

XALQARO NORDIK UNIVERSITETI

Iqtisodiyot va pedagogika fakulteti, Iqtisodiyot va biznesni boshqarish kafedrası

Fan o'qituvchisi: Sabirov Xasan Nusratovich

Mavzu: Geteroskidastlik

Reja:

1. Geteroskidastlik ahamiyati
2. Geteroskedastlikni aniqlash
3. Geteroskedastik-ishlab chiqish standart xatolar
4. Umumiyashtirilgan eng kichik kvadratlar usuli

8.1. Geteroskidastlik ahamiyati

3 va 4-boblardagi oddiy chiziqli regressiya modellari va ko'p regressiya modeli 5 va 6-boblarni bir necha jihatdan umumlashtirish mumkin. Masalan, ushbu modellarning tasodifiy o'zgaruvchilari (y_i yoki e_i) kuzatishlar bo'ylab bir xil o'zgaruvchanlikka ega ekanligiga kafolat bermaydi. Ya'ni, ba'zi kuzatuvlar boshqalardan ko'ra kattaroq dispersiyaga ega bo'lishi mumkin. Bu tasvirlangan holat **heteroskedasticity** deb ataladi.

Quyida oddiy chiziqli regressiya modeli ko'rsatilgan:

$$y_i = \beta_1 + \beta_2 x_i + e_i$$

bu yerda y_i sizning bog'liq o'zgaruvchingiz, x_i i^{th} mustaqil o'zgaruvchi bo'yicha kuzatish, e_i tasodifiy xato va β_1 hamda β_2 siz hisoblamoqchi bo'lgan parametrlardir. Xatolar har qanday parametr uchun qiymati x_i nolga teng va ular bir-biri bilan bog'liq emas.

Ushbu modeldagi farq - xatolarning soni endi u tegishli kuzatuvga bog'liq. Shunday qilib, xato farqi hozir kuzatish pastki belgisi bilan havola qilingan,

$$i = 1, 2, \dots, N.$$

Xatolar haqidagi taxminlar umumlashtiriladi.

$$E(e_i) = 0 \quad \text{var}(e_i) = \sigma_i^2 \quad \text{cov}(e_i, e_j) = 0$$

Ushbu bobda Geteroskidastlikni aniqlashning bir necha usullari ko'rib chiqiladi. Bundan tashqari, statistik jihatdan chiziqli regressiya modelining parametrlarini baholash va

gipotezalarni tekshirish usullari ma'lumotlari Geteroskedastlik bo'lgan parametrlar o'rganiladi.

Hatto eng kichik kvadratlar hisoblagichida xatolar geteroskedastik bo'lsa ham chiziqli modelni baholash uchun ishlatilishi mumkin.

MR3, $var(y_i) = var(e_i) = \sigma^2$ lar buzilgan taqdirda ham xolis va izchildir. Geteroskedastik modelda eng kichik kvadratlardan foydalanish bilan bog'liq muammoning odatdagi o'lchovi aniqlikka (hisoblangan dispersiya-kovariatsiya matritsasi) mos kelmaydi. Bu muammoni hal qilishning bir nechta usullari mavjud.

Birinchisi, kovariatsiyasini baholovchisi bilan birga uning eng kichik kvadratlardan foydalanishda izchil geteroskedastik xatolar bor yoki yo'qligini aniqlashdan iborat. Bu Stata foydalanadigan kovariatsiyaning **robust** baholovchi deb ataladi. Bu quyida 8.2-bo'limda muhokama qilingan. Yana biri bu geteroskedastik va eng kichik kvadratlardan foydalanib modellashtirishdir. Ushbu usul 8.3-bo'limda ko'rib chiqilgan.

Birinchi misol, oziq-ovqat xarajatlari ma'lumotlari eng kichik kvadratlar foydalangan holda modelni baholash uchun ishlatiladi. Ishchi katalogingizni "food.dta" ma'lumotlar to'plamini o'z ichiga olgan katalogga o'zgartiring va ma'lumotlar to'plamini yuklang.

```
cd c:\data\poe4stata
```

```
use food, clear
```

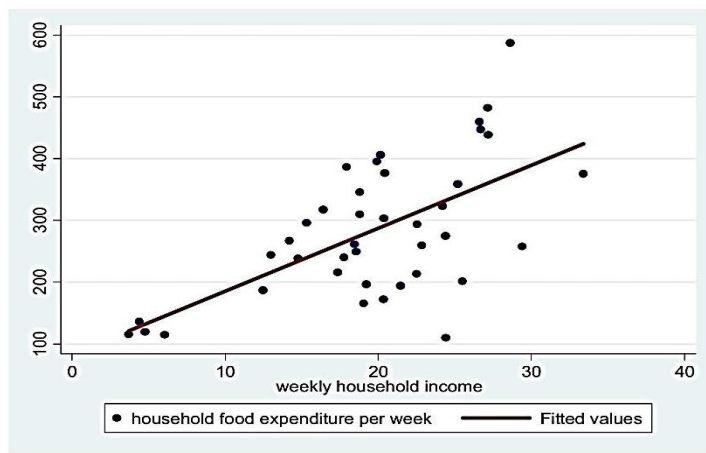
Ishni eng kichik kvadratlar yordamida oziq-ovqat xarajatlari modelini baholashdan boshlang.

```
regress food_exp income
```

Ma'lumotlar va taxminiy regressiya chizig'ini chizish uchun foydali bo'lishi mumkin. Stata da

```
graph twoway (scatter food_exp income) (lfit food_exp income)
```

bir xil grafikda ikkita tarqoq chizma hosil qiladi. Har bir syujetni yaratish buyruqlari ikkita qavs to'plamida joylashgan. Birinchi grafik **food_exp** ning daromadga qarshi ikki **twoway** tarqalish grafigidir. Ikkinchisi **food_exp** ning **income** ga oddiy regressiyasini baholash uchun **lfit** dan foydalaning. Regressiya chizig'i va **food_exp** ning haqiqiy qiymatlari orasidagi og'ishlar eng kichik kvadrat qoldiqlari hisoblanadi.



Grafik *Principle of Econometrics* 4-nashr (POE4) adabiyotidagi 8.2-grafikga juda o'xshaydi.

8.2. Geteroskedastiklikni aniqlash

Modeldagi geteroskedastiklikni aniqlashning bir qancha grafik va statistik usullari mavjud bo'lib, ushbu bo'limda ulardan bir nechtasi muhokama qilamiz.

8.2.1. Qoldiq maydonlari

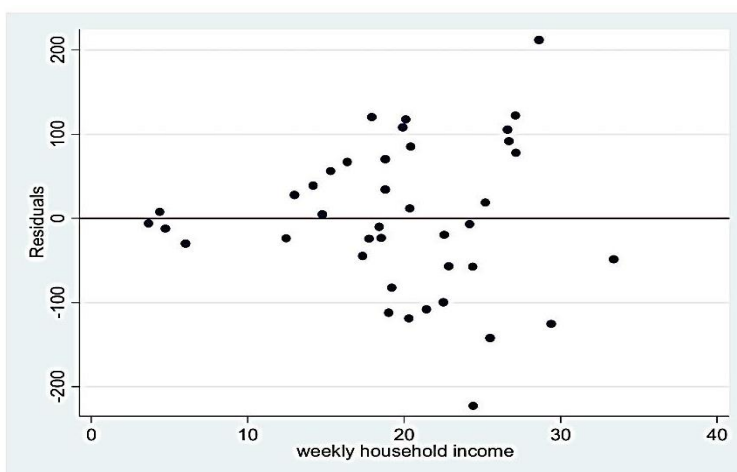
Xatolarning geteroskedastikligini aniqlash usullaridan biri - ularni mustaqil o'zgaruvchining tartiblangan qiymatiga nisbatan chizish. Oldingi bo'limda bir nechta misollar keltirilgan edi. O'zaro bog'liqlikni hisoblashning yana bir usuli - modelni baholash, qoldiqlarni aniqlashdir. Ularni tuzish uchun **graph twoway** dan foydalanish mumkin:

```
regress food_exp income
```

```
predict ehat, res
```

```
graph twoway scatter ehat income, yline(0)
```

quyidagicha grafik hosil bo'ladi:

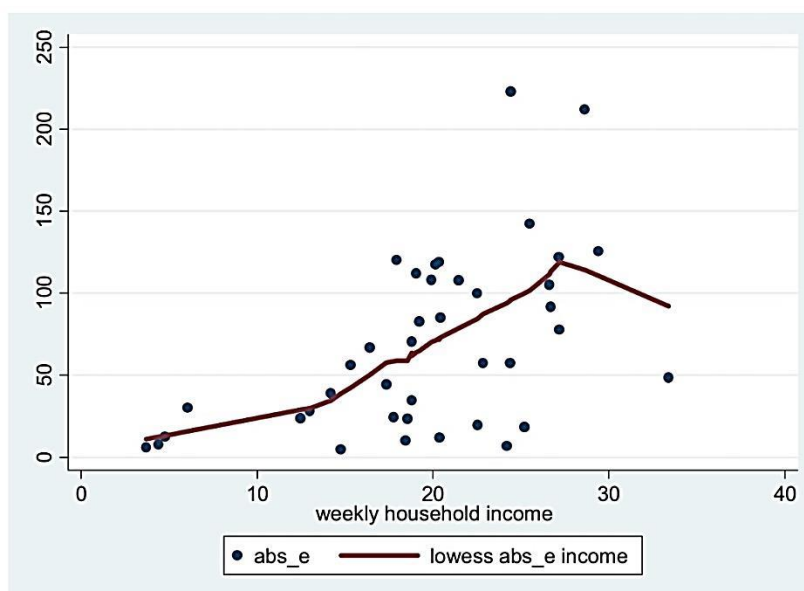


Qizil gorizontaal chiziqni nolga tenglashtirish uchun buyruqqa **yline(0)** birikma qo'shildi. Grafikdan ko'rinib turibdiki, daromadning kattaroq qiymatlari uchun qoldiqlar kattaroqdir. Buni quyidagi testlardan biri yoki bir nechtasi yordamida statistik tasdiqlash mumkin.

Qoldiqlar kattaligi va mustaqil o'zgaruvchi o'rtasidagi bog'liqlikni ko'rsatadigan yana bir grafik usul quyida ko'rsatilgan:

```
generate abs_e = abs(ehat)
```

```
twoway (scatter abs_e income) (lowess abs_e income, lw(thick))
```



Ushbu grafikni yaratish uchun ikkita amal bajarildi.

Birinchidan, eng kichik kvadratlar qoldiqlarining mutlaq qiymati **abs_e** deb nomlangan yangi o'zgaruvchiga yig'iladi. Keyin ular **income** ga scatter sifatida va **lowess** deb ataladigan jarayon orqali hisoblangan mahalliy qiymatli, tekislangan *scatterplot* sifatida chiziladi. Stata ma'lumotlar to'plamidan biz **lowess** ning asosiy g'oyasi yangi o'zgaruvchini (**newvar**) yaratish ekanligini bilib oldik, u bog'liq o'zgaruvchining har bir qiymati uchun y_i mos keladigan tekislangan qiymatni o'z ichiga oladi. Hisoblangan qiymatlar faqat $(x_i; y_i)$ ma'lumotlari va shu nuqtaga yaqin bir nechta ma'lumotlardan foydalangan holda x bo'yicha y ning regressiyasini bajarish orqali olinadi. Pastroqda regressiya shunday tortiladiki, markaziy nuqta $(x_i; y_i)$ eng katta og'irlikni oladi va uzoqroqda joylashgan nuqtalar ($|x_j - x_i|$ masofasidan kelib chiqib) kamroq og'irlik oladi. Keyin taxmin qilingan regressiya chizig'i faqat y_i uchun tekislangan \hat{y}_i qiymatini taxmin qilish uchun ishlatiladi. Qolgan tekislangan qiymatlarni olish uchun jarayon takrorlanadi, ya'ni ma'lumotlarning har bir nuqtasi uchun alohida vaznli regressiya amalga oshiriladi. Shubhasiz, agar sizning ma'lumotlar to'plamingiz katta bo'lsa, bu biroz vaqt talab qilishi mumkin. **Lowess** ma'lumotlarga amal qilishga moyil bo'lganligi sababli kerakli silliqlangan grafik deb aytiladi. Masalan, polinomlarni tekislash usullari globaldir, chunki tarqalish chizig'ining o'ta chap tomonida sodir bo'ladigan narsa o'ta o'ngdagi o'rnatilgan qiymatlarga ta'sir qilishi mumkin.

Grafikdan ko'rinib turibdiki, daromad oshgani sayin qoldiqlar kattalashib, maksimal 28 ga etadi. Eng katta daromadga ega bo'lgan kuzatuv qoldig'i nisbatan kichik va mahalliy darajada tekislangan taxmin chiziqning pastga yo'nalishini boshlashiga sabab bo'ladi.

8.2.2. Lagrange multiplikator testlari

Boshqa joylarda taklif qilingan gomoskedastizmning nol gipotezasi bo'yicha ko'plab testlar mavjud. Lagrange multiplikatorlariga asoslangan ulardan ikkitasini bajarish juda

oddiy va samaralidir. Birinchisi, ba'zan Breusch-Pagan (BP) testi yoki Statada Breusch-Godfrey testi deb ataladi. Ikkinchi sinov White ga beriladi.

Breusch-Pagan testi uchun nol va muqobil gipotezalar:

$$H_0: \sigma_i^2 = \sigma^2 \quad H_1: \sigma_i^2 = \sigma^2 h(\alpha_2 z_{i2} + \dots + \alpha_s z_{is})$$

Nol gipoteza shundan iboratki, ma'lumotlar gomoskedastikdir. Shu bilan bir qatorda, ma'lumotlar modelning dispersiyalari bilan bog'liq bo'lgan ekzogen o'zgaruvchilar bo'lgan $\alpha_{i2}, z_{i3}, \dots, \alpha_{is}$, o'zgaruvchilarga bog'liq bo'lgan tarzda geteroskedastikdir. $h()$ funksiyasi esa belgilanmagan. Bu uning argumentiga, ya'ni z dagi o'zgaruvchilarning chiziqli funktsiyasiga bog'liq bo'lgan har qanday narsa bo'lishi mumkin.

Quyidagi qadamlar orqali amalga oshiriladi:

- Regressiya modelini baholang
- Qoldiqlarni hisoblang
- Qoldiqlarni kvadratga aylantiring
- $z_{i2}, z_{i3}, \dots, \alpha_{is}$ dagi kvadrat qoldiqlarni regressiyaga aylantiring
- Ushbu regressiyadan NR^2 ni hisoblang va uni $\chi^2(S - 1)$ va taqsimotidagi α

darajadagi kritik qiymat bilan solishtiring.

Stata-da oziq-ovqat xarajatlari modelidagi daromadga bog'liq bo'lgan geteroskedastiklik testi:

use food, clear

quietly regress food_exp income

predict ehat, residual

gen ehat2=ehat^2

quietly regress ehat2 income

*di "NR2 = " e(N)*e(r2)*

di "5% critical value = " invchi2tail(e(df_m),.05)

*di "P-value = " chi2tail(e(df_m),e(N)*e(r2))*

Natijada:

NR2 = 7.3844244

5% critical value = 3.8414588

P-value = .00657911

E'tibor bering, Stata mos ravishda $e(N)$, $e(r2)$ va $e(df_m)$ da tanlama hajmini, R^2 va yordamchi regressiyadan erkinlik darajalarini saqlaydi. Odatdagidek, **invchi2tail** 5% kritik qiymatni va **chi2tail** LM statistikasining hisoblangan qiymati bilan bog'liq p-qiymatini olish uchun ishlatiladi.

White ning testi aslida Breusch-Pagan testidagi kichik o'zgarishdir. Nol va muqobil gipotezalar:

$$H_0: \sigma_i^2 = \sigma^2 \quad H_1: \sigma_i^2 \neq \sigma_j^2$$

kamida bitta $i \neq j$ uchun. Bu null bilan qamrab olinganidan boshqa barcha imkoniyatlarni qamrab oluvchi kompozit alternative gipotezadir. Agar siz ma'lumotlaringizdagi geteroskedastiklikning xususiyati haqida hech qanday ma'lumotga ega bo'lmasangiz, bu boshlash uchun yaxshi vaziyatdir. Mazkur test BP testiga juda o'xshaydi. Ushbu testda geteroskedastiklikka bog'liq bo'lgan o'zgaruvchilar ($z_{i2}, z_{i3}, \dots, \alpha_{is}$) har bir ortiqcha bo'lmagan regressor va uning kvadratini hamda regressorlar orasidagi barcha kesishgan natijalarni o'z ichiga oladi. Ma'lumotlar uchun matningizga qarang. Oziq-ovqat xarajatlari modelida faqat bitta doimiy regressor va kesishma mavjud. Shunday qilib, doimiy kvadrat va doimiy daromad o'rtasidagi ko'paytma ortiqcha. Daromad kvadrati bu modelga qo'shish uchun faqat bitta ijobiy o'zgaruvchini qoldiradi. Stata da daromadning kvadrat qiymatini yarating va daromad hamda uning kvadrati bo'yicha modeldan kvadrat qoldiqlarni regressga aylantiring. Ushbu regressiyadan NR^2 ni hisoblang va uni $\chi^2(S - 1)$ taqsimotidagi α darajasidagi kritik qiymat bilan solishtiring. Ushbu kitobda ko'rib chiqilgan barcha *LM* testlarida bo'lgani kabi, N ikkinchi yoki yordamchi regressiyadagi kuzatishlar soni hisoblanadi.

quietly regress ehat2 income c.income#c.income

*di "NR2 = " e(N)*e(r2)*

di "5% critical value = " invchi2tail(e(df_m),.05)

*di "P-value = " chi2tail(e(df_m),e(N)*e(r2))*

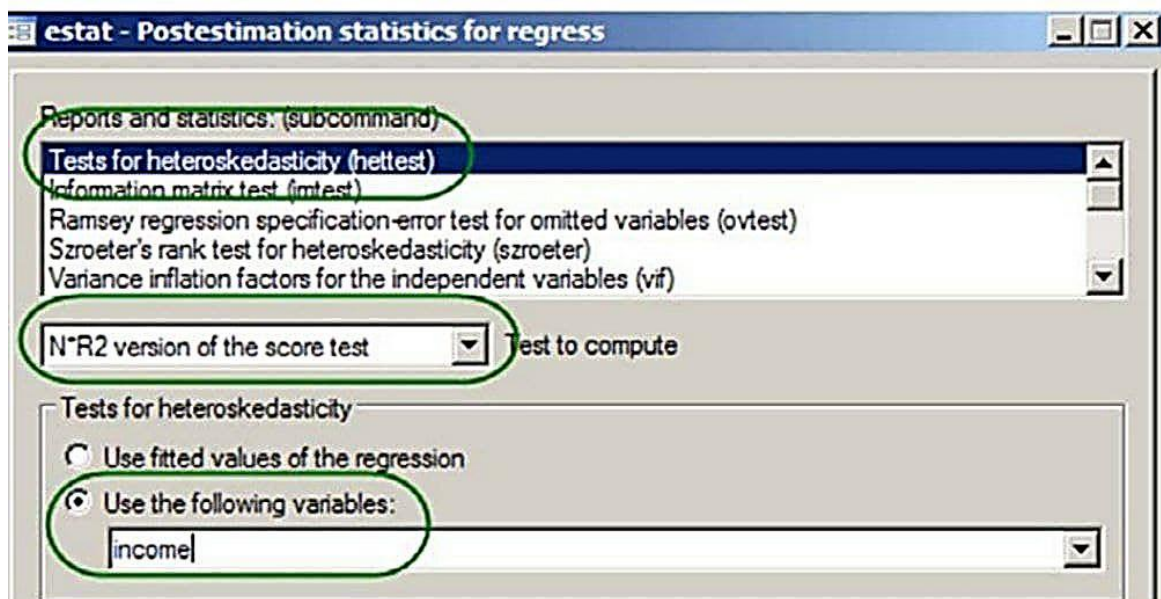
Natija esa quyidagicha:

NR2 = 7.5550786

5% critical value = 5.9914645

P-value = .02287892

Stata ushbu ikkala test statistikasi va ularning p-qiymatlarini hisoblash uchun funktsiyalarni ifodalaydi. Ammo bu yerda ularni topish biroz qiyin. Birinchi jarayon chiziqli regressiyani hisoblashdir. Keyin ochiladigan menyudan **Statistics >Linear models and related>Regression Diagnostics>Specification tests, etc** ni tanlang. So`ngra *LM* testini o'tkazish uchun ro'yxatdan **Tests for heteroskedasticity (hetttest)** ni tanlash uchun sichqonchanning o'ng tomonidagi tugmadan foydalaning. Keyingi ochiladigan menyuda ***N***R*² version of the score test** ni tanlang; bu **estat hottest** buyrug'iga **iid** birikmasini qo'shadi. **Use the following variables** uchun ko'rsatkich tugmasini bosing, so'ng ochiladigan ro'yxatdan kerakli o'zgaruvchi(lar)ni kiriting yoki tanlang. **OK** ni bosing.



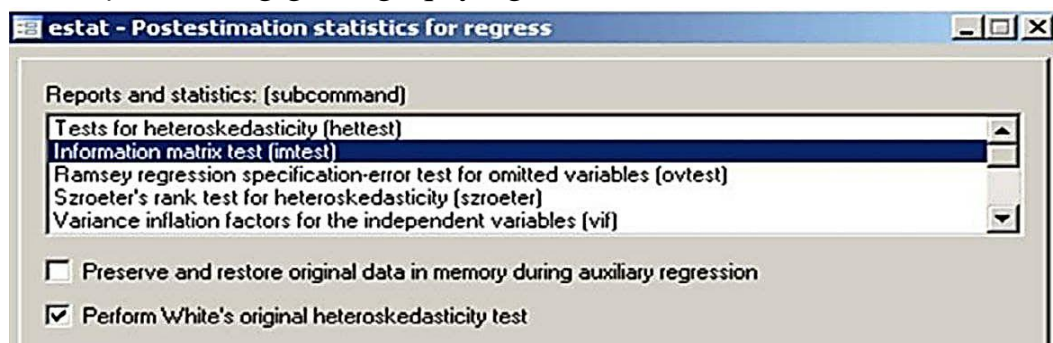
Natijada quyidagilar hosil bo'ladi:

```
. estat hettest income, iid
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
Ho: Constant variance
Variables: income

chi2(1)      =      7.38
Prob > chi2  =      0.0066
```

Stata buyruqlari yordamida olingan natija bilan bir xil natijani ifodalaydi.

White testi xuddi shu dialog oynasi yordamida amalga oshirilishi mumkin. Undan *Information matrix test (imtest)* ni tanlang va quyida ko'rsatilgan *Perform White's original heteroskedasticity test* katagiga belgi qo'ying.



Natijada:

```
. estat imtest, white
White's test for Ho: homoskedasticity
against Ha: unrestricted heteroskedasticity

chi2(2)      =      7.56
Prob > chi2  =      0.0229
```

Ikkala holatlardan eng yaxshisi **Command** oynasi orqali amalga oshirishdir.
quietly regress food_exp income

estat hettest income, iid
estat imtest, white

8.2.3. Goldfeld-Quandt testi

Agar dispersiya ma'lum bir o'zgaruvchiga bog'liqligiga shubha qilsangiz, Goldfeld-Quandt testini o'tkazish mumkin. Ushbu testni tasvirlash uchun o'rtacha ish haqiga ta'lim va tajribaning chiziqli funktsiyasi sifatida baholanadigan misoldan foydalanamiz. Bundan tashqari, agar odam metropolitik hududida yashasa, unga teng bo'lgan sohta o'zgaruvchi kiritilgan. Bu "to'siq" o'zgaruvchisi, ya'ni metro hududlarida yashovchi odamlar ta'lim va tajribadagi o'zgarishlarga (bir xil qiyaliklarda) xuddi shunday munosabatda bo'lishlari kutiladi, ammo ular qishloq joylaridagilarga nisbatan yuqori daromad olishadi (boshqa kesishish).

Sinov ma'lumotlarning ikkita qismidagi taxminiy farqlarni taqqoslaydi. Ushbu misolda metro namunasi uchun xatolik farqi qishloqdagiga teng deb taxmin qilinadi.

$$H_0: \sigma_M^2 = \sigma_R^2 \quad \text{qarshi} \quad H_0: \sigma_M^2 \neq \sigma_R^2$$

Ikki bo'lim eng kichik kvadratlar orqali baholanadi va taxminiy farqlar, $\hat{\sigma}_M^2$ va $\hat{\sigma}_R^2$ olinadi. Sinov statistikasi $F = \hat{\sigma}_M^2 / \hat{\sigma}_R^2$ bo'lib, agar nol gipoteza to'g'ri bo'lsa, $F_{(NM-KM, NR-KR)}$ taqsimotiga ega. $(N_M - K_M)$ faqat metro qismidan erkinlik darajalari va $(N_R - K_R)$ qishloq qismidan.

Birinchi, *cps2.dta* ma'lumotlar to'plamining barcha namunasi ta'lim, tajriba va regressorlar sifatida metro o'zgaruvchisidan foydalangan holda ish haqi modelini baholash uchun ishlatiladi.

use cps2, clear

regress wage educ exper metro

Qishloq kichik namunasi **if metro==0** kvalifikatsiyasi yordamida baholanadi. Stata faqat **metro** nolga yoki birga teng bo'lgan kuzatuvlardan **if** kvalifikatoridan foydalanishni buyuradi; **if** kvalifikatsiyasi regressiyani *after* ishlatsa va har qanday regressiya variantlarini *before* qiladi. Bu holda **if metro == 0** ifodasi faqat shartli (metro nolga teng) qanoatlantiradigan kuzatuvlardan foydalanadi. Ikki marta tenglik belgisi kerak, aks holda Stata siz **metro** o'zgaruvchisiga 0 qiymatini belgilashga harakat qilyapsiz deb o'ylaydi; bitta misolda = Stata ning tayinlash operatori. U o'zining o'ng tomonida joylashgan qiymatni chap tomonidagi nomga belgilaydi; **x=2 assigns** 2 qiymatini **x** o'zgaruvchisiga beradi. Bu yerda kerakli jarayon emas, shuning uchun ikkita teng belgidan foydalaning. Modelni kuzatishlar bilan chegaralangan **metro ==0** bo'lganlar bilan baholang va keyinchalik foydalanish uchun $\hat{\sigma}$ va erkinlik darajalarini saqlang.

regress wage educ exper if metro == 0

scalar rmse_r = e(rmse)

scalar df_r = e(df_r)

Metro kichik na'munasi uchun takrorlang:

```
regress wage educ exper if metro == 1
```

```
scalar rmse_m = e(rmse)
```

```
scalar df_m = e(df_r)
```

Endi Goldfeld-Quandt nisbatini tuzing, 5% kritik qiymatni va GQ ning p-qiymatini oling.

```
scalar GQ = rmse_m^2/rmse_r^2
```

```
scalar crit = invFtail(df_m,df_r,.05)
```

```
scalar pvalue = Ftail(df_m,df_r,GQ)
```

```
scalar list GQ pvalue crit
```

va Stata natijasi:

```
. scalar list GQ pvalue crit  
      GQ = 2.0877623  
      pvalue = 1.567e-09  
      crit = 1.2150333
```

Gomoskedastizm nol gipotezasi muqobillik foydasiga har qanday oqilona darajadagi ahamiyatlilik darajasida (5% yoki 10%) rad etiladi.

Oziq-ovqat xarajatlari misolida

Yana bir misol oziq-ovqat xarajatlari modelidan foydalanadi. Ushbu misolda dispersiya daromadning ortib borayotgan funktsiyasi deb hisoblanadi. Shunday qilib, biz birinchi navbatda ma'lumotlarni daromad (ko'tarilish) bo'yicha saralaymiz va keyin Goldfeld-Quandt testini takrorlaymiz. Qirqta kuzatuv ikkita teng o'lchamdagi bo'limga bo'lingan. Keyin natijaga erishish uchun yuqoridagi amallar takrorlanadi. Avval oziq-ovqat ma'lumotlar to'plamini yuklang va daromad bo'yicha tartiblang

```
use food, clear
```

```
sort income
```

Keyin taxminiy dispersiya va erkinlik darajasini saqlab, dastlabki 20 ta kuzatuvdan foydalanib, **food_exp income** va konstantada regressiya qiling. **in 1/20** iborasi 8.3-bo'limda ishlatilgan **if** kvalifikatsiyasiga o'xshash shartli birikma. Bunday holda, **in** Stataga bir qator kuzatuvlardan foydalanishni buyuradi va **1/20** 1 dan 20 gacha bo'lgan kuzatishlarni ko'rsatish uchun ishlatiladigan sintaksisdir.

```
regress food_exp income in 1/20
```

```
scalar s_small = e(rmse)^2
```

```
scalar df_small = e(df_r)
```

Endi, **food_exp** ni **income** va 21 dan 40 gacha (**in 21/40**) kuzatuvlardan foydalanib, hisoblangan dispersiya va erkinlik darajasini saqlab qolgan holda regressiya qiling.

```
regress food_exp income in 21/40
```

```
scalar s_large = e(rmse)^2
```

```
scalar df_large = e(df_r)
```

Endi Goldfeld-Qvandt statistikasini, uning 5% kritik qiymatini va p-qiymatini hisoblang.

```
scalar GQ = s_large/s_small
```

```
scalar crit = invFtail(df_large,df_small,.05)
```

```
scalar pvalue = Ftail(df_large,df_small,GQ)
```

```
scalar list GQ pvalue crit
```

Natijalar quyidagicha:

```
. scalar list GQ pvalue crit
      GQ = 3.6147557
      pvalue = .00459643
      crit = 2.2171971
```

Yana bir bor, gomoskedastiklik nol gipotezasi muqobillik foydasiga har qanday o'rtacha darajadagi ahamiyatlilik darajasida rad etiladi.

8.3. Geteroskedastik-ishlab chiqish standart xatolar

Eng kichik kvadratlar maydonidagi xatoliklarda geteroskedastik bo'lsa ham, yaxshi natijalarga erishgan taqdirda ham chiziqli modelni baholash uchun ishlatilishi mumkin. Yuqorida aytib o'tilganidek, geteroskedastik modelda eng kichik kvadratlardan foydalanish muammosi shundaki, aniqlikning odatiy baholovchisi (hisoblangan dispersiya-kovariatsiya matritsasi) izchil emas. Ushbu muammoni hal qilishning eng oddiy usuli - bu kesishma va qiyaliklarni baholash uchun eng kichik kvadratlardan foydalanish va xatolar geteroskedastikmi yoki yo'qmi, izchil bo'lgan eng kichik kvadratlar kovariatsiyasini baholovchidan foydalanishdir. Bu Stata ishlatadigan kovariatsiyaning **heteroskedasticity robust** baholovchisidir.

Ushbu misolda oziq-ovqat xarajatlari ma'lumotlari eng kichik kvadratlar yordamida modelni baholash uchun ishlatiladi. Ishchi katalogingizni *food.dta* ma'lumotlar to'plami va ma'lumotlar to'plamini o'z ichiga olgan katalogga o'zgartiring.

```
use food, clear
```

Eng kichik kvadratlar yordamida oziq-ovqat xarajatlari modelini taxmin qilishdan boshlang va hisob-kitoblarni saqlang (Odatda). **Vce(robust)** birikmasi yordamida modelni qayta baholang va natijalarni saqlang (**store White**). Keyin ikkala natijalar to'plamini ekranga chop etish uchun taxminlar jadvali buyrug'idan foydalaning.

```
quietly regress food_exp income
```

```
estimates store Usual
```

```
quietly reg food_exp income, vce(robust)
```

estimates store White

estimates table Usual White, b(%7.4f) se(%7.3f) stats(F)

Natijalar:

```
. estimates table Usual white, b(%7.4f) se(%7.3f) stats(F)
```

variable	Usual	white
income	10.2096 2.093	10.2096 1.809
_cons	83.4160 43.410	83.4160 27.464
F	23.7888	31.8498

Legend: b/se

E'tibor bering, koeffitsient baholari bir xil, ammo hisoblangan standart xatolar boshqacha. Qizig'i shundaki, mustahkam standart xatolar aslida odatdagidan kichikroqdir!

Ishonch oralig'i darajasini regressiya natijasidagi **level (90)** variantidan foydalanib, 90% ga o'zgartirish mumkin.

reg food_exp income, vce(robust) level (90)

Natija esa:

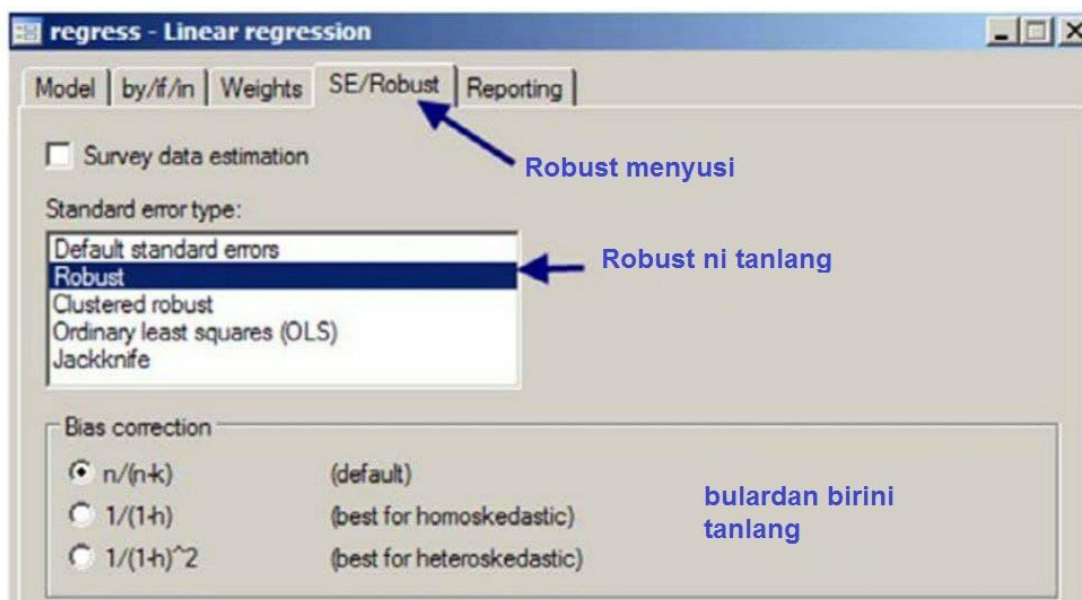
Linear regression

Number of obs = 40
F(1, 38) = 31.85
Prob > F = 0.0000
R-squared = 0.3850
Root MSE = 89.517

food_exp	Coef.	Robust Std. Err.	t	P> t	[90% Conf. Interval]	
income	10.20964	1.809077	5.64	0.000	7.159622	13.25966
_cons	83.416	27.46375	3.04	0.004	37.11337	129.7186

Yoki, ushbu bobning oxiridagi do-faylda ko'rsatilganidek, regressiyadan saqlangan natijalar yordamida ishonch oraliqlarini qo'lda hisoblash mumkin.

Muloqot oynalari bir xil natijalarni olish uchun ishlatilishi mumkin. Tanish **regress – Linear regression** muloqot oynasini ochish uchun **Statistics > Linear models and related > Linear regression** ni tanlang. Odatdagidek bog'liq va mustaqil o'zgaruvchilarni to'ldiring. Muloqot oynasidan chiqishdan oldin **SE/Robust** deb nomlangan belgini tanlang.



standard error type box oynasida **Robust** buyrug`ini tanlang va quyidagi tugmalar yordamida xususiyatini to'g'rilash variantlaridan birini tanlash kerak. Bu yerda biz uni standart qiymatda qoldirdik. Hammasi izchil, lekin har biri kichik namunalarda biroz boshqacha natijalar beradi. Endi **OK**-ni bosing.

Ushbu ishonchli standart xatolar Huber tomonidan taklif qilingan va Uayt tomonidan qayta kashf etilgan geteroskedastik izchil kovariatsiya matritsasi maydoni (**HCCME**) deb ataladigan jarayondan olingan. Ekonometrikada HCCME standart xatolari Uaytning standart xatolari yoki Huber/White standart xatolari deb atalishi ham mumkin.

Geteroskedastik modellarda eng kichik kvadratlar samarasiz bo'lganligi sababli, siz aniqroq bo'lgan boshqa xolis baholovchi bo'lishi mumkin deb o'ylaysiz. Va, bor **Generalized least squares (GLS)** hisoblagichini, hech bo'lmaganda, printsiplial jihatdan olish osonroqdir. Asosan, geteroskedastik modelning GLS mezon bilan turli xil xato dispersiyalari ma'lumotlarni qayta tortish uchun ishlatiladi, shunda ularning barchasi bir xil (gomoskedastik) dispersiyaga ega bo'ladi. Agar ma'lumotlar bir xil darajada o'zgaruvchan bo'lsa, eng kichik kvadratlar samarali bo'ladi!

8.4. Umumiyashtirilgan eng kichik kvadratlar usuli

Agar $var(e_i) = \sigma_i^2$ bo'lsa, e_i ni σ_i ga bo'lish barcha xatolarga bir xil dispersiyani beradi (1 ga teng). Quyida,

$$var(e_i/\sigma_i) = \frac{1}{\sigma_i^2 var(e_i)} = \left(\frac{1}{\sigma_i^2}\right) \sigma_i^2 = 1$$

Model og'irligini o'zgartirish uchun σ_i dan foydalanib kuzatuvlar. Oziq-ovqat xarajatlari modeli uchun bu bo'ladi.

$$\frac{y_i}{\sigma_i} = \beta_1 \frac{1}{\sigma_i} + \beta_2 \frac{1}{\sigma_i} + \frac{e_i}{\sigma_i}$$

Har bir kuzatish, y_i , doimiy va x_i , ushbu kuzatish xatosi bilan bog'liq bo'lgan standart og'ishning o'zaro nisbati bilan o'lchanadi. Bu murakkab tuyuladi, lekin ularning o'zgaruvchan qismini bilsangiz, Stata-da buni qilish ancha oson.

Oziq-ovqat xarajatlari modelidagi tafovut x_i ga proporsional deb faraz qilaylik:

$$\text{var}(e_i) = \sigma_i^2 = \sigma^2 x_i$$

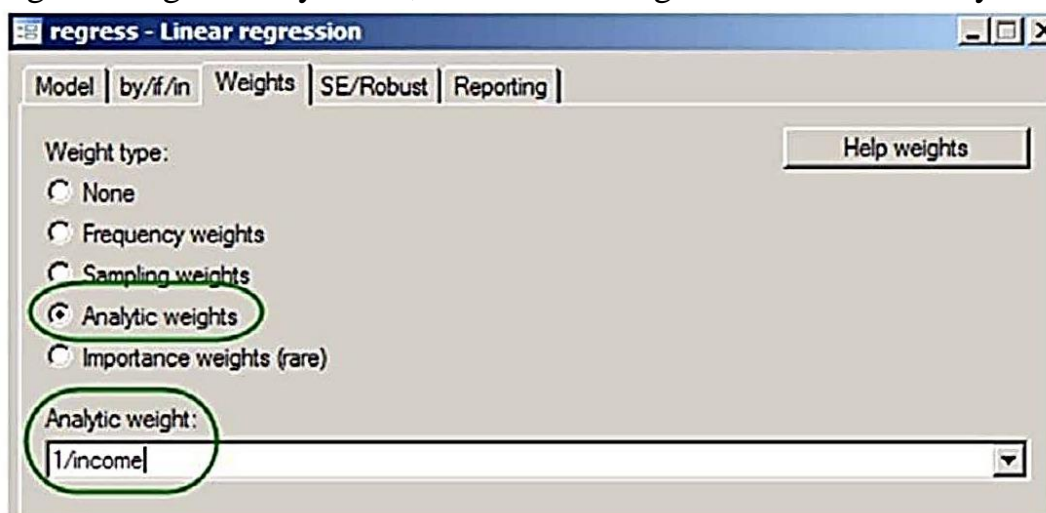
Shunday qilib, har bir kuzatuvga bir xil dispersiyani berish uchun y_i ni, doimiy x_i ni $\sqrt{x_i}$ ga bo'ling. Stata o'zining bir qator jarayonlarida, shu jumladan chiziqli regressiyada qiymatli ma'lumotlar bilan ishlash usulini o'z ichiga oladi. Oziq-ovqat xarajatlari modelini baholash uchun ma'lumotlar $1 / \sqrt{x_i}$ bo'yicha tortilishi kerak, bu Stata'da *analytic weights* yordamida amalga oshiriladi. Analitik og'irliklar kuzatishning *variance* siga *inversely* proporsionaldir. Shunday qilib, oziq-ovqat xarajatlari modelini qayta tortish uchun ishlatiladigan sintaksis quyidagicha:

regress food_exp income [aweight=1/income]

bu yerda **aweight** - analitik og'irliklar uchun Stata buyrug'idir. Standart og'ish olish uchun vaznning kvadrat ildizini olishning hojati yo'q. Kuzatishlarni $\sqrt{x_i}$ ga bo'lish uchun **aweight** ni $1 / x_i$ kiriting.

Bu holatda ham dialoglardan foydalanish oson. Endi tanish bo'lgan **regress – Linear regression** muloqot oynasini ochish uchun **Statistics > Linear models and related > Linear regression** ni tanlang. Odatdagidek bog'liq va mustaqil o'zgaruvchilarni to'ldiring. Muloqot oynasidan chiqishdan oldin **Weights** deb nomlangan yorliqni tanlang.

Analytic weights tugmasini bosing va quyida ko'rsatilganidek, kerakli analitik vaznni maydonchaga kiriting. Bunday holda, biz **income** ning o'zaro nisbatidan foydalanamiz.



OK tugmasini bosing va siz quyidagi natijalarni olasiz:

```
. regress food_exp income [aweight = 1/income]
(sum of wgt is 2.6616e+00)
```

Source	SS	df	MS			
Model	300459.464	1	300459.464	Number of obs =	40	
Residual	200775.783	38	5283.57325	F(1, 38) =	56.87	
				Prob > F =	0.0000	
				R-squared =	0.5994	
				Adj R-squared =	0.5889	
				Root MSE =	72.688	
Total	501235.248	39	12852.1858			

food_exp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
income	10.45101	1.385891	7.54	0.000	7.645419	13.2566
_cons	78.68408	23.78872	3.31	0.002	30.52633	126.8418

GLS tomonidan hisoblangan daromad koeffitsienti (1,386) standart xatosi HCCME (1,809) tomonidan eng kichik kvadratlar uchun ishlab chiqarilganidan sezilarli darajada kichik ekanligini ko'rishingiz mumkin.

8.4.1. Guruhlangan ma'lumotlardan foydalangan holda GLS

Misol ish haqini ta'lim va tajriba funktsiyasi sifatida baholashdan iborat va Goldfeld-Quandt test misolida qo'llanilgan **cps2.dta** ga asoslangan. Ushbu bo'limlarni birlashtirish va umumlashtirilgan eng kichik kvadratlar yordamida parametrlarni baholash strategiyasi juda oddiy. Har bir kichik namunadan modelni baholash uchun foydalaniladi va regressiyaning standart xatosi (**e(rmse)**) saqlanadi. Keyin har bir kichik namuna o'zining taxminiy dispersiyasi (**e(rmse)** ning kvadrat qiymati) bo'yicha tortiladi.

Har bir kichik namunani baholashning bir necha yo'li mavjud. Birinchisi Goldfeld-Quandt test misolida qo'llanilgan, bunda metro namunasi **if metro==1** kvalifikatsiyasi va qishloq uchun esa **if metro==0** bilan tanlangan. Ushbu usul yordamida guruhlangan GLSni bob oxiridagi do-faylida topish mumkin va bu yerda takrorlanmaydi. Ikkinchisi analitik og'irliklar (ya'ni, **aweight** buyrug'i) yordamida ma'lumotlarning kichik namunalarini olish mumkin bo'lgan hiyla ishlatadi. O'zgaruvchilarni 0 yoki 1 ga tortish kichik namunalarni olishning qulay usulidir. Kuzatishni tortish orqali baholash uni baholovchining hisob-kitobidan chiqarib tashlaydi, 1 ga tortilganlar esa uning hisobiga kiradi.

Ma'lumotlarni yuklagandan so'ng, bittadan **metroni** ayirib, qishloq uy xo'jaliklari uchun indikator o'zgaruvchisini yarating (qishloqda 1, aks holda 0). Analitik og'irliklardan foydalangan holda ikkita kichik to'plam regressiyasini bajaring va har birining o'rtacha kvadrat xatosini saqlang. Mana to'liq buyruq:

```
use cps2, clear
```

```
gen rural = 1 - metro
```

```
regress wage educ exper [aweight = rural]
```

```
scalar sr = e(rmse)^2
```

```
regress wage educ exper [aweight = metro]
```

```
scalar sm = e(rmse)^2
```

Ildiz MSElarning saqlangan qiymatlari barcha kuzatuvlarni qayta tortish uchun ishlatilishi mumkin bo'lgan yagona vaznga birlashtiriladi.

$gen\ wtall=(sr*rural) + (sm*metro)$

$regress\ wage\ educ\ exper\ metro\ [aweight = 1/wtall]$

Oxirgi qator GLS tomonidan modelni baholash uchun butun namunadan foydalanadi.

Buning natijasida:

```
. regress wage educ exper metro [aweight = 1/wt]
(sum of wgt is 3.7986e+01)
```

Source	SS	df	MS			
Model	9797.06665	3	3265.68888	Number of obs =	1000	
Residual	26284.1489	996	26.3897077	F(3, 996) =	123.75	
Total	36081.2155	999	36.1173329	Prob > F =	0.0000	
				R-squared =	0.2715	
				Adj R-squared =	0.2693	
				Root MSE =	5.1371	

wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	1.195721	.068508	17.45	0.000	1.061284	1.330157
exper	.1322088	.0145485	9.09	0.000	.1036595	.160758
metro	1.538803	.3462856	4.44	0.000	.8592702	2.218336
_cons	-9.398362	1.019673	-9.22	0.000	-11.39931	-7.397408

Ushbu modelda ish staji va maktabda o'qish yillari koeffitsientlari shahar va qishloqlarda ish haqi oluvchilar uchun bir xil. Ko'rsatkich o'zgaruvchisi, **metro**, shahar ishchilari uchun ijobiy bo'lgan kesishuvni o'zgartirishga imkon beradi.

8.4.2. GLS ning - umumiy holati

Yuqoridagi misolda kuzatishning standart xatosi i (yoki u nimaga mutanosib) ma'lum. Ko'pgina hollarda bu ma'lumot sizga ma'lum bo'lmaydi va siz ma'lumotlardan foydalanib uni taxmin qilishingiz kerak bo'ladi. Bu umumlashtirilgan eng kichik kvadratlarni (GLS) biroz boshqacha narsaga, ya'ni *estimated* yoki *feasible* umumlashtirilgan eng kichik kvadratlariga (FGLS) aylantiradi.

Birinchi qadam ba'zi mustaqil o'zgaruvchilarning funktsiyasi bo'lgan dispersiya uchun modelni tanlashdir. Sizga dispersiya o'zgarishi bilan bog'liq deb hisoblangan ba'zi o'zgaruvchilar kerak bo'ladi va siz dispersiya hamda bu o'zgaruvchilar o'rtasidagi funktsional munosabatni belgilashingiz kerak bo'ladi. Dispersiyaning umumiy modeli eksponensial funktsiyadan foydalanadi

$$\sigma_i^2 = \exp(\alpha_1 + \alpha_2 z_{i2} + \dots + \alpha_s z_{is})$$

bu yerda α_1 mustaqil o'zgaruvchilar va z_{is} parametrlardir. Natural logarifmni olib, eng kichik kvadratlarning kvadratlari qoldiqlarini kuzatilmaydigan σ_i^2 o'rniga qo'yish va xato atamasini qo'shish α_1 ni baholash uchun ishlatilishi mumkin bo'lgan regressiya modelini beradi. Oddiylik uchun, faqat bitta geteroskedastiklik bilan bog'liq z_i o'zgaruvchisini faraz qiling:

$$\ln(\hat{\epsilon}_i^2) = \ln(\sigma_i^2) + v_i = \alpha_1 + \alpha_2 z_i + v_i$$

bu yerda $\hat{\epsilon}_i^2$ sizning asl geteroskedastik regressiya modelingizning eng kichik kvadratlar bahosidan olingan. α_1 ni hisoblash uchun $\mathbf{z}_i = \ln(\mathbf{income})$ va Stata buyrug'i:

$gen\ z = \ln(income)$
 $regress\ food_exp\ income$
 $predict\ ehat,\ residual$
 $gen\ ln_ehat_sq = \ln(ehat*ehat)$
 $reg\ ln_ehat_sq\ z$

income ning natural logarifmi yaratiladi, **food_exp income** va doimiyda regressiya qilinadi, qoldiqlar saqlanadi va kvadratlarning natural logarifmi olinadi. Nihoyat, ular **z** va doimiyda regressiya qilinadi.

Ushbu regressiya quyidagi natijalarni beradi:

```
. reg ln_ehat_sq z
```

Source	SS	df	MS			
Model	54.8255435	1	54.8255435	Number of obs =	40	
Residual	112.530968	38	2.96134126	F(1, 38) =	18.51	
Total	167.356512	39	4.2911926	Prob > F =	0.0001	
				R-squared =	0.3276	
				Adj R-squared =	0.3099	
				Root MSE =	1.7209	

ln_ehat_sq	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
z	2.329239	.5413358	4.30	0.000	1.233362	3.425116
_cons	.9377961	1.583106	0.59	0.557	-2.267034	4.142626

FGLS uchun og'irliklarni olish uchun sizga oxirgi regressiyadan taxmin qilingan qiymatlarning antiloglari kerak. Ya'ni, **lnsig2** regressiyasidan chiziqli taxminlarni oling va eksponensial funktsiya **exp (lnsig2)** yordamida og'irliklarni hosil qiling.

$predict\ lnsig2,\ xb$
 $gen\ wt = exp(lnsig2)$

Endi, regressiya uchun kuzatuvlarni qayta o'lchash uchun uning **aweight** dagi o'zaro nisbatidan foydalaning.

$regress\ food_exp\ income\ [aweight = 1/wt]$

Buning natijasi:

```
. regress food_exp income [aweight=(1/wt)]
(sum of wgt is 6.1600e-02)
```

Source	SS	df	MS			
Model	186108.089	1	186108.089	Number of obs =	40	
Residual	59033.089	38	1553.50234	F(1, 38) =	119.80	
Total	245141.178	39	6285.67122	Prob > F =	0.0000	
				R-squared =	0.7592	
				Adj R-squared =	0.7529	
				Root MSE =	39.414	

food_exp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
income	10.63349	.9715143	10.95	0.000	8.666763	12.60022
_cons	76.05379	9.71349	7.83	0.000	56.38986	95.71773

8.5. Chiziqli ehtimollik modeldagi geteroskedastiklik

7.4-bobda biz ikkita muqobil o'rtasidagi tanlovni tushuntirish uchun chiziqli ehtimollik modelini tuzdik. Bu tanlov y ko'rsatkich o'zgaruvchisi bilan ifodalanishi mumkin,

agar birinchi muqobil tanlangan bo'lsa p ehtimollik bilan bir qiymatni oladi va ikkinchi alternativ tanlansa $1-p$ ehtimollik bilan nol qiymatini oladi.

i^{th} kuzatish uchun regressiya funktsiyasi

$$y_i = E(y_i) + e_i = \beta_1 + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + e_i$$

va $E(y_i) = p_i$, bu muqobilning individual i tomonidan tanlanishi ehtimoli. Buni anglatishi mumkin:

$$var(y_i) = p_i(1 - p_i)$$

bu modelni heteroskedastik qiladi. Mumkin GLS hisoblagichini hisoblash oson. Birinchidan, chiziqli regressiya baholanadi va $\hat{y}_i = \hat{p}_i$ taxminlari olinadi. Bular dispersiya tenglamasiga almashtiriladi.

$$\widehat{var}(y_i) = \hat{p}_i(1 - \hat{p}_i)$$

va nihoyat, ular regressiyada analitik og'irliklar sifatida ishlatiladi. Ba'zan bu bajarilmaydi, chunki taxmin qilingan ehtimollarning bir yoki bir nechtasi (0,1) oraliqdan tashqarida joylashgan.

Misol **coke.dta** ma'lumotlariga asoslangan. Mustaqil o'zgaruvchi, **coke**, agar shaxs Coca-Cola sotib olsa, 1 qiymatini oladi, agar bo'lmasa, 0 ga teng. Coca-Cola sotib olish to'g'risidagi qaror narxning Pepsiga nisbati va Coca-Cola yoki Pepsi uchun displeylar mavjudligiga bog'liq. Agar Coca-Cola displeyi mavjud bo'lsa **disp_coke=1** o'zgaruvchilari, aks holda 0; **disp_pepsi =1**, agar Pepsi displeyi mavjud bo'lsa, aks holda nolga teng.

Birinchidan, ma'lumotlar yuklanadi va umumiy statistik ma'lumotlar topiladi.

use coke, clear

summarize

Natijada:

```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
coke	1140	.4473684	.4974404	0	1
pr_pepsi	1140	1.202719	.3007257	.68	1.79
pr_coke	1140	1.190088	.2999157	.68	1.79
disp_pepsi	1140	.3640351	.4813697	0	1
disp_coke	1140	.3789474	.4853379	0	1
pratio	1140	1.027249	.286608	.497207	2.324675

Keyinchalik, model eng kichik kvadratlar bo'yicha baholanadi va bashoratlar yaratiladi va har qanday salbiy qiymatlar (yoki birdan katta qiymatlar) uchun tekshiriladi.

quietly regress coke pratio disp_coke disp_pepsi

predict p, xb

gen var = p(1-p)*

summarize p var

Natija quyidagicha:

```
. summarize p var
```

Variable	Obs	Mean	Std. Dev.	Min	Max
p	1140	.4473684	.1723611	-.2073211	.7680784
var	1140	.2175476	.0529915	-.2503031	.2499397

Minimal salbiy bo'lgani uchun, qanday davom etish kerakligi haqida qo'shimcha fikr yuritish kerak. Imkoniyatlardan biri $p < 0$ bo'lgan kuzatuvlarni o'tkazib yuborish va modelni baholashdir.

```
reg coke pratio disp_coke disp_pepsi [aweight=1/var] if p > 0
```

Yana bir narsa, p ning kichik musbat chegaradan past bo'lgan barcha qiymatlarini chegara bilan almashtirishdir.

Chegara 0,01 bo'lsin,

```
replace p = .01 if p < .01
```

```
replace var = p*(1-p)
```

```
reg coke pratio disp_coke disp_pepsi [aweight=1/var]
```

Bu yerda **replace** buyrug'i $p=0,01$ ni o'rnatish uchun ishlatiladi, agar u haqiqatan ham ushbu qiymatdan pastga tushsa. Cheklov chegarasidan pastga tushadigan p ning 16 ta qiymati mavjud. Yakuniy imkoniyat eng kichik kvadratlar yordamida modelni baholash va HCCME standart xatolaridan foydalanishdir. Xulosalar samarali bo'lmasa, haqiqiy bo'ladi.

```
reg coke pratio disp_coke disp_pepsi, vce(robust)
```

Ushbu regressiyalarning har biri hisoblab chiqilgan va jadvalga yig'ilgan (batafsil ma'lumot uchun bobning oxiridagi do-faylga qarang).

```
. estimates table LS Robust Trunc Omit, b(%7.4f) se(%7.4f) stats(F N)
```

Variable	LS	Robust	Trunc	Omit
pratio	-0.4009	-0.4009	-0.1652	-0.3859
disp_coke	0.0613	0.0604	0.0444	0.0527
disp_pepsi	0.0772	0.0772	0.0940	0.0760
_cons	0.0344	0.0339	0.0399	0.0353
	-0.1657	-0.1657	-0.1314	-0.1587
	0.0356	0.0344	0.0354	0.0360
	0.8902	0.8902	0.6505	0.8795
	0.0655	0.0653	0.0568	0.0594
F	51.6654	57.0701	36.9728	105.6006
N	1140	1140	1140	1124

Legend: b/se

LS deb nomlangan birinchi ustunda eng kichik kvadratlar taxminlari va odatiy (mos kelmaydigan) standart xatolar mavjud. Keyingi ustunda **robust** buyrug'i orqali yuborilgan geteroskedastiklik bilan mos keladigan standart xatolar bilan eng kichik kvadratlar taxminlari mavjud. Trunc deb nomlangan ustunda 0,01 chegaradan kamroq kuzatuvlar 0,01 ga qisqartirilgan taxminlar mavjud. Oxirgi ustun salbiy bashoratlarni keltirib chiqaradigan kuzatuvlar modeldan olib tashlangan natijalarni ko'rsatadi. Natijalar **Trunc** dan tashqari barcha modellarda mos keladi.

Model nazariy jihatdan geteroskedastik bo'lsa-da, ushbu bobda avval muhokama qilingan Uayt testi orqali namunaning geteroskedastikligini tekshirishga arziydi.

*quietly regress coke pratio disp_coke disp_pepsi
imtest, white*

Natija esa:

```
. imtest, white  
  
white's test for Ho: homoskedasticity  
against Ha: unrestricted heteroskedasticity  
  
chi2(7)      =      25.82  
Prob > chi2  =      0.0005
```

p-qiymati 5% dan ancha past va shuning uchun biz ma'lumotlar muhimlik darajasida geteroskedastik degan xulosaga keldik.

VIII bob mavzularini mustahkamlash uchun savollar

1. Geteroskedasticity deganda nimani tushunasiz?
2. Kovariatsiyasini baholovchisi bilan birga uning eng kichik kvadratlardan foydalanish nima uchun xizmat qiladi?
3. *graph twoway (scatter food_exp income) (lfit food_exp income)* buyrug'i nimaga xizmat qiladi?
4. O'zaro bog'liqlikni tuzish uchun qaysi buyruqdan foydalanish lozim?
5. Breusch-Pagan testi gipotezasini formulada ifodalang.
6. Breusch-Pagan testini amalga oshirish ketma ketligini tushuntiring.
7. Dispersiya ma'lum bir o'zgaruvchiga bog'liqligini tekshirish uchun qanday testni amalga oshirish kerak?
8. Umumiyashtirilgan eng kichik kvadratlar usulini tavsiflab bering.
9. Linear regression muloqot oynasini ochish uchun qanday amallar ketma ketligi amalga oshiriladi?
10. Uayt testini amalga oshirish vazifasini tushuntirib bering.