

Dodatne vežbe: Metod instrumentalnih promenljivih (Slučajne objašnjavajuće promenljive)

Nastavnik: Aleksandra Nojković, nojkovic@gmail.com

Primer 1. Datoteka **Primer 22 (Mroz_data.wf)** sadrži podatke anketnog ispitivanja zaposlenih žena (n=753; preuzeto iz Wooldridge, 2012). U ekonomiji rada je uobičajeno da se zarade (logaritmovani podaci; **lwage**) ocenjuju kao funkcija broja godina školovanja (**educ**), broja godina radnog iskustava (**exper**) i kvadarta godina radnog iskustva (**expersq**). Primenom metoda običnih najmanjih kvadarta (ONK) dobijena je sledeća ocena jednačine zarada žena:

Tabela 1. Ocena jednačine zarada metodom ONK

Dependent Variable: LWAGE
Method: Least Squares
Date: 04/22/20 Time: 20:15
Sample (adjusted): 1 428
Included observations: 428 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.522041	0.198632	-2.628179	0.0089
EDUC	0.107490	0.014146	7.598333	0.0000
EXPER	0.041567	0.013175	3.154906	0.0017
EXPERTSQ	-0.000811	0.000393	-2.062833	0.0397

R-squared	0.156820	Mean dependent var	1.190173
Adjusted R-squared	0.150855	S.D. dependent var	0.723198
S.E. of regression	0.666420	Akaike info criterion	2.035509
Sum squared resid	188.3052	Schwarz criterion	2.073445
Log likelihood	-431.5990	Hannan-Quinn criter.	2.050492
F-statistic	26.28616	Durbin-Watson stat	1.960988
Prob(F-statistic)	0.000000		

Sve uključene objašnjavajuće promenljive imaju očekivan predznak i statistički značajno utiču na nivo zarada. Pri tome, **efekat dodatne godine školovanja iznosi oko 11 %**. U teoriji je poznato da se primenom metoda ONK na jednačinu zarada, dobija **precenjen efekat** obrazovanja na zarade.

Do ovog problema dolazi usled **stohastičke prirode objašnjavajuće promenljive**, odnosno njene korelisanosti sa slučajnom greškom modela (**problem endogenosti**). Naime, promenljivom godine školovanja u model uključujemo ukupne sposobnosti pojedinaca (ambicioznost, razne vidove inteligencije, neformalno obrazovanje itd.), koje nisu merljive. Iz tog razloga, ovu promenljivu možemo tretirati kao približnu promenljivu za ukupne sposobnosti (ili promenljivu koja ukupne sposobnosti meri sa greškom). Alternativno tumačenje problema endogenosti promenljive **educ**, bilo bi da ljudi sa većim sposobnostima ostvaruju veće zarade (izostavljena promenljiva je korelisana sa zavisnom), a pri tome su obično i obrazovaniji (škola im "bolje ide" pa se duže školuju). Dakle, izostavljena promenljiva (sposobnosti) je istovremeno korelisana i sa zavisnom (njen uticaj je sadržan u slučajnoj grešci modela) i sa objašnjavajućom promenljivom koja je uključena u model (**educ**). Navedeno dovodi do korelisanosti objašnjavajuće

promenljive **educ** i slučajne greške modela (narušena je jedna od pretpostavki ONK), a ovaj problem se u ekonometriji još naziva i **neopažena heterogenost** (*engl. unobserved heterogeneity*).

Kako su ocene dobijene metodom ONK u tom slučaju pristrasne i nekonzistentne, problem se prevazilazi primenom metoda instrumentalnih promenljivih (IP). **Ocene dobijene metodom IP su pristrasne** (može biti problem na malim uzorcima), ali **konzistentne**. Potrebno je endogenu promenljivu (**educ**, koja je korelisana sa slučajnom greškom modela) zameniti odgovarajućim instrumentom. Promenljiva (ili više njih) koja se koristi kao instrument mora biti egzogena (ne korelisana sa greškom modela) i korelisana sa promenljivom koju menja. Ova dva uslova koje instrumenti moraju ispunjavati (endogenost i relevantnost instrumenata) čine instrumente validnim.

U razmatranom primeru baziranom **na uporednim mikroekonomskim podacima** (anketa koja obuhvata pojedince/zaposlene žene), ideja **za potencijalne eksterne instrumente** (promenljive sadržane u anketi, koje nisu eksplicitno uključene u model) je da se godine školovanja majke (**matheduc**) i oca (**fatheduc**) koriste kao instrumenti za obrazovanje u modelu zarada. Očekujemo da je školovanje žena korelisano sa obrazovanjem roditelja, kao i da je istovremeno reč o egzogenim promenljivama koje ne utiču na nivo zarada.

Pre korišćenja ovih instrumenta u postupku ocenjivanja, možemo proveriti endogenu prirodu promenljive **educ** primenom Hausman-ovog testa. U jednostavnijoj verziji ovog testa (Wu-ov test), potrebno je rezidualne (**res1**) iz pomoćne regresije promenljive **educ** na instrumente (**matheduc** i **fatheduc**) uključiti u polazni model kako dodatnu promenljivu.

Tabela 1.2. Pomoćna regresija potencijalno endogene promenljive (**educ**) na instrumente

Dependent Variable: EDUC				
Method: Least Squares				
Date: 04/22/20 Time: 19:53				
Sample: 1 753				
Included observations: 753				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.975657	0.225668	39.77374	0.0000
MOTHEduc	0.183279	0.026217	6.990911	0.0000
FATHEduc	0.183418	0.024714	7.421766	0.0000
R-squared	0.244970	Mean dependent var	12.28685	
Adjusted R-squared	0.242956	S.D. dependent var	2.280246	
S.E. of regression	1.984002	Akaike info criterion	4.212085	
Sum squared resid	2952.198	Schwarz criterion	4.230508	
Log likelihood	-1582.850	Hannan-Quinn criter.	4.219182	
F-statistic	121.6689	Durbin-Watson stat	1.961485	
Prob(F-statistic)	0.000000			

Napomena: Ovaj rezultat se kasnije koristi za primenu Stock-Watson-ovog testa.

Polazni model proširujemo uključivanjem reziduala iz pomoćne regresije (Tabela 1.2.), što je prikazano u nastavku.

Tabela 1.3. Provera endogenosti promenljive **educ** – Wu-ov test (II verzija Hausman-ovog testa)

Dependent Variable: LWAGE
 Method: Least Squares
 Date: 04/22/20 Time: 19:57
 Sample (adjusted): 1 428
 Included observations: 428 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.026967	0.378398	0.071266	0.9432
EDUC	0.063404	0.029483	2.150561	0.0321
EXPER	0.041537	0.013146	3.159686	0.0017
EXPERSQ	-0.000841	0.000393	-2.140500	0.0329
RES1	0.056370	0.033097	1.703166	0.0893
R-squared	0.162563	Mean dependent var		1.190173
Adjusted R-squared	0.154644	S.D. dependent var		0.723198
S.E. of regression	0.664931	Akaike info criterion		2.033348
Sum squared resid	187.0226	Schwarz criterion		2.080768
Log likelihood	-430.1365	Hannan-Quinn criter.		2.052076
F-statistic	20.52819	Durbin-Watson stat		1.931082
Prob(F-statistic)	0.000000			

Statistička značajnost ove promenljive (**res1**), potvrda je potrebe za korišćenjem instrumenata (endogene prirode promenljive **educ**). Pored navedene verzije, provera endogenosti može se sprovesti i primenom Hausman-ovog testa koji se zasniva na sledećem poređenju ocena ONK i ocena IP (kao instrumenti se koriste obrazovanje amajke i oca). Navedenu verziju testa je u program Eviews moguće sprovesti nakon ocenjivanja jednačine **metodom IP**. Sam postupak sprovođenja metode IP objašnjen je u nastavku, nakon čega je potrebno slediti gotovu proceduru na putanji *View-IV Diagnostic & Tests-Regressor- Endogeneity Test*).

Tabela 1.4. Provera endogenosti promenljive **educ** – Hausman-ov test (I verzija)

Endogeneity Test
 Equation: EQ02_ONK
 Endogenous variables to treat as exogenous: EDUC
 Specification: LWAGE C EDUC EXPER EXPERSQ
 Instrument specification: C MOTHEDUC FATHEDUC EXPER EXPERSQ
 Null hypothesis: EDUC are exogenous

	Value	df	Probability
Difference in J-stats	2.780836	1	0.0954
J-statistic summary:			
	Value		
Restricted J-statistic	3.164752		
Unrestricted J-statistic	0.383916		

Razlika u ocenama dobijenim primenom dva metoda potvrda je endogene prirode promenljive **educ** (na nivu značajnosti 10%).

Vraćamo se ocenjivanju polazne jednačine zarada primenom metoda IP. Umesto originalne promenljive **educ**, koristi se njena ocena iz prvog stepena (obrazovanje žene ocenjeno kao funkcija obrazovanja roditelja, Tabela 1.2.). Otuda je metod IP u literaturi poznat i kao metod dvostepenih najmanjih kvadrata (2SNK; *engl. two-stage least squares*, TSLS).

Tabela 1.5. Ocena jednačine zarada metodom IP

Dependent Variable: LWAGE
 Method: Two-Stage Least Squares
 Date: 04/22/20 Time: 20:05
 Sample (adjusted): 1 428
 Included observations: 428 after adjustments
 Instrument specification: **MOTHEduc FATHEDUC EXPER EXPERSQ**
 Constant added to instrument list

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.048100	0.400328	0.120152	0.9044
EDUC	0.061397	0.031437	1.953024	0.0515
EXPER	0.044170	0.013432	3.288329	0.0011
EXPERSQ	-0.000899	0.000402	-2.237993	0.0257
R-squared	0.135708	Mean dependent var		1.190173
Adjusted R-squared	0.129593	S.D. dependent var		0.723198
S.E. of regression	0.674712	Sum squared resid		193.0200
F-statistic	8.140709	Durbin-Watson stat		1.945659
Prob(F-statistic)	0.000028	Second-Stage SSR		212.2096
J-statistic	0.374538	Instrument rank		5
Prob(J-statistic)	0.540541			

Napomena: U padajućem meniju rasoloživih metoda ocenjivanja programa Eviews odabrira se postupak TSLS. U listi instrumenata navode se promenljive mothereduc, fatheduc (kao instrumenti za educ), ali i exper i expersq, koje su egzogene i formalno ih možemo tretirati kao da su same sebi instrumenti.

Uočavamo da se primenom metoda IP dobija znatno manji efekat dodatne godine školovanja na zaradu, od oko 6%.

U nastavku, nameće se pitanje validnosti ovako odabranih instrumenata, koje razmatra dve različite teme. Prvo se proverava jačina (relevantnosti) instrumenata. U tom cilju analiziramo rezultate iz pomoćne regresije endogene promenljive na instrumente (Tabela 1.2.). Reč je o Stock-Watson-ovom testu, gde dobijena vrednost F- statistike iznad 10 ($F=121,67$) govori u prilog tome da su instrumenti relevantni. Drugo pitanje se tiče egzogenosti dobijenih instrumenata, o čemu potvdu dobijamo primenom Hansen-Sargan-ovg J test. U tom cilju, potrebno je rezidualne iz polazne jednačine zarada (Tabela 1.1.) oceniti u funkciji potencijalnih instrumenata, što je prikazano u nastavku:

Tabela 1.6. Pomoćna regresija za proveru egzogenosti instrumenata (J-test)

Dependent Variable: RES11
 Method: Least Squares
 Date: 04/22/20 Time: 20:33
 Sample (adjusted): 1 428
 Included observations: 428 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.155268	0.104537	1.485298	0.1382
MOTHEDUC	-0.013716	0.011662	-1.176125	0.2402
FATHEDUC	-0.002753	0.010949	-0.251416	0.8016
R-squared	0.005988	Mean dependent var		-5.41E-16
Adjusted R-squared	0.001310	S.D. dependent var		0.664075
S.E. of regression	0.663640	Akaike info criterion		2.024831
Sum squared resid	187.1777	Schwarz criterion		2.053283
Log likelihood	-430.3138	Hannan-Quinn criter.		2.036068
F-statistic	1.280025	Durbin-Watson stat		1.937153
Prob(F-statistic)	0.279100			

J- statistika se formira na sledeći način: $J = nR^2$ (R^2 iz pomoćne regresije prikazane u Tabeli 1.6.). U našem primeru: $J=428*0.005988=2.56 < \chi_1^2=3.841$ (broj stepeni slobode kritične vrednosti je razlika između broja instrumenata (ovde 5: c, mothereduc, fathereduc, exper i expersq) i broja promenljivih u polaznoj regresiji (ovde 4: c, educ, exper i expersq)). Dakle, na osnovu rezultata J-testa zaključujemo da su instrumenti egzogeni, te da ispunjavaju oba uslova validnosti.

Na nešto drugačijem postupku se zasniva izračunavanje vrednosti J-statistike u okviru gotove procedure sprovođenja metode IP u program Eviews. Naime, u donjem levom delu izlazne tabele postupka IP (Tabela 1.5.) nalazi se vrednost J-statistike i pridružena verovatnoća (0.37 (0.54)), na osnovu koje dolazimo do istovetnog zaključka o egzogenosti korišćenih instrumenata.

Dodatno – podrazumeva samostalan rad studenata: Sprovesti opisani postupak provere validnosti instrumenata i ponovo oceniti model metodom IP, koristeći dodatni instrument za promenljivu **educ** - obrazovanje supruga (**huseduc**), pored obrazovanja roditelja (**motheduc i fatheduc**). **Rešenje:** ocena dodatne godine školovanja je oko 8%.

Primer 2. U datoteci **Primer 23 (Greene.wf1)** nalaze se kvartalni makroekonomski podaci vremenskih serija o realnoj potrošnji (**realcons**) i dohotku (**realgdp**) u SAD za period 1950:1 – 1985:4 (preuzeto iz Greene, 2008). Ocena metodom ONK data je u nastavku:

Tabela 2.1. Ocena potrošnje u SAD metodom ONK

Dependent Variable: REALCONS
Method: Least Squares
Date: 04/22/20 Time: 22:05
Sample: 1950Q1 1985Q4
Included observations: 144

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-135.7486	10.08953	-13.45440	0.0000
REALGDP	0.685201	0.002790	245.6048	0.0000
R-squared	0.997651	Mean dependent var		2203.997
Adjusted R-squared	0.997635	S.D. dependent var		820.0615
S.E. of regression	39.88110	Akaike info criterion		10.22347
Sum squared resid	225851.2	Schwarz criterion		10.26472
Log likelihood	-734.0901	Hannan-Quinn criter.		10.24023
F-statistic	60321.70	Durbin-Watson stat		0.256249
Prob(F-statistic)	0.000000			

Ova regresija ocenjena je uz pretpostavku da dohodak utice na potrošnju i da je u pitanju jednosmerna veza. Međutim, postavlja se pitanje mogućeg uzajamnog delovanja ovih promenljivih, odnosno prisustvo problema simultane međuzavisnosti promenljivih. Formalno, reč je o narušenoj pretpostavci KLRM o nekorelisanosti objašnjavajuće promenljive i slučajne greške (pretpostavka o egzogenosti objašnjavajuće promenljive), čije su posledice pristrasne i nekonzistentne ocene ONK. Problem endogenosti proverava se primenom Hauman-ovog testa, a u nastavku razmatramo potencijalne instrumente koji bi se koristili u daljoj analizi.

Imajući u vidu da je ovde reč o makroekonomskim podacima vremenskih serija, kao prirodni instrumenti se nameću **pomaknute vrednosti** objašnjavajuće ili objašnjavajuće i zavisne promenljive (dohotka ili dohotka i potrošnje). U ovakvoj vrsti istraživanja, prilikom ocenjivanja **pojedinačnih jednačina**, nije jednostvano naći dobre eksterne instrumente (nešto drugačiji pristup sledi u delu o Sistemima simultanih jednačina). Međutim, realno je pretpostaviti da su pomaknute vrednosti promenljivih visoko korelisane sa tekućim (relevantni ili jaki instrumenti), a da istovremeno nisu korelisani sa slučajnom greškom u istom vremenskom periodu ili istoj docnji (ovo je ispunjeno ukoliko u modelu ne postoji autokorelacija).

U našem primeru, kao potencijalne instrumente za objašnjavajuću promenljivu možemo koristiti pomaknute vrednosti objašnjavajuće i zavisne promenljive. Pretpostavićemo da se kao instrumenti za tekući dohodak uvode dohotka i potrošnja iz prethodnog peiroda. Ocena metodom IP data je u sledećoj tabeli:

Tabela 2.2. Ocena potrošnje u SAD metodom IP

Dependent Variable: REALCONS
Method: Two-Stage Least Squares
Date: 04/22/20 Time: 22:51
Sample (adjusted): 1950Q2 1985Q4
Included observations: 143 after adjustments
Instrument specification: REALGDP(-1) REALCONS(-1)
Constant added to instrument list

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-140.2664	10.04621	-13.96212	0.0000
REALGDP	0.686334	0.002770	247.7508	0.0000
R-squared	0.997708	Mean dependent var		2212.004
Adjusted R-squared	0.997692	S.D. dependent var		817.2748
S.E. of regression	39.26371	Sum squared resid		217371.0
F-statistic	61380.45	Durbin-Watson stat		0.265831
Prob(F-statistic)	0.000000	Second-Stage SSR		220740.7
J-statistic	106.6033	Instrument rank		3
Prob(J-statistic)	0.000000			

U nastavku, proveravamo da li je dohodak endogena promenljiva, odnosno da li je uopšte potrebno umesto originalne promenljive koristiti instrumente. Navedeno možemo proveriti primenom Hausman-ovog testa.

Tabela 2.3. Provera endogenosti promenljive **realgdp** – Hausmann-ov test (I – verzija)

Endogeneity Test
Equation: EQ01
Endogenous variables to treat as exogenous: REALGDP
Specification: REALCONS C REALGDP
Instrument specification: C REALGDP(-1) REALCONS(-1)
Null hypothesis: REALGDP are exogenous

	Value	df	Probability
Difference in J-stats	13.28446	1	0.0003

J-statistic summary:
Value _____

Restricted J-statistic 119.8972
 Unrestricted J-statistic 106.6127

Rezultat Hausman-ovog testa potvrđuje endogenu prirodu promenljive dohotka, odnosno konzistentnost ocena dobijenih metodom IP. Alternativno, ukoliko potencijalno endogenu promenljivu dohotka regresiramo na instrumente, uključivanjem reziduala iz pomoćne regresije (**res1**) dolazimo do istog zaključka pomoću Wu-ove verzije testa endogenosti.

Tabela 2.4. Provera endogenosti promenljive **realgdp** – Hausmann-ov test (II – verzija)

Dependent Variable: REALCONS
 Method: Least Squares
 Date: 04/22/20 Time: 23:08
 Sample (adjusted): 1950Q2 1985Q4
 Included observations: 143 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-139.5242	9.763951	-14.28972	0.0000
REALGDP	0.686211	0.002693	254.8280	0.0000
RES1	-0.265701	0.087842	-3.024750	0.0030
R-squared	0.997849	Mean dependent var		2212.004
Adjusted R-squared	0.997818	S.D. dependent var		817.2748
S.E. of regression	38.17432	Akaike info criterion		10.14296
Sum squared resid	204019.1	Schwarz criterion		10.20512
Log likelihood	-722.2216	Hannan-Quinn criter.		10.16822
F-statistic	32472.57	Durbin-Watson stat		0.254753
Prob(F-statistic)	0.000000			

U ovom primeru ne razmatramo pitanje validnosti instrumenata (najverovatnije postoji problem nestacionarnosti podataka razmatranih vremenskih serija, što trenutno nije predmet našeg interesovanja!).