

## XALQARO NORDIK UNIVERSITETI

Iqtisodiyot va pedagogika fakulteti,  
Iqtisodiyot va biznesni boshqarish kafedrası

Fan o'qituvchisi: Sabirov Xasan Nusratovich

---

**Mavzu: Tasodifiy regressorlar va momentga asoslangan baholash**

### Reja:

1. Ikki bosqichli eng kichik kvadratlar.
2. Qo'shimcha o'zgaruvchilar bilan IV ni baholash.
3. Endogenlik bo'yicha HAUSMAN testi.
4. Cragg-Donald F-statistikasini hisoblash.

### 10.1. Ish haqi tenglamasining eng kichik kvadratlarini baholash

Principle of Econometrics 4-nashr adabiyotida 10-bobdagi misol Tomas Mrozning turmushga chiqqan ayollarning mehnat tajribasi haqidagi ma'lumotlardan foydalanadi. Jurnal faylini, *mroz.dta* ma'lumotlar faylini oching va ma'lumotlarni tekshiring.

```
log using chap10_wage,  
replace text use mroz,  
clear  
describe  
summarize
```

log-chiziqli ish haqi tenglamasini baholash uchun ishlaydigan ayollarning ish haqi ma'lumotlaridan foydalanamiz.

$$\ln(WAGE) = \beta_1 + \beta_2 EDUC + \beta_3 EXPER + \beta_4 EXPER^2 + e$$

Ma'lumotlar faylida ishlamaydigan ayollarni yo'q qilish uchun biz **drop** buyrug'idan foydalanamiz. Ish kuchi ishtirokini identifikatsiya qiluvchi o'zgaruvchi **lfp** ni tashkil etadi, agar ayol ish kuchida bo'lsa 1 ga, agar u bo'lmasa 0 ga teng bo'ladi. Keyin ish haqi, ta'lim va tajriba (exper) asosiy o'zgaruvchilarni umumlashtiring.

```
drop agar lfp==0  
summarize wage educ exper
```

```
. summarize wage educ exper
```

variable	Obs	Mean	Std. Dev.	Min	Max
wage	428	4.177682	3.310282	.1282	25
educ	428	12.65888	2.285376	5	17
exper	428	13.03738	8.055923	0	38

ln (wage) va tajriba kvadrati o'zgaruvchilarini yarating.

**gen lwage = ln(wage)**

**gen exper2 = exper^2**

Eng kichik kvadratlar yordamida regressiya modelini toping.

**reg lwage educ exper exper2**

Natija:

```
. reg lwage educ exper exper2
```

Source	SS	df	MS			
Model	35.0222967	3	11.6740989	Number of obs =	428	
Residual	188.305145	424	.444115908	F( 3, 424) =	26.29	
Total	223.327442	427	.523015086	Prob > F =	0.0000	
				R-squared =	0.1568	
				Adj R-squared =	0.1509	
				Root MSE =	.66642	

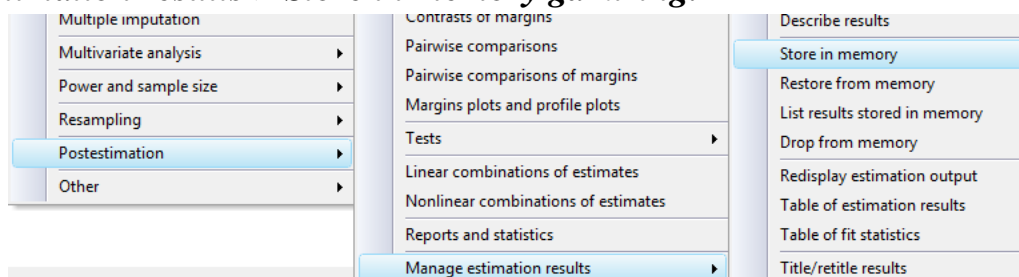
lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	.1074896	.0141465	7.60	0.000	.0796837	.1352956
exper	.0415665	.0131752	3.15	0.002	.0156697	.0674633
exper2	-.0008112	.0003932	-2.06	0.040	-.0015841	-.0000382
_cons	-.5220406	.1986321	-2.63	0.009	-.9124667	-.1316144

Keyingi maqsadlar uchun biz bu regressiya natijalarini *estimates store* buyrug'i yordamida saqlaymiz.

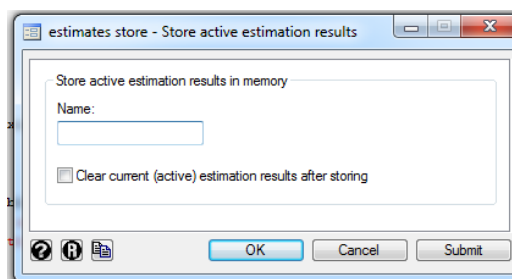
Ushbu *post-estimation* buyrug'i natijalarni Stata xotirasida saqlaydi va keyinroq foydalanish uchun chiqarilishi mumkin. Hisob-kitoblarni "eng kichik kvadratlar" uchun "ls" nomi ostida saqlang. Buyruq quyidagicha:

**estimates store ls**

Buyruqni unutilib qoldirgan bo'lsangiz *Statistics > Postestimation ni tanlang*. So`ngra *Manage estimation results > Store in memory ga kiring*.



Olingan muloqot oynasida saqlangan natijalar uchun "ls" nomini kiriting.



## 10.2. Ikki bosqichli eng kichik kvadratlar

Ushbu misolda biz *educ* ta'lim o'zgaruvchisini endogen deb hisoblashimiz mumkin, chunki u regressiya xatosi atamasidagi qobiliyat va boshqa omillar bilan bog'liq bo'lishi mumkin. Instrumental o'zgaruvchilarni baholash ikki bosqichli eng kichik kvadratlar sifatida ham tanilgan, chunki taxminlarni ikki bosqichda olish mumkin. Ta'lim uchun birinchi bosqich tenglamasini baholang, *educ*, shu jumladan o'ng tomonda tushuntirish o'zgaruvchilari sifatida kiritilgan ekzogen o'zgaruvchilar *exper* va *exper2* hamda modelga kiritilmagan *mothereduc* instrumental o'zgaruvchisidan foydalanamiz.

*reg educ exper exper2 mothereduc*

**Natijada:**

```
. reg educ exper exper2 mothereduc
```

Source	SS	df	MS			
Model	340.537834	3	113.512611	Number of obs =	428	
Residual	1889.65843	424	4.45674158	F( 3, 424) =	25.47	
Total	2230.19626	427	5.22294206	Prob > F =	0.0000	
				R-squared =	0.1527	
				Adj R-squared =	0.1467	
				Root MSE =	2.1111	

	educ	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
	exper	.0488615	.0416693	1.17	0.242	-.0330425	.1307655
	exper2	-.0012811	.0012449	-1.03	0.304	-.003728	.0011659
	mothereduc	.2676908	.0311298	8.60	0.000	.2065029	.3288787
	_cons	9.775103	.4238886	23.06	0.000	8.941918	10.60829

Birinchi bosqichdagi regressiyaning asosiy elementi shundan iboratki, *mothereduc* statistik ahamiyatga ega tushuntirish o'zgaruvchisi bo'lib, t-statistik > 3,3 yoki F-qiymati > 10. Ushbu bobning 10.6-bo'limida F-testning muhim qiymatlari haqida ko'proq ma'lumot beriladi. F-test qiymatlari yordamida olinadi

*test mothereduc*

```
. test mothereduc
( 1) mothereduc = 0
F( 1, 424) = 73.95
Prob > F = 0.0000
```

Birinchi bosqich tenglamasidan o'rnatilgan qiymatni oling va uni *ln(wage)* tenglamada *educ* o'rniga tushuntiruvchi o'zgaruvchi sifatida foydalaning.

*predict educ\_hat*

*reg lwage educ\_hat exper exper2*

Olingan koeffitsient baholari to'g'ri IV/2SLS hisob-kitoblaridir, ammo quyida keltirilgan standart xatolar, t-statistik ma'lumotlar, p-qiymatlari va intervalli taxminlar to'g'ri emas.

```
. reg lwage educat exper exper2
```

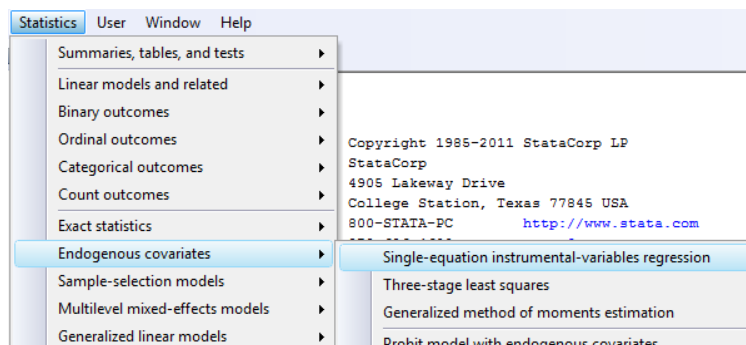
Source	SS	df	MS			
Model	10.181204	3	3.39373467	Number of obs =	428	
Residual	213.146238	424	.502703391	F( 3, 424) =	6.75	
Total	223.327442	427	.523015086	Prob > F =	0.0002	
				R-squared =	0.0456	
				Adj R-squared =	0.0388	
				Root MSE =	.70902	

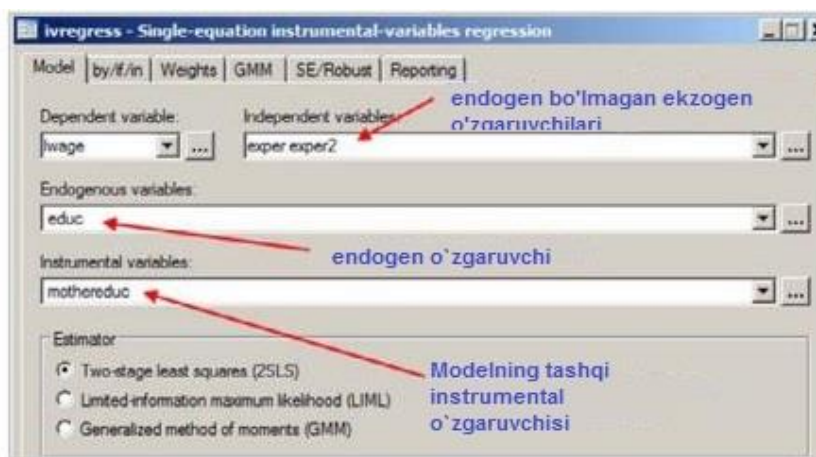
lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educat	.0492629	.0390562	1.26	0.208	-.0275049	.1260308
exper	.0448558	.0141644	3.17	0.002	.0170147	.072697
exper2	-.0009221	.000424	-2.17	0.030	-.0017554	-.0000887
_cons	.1981861	.4933427	0.40	0.688	-.7715157	1.167888

Instrumental o'zgaruvchan baholashni amalga oshirishda har doim shu maqsadda ishlab chiqilgan dasturiy ta'minotdan foydalaning. Stata 11 da bu buyruq *ivregress* hisoblanadi. Ushbu kuchli buyruqning imkoniyatlarini to'liq tavsiflash uchun *help ivregress* ga kiring. Muloqot oynasi yordamida ivregressni amalga oshirish uchun quyidagi ketma-ketlikni bajaring.

*Statistics>Endogenous covariates>Single-equation instrumental-variables regression*



Shu bilan bir qatorda Buyruqlar oynasiga *db ivregress* ni kiriting. Quyida ko'rsatilgandek to'ldiring va OK tugmasini bosing.



Natijada:

```
. ivregress 2sls lwage exper exper2 (educ = mothereduc)
Instrumental variables (2SLS) regression
Number of obs = 428
Wald chi2(3) = 22.25
Prob > chi2 = 0.0001
R-squared = 0.1231
Root MSE = .67642
```

	lwage	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
educ		.049263	.0372607	1.32	0.186	-.0237666	.1222925
exper		.0448558	.0135132	3.32	0.001	.0183704	.0713413
exper2		-.0009221	.0004045	-2.28	0.023	-.0017148	-.0001293
_cons		.1981861	.4706623	0.42	0.674	-.7242952	1.120667

```
Instrumented: educ
Instruments: exper exper2 mothereduc
```

Statadagi buyrug'i esa quyidagicha:

***ivregress 2sls lwage exper exper2 (educ=mothereduc)***

***ivregress*** dan keyin instrumental o'zgaruvchilarni baholash uchun zarur bo'lgan 2sls variantidan foydalaning. Bog'liq o'zgaruvchi ***lwage***, keyin esa tushuntirish o'zgaruvchilari keladi. Endogen o'zgaruvchilar qavslar ichida modeldan tashqaridagi instrumental o'zgaruvchilar ro'yxati bilan (***educ=mothereduc***) kabi joylashtiriladi. ***Endogen*** bo'lgan har qanday tushuntirish o'zgaruvchisi uchun bizda qavslar ichidagi bayonot mavjud, ya'ni

(***varlist2 = varlist\_iv***)

Bu yerda:

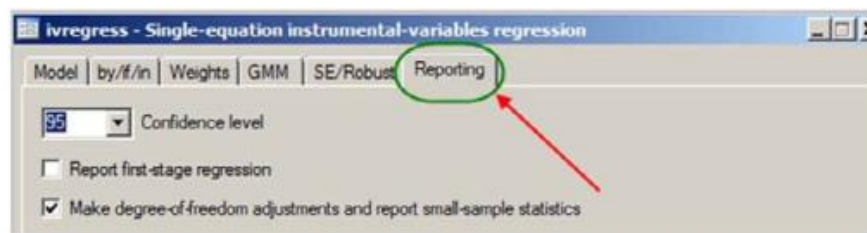
***varlist2***- o'ng tomondagi barcha endogen o'zgaruvchilar ro'yxati

***varlist\_iv***- modelda mavjud bo'lmagan barcha instrumental o'zgaruvchilar ro'yxati.

Endogen bo'lmagan tushuntirish o'zgaruvchilari qavs ichidagi ifodadan oldin yoki keyin sanab o'tiladi. Dialog oynasi yondashuvidan foydalanib, bu buyruq oxirida joylashtiriladi, lekin u erksiz o'zgaruvchidan keyin istalgan joyda paydo bo'lishi mumkin.

Koeffitsient baholari IV baholar bo'lib, standart xatolar to'g'ri hisoblab chiqilgan. Xabar qilingan test statistikasi "z" bilan belgilanadi, chunki IV baholovchilar asimptotik xususiyatlarga ega va katta namunalarda t-statistika standart normal taqsimotga yaqinlashadi va Z-statistika mos keladi.

ko'proq "konservativ" yondashuvni qo'llash uchun biz erkinlik darajalarini to'g'rilyadigan, biroz kattaroq standart xatolarni va shuning uchun biroz kattaroq p-qiymatlarini keltirib chiqaradigan t-statistikasini hisoblaymiz. Muloqot oynasida ***Reporting*** yorlig'idan foydalaning va erkinlik darajasini sozlash (***degrees-of-freedom adjustments***) variantini tanlang.



Ekvivalent, buyruqda biz ***small*** variantni qo'shamiz.

***ivregress 2sls lwage (educ=mothereduc) exper exper2, small***

E'tibor bering, bu buyruqda biz erksiz o'zgaruvchidan keyin ( $educ=mothereduc$ ) qo'yganimiz. Ta'kidlanganidek, bu joylashtirish tadqiqotchining ixtiyorida.

```
. ivregress 2sls lwage (educ=mothereduc) exper exper2, small
```

Instrumental variables (2SLS) regression

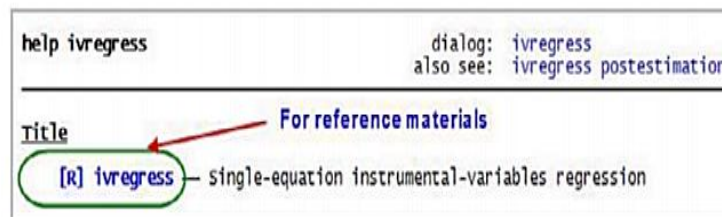
Source	SS	df	MS			
Model	27.4983827	3	9.16612758	Number of obs =	428	
Residual	195.829059	424	.461860988	F( 3, 424) =	7.35	
Total	223.327442	427	.523015086	Prob > F =	0.0001	
				R-squared =	0.1231	
				Adj R-squared =	0.1169	
				Root MSE =	.6796	

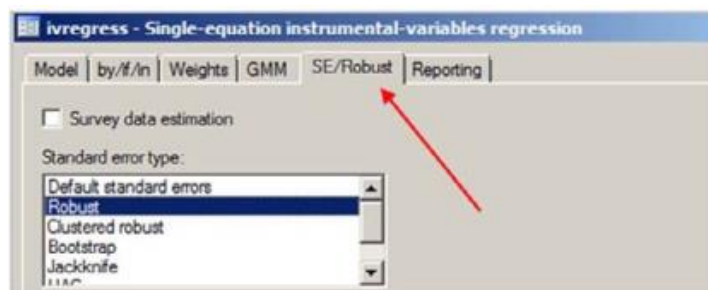
lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	.049263	.037436	1.32	0.189	-.0243204	.1228463
exper	.0448558	.0135768	3.30	0.001	.0181696	.0715421
exper2	-.0009221	.0004064	-2.27	0.024	-.0017208	-.0001233
_cons	.1981861	.4728772	0.42	0.675	-.7312895	1.127662

Instrumented: educ  
Instruments: exper exper2 mothereduc

E'tibor bering, endi t-statistik ma'lumotlar dispersiya jadvali tahlili bilan birga odatiy tarzda taqdim etiladi. Kvadratlarning tushuntirilgan yig'indisi uchun odatiy formulalar (regressiya tufayli) IV baholash bilan mos kelmaydi. Biroq, kvadratlar yig'indisi qo'shiladi, chunki Stata  $SS\_Model = SS\_Total - SS\_Residual$  ni belgilaydi. Bunday ma'lumotlarni to'liq Stata hujjatlarini o'qish orqali topish mumkin. Ushbu material ilg'or va matritsa algebrasidan foydalanadi.



Mroz ma'lumotlari kabi kesma ma'lumotlar uchun, biz, shuningdek, *ivregress* ga *vce(robust)* variantini qo'shish orqali Oq heteroskedastlikni tuzatishdan foydalangan holda, heteroskedastiklik haqida tashvishlanishimiz mumkin. Instrumental o'zgaruvchilar standart xatolar "robust" bo'lishi mumkin. SE/Robust yorlig'idagi ivregress dialog oynasida "Robust" ni tanlang.



Buyruqlar qatoriga kiriting

*ivregress 2sls lwage (educ=mothereduc) exper exper2, vce(robust) small*

```
. ivregress 2sls lwage (educ=mothereduc) exper exper2, vce(robust) small
```

Instrumental variables (2SLS) regression

Number of obs =	428
F( 3, 424) =	5.50
Prob > F =	0.0010
R-squared =	0.1231
Adj R-squared =	0.1169
Root MSE =	.6796

lwage	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
educ	.049263	.0380396	1.30	0.196	-.0255067	.1240326
exper	.0448558	.0156038	2.87	0.004	.0141853	.0755264
exper2	-.0009221	.0004319	-2.14	0.033	-.001771	-.0000732
_cons	.1981861	.4891462	0.41	0.686	-.7632673	1.159639

Instrumented: educ  
Instruments: exper exper2 mothereduc

Standart xatolar endi "Robust Std. Err" deb nomlanadi. E'tibor bering, ishonchli standart xatolar odatdagi standart xatolardan biroz kattaroqdir, bu odatiy natijadir. Umumiy F-testi ham mustahkam kovariatsiya matritsasiga asoslanadi.

### 10.3. Qo`shimcha o`zgaruvchilar bilan IV ni baholash

O`zgaruvchilar sonini ko'paytirish *ivregress* sintaksisini oddiy o'zgartirishni talab qiladi. Aytaylik, "*mothereduc*" dan tashqari, biz "*fathereduc*" dan o`zgaruvchi sifatida foydalanamiz. Bizning o`zgaruvchilarimiz ta'lim bilan yetarli darajada bog'liqligini tekshirish uchun birinchi bosqich tenglamasini baholaymiz. Modeldan tashqari o`zgaruvchilarning ahamiyatini sinab ko'ring. Bizda faqat bitta endogen tushuntirish o'zgaruvchisi borligi sababli, biz faqat bitta instrumental o'zgaruvchini talab qilamiz. Agar biz "*mothereduc*" va "*fathereduc*" ni alohida ko'rib chiqsak, ularning ahamiyatini tekshirish uchun t-testlaridan foydalanishimiz mumkin. Eslatib o'tamiz, shunchaki ahamiyat yetarli emas. T-testlar uchun biz 3,3 dan ortiq qiymatlarni qidiramiz.

*reg educ exper exper2 fathereduc*

Natija esa:

```
. reg educ exper exper2 fathereduc
```

Source	SS	df	MS			
Model	391.477157	3	130.492386	Number of obs =	428	
Residual	1838.7191	424	4.33660166	F( 3, 424) =	30.09	
Total	2230.19626	427	5.22294206	Prob > F =	0.0000	
				R-squared =	0.1755	
				Adj R-squared =	0.1697	
				Root MSE =	2.0825	

educ	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
exper	.0468243	.0411074	1.14	0.255	-.0339754	.127624
exper2	-.0011504	.0012286	-0.94	0.350	-.0035652	.0012645
fathereduc	.2705061	.0288786	9.37	0.000	.2137431	.3272691
_cons	9.887034	.3956078	24.99	0.000	9.109438	10.66463

Agar ikkala vositani birgalikda ishlatsak, ularning individual va umumiy ahamiyatini tekshiramiz:

*reg educ exper exper2 mothereduc fathereduc*  
*test mothereduc fathereduc*

```
. test mothereduc fathereduc
( 1) mothereduc = 0
( 2) fathereduc = 0
F( 2, 423) = 55.40
Prob > F = 0.0000
```

yoki bu testni heteroskedastikaga chidamli qilish uchun

```
reg educ exper exper2 mothereduc fathereduc, vce(robust)
test mothereduc fathereduc
```

Natija:

```
. test mothereduc fathereduc
( 1) mothereduc = 0
( 2) fathereduc = 0
F( 2, 423) = 49.53
Prob > F = 0.0000
```

Muhimlikni qo'shma testda muqobil gipoteza shundan iboratki, o'zgaruvchilardan kamida bittasi muhim va bitta endogen o'zgaruvchi mavjud bo'lganda biz faqat bittasini talab qilamiz. F-testi uchun adekvat asbob uchun minimal chegara qiymati taxminan 10 ni tashkil qiladi. Bizning asboblarning kuchli ekanligiga ishonch hosil qilsak, endi ikkita instrumental o'zgaruvchi yordamida IV baholashni amalga oshirishimiz mumkin.

```
ivregress 2sls lwage (educ=mothereduc fathereduc) exper exper2, small
estimates store iv
```

Natija quyidagicha:

```
. ivregress 2sls lwage (educ=mothereduc fathereduc) exper exper2 , small
```

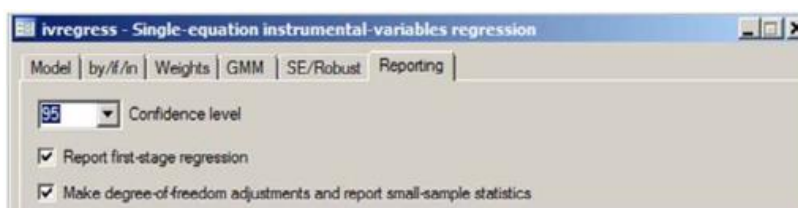
Instrumental variables (2SLS) regression

Source	SS	df	MS	Number of obs =	428
Model	30.3074259	3	10.1024753	F( 3, 424) =	8.14
Residual	193.020016	424	.455235886	Prob > F =	0.0000
Total	223.327442	427	.523015086	R-squared =	0.1357
				Adj R-squared =	0.1296
				Root MSE =	.67471

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
educ	.0613966	.0314367	1.95	0.051	-.0003945 .1231878
exper	.0441704	.0134325	3.29	0.001	-.0177679 .0705729
exper2	-.000899	.0004017	-2.24	0.026	-.0016885 -.0001094
_cons	.0481003	.4003281	0.12	0.904	-.7387745 .834975

```
Instrumented: educ
Instruments: exper exper2 mothereduc fathereduc
```

E'tibor bering, biz kelajakda foydalanish uchun instrumental o'zgaruvchilar taxminlarini saqladik. Hisoblashdan keyingi *estat firststage* buyrug'i birinchi bosqich F-statistik qiymatini ishlab chiqaradi. ivregress dialog oynasida Reporting ko'rinishida ushbu parametrlarni tanlang.





Hisoblashdan keyingi buyruq:

*estat firststage*

```
. estat firststage
First-stage regression summary statistics
```

Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	F(2,423)	Prob > F
educ	0.2115	0.2040	0.2076	55.4003	0.0000

```

Minimum eigenvalue statistic = 55.4003
Critical values
Ho: Instruments are weak
# of endogenous regressors: 1
# of excluded instruments: 2
2SLS relative bias
5% 10% 20% 30%
(not available)
2SLS Size of nominal 5% wald test 19.93 11.59 8.75 7.25
LIML Size of nominal 5% wald test 8.68 5.33 4.42 3.92

```

F-qiymatiga **Minimum eigenvalue statistic** nomi beriladi. Ushbu terminologiya va statistik ma'lumotlarning quyida keltirilgan kritik qiymatlarning foydaliligi ushbu bobning 10.6-bo'limda to'liq tushuntiriladi. IV bahosi yordamida heteroskedastikaga nisbatan mustahkam bo'lishi mumkin.

*ivregress 2sls lwage (educ=mothereduc fathereduc) exper exper2, vce(robust) small*

**Natija:**

```
. ivregress 2sls lwage (educ=mothereduc fathereduc) exper exper2 , vce(robust)
> small
Instrumental variables (2SLS) regression
Number of obs = 428
F( 3, 424) = 6.15
Prob > F = 0.0004
R-squared = 0.1357
Adj R-squared = 0.1296
Root MSE = .67471
```

lwage	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]
educ	.0613966	.0333386	1.84	0.066	-.0041329 .1269261
exper	.0441704	.0155464	2.84	0.005	.0136128 .074728
exper2	-.000899	.0004301	-2.09	0.037	-.0017443 -.0000536
_cons	.0481003	.4297977	0.11	0.911	-.7966992 .8928998

```

Instrumented: educ
Instruments: exper exper2 mothereduc fathereduc

```

```
. estat firststage
First-stage regression summary statistics
```

Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	Robust F(2,423)	Prob > F
educ	0.2115	0.2040	0.2076	49.5266	0.0000

Kuchli IV baholashdan so'ng, avtomatik birinchi bosqich buyrug'i, *estat firststage*, kritik qiymatlar haqida xabar bermaydi, chunki ular heteroskedastika ostida haqiqiy emas.

### 10.3.1. Qisman korrelyatsiyalarni tasvirlash

O'zgaruvchi kuchini muhokama qilishda eshutilgan iboralardan biri bu "qisman korrelyatsiyalar". Bu nima degani? Soddalashtirish uchun bizda bitta instrumental o'zgaruvchi mothereduc bo'lgan holatni ko'rib chiqaylik. *Ivregress* dan keyingi *estat firststage* natijalarining bir qismini ko'rib chiqamiz.

*ivregress 2sls lwage (educ=mothereduc) exper exper2, small estat firststage*

. estat firststage

First-stage regression summary statistics

Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	F(1,424)	Prob > F
educ	0.1527	0.1467	0.1485	73.9459	0.0000

Qiziqishning qisman korrelyatsiyasi bu *educ* endogen o'zgaruvchisi va *mothereduc* instrumental o'zgaruvchisi o'rtasidagi boshqa ekzogen o'zgaruvchilarning *exper* va *exper2* ta'sirini yo'qotgandan keyingi bog'liqlikdir.

O'zgaruvchi kuchini endogen o'zgaruvchi va bitta instrumental o'zgaruvchi o'rtasidagi qisman korrelyatsiya bilan o'lchash mumkin. *exper* va *exper2* ta'siri ushbu o'zgaruvchilar bo'yicha *educ* va *mothereduc* ni regressiya qilish va eng kichik kvadratlar qoldiqlarini hisoblash tomonidan olib tashlanadi. Qoldiqlar *exper* va *exper2* ta'sirini olib tashlaganidan keyin qolgan narsalarni o'z ichiga oladi.

*reg educ exper exper2*  
*predict v1, r*

*reg mothereduc exper exper2*  
*predict v2, r*

Bashoratli bayonotlardagi "r" varianti qoldiqlar uchun qisqa. Ushbu qoldiqlar orasidagi bog'liqlik:

*correlate v1 v2*

. correlate v1 v2  
(obs=428)

	v1	v2
v1	1.0000	
v2	0.3854	1.0000

Ushbu korrelyatsiya kvadrati birinchi navbatda *correlate* dan keyin saqlangan narsalarni ko'rish uchun *return list* dan foydalanib olinadi.

. return list

scalars:

r(N) = 428  
r(rho) = .3853595047039399

Keyin korrelyatsiya kvadrati bilan ko'rsatiladi

*di "partial correlation = "r(rho)^2*

. di "partial correlation = "r(rho)^2  
partial correlation = .14850195

*estat firststage* chiqishida bu *Partial R-sq* deb ataladi. Nima uchun u R-kvadrat deb ataladi? *v1* ni *v2* da regressiya qiling, chunki *v1* qoldiqlarining o'rtacha qiymati nolga teng.

*reg v1 v2, noconstant*

```
. reg v1 v2, noconstant
```

Source	SS	df	MS	
Model	329.557956	1	329.557956	Number of obs = 428
Residual	1889.65843	427	4.42542958	F( 1, 427) = 74.47
Total	2219.21639	428	5.18508502	Prob > F = 0.0000
				R-squared = 0.1485
				Adj R-squared = 0.1465
				Root MSE = 2.1037

v1	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
v2	.2676908	.0310202	8.63	0.000	.2067194 .3286622

E'tibor bering, konstantani o'z ichiga olmaydigan bu regressiyadan R-kvadrat 0,1485 ga to'g'ri keladi.

Korrelyatsiyalar va kovarians o'rtasidagi bog'liqlik yuqoridagi regressiya koeffitsientini tushunishimizga yordam beradi. v1 va v2 orasidagi odatiy kovariatsiya yordamida olinadi.

*correlate v1 v2, covariance*  
*return list*

```
. correlate v1 v2, covariance
(obs=428)
```

	v1	v2
v1	5.19723	
v2	2.88317	10.7705

```
. return list
```

```
scalars:
```

```

r(N) = 428
r(cov_12) = 2.883171450907586
r(Var_2) = 10.77052844798386
r(Var_1) = 5.197228071400259
```

Bu qiymatlardan regressiya koeffitsienti va korrelyatsiyani hisoblashimiz mumkin.

*di "partial LS coefficient = " r(cov\_12)/r(Var\_2)*

*di "partial correlation = " r(cov\_12)/sqrt(r(Var\_2)\*r(Var\_1))*

```
. di "partial LS coefficient = " r(cov_12)/r(Var_2)
partial LS coefficient = .26769081
```

```
. di "partial correlation = " r(cov_12)/sqrt(r(Var_2)*r(Var_1))
partial correlation = .3853595
```

#### 10.4. Endogenlik bo'yicha HAUSMAN testi

Bizning tushuntirish o'zgaruvchilarimiz orasida endogen regressor mavjudligini har doim ham bilmaymiz. Hausman protsedurasi tushuntirish o'zgaruvchining endogen yoki yo'qligini empirik tekshirish usulidir.

$y = \beta_1 + \beta_2 X + e$  regressiyada x ning e bilan bog'liqligini bilishni istaymiz.  $Z_1$  va  $Z_2$  x uchun instrumental o'zgaruvchilar bo'lsin. Xatolik bilan bog'liq bo'lishi mumkin bo'lgan

har bir o'zgaruvchi uchun minimal darajada bitta instrument talab qilinadi. Keyin quyidagi amallarni bajaring:

1.  $x = y_1 + \theta_1 z_1 + \theta_2 z_2 + v$  Modelni oddiy (ya'ni i.e- 2SLS emas) eng kichik kvadratlar bo'yicha baholang va  $\hat{v} = x - \hat{y}_1 - \theta_1 z_1 - \theta_2 z_2$  qoldiqlarni oling. Agar endogenlik tekshirilayotgan bir nechta tushuntiruvchi o'zgaruvchilar mavjud bo'lsa, har bir regressiyada mavjud bo'lgan barcha instrumental o'zgaruvchilardan foydalanib, har biri uchun ushbu baholashni takrorlang.

2. 1-qadamda hisoblangan qoldiqlarni izohli o'zgaruvchi sifatida dastlabki regressiyaga kiriting,  $y = \beta_1 + \beta_2 x + \delta \hat{v} + e$ . Ushbu sun'iy regressiya"ni eng kichik kvadratlar bo'yicha baholang, va ahamiyatlilik gipotezasi uchun odatdagi t-testdan foydalaning.

$H_0: \delta = 0$  (x va e o'rtasida hech qanday bog'liqlik yo'q)

$H_1: \delta \neq 0$  (x va e o'rtasidagi bog'liqlik mavjud)

*educ* endogen ekanligini va regressiya xatosi atamasi bilan bog'liqligini tekshirish uchun biz yuqorida tasvirlangan regressiyaga asoslangan Hausman testidan foydalanamiz. Testni amalga oshirish uchun *educ* uchun birinchi bosqich tenglamasini eng kichik kvadratlar, shu jumladan barcha ekzogen o'zgaruvchilar, o'ng tomonda *mothereduc* va *fathereduc* instrumental o'zgaruvchilari yordamida baholang. Qoldiqlarni saqlang:

*reg educ exper exper2 mothereduc fathereduc*  
*predict vhat, residuals*

Hisoblangan qoldiqlarni  $\ln(\text{wage})$  tenglamasiga qo'shimcha tushuntirish o'zgaruvchisi sifatida qo'shish va ahamiyatini standart t-test yordamida tekshiramiz.

*reg lwage educ exper exper2 vhat*

. reg lwage exper exper2 educ vhat

Source	SS	df	MS			
Model	36.2573098	4	9.06432745	Number of obs =	428	
Residual	187.070132	423	.442246175	F( 4, 423) =	20.50	
Total	223.327442	427	.523015086	Prob > F =	0.0000	
				R-squared =	0.1624	
				Adj R-squared =	0.1544	
				Root MSE =	.66502	

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
exper	.0441704	.0132394	3.34	0.001	.0181471	.0701937
exper2	-.000899	.0003959	-2.27	0.024	-.0016772	-.0001208
educ	.0613966	.0309849	1.98	0.048	.000493	.1223003
vhat	.0581666	.0348073	1.67	0.095	-.0102502	.1265834
_cons	.0481003	.3945753	0.12	0.903	-.7274721	.8236727

Agar heteroskedastiklikga ikkilangan bo'lsa, u holda *vce(robust)* parametrini qo'shish orqali mustahkam standart xatolarni hisoblaymiz.

*reg lwage educ exper exper2 vhat, vce(robust)*

```
. reg lwage exper exper2 educ vhat, vce(robust)
```

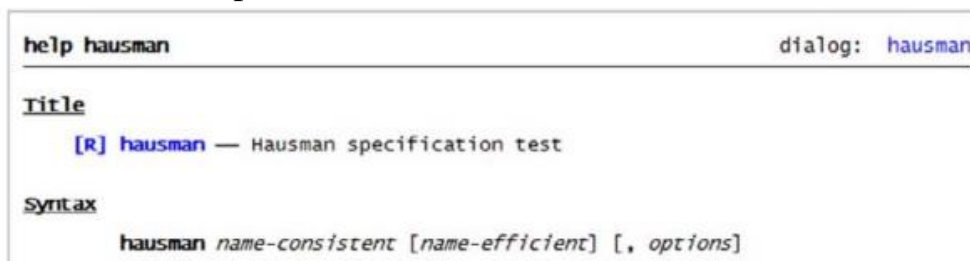
Linear regression

Number of obs = 428  
 F( 4, 423) = 21.52  
 Prob > F = 0.000  
 R-squared = 0.1624  
 Root MSE = .66502

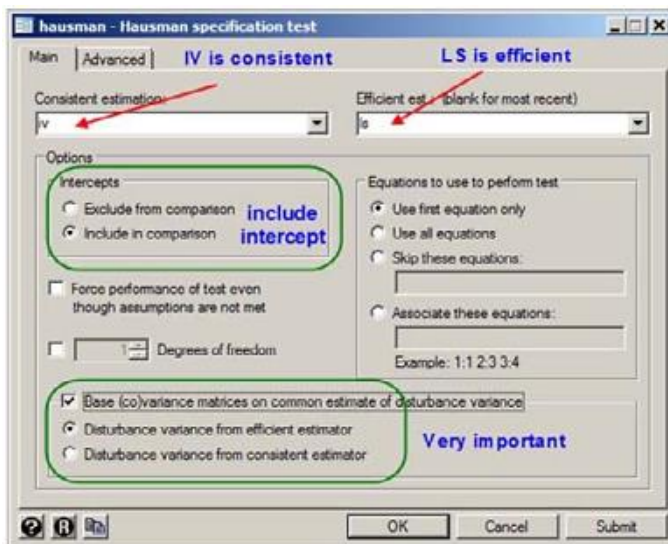
lwage	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
exper	.0441704	.0151219	2.92	0.004	.0144469	.0738939
exper2	-.000899	.0004152	-2.16	0.031	-.0017152	-.0000828
educ	.0613966	.0326667	1.88	0.061	-.0028127	.125606
vhat	.0581666	.0364135	1.60	0.111	-.0134073	.1297405
_cons	.0481003	.4221019	0.11	0.909	-.7815781	.8777787

Ushbu testlar *educ* ning taxminan 10% ahamiyatga ega bo'lgan darajada endogen ekanligini ko'rsatadi.

Biz aksariyat hollarda regressiyaga asoslangan sinovni afzal ko'ramiz. Stata ning "avtomatik" Hausman testini amalga oshirish uchun biz avval saqlangan instrumental o'zgaruvchilar (*Iv*) va eng kichik kvadratlar (*Is*) baholarini taqqoslaymiz. *Help hausman* yordamida biz sintaksisni topamiz.



Muloqot oynasi yondashuvidan foydalanish uchun *db hausman* kiriting yoki *Statistics > Postestimation > Tests > Hausman specification test* ketma-ketlikni bajaring.



Avtomatik test eng kichik kvadratlar hisoblagichi o'rtasidagi *contrast test* bo'lib, agar POE4 qo'llanmaning 10.1-bo'limida keltirilgan taxminlar bo'lsa, u eng yaxshi chiziqli holis va samarali bo'ladi. Agar regressor endogen bo'lsa instrume, u holda eng kichik kvadratlar hisoblagichi mos kelmaydi, lekin instrumental o'zgaruvchilar baholovchisi izchil bo'ladi. Bu contrast testi heteroskedastlik sharoitida haqiqiy emas, chunki test eng kichik kvadratlar hisoblagichining samarali bo'lishiga asoslanadi. Agar heteroskedastlik mavjud bo'lsa, eng

kichik kvadratlar samarali emas, chunki Gauss-Markov teoremasi bajarilmaydi. Bu heteroskedastik ma'lumotlar bilan qo'llanilishi mumkin bo'lgan regressiyaga asoslangan testning afzalliklaridan biridir. "*Consistent Estimation*" ochilgan ro'yxatida "*iv*" ni va "*Efficient Estimation*" ro'yxatida "*ls*" ni tanlang.

Biz ko'rsatadigan boshqa tanlovlar ushbu kontrast testi iloji boricha yaxshi ishlashi uchundir. taqqoslashda kesishuvni o'z ichiga oladi va eng muhimi, xato dispersiyasining umumiy bahosiga, eng kichik kvadratlar hisob-kitoblariga va qoldiqlarga asoslangan  $\sigma^2$  ni baholashga asoslanadi. "Endogenlik yo'q" degan nol gipoteza to'g'ri bo'lsa, bu baholovchi to'g'ri bo'ladi.

Stata *Result* oynasida siz tushunmaydigan va ushbu kitob doirasidan tashqarida bo'lgan juda ko'p so'zlar mavjud. Ular sizning nuqtai nazarangizdan asosiy natija shundaki, Hausman testi 1 daraja erkinlik bilan chi-kvadrat statistikasi hisoblanadi. Chi-kvadrat qiymati uning p-qiymati bilan birga berilgan. Testning ushbu versiyasiga asoslanib, biz *educ* 10% ahamiyatlilik darajasida regressiya xatosi bilan bog'liq degan xulosaga keldik.

$$\begin{aligned} \text{chi2}(1) &= (b-B)' [(V_b-V_B)^{-1}] (b-B) \\ &= 2.78 \\ \text{Prob>chi2} &= 0.0954 \end{aligned}$$

Ma'lum bo'lgan Stata buyrug'i:

***hausman iv ls, constant sigmamore***



Constant iborasi qo'shiladi, shunda contrast kesishish atamasini o'z ichiga oladi. Sigmamore iborasi Stata ni har ikkala baholovchi uchun xato dispersiyasini baholashda eng kichik kvadratlar qoldiqlaridan foydalanishga ta'siri uchun kiritilgan. Bu Stata regressiyaning o'ng tomonidagi endogen o'zgaruvchilar soni bo'lgan hausman testi uchun erkinlik darajalarining to'g'ri sonini hisoblashini ta'minlaydi.

### 10.5. Qo'shimcha o'zgaruvchilarning haqiqiylikini tekshirish

Ortiqcha yoki haddan tashqari aniqlangan asboblarning haqiqiyligi uchun *LM* testi *IV/2SLS* qoldiqlari bog'liq o'zgaruvchi sifatida va barcha instrumental hamda ekzogen o'zgaruvchilar tushuntirish o'zgaruvchilari sifatida sun'iy regressiyadan  $NR^2$  sifatida hisoblanadi. Shu maqsadda ushbu taxmindan *IV/2SLS* qoldiqni hisoblang.

***quietly ivregress 2sls lwage (educ=mothereduc fathereduc) exper exper2, small predict ehat, residuals***

Chunki biz buni allaqachon ko'rganmiz, biz *ivregress*dan oldin Stata buyrug'ini qo'shish orqali taxminiy natijani bostirishimiz mumkin. Endi barcha ekzogen o'zgaruvchilar va instrumental o'zgaruvchilar bo'yicha ehat regressini topamiz.

***reg ehat exper exper2 mothereduc fathereduc***

*Qaysi elementlar regressiyadan saqlanganligini eslash uchun ereturn list-dan foydalaning. Keyin NR<sup>2</sup> ni hisoblang.*

***ereturn list***

***scalar nr2 = e(N)\*e(r2)***

*Invchi2tail* yordamida chi-kvadrat (1) 95-protsentilni hisoblang, bu qo'shimcha o'zgaruvchilarning haqiqiylikini tekshirish uchun 0,05 kritik qiymat bo'ladi. Qo'shimcha instrumentlar adabiyotda "haddan tashqari identifikatsiyalash cheklovlari" deb ham ataladi. Bu yerda erkinlik darajalari soni 1 ga teng, chunki bitta ortiqcha o'zgaruvchi mavjud.

***scalar chic = invchi2tail(1,.05)***

*Chi2tail* yordamida testning p-qiymatini hisoblash.

***scalar pvalue = chi2tail(1,nr2)***

So'ngra natijalarni ifodalaymiz:

***di "R<sup>2</sup> from artificial regression = " e(r2)***

***di "NR<sup>2</sup> test of overidentifying restriction = " nr2***

***di "Chi-square critical value 1 df, .05 level = " chic***

***di "p value for overidentifying test 1 df, .05 level = " pvalue***

```
. di "R2 from artificial regression = " e(r2)  
R2 from artificial regression = .00088334
```

```
. di "NR2 test of overidentifying restriction = " nr2  
NR2 test of overidentifying restriction = .37807151
```

```
. di "Chi-square critical value 1 df, .05 level = " chic  
Chi-square critical value 1 df, .05 level = 3.8414588
```

```
. di "p value for overidentifying test 1 df, .05 level = " pvalue  
p value for overidentifying test 1 df, .05 level = .53863714
```

Stata-dan foydalanib, biz xuddi shu LM testini yaratishimiz mumkin. IV/2SLS hisob-kitoblarini yana *quietly* ishlatib, chunki biz bu natijalarni tekshirib chiqdik, keyin esa *estat overid* ni.

***quietly ivregress 2sls lwage (educ=mothereduc fathereduc) exper***

***exper2, small***

***estat overid***

Natijaviy *Sargan (score) testi* LM test, *NR<sup>2</sup>*, yuqoridan.

```
. quietly ivregress 2sls lwage (educ=mothereduc fathereduc) exper exper2, small
```

```
. estat overid
```

Tests of overidentifying restrictions:

```
Sargan (score) chi2(1) = .378071 (p = 0.5386)  
Basman chi2(1) = .373985 (p = 0.5408)
```

## 10.6. Kuchsiz o`zgaruvchilar uchun sinov

Kuchsiz o`zgaruvchilar uchun F-testi tenglamaning o'ng tomonida bir nechta endogen o'zgaruvchilarga ega modellar uchun mos emas. Masalan, bizda ikkita endogen o'zgaruvchi va ikkita instrumental o'zgaruvchi bor deylik. Instrumental o'zgaruvchilarni baholash uchun bizga ikkita tashqi instrumental o'zgaruvchilar kerak edi. Birinchi bosqich F-test yondashuvidan foydalanib, biz ikkita birinchi bosqich tenglamalarini baholaymiz va ikkita instrumental o'zgaruvchining umumiy ahamiyatini tekshiramiz.

Birinchi bosqich F-testlari muqobil gipoteza sifatida o`zgaruvchilardan kamida bittasi tegishli, kuchli vositadir. Aytaylik, bizning ikkita vositamizdan faqat bittasi endogen o'zgaruvchilar bilan bog'liq. Shuning uchun haqiqatda bizda bitta o`zgaruvchi bor. F-testi qo'shimcha nol gipotezani rad etadi, bu bizni ikkita o`zgaruvchimiz borligiga ishonishimizga olib keladi.

**Canonical correlations** yordamida tenglama bir nechta endogen o'zgaruvchilarga ega bo'lsa, zaif o`zgaruvchilarni aniqlash muammosini hal qilish mumkin. Kanonik korrelyatsiyalar - bu ikki o'zgaruvchi o'rtasidagi korrelyatsiya haqidagi odatiy tushunchaning umumlashtirilishi va ikkita o'zgaruvchi to'plami o'rtasidagi bog'lanishni tavsiflashga urinishdir. Kanonik korrelyatsiyalarni batafsil muhokama qilish bu tahlil doirasidan tashqarida. Ko'p o'lchovli tahlil bo'yicha kitobga murojaat qiling, ammo tushuntirishlar matritsa algebrasini o'z ichiga oladi.

N tanlama hajmini, B o'ng tomondagi endogen o'zgaruvchilar sonini, G tenglamaga kiritilgan ekzogen o'zgaruvchilar sonini (shu jumladan kesmani), L modelga kiritilmagan "tashqi" o`zgaruvchilar sonini bildirsin.

Agar birinchi o'zgaruvchilar to'plamida ikkita o'zgaruvchi va ikkinchi to'plamda ikkita o'zgaruvchi bo'lsa, u holda ikkita kanonik korrelyatsiya mavjud:  $r_1$  va  $r_2$ . Agar birinchi guruhda B o'zgaruvchilari ( $x_1=1, x_2, \dots, x_G$  ekzogen o'zgaruvchilarning ta'siri bilan endogen o'zgaruvchilar olib tashlangan) va ikkinchi guruhda  $L \geq B$  o'zgaruvchilari ( $x_1=1, x_2, \dots, x_G$  ta'siri bo'lgan o`zgaruvchilar guruhi) bo'lsa. olib tashlangan), B mumkin bo'lgan kanonik korrelyatsiya,  $r_1 \geq r_2 \geq \dots \geq r_B, r_B$  minimal kanonik korrelyatsiya. Zaif identifikatsiya uchun test, o`zgaruvchilar endogen regressorlar bilan korrelyatsiya qilinganda yuzaga keladigan vaziyat, lekin faqat zaif, **Cragg-Donald F-test** statistikasiga asoslanadi.

$$Cragg - Donald F = [(N - G - B)/L] \times [r_B^2 / (1 - r_B^2)]$$

Cragg-Donald statistikasi endogen o'zgaruvchilar soni  $B = 1$  bo'lsa, odatdagi zaif o`zgaruvchilar F-testiga kamayadi. Ushbu test statistikasi uchun kritik qiymatlar **Jeyms Stok** va **Motohiro Yogo** (2005) tomonidan jadvalga kiritilgan, shuning uchun biz o`zgaruvchilar kuchsizligi haqidagi nol gipotezani zaif o`zgaruvchilarning ikkita o'ziga xos oqibati bo'yicha ular bo'lmagan muqobilga qarshi sinab ko'rishimiz mumkin.

**Nisbiy moyillik:** kuchsiz o`zgaruvchilar mavjud bo'lganda, IV baholovchida noaniqlik miqdori katta bo'lishi mumkin. **Stock** va **Yogo** endogen o'zgaruvchilarning koeffitsientlarini baholashda moyillikni hisobga oladi. Ular eng kichik kvadratlar



smetatorining moyilligiga nisbatan IV baholovchining maksimal egilishini tekshiradilar. Stock va Yogo ta'limga qaytishni taxmin qilish misolini keltiradi. Agar tadqiqotchi eng kichik kvadratlar baholovchisi maksimal 10% ga tegadi, deb hisoblasa va agar nisbiy moyillik 0,1 bo'lsa, u holda IV baholovchining maksimal moyilligi 1% ni tashkil qiladi.

**Rad etish darajasi (sinov hajmi):** Endogen regressorlarga ega modelni baholashda endogen o'zgaruvchilarning koeffitsientlari haqidagi gipotezalarni sinab ko'rish ko'pincha qiziqish uyg'otadi.  $\alpha=0,05$  ahamiyatlilik darajasini tanlasak, takroriy namunalarda haqiqiy nol gipoteza vaqtning 5% rad etilishini kutamiz. Agar o'zgaruvchilar zaif bo'lsa, test o'lchami deb ham ataladigan nol gipotezaning haqiqiy rad etish darajasi kattaroq bo'lishi mumkin. Stock va Yogo ning ikkinchi mezon, agar biz  $\alpha =0,05$  ni tanlasak, haqiqiy nol gipotezani rad etishning maksimal darajasidir. Masalan, biz 5% darajasidagi test uchun maksimal 10% rad etish darajasini qabul qilishga tayyor bo'lishimiz mumkin, lekin 5% darajasidagi test uchun 20% rad etish darajasini qabul qilishga tayyor bo'lmasligimiz mumkin.

O'zgaruvchilar kuchsizligi haqidagi nol gipotezani sinab ko'rish uchun, ular bo'lmagan muqobilga qarshi, biz **Cragg-Donald** F-test statistikasini kritik qiymatga solishtiramiz. POE4 adabiyotida 10E.1-jadval va 10E.2-jadvalda keltirilgan qiymat Stata-ga kiritilgan. *ivregress*dan keyin *estat firststage* ishlatilsa, ushbu bobning 10.3-bo'limida ko'rsatilganidek, ushbu muhim qiymatlar xabar qilinadi. Qadamlar quyidagilardir:

1. Birinchi navbatda maksimal nisbiy moyillik yoki maksimal test o'lchami mezonini tanlang. Shuningdek, siz qabul qilishga tayyor bo'lgan maksimal nisbiy tarafkashlik yoki maksimal sinov hajmini tanlashingiz kerak.

2. **Cragg-Donald** F-test statistikasi jadvaldagi kritik qiymatdan kattaroq bo'lsa, o'zgaruvchilar kuchsizligi haqidagi nol gipotezani rad eting. Agar F-test statistikasi kritik qiymatdan katta bo'lmasa, o'zgaruvchilar kuchsizligi haqidagi nol gipotezani rad qilmang.

Mroz ma'lumotlaridan foydalanib, quyidagi HOURS ta'minot tenglama spetsifikatsiyasini ko'rib chiqing.

$$HOURS = \beta_1 + \beta_2 MTR + \beta_3 EDUC + \beta_4 KIDSL6 + \beta_5 NWIFEINC + e$$

O'zgaruvchi  $NWIFEINC = (FAMINC - WAGE \times HOURS)/1000$  - xotinning daromadidan boshqa manbalarga tegishli bo'lgan uy xo'jaliklarining daromadi. O'zgaruvchan  $MTR$ -bu xotin oldida turgan marginal soliq stavkasi, shu jumladan ijtimoiy sug'urta soliqlari. Bu tenglamada  $MTR$ ,  $KIDSL6$  va  $NWIFEINC$  da koeffitsient belgilari manfiy va  $EDUC$  koeffitsient noaniq belgiga ega.

Marginal soliq stavkasi  $MTR$  va ta'lim  $EDUC$ ni endogen deb hisoblang, shuning uchun  $B=2$ .  $L=2$  bo'lishi uchun ona va otaning ta'limini,  $MOTHEREDUC$  va  $FATHEREDUC$  ni o'zgaruvchilar sifatida ishlatning. Boshlanish uchun ma'lumotlarni o'chiring va kerakli o'zgaruvchilarni yarating.

*use mroz, clear*

*drop if lfp==0 gen lwage=ln(wage)*

*gen nwifeinc = (faminc-wage\*hours)/1000*

*gen exper2 = exper^2*

*MTR* va *EDUC* uchun birinchi bosqich tenglamalari va testlari yordamida olinadi.

*reg mtr mothereduc fathereduc kidsl6 nwifeinc*

*test mothereduc fathereduc*

*Natijada:*

```
. test mothereduc fathereduc
( 1) mothereduc = 0
( 2) fathereduc = 0
      F( 2, 423) = 8.14
      Prob > F = 0.0003
```

*reg educ mothereduc fathereduc kidsl6 nwifeinc*

*test mothereduc fathereduc*

*Natija:*

```
. test mothereduc fathereduc
( 1) mothereduc = 0
( 2) fathereduc = 0
      F( 2, 423) = 49.02
      Prob > F = 0.0000
```

Sinov natijalari shuni ko'rsatadiki, o'zgaruvchilar *EDUC* uchun yuqorida ko'rganimizdek kuchli birinchi bosqich kuchsiz o'zgaruvchi F-test statistikasi 49.2. *MTR* uchun bu ikki o'zgaruvchi kamroq kuchli. *FATHEREDUC* 5% darajasida muhim va birinchi bosqich kuchsiz o'zgaruvchi F-test statistikasi 8,14 bo'lib, p-qiymati 0,0003 ga teng. Bu  $F > 10$  qoidasini qoniqtirmasa ham, shunday "yaqin" va biz bu ikki o'zgaruvchi yetarli darajada kuchli degan xulosaga kelgan bo'lishimiz mumkin.

*Instrumental o'zgaruvchilarni baholash va birinchi bosqich statistikasi yordamida olinadi.*

*ivregress 2sls hours (mtr educ = mothereduc fathereduc) kidsl6*

*nwifeinc, small*

*estat firststage*

*Instrumental o'zgaruvchilar natijalari:*

```
. ivregress 2sls hours (mtr educ = mothereduc fathereduc) kidsl6 nwifeinc, small
> 1
```

Instrumental variables (2SLS) regression

Source	SS	df	MS			
Model	-1.0343e+09	4	-258577372	Number of obs =	428	
Residual	1.2916e+09	423	3053476.37	F( 4, 423) =	0.79	
Total	257311020	427	602601.92	Prob > F =	0.5329	
				R-squared =	.	
				Adj R-squared =	.	
				Root MSE =	1747.4	

hours	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
mtr	29709.47	90487.78	0.33	0.743	-148152.2	207571.2
educ	258.559	846.0142	0.31	0.760	-1404.356	1921.474
kidsl6	-1144.478	2510.194	-0.46	0.649	-6078.485	3789.529
nwifeinc	149.2325	470.5173	0.32	0.751	-775.6108	1074.076
_cons	-24491.6	79689.72	-0.31	0.759	-181128.8	132145.6

Instrumented: mtr educ  
Instruments: kidsl6 nwifeinc mothereduc fathereduc

birinchi bosqich natijalari esa:

```
. estat firststage
```

Shea's partial R-squared			
Variable	Shea's Partial R-sq.	Adj. Partial R-sq.	
mtr	0.0005	-0.0066	
educ	0.0024	-0.0046	

Minimum eigenvalue statistic = .100568

Critical values		# of endogenous regressors: 2			
Ho: Instruments are weak		# of excluded instruments: 2			
2SLS relative bias		5%	10%	20%	30%
		(not available)			
2SLS Size of nominal 5% wald test		10%	15%	20%	25%
LIML Size of nominal 5% wald test		7.03	4.58	3.95	3.63

Aytaylik, biz 5% test uchun maksimal 15% sinov hajmini qabul qilishga tayyormiz. Stata 2SLS dan foydalanganda kuchsiz o`zgaruvchi testi uchun kritik qiymat 4,58 ni tashkil qiladi. LIML uchun muhim qiymatlarga e'tibor bermang. Ushbu qiymatlar 11-bobda tushuntiriladi. Shuningdek, POE4 adabiyotidagi 10E.1-jadvalga qarang. Cragg-Donald F-testning statistik qiymati atigi 0,101 ni tashkil etadi, bu 15% maksimal sinov hajmi uchun (MTR va EDUC bo'yicha 5% sinov uchun) kritik qiymatdan 4,58 ancha past. Qulay birinchi bosqich F-test qiymatlariga qaramay, o`zgaruvchilar kuchsizligi haqidagi nol gipotezani rad eta olmaymiz. HOURS ta'minot tenglamasining taxminlari 10E.4-jadval, POE4, 439-betda keltirilgan Model (1) va Model (2)dagilardan keskin farq qiluvchi parametr baholarini ko'rsatadi va juda kichik t-statistik qiymatlar juda katta ekanligini bildiradi. Standart xatolar, kuchsiz o`zgaruvchilar mavjudligida instrumental o'zgaruvchilarni baholashning yana bir natijasi. Boshqa modellar ushbu bobning oxiridagi 10-bob do-faylda tasvirlangan.

### 10.7. Cragg-Donald F-statistikasini hisoblash

Cragg-Donald F-statistikasini hisoblashni tasvirlash uchun yuqoridagi 10.6-bo'limda tasvirlangan modeldan foydalaning, o`zgaruvchilar sifatida *Mothereduc* va *Fathereduc*. Qaytarilgan ro'yxat yordamida erkinlik darajalarini saqlang, N-G-B.

***ivregress 2sls hours (mtr educ = mothereduc fathereduc) kidsl6***

***nwifeinc, small***

***ereturn list***

***scalar df\_r = e(df\_r)***

*Kidsl6* va *nwifeinc* ning endogen o'zgaruvchilardan va o`zgaruvchilardan ta'sirini yuqoridagi 10.3.1-bo'limda ko'rsatilgan jarayondan foydalanib, qisman olib tashlang.

***reg mtr kidsl6 nwifeinc***

***predict mtr, r***

***reg educ kidsl6 nwifeinc***

***predict educr, r***

*reg mothereduc kidsl6 nwifeinc*

*predict mothereducr, r*

*reg fathereduc kidsl6 nwifeinc*

*predict fathereducr, r*

*Kanonik korrelyatsiyalar Stata tomonidan Canon buyrug'i yordamida hisoblanadi. Canon help ga qarang.*

```
help canon                                dialog: canon
                                           also see: canon postestimation

title
  [mv] canon — Canonical correlations

Syntax
  canon (varlist1) (varlist2) [if] [in] [weight] [, options]
```

*Muloqot oynasidan foydalanish uchun yuqoridagi havolani bosib yoki quyidagi ketma-ketlikni bajaring.*

*Statistics > Multivariate analysis > MANOVA, multivariate regression, and related > Canonical correlations*

*Biz boshqa ekzogen o'zgaruvchilarning ta'sirini olib tashlagan ikkita endogen o'zgaruvchidan va ikkita instrumentdan kanonik korrelyatsiyalarni topish buyrug'i:*

*canon (mtrr educr) (mothereducr fathereducr)*

*Boshqa natijalar mavjud bo'lsa-da, kanonik korrelyatsiyalarda ifodalaniadi.*

```
Canonical correlations:
0.4356  0.0218
```

*Saqlangan natijalar va minimal kanonik korrelyatsiya yordamida quyidagilar topiladi.*

*ereturn list*

*matrix r2=e(ccorr)*

*di "Calculation of Cragg-Donald statistic*

*" di "The canonical correlations "*

*matrix list r2*

*scalar mincc = r2[1,2]*

*di "The minimum canonical correlation = " mincc*

```
. di "The minimum canonical correlation = " mincc
The minimum canonical correlation = .02180077
```

*Cragg-Donald F-statistikasini hisoblash oddiy va ivragressdan keyingi estat firststage yordamida avtomatik hisoblash bilan bir xil bo'ladi. Maxrajdagi "2"-instrumental o'zgaruvchilar soni, L.*

*scalar cd = df\_r\*(mincc^2)/(2\*(1-mincc^2))*

*di "The Cragg-Donald F-statistic = " cd*

```
. di "The Cragg-Donald F-statistic = " cd
The Cragg-Donald F-statistic = .10056813
```

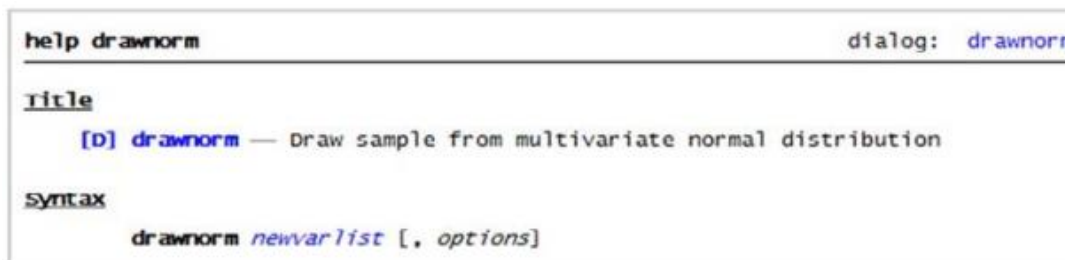
## 10.8. Simulyatsiya tajribasi

Principles of Econometrics 4-nashri darsligada, 10F.2-ilovada IV/2SLS baholovchisining namuna olish xususiyatlarini ko'rsatish uchun simulyatsiya tajribasi o'tkaziladi. Simulyatsiyada biz  $y=x+e$  ma'lumotlar ishlab chiqarish jarayonidan foydalanamiz, shuning uchun kesishma parametri 0 va qiyalik parametri 1 bo'ladi. Birinchi bosqich regressiyasi  $x=\pi z_1+\pi z_2+\pi z_3+v$ . E'tibor bering, bizda  $L=3$  instrumentlar mavjud, ularning har biri mustaqil standart normal  $N(0,1)$  ga ega.  $\pi$  parametri instrumental o'zgaruvchi kuchini nazorat qiladi. Agar  $\pi = 0$  bo'lsa, instrumental o'zgaruvchilar  $x$  bilan korrelyatsiya qilinmaydi va instrumental o'zgaruvchilarni baholash muvaffaqiyatsiz bo'ladi.  $\pi$  qanchalik katta bo'lsa, asboblar shunchalik kuchli bo'ladi. Nihoyat, biz  $x$  ning endogenligini boshqaradigan  $\rho$  korrelyatsiyasi bilan standart normal taqsimotlarga ega bo'lish uchun tasodifiy  $e$  va  $v$  xatolarini yaratamiz. Agar  $\rho=0$  bo'lsa,  $x$  endogen emas.  $\rho$  qanchalik katta bo'lsa, endogenlik shunchalik kuchli bo'ladi. Biz  $N=100$  o'lchamdagi 10 000 ta namunani yaratamiz va keyin bir nechta stsenariylar ostida eng kichik kvadratlar (LS) va IV/2SLSni sinab ko'ramiz. Biz  $\pi = 0,1$  (kuchsiz instrumental o'zgaruvchilar) va  $\pi = 0,5$  (kuchli instrumental o'zgaruvchilar) bo'lsin. Biz  $\rho = 0$  ( $x$  ekzogen) va  $\rho = 0,8$  ( $x$  yuqori endogen) ga ruxsat beramiz.

Simulyatsiya barcha xotirani tozalash va simulyatsiyani boshqaradigan global konstantalarni belgilash bilan boshlanadi.

```
clear all global numobs 100 // number of simulated sample observations  
global pi 0.1 // first stage parameter controls IV strength  
global rho 0.8 // rho controls endogeneity  
set seed 1234567 // random number seed  
set obs $numobs
```

Yuqoridagi to'plam buyruqni takrorlaganimizda biz bir xil pseudo-tasodifiy raqamlar ketma-ketligini va shuning uchun bir xil natijalarni olishimizni ta'minlaydi. Pseudo-tasodifiy raqamlar va tasodifiy qiymatlarini tushuntirish uchun *POE4*, B ilovasiga qarang. Simulyatsiya tajribasining asosiy komponenti  $e$  va  $v$  xato atamalari o'rtasidagi bog'liqlikdir. Korrelyatsiya qilingan tasodifiy sonlarni yaratish *drawnorm* Stata buyrug'i yordamida amalga oshiriladi. *Help drawnorm*-dan biz asosiy sintaksis va variantlarni topamiz.



```
help drawnorm                                     dialog: drawnorm  
  
title  
[D] drawnorm — Draw sample from multivariate normal distribution  
  
syntax  
drawnorm newvarlist [, options]
```

Variantlar orasida

$n(\#)$	<i># of observations to be generated</i>
$cov(\text{matrix/vector})$	<i>covariance matrix</i>

Ya'ni, biz kuzatuvlar sonini, kovariantlar va dispersiyalarni biz tanlagan har qanday narsa bo'lishi uchun belgilashimiz mumkin. Biz kovarian matritsiasidan foydalanamiz, ya'ni

$$\sum \begin{bmatrix} var(e) & cov(e, v) \\ cov(e, v) & var(v) \end{bmatrix} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

Ushbu matritsadan foydalanib belgilab oling:

*matrix sig = (1, \$rho | \$rho, 1) // corr(e,v)*

*drawnorm e v, n(\$numobs) corr(sig) // e & v values*

Instrumental o'zgaruvchilar uchun biz 3 ta mustaqil normal tasodifiy o'zgaruvchilardan foydalanamiz.

*gen z1 = rnormal()*

*gen z2 = rnormal()*

*gen z3 = rnormal()*

Yuqorida qayd etilgan ma'lumotlarni yaratish jarayonidan foydalanib, biz x va y ni yaratamiz. Xato shartlari \$rho korrelyatsiyasi bilan o'zaro bog'liq.

*generate x = \$pi\*z1 + \$pi\*z2 + \$pi\*z3 + v*

*generate y = x + e*

*correlate x e*

Biz topgan 100 ta tasodifiy qiymatlar namunaviy korrelyatsiyaga ega.

```
. correlate x e
(obs=100)
```

	x	e
x	1.0000	
e	0.7960	1.0000

Simulyatsiya qilingan ma'lumotlardan foydalangan holda birinchi bosqich regressiyasi instrumental o'zgaruvchilarning kuchsiz ekanligini ko'rsatadi, chunki biz \$pi = 0,1 ni o'rnatdik.

*reg x z1 z2 z3*

```
. reg x z1 z2 z3
```

Source	SS	df	MS			
Model	1.69113235	3	.563710782	Number of obs =	100	
Residual	115.728748	96	1.20550779	F( 3, 96) =	0.47	
Total	117.419881	99	1.1860594	Prob > F =	0.7056	
				R-squared =	0.0144	
				Adj R-squared =	-0.0164	
				Root MSE =	1.098	

x	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
z1	.1141584	.098892	1.15	0.251	-.0821408	.3104575
z2	.03231	.1214898	0.27	0.791	-.2088453	.2734652
z3	.0217676	.1132988	0.19	0.848	-.2031288	.2466641
_cons	.0174734	.1106386	0.16	0.875	-.2021425	.2370893

Koeffisientning quyidagi eng kichik kvadratlari haqiqiy qiymatdan uzoqdir, chunki x o'zgaruvchisi biz yaratgan kuchli endogendir.

*reg y x*

Natija:

```
. reg y x
```

Source	SS	df	MS			
Model	369.721081	1	369.721081	Number of obs =	100	
Residual	40.7364595	98	.415678159	F( 1, 98) =	889.44	
Total	410.45754	99	4.14603576	Prob > F =	0.0000	
				R-squared =	0.9008	
				Adj R-squared =	0.8997	
				Root MSE =	.64473	

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x	1.77446	.0594987	29.82	0.000	1.656387	1.892534
_cons	.1243801	.0644743	1.93	0.057	-.003567	.2523272

Bizda kuchsiz instrumental o'zgaruvchilar borligi sababli, IV/2SLS koeffisientning taxmini ham haqiqiy qiymatga yaqin emasligini ko'ramiz.

```
ivregress 2sls y (x=z1 z2 z3), small
```

Natija esa:

```
. ivregress 2sls y (x=z1 z2 z3), small
```

Instrumental variables (2SLS) regression

Source	SS	df	MS			
Model	353.009983	1	353.009983	Number of obs =	100	
Residual	57.4475569	98	.586199561	F( 1, 98) =	5.63	
Total	410.45754	99	4.14603576	Prob > F =	0.0196	
				R-squared =	0.8600	
				Adj R-squared =	0.8586	
				Root MSE =	.76564	

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x	1.397208	.5887541	2.37	0.020	.2288449	2.565572
_cons	.1268431	.0766601	1.65	0.101	-.0252863	.2789726

Instrumented: x  
Instruments: z1 z2 z3

Biz simulyatsiya uchun asos bo'lib xizmat qilish uchun dasturdan foydalanamiz. Dasturning birinchi qismida biz global makrolar tomonidan boshqariladigan bir xil ma'lumotlarni yaratish jarayoniga ega.

```
program ch10sim, rclass
```

```
version 11.1
```

```
drop _all
```

```
set obs $numobs
```

```
matrix sig = (1, $rho \ $rho, 1)
```

```
drawnorm e v, n($numobs) corr(sig)
```

```
gen z1 = rnormal()
```

```
gen z2 = rnormal()
```

```
gen z3 = rnormal()
```

```
* DGP generate x = $pi*z1 + $pi*z2 + $pi*z3 + v
```

```
generate y = x + e // structural equation
```

*\* first stage regression using all IV reg x z1 z2 z3*

Ushbu dasturni bajarish jarayonida qiymatlar baholashdan keyingi natijalardan "qaytariladi"

*return scalar rsq = e(r2) // first stage  $r^2$*

*return scalar F=e(F) // first stage F*

*predict vhat, r*

Koeffisient 1 ga teng bo'lgan haqiqiy nol gipotezaning 5% li testi uchun eng kichik kvadratlarning qiyalik bahosi (*bols*) va t-qiymati (*tols*) test natijasi (*rols*) kabi qaytariladi. Agar nol gipoteza rad etilsa, rollarning qiymati 1 ga, aks holda 0 ga teng. POE4 qo'llanmaning 10.1-bobidagi chiziqli regressiya modelining farazlariga ko'ra, test 5% vaqt oralig'ida haqiqiy nol gipotezani rad qilishi kerak.

*\* OLS*

*reg y x*

*return scalar bols = \_b[x]*

*return scalar seols = \_se[x]*

*return scalar tols = (\_b[x]-1)/\_se[x]*

*return scalar rols = abs(return(tols))>invttail(\$numobs-2,.025)*

Regressiyaga asoslangan Hausman testi statistikasi *haust*. 5% li test natijasi - haus, agar endogenlik yo'qligi haqidagi nol gipoteza rad etilsa, u 1 qiymatini oladi.

*\* Hausman*

*reg y x vhat*

*return scalar haust = \_b[vhat]/\_se[vhat]*

*return scalar haus = abs(return(haust))>invttail(\$numobs-3,.025)*

IV/2SLS qiyalik bahosi (*b2sls*), haqiqiy nol gipotezasini tekshirish uchun t-qiymati. Koeffisient 1 ga teng va test natijasi (*r2sls*) qaytariladi. Agar null bo'lsa, test natijasi *r2sls* si 1 ga teng, gipoteza rad etiladi va aks holda nolga teng bo'ladi. Agar instrumental o'zgaruvchilar tegishli va haqiqiy bo'lsa, test haqiqiy nol gipotezani 5% hollarda rad etishi kerak.

*\* 2sls*

*ivregress 2sls y (x=z1 z2 z3), small*

*return scalar b2sls = \_b[x]*

*return scalar se2sls = \_se[x]*

*return scalar t2sls = (\_b[x]-1)/\_se[x]*

*return scalar r2sls = abs(return(t2sls))>invttail(\$numobs-2,.025)*

Dastur yakuni bilan xulosa qilinadi. Simulyatsiya aslida Stata simulyatsiyasi buyrug'i yordamida amalga oshiriladi.

*simulate rsqf = r(rsq) Fr=r(F) bolsr=r(bols) seolsr=r(seols) ///*

*rolsr=r(rols) b2slsr=r(b2sls) se2slsr=r(se2sls) ///*



```
t2slsr=r(t2sls) r2slsr=r(r2sls) hausr=r(haus), ///
reps(10000) nodots nolegend seed(1234567): ch10sim
```

Qaytarilgan qiymatlarga tayinlangan o'zgaruvchilar nomlari *rsqf*, *fr*, *bolsr* va boshqalar. *ch10sim* dasturining 10 000 ta eksperimental nusxalari mavjud. Simulyatsiyadan so'ng biz natijalarni saqlash uchun global parametr qiymatlaridan foydalanamiz.

```
di " Simulation parameters"
```

```
di " rho " $rho
```

```
di " N " $numobs
```

```
di " pi " $pi
```

```
. di " Simulation parameters"
Simulation parameters

. di " rho = " $rho
rho = .8

. di " N = " $numobs
N = 100

. di " pi = " $pi
pi = .1
```

Bizni dastlabki bosqichdagi o'rtacha F-qiymati qiziqtiradi, bu instrumental o'zgaruvchi kuchining ko'rsatkichidir.

```
di " average first stage F"
```

```
mean Fr
```

Mean estimation		Number of obs = 10000		
	Mean	Std. Err.	[95% Conf. Interval]	
rsqf	.0576316	.000391	.0568652	.0583981

Har bir baholovchi uchun Stata *tabstat* buyrug'i yordamida uning o'rtacha qiymati va standart og'ishini hisoblaymiz. Bu statistik ma'lumotlarni chiroyli jadvalda hisobot berishga ruxsat berganidek, umumlashtirish uchun qulay alternativadir. Muloqot oynasida quyidagi ketma-ketlikni amalga oshiring.

*Statistics > Summaries, tables, and tests > Tables > Table of summary statistics (tabstat)*

Sintaksis va muhim variantlar *help tabstat* buyrug'i orqali to'liq ifodalanadi.

```
help tabstat dialog: tabstat

Title
[R] tabstat — Display table of summary statistics

Syntax
tabstat varlist [if] [in] [weight] [, options]

options description
Main
by(varname) group statistics by variable
statistics(statname [...]) report specified statistics
```

Agar baholovchi xolis bo'lsa, o'rtacha taxminiy qiymat haqiqiy qiymatga yaqin bo'lishi kerak, 1. Rad etish darajasi o'zgaruvchisining o'rtacha qiymati (*rols*) haqiqiy nol gipotezaning haqiqiy rad etish tezligini ko'rsatadi. *Mseols* ning o'rtacha qiymati baholashning "o'rtacha kvadrat xatosi" hisoblanadi. Bu empirik analogidir.

$$MSE = E[(b_2 - \beta_2)^2] = var(b_2) + [bias(b_2)]^2$$

O'rtacha kvadrat xatosi o'rtacha hisobda haqiqiy parametr qiymatiga qanchalik yaqin ekanligi haqidagi savolga javob beradi. Va nihoyat, biz Hausman testining o'rtacha rad etish darajasini ko'rib chiqamiz.

*di " OLS "*

$$gen\ mseols = (bolsr-1)^2$$

*tabstat bolsr seolsr rolsr mseols hausr, stat(mean sd)*

```
. tabstat bolsr seolsr rolsr mseols hausr, stat(mean sd)
```

stats	bolsr	seolsr	rolsr	mseols	hausr
mean	1.776194	.0612667	1	.6061978	.3841
sd	.0609983	.0061937	0	.0947578	.4864061

Eng kichik kvadratlar taxminlarining o'rtacha qiymati 1,776 ni tashkil etadi, bu 1 ning haqiqiy qiymatiga yaqin emas. Haqiqiy nol gipotezaning t-testi nominal 5% stavka o'rniga 100% vaqtni rad etadi. Hausman regressiyasiga asoslangan test endogenlik yo'qligi haqidagi (noto'g'ri) nol gipotezani 38% hollarda rad etadi.

*di " 2sls "*

$$gen\ mse2sls = (b2slsr-1)^2$$

*tabstat b2slsr se2slsr r2slsr mse2sls, stat(mean sd)*

IV/2SLS hisoblagichi uchun o'xshash qiymatlar:

```
. tabstat b2slsr se2slsr r2slsr mse2sls, stat(mean sd)
```

stats	b2slsr	se2slsr	r2slsr	mse2sls
mean	1.331058	.8850129	.2886	1.008766
sd	.9482915	35.54677	.4531342	58.61448

Agar biz  $\pi$  ni 0,5 ga o'rnatish orqali instrumental o'zgaruvchi uzunligini oshirsak, eng kichik kvadratlarni baholashda biz hech qanday yaxshilanishni topmaymiz.

```
. tabstat bolsr seolsr rolsr mseols hausr, stat(mean sd)
```

stats	bolsr	seolsr	rolsr	mseols	hausr
mean	1.456824	.0608191	1	.2124129	1
sd	.0610322	.0061183	0	.0560244	0

Biroq, IV/2SLS hisoblagichi uchun biz kuchliroq instrumental o'zgaruvchilar tufayli katta yaxshilanishlarni topamiz.

```
. tabstat b2s1sr se2s1sr r2s1sr mse2s1s, stat(mean sd)
```

stats	b2s1sr	se2s1sr	r2s1sr	mse2s1s
mean	1.011116	.11695	.0636	.0139068
sd	.1174081	.0274133	.2440512	.0227001

### **X bob mavzularini mustahkamlash uchun savollar**

1. Regressiya natijalarini saqlash qaysi buyruq orqali amalga oshiriladi?
2. Agar *estimates store ls buyrug'i* unitilsa qanday amal bajarilishi lozim?
3. *Ivregress* buyruqning imkoniyatlarini to'liq tavsiflash uchun qayday buyruqga murojaat etish kerak?
4. Standart xatolar Stata oynasida qanday tasvirlanadi?
5. Qo'shimcha o'zgaruvchilar bilan ishlashda individual va umumiy ahamiyatini tekshirish qanday amalga oshiriladi?
6. F-statistik qiymatni qaysi amaldan foydalanib ifoda etish mumkin?
7. F-qiymatiga qanday nim beriladi?
8. O'zgaruvchi kuchini qanday o'lchash mumkin?
9. HAUSMAN testi qandau usul?
10. Mustahkam standart xatolari qanday hisoblanadi?
11. *ereturn list buyrug'i* nimaga xizmat qiladi?
12. Canonical correlations nima uchun qo'llaniladi?
13. Cragg-Donald F-statistikasini hisoblash qanday amalga oshiriladi?