OCENJIVANJE PARAMETARA NORMALNOG LINEARNOG REGRESIVNOG MODELA

Ivan Pavkov[*]

Sažetak: *Ekonometrijska istraživanja imaju za cilj proveru važećih ekonomskih zakonitosti i predviđanje vrednosti ekonomskih varijabli. Normalni linearni regresivni model pretpostavlja linearnu funkcionalnu zavisnost dve varijable, pri čemu je uticaj drugih pojava na vrednost zavisne varijable zanemarljiv u odnosu na uticaj nezavisne varijable. Nakon ocenjivanja nepoznatih parametara ovog modela, u mogućnosti smo da odredimo regresijsku liniju uzorka koja predstavlja ocenu regresijske linije populacije i na osnovu koje možemo izvršiti predviđanje.*

Ključne reči: *ekonomska teorija, ekonometrija, normalni linearni regresivni model, regresijska linija uzorka, predviđanje*

Abstract: *Aim of an econometric research is either evaluation of valid economic laws or prediction of economic variables' values. Assumption of normal linear regressive model is linear functional dependance of two variables, while other items have no significant influence on dependent variable value compared to influence of independent variable. After estimating unknown parameters of this model, sample regression line can be determined and it is, in fact, an estimation of population regression line. Finally, prediction is obtained using sample regression line.*

Key words: *economic theory, econometry, normal linear regressive model, sample regression line, prediction*

Uvod

Odnos između varijabli X i Y opisan sa Y= f(X) nazivamo determinističkim ako svakoj vrednosti varijable X odgovara tačno jedna vrednost varijable Y. Za varijable X i Y koje su u determinističkom odnosu, kažemo da Y funkcionalno zavisi od X ako f(X) nije konstanta za sve vrednosti X.

Odnos između varijabli X i Y opisan sa Y= f(X) nazivamo stohastičkim ako svakoj vrednosti varijable X pridružujemo celokupnu raspodelu varijable Y. Za varijable X i Y koje su u stohastičkom odnosu, kažemo da Y funkcionalno zavisi od X ako raspodela varijable Y nije ista za sve vrednosti X.

Matematička statistika se zasniva na pretpostavci kontrolisanog eksperimenta u kome se menjaju isključivo vrednosti varijabli čiji se odnosi ispituju. Međutim, ekonomski procesi najčešće se odvijaju u uslovima nekontrolisanog eksperimenta, te se za njihovo proučavanje koriste prilagođene metode matematičke statistike, tzv. ekonometrijske metode.

Ekonomska teorija, pojednostavljeno gledano, u svojoj suštini predstavlja sistem deterministički opisanih odnosa između varijabli. Sa druge strane, u ekonometriji se isključivo bavimo stohastičkim odnosima čiji je najjednostavniji oblik jednostavni linearni regresijski model:

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

gde je X nezavisna varijabla, Y zavisna varijabla, ε slučajno odstupanje, α i β nepoznati parametri regresije. Indeks i se odnosi na i-to opažanje.

Osnovne pretpostavke koje se pridružuju jednostavnom linearnom regresijskom modelu su:

1. ε_i ima normalnu raspodelu
2. $E(\varepsilon_i) = 0$

[*] Ivan Pavkov, Diplomirani inženjer matematike

3. $Var(\varepsilon_i) = \sigma$

4. $Cov(\varepsilon_i, \varepsilon_j) = 0 \ (i \neq j)$

5. X je nestohastička varijabla sa fiksnim vrednostima u ponovljenim uzorcima i za bilo koji uzorak veličine n suma $\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ je različita od nule i $\lim \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ je konačan broj.

Jednostavni linearni regresijski model zajedno sa pretpostavkama 1.-5. naziva se normalni linearni regresijski model.

Osnovni preduslov uspešnog ekonometrijskog istraživanja je specifikacija adekvatnog ekonometrijskog modela koja se postiže formiranjem odgovarajućeg matematičkog oblika funkcije, uvrštavanjem bitnih i izostavljanjem nebitnih varijabli.

Pretpostavke 1.- 4. se odnose na definisanje raspodele odstupanja. Dakle, za svaku vrednost varijable X odstupanje ima normalnu raspodelu oko nule sa jednakim varijansama σ^2, pri čemu odstupanja koja odgovaraju dvema različitim vrednostima varijable X nisu u korelaciji.

Pretpostavka 5. znači da sve vrednosti varijable X u uzorku ne smeju biti međusobno jednake i da ne mogu beskonačno ni rasti ni opadati sa povećanjem veličine uzorka.

Iz pretpostavki 1.- 4., očigledno je da je jedini nepoznati parametar raspodele odstupanja njena varijansa σ^2.

Drugim rečima, opisani normalni linearni regresijski model ima tri nepoznata parametra: α, β i σ^2.

Pokušajmo sada da odredimo raspodelu zavisne varijable Y. Vidimo da važi:

$$E(Y_i) = E(\alpha + \beta X_i + \varepsilon_i) = \alpha + \beta X_i$$

$$Var(Y_i) = E[Y_i - E(Y_i)]^2 = E[(\alpha + \beta X_i + \varepsilon_i) - (\alpha + \beta X_i)]^2 = E(\varepsilon_i^2) = \sigma^2$$

$$\text{Za } i \neq j,\ Cov(Y_i, Y_j) = E(Y_i Y_j) - E(Y_i)E(Y_j) = E(Y_i)E(Y_j) - E(Y_i)E(Y_j) = 0$$

Iz predhodne tri jednakosti zaključujemo da Y_i, $i=1,\ldots,n$ imaju normalne raspodele, svaka sa očekivanjem $\alpha+\beta X_i$, $i=1,\ldots,n$ i da bilo koje dve nisu u korelaciji.

Kada se ocene vrednosti α i β sa $\hat{\alpha}$ i $\hat{\beta}$ dobijamo regresijsku liniju uzorka $\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$, gde je $\hat{Y}_i$ prilagođena vrednost promenljive Y_i.

Regresijska linija uzorka $Y_i = \hat{\alpha} + \hat{\beta} X_i + e_i$, gde su sa e_i označeni reziduali $e_i = Y_i - \hat{Y}_i$, zapravo predstavlja ocenu regresijske linije populacije $Y_i = \alpha + \beta X_i + \varepsilon_i$, pri čemu se reziduali e_i mogu smatrati ocenama odgovarajućih odstupanja ε_i.

Ocenjivanje parametara metodom najmanjih kvadrata

Princip metode najmanjih kvadarata (LSE) je minimalizacija sume kvadratnih odstupanja opaženih vrednosti od njihove sredine.

Znamo da su ocene parametara α i β metodom najmanjih kvadrata:

$$\hat{\beta} = \frac{\Sigma(X_i - \overline{X})(Y_i - \overline{Y})}{\Sigma(X_i - \overline{X})^2}$$

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

Primer:

Posmatrajmo podatke o cenama i količinama prodatih akumulatora u jednoj prodavnici u toku deset uzastopnih dana date u tabeli 1. Neka nezavisna varijabla X_i bude obračunata cena i-tog dana, a zavisna varijabla Y_i broj prodatih akumulatora i-tog dana.

X_i [din/kom]	Y_i [kom]
1500	30
1400	35
1300	50
1100	55
1100	50
950	60
850	70
800	70
750	80
750	70

Tabela 1

Pretpostavimo da je funkcija potražnje oblika $Y_i = \alpha + \beta X_i + \varepsilon_i$ i da su zadovoljeni uslovi 1. – 5. normalnog linearnog regresijskog modela.

Lako dobijamo da je:

$$\overline{X} = 1050$$

$$\overline{Y} = 57$$

Ocenimo sada parametre α i β metodom najmanjih kvadrata

$$\hat{\beta} = \frac{-12150 - 7700 - 1750 - 100 - 350 - 300 - 2600 - 3250 - 6900 - 3900}{202500 + 122500 + 62500 + 2500 + 2500 + 10000 + 40000 + 62500 + 90000 + 90000}$$

$$\hat{\beta} = \frac{-39000}{685000}$$

$$\hat{\beta} = -0{,}057$$

$$\hat{\alpha} = 57 - (-0{,}057) \cdot 1050 = 57 + 59{,}85 = 116{,}85$$

Dakle, ocenjena regresijska linija uzorka je $\hat{Y}_i = 116{,}85 - 0{,}057\ X_i$ i predstavlja ocenu krive potražnje.

Najbolje linearno nepristrasno ocenjivanje

Metoda najboljeg linearnog nepristrasnog ocenjivanja (BLUE) zahteva da ocene parametara α i β, $\hat{\alpha}$ i $\hat{\beta}$, budu linearne kombinacije opažanja iz uzorka, da budu nepristrasne i da njihova varijansa bude manja od varijanse bilo koje druge linearne nepristrasne ocene.

Ocene za α i β su identične ocenama dobijenim metodom najmanjih kvadrata.

Međutim, metodom najboljeg linearnog nepristrasnog ocenjivanja se kao usputni rezultat određivanja ocena za α i β dobijaju ocene za njihove varijanse:

$$\mathrm{Var}(\hat{\beta}) = \sigma^2 \frac{1}{\Sigma(X_i - \overline{X})^2}$$

$$\mathrm{Var}(\hat{\alpha}) = \sigma^2 \left(\frac{1}{n} + \frac{\overline{X}^2}{\Sigma(X_i - \overline{X})^2}\right)$$

Napomena

Za ocene parametara α i β dobijene metodom najboljeg linearnog nepristrasnog ocenjivanja koristili smo iste oznake kao kod ocenjivanja metodom najmanjih kvadrata ($\hat{\alpha}$ i $\hat{\beta}$), s obzirom da obe metode daju identične ocene ovih parametara. Međutim, metodom najboljeg linearnog ocenjivanja dobijamo i ocene varijansi za $\hat{\alpha}$ i $\hat{\beta}$, tako da ih možemo izračunati koristeći podatke o cenama i količinama prodatih akumulatora datih u tabeli 1.

$$\mathrm{Var}(\hat{\alpha}) = \sigma^2 \left(\frac{1}{10} + \frac{1050^2}{685000}\right)$$

$$\mathrm{Var}(\hat{\alpha}) = 1.71\,\sigma^2$$

$$\mathrm{Var}(\hat{\beta}) = \sigma^2 \frac{1}{685000}$$

$$\mathrm{Var}(\hat{\beta}) = 1{,}46 \cdot 10^{-6}\,\sigma^2$$

Smanjenje varijanse ocena za α i β, tj. poboljšanje njihove efikasnosti može se postići optimalnim izborom vrednosti varijable X, što je u praksi često nemoguće postići.

Ocenjivanje metodom maksimalne verodostojnosti

Kod maksimalno verodostojnog ocenjivanja (MLE) uzimaju se one vrednosti parametara koje bi najčešće generisale dati uzorak.

Ocene parametara α i β dobijene metodom maksimalno verodostojnog ocenjivanja su, takođe, identične odgovarajućim ocenama dobijenim metodom najmanjih kvadrata. ..

Metodom maksimalno verodostojnog ocenjivanja dobija se ocena varijanse odstupanja ε_i (tj. ocena σ^2):

$$\hat{\sigma}^2 = \frac{1}{n}\sum(Y_i - \hat{\alpha} - \hat{\beta}X_i)^2$$

S obzirom da se pomoću sve tri navedene metode dobijaju iste ocene za α i β ($\hat{\alpha}$ i $\hat{\beta}$), a kod metode BLUE se zahteva nepristrasnost, zaključujemo da su očekivanja ovih ocena jednaka stvarnim vrednostima α i β. Takođe, jasno je da $\hat{\alpha}$ i $\hat{\beta}$, kao linearne kombinacije nezavisnih normalnih varijabli i same imaju normalnu raspodelu.

Dakle:

$$\hat{\alpha} \sim \mathrm{N}\left[\alpha, \sigma^2\left(\frac{1}{n} + \frac{\overline{X}^2}{\Sigma(X_i - \overline{X})^2}\right)\right]$$

$$\hat{\beta} \sim \mathrm{N}\left[\beta, \frac{\sigma^2}{\Sigma(X_i - \overline{X})^2}\right]$$

Koeficijent determinacije

Koeficijent determinacije (R^2) se računa prema sledećoj formuli:

$$R^2 = \frac{\Sigma(Y_i - \overline{Y})(\hat{Y}_i - \overline{Y})^2}{\Sigma(Y_i - \overline{Y})^2 \Sigma(\hat{Y}_i - \overline{Y})^2}$$

Ovaj koeficijent se uzima kao mera koliko se dobro regresijska linija uzorka prilagođava opažanjima. Jasno je da $R^2 \in [0,1]$, tako da koeficijent determinacije zapravo odražava procentualni udeo varijacije varijable Y koja se može pripisati varijabli X. Drugim rečima, R^2 je indikator učešća objašnjenog variranja Y u ukupnom.

U slučaju dobijanja niske vrednosti za koeficijent determinacije zaključujemo da se regresijska linija loše prilagođava opažanjima. Uzrok tome može biti pogrešna specifikacija regresijske jednačine, prevelik uticaj slučajnog odstupanja ili odsustvo funkcionalne zavisnosti varijabli X i Y.

Predviđanje

U slučaju ovde izloženog linearnog regresijskog modela, predviđa se vrednost zavisne varijable Y_0 za datu vrednost nezavisne varijable X_0.

Kada bi parametri populacije bili poznati, vrednost Y_0 bi se predviđala njenim očekivanjem:

$$E(Y_0) = \alpha + \beta X_0$$

Međutim, vrednost $E(Y_0)$ je nepoznata i može se oceniti sa $\hat{\alpha} + \hat{\beta} X_0$.

Razlika između stvarne i predviđene vrednosti zavisne varijable Y_0 naziva se greškom predviđanja:

$$Y_0 - \hat{Y}_0 = (\alpha + \beta X_i + \varepsilon_i) - (\hat{\alpha} + \hat{\beta} X_0)$$

Predviđanja unutar raspona u kome smo varirali nezavisnu varijablu, tj. predviđanja vrednosti Y koje odgovaraju vrednostima X blizu $\overline{X}$ su pouzdanija nego predviđanja vrednosti Y koje odgovaraju vrednostima X udaljenijim od $\overline{X}$. Drugim rečima, što više rizikujemo sa predviđanjima napuštajući raspon našeg iskustva, prognoza je manje realna.

Zaključak

Rezultati dobijeni ekonometrijskim istraživanjima mogu korisno poslužiti za testiranje važećih zakonitosti ekonomske teorije, ali i za predviđanje budućih vrednosti ekonomskih varijabli što može biti od ogromnog značaja pri donošenju važnih odluka iz oblasti ekonomske politike.

Literatura

[1] Greenberg E., Webster C.E.,Jr., (1983) *Advanced Econometrics*, New York, Wiley

[2] Jednak J., Tomić R., (2007) *Osnovi ekonomije*, Novi Sad, Alfa-graf

[3] Kiš T., Čileg M., Vugdelija D., Sedlak O., (2005) *Kvantitativni metodi u ekonomiji*, Subotica, Ekonomski fakultet Subotica

[4] Stojaković M., (2003) *Matematička statistika*, Beograd, Vedes

[5] www.sr.wikipedia.org