

Mikroekonometrija

Master studije

Predavač: Aleksandra Nojković

Beograd, školska 2021/22

Struktura predavanja

- Uvodna razmatranja o modelima specifične zavisne promenljive
- Modeli diskretnog izbora
- Modeli binarnog izbora
 - Specifikacija
 - Ocenjivanje
 - Zaključivanje

Modeli specifične zavisne promenljive

- Dve klase ekonometrijskih modela čije su zavisne promenljive specifične:
 - 1) Modeli kvalitativne zavisne promenljive
 - Modeli binarnog izbora
 - Modeli višestrukog izbora (poređanog ili nepoređanog)
 - Modeli prebrojivih događaja
 - 2) Modeli zavisne promenljive sa ograničenjim
 - Modeli odsečene i cenzurisane funkcije raspodele zavisne promenljive
 - Modeli korekcije uzoračkog izbora
 - Modeli trajanja

Primena modela specifične zavisne promenljive

- Predstavljaju standardni metodološki okvir mikroekonometrijske analize, odnosno analize odluka koje donose pojedinci, domaćinstva ili preduzeća.
- Intenzivno se primenjuju u analizi uporednih podataka i najveću primenu nalaze u istraživanjima tržišta rada, ponašanju potrošača i ekonomiji saobraćaja.
- U novim ekonometrijskim istraživanjima otvoreno je pitanje primene ekonometrijskih metoda modela specifične zavisne promenljive u analizi strukturnih odnosa makroekonomskih vremenskih serija i podataka panela.

Modeli diskretnog izbora

- Modeli diskretnog izbora su modeli u kojima zavisna promenljiva uzima dva ili više modaliteta, a još ih nazivamo i modelima kvalitativnog odgovora/ishoda (***qualitative response (QR) models***).
- Modeli ove grupe mogu biti modeli binarnog i modeli višestrukog izbora (***binominal*** and ***multinomial*** models), u zavisnosti da li izbor vrši između dve ili više mogućih alternativa.
- Modeli višestrukog izbora se dalje dele na modele nepoređanog i modele poređanog izbora, kao i modele prebrojivih podataka/događaja (zavisna promenljiva uzima cele, nenegativne vrednosti).

Literatura

- Greene, W.H. (2018), *Econometric Analysis*, 8th edition, Pearson.
- Wooldridge, J. M. (2011), *Econometric Analysis of Cross Section and Panel Data*, 2nd edition, The MIT Press.
- Maddala, G.S. (1983), *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press.
- Nojković, A. (2017). *Odabrane ekonometrijske teme: metodologija i primena*, Ekonomski fakultet, Beograd.
- **Dodatno:** Literatura preglednog sadržaja o QR modelima: Amemiya (1981, 1985), Dhrymes (1984) and McFadden (1984).

Primeri promenljivih diskretnog izbora

- **Broj** nedozvoljenih minusa koje je vlasnik kreditne kartice ostvario u određenom periodu, broj poseta lekaru, broj avionskih/pomorskih nesreća i slično: $y = 0, 1, 2, \dots$. Ovo su primeri **prebrojivih** događaja (preovlađuje broj ishoda nula, a zavisna promenljiva uzima **diskretne vrednosti**).
- Odluka pojedinaca o učešću u radnoj snazi ili članstvu u sindikatu: y uzima vrednost 1 ako je pojedinac zaposlen (član sindikata), a 0 u suprotnom slučaju. Postoji i drugi primeri izbora **između dve alternative**. Šta opredeljuje pojedinca da kupi automobili ili prestane da puši? Koji faktori utiču na odluku o rađanju dece ili odluku banke da odobri stambeni kredit? Koje determinante opredeljuju međunarodne fin. institucije da uđu u kreditni aranžman ili odobre pomoć nekoj zemlji?

Primeri promenljivih diskretnog izbora (nastavak)

- **Stav pojedinaca** u istraživanjima javnog mnjenja (npr. 0 – neslaganje u potpunosti do 4 – slaganja u potpunosti), **istraživanje zaposlenosti** (pojedinaac može biti nezaposlen, zaposlen sa nepunim ili punim radnim vremenom), **nivo obrazovanja** ili **preferencije** potrošača su primeri poređanih alternativa (rang dodeljen zavisnoj promenljivoj je **ordinalnog karaktera**)
- **Izbor zanimanja** (npr. 1-nekvalifikovan radnik do 5-stručnjak), **izbor načina plaćanja** (gotovina ili različite vrste platnih kartica) ili **izbor prevoznog sredstva** (npr. 1-automobil, 2-autobus i 3-voz), su primeri izbora između nepoređanih alternativa. Ovi vrednosti zavisne promenljive ne predstavljaju ni prebrojive podatke ni dodeljene rangove.

Modeli binarnog izbora

- Analiziramo model kojim se opisuje odluka banke da prihvati ili odbije zahtev za odobravanje stambenog kredita ($y=1$ ako je zahtev odbijen, odnosno 0 pri surotnoj odluci).
- **Linearni model verovatnoće, LMV** (*Linear probability model, LPM*) je specijalan slučaj višestrukog linearnog regresionog modela sa 0/1 zavisnom promenljivom:

$$y_i = \beta'x_i + \varepsilon_i,$$

pri čemu je x vektor objašnjavajućih promenljivih, dok je β vektor nepoznatih parametara kojima se opisuje uticaj na verovatnoću realizacije ishoda $y=1$ do koje dovodi promena objašnjavajućih promenljivih za jednu jedinicu.

Nedostaci LMV:

1. Prisutna heteroskedastičnost:

$$\text{var}(\varepsilon_i) = \beta'x_i(1 - \beta'x_i).$$

2. Ne možemo ograničiti $\beta'x_i$ u intervalu $[0, 1]$. Ovi modeli često predviđaju nelogične verovatnoće (negativne vrednosti i vrednosti veće od 1), kao i negativne varijanse.
3. Verovatnoća ishoda 1 raste linearno sa rastom X (β_i promena verovatnoće $Y=1$ sa rastom X_i za jednu jedinicu)

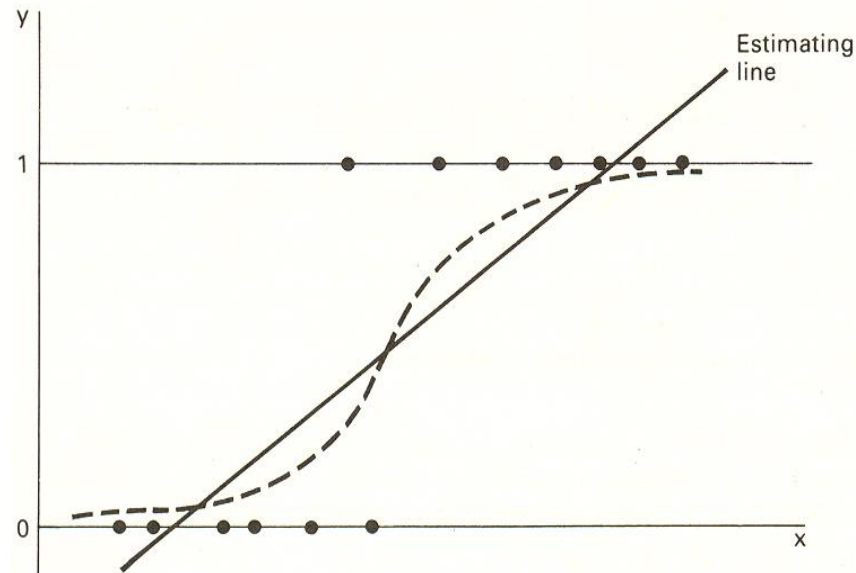
Rešenje prvog nedostatka (neefikasnih ocena LMV)

- Asimptotski efikasan metod ocenjivanja su ponderisani najmanji kvadrati (*weighted least sqears*).
- Postupak se sprovodi u dva koraka:
 1. Prvo se jednačina LMV oceni metodom običnih najmanjih kvadrata, a zatim se izračunaju vrednosti pondera: $w_i = [\hat{y}_i(1 - \hat{y}_i)]^{1/2}$.
 2. Ponovnom primenom običnih najmanjih kvadrata oceni regresija y_i / w_i na x_i / w_i .
- Mogu se dobiti negativne vrednosti pondera w_i (mala verovatnoća u velikim uzorcima).

Problemi koji ostaju u LMV

- Kako vrednosti zavisne promenljive nisu normalno raspodeljene, nijedan metod ocenjivanja linearan po y u opštem slučaju nije efikasan.
- R^2 u ovim modelima su daleko ispod vrednosti 1 i u većini praktičnih istraživanja se kreću od 0.2 do 0.6.
- Izračunata verovatnoća se u praksi često nalazi izvan granica intervala $(0, 1)$. Kao moguće rešenje predlaže se pripisivanje vrednosti 0 i 1, za dobijene negativne vrednosti, odnosno vrednosti zavisne promenljive veće od jedan, redom.
- Najveći nedostatak ovih modela sadržan u pretpostavci da rast x dovodi do **linearnog rasta** y .

Novi modeli verovatnoće prilagođeni binarnoj zavisnoj promenljivoj



- Potrebno je obezbediti da se ocenjena verovatnoća nađe unutar intervala 0-1 (prikazano isprekidanom linijom).
- Odgovarajuće funkcije raspodele koje se najčešće koriste su: **normalna** funkcija raspodele (***cumulative normal function***) i **logistička** funkcija raspodele (***logistic function***)
- Pored toga što ovi modeli obezbeđuju da se predviđene verovatnoće nađu **u intervalu 0-1**, uticaj jedinične promene objašnjavajuće promenljive na verovatnoću pozitivnog ishoda **nije linearan**, već zavisi od stepena strmosti funkcije raspodele za date vrednosti x (verovatnoća se približava nuli, kao i jedinici po sve sporijoj stopi).

Primer 1: Odluka banke o odobrenju kredita

- Izvor: J. Stock and M. Watson, *Introduction to Econometrics*, Addison Wesley, Pearson International Edition, 2003.
- HMDA data (Home Mortgage Disclosure Act) su podaci koji se odnose na zahteve za odobravanje hipoteka podnetih u oblasti Bostona (Massachusetts) tokom 1990 godine.
- 28% prijava koje su podneli afro-amerikanci je odbijeno, dok je to slučaj sa samo 9% prijava podnetih od strane belaca. Dokaz rasne diskriminacije?
- LMV: $Pr (Y=1 | X_1, X_2, \dots, X_k) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$.
- Probit Model: $Pr (Y=1 | X) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)$.
- Logit Model: $Pr (Y=1 | X) = \Lambda(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k) =$

$$= \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)}} \cdot$$

FIGURE 9.1 Scatterplot of Mortgage Application Denial and the Payment-to-Income Ratio

Mortgage applicants with a high ratio of debt payments to income (*P/I ratio*) are more likely to have their application denied (*deny* = 1 if denied, *deny* = 0 if approved). The linear probability model uses a straight line to model the probability of denial, conditional on the *P/I ratio*.

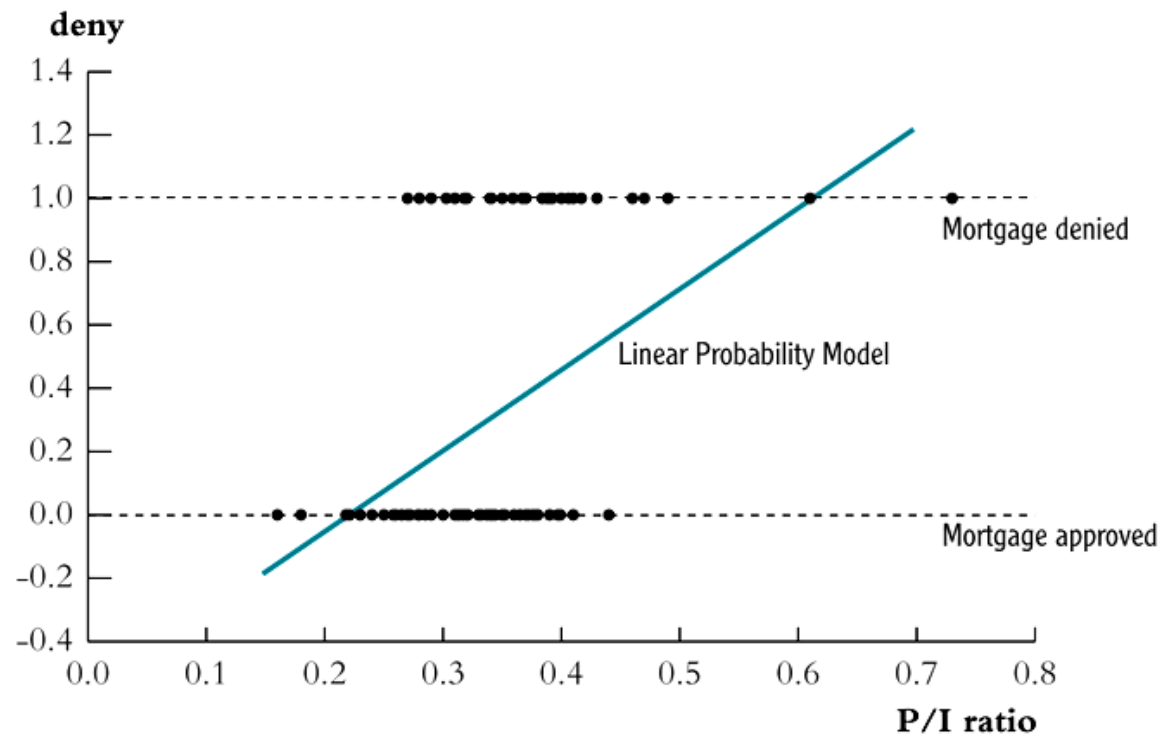
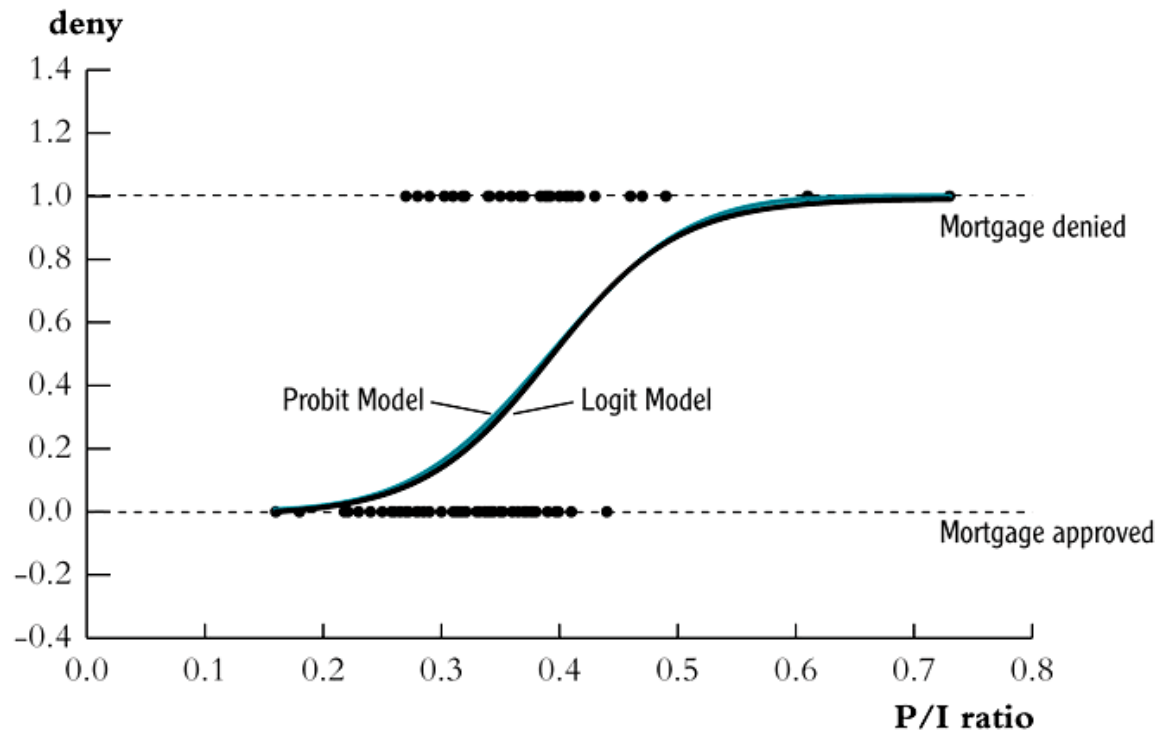


FIGURE 9.3 Probit and Logit Models of the Probability of Denial, Given the P/I Ratio

These logit and probit models produce nearly identical estimates of the probability that a mortgage application will be denied, given the payment-to-income ratio.



Nelinearni modeli

- Model uslovne verovatnoće definisan je kao:

$$E[y | x] = 0 [1 - F(\beta'x)] + 1 [F(\beta'x)] = F(\beta'x)$$

- Opredeljenje za ma koju funkciju raspodele verovatnoće **ne obezbeđuje** da ocenjeni parametri modela predstavljaju marginalne efekte koje interpretiramo na način uobičajen za linearne modele. Marginalni efekti se u opštem slučaju izračunavaju kao:

$$\frac{\partial E[y | x]}{\partial x} = \left\{ \frac{dF(\beta'x)}{d(\beta'x)} \right\} \beta = f(\beta'x) \beta,$$

gde je $f(\cdot)$ funkcija gustine koja odgovara funkciji raspodele $F(\cdot)$.

Izračunavanje marginalnih efekata

- Za slučaj normalne funkcije raspodele marginalni efekti se izračunavaju kao: $[\Phi(\beta'x) \beta]$, dok je za slučaj logit modela: $[\Lambda(\beta'x) (1 - \Lambda(\beta'x)) \beta]$.
- Znak koeficijenta zaista **odgovara smeru promene** verovatnoće.
- Uobičajeno je da se marginalni efekti izračunavaju za vrednosti aritmetičkih sredina vektora x (i vrednost 0 veštačkih objašnjavajućih promenljivih) ili pak za neku drugu vrednost x od interesa.
- Marginalni efekat se može dobiti i kao aritmetička sredina marginalnih efekata izračunatih za vrednost svake pojedinačne opservacije u uzorku.

Nelinearni modeli – alternativni pristup

- Zasniva se na primeni modela indeksne funkcije $\beta'x_i$ (*index function models*). U model uvodimo pojam latentne promenljive y_i^* koja definisana kao:

$$y_i^* = \beta'x_i + \varepsilon_i .$$

Ova promenjiva se ne opaža u praksi, a ono što opažamo je veštačka, indikator promenljiva y koja je definisana na sledeći način:

$$y_i = 1 \text{ if } y_i^* > 0 \text{ i } y_i = 0 \text{ if } y_i^* \leq 0 .$$

U ovoj formulaciji modela, $F(\beta'x_i)$ nije $E(y_i/x_i)$ kao u LMV, već je $E(y_i^*/x_i)$.

- Sada je verovatnoća da je $Y = 1$:

$$E[y_i^* | x] = Pr[y_i = 1 | x]$$
$$Pr(Y_i = 1) = Prob(\varepsilon_i > -\beta'x_i) = 1 - F(-\beta'x_i) ,$$

gde je F odgovarajuće funkcija raspodele slučajne greške ε .

Logit and probit model

- Ukoliko pretpostavimo logističku funkciju raspodele za slučajnu grešku ε_j , dolazimo do specifikacije **logit modela**:

$$Prob(Y = 1) = e^{\beta'x} / (1 + e^{\beta'x}) = \Lambda(\beta'x),$$

pri je $\Lambda(\cdot)$ oznaka logističke funkcije raspodele.

- U **probit modelu** (još se naziva i normit model) pretpostavljamo normalnu standardizovanu funkciju raspodele slučajne greške ($\varepsilon_j : N(0,1)$) iz čega sledi:

$$Prob(Y = 1) = \int_{-\infty}^{\beta'x} \frac{1}{\sqrt{2\pi}} \exp\left[-\left(t^2/2\right)\right] dt = \int_{-\infty}^{\beta'x} \phi(t) dt = \Phi(\beta'x)$$

gde je $\Phi(\cdot)$ uobičajena notacija za standardizovanu normalnu raspodelu.

Logit ili Probit model?

- Osim što je matematički jednostavniji za ocenjivanje, logit model rešavanjem po $\beta'x$ daje sledeću relaciju:

$$\left[\frac{P}{(1-P)} \right] = e^{\beta'x}$$

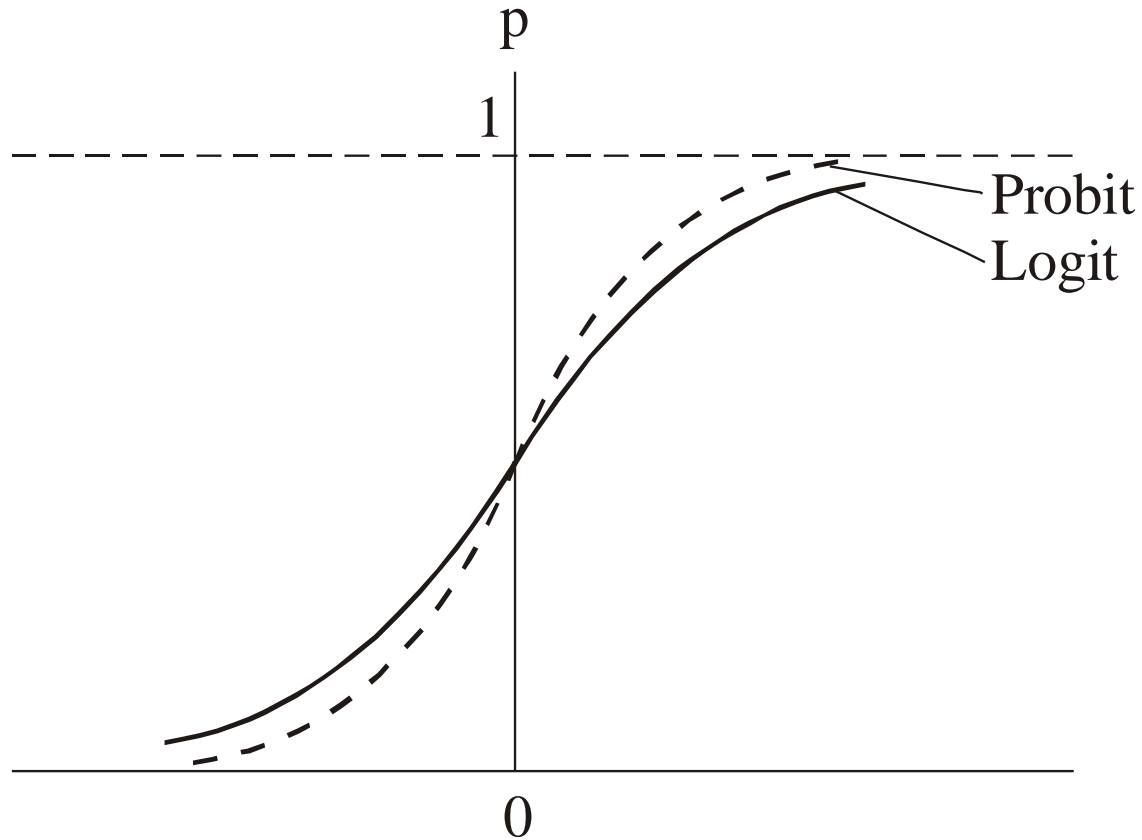
gde je P_i verovatnoća realizacije pozitivnog ishoda, $y=1$, dok je $(1-P_i)$ verovatnoća realizacije suprotnog događaja, $y=0$. Količnik $[P_i / (1-P_i)]$ je poznat “**odds ratio**” (odnos verovatnoća prvog i drugog izbora). Prirodni logaritam odnosa ovih verovatnoća je:

$$\ln [P / (1-P)] = \beta'x,$$

iz čega se uočava veoma jednostavna interpretacija da zavisna promenljiva uzima **pozitivnu vrednost** ako je **verovatniji prvi izbor**, a negativnu vrednost u suprotnom slučaju (**logističke regresije**).

- Dve raspodele veoma bliske, osim na krajevima, gde logistička funkcija raspodele ima “teže” repove (logistička raspodela je zapravo bliža t raspodeli sa sedam stepeni slobode).

Funkcije raspodele logit i probit modela



Ocenjivanje modela binarnog izbora

- Sa izuzetkom LMV, modeli binarnog izbora se **ocenjuju metodom maksimalne verodostojnosti (MV)**.
- Svaka opservacija se tretira kao jedno izvlačenje iz Bernoulli-eve raspodele sa verovatnoćom pozitivnog odgovora $F(\beta'x_i)$, tada se **združena** funkcija verodostojnosti **nezavisnih** opservacija definiše kao:

$$\text{Pr ob } (Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n) = \prod_{y_i=0} [1 - F(\beta'x_i)] \prod_{y_i=1} F(\beta'x_i),$$

odnosno, $L(x; \beta) = \prod [F(\beta'x_i)]^{y_i} [1 - F(\beta'x_i)]^{1-y_i}$, gde je $i = 1, \dots, n$.

- Na osnovu principa maksimalne verodostojnosti za ocenu parametra β se uzima ona statistika za koju $L(x; \beta)$, pri fiksiranom x , dostiže svoju najveću vrednost, tj. za koju je: $L(x; \hat{\beta}) = \sup_{\beta \in B} L(x; \beta)$, kada supremum postoji pri bilo kojoj dopustivoj vrednosti β iz prostora B .

Ocenjivanje modela binarnog izbora (nastavak)

- U slučaju kada je L dva puta diferencijabilna u odnosu na β , do ocene maksimalne verodostojnosti možemo relativno lako da dođemo određivši sve lokalne maksimume funkcije L iz uslova:

$$\partial L(x; \beta) / \partial \beta = 0, \quad \partial^2 L(x; \beta) / \partial \beta \partial \beta' < 0.$$

Za ocene maksimalne verodostojnosti uzimamo one vrednosti β za koje se postiče apsolutni maksimum.

- Kako logaritam funkcije verodostojnosti:

$$\log L_n = \log L_n(x; \beta) = \sum_{i=1}^n \{y_i \log F(\beta' x_i) + (1 - y_i) \log [1 - F(\beta' x_i)]\},$$

dostiže ekstremne vrednosti za iste vrednosti β za koje to postiže i funkcija L rešenja se, zbog jednostavnije operacionalizacije, traže iz jednačine:

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^n \frac{y_i - F_i}{F_i(1 - F_i)} f_i x_i = \sum_{i=1}^n \left[\frac{y_i f_i}{F_i} + (1 - y_i) \frac{-f_i}{(1 - F_i)} \right] x_i = 0$$

gde je f_i funkcija gustine ($\partial F_i / \partial (\beta' x_i)$) koja odgovara odabranoj funkciji raspodele F_i .

Ocenjivanje modela binarnog izbora (nastavak)

- Uslov za pronalaženje maksimuma log L dat je kao:

$$\partial \log L(\beta) / \partial \beta = 0 \quad \text{i} \quad \partial^2 \log L(\beta) / \partial \beta \partial \beta' < 0.$$

- Uslov prvog reda za maksimiziranje log L je jednačina verodostojnosti $S_n(\beta)$ poznata i kao *funkcija ostvarenih pogodaka (score function)*:

$$S_n(\beta) = \frac{\partial \log L_n}{\partial \beta} = 0,$$

- Jednačine verodostojnosti logit i probit modela su nelinearne po β , što zahteva primenu nekog od poznatih iterativnih metoda optimizacije.

- Najčešće se primenjuje Newton-Raphson-ov metod, pri čemu se za oba modela jednostavno pokazuje da je ispunjen uslov globalne konkavnosti (Hessian matrica je **uvek negativno definitna**:

$$H_n(\beta) = \partial^2 \log L_n(\beta) / \partial \beta \partial \beta' < 0).$$

Ocenjivanje modela binarnog izbora (nastavak)

- Asimptotske ocene kovarijante matrice ocena MV moguće je izračunati na tri načina:

1. Kao negativna inverzna Hessian matrica za ocene $MV\hat{\beta}$:

$$A_{\text{var}}(\hat{\beta}) = \left[- \frac{\partial^2 \ln L(\hat{\beta})}{\partial \hat{\beta} \partial \hat{\beta}'} \right]^{-1} = \hat{V}$$

2. The Berndt, Hall, Hall, and Hausman (BHHH) ocene:

$$B = \sum_i g_i^2 x_i x_i'$$

gde je g_i izračunato na bazi prvih izvoda logit i probit modela.

3. Ocena zasnovana na očekivanoj vrednosti Hessian matrice : $H = E(H)$.

Zaključivanje u modelima binarnog izbora

- Ocene MV su **konzistentne** i poseduju asimptotski **normalnu raspodelu**
- ***Testiranje hipoteza:***
 - Za testiranje ograničenja na pojedinačne parametre modela koriste se kritične vrednosti standardizovane normalne raspodele (z).
 - Testiranje složenijih hipoteza, kojima se zahteva da koeficijenti u ocenjenom modelu zadovoljavaju izvesna linearna ili nelinearna ograničenja, sprovodi se primenom sledećih, asimptotski međusobno ekvivalentnih testova: **Wald-ovog (W)**, **testa količnika verodostojnosti (likelihood ratio, LR)** i **testa Lagrange-ovog multiplikatora (LM)**. Pod pretpostavkom da je tačna nulta hipoteza sve tri test - statistike su *asimptotski ekvivalentne i imaju X^2 raspodelu*, sa brojem stepeni slobode koji je jednak broju ograničenja.

Zaključivanje u modelima binarnog izbora (nastavak)

- Analiza stabilnosti parametara sprovodi se primenom testa koji odgovara Chow testu u klasičnom linearnom regresionom modelu.
- Pretpostavimo da se u vektoru egzogenih promenljivih nalazi bar jedna veštačka promenljiva i da testiramo nultu hipotezu da su ocene preostalih K koeficijenata u modelu nepromenjene bez obzira na vrednost koju uzima veštačka promenljiva.
- Odgovarajuća statistika količnika verodostojnosti, kojom se testira K ograničenja na parametre, formira se polazeći od izraza :

$$LR = -2(\ln\hat{L}_O - \ln\hat{L}_B),$$

pri čemu se oznaka (O) odnosi na logaritam funkcije verodostojnosti dobijen za sve podatke u uzorku ($n=n_1+n_2$), dok se oznaka (B) odnosi na zbir logaritama funkcija verodostojnosti dva zasebna modela, koji su ocenjeni odvojeno na bazi n_1 i n_2 opservacija (modeli bez ograničenja).

Wald-ov test

- Pretpostavimo da se J međusobno funkcionalno nezavisnih ograničenja (linearnih ili nelinearnih) može predstaviti sledećom nultom hipotezom: $H_0: R(\beta) = q$,
gde je q poznata vektor kolona od J elemenata, pri čemu je uslov da je $J < K$ (K broj nepoznatih parametara u modelu).
- Poznata je matrica C dimenzija ($J \times K$) i ranga J , definisana kao:

$$C = \frac{\partial R(\beta)}{\partial \beta'}$$

- Za sprovođenje Wald-ovog testa potrebno je oceniti model bez ograničenja, a zatim testirati da li ocenjeni koeficijenti značajno odstupaju od zahtevanih restrikcija.
- Navedeni skup ograničenja testira se koristeći statistiku:

$$W = \left(R(\hat{\beta}) - q \right)' \left[C \hat{V} C' \right]^{-1} \left(R(\hat{\beta}) - q \right),$$

gde je \hat{V} asimptotska kovarijantna matrica ocena maksimalne verodostojnosti a C matrica prvih izvoda izračunate za ovako ocenjene vrednosti parametara.

LR test

- Podrazumeva ocenjivnije i modela sa ograničenjem i modela bez ograničenja i zasniva se na sledećoj statistici:

$$LR = -2(\ln\hat{L}_O - \ln\hat{L}_B),$$

gde se oznake (O) i (B) odnose na funkcije verodostojnosti izračunate sa i bez ograničenja, respektivno.

- Ovim testom možemo testirati hipotezu da su svi parametri probit i logit modela (osim konstantnog člana) jednaki nuli, odnosno $H_0: \beta_0 = \beta_1 = \dots = \beta_K = 0$.
- Prihvatanje nulte hipoteze značilo bi da ni jedna od eksplanatornih promenljivih u modelu nema uticaja na verovatnoću izbora (poput F testa u klasičnom linearnom regresionom modelu)

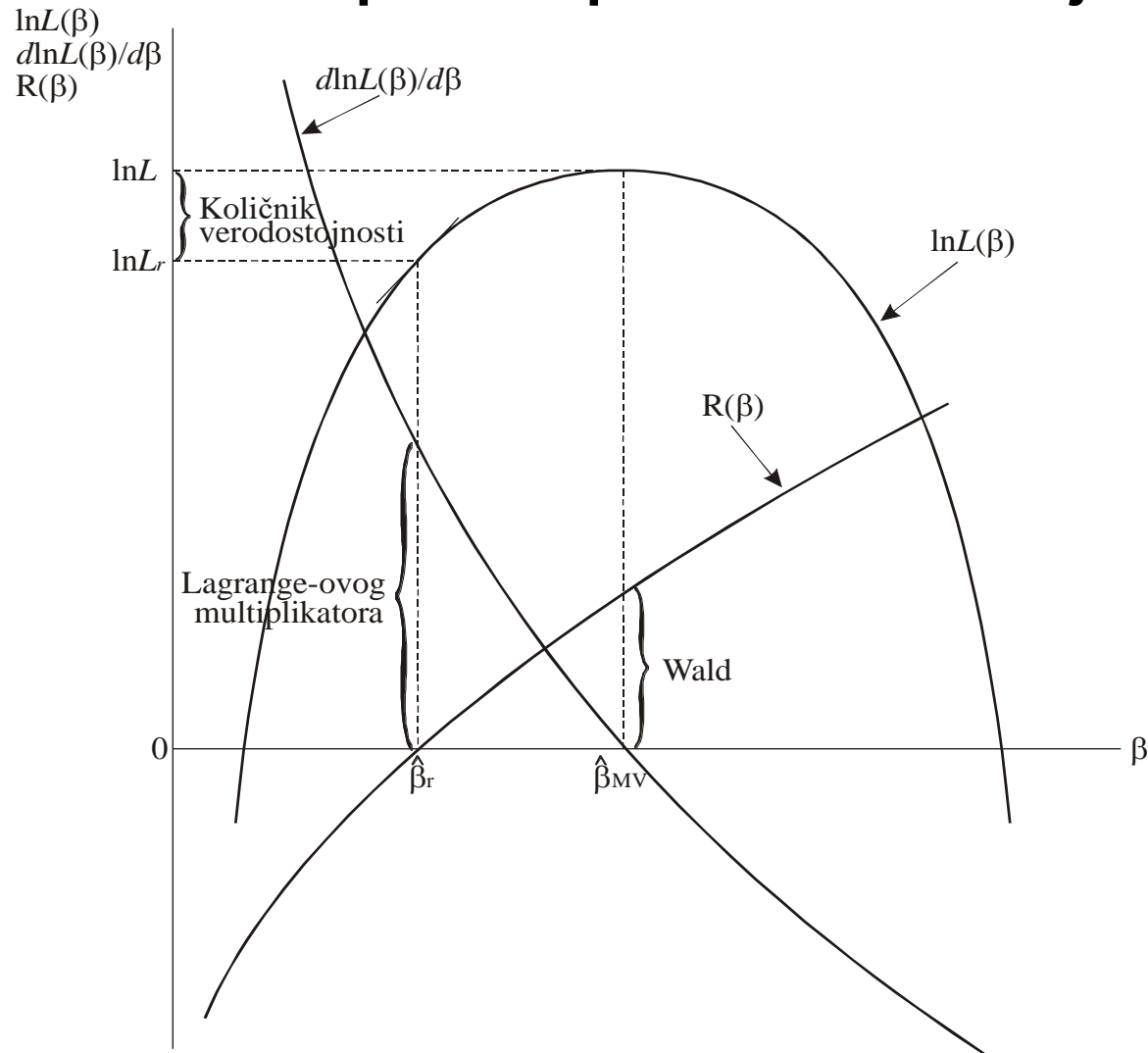
LM test

- Zasniva se samo na modelu koji je ocenjen pod ograničenjem.
- Ako je ograničenje definisano nultom hipotezom tačno, ocena dobijena pod datim ograničenjem ($\hat{\beta}_o$) ne razlikuje se značajno od vrednosti koja maksimizira logaritam funkcije verodostojnosti, odnosno nagib funkcije $\ln L$ u toj tački se ne razlikuje statistički značajno od nule (H_o : prvi izvod u ovoj tački je jednak nuli).
- Statistika Lagrange-ovog multiplikatora je:

$$LM = g'Vg ,$$

gde je g vektor prvih izvoda izračunat pod ograničenjem nulte hipoteze, a V je asimptotska ocena kovarijantne matrice ocena MV , dobijena na jedan od tri predložena načina.

Odnos tri postupka testiranja



- Pretpostavka je da su ograničenja na parametre predstavljena sledećom nultom hipotezom: $H_0: R(\beta) = 0$.

Poređenje ocena LMV, probit and logit modela

- Logistička raspodela ima varijansu $\pi^2/3$, tako da ocene za β dobijene u logit modelu množenjem sa $3^{1/2}/\pi$ postaju uporedive sa ocenama dobijenim u probit modelu. Amemiya (1981) je kao predložio sledeću vezu:

$$\beta_{probit} \approx 0.625 \beta_{logit},$$

Dakle, potrebno je ocene iz logit modela podeliti sa 1.6 (umesto sa $\pi / 3^{1/2} = 1.8$) čime se postiže bolja aproksimacija ocena dobijenih iz ova dva modela.

- Isti autor je predložio i sledeću vezu sa ocenama dobijenim u LMV (LPM):

$$\begin{aligned} \beta_{LMV} &\approx 0.25 \beta_{logit} \quad (\text{osim za konstantu}) \\ \beta_{LMV} &\approx 0.25 \beta_{logit} + 0.5 \quad (\text{za konstantu}). \end{aligned}$$

Pokazatelji kvaliteta modela binarnog izbora

- **Pseudo – R^2** ili **McFadden-ov indeks** količnika verodostojnosti (McFadden's likelihood ratio index), koji poredi vrednost logaritma funkcije verodostojnosti u ocenjenog modela (L) i modela u kome je prisutna samo konstanta (isključene su sve objašnjavajuće promenljive iz modela, L_0):

$$pseudo-R^2 = LRI = 1 - (\ln L / \ln L_0).$$

Vrednost ovog pokazatelja se kreće u intervalu od 0 do 1, podseća na koeficijent determinacije linearnih modela ali nema tako direktnu interpretaciju.

- **Procenat tačnih predviđanja modela (engl. Expectation-Prediction Evaluation)**, koji je prestavljen tabelom *pogodaka* i *promašaja* dimenzija 2x2 u slučaju binarnog modela, a kao pravilo predviđanja koristi se:

$$y = 1 \text{ if } F > F^* \text{ and } 0 \text{ otherwise.}$$

Najčešće se kao granica predviđanja koristi $F^* = 0.5$ (u tom slučaju model predviđa ishod 1 ako je verovatnoća realizacije ovog ishoda veća). Problem ovog pokazatelja uočljiv u nebalansiranim uzorcima (mali broj ishoda 1 u uzorku).

Pokazatelji kvaliteta modela binarnog izbora (nastavak)

- Dva testa koje koriste X^2 -statistiku pri testiranju saglasnosti modelom ocenjenih vrednosti sa pravim podacima.
- Razlika ova dva testa je u grupi opservacija koja je relevantna i koja je asimptotska raspodela testa.
 - 1) Hosmer-Lemeshow test posmatra tačna predviđanja za $y_i=1$.
 - 2) Andrews-ov test je opštiji, tako da su opservacije grupisane i poređene po bilo kom obeležju.
- Moguće je dobiti različit rezultat testiranje primenom ova dva testa (kombinuju se sa drugim pokazateljima kvaliteta).

Greške specifikacije modela binarnog izbora

- Najčešće greške specifikacije su: izostavljanje relevantne objašnjavajuće promenljive i prisustvo heteroskedastičnosti u modelu.
- Posledice navedenih grešaka specifikacije u probit i logit modelima su:
 - 1) Izostavljanjem X_2 iz modela u koji je opravdano uključiti X_1 i X_2 (odnosno, ako je $\beta_2 \neq 0$), ocena maksimalne verodostojnosti postaje asimptotski pristrasna:

$$\text{plim } c_1\beta_1 + c_2\beta_2,$$

gde su c_1 i c_2 komplikovane funkcije nepoznatih parametara, što ima za posledicu **neKonzistentnost** ocene parametra β_1 , čak i slučaju nezavisnosti promenljivih X_1 i X_2 .

2) Posledice prisustva heteroskedastičnosti u modelu su: **neKonzistentnost** ocena maksimalne verodostojnosti i **nepouzdanost** asimptotske ocene kovarijantne matrice. Heteroskedastičnost se **veoma često** javlja u mikroekonometrijskim podacima.

Greške specifikacije modela binarnog izbora (II)

- Za potrebe analize specifikacije modela moguće je koristiti sve tri test statistike: Wald-ovu, LR i LM.
- Za proveru opravdanosti uključivanja seta prom. X_2 Najjednostavniji je test Lagrange-ovog multiplikatora (zasniva se samo na ocenama modela pod ograničenjem datim nultom hipotezom). Nulta hipoteza o polaznoj specifikaciji modela (na primer, model koji sadrži samo objašnjavajuću promenljivu X_1) proverava se protiv alternativne specifikacije modela (u modelu je opravdano prisustvo dodatog regresora X_2) sledećom test statistikom:

$$LM = g_0'V^{-1}g_0 ,$$

gde je g_0 vektor prvih izvoda logaritma funkcije verodostojnosti alternativnog modela, ocenjen pod ograničenjem datim nultom hipotezom, a V^{-1} predstavlja ocene asimptotske kovarijantne matrice izačunte na bazi ocena maksimalne verodostojnosti polaznog modela.

Ostale greške specifikacije

- neispunjenost pretpostavke o normalnom rasporedu slučajne greške
- autokorelacija
- pristrasnost u izboru uzorka
- neispunjenost pretpostavke egzogenost promenljivih
- neopravdanost pretpostavke o nezavisnosti irelevantnih alternativa (videti logit model višestrukog izbora).

Primer 1: Odluka banke o odobrenju kredita (nastavak)

TABLE 9.1 Variables Included in Regression Models of Mortgage Decisions

Variable	Definition	Sample Average
Financial Variables		
<i>P/I ratio</i>	Ratio of total monthly debt payments to total monthly income	0.331
<i>housing expense-to-income ratio</i>	Ratio of monthly housing expenses to total monthly income	0.255
<i>loan-to-value ratio</i>	Ratio of size of loan to assessed value of property	0.738
<i>consumer credit score</i>	1 if no “slow” payments or delinquencies 2 if one or two slow payments or delinquencies 3 if more than two slow payments 4 if insufficient credit history for determination 5 if delinquent credit history with payments 60 days overdue 6 if delinquent credit history with payments 90 days overdue	2.1
<i>mortgage credit score</i>	1 if no late mortgage payments 2 if no mortgage payment history 3 if one or two late mortgage payments 4 if more than two late mortgage payments	1.7
<i>public bad credit record</i>	1 if any public record of credit problems (bankruptcy, charge-offs, collection actions) 0 otherwise	0.074

TABLE 9.1 Variables Included in Regression Models of Mortgage Decisions

Variable	Definition	Sample Average
<i>Additional Applicant Characteristics</i>		
<i>denied mortgage insurance</i>	1 if applicant applied for mortgage insurance and was denied, 0 otherwise	0.020
<i>self-employed</i>	1 if self-employed, 0 otherwise	0.116
<i>single</i>	1 if applicant reported being single, 0 otherwise	0.393
<i>high school diploma</i>	1 if applicant graduated from high school, 0 otherwise	0.984
<i>unemployment rate</i>	1989 Massachusetts unemployment rate in the applicant's industry	3.8
<i>condominium</i>	1 if unit is a condominium, 0 otherwise	0.288
<i>black</i>	1 if applicant is black, 0 if white	0.142
<i>deny</i>	1 if mortgage application denied, 0 otherwise	0.120

TABLE 9.2 Mortgage Denial Regressions Using the Boston HMDA Data**Dependent Variable: deny = 1 If Mortgage Application Is Denied, = 0 If Accepted; 2,380 observations.**

Regression Model	LPM	Logit	Probit	Probit	Probit	Probit
Regressor	(1)	(2)	(3)	(4)	(5)	(6)
<i>black</i>	0.084** (0.023)	0.688** (0.182)	0.389** (0.098)	0.371** (0.099)	0.363** (0.100)	0.246 (0.448)
<i>P/I ratio</i>	0.449** (0.114)	4.76** (1.33)	2.44** (0.61)	2.46** (0.60)	2.62** (0.61)	2.57** (0.66)
<i>housing expense-to-income ratio</i>	-0.048 (.110)	-0.11 (1.29)	-0.18 (0.68)	-0.30 (0.68)	-0.50 (0.70)	-0.54 (0.74)
<i>medium loan-to-value ratio</i> (0.80 ≤ <i>loan-value ratio</i> ≤ 0.95)	0.031* (0.013)	0.46** (0.16)	0.21** (0.08)	0.22** (0.08)	0.22** (0.08)	0.22** (0.08)
<i>high loan-to-value ratio</i> (<i>loan-value ratio</i> ≥ 0.95)	0.189** (0.050)	1.49** (0.32)	0.79** (0.18)	0.79** (0.18)	0.84** (0.18)	0.79** (0.18)
<i>consumer credit score</i>	0.031** (0.005)	0.29** (0.04)	0.15** (0.02)	0.16** (0.02)	0.34** (0.11)	0.16** (0.02)
<i>mortgage credit score</i>	0.021 (0.011)	0.28* (0.14)	0.15* (0.07)	0.11 (0.08)	0.16 (0.10)	0.11 (0.08)
<i>public bad credit record</i>	0.197** (0.035)	1.23** (0.20)	0.70** (0.12)	0.70** (0.12)	0.72** (0.12)	0.70** (0.12)
<i>denied mortgage insurance</i>	0.702** (0.045)	4.55** (0.57)	2.56** (0.30)	2.59** (0.29)	2.59** (0.30)	2.59** (0.29)

(Table 9.2 continued)

TABLE 9.2 Mortgage Denial Regressions Using the Boston HMDA Data**Dependent Variable: deny = 1 If Mortgage Application Is Denied, = 0 If Accepted; 2,380 observations.**

Regression Model	LPM	Logit	Probit	Probit	Probit	Probit
Regressor	(1)	(2)	(3)	(4)	(5)	(6)
<i>self-employed</i>	0.060** (0.021)	0.67** (0.21)	0.36** (0.11)	0.35** (0.11)	0.34** (0.11)	0.35** (0.11)
<i>single</i>				0.23** (0.08)	0.23** (0.08)	0.23** (0.08)
<i>high school diploma</i>				-0.61** (0.23)	-0.60* (0.24)	-0.62** (0.23)
<i>unemployment rate</i>				0.03 (0.02)	0.03 (0.02)	0.03 (0.02)
<i>condominium</i>					-0.05 (0.09)	
<i>black × P/I ratio</i>						-0.58 (1.47)
<i>black × housing expense-to-income ratio</i>						1.23 (1.69)
<i>Additional credit rating indicator variables</i>	no	no	no	no	yes	no
<i>constant</i>	-0.183** (0.028)	-5.71** (0.48)	-3.04** (0.23)	-2.57** (0.34)	-2.90** (0.39)	-2.54** (0.35)

(Table 9.2 continued)

(Table 9.2 continued)

F-statistics and p-values Testing Exclusion of Groups of Variables

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Applicant single; HS diploma; industry unemployment rate</i>				5.85 (<0.001)	5.22 (0.001)	5.79 (<0.001)
<i>Additional credit rating indicator variables</i>					1.22 (0.291)	
<i>Race interactions and black</i>						4.96 (0.002)
<i>Race interactions only</i>						0.27 (0.766)
<i>Difference in predicted probability of denial, white vs. black (percentage points)</i>	8.4%	6.0%	7.1%	6.6%	6.3%	6.5%

These regressions were estimated using the $n = 2,380$ observations in the Boston HMDA data set described in Appendix 9.1. The linear probability model was estimated by OLS, and probit and logit regressions were estimated by maximum likelihood. Standard errors are given in parentheses under the coefficients and p -values are given in parentheses under the F -statistics. The change in predicted probability in the final row was computed for a hypothetical applicant whose values of the regressors, other than race, equal the sample mean. Individual coefficients are statistically significant at the *5% or **1% level.

Primer 2. Efekat nove metode učenja (PSI) na rezultat ispita iz makroekonomije

- *Izvor:* L. Spector and M. Mazzeo, “Probit Analysis and Economic Education”, *Journal of Economic Education*, vol.11, 1980, pp.37-44.
- LPM:
$$Y_i = \beta_1 + \beta_2 \text{GPA}_i + \beta_3 \text{TUCE} + \beta_4 \text{PSI} + \varepsilon_i,$$

gde je $Y=1$ ukoliko je na završnom ispitu dobijena ocena A, a $Y=0$ ukoliko je ocena B ili C, GPA – ukupna prosečna ocena studenta, TUCE – ocena na preliminarnom testiranju na početku semestra i PSI – binarna promeljiva, koja uzima vrednost 1 ako je student učio po novoj metodi, a 0 u suprotnom slučaju.

- Problem se može predstaviti sledećim modelom binarnog izbora:

$$P(Y=1 \mid \text{GPA}, \text{TUCE}, \text{PSI}) = p_i = \beta_1 + \beta_2 \text{GPA}_i + \beta_3 \text{TUCE} + \beta_4 \text{PSI} + \varepsilon_i,$$

pri čemu je P oznaka za verovatnoću.

Rezultati ocenjivanja nove metode učenja na 32 opservacije

TABLE 23.1 Estimated Probability Models

<i>Variable</i>	<i>Linear</i>		<i>Logistic</i>		<i>Probit</i>		<i>Gumbel</i>	
	<i>Coefficient</i>	<i>Slope</i>	<i>Coefficient</i>	<i>Slope</i>	<i>Coefficient</i>	<i>Slope</i>	<i>Coefficient</i>	<i>Slope</i>
Constant	-1.498	—	-13.021	—	-7.452	—	-10.631	—
GPA	0.464	0.464	2.826	0.534	1.626	0.533	2.293	0.477
TUCE	0.010	0.010	0.095	0.018	0.052	0.017	0.041	0.009
PSI	0.379	0.379	2.379	0.450	1.426	0.468	1.562	0.325
$f(\bar{x}'\hat{\beta})$	1.000		0.189		0.328		0.208	

TABLE 23.2 Estimated Coefficients and Standard Errors (standard errors in parentheses)

<i>Variable</i>	<i>Logistic</i>				<i>Probit</i>			
	<i>Coefficient</i>	<i>t Ratio</i>	<i>Slope</i>	<i>t Ratio</i>	<i>Coefficient</i>	<i>t Ratio</i>	<i>Slope</i>	<i>t Ratio</i>
Constant	-13.021 (4.931)	-2.641	—	—	-7.452 (2.542)	-2.931	—	—
GPA	2.826 (1.263)	2.238	0.534 (0.237)	2.252	1.626 (0.694)	2.343	0.533 (0.232)	2.294
TUCE	0.095 (0.142)	0.672	0.018 (0.026)	0.685	0.052 (0.084)	0.617	0.017 (0.027)	0.626
PSI	2.379 (1.065)	2.234	0.456 (0.181)	2.521	1.426 (0.595)	2.397	0.464 (0.170)	2.727
log-likelihood	-12.890				-12.819			

Primer 3. Učešće žena na tržištu rada u Srbiji

- 2002 ARS, 1528 zaposlenih žena od ukupno 4376 žena uzrasta od 18 do 65 godina u uzorku
- Zavisna promenljiva Y je binarna, uzima vrednosti 1/0 za odgovore žena da/ne, na pitanje da li su zaposlene ili ne.
- Objašnjavajuće promenljive :
age = godine starosti
m_stat1 = bračni status; binarna promenljiva, uzima vrednost 1 za žene koje nisu u braku, a 0 u suprotnom slučaju.

educ = godine školovanja žene, ili
sc1 = završena osnovna škola
sc2 = završena srednja škola
sc3 = završena viša škola
sc4 = završen fakultet ili više (magistri i doktori nauka).

Primer 3: Model učešća žena u radnoj snazi (ARS Srbija)

Objašnjavajuće promenljive	LMV (ONK)	Logit (MV)	Probit (MV)
konstanta	-1.8867 (0.0627)	-17.7344 (0.6633)	-10.1040 (0.3995)
godine	0.0872 (0.0036)	0.6458 (0.0311)	0.3676 (0.0177)
godine ²	-0.0010 (0.00004)	-0.0077 (0.00037)	-0.0044 (0.00021)
obrazovanje	0.0573 (0.0023)	0.4406 (0.0201)	0.2528 (0.0118)
br.status	-0.0510 (0.0162)	-0.3184 (0.1109)	-0.1701 (0.0644)
Broj opservacija 0/1	2419/1368	2419/1368	2419/1368
Procenat tačnih predviđanja	67.54	79.93	79.93
Log-verodostojnosti	/	-1669.088	-1673.297
LR statistika (<i>prob</i>)	/	1616.174 (0.000)	1607.756 (0.000)
Wald χ^2 (<i>prob</i>)	/	816.020 (0.000)	834.790 (0.000)
Pseudo R ²	0.3246	0.3262	0.3245

Izračunavanje marginalnih efekata (probit)

- $Prob(posao=1)=F(konstanta, godine, godine^2, br. status, obrazovanje)$
- Vrednosti aritmetičkih sredina:

promenljiva	godine	godine ²	obrazovanje	br. status
ar. sredina	41.332	1906.571	10.292	0.638

- Za numeričke promenljive $= f(\beta'x)\beta$ (npr. obrazovanje = $0.32951*0.2528$)
- Za binarne promenljive (bračni status= $0.243-0.300=-0.057$):

$$\begin{aligned}
 & Prob (posao=1 / \overline{godine}, \overline{godine^2}, \overline{obrazovanje}, br.status = 0) \\
 & = \Phi(-9.9339+0.3676*41.332-0.0044*1906.571+ \\
 & \quad +0.2528*10.292+0.1701*0) = \\
 & = \Phi (-0.5254) = 1 - \Phi (0.5254) = 1-0.7002 = 0.300.
 \end{aligned}$$

$$\begin{aligned}
 & Prob (posao=1 / \overline{godine}, \overline{godine^2}, \overline{obrazovanje}, br.status = 1) \\
 & = \Phi(-9.9339+0.3676*41.332- \\
 & \quad -0.0044*1906.571+0.2528*10.292+0.1701*1) = \\
 & = \Phi (-0.6954) = 1 - \Phi (0.6954) = 1-0.7570 = 0.243.
 \end{aligned}$$

Ocene koeficijenata i izračunati marginalni efekti: LMV, logit i probit model

Nezavisne promjenjive	LMV		Logit		Probit	
	Koeficijent	Marg. efekat	Koeficijenti	Marg. efekat	Koeficijenti	Marg. efekat
konstanta	-1.8357	/	-17.734	/	-10.1040	/
godine	0.0872	0.0872	0.646	0.1218	0.3676	0.1211
godine ²	-0.0010	-0.0010	-0.008	-0.0015	-0.0044	-0.0014
obrazovanje	0.0573	0.0573	0.441	0.0831	0.2528	0.0833
br.status	-0.0510	-0.0510	-0.318	-0.0614	-0.1701	-0.0568
$f(\hat{\beta}x)$	1.000		0.1886		0.3294	