H_0: $\beta_j = 0$, H_1: $\beta_j \neq 0$ $j=1,2,...,k-1.$

- Diskriminacija između postavljenih hipoteza realizuje se primenom t - testa, primenom test-statistika :

$$test-statistika = \frac{b_j - \beta_j}{s_{b_j}} : t\,(n-k).$$

- Odnosno, značajnost pojedinačnog uticaja svake od objašnjavajucih promenljivih na zavisnu promenljivu se proverava kao:

$$test-statistika = \frac{b_j}{s_{b_j}} : t\,(n-k).$$

- **Pokazati** da ova test-statistika poseduje t-raspodelu sa (n-k) stepeni slobode...

Specifičan tip t-testa: *t*-odnos

- Kao i u slucaju jednostavnog modela, i u višestrukoj regresiji se koristi test-statistika oblika:

$$test - statistika = \frac{b_i - \beta_i *}{s_{b_i}} : t\,(n-k).$$

- Pretpostavimo da je hipoteza od interesa:

  H_0: $\beta_i = 0$, H_1: $\beta_i \neq 0$ $i=1,2,...,k$.
- U uslovima validnosti nulte hipoteze test-statistika je:

$$test - statistika = \frac{b_i}{s_{b_i}} : t\,(n-k).$$

- Buduci da je u pitanju kolicnik ocene i odgovarajuce standardne greške te ocene, ova statistika se naziva ***t*-odnos.** Na ovaj način se proverava značajnost pojedinačnog uticaja svake od objašnjavajucih promenljivih na zavisnu promenljivu.

Korelacija u jednostavnom i višestrukom (dvostrukom) modelu

- Za jednostavni model važi:

$$r_{xy} = \hat{\beta}\frac{s_x}{s_y} = b\frac{s_x}{s_y}.$$

- U višestrukom modelu sa npr. dve objašnjavajuće promeljive X_1 i X_2, važi da *koeficijent parcijalne korelacije između Y i X_1* (koeficijent prvog reda ili koef. korelacija po odbitku uticaja X_2) odgovara znaku ocenjenog regresionog koeficijenta b_1:

$$r_{YX_1 \bullet X_2} = \frac{r_{YX_1} - r_{YX_2} r_{X_1X_2}}{\sqrt{1 - r_{YX_2}^2}\sqrt{1 - r_{X_1X_2}^2}}.$$

- **Objasniti...**

Koeficijent determinacije dvostrukom modelu

- Ukupne varijacije se i u dvostrukom modelu mogu zapisati kao:

$$\underbrace{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}_{\text{ukupne var ijacije}} = \underbrace{\sum_{i=1}^{n}\left(\hat{Y} - \bar{Y}\right)^2}_{\text{objaš . var ijacije}} + \underbrace{\sum_{i=1}^{n} e_i^2}_{\text{neobjaš . var ijacije}}$$

- Koeficijent determinacije R^2 deo ukupnih varijacija zavisne promenljive objašnjen kretanjem svih nezavisnih promenljivih, odnosno za dve objašnjavajuće promenljive definiše se kao:

$$R^2 = \frac{\sum_{i=1}^{n} \hat{y}_i^2}{\sum_{i=1}^{n} y_i^2} = \frac{b_1 \sum_{i=1}^{n} x_{1i} y_i + b_2 \sum_{i=1}^{n} x_{2i} y_i}{\sum_{i=1}^{n} y_i^2}.$$

- **Pokazati...**

Testiranje statističke značajnosti cele regresije

- Hipoteze od interesa:

$$H_0: \beta_1 = \beta_2 = \ldots = \beta_{k-1} = 0 \Leftrightarrow H_0: R^2 = 0$$

$$H_1: \text{hipoteza}\, H_0 \text{ nije tacna} \Leftrightarrow H_1: R^2 \neq 0$$

- Nulta hipoteza: regresija nije statistički značajna (zajednički uticaj objašnjavajućih promenljivih nije statistički značajan).

- Alternativna hipoteza: objašnjavajuće promenljive ostvaruju statistički značajan uticaj na kretanje zavisne promenljive (bar jedan od parametara je značajno različit od nule).

Ispitivanje kvaliteta regresije na osnovu koeficijenta determinacije

- Relevantna statistika je:

$$F_{n-k}^{k-1} = \frac{\text{Objasnjeni varijabilitet / (k - 1)}}{\text{Neobjasnjeni varijabilitet}/(n-k)}$$

$$F_{n-k}^{k-1} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

- Pravilo odlučivanja:
  - Ako je izračunata vrednost date statistike veća od kritične vrednosti *F*-raspodele sa *k-1* i *n-k* stepeni slobode, tada se nulta hipoteza odbacuje uz izabrani nivo značajnosti.

- Različite situacije sa stanovišta statističke značajnosti ocena koef. determinacije i pojedinačnih regresionih koeficijenata (**objasniti!).**

Korelacija u višestukom KLRM

- Uopštenjem za polazni model sa ukupno k parametara (k-1 parametrom parcijalnih nagiba i sl. članom), koeficijent determinacije R^2 se izračunava kao:

$$R^2 = \frac{\sum_{i=1}^{n} \hat{y}_i^2}{\sum_{i=1}^{n} y_i^2} = \frac{b_1 \sum_{i=1}^{n} x_{1i} y_i + b_2 \sum_{i=1}^{n} x_{2i} y_i + \ldots + b_{k-1} \sum_{i=1}^{n} x_{k-1i} y_i}{\sum_{i=1}^{n} y_i^2}.$$

- R^2 se uvek povećava sa dodavanjem novih objašnjavajućih promenljivih.

- R^2 je krajnje nepouzdan pokazatelj u regresionoj analizi vremenskih serija kada vrednost, na primer 0.999, ne mora nužno pokazivati ništa.

Korigovani koeficijent determinacije R^2

- Koriguje se koeficijent determinacije sa ciljem dobijanja pokazatelja koji se neće neopravdano povećavati sa rastom broja objašnjavajućih promenljivih.
- Novi pokazatelj: korigovani koeficijent determinacije $\overline{R}^2$

$$R^2 = 1 - \frac{\sum_i e_i^2}{\sum_i y_i^2}$$

$$\overline{R}^2 = 1 - \frac{\sum_i e_i^2 / (n-k)}{\sum_i y_i^2 / (n-1)} = 1 - \frac{n-1}{n-k}\left(1 - R^2\right)$$

- Korigovani koeficijent determinacije je uvek manji od običnog koeficijenta determinacije. Koeficijenti su jednaki samo za jednostavni model bez slobodnog člana.

Kriterijumi za izbor optimalnog skupa objašnjavajućih promenljivih

- Skup promenljivih **koji maksimizira** vrednost korigovanog koef. determinacije, u isto vreme **minimizira** s^2.

- Navedeno je posledica relacije:

$$s^2 = \left(1 - \bar{R}^2\right) \frac{\sum y_i^2}{n-1}.$$

Kriterijumi za izbor optimalnog skupa objašnjavajućih promenljivih (nastavak)

- **Informacioni kriterijum (VS)** je zbir dve komponente koje različito reaguju na promenu broja parametara modela (K):

$$IC(K) = \ln(s^2) + g(K/n).$$

  - Model sa najmanjom vrednošću IC je optimalan uz uslov da su valjane sve pretpostavke KLRM
  - AIC – Akaikeov informacioni kriterijum (g=2)
  - SC – Švarcov informacioni kriterijum (g=ln(n))
  - HQC – Hana-Kvinov kriterijum (g=2lnln(n)).

Testiranje opštih linearnih ograničenja na parametre

- Ekonomski kriterijumi često zahtevaju da koeficijenti u ocenjenom modelu zadovoljavaju izvesna ograničenja (npr. *konstantni prinosi u Cobb-Douglasovoj funkciji*; *odsustvo iluzije novca* (zbir elastičnosti tražnje s obzirom na nominalni dohodak i cene je jednak nuli i sl.)).

- Dva alternativna postupka:

1) Oceniti *f*-ju **ne vodeći računa o ograničenjima**, a zatim testirati da li ocenjeni koeficijenti zadovoljavaju zahtevane restrikcije.
2) Oceniti **model sa inkorporiranim ograničenjima**, a zatim testirati značajnost razlike između ocena takvog modela (pod ograničenjem) i modelom bez ograničanje.

Testovi linearnih ograničenja na parametre

- Pri testiranju *jednog ograničenja* (jednostavna nulta hipoteza) koristi se t-test (ili alternativno, F-test).
- U slučaju testa *više ograničenja* (kad je nulta hipoteza složena) F-statistika.

Prvi postupak testiranja: t-test
Testiranje jednog linearnog ograničenja

- Posmatramo nultu hipotezu kojom se tvrdi da je istinito linearno ograničenje oblika:

$$H_0 : r_0\beta_0 + r_1\beta_1 + \ldots + r_{k-1}\beta_{k-1} = RB = q.$$

- Alternativna hipoteza je:

$$H_1 : r_0\beta_0 + r_1\beta_1 + \ldots + r_{k-1}\beta_{k-1} = RB \neq q,$$

pri čemu su parametri r_0, r_1, ..., r_{k-1} poznati.

- Primenom metoda ONK dobijajnu se ocene b_0, b_1,..., b_{k-1} i obrazuje se ocena linearne kombinacije prema nultoj hipotezi:

$$r_0 b_0 + r_1 b_1 + \ldots + r_{k-1} b_{k-1} = \hat{q}.$$

Testiranje jednog linearnog ograničenja (nastavak)

Skalarni oblik ove ocene varijanse je:

$$s^2\left(\hat{q}\right) = r_0^2 s_{b_0}^2 + r_1^2 s_{b_1}^2 + \cdots + r_{k-1}^2 s_{k-1}^2 + 2r_0 r_1 \,\hat{\mathrm{cov}}\left(b_o, b_1\right) + \cdots + 2r_{k-2} r_{k-1} \,\hat{\mathrm{cov}}\left(b_{k-2}, b_{k-1}\right).$$

- Nulta hipoteza se testira na osnovu t_{n-k} raspodele:

$$t_{n-k} = \frac{\hat{q} - q}{s\left(\hat{q}\right)}.$$

Testiranje g=1 ograničenja

- Ukoliko je potrebno testirati samo jedno ograničenje na parametre modela (g=1), kao alternativa t-testu, može se koristiti F^1_{n-k} raspodela.

- F test statistika za proveru tačnosti jednog ograničenja definisanog nultom hipotezom

$$F^1_{n-k} = \frac{(n-k)}{1}\frac{(R^2 - R_o^2)}{1-R^2},$$

gde se oznaka (O) odnosi na model sa ograničenjem, dok se za model bez ograničenja ponekad koristi oznaka (B), *g* je broj ograničenja (ovde g=1).

Drugi postupak testiranja (više ograničenja): F-test

- Distinkcija između nulte i alternativne hipoteze vrši se pomoću sledeće F statistike:

$$F_{n-k}^{g} = \frac{(n-k)}{g}\frac{(R^2 - R_o^2)}{1 - R^2},$$

gde se oznaka (O) odnosi na model sa ograničenjem, dok se za model bez ograničenja ponekad koristi oznaka (B), ***g* je broj ograničenja.**

- Na ovaj način se testira važenje **većeg broj ograničenja** (značajnost podskupa parametara modela predstavlja specijalan slučaj ovog testa).

Primer

- Ako je Cobb-Douglasova proizvodna funkcija oblika $Y=\beta_0 K^{\beta 1} L^{\beta 2} e^u$, gde je Y proizvodnja, K kapital i L rad, ocenjena na bazi podataka za SAD za period 1Q1960-1Q1991 (primer: Asteriou and Hall, 2015):

$$\log \hat{Y} = 4.514 + \underset{(0.008)}{0.383} \log K + \underset{(0.088)}{0.624} \log L \qquad R^2 = 0.968;$$

$$\sum e^2 = \sum e_B^2 = 0.3315.$$

- T=125

Primer (nastavak)

a) Testirati statističku značajnost pojedinačnog uticaja rada i kapitala na proizvodnju, na nivou značajnosti $\alpha = 0{,}05$.

b) Testirati statističku značajnost istovremenog uticaja objašnjavajućih promenljivih na zavisnu, na nivou značajnosti $\alpha = 0{,}05$.

c) Testirati hipotezu da rad i kapital **ostvaruju jednak efekat** na kretanje proizvodnje.

d) Testirati hipotezu da **rad ostvaruje dvostruko manji efekat na kretanje** proizvodnje u odnosu na kapital.

e) Testirati hipotezu da su u datoj grani industrije **prinosi konstantni** ($\beta_1+\beta_2=1$). Predložiti pojednostavljenje funkcije ako se ocenjuje pod datim ograničenjem.

Primer: CD funkcija (nastavak)

- Model bez ograničenja:

$$\log \hat{Y} = 4.514 + \underset{(0.008)}{0.383} \log K + \underset{(0.088)}{0.624} \log L \qquad R^2 = 0.968;$$

$$\sum e^2 = \sum e_B^2 = 0.331497,$$

cov(b1,b2)=-0.0003372;

- Model ocenjen pod ograničenjem (pod e)H_0: $\beta_1+\beta_2=1$):

$$(\log \hat{Y} - \log L) = 4.533 + \underset{(0.007116)}{0.3834} (\log K - \log L) \qquad R^2 = 0.9593; \sum e_0^2 = 0.331516.$$

- Rezultat Wald-ovog testa:

t-statistic	0.082798	122	0.9341
F-statistic	0.006855	(1, 122)	0.9341