Contents

Chapter 1

Uvod

1.1 Šta je to ekonometrija?

Arthur S. Goldberger, Econometric Theory, Wiley, New York, 1964:
Ekonometrija može da se definiše kao društvena nauka u kojoj se alat ekonomske teorije, matemetike i statističkog zaključivanja primenjuje za analizu ekonomskih fenomena.

Kao rezultat odredenog pogleda na ulogu ekonomije, ekonometrija primenjuje matematičku statistiku u obradi ekonomskih podataka, sa ciljem da se omogući empirijska potvrda modela koji su konstruisani primenom ekonomske matematike, te u cilju dobijanja numeričkih rezultata.

1.2 Ekonometrija i ekonometrijski modeli

Prvi zadatak sa kojim se ekonometrista suočava, jeste formulisanje ekonometrijskog modela. Zato je važno definisati pojam modela. Model je pojednostavljena predstava nekog procesa iz realnog sveta. Na primer, tvrdnju da potražnja pomorandži zavisi od cene pomorandži, smatramo pojednostavljenom. Pre svega zato, jer se tu mora uzeti u obzir i niz drugih faktora, kao što su kupovna moc potrošaca, porast svesti o važnosti zdrave ishrane, fluktuacije u ceni goriva, itd. Sa druge strane, ovom nizu uticajnih faktora teško je sagledati kraj, pa otud vlada večiti sukob izmedu zagovornika kompleksnih i pojednostavljenih modela. U tom smislu, moguće je razlikovati dva osnovna prilaza formiranju modela.

Prvi prilaz zagovara primenu jednostavnijeg modela na početku, pri čemu se postepenim rafiniranjem i usavršavanjem početnog modela, dolazi do složenijeg modela. Drugi prilaz, kog u šali možemo nazvati i ,,slonovskim"

, za polazište uzima jedan generalizovani model, koji, na osnovu prikupljenih podataka treba postepeno smanjivati, tj. svoditi na specificnu oblast primene. Bez obzira na svoju složenost, model nikada ne može ostvariti idealnu realističnost u opisu posmatrane pojave, pa je tako daleko važnije u kojoj meri daje dobre aproksimacije. Stoga je, prilikom kreiranja modela koji opisuje neku pojavu, neophodno uključiti sve one promenljive za koje smatramo da su značajne, dok one preostale promenljive svrstavamo u grupu ,,poremećaja".

Ekonomski ili ekonometrijski model?

Pomenuta izbirljivost, tj restriktivnost, prilikom izbora skupa promenljivih koje ce biti uzete u obzir kao uticajni faktora prilikom formiranja modela, dovode nas do važnog pitanja o razlici izmedu ekonomskog i ekonometrijskog modela. Naime, ekonomski model čini skup pretpostavki koje, sa manjom ili većom preciznošcu, opisuju ponašanje ekonomskog sistema ili jednog njegovog segmenta. Sa druge strane, ekonometrijski model cine jednačine koje opisuju ponašanje sistema, a koje su nastale na osnovu ekonomskog modela. U ovim jednacinama figurišu:

- promenljive koje reprezentuju značajne faktore uticaja i

- ,,poremecaji", u koje se svrstavaju sve ,,manje bitne" promenljive kao i nepredvidene okolnosti.

Pored toga, ekonometrijski model obuhvata i raspodelu verovatnoće pojave takvih ,,poremećaja" i grešaka merenja koje su nastale prilikom posmatranja značajnih promenljivih.

Primer: Najjednostavniji slučaj ekonometrijskog modela
Posmatrano na primeru najjednostavnijeg modela tražnje, ekonometrijski model bi, prema gore navedenom, činile:

- jednacina koja opisuje ponašanje, oblika: $q = a + p + u$ pri čemu je:
  q - količina tražene robe
  p - cena robe
  u - faktor poremećaja

- raspodela verovatnoće pojave poremećaja, pri čemu važi $E(u/p) = 0$, gde u ima nezavisnu i normalnu raspodelu, sa srednjom vrednosti 0 i varijansom σ^2.

Ovako formiran ekonometrijski model, pruža dobru polaznu osnovu za proveru valjanosti ekonomskog modela, kao i za različita predvidjanja i analizu ekonomske politike.

1.3 Ekonometrijska metodologija

U svom uprošcenom obliku, ekonometrijska metodologija može se svesti na tri kljucna koraka:

1. Formulisanje ekonometrijskog modela u formi ekonomskog modela koji je proverljiv u praksi. S obzirom na različite mogućnosti izbora funkcionalne forme budućeg ekonometrijskog modela, kao i specifikacije stohastičke strukture promenljivih, na osnovu jednog ekonomskog zakona, uvek postoji nekoliko mogućih formulacija ekonometrijskog zakona. Taj korak nazivamo fazom specifikacije.

2. Testiranje modela uz pomoć prikupljenih podataka. Ovaj korak nazivamo fazom zaključivanja.

3. Primenu kreiranog modela u predvidjanju i kreiranju ekonomske politike.

Zbog, često vrlo složenih, matematičkih proračuna koje nije moguće sprovoditi bez primene računara, ekonometrijska praksa je prošla kroz dve karakteristične faze u svom razvoju.

U prvoj fazi, u rasponu od kasnih četrdesetih, pa sve do kraja šezdesetih godina dvadesetog veka, akcenat je bio na fazi zaključivanja, pri čemu su ekonometristi glavninu vremena posvećivali statističkim proverama modela i iznalaženju novih algoritama za proračun, dok su greške merenja ostajale u drugom planu. Šematski prikaz karakterističnih koraka u kreiranju i primeni ekonometrijskih modela iz ovog perioda, dat je na slici 1.

Slika 1. Karakteristicni koraci ekonometrijske analize iz prve faze u razvoju ekonometrije

Ubrzani razvoj računarskog hardvera početkom sedamdesetih, značajno je olakšao i ubrzao realizaciju koraka iz domena zaključivanja, čime je akcenat pomeren na fazu specifikacije. Na taj način su greške u specifikaciji, kao i greške u merenju, koje prate korak odabiranja promenljivih, dospele u žižu interesovanja. Ova faza šematski je prikazana na slici 2.

Slika 2. Karakteristični koraci ekonometrijske analize iz druge faze razvoja ekonometrije

Iako slike 1 i 2 predstavljaju uprošcene metodološke sheme koje je ekonometrija primenjivala u karakterističnim fazama razvoja, na osnovu njih je moguće konstatovati nekoliko značajnih poboljšanja koje je donela druga faza razvoja.

Ukoliko zanemarimo promene koje su nastale eliminacijom, odnosno, sažimanjem nekih koraka, lako je uočiti postojanje povratnih veza (slika 2) koje sada omogućuju sledeće:

1. Primenu rezultata dobijenih ekonometrijskom analizom, u cilju usavršavanja polazne ekonomske teorije (slika 2, linija A).
2. Primenu rezultata dobijenih u fazi testiranja i dijagnosticiranja ponašanja ekonometrijskog modela, radi modifikacije i poboljšavanja ekonometrijskog modela (slika 2, linija B).
3. Rafiniranje ulaznih podataka koji se prikupljaju ,,na terenu", na osnovu uočenog ponašanja ekonometrijskog modela (slika 2, linija C).

1.4 Kako testirati ekonomsku teoriju?

Kako je vec napomenuto, osnovna namena ekonometrijskih modela jeste testiranje ekonomskih teorija. Postavlja se, medjutim, pitanje, šta usvojiti kao pokazatelj valjanosti jedne ekonomske teorije. S tim u vezi postoje dva prilaza:

- prilaz koji je usmeren ka potvrdi ekonomske teorije i
- prilaz koji se oslanja na tačnost predvidanja ekonomske teorije

Kada je u pitanju prvi prilaz, ekonomska teorija je potvrdena ukoliko su tačni predznaci koeficijenata koji figurišu u ekonometrijskom modelu. Ekonomska praksa, medjutim, ukazuje na urodjenu slabost ovakvog prilaza testiranju. Naime, neretko se dogada da dve ekonomske teorije, čija je valjanost potvrdjena ovim postupkom testiranja, daju kontradiktorne rezultate, pri čemu nije moguće utvrditi koja od njih daje ispravne zaključke.

Poslednjih godina, veće poverenje dobija prilaz koji za kriterijum testiranja usvaja tačnost predvidjanja ekonomske teorije. Na osnovu ovog prilaza, rezultati testiranja slede na osnovu poredjenja ekonometrijskog modela sa njegovim aktuelnim konkurentima ili prethodnicima. Pri tom je najbolji onaj model koji daje najbolje prognosticke rezultate.

Chapter 2

Ocenjivanje

Ekonometričare više zaokuplja problem ocenjivanja od problema testiranja hipoteze. Teorija ocenjivanja za razne vrste ekonomskih modela dosta se dobro razvila, premda, mnoge poteškoće ostaju.

Tačkasta ocena - koristi apriori informacije (npr. o raspodeli) i informacije iz uzorka $(x_1, x_2, \ldots, x_n)$ iz uzorka radi izračunavanja vrednosti koje će, na neki način, biti naš najbolji pokušaj u pogadjanju prave vrednosti parametra koji nas zanima.

Intervalna ocena - koriste iste informacije za odredjivanje intervala.

Potrebno je:

I odrediti šta znači "najbolji pokušaj"

II konstruisati ocenjivače koji će zadovoljiti te kriterijume ili barem biti blizu

Problem I svodi se na odredjivanje različitih svojstava ocene koja se mogu smatrati poželjnim. Kako je ocenjivač slučajna promenljiva čija se vrednost razlikuje od uzorka do uzorka, njegova svojstva su, zapravo, svojstva njegove distribucije iz uzorka

$$\hat{\theta} = \hat{\theta}(x_1, x_2, \ldots, x_n)$$

Osnovna svojstva ocenjivača $\hat{\theta}$ su

$$E(\hat{\theta}) \quad \text{i} \quad Var(\hat{\theta}) = E(\hat{\theta}^2 - E(\hat{\theta}))^2 = E(\hat{\theta}^2) - E(\hat{\theta})^2.$$

Standardna devijacija ili odstupanje $\sqrt{Var(\hat{\theta})}$ poznata je kao standardna greška od $\hat{\theta}$.

$\hat{\theta} - \theta$	greška ocenjivača
$E(\hat{\theta}) - \theta$	pristrasnost ocene
$MSE = E((\hat{\theta} - \theta)^2)$	srednja kvadratna greška

Srednja kvadratna greška razlikuje se od varijanse jer varijansa meri disperziju distribucije ocene oko sredine $E(\hat{\theta})$, a MSE meri rasturanje oko prave vrednosti parametra θ.

$$
\begin{aligned}
MSE(\hat{\theta}) &= (\hat{\theta} - \theta)^2 = E(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta)^2 \\
&= (E(\hat{\theta} - E(\hat{\theta}))^2 + E(E(\hat{\theta}) - \theta)^2 + 2E((\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta)) \\
& E(\hat{\theta} - E(\hat{\theta}))^2 + E(E(\hat{\theta}) - \theta)^2 = var(\hat{\theta}) + kvadratpristrasnosti
\end{aligned}
$$

jer je

$$E((\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta)) = E(\hat{\theta}E(\hat{\theta}) - E(\hat{\theta})^2 + \theta E(\hat{\theta}) - \hat{\theta}\theta)$$

$$= E(\hat{\theta})^2 - E(\hat{\theta})^2 + E(\theta)E(\hat{\theta}) - E(\hat{\theta})E(\theta) = 0$$

Svojstva ili osobine ocenjivača mogu se podeliti u dve velike grupe zavisno od veličine uzorka.

I
- svojstva ocenjivača kod konačnih uzoraka (odnose se na svojstva distribucija ocenjivača zasnovanog na fiksnoj veličini uzorka)
- ocene se računaju na osnovu bilo kojeg broja opažanja
- često se zovu i svojstva ocena kod malih uzoraka jer se mogu primeniti čak i kad su uzorci mali

II
- asimptotska svojstva ili svojstva ocenjivača kod velikih uzoraka ograničavaju se na distribucije iz uzorka čija veličina teži beskonačno
- ova svojstva važe samo približno kad je veličina uzorka velika, a možda uopšte ne važe kad su uzorci mali

2.1 Svojstva ocenjivača kod malih uzoraka

- Ocenjivač $\hat{\theta}$ je **nepristrasan** za θ ako važi $E(\hat{\theta}) = \theta$. Problem je kada imamo nepristrasnu ocenu sa velikom varijansom. Kada imamo malu varijansu, ali pristrasnu ocenu a stepen pristrasnosti nije poznat ocenivač je veoma loš.
- Ocenjivač $\hat{\theta}$ je **efikasan** za θ ako važi
  1. $\hat{\theta}$ je nepristrasan

2. $Var(\hat{\theta}) \leq Var(\tilde{\theta})$, gde je $\tilde{\theta}$ bilo koja druga ocena od θ

U statističkoj literaturi ne postoji opšte slaganje o definiciji efikasnosti. U praksi se koristi relativna efikasnost. Primena Rao-Kramerove nejednakosti ima prednosti (primer: sredina kod normalne raspodle) i mane (primer: varijansa normalne raspodele).

- Specijalno, $\hat{\theta}$ je **najbolji linearni nepristrasan ocenjivač (BLUE)** od θ ako važi:

  1. $\hat{\theta}$ je linearna funkcija opažanja iz uzorka $(x_1, x_2, \ldots, x_n)$
  2. $\hat{\theta}$ je nepristrasan
  3. $Var(\hat{\theta}) \leq Var(\tilde{\theta})$, $\tilde{\theta}$ je bilo koji drugi linearni nepristrasan ocenjivač

Slično se može definisati i najbolji kvadratni nepristrasni ocenjivač (BQU). Na kraju, primetimo da se svojsta ocenjivača mogu ispitivati samo ako sredina i varijansa ocenjivača postoje.

2.2 Asimptotska svojstva

U opštem slučaju, distribucija (raspodela) nekog ocenjivača na osnovu jedne veliicne uzorka razlikuje se od raspodele istog ocenjivača na osnovu druge veličine uzorka. Naprimer, normalna raspodela je asimptotska (granična) raspodela sredine uzorka (na osnovu centralne granične teoreme)).

Pod asimptotskom raspodelom ne smatra se konačni oblik raspodele, koji može biti degenerisan, već oblik koji raspodela poprima u poslednjem delu svog puta prema padu (ako se to dogadja).

Postojanje i oblik asimptotske raspodele:

- neki ocenjivači imaju raspodelu koja je istog oblika bez obzira na veličinu uzorka, a taj je oblik poznat (primer: sredina uzorka kao ocenjivača sredine normalne populacije, $\bar{X} : \mathcal{N}(\mu, \frac{\sigma^2}{n})$

- neki ocenjivači imaju raspodelu koja je poznata za svaku veličinu uzorka, ali nije uvek istog oblika, asimptotska raspodela tih ocenjivača je raspodela dobijena za $n \to \infty$ (primer: raspodela proporcije uspeha u uzorku je binomna, ali za $n \to \infty$ konvergira ka normalnoj)

- za neke ocenjivače raspodela ne mora nužno biti poznata za svaku veličinu uzorka, ali je poznata kada $n \to \infty$ (primer: sredina uzorka kao ocenjivač sredine ne-normalne raspodele)

Ove tri vrste ocenjivača pokrivaju veći deo problema ocenjivanja koji se javlja u ekonomiji.

Asimptotske raspodele mogu se opisati pomoću njihovih momenata, od kojih su najvažniji asimptotska sredina i asimptotska varijansa.

Pod pretpostavkom da $E(\hat{\theta})$ postoji asimptoska sredina je $\lim_{n\to\infty} E(\hat{\theta})$. Slično je asimptotska varijansa data sa $\lim_{n\to\infty} Var(\hat{\theta})$. Problem nastaje kada važi $\lim_{n\to\infty} Var(\hat{\theta}) = 0$. Pravilno bi bilo reći varijansa asimptotske raspodele,

$$asimptot.Var(\hat{\theta}) = \frac{1}{n} \lim_{n\to\infty} E(\sqrt{n}(\hat{\theta} - \theta))^2.$$

Svojstva:

- $\hat{\theta}$ je asimptotske nepristrasan ocenjivač od θ ako je $\lim_{n\to\infty} E(\hat{\theta}) = \theta$. Lako se pokazuje da ako je ocenjivačnepristrasan da je tada i asimptotske nepristrasan. Obrnuto ne mora da važi, kao što je u slučaju $\hat{\sigma}^2 = \frac{1}{n}\sum(x_i - \bar{x})^2$.
- Uvedimo osnaku

  $$plim\ \hat{\theta} = \theta^* \Leftrightarrow \lim_{n\to\infty} P(|\hat{\theta} - \theta^*| \leq \varepsilon) = 1 \quad \text{slučaj degenerisane raspodele,}$$

  tj. $\hat{\theta}$ konvergira u verovatnoći ka θ^*. Ocenjivač se smatra **konzistentnim** ako posrne (padne) u jednu tačku prave vrednosti parametra, tj

  $$\hat{\theta} \quad \text{je konzistentan ocenjivač od } \theta \text{ ako važi} \quad plim\hat{\theta} = 0.$$

  Praćenje konzistentnosti: ako porast veličine uzorka prati smanjenje pristrasnosti (ako je ima) i varijanse, i ako se približava 0, kad $n \to \infty$, tada je taj ocenjiva č konzistentan, pa važi

  $$\text{Ako je } \hat{\theta} \text{ ocenjivač od } \theta \text{ i ako je } \lim_{n\to\infty} MSE(\hat{\theta}) = 0 \text{ tada je } \hat{\theta} \text{ konzistentan.} \quad (*)$$

  Obrnuto ne mora da važi, npr. konzistentan ocenjivač čija sredina ili varijansa ne postoje.
  Ako postoje konačna asimptotska sredina i varijansa tada je (*) i dovoljan uslov.
  Ocenjivačkoji je konzistentan je asimptotski nepristrasan.
  Bilo kakva neprekidna funkcija konzistentnog ocenjivača i sama je konzistentan ocenjivač. Ova osobina poznata je pod imenom "prenos konzistentnosti", dok se naprimer nepristrasnost ne prenosi.
  Konzistentnost kvadrata greške znači da je (*) poterban i dovoljan uslov.

- Ocenjvač $\hat{\theta}$ je **asimptotski efikasan** ako su zadovoljeni sledeći uslovi:
  1. $\hat{\theta}$ ima asimptotsku distribuciju sa konačnom sredinom i varijansom
  2. $\hat{\theta}$ je konzistentan
  3. Nijedan konzistentan ocenjivač od θ nema manju asimptotsku varijansu od $\hat{\theta}$

  Iz efiksanosti uvek sledi asimptotska efikasnost.

Tabelarni prikaz svojstava ocenjivača:

svojstva kod konačnih (malih) uzoraka	asimptotska svojstva (kod velikih uzoraka)
nepristrasnost	asimptotska nepristrasnost
efikasnost	konzistentnost
BLUE	asimptotska efikasnost

2.3 Metode ocenjivanja

- **METOD MOMENATA:** moment raspodele populacije ocenjujemo odgovarajućim momentom uzorka; ovako dobije ocenjivači su konzistentni i imaju asimptotsku normalnost.
- **MOETOD NAJMANJIH KVADRATA:** (LSE,OLE) U ovom slučaju svojstva se moraju posebno ispitivati. Veliki deo moderne ekonometrije duguje svoje postojanje otkriću da su u mnogim ekonometrijskim modelima ocenjivači dobijeni ovom metodom nekonzistentni.
- METOD MAKSIMALNE VERODOSTOJNOSTI: (MLE) Ovaj metod temelji se na relativno jednostavnoj zamisli da različite populacije daju različite uzorke, te da za bilo koji uzorak vredi da je je verovatnije da potiče iz nekih populacija nego iz drugih. Ako slučajna promenljiva X ima raspodelu verovatnoće ili gustine $f(x)$ koja sadrži parametre $\theta_1, \theta_2, \ldots, \theta_k$ i ako imamo uzorak $(x_1, x_2, \ldots, x_n)$, tada su ocenjivači maksimalne verodostojnosti one vrednosti parametra θ_i koja će najčešće generisati dati uzorak. Funkcija verodostojnosti je

  $$L = f(x_1)f(x_2)\cdots f(x_n)$$

  Pri vrlo opštim uslovima ocenjivači maksimalne verodostojnosti poseduju sledeća svojstva:

  1. konzistentnost kvadrata greške

2. asimptotska efikasnost
3. asimptotska distribucija je normalna
4. formula za odredjivanje asimptotskih varijansi je dostupna

Informacijska matrica:

$$I_k = \begin{bmatrix} \frac{\partial^2 L}{\partial \theta_1^2} & \frac{\partial^2 L}{\partial \theta_1 \theta_2} & \cdots & \frac{\partial^2 L}{\partial \theta_1 \theta_k} \\ \frac{\partial^2 L}{\partial \theta_2 \theta_1} & \frac{\partial^2 L}{\partial \theta_2^2} & \cdots & \frac{\partial^2 L}{\partial \theta_2 \theta_k} \\ & \cdots & \cdots & \\ \frac{\partial^2 L}{\partial \theta_k \theta_1} & \frac{\partial^2 L}{\partial \theta_k \theta_2} & \cdots & \frac{\partial^2 L}{\partial \theta_k^2} \end{bmatrix}$$

Kod konačnih uzoraka za ocenu asimptotskih varijansi koristimo dijagonalne elemente I_k^{-1}.
Iako su sva svojstav 1-4 samo asimptotska, u mnogim slučajevima to je često sve čemu se možemo nadati.

- **METOD NAJBOLJEG NEPRISTRASNOG OCENJIVAČA:** (BLU)

  1. $\hat{\theta} = \sum a_i X_i$
  2. $E(\hat{\theta}) = \theta \Rightarrow \sum a_i = 1$
  3. $Var(\hat{\theta}) = Var(\sum a_i X_i) = \sum a_i^2 Var(X_i) = \sigma^2 \sum a_i^2$

  Analizom prethodnih koraka uvidjamo da smo odredjivali min $\sum a_i^2$ uz uslov $\sum a_i = 1$ metodom Lagranžovih multiplikatora

Sva četiri navedena metoda daju isti ocenjivač za sredinu uzorka koji ima sva optimalna svojstva.

Chapter 3

Jedostruka ili prosta regresija

Jedna od najčešće korišćenih analiza u ekonometriji je regresiona analiza.

Regresiona analiza se bavi opisom i ocenom izmedju date zavisne promenljive (y) i jedne ili više nezavisnih promenljivih $(x_1, x_2, \ldots, x_n)$.

Primer 1:

y	-prodaja
x	-ulaganje u reklamu

Primer 2:

y	-potrošnja porodice na dobra i usluge
x_1	-porodični budžet
x_2	-porodične hartije od vrednosti
x_3	-veličina porodice

Različiti nazivi obeležja u regresionoj analizi

	y	$x_1, x_2, \ldots, x_n$
(a)	predictand	predictors
(b)	regressand	regressors
(c)	explained variable	explanatory variables
(d)	dependent variable	independent variables
(e)	effect variable	causal variables
(f)	endogenous variable	exogenous variables
(g)	target varible	control variable

Postoje dve vrste relacija izmedju promenljivih:

1. **deterministička** ili matematička veza koja daje jedinstvenu vrednost za y $y = 2500 + 100x - x^2$

2. **stohastička** ili statistička veza koja ne daje jedinstvenu vrednost za y za dato x:

$$y = 2500 + 100x - x^2, \qquad u = \begin{cases} +500 & \text{sa verovatnoćom } 0.5 \\ -500 & \text{sa verovatnoćom } 0.5 \end{cases}$$

Ako je $u : \mathcal{N}(0,1)$ tada za svako x imamo normalnu raspodelu za y .
Primer: UBACITI SLIKU

3.1 Jednačina linearne regresije

$$y = \alpha + \beta x + u$$

α, β	- regresioni koeficijenti
$\alpha + \beta x$	- determinističi deo
u	- greška (stohastički deo)

Zašto se uvodi greška:

1. nepredvidljivost ljudskog ponašanja

2. efekat velikog broja izostavljenih promenljivih

3. greška merenja

3.2 Standardne pretpostavke

1. **normalnost:** u_i ima normalnu raspodelu za svako i

2. **sredina je nula:** $E(u_i) = 0$ za svako i

3. **homoskedastičnost:** $Var(u_i) = \sigma^2$ za svako i

4. **odsustvo autokorelacije:** $Cov(u_i, u_j) = 0, \quad i \neq j$

5. $Cov(u_i, x_i) = 0$, za svako i

6. **nestohastičnost obeležja** X**:** $\frac{1}{n}\sum(X_i - \bar{X})^2 \neq 0$ i konačna za $n \to \infty$

Kako je $E(u_i) = 0$, to je

$$E(y_i) = \alpha + \beta x_i. \tag{3.1}$$

Relacija 3.1 naziva se **regresiona funkcija populacije**.

Odredjivanjem ocena za α i β dobijamo **regresionu funkciju uzorka**

$$E(y_i) = \hat{\alpha} + \hat{\beta} x_i. \tag{3.2}$$

Parametre α i β možemo oceniti jednom od predloženih metoda.

3.3 Metod momenata

$$\begin{array}{lcl} E(u) = 0 & \Leftrightarrow & \frac{1}{n}\hat{u}_i = 0 \Leftrightarrow \sum(y_i - \hat{\alpha} - \hat{\beta}x_i) = 0 \\ cov(x,u) & \Leftrightarrow & \frac{1}{n}\sum x_i\hat{u}_i = 0 \Leftrightarrow \sum x_i(y_i - \hat{\alpha} - \hat{\beta}x_i) = 0 \end{array}$$

3.4 Metod najmanjih kvadrata

$$\min_{\alpha,\beta} Q = \min_{\alpha,\beta}(\sum(y_i - \hat{\alpha} - \hat{\beta}x_i)^2)$$

$$\begin{array}{l} \bar{y} = \hat{\alpha} + \hat{\beta}\bar{x} \\ \sum y_i x_i = \\ ha\sum x_i + \hat{\beta}\sum x_i^2 \end{array} \tag{3.3}$$

Uvodeći oznake

$$\begin{array}{l} S_{yy} = \sum(y_i - \bar{y})^2 = \sum y_i^2 - n\bar{y}^2 \\ S_{xx} = \sum(x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2 \\ S_{xy} = \sum(x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - n\bar{x}\bar{y} \end{array}$$

dobijamo

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} \quad \text{i} \quad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}. \tag{3.4}$$

Sada možemo izraziti kvadratnu grešku (RSS =rezidual sum of square)

$$\begin{aligned} RSS &= \sum(y_i - \hat{y}_i) = \sum(y_i - \hat{\alpha} - \hat{\beta}x_i)^2 \\ &= \sum(y_i - \bar{y} + \hat{\beta}\bar{x} - \hat{\beta}x_i)^2 \\ &= \sum(y_i - \bar{y} - \hat{\beta}\bar{x} - \hat{\beta}x_i)^2 \\ &= \sum(y_i - \bar{y})^2 + \hat{\beta}^2\sum(x_i - \bar{x})^2 - 2\hat{\beta}\sum(y_i - \bar{y})(x_i - \bar{x}) \\ &= S_{yy} + \frac{S_{xy}^2}{S_{xx}^2}S_{xx} - 2\frac{S_{xy}}{S_{xx}}S_{xy} \\ &= S_{yy} - \frac{S_{xy}^2}{S_{yy}}. \end{aligned}$$

Uvodeći oznake $TSS = S_{yy}$ (total sum of squares) i $ESS = \dfrac{S_{xy}^2}{S_{yy}}$ (explained sum squared), dobijamo

$$TSS = RSS + ESS.$$

Ubaciti sliku

Koeficijent determinacije je

$$r_{xy}^2 = \frac{ESS}{TSS} = \frac{TSS - RSS}{TSS} = \frac{\hat{\beta} S_{xy}}{S_{yy}}.$$

Ako je r_{xy}^2 blizu 1 tada x dobro "objašnjava" y.

Napomena: Obrnuta regresija koristi se za ocenu β kada pri merenju i x i y pravimo grešku. U ekonomiji često nije jasno koji smer regresije treba izabrati.

3.5 Statistički zaključci u linearnom regresijonom modelu

Prilikom odredjivanja ocena za α i β nije bilo potrebe za uslovom na raspodelu za grešku u_i. Medjutim, za dobijanje intervala za ocenu parametra, kao i za testove u vezi sa njima neophodno je pretpostaviti neke uslove za u_i.

Neka važe pretpostavke:

1. $E(u_i) = 0$
2. $var(u_i) = \sigma^2$
3. u_i i u_j su nezavisne za $i \neq j$
4. x_i su nestohastičke promenljive

Pod ovim pretpostavkama ocene $\hat{\alpha}$ i $\hat{\beta}$ dobijene metodom najmanjih kvadrata su

1. nepristrasne
2. BLUE tj. imaju najmanju varijansu medju svim centriranim linearnim ocenama

Da bismo odredjivali intervale poverenja za nepoznate parametre neophodno je uvesti i pretpostavku

5. u_i imaju normalnu raspodelu za svako i

3.6 Intervali poverenja za α i β

Pretpostavimo da su ispenjene pretpostavke 1-5. Tada posmatramo model

$$y_i = \alpha + \beta x_i + u_i, \qquad u_i : \mathcal{N}(0, \sigma^2$$

i ocene

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}}, \qquad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}.$$

Koristimo tvrdjenje: *Ako su $y_1, y_2, \ldots, y_n$ nezavisne sa normalnom raspodelom $\mathcal{N}(0, \sigma^2)$ i ako su $L_1 = \sum c_i y_i$ i $L_2 = \sum d_i y_i$ dve linearne funkcije od y_i, tada L_1 i L_2 imaju zajedničku normalni raspodelu i važi*

$$var(L_1) = \sigma^2 \sum c_i^2 \quad i \quad L_2 = \sum d_i y_i$$

i

$$cov(L_1, L_2) = E(L_1 L_2) - E(L_1)E(L_2) = \ldots = \sigma^2 \sum c_i d_i.$$

Sada raspisujemo ocene $\hat{\beta}$ i $\hat{\alpha}$ kao funkcije od $y_1, \ldots, y_n$. Prvo primetimo,

$$S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum (x_i - \bar{x}) y_i - \bar{y} \sum (x_i - \bar{x}) = \sum (x_i - \bar{x}) y_i.$$

Sada dobijamo

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2} = \sum c_i y_i,$$

gde smo uveli oznaku $c_i = \dfrac{x_i - \bar{x}}{S_{xx}}$.

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = \frac{1}{n} \sum y_i - \bar{x} \sum c_i y_i = \sum (\frac{1}{n} - \bar{x} c_i) y_i = \sum d_i y_i,$$

gde je $d_i = \dfrac{1}{n} - \bar{x} \dfrac{x_i - \bar{x}}{S_{xx}}$. Sada izračunavamo varijanse za $\hat{\alpha}$ i $\hat{\beta}$,

$$\begin{aligned} var(\hat{\beta}) &= var(\textstyle\sum c_i y_i) = \sum c_i^2 var(y_i) \\ &= \sigma^2 \sum c_i^2 = \sigma^2 \frac{\sum (x_i - \bar{x})}{S_{xx}^2} \\ &= \frac{\sigma^2}{\sum (x_i - \bar{x})^2} \end{aligned}$$

i

$$\begin{aligned}
var(\hat{\alpha}) &= \sigma^2 \sum d_i^2 \\
&= \sigma^2 \sum (\frac{1}{n^2} - 2\frac{\bar{x}(x_i - \bar{x})}{S_{xx}n} + \bar{x}^2 \frac{(x_i - \bar{x})^2}{S_{xx}^2}) \\
&= \sigma^2 (\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}
\end{aligned}$$

Kovarijansa za $\hat{\alpha}$ i $\hat{\beta}$ je

$$\begin{aligned}
cov(\hat{\alpha}, \hat{\beta}) &= \sigma^2 \sum \frac{(x_i - \bar{x})}{S_{xx}} (\frac{1}{n} - \bar{x}\frac{x_i - \bar{x}}{S_{xx}}) \\
&= \sigma^2 \sum (\frac{1}{n}\frac{x_i - \bar{x}}{S_{xx}} - \sum \bar{x}\frac{(x_i - \bar{x})^2}{S_{xx}^2}) \\
&= \sigma^2 (-\frac{\bar{x}}{S_{xx}}).
\end{aligned}$$

Sada odredjuje očekivanje za $\hat{\alpha}$ i $\hat{\beta}$,

$$E(\hat{\beta}) = E(\sum c_i y_i) = \sum c_i E(y_i) = \sum c_i(\alpha + \beta x_i) = \beta,$$

$$E(\alpha) = E(\bar{y} - \hat{\beta}\bar{x}) = \alpha + \beta\bar{x} - \beta\bar{x} = \alpha.$$

Iz prethodne analize zaključujemo:

1. $\hat{\alpha}$ ima normalnu raspodelu $\mathcal{N}(\alpha, \sigma^2(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}))$
2. $\hat{\beta}$ ima normalnu raspodelu $\mathcal{N}(\beta, \frac{\sigma^2}{S_x x})$
3. $cov(\hat{\alpha}, \hat{\beta}) = \sigma^2(-\frac{\bar{x}}{S_{xx}})$

Primetimo da navedeni zaključci važe kada je σ^2 poznato. To, naravno, nije čest slučaj, pa je potrebno razmotriti slučaj kada ocenjujemo parametar σ^2. Ovde navodimo tvrdjenje bez izvodjenja, tj. dokaza. Ocena za σ^2 je statistika

$$\hat{\sigma^2} = \frac{RSS}{n-2},$$

čija je raspodela

$$\frac{RSS}{\sigma^2} : \chi^2_{n-2}.$$

Koristeći poznata tvrdjenja o χ^2 raspodeli možemo izvesti zaključke o raspodelama za ocene $\hat{\alpha}$ i $\hat{\beta}$ kada σ^2 nije poznata, tj. kada koristimo ocenu $\hat{\sigma^2}$. Ocena varijanse za $\hat{\beta}$ u ovom sluãju je $\dfrac{RSS}{(n-2)S_{xx}}$, a

$$\sqrt{\frac{\hat{\sigma^2}}{S_{xx}}}$$

nazivamo standardnom greškom i označavamo sa $SE(\hat{\beta})$.

Slično dobijamo i standardnu grešku $SE(\hat{\alpha})$, pa možemo zaključiti

$$\frac{\alpha-\hat{\alpha}}{SE(\hat{\alpha})} \text{ i } \frac{\hat{\beta}-\beta}{SE(\hat{\beta})} \text{ imaju } t \text{ raspodelu sa } (n-2) \text{ stepena slobode.}$$

Prethodno tvrdjenje koristimo za odredjivanje intervala poverenja za α i β. $\hat{\sigma}$ nazivamo standardna greška regresije (standard error of the regression) i označavamo sa SER (ponekad i SEE).

3.7 Analiza varijanse za jednostruku regresiju

Source of variation	sum of square	degrees of freedom	mean square
x	$ESS = \hat{\beta}S_{xy}$	1	$ESS/1$
Residual	$RSS = S_{yy} - \hat{\beta}S_{xy}$	$n-2$	$RSS/(n-2)$
Total	$TSS = S_{yy}$	$n-1$	

Tumačenje:

- po pretpostavci $\frac{RSS}{\sigma^2}$ ima χ^2_{n-2} raspodelu;
- $\frac{ESS}{\sigma^2}$ ima χ^2_1 raspodelu samo ako je $\beta = 0$;
- pod pretpostavkom $\beta = 0$ statistika $F = \dfrac{\frac{ESS}{\sigma^2}}{\frac{RSS}{\sigma^2(n-2)}}$ ima F raspodelu $F_{1,(n-2)}$, pa se F statistika može koristiti za testiranje hipoteze $\beta = 0$

Medjutim, analiza varijanse je od velikog značaja kod višestruke regresije.

Možemo izvesti formulu koja povezuje koeficijent determinacije i vrednost t statistike

$$\begin{aligned} ESS &= \hat{\beta} S_{xy} = \frac{S_{xy}}{S_{xx}} S_{xy} = \ldots = r^2 S_{yy} \\ RSS &= (1 - r^2) S_{yy} \\ F &= \frac{ESS}{RSS/(n-2)} = \frac{r^2}{(1-r^2)/(n-2)} \end{aligned}$$

Kako je $F = t^2$ to dobijamo

$$r^2 = \frac{t^2}{t^2 + (n-2)},$$

gde je t t-vrednost za testiranje hipoteze $H(\beta = 0)$.

3.8 Predvidjanje primenom proste regresije

Predvidjena vrednost je $\hat{y_0} = \hat{\alpha} + \hat{\beta} x_0$, a stvarna vrednost je $y_0 = \alpha + \beta x_0 + u_0$, pa je greška predvidjanja

$$\hat{y_0} - y_0 = (\hat{\alpha} - \alpha) + (\hat{\beta} - \beta) x_0 - u_0.$$

Kako je $E(\hat{\alpha} - \alpha) = 0$ i $E(\hat{\beta} - \beta) = 0$ i $E(u_0) = 0$ to imamo

$$E(\hat{y_0} - y_0) = 0,$$

što znači da je predikcija nepristrasna.

Odredjujemo varijansu greške

$$\begin{aligned} Var(\hat{y_0} - y_0) &= Var(\hat{\alpha} - \alpha) + x_0^2 var(\hat{\beta} - \beta) + 2x_0 cov(\hat{\alpha} - \alpha, \hat{\beta} - \beta) + var(u_0) \\ &= \sigma^2 (\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}) + x_0^2 \frac{\sigma^2}{S_{xx}} + 2x_0 \sigma^2 (-\frac{\bar{x}}{S_{xx}}) + \sigma^2 \\ &= \sigma^2 (1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}). \end{aligned}$$

Analizom prethodne relacije uočavamo da varijansa raste kako se x_0 udaljava od $\bar{x}$.

Slika

Predvidjanje očekivane vrednosti

Ponekad za dato x_0 ne želimo da predvidimo y_0, već $E(y_0)$. Kako je $E(y_0) = \alpha + \beta x_0$ možemo $E(y_0)$ predvideti koristeći $\hat{E}(y_0) = \hat{\alpha} + \hat{\beta} x_0$, što je isto kao i kod predvidjanja $\hat{y_0}$. Znači predvidjena vrednost za y_0 i $E(y_0)$ je ista, ali je zato varijansa (greška) drugačija. Greška predikcije sada je

$$\hat{E}(y_0) - E(y_0) = (\hat{\alpha} - \alpha) + (\hat{\beta} - \beta)x_0,$$

pa je varijansa

$$Var(\hat{E}(y_0) - E(y_0)) = \sigma^2(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}),$$

odnosno

$$SE(\hat{E}(y_0) = \hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}.$$

3.9 Bacanja

Često se dešava da ocena regresionih parametara zavisi od nekoliko ekstremnih observacija, odnosno podataka. Takve podatke zovemo outliers ili bacanja.

Ovaj problem se može detektovati analizom greške regresionog modela. Ukoliko se pojave bacanja mogu se dovesti u sumnju i neke od standardnih pretpostavke.

Bacanje ili outlier je podatak koji je "daleko" od preostalih podataka. Ovaj podatak je obično dobijen pod uticajem nekog neobičnog faktora. U jednostrukoj regresiji možemo nacrtati regresionu pravu i prepoznati bacanja. Medjutim, kod višestruke regresije takva analiza je nemoguća, pa ćemo u tom slučaju analizirati grešku $\hat{u_i}$. U nekim situacijama outliers možemo isključiti, ali je to u globalu pogrešno.

Chapter 4

Narušavanje standardnih pretpostavki

Ako su ispunjene standarde pretpostavke ocenjivači

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}, \qquad \hat{\beta} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})2} = \frac{S_{xy}}{S_{xx}}, \qquad \hat{\sigma^2} = \frac{\sum u_i^2}{n-2} = \frac{\sum(y_i - \hat{y}_i)^2}{n-2} = \frac{RSS}{n-2}$$

imaju sledeća svojstva

1. ocenjivači su nepristrasni
2. ocenjivači su efikasni
3. ocenjivači su BLUE
4. ocenjivači su konzistenti
5. ocenjivač $\hat{\alpha}$ ima normalnu raspodelu $\mathcal{N}(\alpha, \sigma_\alpha^2)$, gde je

$$\sigma_\alpha = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right) \qquad se(\hat{\alpha}) = \sqrt{\hat{\sigma^2}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}$$

6. ocenjivač $\hat{\beta}$ ima normalnu raspodelu $\mathcal{N}(\beta, \sigma_\beta^2)$, gde je

$$\sigma_\beta = \frac{\sigma^2}{S_{xx}} \qquad se(\hat{\beta}) = \sqrt{\frac{\hat{\sigma^2}^2}{S_{xx}}}$$

7. $\operatorname{cov}(\hat{\alpha}, \hat{\beta}) = \sigma^2 \left(-\frac{\bar{x}}{S_{xx}}\right)$

8. $\frac{(n-2)\hat{\sigma^2}^2}{\sigma^2}$ ima χ^2 raspodelu sa $n-2$ stepena slobode

9. raspodela $(\hat{\alpha}, \hat{\beta})$ je nezavisna od $\hat{\sigma^2}^2$

4.1 Ne-normalnost

4.2 Šta se dešava?

- Pretpostavka o normalnosti korišćena je za dokaz efikasnosti ocenjivača, odnosno intervala poverenja i testiranja hipoteza.
- Nema ozbiljne posledice na ocenjivače, mada ekstremna odstupnja sa malom verovatnoćom imaju neopravdano jak uticaj.
- ocenjivač$\hat{\sigma^2}^2$ je i dalje nepristrasan i konzistentan;
- $\hat{\alpha}$ i $\hat{\beta}$ su asimptotski normalni i asimptotski vrede intervali poverenja i testovi značajnosti;

4.3 Kako je detektovati?

1. **grafički prikaz**

2. χ^2 **goodness of fit test** na osnovu uzorka odredjujemo klasični linearni model i izračunavamo rezidual $\hat{u}$ ($var(\hat{u}) = \frac{1}{n-1}\sum(\hat{u}_i - \bar{\hat{u}})^2 = \frac{1}{n-1}\sum \hat{u}_i^2$)

3. **uporedjivanja koeficijenata regresije**

   Uporedjujemo koeficijente dobijene primenom metoda najmanjih kvadrata i minimiziranjem sume apsolutnih ostupanja

4. **Jarque-Bera (JB) test of normality**

   Ocenjujemo *simetričnost* (skewness) i *sploštenost*(kurtosis); kod normalne raspodele je $S = 0$ i $K = 3$

$$\sqrt{S} = \frac{\nu_3}{\nu_2^{3/2}} \qquad K = \frac{\nu_4}{\nu_2^2}.$$

   Ocene za $\hat{S}$ i $\hat{}K$ dobijamo zamenom momenata njihovim ocenama iz uzorka tj. sa $\hat{\nu}_r = \frac{1}{n}\sum u_i^r$.

Koristimo test statistiku

$$JB = n\left(\frac{S^2}{6} + \frac{(K-3)^2}{24}\right).$$

napomena: ako rezidual ima normalnu raspodelu tada JB statistika ima χ^2 raspodelu sa 2 stepena slobode

4.4 Sredina različita od nule

I slučaj: $E(u_i) = \nu$

Ocena za β ostaje sa svim osobinama, ali ocenu za α ne možemo dobiti odvojeno od ocene za ν.

II slučaj: $E(u_i) = \nu_i$

$E(Y_i)$ se ne menja samo zbog promene X_i, što znači da veza izmedju Y_i i X_i nije korektna.

4.5 Heteroskedastičnost

$$Var(u_i) = \sigma_i^2, \qquad \text{ili} \qquad E(u_i^2) = \sigma_i^2$$

- sreće se kod podataka koje dobijamo u preseku vremena;

4.6 Šta se dešava?

1. ocenjivači su i dalje linearni;
2. ocenjivači su i dalje nepristrasni;
3. ocenjivači nisu BLUE;
4. ocenivači nisu efikasni; ($var(\hat{\beta}) = \dfrac{\sum(x_i-\bar{x})^2\sigma_i^2}{(\sum(x_i-\bar{x})^2)^2}$)
5. ocenjivači $\hat{\alpha}$ i $\hat{\beta}$ su konzistentni;
6. ocenjivači $\hat{\alpha}$ i $\hat{\beta}$ nisu asimptotski efikasni;
7. ocenjivač $\hat{\sigma}^2$ je pristrasan;

8. intervali poverenja i testovi značajnosti više ne važe jer su ovenjivači varijansi za $\hat{\alpha}$ i $\hat{\beta}$ pristrasni i ne mogu se primenjivati;

GENERALIZOVAN (PONDERISAN) METOD NAJMANJIH KVADRATA

$$S^* = \sum w_i(y_i - \alpha - \beta x_i)^2$$

$$\tilde{\beta} = \frac{\sum w_i(x_i - \tilde{x})(y_i - \tilde{y})}{\sum w_i(x_i - \tilde{x})^2}; \qquad \tilde{\alpha} = \tilde{y} - \tilde{\beta}\tilde{x},$$

gde je $\tilde{x} = \dfrac{\sum w_i x_i}{\sum w_i}$, $\quad \tilde{y} = \dfrac{\sum w_i y_i}{\sum w_i}$

4.7 Kako rešiti problem?

4.7.1 Pretpostavke o σ_i^2

I **multiplikativna heteroskedastičnost**

$$\sigma_i^2 = \sigma^2 Z_i^\delta$$

δ meri jačinu heteroskedastičnosti; u praksi se često stavlja $Z_i = X_i$;
- primenom metode maksimalne verodostojnosti možemo oceniti ětiti nepoznata parametra; tako dobijene ocene imaju sva asimptotska svojstva;

II **multiplikativna heteroskedastičnost**

$$e_i^2 = a + bX_i + cX_i^2 + v_i,$$

$v_i = e_i^2 - \sigma_i^2$ ....

4.7.2 Ocenjivanje σ_i^2

- pretpostvaka: za svako X_i, $i = 1, ..., m$ imamo n_i opažanja zavisne promenljive Y_i, tj. dobijamo skup opažanja $Y_{i1}, Y_{i2}, ..., Y_{in_i}$

- primenom metode maksimalne verodostojnosti ocenjujemo $m+2$ parametra;

- rešenje odgovarajućeg sistema može se dobiti i iterativnim postupkom; ali je najjednostavnije primeniti

$$s_i^2 = \sum_{j=1}^{n_i} \frac{(Y_{ij} - \bar{Y}_i)^2}{(n_i - 1)}$$

s_i^2 je konzistentan ocenjivč od σ_i^2;
Da li ostaju na snazi sva svojstva ocenjivača MMV? (za $n_i \geq 5$ gubici su jako mali)

4.8 Testovi heteroskedastičnosti

$$H_0: \ \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_m^2$$

4.8.1 Goldfeld-Quandtov test

Test se zasniva na ideji da je varijansa odstupanja jednog dela elemenata iz uzorka generisana u uslovima homoskedastičnost. Test homoskedastičnosti svodi se na test jednakosti dve varijanse.

Golfeld-Quandtova statistika

$$\frac{s_2^2}{s_1^2} : F_{n_2-2, n_1-2},$$

gde je

$$s_1^2 = \frac{\sum_{i=1}^{n_1} (Y_i - \hat{\alpha}_1 - \hat{\beta}_1 X_i)^2}{n_1 - 2}$$

$$s_2^2 = \frac{\sum_{i=n_1+p}^{n_1+p+n_2} (Y_i - \hat{\alpha}_2 - \hat{\beta}_1 X_2)^2}{n_2 - 2}$$

GQ teste je egzaktan, ali ne jako moćan (velika verovatnoća prihvatanja H_0 kada je pogrešna);
preporučuje se kada se elementi uzorka mogu poredjati po rastućoj varijansi odstupanja (npr. kada je varijansa povezana sa X);
eksperimentalno je dobijeno da treba izostaviti oko 1/6 opažanja; (ovo je velika mana testa);

4.8.2 Breusch-Paganov test ili test Lagranžovih multiplikatora

Zasniva se na pretpostavci da se ocene metoda najmanjih kvadrata i metoda maksimalne verodostojnosti u slučaju homoskedastičnosti nebi smele jako razlikovati.
alternativna hipoteza je

$$H_A(\sigma_i = g(\gamma_0 + \gamma_1 Z_{i1} + \cdots + \gamma_p Z_{ip}) \qquad i = 1, 2, ..., n)$$

gde je g neprekidno diferencijabilna funkcija;

Test je prilično osetljiv na neznatna narušavanja normalnosti;

4.8.3 Whiteov test

Zasniva se na poredjenju varijansi ocenjivača dobijenih u uslovima homoskedastičnosti iheteroskedastičnosti.

Chapter 5

Dodatak

$$\frac{P(H_1|D)}{P(H_2|D)} = \frac{P(D|H_1)}{P(D|H_2)} \cdot \frac{P(H_1)}{P(H_2)}$$

Izraz $\frac{P(H_1|D)}{P(H_2|D)}$ naziva se **posterior odds** naknadni izgledi,izraz $\frac{P(H_1)}{P(H_2)}$ **prior odds** prethodni izgledi, a $\frac{P(D|H_1)}{P(D|H_2)}$ **likelihood ratio**.

- **random variables** - slučajna promenljiva (za svako $a \in R$ postoji $P(X < a)$: diskretne, neprekidne i ostale;
- **probability distribution** - zakon raspodele;
- **probability density distribution** (skraćeno p.d.f.), u oznaci $f(x)$ - funkcija gustine;
- **cumulative distribution function** (skraćeno c.d.f.), u oznaci $F(x)$ - funkcija raspodele;
- **joint distribution** - zajednička raspodela $f(x, y)$;
  marginal distribution - marginalne raspodele $f(x)$ i $f(y)$;
  conditional distribution - uslovne raspodele $f(x|y)$, $f(y|x)$
  veza
  $$f(x, y) = f(x)f(y|x) = f(y)f(x|y)$$
- važne raspodele
  - **normal distribution:** $\mathcal{N}(\mu, \sigma^2)$ $\quad f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)}, \quad x \in R$

- χ^2**-distribution:** χ_n^2 $Z = \sum_{i=1}^{n} X_i^2, \quad X_i : \mathcal{N}(0,1)$ osobina aditivnosti $Z_m + Z_n = Z_{m+n}$, (nezavisne slučajne promenljive)
- **t-distribution:** t_n Ako je $X : \mathcal{N}(0,1)$ i $Y : \chi_n^2$ tada $T = \frac{X}{\sqrt{\frac{Y}{n}}}$
- **F-distribution** $F_{n1,n2}$**:** Ako je $Y_1 : \chi_{n1}^2 \quad Y_2 : \chi_{n2}^2 \quad Z = \frac{\frac{Y_1}{n1}}{\frac{Y_2}{n2}}$

Posmatracemo tri važne statisticke analize:

1. tačkasta ocena $\hat{\theta} = g(X_1, X_2, \ldots, X_n)$
   osobine
   - **unbiasedness**=nepristrasnost (centriranost) $(E(g) = \theta)$
     (**mean-squared error (MSE)**=$(bias)^2 + variance$)
   - **efficiency**=efikasnost
   - **consistency**=konzistentnost (postojanost) $\lim_{n\to\infty} P(|\theta - \hat{\theta}_n| < \varepsilon) = 1$
2. intervalna ocena
3. testiranje hipoteza