

AMALIY EKONOMETRIKA

Nuriddin **MURADULLAYEV**

Xasan **SABIROV**





XALQARO NORDIK UNIVERSITETI

N.N.MURADULLAYEV,
X.N.SABIROV

AMALIY EKONOMETRIKA

O'quv qo'llanma

TOSHKENT – 2024

UO'K: 519.862.6(075.8)

KBK: 65.6ya72

**Muradullayev N.N., Sabirov X.N. Amaliy
ekonometrika o'quv qo'llanma. - T.: XNU, 2024. –458 bet.**

Mazkur o'quv qo'llanmada ekonometrik modellashtirish jarayoni va unda Stata dasturining imkoniyatlari yoritib berilgan. O'quv qo'llanmada ekonometrik modellashtirish, Stata dasturiga ma'lumotlarni kiritish, ekonometrik modellarni tuzish, olingan natijalarni turli testlar yordamida tekshirish hamda prognozlash masalalari keltirilgan. O'quv qo'llanma bakalavr va magistratura talabalari uchun mo'ljallangan bo'lib, u ekonometrika sohasidagi fanlarni o'rganishda tavsiya etiladi.

Bundan tashqari mazkur o'quv qo'llanmadan professor-o'qituvchilar hamda tadqiqotchilar ham foydalanishlari mumkin.

Mas'ul muharrir:

Baxtiyor Salimov – Toshkent davlat iqtisodiyot universiteti
professori, iqtisodiyot fanlari doktori

Taqrizchilar:

Sherzod Mustafakulov – Xalqaro Nordik universiteti rektori, iqtisodiyot
fanlari doktori, professor

Piraddin Allayarov – Toshkent davlat iqtisodiyot universiteti,
"Iqtisodiyotda matematik metodlar" kafedrası dotsenti.

ISBN 978-9910-686-97-9

© «Dimal» nashriyoti, 2024.

Foydalanuvchilarga!

Qadrdon kitobxonlar, avvalambor "Amaliy ekonometrika" nomli o'quv qo'llanmadan foydalanayotganingiz uchun sizga tashakkur bildiramiz. Ushbu o'quv qo'llanma ta'lim ehtiyojidan kelib chiqqan holda yaratildi. O'quv qo'llanma ekonometrika va Stata dasturi bilan bog'liq amaliy muammolarni hal qilish va ma'lumotlar to'plamlari bilan ishonchli ishlashga imkon beruvchi amaliy va qulay ko'nikmani hosil qilishni ta'minlaydi.

Umid qilamizki, ushbu o'quv qo'llanmadan foydalangan holda ekonometrika haqida tushunchaga ega bo'lasiz va Stata dasturi imkoniyatlaridan foydalanishni o'rganasiz. Bu qo'llanmaning bir turi bo'lib, Stat Data Release 12 dasturidan foydalangan holda qo'llanmadagi misollarni qanday bajarish kerakligini buyruqlar va menyular asosida ko'rsatib beradi. Ushbu kitob ekonometrika fanini o'rganayotgan talabalar, shuningdek, o'qituvchilar va Stata dasturidan ekonometrik va statistik tahlil uchun foydalanmoqchi bo'lganlar uchun foydalidir.

Stata bugungi kunda juda ko'p turli xil ilmiy fanlarda qo'llaniladigan juda kuchli dastur hisoblanadi. Ushbu Web sayt: <http://www.stata.com> orqali juda ko'p ma'lumotlarni topishingiz mumkin. UCLA-saytda vizual manbalardan biri: <http://www.ats.ucla.edu/stat/stata/> hisoblanadi. Ushbu web saytdan foydalanishingizni tavsiya qilamiz. Ushbu Stata kompyuter dasturidan tashqari, shunga o'xshash EViews, Excel, Gretl va Shazam dasturiy ta'minot paketlari mavjud.

Biz ushbu o'quv qo'llanmaga izohlarni va takomillashtirish bo'yicha takliflarni mamnuniyat bilan qabul qilamiz. Savollarimizga javob berish va qo'llanmani takomillashtirishga beradigan takliflaringizni quyidagi elektron pochta manziliga yozishingizni so'raymiz. Navbatdagi izlanishlarimizda Stata dasturiy paketining ekonometrik modellashtirishdagi qo'shimcha imkoniyatlarini yoritish, shunga o'xshash dasturiy ta'minot paketlarining qiyosiy tahlili rejalashtirilgan va sizlar bergan takliflaringiz orqali takomillashtiriladi.

Mualliflar

MUQADDIMA

Bugungi kunda ekonometrik usullar iqtisodiy tahlil va prognozlashning eng muhim instrumentiga aylanib bormoqda. Ekonometrik tahlilda statistik axborotlarni qayta ishlash va tahlil qilishda bir qator universal dasturiy mahsulotlar mavjud. Ko'rib chiqilayotgan masalalar ko'lamiga qarab, ular ekonometrik usullarni o'rganishda nafaqat talabalarga, balki statistik ma'lumotlardan foydalanib, iqtisodiy tahlil va prognozlash masalalarini hal qilishda ilmiy izlanuvchilar hamda iqtisodchilarga ham foydali bo'lishi mumkin.

Mazkur o'quv qo'llanmada barcha asosiy ekonometrik hisoblar Stata amaliy paketida amalga oshirish ko'zda tutilgan. Stata dasturini tanlash undan foydalanishning qulayligi va ekonometrik modellashtirish amaliyotida keng qo'llanishi bilan asoslanadi.

Stata dasturi ekonometrik modellashtirishning barcha bosqichlarini o'zida mujassamlashtirgan. Unda ma'lumotlarni kiritish, ular asosida tavsifiy statistiklarni o'tkazish, korrelatsion tahlil, juft va ko'plikdagi regression tahlillarni o'tkazish, fazoviy va vaqtli qatorlar bo'yicha turli xil grafiklarni olish, olingan natijalarni turli testlar yordamida tekshirish imkoniyati mavjud. Yanada aniqroq to'xtaladigan bo'lsak, ushbu qo'llanma 13 ta bobdan iborat. Birinchi bob ekonometrika va Stata dasturi haqida qisqacha ma'lumot deb nomlanadi. Bu bobda ekonometrikaning maqsad va vazifalari, iqtisodiyotni ekonometrik modellashtirishning zarurligi, ekonometrik model tushunchasi, turlari va undagi o'zgaruvchilar hamda STATA dasturini ishga tushirish, o'chirish, ma'lumotlarni yuklash, tasvirlash va natijalarni saqlash haqida tushunchalar ifodalangan.

Ikkinchi bob regression model tuzish Stata dasturida deb nomlangan va bu qismda juft chiziqli regression modelni yaratish, tahliliy statistik ma'lumotlarni olish, ma'lumotlar asosida diagrammalar va grafiklar tuzish, hisoblangan qiymat va qoldiqlarni aniqlash, elastiklik qiymatini hisoblashni, variatsiya va kovarvatsiyani

hisoblashni, prognoz qiymatlarni topishni, kvadratik va logarifmik modellarni tuzishni hamda sifat o'zgaruvchilardan foydalanish imkoniyatlari tushuntirib o'tilgan. Uchinchi bob esa intervalli baholash va gipoteza testlari deb nomlandi hamda t taqsimotning kritik qiymatlarini hisoblash, gipoteza testlarini va p qiymatlarni hisoblash jarayonlariga to'xtalib o'tilgan. To'rtinchi bob chiziqsiz regressiyalar va modellashtirish masalalari deb nomlandi. Bu bobda asosan siz eng kichik kvadratlarni hisoblash, $\hat{Y}$ hisoblangan qiymat oralig'ini yaratish, korrelatsiya va determinatsiya R^2 , polinomial modellar, Jarque-Bera testi, x^2 -kvadrat taqsimotining kritik qiymatlari, Log-liner va Log-log modellari haqida bilim ko'nikmalarga ega bo'lasiz. Beshinchi bobni ko'p omilli regression modellar deb nomladik va bu bobda eng kichik kvadratlar bashorati, parametrlarning chiziqli birikmasi uchun ishonch oralig'i, o'zaro ta'sirlar uchun omil o'zgaruvchilardan foydalanish va modellar orasidagi moslik tushunchalarini ifodaladik.

Shuningdek, keyingi boblarda Stata dasturida alohida o'ringa ega bo'lgan ko'p omilli modellarni baholash, o'zgaruvchilar indikatoridan foydalanish, geterokidastiklik, Lagrange multiplikator testlari, Goldfeld-Quandt testi, HAC standart xatolar, dinamik qatorlar regressiyasi, Okun qonuni, AR modeli bilan prognoz qilish, Prais-Winsten FGLS testi, tasodifiy regressorlar va momentga asoslangan baholash, autoregressiya, taklif va talab modellari, Engle-Granger Testi, VEC, VAR, ARCH modellari to'g'risida ketma-ket mavzular yoritilgan. Foydalanuvchiga yana shuni ma'lum qilamizki, dasturda keltirilgan qulay menyu yordamida turli buyruqlar bilan ishlash, natijalarni taqqoslash uchun oynalarni yonma-yon qo'yish, yangi vaqtli qatorlarni tuzish, olingan natijalarni ishchi faylda obyekt sifatida saqlash va zarur vaqtda ulardan foydalanishingiz mumkin.

I BOB. AMALIY EKONOMETRIKA VA STATA DASTURI HAQIDA QISQACHA MA'LUMOT

1.1. Amaliy ekonometrika fanining predmeti

“Amaliy ekonometrika” fani milliy iqtisodiyotning barcha tarmoqlarini kompleks tahlil qilish, barqaror iqtisodiy o'sishning ko'p variantli modellarini yaratish, milliy iqtisodiyotni modernizatsiya qilish va taqchil resurslardan optimal foydalanish yo'nalishlarini aniqlab beradi.

“Amaliy ekonometrika” fani:

- iqtisodiyot bo'yicha chuqur fundamental tadqiqotlar olib borishga, iqtisodiy rivojlanish modellarini tuzishga imkon beradi;

- milliy iqtisodiyotni yaxlit tizim tarzidagi matematik modellarni yaratish orqali unga tashqi ta'sirlar, jahonda yuz berayotgan inqirozning biznes sikllariga ta'sirini aniqlay oladigan va qarshi choralarni belgilash strategiyalarini ishlab chiqishga imkon beradi;

- qo'yilgan muammoni nafaqat analitik ifodalab qolmasdan, balki axborot-kommunikatsiya texnologiyalari asosida iqtisodiy jarayonlarning matematik modellarini tuza oladigan, endogen va ekzogen omillar ta'sirini aniqlay oladigan milliy iqtisodiyotdagi tarmoqlar va korxonalarning biznes jarayonlarini modellashtirishga ko'maklashadi;

- tadqiq qilinayotgan obyekt, korxona, tarmoqning yoxud milliy iqtisodiyotning deterministik va stoxastik modellarini tuzishga imkon berib, ushbu jarayonlarni sifat va son jihatidan tizimli boshqara oladigan va prognoz qila oladigan modellarini yaratadi;

- tadqiq qilinayotgan obyektni samarali boshqarish uchun o'tkazilgan tadqiqotlar asosida maslahat beruvchi takliflar yoki boshqaruv qarorlarini ishlab chiqishga imkon beradi.

“Amaliy ekonometrika” fanining predmeti – bu korxonalarning xo'jalik jarayonlari, obyektiv va subyektiv omillar ta'siri ostida

shakllanayotgan ijtimoiy-iqtisodiy jarayonlar, samaradorlik va pirovard moliyaviy natijalarni iqtisodiy-matematik modellar va ma'lumotlar tizimi orqali ifodalangan asosiy ko'rsatkichlar, ularning barqarorligi va tebranish xususiyatlarini o'rganishdir. Xo'jalik jarayonlariga ta'sir etuvchi tashqi va ichki omillarni o'rganish hamda ularni miqdoriy baholash ham fanning predmeti hisoblanadi. Bularga misol qilib narx omilini keltirish mumkin. Narx omiliga doimo o'zgarib turuvchi narxlar, tariflar, stavkalar kiradi. Bozor sharoitida narxning shakllanishi umuman stixiyali jarayondir. Ularni shakllanishi qiymat qonuni va bozor qonuniyatlari talablari ostida shakllanadi, narxni o'zgarishi bozor ishtirokchilari faoliyatlariga turlicha ta'sir o'tkazadi. Ularning ta'sirini o'rganish va iqtisodiy tahlil qilish murakkab iqtisodiy hisob-kitoblarni bajarishni talab qiladi.

"Amaliy ekonometrika" fani iqtisodiy tahlil predmeti bo'lib, subyektiv ichki omillar ta'siri ostida shakllangan xo'jalik faoliyati va uning pirovard natijalari ham hisoblanadi. Ichki omillar inson faoliyati bilan bog'liq bo'lgani uchun iqtisodiy tahlilda unga katta ahamiyat beriladi. Xo'jalik faoliyatini mohirona boshqarish obyektiv sharoit va obyektiv omillardan to'g'ri foydalanib, biznes rejani to'la bajarish imkoniyatini yaratadi. Buning uchun "Amaliy ekonometrika" matematik usullar yordamida xo'jalik mexanizmini boshqarish asboblari (instrumentlari) to'g'ri foydalaniladi. Subyektiv omillar (ichki), uddaburon rahbar, ishlab chiqarishni to'g'ri malakali boshqarish, iqtisod, moliya faoliyatini to'g'ri tashkil qilish, o'z sohasini chuqur egallash iqtisodiy va tashkiliy tayyorgarlik asoslaridir.

Obyektiv va subyektiv omillar ta'siri ostida shakllanayotgan xo'jalik jarayonlari va ularning natijalari mos ravishda korxonaning iqtisodiy ko'rsatkichlar tizimida o'z ifodasini topadi va tahlil predmeti hisoblanadi. Ma'lumotlar tizimi tez o'zgaruvchan – dinamik xarakterga ega; ular kirish ma'lumotlari to'plami, oraliq qayta ishlash natijalari, chiqish ma'lumotlari va pirovard natijalarni o'z ichiga oladi hamda boshqarish tizimiga kiradi. Bu ma'lumotlarni samarali tashkil etish va boshqariladigan ma'lumotlar oqimini shakllantirish iqtisodiy tahlil va boshqarishning optimal qarorlarini qabul qilishda xizmat qiladi.

Milliy iqtisodiyotda ekonometrik usullarni qo'llashdan asosiy maqsad ommaviy hodisalar va jarayonlar haqida ularni kuzatish yoki eksperimentlar natijasida olingan ma'lumotlar asosida xulosalar hosil qilish va ular asosida ekonometrik modellar tuzishdan iborat. Bu statistik xulosalar alohida tajribalarga tegishli bo'lmasdan, balki tadqiq qilinayotgan hodisani keltirib chiqaruvchi shart-sharoitlarning doimiy ekanligi farazidagi shu hodisaning umumiy tavsiflari (ehtimolliklari, taqsimot qonunlari va ularning parametrlari, matematik kutilishlari va h.k.) haqidagi da'volardan iborat.

Ommaviy tasodifiy hodisalar bo'ysunadigan qonuniyatlarni aniqlash statistik ma'lumotlarni kuzatish natijalarini ehtimollar nazariyasi uslublari bilan o'rganishga asoslanadi.

1.2. "Amaliy ekonometrika" fanining ketma-ket vazifalari

"Amaliy ekonometrika" fanining vazifasi kuzatishlar yoki maxsus o'tkazilgan eksperimentlar natijasida olingan statistik ma'lumotlarni to'plash va guruhlash usullarini o'rganish hisoblanadi.

"Amaliy ekonometrika" fanining vazifasi:

a) hodisaning noma'lum ehtimolligini baholash; noma'lum taqsimot funksiyasini baholash; ko'rinishi ma'lum bo'lgan taqsimotning parametrlarini baholash; tasodifiy miqdorning boshqa bitta yoki bir nechta tasodifiy miqdorlarga bog'liqligini baholash va h.k.;

b) noma'lum taqsimotning ko'rinishi haqidagi yoki ko'rinishi ma'lum bo'lgan taqsimot parametrlarining miqdori haqidagi statistik gipotezalarni tekshirish kabi tadqiqot maqsadlariga bog'liq ravishda statistik ma'lumotlarni tahlil qilish usullarini ishlab chiqishdan iborat.

Har qanday fan keng ma'noda tabiat, jamiyat va tafakkur haqidagi bilimlar to'plamidan iboratdir. Bu to'plam har qanday tarixiy bosqichda erishilgan va tushunib yetilgan tabiat va jamiyatning obyektiv qonunlariga mos keladi. "Amaliy ekonometrika" ham shu

kungacha iqtisodiyot va matematika fanlari tomonidan erishilgan nazariy yutuqlarni amaliy faoliyatda qo'llash asosida shakllangan fandir. Iqtisodiyotning talablariga mos ravishda matematika usullari rivojlangan bo'lsa, matematika kirib borgan soha o'z masala va muammolariga aniq va to'g'ri javob olgan va rivojlangan.

"Amaliy ekonometrika" fani maxsus bilimlar tizimidan iborat bo'lib, u quyidagilar bilan shug'ullanadi:

- obyektiv iqtisodiy qonunlar va subyektiv omillar ta'siri ostida shakllanayotgan iqtisodiy jarayonlar va ularning o'zaro bog'lanishini miqdoriy tomonlarini o'rganish;

- biznes rejalarini ilmiy asoslash va ularni bajarilishini obyektiv baholash;

- iqtisodiyotga ijobiy va salbiy ta'sir etuvchi omillarni aniqlash va ularning ta'sirini miqdoriy baholash;

- xo'jalikning rivojlanish tendensiyalari va proporsiyalarini ochib berish, foydalanilmagan ichki imkoniyatlar va zaxiralarni aniqlash;

- ilmiy asoslangan bashoratlarni ishlab chiqish;

- milliy iqtisodiyotning tarmoqlari va sohalari asosiy ko'rsatkichlarining istiqboldagi holatini aniqlash va baholash;

- ilg'or tajribalardan foydalanish va optimal boshqarish qarorlarini qabul qilish.

Iqtisodiy jarayonlarni o'rganish alohida xo'jalik elementlari, faktlar, hodisa va holatlarni kuzatishdan boshlanadi va ular birgalikda xo'jalik jarayonini, xo'jalik faoliyati mazmunini u yoki bu boshqarish tizimida ifodalanadi. Ammo shu bilan birga iqtisodiy jarayonni to'liq, butunligicha o'rganish kerak bo'ladi, chunki bir vaqtning o'zida umumiy yo'nalish va uni ta'minlash uchun zarur bo'lgan sharoitlar va faoliyatlar tekshiriladi.

Iqtisodiy tahlilni olib borishda xo'jalik jarayonlari birgalikda, o'zaro bog'langan holda va o'zaro munosabatlari kelishilgan holda o'rganiladi. Iqtisodiyotdagi o'zaro bog'lanishlar, bir-biridan kelib chiqishlar va birgalikdagi munosabatlar hamda ularni miqdoriy baholash tahlilning eng ahamiyatli masalalari hisoblanadi. Bog'lanishlar sabablari barcha xo'jalik asoslari, hodisa va sharoitlarni qamrab oladi. Bu bog'lanishlarsiz xo'jalik faoliyati davom eta

olmaydi, rivojlanmaydi, to'xtab qoladi.

Sababli yoki omilli tahlilni iqtisodiy jarayonlarda qo'llash natijasidan shu narsa kelib chiqadiki, xo'jalik faoliyati bilan bog'liq har bir sabab, har bir omil o'ziga yarasha baho oladi. Shu maqsadda dastavval sabab-oqibat omillari o'rganiladi, buning uchun ular guruhlar bo'yicha tavsiflanadi: mohiyatli va mohiyatsiz, asosiy va qo'shimcha, aniqlovchi va aniqlamaydigan omillarga ajratiladi. Bundan keyin eng avvalo xo'jalik jarayonlariga mohiyatli, asosiy va aniqlovchi omillarning ta'siri o'rganiladi. Mohiyatsiz, qo'shimcha va aniqlanmaydigan omillarni o'rganish, agar zarur bo'lsa ikkinchi navbatda olib boriladi. Iqtisodiy jarayonlarga ta'sir etuvchi barcha omillarning ta'sirini o'rganish juda ham qiyin masala va amalda hamma vaqt ham zarur bo'lmaydi.

Iqtisodiy jarayonlarning bajarilishini aniqlovchi, ta'sir o'tkazuvchi asosiy sabablarni ochish va tushunib yetish, ularning ta'siri va o'zaro ta'sirini aniqlash – bu tahlil qilinayotgan obyektning xo'jalik faoliyati xususiyatlariga tushunib yetishdir. Tahlil jarayonida xo'jalik faoliyatiga ta'sir etuvchi asosiy omillar aniqlanib va xarakterlanib qolmasdan, balki ularning ta'sir darajasi ham miqdoriy o'lchanadi.

"Amaliy ekonometrika" fanining ahamiyati, mazmuni va predmetlari uning oldiga qo'yilgan vazifalarni belgilaydi va bu vazifalarga quyidagilar kiradi:

- korxonaning biznes rejasi va me'yorlarini ilmiy-iqtisodiy asoslash;

- biznes reja va me'yorlarning bajarilishini obyektiv va har tomonlama o'rganish;

- moliyaviy, mehnat va moddiy resurslardan samarali foydalanish yo'llarini aniqlash;

- tijorat rejasining bajarilishini nazorat qilish;

- ichki zaxiralarni aniqlash va miqdoriy o'lchash (ishlab chiqarishning barcha bosqichlarida);

- ishlab chiqarishning texnik-iqtisodiy ko'rsatkichlarini prognozlash;

- boshqarish yechimining optimalligini sinab ko'rish va tekshirish (iyerarxiya pog'onasining barcha bo'g'inlarida).

Birinchi vazifa – biznes reja va me'yorlarni ilmiy-iqtisodiy asoslash birinchi navbatda xo'jalik faoliyatini har tomonlama retrospektiv tahlilini amalga oshirish bilan erishiladi. Ma'lum bir davr uchun vaqtli qatorlarni tuzish, xo'jalikning rivojlanishidagi ma'lum iqtisodiy qonuniyatlarni belgilash imkoniyatini beradi. Shundan so'ng xo'jalik faoliyatiga o'tgan davrda ta'sir o'tkazgan va kelgusida ahamiyatli ta'sir o'tkazishi mumkin bo'lgan asosiy omillar aniqlanadi. Joriy davrda xo'jalik faoliyati tahliliga alohida e'tibor qilinadi, chunki, u rejalashtirish oldi davri bo'lib hisoblanadi. Retrospektiv tahlil xulosalari joriy kuzatishlar bilan birgalikda umumlashtirilgan holda rejalashtirish hisob-kitoblarida foydalaniladi. Rejalashtirish ishlari joriy hisobot davridan oldin bajarilgani uchun reja oldi tahliliga zarurat tug'iladi: kutilayotgan rejalarning bajarilishi, qoidaga ko'ra to'rtinchi kvartalda bajariladi.

Retrospektiv va joriy tahlil kelgusiga (perspektiv) tahlil qilish bilan tugallanadi va uning ko'rsatkichlari rejalashtirilgan – tahlil qilingan ko'rsatkichlarga to'g'ri chiqadi.

Barcha hollarda pirovard ishlab chiqarish moliyaviy natijalari qiyosiy tahlil usullaridan, ilg'or korxonalarining ijtimoiy-iqtisodiy samaradorlik ko'rsatkichlaridan foydalaniladi. Shunday qilib, biznes reja har tomonlama zarur iqtisodiy hisob-kitoblar bilan asoslanadi.

Ikkinchi vazifa – hisob-kitob va hisobot ma'lumotlari bo'yicha biznes-rejalarning bajarilishini va me'yorlarga rioya qilishni obyektiv va har tomonlama o'rganish.

Sanoat korxonalarida tahlil jarayonida ishlab chiqarish dasturining, asosiy mahsulotning miqdori va assortimenti, navi, komplektligi va sifati bo'yicha bajarilishi, ishlab chiqarishning barqarorligi, sotish rejasi, yetkazib berish shartnomasining bajarilishi o'rganiladi.

Uchinchi vazifa – mehnat, moddiy va moliyaviy resurslardan foydalanishning iqtisodiy samarasini aniqlash. Korxonalarda mehnat vositalari va predmetlarini (bino va inshootlar, texnologiyalar, asbob-uskuna, xom ashyo va materiallar); ishchi kuchi (ishchilar soni va ixtisoslik tarkibi, asosiy, yordamchi, xizmat ko'rsatuvchi va boshqarish personal, mehnat unumdorligi va boshqalar); moliyaviy

resurslar (shaxsan o'zining va qarzga olingan, asosiy va aylanma mablag'lar) umuman barchasi bo'yicha izlanishlar olib boriladi.

To'rtinchi vazifa – tijorat hisobi va pirovard moliyaviy natijalar tahlilini nazorat qilishdir (uni to'liq va to'liq bo'lmagan shakllarida). Korxonaning barcha ishlab chiqarish faoliyati va ularning moliyaviy natijalari eng avvalo tijorat hisobi tamoyillariga rioya qilish bilan bog'liqdir, u esa o'z navbatida ishlab chiqarish munosabatlarini ifodalab, ular shakllanayotgan bozor munosabatlari talablariga to'liq javob beradi. Moddiy tomondan manfaatdorlik tamoyili tijorat hisobining ajralib turadigan xususiyati bo'lib, korxona va jamoa, shaxsiy va jamiyat manfaatlarining zarur bo'lgan moslashuvni ta'minlaydi. Tijorat hisobi mehnatga yarasha taqsimotning iqtisodiy tamoyilini amalga oshiruvchi ahamiyatli vosita sifatida namoyon bo'ladi. Tijorat hisobi bir mulkchilik shaklidagi korxonalar, turli mulkchilik shaklidagi korxonalar va davlat korxonalari o'rtasidagi turli munosabatlarni (o'zaro manfaatli) shakllantirishni ta'minlaydi. Qonun doirasida korxonalarning mustaqilligi va davlat oldida soliqqa tortish bo'yicha mas'uliyati, resurslardan samarali foydalanish barcha tijorat hisobining ajralib turadigan xususiyatlaridir. Bularning barchasi hayotga bozor iqtisodiyoti talablarini joriy qilishning zaruriyatini ko'rsatadi.

Korxonalarning tijorat hisobi va moliyaviy natijalarini baholashda miqdoriy va sifat ko'rsatkichlaridan foydalaniladi, sanoat korxonasining foyda miqdori faqat ishlab chiqarilayotgan mahsuloti miqdori va assortimentiga bog'liq bo'lmaydi, balki ko'p jihatdan uning tannarxiga ham bog'liqdir.

Korxonaning tijorat hisobi va moliyaviy natijalarini to'g'ri baholash o'rganilayotgan ko'rsatkichlarga ta'sir o'tkazuvchi omillarning bog'langan va bog'lanmaganlarga ajratishni talab qiladi. Tashqi, bog'lanmagan omillarni hisobga olmaslik, korxonaning jamoasi mehnatini to'g'ri, aniq talqin etish imkonini beradi.

Korxonaning ichki bo'limlari faoliyatini tahlil qilishda analitik yo'nalishlar birmuncha qisqaradi. Xususan, turli resurslarni yetkazib berish shartnomalarini tuzish, bank va budjet hisob-kitoblari bilan

bog'liq pozitsiyalar qisqaradi, moliyaviy natijalar tahlili imkoniyatlari chegaralanadi.

"Amaliy ekonometrika" fanining beshinchi vazifasi – ishlab chiqarish jarayonining barcha bosqichlarida ichki imkoniyatlar zaxirasini izlab topish va miqdoriy o'lchashdir.

"Amaliy ekonometrika" amaliy fan bo'lib, undan haqiqiy – real foyda kelsagina o'zini to'la oqlashi mumkin. "Amaliy ekonometrika" fanining haqiqiy foydasi ishlab chiqarishning barcha bo'g'inlarida foydalanilmayotgan zaxiralarni aniqlashdan iborat. Iqtisodiyotning o'sish sur'atining ortishi xo'jalikdagi mavjud imkoniyatlardan barchasi harakatga keltirilganligidan iqtisodiyotning asosiy tamoyili – eng yuqori natijalarga eng kichik xarajatlar bilan erishishni yo'lga qo'yishga to'g'ridan to'g'ri bog'liqdir. Bu muammolarni to'g'ri aniqlashda va miqdoriy baholashda iqtisodiy tahlilga matematik usullar asosiy qurol bo'lib xizmat qiladi.

Iqtisodiy tahlil jarayonida zaxiralarni aniqlash korxona bo'limlarini, bir turdagi korxonalarning rejani bajarishlarini hamda mamlakat va xorijdagi ilg'or tajribalardan to'la foydalanishni qiyoslash yo'li bilan o'rganilib topiladi. Ishlab chiqilgan rejani tanqidiy baholash ham uni ortig'i bilan bajarish imkoniyatini ko'rsatib berishi mumkin.

"Amaliy ekonometrika" fanining oltinchi vazifasi – boshqarish yechimlarini optimalligini asoslash va sinashdan (tekshirishdan) iborat. Boshqarishning barcha pog'onalarida xo'jalik faoliyatining muvaffaqiyatlari boshqarish darajasi bilan, o'z vaqtida qabul qilingan boshqarish qarorlari bilan ham to'g'ridan to'g'ri bog'liq. To'g'ri boshqarish yechimini qabul qilish, uni ratsional va samaraliligini aniqlash faqat dastlabki iqtisodiy tahlil qilish asosidagina bajarish mumkin.

Amaliy ekonometrika vazifalari yuqorida keltirilgan vazifalar bilan chegaralanmaydi. Xo'jalik vaziyatlarining ko'p qirraligi va ko'p variantligi, uning oldiga avtonom xarakterga ega bo'lgan ko'pgina vazifalarni qo'yadi. Ularni yechish uchun umumiy va xususiy analitik usullardan foydalanish zarur.

1.3. Matematik modellarni ishlab chiqish va masalalar yechish ketma-ketligi

Bu bo'limda har qanday loyihani bajarishning iqtisodiy-matematik, xususan ekonometrik modellarini ishlab chiqish va ulardan foydalanish xususiyatlarini hisobga olgan holda umumiy tamoyillari ko'rib chiqilgan. Agar ilgari sanab o'tilgan bosqichlar ketma-ket bajariladi deb taxmin qilingan bo'lsa, endi oldingi bosqichlarga qaytish normal hisoblanadi. Masalan, konseptual modelni ishlab chiqishda yangi ma'lumotlarga ehtiyoj bo'lishi mumkin va umuman bu bosqich ma'lumotlarni to'plashdan oldin bo'lishi mumkin.

1. Masalaning qo'yilishi. Ehtiyojlarni tushunish, ishning maqsadini, kutilayotgan natijalarni, mavjud resurslarni (pul, texnik, inson, huquqiy), bajarilishi kutilayotgan ish hajmini ifodalash; mavjud ishlanmalarni va dasturiy ta'minotni, sotib olish yoki yetishmayotgan ishlanmalarni baholash; ishlanmaning maqsadga muvofiqligi haqida qaror qabul qilish; texnik topshiriq, kalendar reja, narx kelishuvini ishlab chiqish kerak. Masalaning qo'yilishi yoki chiqish parametrlarining noto'g'riligi katta vaqt va pul sarflanishiga olib kelishi mumkin. Ushbu kursning asosiy maqsadlaridan biri - ma'lumotlarni qayta ishlash va iqtisodiy model tuzish masalalarini to'g'ri shakllantirish imkonini berishdir.

2. Mavzu obyektini tekshirish, axborotni to'plash va sifatini baholash. Modelning adekvatligi ham, modellashtirish natijalarining ishonchliligi ham modellashtirish obyekti haqidagi dastlabki ma'lumotlarning sifatiga bog'liq. Balki o'rganilayotgan obyektlarni guruhlariga bo'lish kerak bo'ladi va har birining o'z qonuniyatlari bo'ladi. Noto'g'ri yoki anomal (g'ayritabiyy) ma'lumotlar (masalan, bayram kunlaridan oldingi savdo haqida) bo'lishi mumkin. Ularni ajratish va alohida ko'rib chiqish imkoniyatiga ega bo'lish kerak. Ular odatda grafiklarda ko'rinadi. Nazariy asoslar ham obyekt haqida ma'lumot bo'lib, maqsadli ma'lumotlar to'planishiga yordam beradi. Ekonometrik tadqiqotlarda ma'lumotlar uch guruhga bo'linadi: o'zbek tilidagi darsliklarda "fazoviy" deb yuritiladigan ko'ndalang

kesimli ma'lumotlar (**cross-sectional data**); **vaqtli qatorlar (time series)**; **panel ma'lumotlari (panel data)**, vaqtli qatorlarni o'z ichiga oladi.

3. Konseptual modelni tuzish. Tizimda o'rganilayotgan jarayonlarning xarakteriga qarab modellashtirishning barcha turlarini deterministik va stoxastik, statik va dinamik, diskret, uzluksiz va diskret uzluksizliklarga bo'lish mumkin. **Deterministik modellashtirish** deterministik jarayonlarni, ya'ni tasodifiy ta'sirlar yo'qligi taxmin qilinadigan jarayonlarni ko'rsatadi; **stoxastik modellashtirish** ehtimoliy jarayon va hodisalarni ko'rsatadi; **ekonometrik modellashtirish** bu turga kiradi. Bu holda tasodifiy jarayonning bir qator amalga oshishlari tahlil qilinadi va o'rtacha xarakteristikalar, ya'ni o'xshash amalga oshishlar to'plami baholanadi. **Statik modellashtirish vaqtning** qandaydir momentida obyekt xatti-harakatini tasvirlash uchun ishlatiladi, ekonometrikada bunday modellar fazoviy modellar deyiladi. **Dinamik modellashtirish** obyektning vaqt davomidagi xatti-harakatini aks ettiradi. Ekonometrikada vaqtli qatorlar va ularning to'plamlari (panelnie dannie) o'rganiladi. **Diskret modellashtirish** diskret deb qabul qilingan jarayonlarni tasvirlash uchun ishlatiladi, uzluksiz modellashtirish tizimlarda uzluksiz jarayonlarni aks ettirish imkonini beradi, **diskret uzluksiz modellashtirish** esa diskret va uzluksiz jarayonlarning mavjudligini ta'kidlaydigan holatlar uchun qo'llaniladi. Ekonometrika diskret ma'lumotlarga asoslangan, lekin natijasi uzluksiz funksiyadir.

Konseptual model bosqichida real obyektning norasmiy tavsifidan (matnlar, chizmalar, sxemalar, diagrammalar va hokazo) foydalanish mumkin, lekin rasmiylashtirilgan texnologiyalardan: tizimli tahlil, Universal Modellashtirish Tili va boshqalardan ham foydalaniladi.

Konseptual modelni qurishning asosiy bosqichlari:

- faraz va takliflar qilish;
- modelning parametrlari va o'zgaruvchilarini aniqlash;
- tizim samaradorligi ko'rsatkichlari va mezonlarini tanlashni asoslash;

■ modelning mazmunli tavsifini tayyorlash.

Tizim modelini qurishdagi gipotezalar tadqiqotchining masalani tushunishidagi "bo'shliqlar"ni to'ldirishga xizmat qiladi. Taxminlar modelni soddalashtirish imkonini beradi. Tizim modeli bilan ishlash jarayonida modellashtirishdan olingan natijalar va obyekt haqidagi yangi ma'lumotlarga qarab, bu kichik bosqichga ko'p marta qaytish mumkin.

Parametrlar va o'zgaruvchilarni aniqlashda kirish, chiqish va boshqarish o'zgaruvchilari hamda tizimning tashqi va ichki parametrlari ro'yxati tuziladi.

Tizimning tanlangan ko'rsatkichlari va samaradorlik mezonlari tizimning ishlash maqsadini aks ettirishi va tizimning o'zgaruvchilari va parametrlari funksiyalarini ifodalashi kerak. Ekonometrikada o'zgaruvchilarning asosiy turlari:

■ prognozlash ekonometrikaning asosiy vazifalaridan biri bo'lgan endogen yoki bog'liq o'zgaruvchilar;

■ **ekzogen** yoki ta'sir qiluvchi o'zgaruvchilar; tizimga nisbatan tashqi bo'lishi mumkin (dollar kursi, hisob stavkasi, vaqt) yoki biz ularni boshqarishimiz mumkin: turli maqsadlar uchun xarajatlar;

■ **lag**: o'tgan vaqt oralig'idagi o'zgaruvchilar; kecha biz ularni prognozlashga harakat qildik, lekin bugun biz ularni bilamiz.

Ekzogen va lag o'zgaruvchilarni **oldindan belgilangan** deb ham atashadi. Bundan tashqari, biz keyinchalik ko'rib chiqadigan fiktiv, almashtiruchi va instrumental o'zgaruvchilar mavjud.

Konseptual modelni ishlab chiqish tizim yoki jarayonni tasvirlovchi matnlar, formulalar, grafiklar, rasmlar, jadvallarni o'z ichiga olishi mumkin bo'lgan mazmunli tavsif tuzish bilan yakunlanadi va keyingi harakatlar uchun asosiy hujjat sifatida foydalaniladi.

4. Masalalarning formal tavsifi, strukturaviy modelni tuzish: tenglamalar sistemasi, ayniyatlar, cheklanishlar-tengliklar va cheklanishlar-tengsizliklar.

5. Masalani yechish algoritmini ishlab chiqish. Algoritm–masalani yechishga olib keladigan dastlabki va oraliq ma'lumotlarni qayta ishlash jarayonini yuqori darajada aniqlaydigan amallarning

sonli ketma-ketligi. Ekonometrikada – bu strukturaviy modelni keltirilgan shaklga aylantirishdir: tenglama yoki tengliklar tizimi, unda endogen (prognoz qilinuvchi) o'zgaruvchilar chap tomonda, ekzogen va lag o'zgaruvchilar esa o'ng tomonda bo'ladi. Bu bosqich juda ko'p hisob-kitoblarni talab qiladi. Hozirgi kunda bunday masalalarni yechish uchun ko'plab dasturlar turli amaliy dasturlar paketlari: Excel, MatCad, MatLab, Stata, EVIEWS va boshqa servislar ko'rinishida rasmiylashtirilgan, masalani yechish odatda paket, xizmat ko'rsatish, uni sozlash va foydalaniladigan ma'lumotlar bilan tutashtirish bilan almashtiriladi, odatda interaktiv grafik rejimda: formulalarni kiritish, cheklovlarni belgilash va hokazo. Yechish usulini, dasturiy ta'minotni (va haqiqiy loyihada dasturchini) to'g'ri tanlash loyihaaning vaqtini va narxini o'n barobar kamaytirishi mumkin. Taklif etilgan masalalarni yechish texnologiyalari buni amalga oshirishga imkon beradi.

6. Testlarga tekshirish. Sintaksis xatolariga ega bo'lmagan dastur mantiqiy xatolarga ega bo'lishi va noto'g'ri natijalar berishi mumkin. Shuning uchun alohida bloklar ham, umuman olganda dasturning o'zi ham ma'lum yechimlar bilan test masalalari yordamida sinalishi kerak. Ekonometrikaning 90%ini modelning ishonchliligini va uning parametrlarini baholash usullari tashkil etadi. Ekonometrik modelning sifat ko'rsatkichlari: determinatsiya koeffitsiyenti R^2 , F Fisher statistikasi, tenglamalar koeffitsiyentlari uchun Studentning t statistikasi, Darbin-Uotsonning DWavtokorrelyasiya uchun testi, Goldfeld-Kvandtning GQ geteroskedastikli uchun testi, ekzogen o'zgaruvchilarning korrelatsion matritsasi bo'yicha multikollinearlikni aniqlash, shuningdek prognoz xatoligini baholash va modelning adekvatligini tekshirish.

7. Modellastirish natijalarini rasmiylashtirish va talqin qilish model bilan mashina eksperimenti natijasida olingan ma'lumotlardan asl obyektning ishlash jarayoniga doir xulosalarga o'tishga qaratilgan. Modellastirish natijalarini jadvallar, grafiklar, diagrammalar, sxemalar va h.k. ko'rinishda taqdim etilishi mumkin. Ko'p hollarda, tizimni modellastirishning natijalarini grafiklar aniqroq ko'rsatsa-da, eng oddiy shakl jadvallar hisoblanadi.

Boshliqqa yaxshi tayyorlangan jadvallarini keltiring, u biror qaror qabul qilishi mumkin bo'ladi va sizdan mamnun bo'ladi. O'qituvchiga esa masalalar yechimini tekshirish qulay bo'ladi.

1.4. STATA dasturini ishga tushirish va o'chirish

Stata dasturini bir necha usul bilan ishga tushurish mumkin. Birinchisi, kompyuteringiz ish stolida (1.1-rasm) ikki marta bosishingiz mumkin bo'lgan yorliq bo'lishi mumkin va Stata / SE Release 12 uchun quyidagicha ko'rinishda bo'ladi:



1.1-rasm. Stata dasturining yorlig'i.

Stata dasturining oldingi versiyalariga o'xshash ko'rinishdagi belgilari mavjud, lekin, bu albatta, boshqa raqamga ya'ni "StataSE-12" qo'shimcha funksiyalariga ega. Shu bilan bir qatorda, Windows menusidan foydalanib, **Start > All Programs > Stata 12** ni bosishingiz mumkin.

Ikkinchi usul esa * **.dta** kengaytmasi bilan Stata ma'lumotli faylini topish va uni ikki marta bosib, Stata dasturini ishga tushirish mumkin.

Dasturning ko'rinishi.

Stata dasturi ishga tushirilgandan so'ng, oynalarni o'z ichiga olgan Stata dasturinig asosiy oynasi paydo bo'ladi (1.2-rasm). Bu asosiy oynada quyidagicha oynalar mavjud:

Command (buyruq) – bunda Stata buyruqlari yoziladi.

Results (natijalar) – buyruqlar natijalari va xato buyruqlar bu oynada ko'rinadi.

Review (ko'rib chiqish) – yaqinda bajarilgan buyruqlar ro'yxati oynasi.

Variables (o'zgaruvchilar) – ma'lumotlar va teglardagi o'zgaruvchining nomlari bu oynada ko'rinadi (yaratilgan bo'lsa).

Yuqorida "Stata"da ochiladigan menyulardir. Biz ulardan ko'pchiligini o'rganib chiqamiz. Chap pastki burchakda "Stata"da grafikalar, ma'lumotlar fayllari va hokazolarni saqlaydigan ishchi katalogning mavjud joyidir.



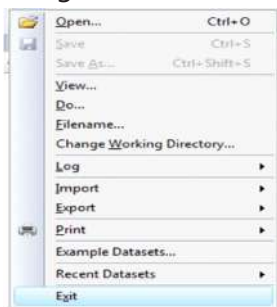
1.2-rasm. Stata dasturi oynasi.

STATA dasturini o'chirish

Stata dasturini to'xtatish uchun fayl ni bosib, ya'ni:



va File menyusidagi Exit ni tanlab bosish kifoya.



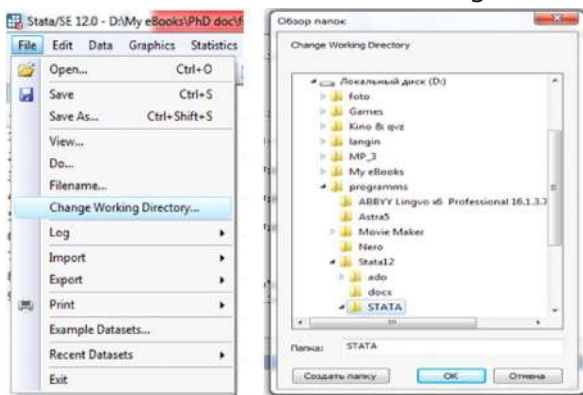
Biz File > Exit kabi ketma-ket bosish buyruqlarini bildiramiz. Shu bilan bir qatorda, shunchaki buyruqlar oynasiga "exit" so'zini yozib, Enter klavishini bosish ham kifoya qiladi:



Ishchi fayl

Ma'lumotni qulay faylga ko'chirishingiz kerak. Buni qanday amalga oshirilishi kompyuter tizimiga bog'liq bo'ladi. Ushbu Windows bazasidagi barcha ma'lumot va natijalar fayllari uchun biz **d:\programms\Stata12\STATA** pastki katalogidan foydalanamiz. Biz buni qulay bo'lishi uchun qilyapmiz va agar siz laboratoriya sharoitida bo'lsangiz, bu juda yaxshi tanlov emas.

Agar siz laboratoriya xonalaridagi kompyuterda ishlayotgan bo'lsangiz, siz "**fleshka**" yoki "**Hard**" kabi saqlash qurilmasiga ega bo'lishingiz kerak. Ular Stata dasturidagi ma'lumotlar faylli va dastur fayllarini saqlash uchun yetarlicha katta bo'lishi kerak. Ishchi faylda ishlash uchun **File (Fayl) > Working Directory of Changen**i oching. Natijada fayllar oynasida siz tanlagan joyga o'ting va **OK** tugmasini bosing. Ushbu ketma-ketlik 1.3-rasmda izohlangan.



1.3-rasm. Stata ischi faylida ishlash.

Mazkur ishni Stata dasturida **Command** oynasi orqali ham bajarish mumkin, ya'ni: `cd "D:\programms\Stata12\STATA"`.

Yuqoridagi buyruqni **Command** oynasiga kiriting va **Enter** tugmasini bosing.

Command
`cd "D:\programms\Stata12\STATA"`

Mazkur buyruq natijasi esa 1.3-rasmda ko'rsatilgan:



1.4-rasm. Stata dasturi menyusi.

Results (natijalar) oynasida buyruq qayta aks ettiriladi va u **Review** (tadqiq) oynasida ham ko'rinadi. Yangi yo'l ekranning chap pastki qismida ko'rsatiladi.

1.5. STATA dasturida ma'lumotlarni kiritish va yuklash

Stata ma'lumotlar fayllarini ochish yoki yuklashning bir necha yo'llari mavjud. Ularning ikki usulini tushuntiramiz:

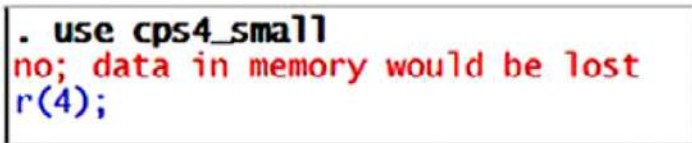
Fayldan foydalanish. Stata dasturi ishga tushgandan so'ng, Stata ma'lumotlar fayllarini saqlagan joyingizga o'zingizning ishchi

katalogingizni joylashtiring. So'ngra Command oynasida `use cps4_small` dan foydalaning va Enter tugmasini bosing.



1.5-rasm. Stata dasturida `use cps4_small` buyrug'idan foydalanish.

Agar ma'lumotlar faylingiz allaqachon ochiq bo'lsa va uni qandaydir tarzda o'zgartirgan bo'lsangiz, bunda Stata buyruqni xato ma'lumot deb quyidagicha javob beradi:



1.6-rasm. Stata dasturida xato ma'lumot haqidagi xabar.

r (4) tugmasini bosganingizda xato ma'lumot haqidagi xabarni Viewer oynasida o'qishingiz mumkin. Ba'zan bu foydali bo'ladi. Viewer oynasini yopish uchun X ni bosing.

Bu xususiyat siz saqlashingiz mumkin bo'lgan ma'lumotlar faylidagi o'zgarishlarni yo'qotishingizga to'sqinlik qiladi. Agar shunday bo'lsa, avvalgi ma'lumot faylini saqlashingiz yoki quyidagi buyruqni kiritishingiz mumkin:

«clear»

Yuqoridagi buyruq Stata xotirasida bo'lgan narsalarni yo'q qiladi. Ma'lumotlar faylini ochish va xotirani tozalash zarur bo'lsa, quyidagini kiriting:

`use cps4_small, clear`

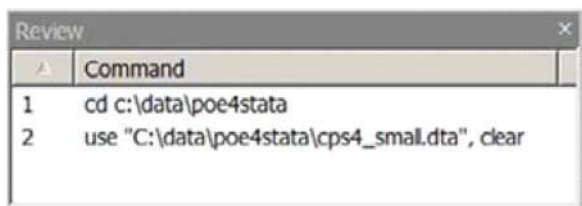
Keyingi usul **uskunalar panelini ishlatish**. Stata data faylini ochishda uskunalar satridagi Open (ochish) belgisini bosing.



1.7-rasm. Stata dasturida uskunalar paneli.

Siz ochmoqchi bo'lgan faylni toping, tanlang va Open buyrug'ini bosing. Review oynasida ko'rsatilgan Stata buyrug'i ko'rsatiladi. use "C:\data\poe4stata\cps4_small.dta", clear.

Ma'lumotlar faylini Stata dasturida ochishda use buyrug'i bilan erishiladi. Ma'lumotlar faylining yo'li ekranning chap pastki qismida ko'rsatiladi. Yangi ma'lumotni ochish mavjud ma'lumotlarning xotiradan o'chirilishini ko'rsatadi.



1.8-rasm. Stata dasturida uskunalar paneli.

O'zgaruvchilar oynasi (variables window).

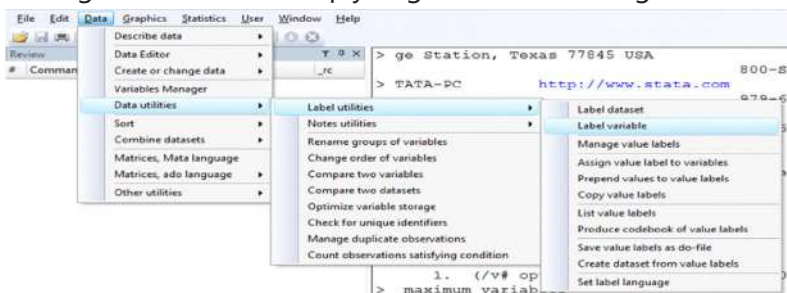
O'zgaruvchilar oynasida stata ma'lumotlari faylining parametrlari beriladi (1.9-rasm). Shuningdek, o'zgaruvchilarning izohlari ko'rsatiladi, agar ular mavjud bo'lsa, ularning turi va formati ham ko'rsatilgan bo'ladi.

Name	Label	Type	Format
wage	earnings per hour	double	%10.0g
educ	years of education	byte	%8.0g
exper	post education years experience	byte	%8.0g
hrrswk	usual hours worked per week	byte	%8.0g
married	= 1 if married	byte	%8.0g
female	= 1 if female	byte	%8.0g
metro	= 1 if lives in metropolitan area	byte	%8.0g
midwest	= 1 if lives in midwest	byte	%8.0g
south	= 1 if lives in south	byte	%8.0g
west	= 1 if lives in west	byte	%8.0g
black	= 1 if black	byte	%8.0g
asian	= 1 if asian	byte	%8.0g

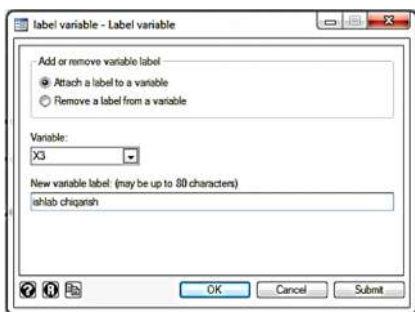
1.9-rasm. O'zgaruvchilar oynasida stata ma'lumotlari faylining parametrlari.

Izohlar foydalidir va ularni osongina qoʻshilishi, oʻzgartirilishi yoki oʻchirilishi mumkin.

Bitta oʻzgaruvchiga izoh joylashtirish. Stataning yuqoridagi menusidan **Data > Data utilities > Label utilities > Label variable** ni tanlang. Bu ketma-ketlik quyidagi rasmda izohlangan.



Natijada paydo boʻlgan oynada, yangi oʻzgaruvchiga izoh yozish uchun **Attach a label to a variable** ni tanlab, soʻngra oʻzgaruvchi tanlanadi va **New variable label** joyiga oʻzingiz xohlagan izoh qoldirishingiz mumkin (1.10-rasm). Undan keyin **OK** tugmasini bosish kifoya.



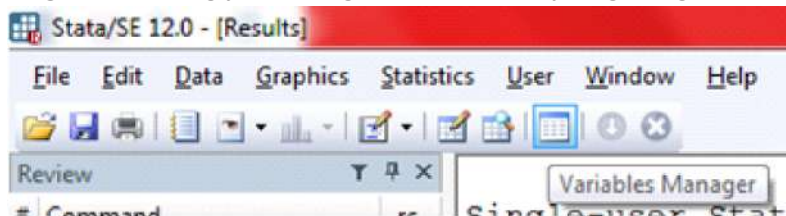
1.10-rasm. Oʻzgaruvchiga izoh yozish oynasi.

Yuqoridagi ketma-ketlik oʻrniga siz Command oynasiga quyidagini kiritishingiz mumkin. Bu quyidagicha boʻladi: label variable X3 "ishlab chiqarish"

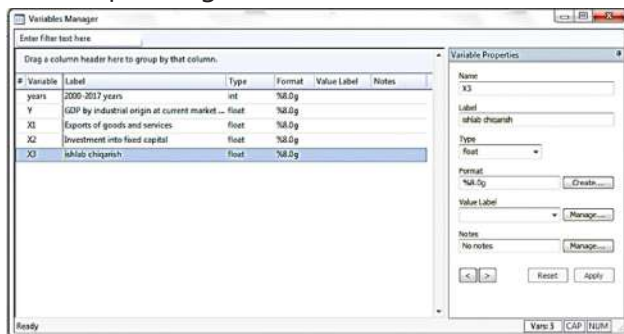
Command oynasiga buyruqni yozganingizdan soʻng **Enter** ni bosing. Bu esa izohni yaratadi va X3 uchun yangi izoh yoziladi.

Muloqot oynasidan siz yozgan izohni olib tashlashni ham tanlashingiz mumkin.

Shuningdek, izohlar menejeridan ham foydalanish mumkin. Izohlar boshqaruvchisiga boshqarish uchun yagona joy – **Variables Manager**. Stataning yuqoridagi menusidan quyidagi belgini bosing.



Variables Manager o'zgaruvchilar xususiyatlarini ochish uchun **Variable Properties** ni bosing. Bu yerda o'zgaruvchining nomini o'zgartirish, izohlar qo'shish va o'zgaruvchining formatini va guruhlarini boshqarishingiz mumkin (12-rasm).



1.11-rasm. O'zgaruvchilar oynasi.

1.6. Ma'lumotlarni tasvirlash va tasviriy statistikasini olish

Ma'lumotlar faylini har doim ko'rish va tasvirlash uchun bir necha usullar mavjud. Avval quyidagi buyruqni kiriting: **describe**

Bu esa xotiradagi ma'lumotlar majmuasining xulosasini, shu jumladan o'zgaruvchilar ro'yxati, ular haqidagi ma'lumotlar va

ularning izohlarini tasvirlaydi. Natijalarning bir qismini quyidagi rasmda ko'rishingiz mumkin.

```
. describe
```

Contains data from 2017.dta
 obs: 18
 vars: 5 1 Mar 2019 15:38
 size: 324

variable name	storage type	display format	value label
years	int	%8.0g	2000-2017 years
Y	float	%8.0g	GDP by industrial origin at current market prices
X1	float	%8.0g	Exports of goods and services
X2	float	%8.0g	Investment into fixed capital
X3	float	%8.0g	Manufacturing

Sorted by:

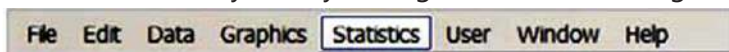
Keyin quyidagi buyruqni kiriting: **summarize**

Natijalar oynasida biz qisqacha xulosa statistikasini ko'ramiz. Bu esa quyidagi rasmda tasvirlangan:

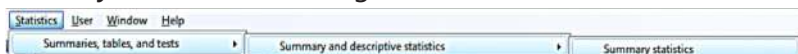
```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
years	18	2008.5	5.338539	2000	2017
Y	18	73.49739	75.99263	3.256	254.043
X1	18	19.23783	18.00062	.864	73.392
X2	18	17.95961	19.63861	.744	68.424
X3	18	13.18078	12.69302	.462	42.611

Agarda Stata buyrug'ini unutib qo'ysangiz, yuqoridagi menyulardan foydalanib, qisqacha xulosa statistikasini olishingiz mumkin. Siz Stata menyusi ro'yxatidagi Statistics ni bosing.

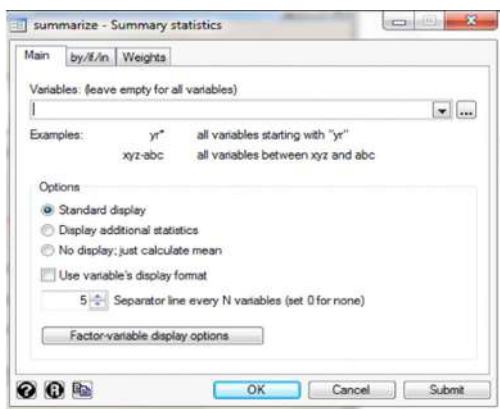


Siz undan foydalanishingiz mumkin bo'lgan statistik tahlillarning uzoq ro'yxatini ko'rasiz. Keyin undan **Summaries, tables, and tests** ni tanlang, so'ngra **Summary and descriptive statistics** va **Summary statistics** ni tanlang.



Ko'p variantli ko'rsatadigan dialog oynasi ochiladi. Asosiy

abstrakt statistikasi uchun hech qanday tanlov kerak emas. OK-ni tanlang. Stata avtomatik ravishda ma'lumotlar to'plamidagi barcha o'zgaruvchilar uchun qisqacha xulosa statistikasini taqdim etadi. **Variables** joyiga o'zingizga kerakli o'zgaruvchilar nomlarini yozish orqali yoki individual o'zgaruvchilarni tanlashingiz ham mumkin. **Standard display** kuzatuvlar soni, arifmetik o'rtacha, standart og'ish, o'zgaruvchilarning minimal va maksimal qiymatini beradi.



Tasviriy statistikaning sintaksisi

Quyidagi misollarni sintaksis xususiyatlaridan foydalanib ko'rib chiqamiz. Shuning uchun **Command** oynasiga quyidagi buyruqni kiriting va Enter tugmasini bosing.

Command

summarize y, detail

Summarize Y, detail buyrug'i o'zgaruvchan **Y(GDP)** uchun batafsil abstrakt statistikasini hisoblaydi. Eng kichik va eng kattalikdagi **Y** o'zgaruvchining foizlari haqida ma'lumot olishingiz mumkin bo'lgan qo'shimcha ma'lumotlar statistikasi (masalan, skewness va kurtosis) bilan birga ko'rsatiladi. Eslatib o'tamiz, Stata oldingi nuqta (.) bilan berilgan buyruqqa qaytadi.

```
. summarize Y, detail
```

GDP by industrial origin at current market prices

Percentiles		Smallest		
1%	3.256	3.256		
5%	3.256	4.925		
10%	4.925	7.45	Obs	18
25%	12.261	9.844	Sum of Wgt.	18
50%	44.173		Mean	73.49739
		Largest	Std. Dev.	75.99263
75%	120.862	145.846		
90%	199.993	171.808	Variance	5774.881
95%	254.043	199.993	Skewness	1.02485
99%	254.043	254.043	Kurtosis	2.923281

Natijalar oynasining quyi chap qismida **more** ko'rsangiz, natijada: Results oynasi to'lgan bo'lsa, to'xtatib turadi va qo'shimcha natijalar paydo bo'lishi uchun **more** so'zini bosishingiz kerak.

1.7. Natijalarni saqlash usuli

Nusxa olish va joylashtirish. Umumiy tartib sifatida tavsiya etilmasa ham, natijalarni **Result** oynasini belgilab olib sichqonchani o'ng tugmasini bosing.

Bu sizning matni matn sifatida nusxalash (**Ctrl + C**) va qisqa (**Ctrl + V**) yoki pastki belgini bosish orqali hujjatga joylashtirishni beradi.

```
. summarize Y, detail
```

GDP by industrial origin at current market prices

Percentiles		Smallest		
1%	3.256	3.256		
5%	3.256	4.925		
10%	4.925	7.45	Obs	18
25%	12.261	9.844	Sum of Wgt.	18
50%	44.173		Mean	73.49739
		Largest	Std. Dev.	75.99263
75%	120.862	145.846		
90%	199.993	171.808	Variance	5774.881
95%	254.043	199.993	Skewness	1.02485
99%	254.043	254.043	Kurtosis	2.923281



Agar siz matnni qayta ishlash hujjatiga joylashtirsangiz, to'g'ri joylashtirilgan Stata natijalari noto'g'ri va o'qish uchun qiyinlashishi

mumkin. Natijalarning bir qismi o'xshash bo'lishi mumkin. Bu quyidagi ko'rinishda bo'ladi:

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
years	8	2008.5	5.338539	2000	2017
Y	18	73.49739	75.99263	3.256	254.043
X1	18	19.23783	18.00062	.864	73.392
X2	18	17.95961	19.63861	.744	68.424
X3	18	13.18078	12.69302	.462	42.611

Bu shriftni o'zgartiradigan matn protsessorga bog'liq. Siz standart matn uchun Times New Roman shriftini ishlatishingiz mumkin bo'lsa, Statada natijalar chiqishi uchun Courier New shriftidan foydalanilgan. Shrift hajmini moslash uchun 8 yoki 9 hajmgacha kamaytirishingiz kerak bo'ladi. Shunda natija qisman to'g'ri chiqishi mumkin:

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
years	18	2008.5	5.338539	2000	2017
Y	18	73.49739	75.99263	3.256	254.043
X1	18	19.23783	18.00062	.864	73.392
X2	18	17.95961	19.63861	.744	68.424
X3	18	13.18078	12.69302	.462	42.611

Belgilangan materialni rasm sifatida nusxalash sizning ma'lumotlaringizni hujjatga joylashtirilganda to'g'ri va aniq ko'rinishdagi rasm bo'lib joylashadi. Rasm sifatida joylashishi, bu sizning ma'lumotlaringizni tahrir qilish imkoniga ega bo'lmaysiz.

Birinchi bob mavzularini mustahkamlash uchun savollar:

1. Ekonometrikaning boshqa fanlar bilan o'zaro aloqasi haqida qanday tushunchaga ega bo'ldingiz?
2. An'anaviy ekonometrik metodologiyalar qaysi yo'nalishlarda olib boriladi?
3. Ekonometrik modellashtirish va modellarning ahamiyati haqida fikr yuriting.
4. Ekonometrikani asosiy maqsad va vazifalari haqida so'zlab bering.
5. Iqtisodiy model va ekonometrik modellarni ta'riflang.
6. Iqtisodiyotni ekonometrik modellashtirishning zarurligi qay darajada deb o'ylaysiz?
7. STATA dasturini ishga tushirish va o'chirish ketma-ketligini yozing.
8. STATA dasturida ma'lumotlarni kiritish va yuklash qay tartibda amalga oshiriladi?
9. Ma'lumotlar faylini tasvirlash uchun qanday buyruq kiritiladi?
10. Tasviriyl statistikaning sintaksis xususiyatlari haqida so'zlab bering.

Birinchi bob mavzularida qo'llanilgan buyruqlar va atamalar

cd	do selected	results window
cdf	do-file	review window
clear	do-file editor	summarize
command syntax	drop	summarize variable
command window	exit	summarize, detail
current path	factor variables	generate syntax
data browser	help	text format
data utilities	help command	two-way graph
definition files	label	use
density functions	log close	use "data file", clear
describe	log file	variable manager

dialog box, db
display

variables window
working directory

Birinchi bob mavzularini mustahkamlash uchun masalalarning do-filelari

Ma'lumotlar buyrug'i: cd c:\data\poe4stata

1-masala. log file ni ochish.

log using chap01, replace text

* Datalardan foydalanish

use cps4_small, clear

describe

* O'zgaruvchilarga izohlar yozish.

label variable wage "earnings per hour"

* Tasviriy statistik ma'lumotlarni olish.

summarize

summarize wage, detail

summarize if exper >= 10

summarize in 1/50

summarize wage in 1/50, detail

summarize wage if female == 1 in 1/500, detail

II BOB. REGRESSION MODELLAR TUZISH STATA DASTURIDA

2.1. Klassik chiziqli modelni yaratish uchun ma'lumotlarini yuklash

“Principles of Econometrics” nomli adabiyotning 4-nashrning I bobining bir necha rejalarida, uy xo'jaliklarining haftalik xarajatlari bilan daromadlari o'rtasidagi munosabatlarning iqtisodiy modeli ko'rsatib o'tilgan va shu asnoda stata dasturida oziq-ovqat sanoatining dinamik ma'lumotlari asosida model tuzishga harakat qilamiz. Birinchidan Stata dasturini ishga tushiramiz va ishchi katalogini kiritamiz. Buning qanday qilinishi kompyuterning operatsion tizimiga va stata dasturining do-file lari joylashgan joyiga bog'liq.

Ishchi katalog turi quyidagicha: `cd "C:\Users\User\Documents\stata"` yuqoridagi ishchi katalogni buyruqlar oynasiga kiriting va Enter ni bosing yoki Stata ochiladigan menyusidagi **File > Change Working Directory** katalogini tanlang.

Yangi muammoga duch kelsangiz, ochiq bo'lgan har qanday jurnal faylini o'chirib tashlashingiz kerak. Yangi bir stata log faylini boshlash uchun uskunalar paneli belgisiga bosing.



Bir kunlik faylingiz ochiq bo'lsa, sizga ba'zi imkoniyatlarni beradigan dialog paydo bo'ladi. Yangi jurnal faylini ishga tushirishdan avval siz eski faylni yopishingiz kerak yoki buyruq oynasida quyidagi buyruqni kiriting: **log close**

Stataning xotirasini tozalash uchun quyidagini kiriting: **clear**

2.1.1. Kundalik log faylni ishga tushirish va ochish

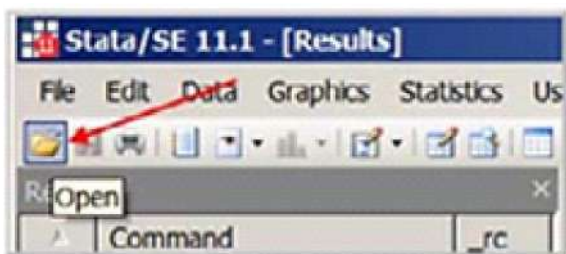
Stata log faylini ishga tushirish yoki yopish uchun uskunalar paneli belgisiga bosing yoki quyidagi buyruqni kiriting:

log using chap02, replace text

Ushbu **log** fayli mavjud katalogda matn formatida ochiladi. Variantni almashtirish **chap02.log** ning avvalgi versiyasiga yozilib, o'chiriladi.

Izoh: Foydalanuvchilar har bir bo'lim yoki bo'limning bir qismi uchun log faylini ochishlari kerak. Oldingi boblar uchun log fayllarini ochish va yopishni sizga eslatib qo'yamiz, lekin keyingi boblarda bajarib bo'lmaydi. Kundalik log fayllardan foydalanishni odat qiling.

Oziq-ovqat xarajatlari misoli haqidagi ma'lumotlar stata nomli faylida va food_exp faylida joylashgan. Stata ma'lumotlar faylini ochish uchun uskunalar panelidagi **Open** (foydalanish) tugmasini bosing.



Food_exp.dta toping, uni tanlang va **Open** tugmasini bosing yoki buyruqlar oynasida joriy papkadan ma'lumotlar faylini ochish uchun quyidagi buyruqni kiriting: **use food_exp**

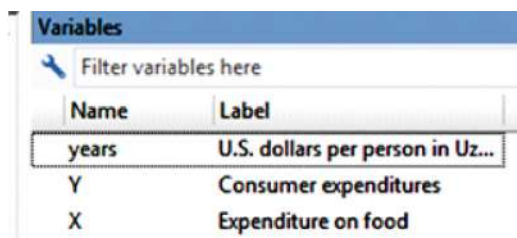
Agar **Stata** xotirasini bir vaqtning o'zida yangi ma'lumotlar faylini ochish hamda tozalash zarur bo'lsa, quyidagi buyruqni kiriting: **use food_exp, clear**

Oldindan ochilgan ma'lumotlar to'plamini xotiradan o'chiradi. Biroq, yangi ma'lumotlar faylini ochishdan oldin "tozalovchi" ma'lumot faylini amalga oshirish xavfsizroqdir.

Stata dasturida internet saytidan ma'lumotlarni yuklash ham mumkin. Buning uchun quyidagi buyruqni kiring:

use <http://www.stata.com/texts/s4poe4/food>

O'zgaruvchilar oynasida ikkita parametr ro'yxatga kiritiladi: Y va X izohi bilan birga. O'zgaruvchilar Turi va Format haqida boshqa ma'lumotlar ham paydo bo'lishi mumkin. Biz ustunlarni to'ldirish uchun tanladik.



Name	Label
years	U.S. dollars per person in Uz...
Y	Consumer expenditures
X	Expenditure on food

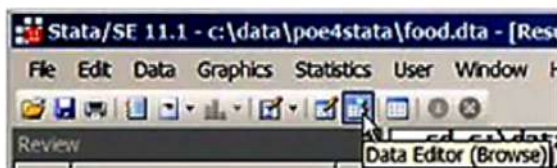
2.1.2. Ma'lumotlarni tasvirlash va ro'yxatlash

Har bir yangi masalani boshlashdan oldin ma'lumotni ko'rib tekshirib chiqish kerak. Buyruqlar oynasiga quyidagi buyruqni kiriting: **describe**

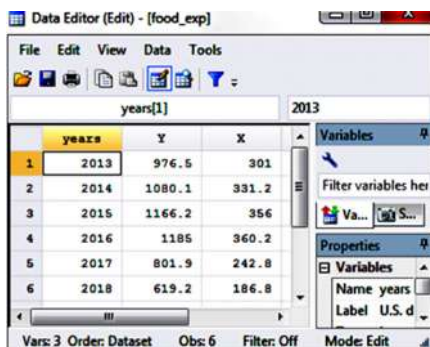
Ushbu o'zgaruvchilar haqida ko'proq ma'lumot olish uchun Buyruqlar oynasida **help describe** buyrug'ini kiriting. Oddiy sarlavha uchun hech narsa talab qilinmaydi, shuning uchun **OK** tugmasini bosing.

Contains data from food_exp.dta				
obs:	6			
vars:	3			5 Nov 2019 16:24
size:	60			
variable name	storage type	display format	value label	variable label
years	int	%8.0g		U.S. dollars per person in Uzbekistan
Y	float	%8.0g		Consumer expenditures
X	float	%8.0g		Expenditure on food

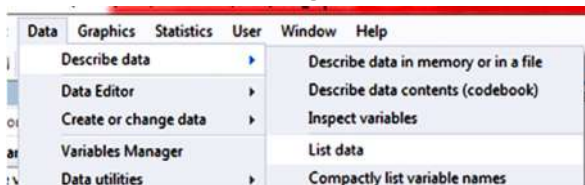
Yuqoridagi ma'lumotlar fayli **food_exp.dta** haqida umumiy ma'lumot. Biz ma'lumotlarimizni tekshiramiz. Ma'lumotlar brauzeri (**Data Editor**) dan foydalaning.



Ma'lumotlar qiymatlarini ko'rish imkonini beruvchi elektron ko'rinishi ochiladi.



Data Editor (Browser)ni yopish uchun yuqoridagi "x" tugmasini bosing. Agarda siz u ma'lumotlarni chop qilishni yoki ba'zilarini ro'yxatlamochi bo'lsangiz yuqoridagi menu oynasidan **Data > Describe data > List data** ni bosing.



Ochilgan dialog oynasida, o'zgaruvchini tanlang va Natijalar oynasida barcha ma'lumotlarni ro'yxatlash uchun OK tugmasini bosing. Stata buyrug'ini ro'yxatlash. Ro'yxatlash buyrug'i sintaksisi quyidagicha: **list [varlist] [if] [in] [, options]**

Muayyan o'zgaruvchilar qiymatlarini ro'yxatlash uchun o'zgaruvchilarning nomlarini kiriting. Ro'yxatlangan qiymatlar oralig'i muayyan chiziqlarni ko'rsatish uchun mantiqiy "if" yoki "in" yordamida o'zgartirilishi mumkin. Misol uchun:

list in 1/5

list Y in 1/5

list Y if X <= 360

Natijalar oynasida quyidagilarni ko'rsatadi:

list in 1/5

list Y in 1/5

list Y if X <= 360

	years	Y	X
1.	2013	976.5	301
2.	2014	1080.1	331.2
3.	2015	1166.2	356
4.	2016	1185	360.2
5.	2017	801.9	242.8

	Y
1.	976.5
2.	1080.1
3.	1166.2
4.	1185
5.	801.9

	Y
1.	976.5
2.	1080.1
3.	1166.2
5.	801.9
6.	619.2

Natijalar oynasi to'lsa, siz **more** so'zini ko'rasiz va ko'proq ma'lumotni ko'rish uchun uni bosishingiz kerak, shunda barcha ma'lumotlarni topa olasiz. Bu **more** so'zi pauzani bildiradi. Yana bosing yoki bo'sh joyni bosing. **Stata log** faylini ishga tushirgandan keyin **set more off** buyrug'ini kiritgan bo'lsa, pauza xususiyatini o'chirib qo'yadi.

2.2. Ma'lumotlarning tahliliy statistikasi

Keling, ma'lumotlarning ta'rif faylida bildirilgan bir xil xulosa statistik qiymatlarga ega ekanligini tekshiring. Yuqoridagi menyulardan foydalanib ishlating:

Statistics > Summaries, tables, and tests > Summary and descriptive statistics > Summary statistics

Natijada muloqot oynasidagi ma'lumotlar to'plamidagi barcha o'zgaruvchilar bo'yicha abstrakt statistika uchun o'zgaruvchilarni tanlang va OK tugmasini bosing. Bundan tashqari, dialog oynasini ochish uchun **db summarize** yoki **db su** buyruqlarini kiritishingiz ham mumkin, shunda dialog oynasi paydo bo'ladi yoki buyruqlar oynasida quyidagi buyruqni kiriting:

Summarize

summarize

Variable	Obs	Mean	Std. Dev.	Min	Max
years	6	2015.5	1.870829	2013	2018
Y	6	971.4833	222.7953	619.2	1185
X	6	296.3333	68.90558	186.8	360.2

Xulosa statistika buyrug'i sintaksisi quyidagicha: **summarize [varlist] [if] [in] [weight] [, options]**

Asosiy variant bizga batafsil abstrakt statistikani olish imkonini beradi. Stata buyrug'i oynasida quyidagini kiriting: **summarize Y, detail**

```
. summarize Y, detail
```

Consumer expenditures				
Percentiles		Smallest		
1%	619.2	619.2		
5%	619.2	801.9		
10%	619.2	976.5	Obs	6
25%	801.9	1080.1	Sum of Wgt.	6
50%	1028.3		Mean	971.4833
		Largest	Std. Dev.	222.7953
75%	1166.2	976.5		
90%	1185	1080.1	Variance	49637.74
95%	1185	1166.2	Skewness	-.5982541
99%	1185	1185	Kurtosis	1.936811

Natijalar oynasida ma'lumotlarni eng ko'p miqdori, eng kichik va eng katta kuzatishlar, kuzatuvlar soni (Obs) va e'tiborga olinmaydigan WGT summasi ko'rsatilgan. Stata siz tushunmagan ko'p narsalarni e'lon qiladi. Haqiqat siz bilgan narsalarni aniqlashga qodir. Misol uchun, natijalar o'z ichiga quyidagilarni oladi:

Mean 971.4833
Std. Dev. 222.7953

Variance 49637.74

Food_exp o'zgaruvchisi uchun xulosa statistikasi:

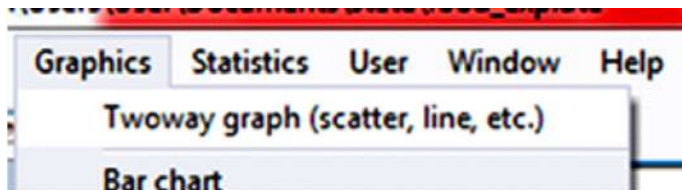
O'rtacha qiymat,
$$\bar{y} = \sum y_i / N$$

Std. Dev. Variantsning kvadrat ildizlari bo'lgan namunali standart og'ish:

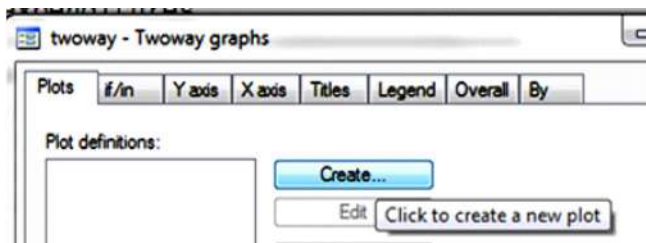
Oddiy variatsiyaning turi, $var(y) = \sum (y_i - \bar{y})^2 / N - 1$
Skewness va Kurtosis qiymatlari keyinroq muhokama qilinadi.

2.3. Ma'lumotlar asosida scatter diagrammasini yaratish usullari

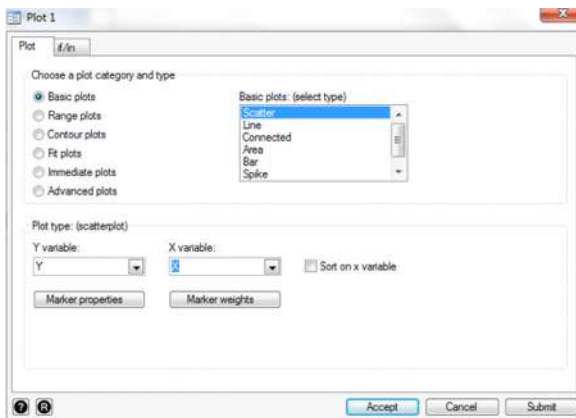
Oddiy regressiya modelida ma'lumotlar qiymatlarini Scatter diagrammasida chizish muhimdir. Stata ochiladigan menyusida **Graphics> Twoway graph (scatter, line, etc.)** ni tanlang. Qo'shimcha tafsilotlarni bilish uchun buyruqlar oynasiga **help twoway** buyrug'ini kiriting.



Muloqot oynasida **Create** tugmasini bosing.



Natijada paydo bo'lgan dialog oynasida **Basic plots**, **Scatter** ni tanlab, keyin ochiladigan o'qlarni foydalanib, Y o'zgaruvchisini (vertikal o'q) va X o'zgaruvchisini (gorizontal o'q) tanlang.



Submit tugmasini bosganingizda Scatter diagrammasi yaratiladi. Stata buyruqlar oynasiga quyidagi buyruqni kiritib yaratsa ham bo'ladi: **twoway (scatter Y X)**

Agar **Accept** tugmasini bossangiz Plot-1-buyruq belgilari oynasida paydo bo'ladi va **OK** tugmasini bosganingizda grafika yaratiladi.

Grafikni stata grafik kengaytmasi ***.gph** dan foydalanib, ya'ni diskda standart katalogga saqlash uchun: **graph save food1, replace**

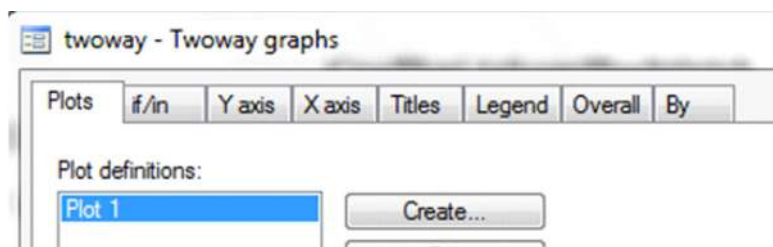
Grafikni va saqlashni save opsiyasidan foydalanib, bir qadamda bajarish mumkin: **twoway (scatter Y X), saving(food1, replace)**

Diskni saqlashning o'rniga xotiraga nom variantini ishlatishingiz ham mumkin. Agar siz laboratoriya sharoitida bo'lsangiz, bu qulay bo'lishi mumkin.

twoway (scatter Y X), name(food1, replace)

Grafikni takomillashtirish

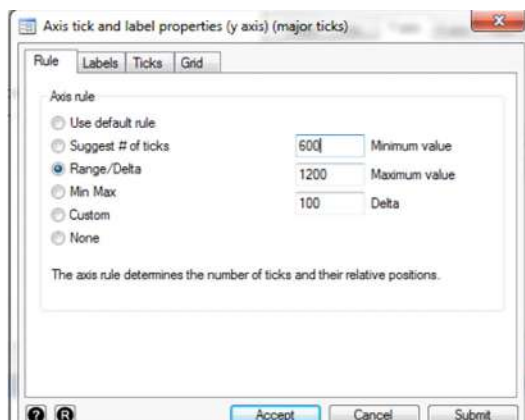
Grafikni kuchaytirish uchun 1-satrd dialog oynasida Accept ni bosing. Grafika dialog oynasida bu Plot 1 deb ataluvchi plot ta'rifi yoki profilini yaratadi.



Y o'qi yorlig'ini bosing. Natijada dialog oynasidan quyidagini tanlang.

Major tick/label properties

Bir nechta variant mavjud, ammo vertikal o'qning oralig'ini belgilaylik.

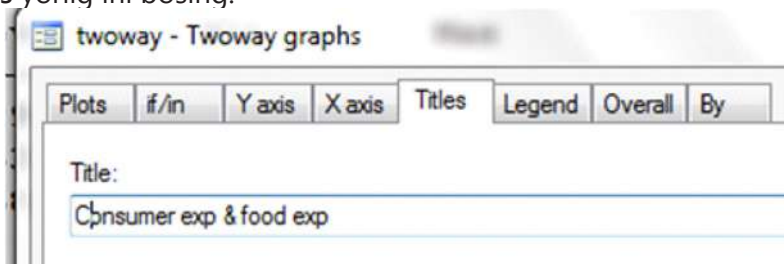


Range / Delta ni bosing. Xulosa statistikasidan **Y** ning minimal va maksimal qiymatlarini ko'rgandik. Ma'lumotlar butunligini ko'rishimiz uchun Minimal qiymat 600 va Eng yuqori 1200 qiymatini kiriting. **Delta** o'qlar uchun o'lchov birliklari belgilanadigan belgilar orasidagi bo'shliq. Ushbu qiymatni 100 ga to'g'rilang. **Accept** tugmasini bosing.

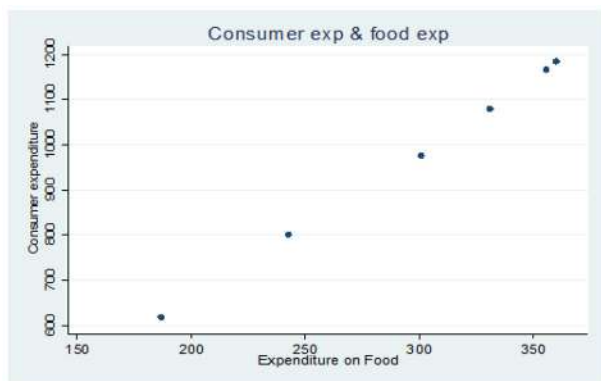
Ushbu jarayonni **X** o'qi uchun maksimal qiymat 370 va **Delta** 50 dan foydalanib takrorlang. **Accept** tugmasini bosing.



Grafikka nom qo'shish uchun **Twoway Graph** dialog oynasidagi **Titles** yorliq'ini bosiq.



Grafik yaratish va oynani yopish uchun **OK** tugmasini bosiq. Bu Y o'qi va X o'qi bilan chiroyli nuqtalarni ishlab chiqaradi va ma'lumotlar to'plamidagi o'zgaruvchilar izohlari bilan belgilanadi.



Ushbu nuqtali grafikni yaratish uchun foydalaniladigan Stata buyrug'i quyidagicha:

```
twoway (scatter Y X), ytitle(Consumer expenditure) ///  
ylabel(600(100)1200) xtitle(Expenditure on Food) ///  
xlabel(150(50)370) title(Consumer exp & food exp)
```

Ushbu buyruqdagi "///" belgilar keyingi qatordan buyruqni davom ettiradi. Buyruqlar uzunligi yoki buyruqning muayyan qismlaridan keyin ularga izohlar kiritilishi foydalidir. Fikrlarni ushbu belgilar orqali "/" * * / " ifodalanadi va bu foydalidir, chunki u har qanday joyga joylashtirilishi mumkin va tarkib Stata dasturi tomonidan e'tiborga olinmaydi. Muqobil sharhning buyruqlarini muhokama qilish uchun **help comments** buyrug'ini kiriting.

twoway (scatter Y X) oddiy nuqtali grafik uchun ishlatiladigan buyruqlar.

Bu vergul muhim ahamiyatga ega va qo'llaniladigan variantlarni bildiradi.

ylabel(600(100)1200) Y o'qi oralig'ini, 600 dan 1200 gacha va 100 asosiy nuqtalar oralig'ini bildiradi.

xlabel(150(50)370) 150-370 oralig'ini belgilaydi va X o'qi uchun 50 sonini oralig'i hisoblanadi.

title(Consumer exp & food exp) asosiy grafik nomini bildiradi.

Yana bir marta siz ikkita buyruqqa saqlash parametrini qo'shishingiz yoki xotiraga saqlash uchun variant nomini ishlatishingiz yoki grafikni saqlash buyrug'idan foydalanishingiz mumkin. Grafikni saqlash buyruqlaridan ushbu bobda faylda foydalanamiz.

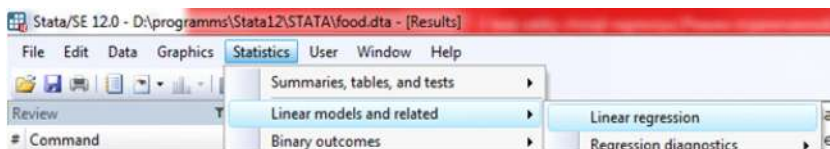
2.4. STATA dasturida regression model yaratish usuli

Oddiy chiziqli regressiya modeli:

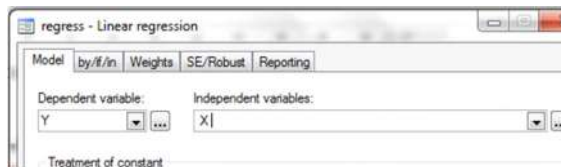
$$y = \beta_1 + \beta_2 x + e$$

Asosiy omil, y (**consumer expenditure**) va ta'sir etuvchi omil x (**expenditure on food**) bo'yicha ma'lumot berilgan holda, noma'lum parametrlarni topish va baholash uchun Stata dasturidan foydalanamiz. Regression tahlilni amalga oshirish uchun yuqoridagi menyudan foydalanilgan holda quyidagicha amalga oshiramiz:

Statistics > Linear models and related > Linear regression



Regress - Linear regression muloqot oynasida **Y** ga bog'liq bo'lgan o'zgaruvchini tanlang. Bu esa chap tomonda joylashgan regressiya modelidagi asosiy omildir. Ta'sir etuvchi (o'ng tomon) o'zgaruvchisi sifatida **X** ni tanlang (yoki kiriting). So'ngra **OK** tugmasini bosing. Stata dasturi avtomatik tarzda hisob-kitoblarni amalga oshiradi.



Shu bilan bir qatorda, quyidagi buyruq orqali amalga oshirish ham mumkin: **regress depvar [indepvars] [if] [in] [weight] [options]**

Help regress buyrug'i orqali qo'shimcha ma'lumotga ega bo'lasiz. Bu shuni ko'rsatadiki, asosiy omil **depvar** birinchi bo'lib joylashtiriladi, keyin ta'sir etuvchi o'zgaruvchi yoki o'zgaruvchilar (ko'p ommilli regressiya uchun), **indepvars** lar kiritiladi. Agar xohlasangiz siz **if** yoki **in** bilan cheklash, kuzatishlardagi hajmini belgilash yoki ulardan foydalanish imkoniyatlarini qo'llashingiz mumkin. Ushbu variantlarni (**if** va **in** buyruqlarini) kelgusi boblarda ko'proq muhokama qilamiz.

Iste'mol va oziq-ovqat xarajatlari uchun oddiy regressiya buyrug'i mavjud: **regress Y X**

buni qisqartirilishi ham mumkin: **reg Y X**

Natijalar oynasida regress uchun Stata buyrug'i berilgan va **regression** natijalari ko'rsatiladi. Bu narsalarni hammasini o'rganamiz, ammo hozircha **Y** va **Coef** deb nomlangan birinchi ikkita ustunni ko'rib chiqamiz.

. regress Y X

Source	SS	df	MS	Number of obs	=	6
Model	248002.086	1	248002.086	F(1, 4)	=	5315.64
Residual	186.620714	4	46.6551785	Prob > F	=	0.0000
				R-squared	=	0.9992
				Adj R-squared	=	0.9991
Total	248188.707	5	49637.7414	Root MSE	=	6.8305

Y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
X	3.232126	.0443313	72.91	0.000	3.109042 3.355209
_cons	13.69666	13.42954	1.02	0.365	-23.58972 50.98304

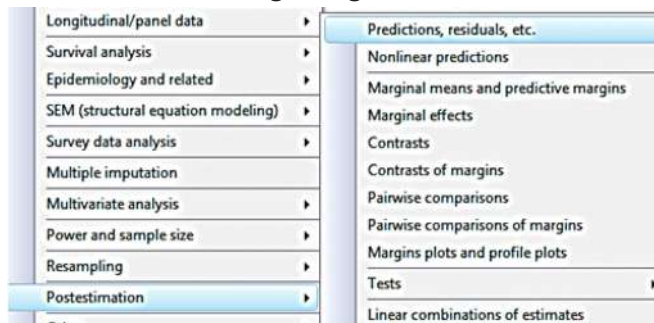
Birinchi ustunda o'zgaruvchilarning nomlari berilgan. Kompyuter dasturlari biz " β_1 " va " β_2 " deb chiziqli regressiya tenglamasida o'zgaruvchilarning nomlarini belgilab olganimizni bilmaydi va shuning uchun " β_1 " va " β_2 " lar o'rniga o'zgaruvchilarning nomlari bilan ifodalaydi. Stata birinchi navbatda tenglama parametrlarni **Coef.** so'zi ostida beradi. **Y** endogen **cons**, ya'ni "doimiy" yoki "asosiy omil" deb nomlanadi, bu **y endogen** uchun yana bir umumiy nom.

2.4.1. Hisoblangan qiymat (Y) va qoldiqlar

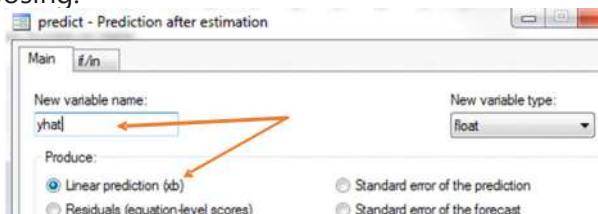
Hisoblangan y qiymatlari va qoldiqlarni "postestimation" buyruqlaridan foydalanib topiladi. Ularni "postestimation" parametri deb atashadi, chunki ular regressiya modelini baholashadi. Bunga qo'shimcha ma'lumotni, **help regress** buyrug'ini kiritgandan so'ng **postestimation** parametrlari havolasini oynaning o'ng tomoni yuqorisida va burchagida ko'rishingiz mumkin.



Stata dasturidagi yuqoridagi **menu** lar orqali bajarish yo'li:
Statistics > Postestimation > Predictions, residuals, etc.
Statistics menyusini bosing so'ngra:



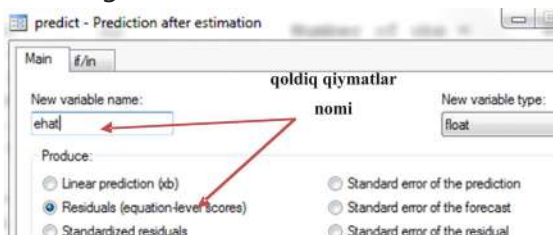
paydo bo'lgan muloqot oynasida bir nechta muqobillar mavjud. Hisoblangan qiymatni ($\hat{y}$) hisoblash uchun **Linear prediction** ni belgilang va hisoblangan qiymatlar nomini kiriting, so'ngra **Ok** tugmasini bosing.



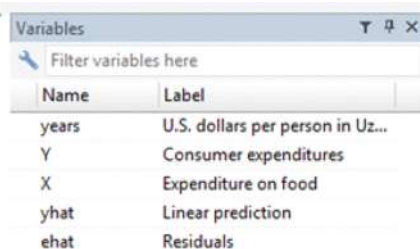
Qoldiqlarni olish uchun:

$$\hat{e}_i = y_i - \hat{y}_i = y_i - b_1 - b_2 x_i$$

dialog oynasini qayta oching, **Residuals (equation-level scores)** ni belgilang va qoldiqlarning nomini kiriting, ya'ni **ehat** deb so'ngra **Ok** tugmasini bosing.

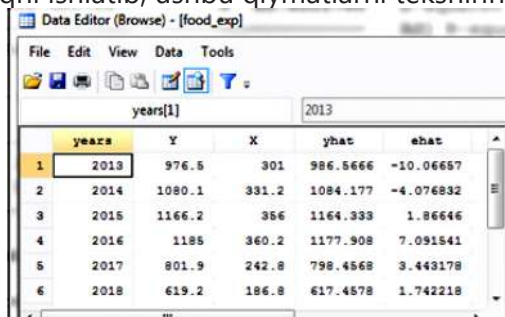


Ushbu buyruqlar o'zgaruvchilari oynasida ko'rinadi, ya'ni ikkita yangi o'zgaruvchi, **yhat** va **ehat** paydo bo'ladi.



Name	Label
years	U.S. dollars per person in Uz...
Y	Consumer expenditures
X	Expenditure on food
yhat	Linear prediction
ehat	Residuals

Ma'lumotlar brauzerini (**Data editor (Browse)**) tanlab, yoki quyidagi buyruqni ishlatib, ushbu qiymatlarni tekshiring: **browse**



	years	Y	X	yhat	ehat
1	2013	976.5	301	986.5666	-10.06657
2	2014	1080.1	331.2	1084.177	-4.076832
3	2015	1166.2	356	1164.333	1.86646
4	2016	1185	360.2	1177.908	7.091541
5	2017	801.9	242.8	798.4568	3.443178
6	2018	619.2	186.8	617.4578	1.742218

Hisoblangan $\hat{y}$ qiymatlari va qoldiqlarni hisoblaydigan Stata komandalari asosiy **predict** buyrug'i yordamida taxmin qilinadi. Ma'lumotlar brauzerini yoping va buyruqlar oynasiga quyidagini kiriting:

help predict

Predict buyrug'ining asosiy sintaksi:

predict [type] newvar [if] [in] [, single_options]

Nimalar kerak: o'zgaruvchining nomi, newvar va turi. Hisoblangan $\hat{y}$ qiymatlarini olish uchun "**x** marta **b**" uchun qisqa bo'lgan **xb** qisqartmani ishlatib: **predict yhat, xb**

Bashoratlarni olish, aslida, bu buyruq uchun amrimahol, shuning

uchun biz quyidagidan foydalanganimiz afzal: **predict yhat**

Qolgan qoldiqlarni olish uchun options ga residualsdan foydalaning. Oziq-ovqat sarf-xarajatlari modeli uchun bu: **predict ehat, residuals**

Qoldiqlarni bir necha turlarda yozish mumkin, minimal darajasiga qisqartmasi **r** yoki **res**, yoki **resid** ga o'xshab biroz qisqartirilishi mumkin.

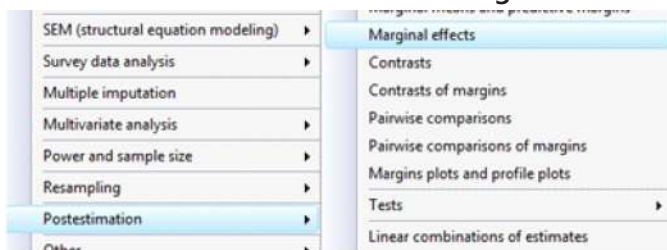
2.4.2. Elastiklikni hisoblash

Parametr taxminlarini va o'zgaruvchilar uchun umumiy statistikani hisobga olgan holda, iste'mol sarf-xarajatlarning oziq-ovqat xarajatlariga nisbatan moslashuvchanligi kabi boshqa miqdorlarni osonlikcha hisoblashimiz mumkin:

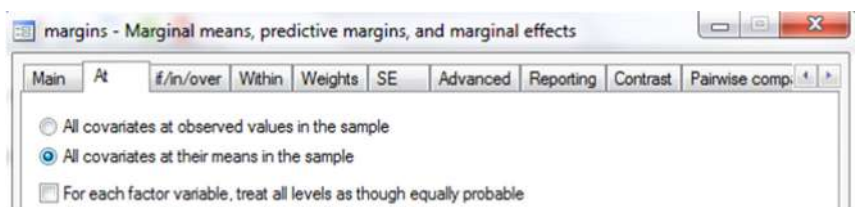
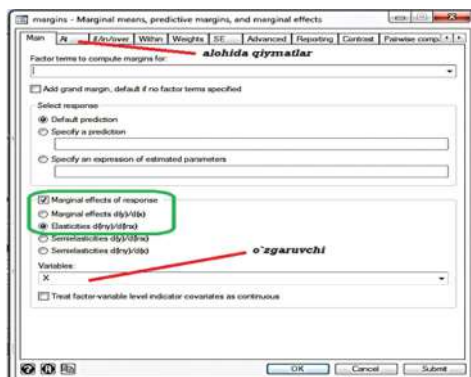
$$\hat{\varepsilon} = b_2 \cdot \frac{\bar{x}}{\bar{y}} = 3.2321 \times \frac{296.33}{971.48} = 0.98$$

Stata dasturining **postestimation** komandalaridan biri bu elastiklikni avtomatik ravishda hisoblash imkonini beradi. Bu quyidagicha:

Statistics > Postestimation > Marginal effects.



Ochilgan muloqot oynasida **Elasticities** va **Variable** ni tanlang. Bizning oddiy regressiya modelimizda bitta o'zgaruvchi, ya'ni **X** ni kiriting. Elastiklikni baholash uchun **At menu** sini tanlang va **All covariates at their means in the sample** ni belgilang.



Natijalar oynasida biz shartli marginal ta'sirlarni ko'ramiz:

```
. margins, eyex( X) atmeans
```

Conditional marginal effects Number of obs = 6
Model VCE : OLS

Expression : Linear prediction, predict()
ey/ex w.r.t. : X
at : X = 296.3333 (mean)

	Delta-method				
	ey/ex	Std. Err.	z	P> z	[95% Conf. Interval]
X	.9859013	.0138154	71.36	0.000	.9588236 1.012979

Margins dialog oynasi juda kuchli va siz tushunmaydigan hisob-kitoblarni amalga oshiradi. Biz uni bir nechta nuqtalar qilish uchun taqdim etamiz. Iqtisodiy tahlilda bir o'zgaruvchining natijasi boshqa bir o'zgaruvchiga ta'sirini ko'rasiz, shuningdek barchasi o'zgaras bo'lsa, bu **marginal effect** deb ataladi. Regression modeldan, bu qiyalik parametridir va oddiy regresion modeli hosilasi dy/dx hisoblanadi. Stata bu qiymatni murakkab modellarda ham hisoblab chiqadi va biz ushbu xususiyatga keyinroq ko'rib chiqamiz.

Linear (to'g'ri chiziqli) regressiya modelida elastiklik quyidagicha:

$$\varepsilon = \frac{\Delta y/y}{\Delta x/x} = \frac{\Delta y}{\Delta x} \cdot \frac{x}{y}$$

Deltani lotin shaklida "d" harfi bilan o'zgartiring:

$$\varepsilon = \frac{\Delta y/y}{\Delta x/x} = \frac{\Delta y}{\Delta x} \cdot \frac{x}{y}$$

To'g'ri chiziqli regressiya modelida $d \ln(y) = dy/y$ va $d \ln(x) = dx/x$ qilib hisoblanadi:

$$\varepsilon = \frac{d \ln(y)}{d \ln(x)} = \frac{dy/y}{dx/x}$$

Regressiyadan keyin ushbu elastiklik uchun Stata post-estimation buyrug'i quyidagicha: **margins, eyex(income) atmeans**

Stata faqat elastiklik haqida xabar bermasdan, shuningdek bu miqdor uchun standart xatolikni va ishonch oralig'ini hisoblaydi. Bu miqdor hisoblab chiqilgan oziq-ovqat mahsulotlariga xarajat qiymatini ko'rsatadi, bu esa namunadagi o'rtacha qiymatdir. Odatda moslashuvchanlik, "vositalar nuqtasida" $(x,y)=(296.33;971.48)$ da hisoblab chiqiladi, chunki u regression yo'nalishdagi vakolatli nuqtadir.

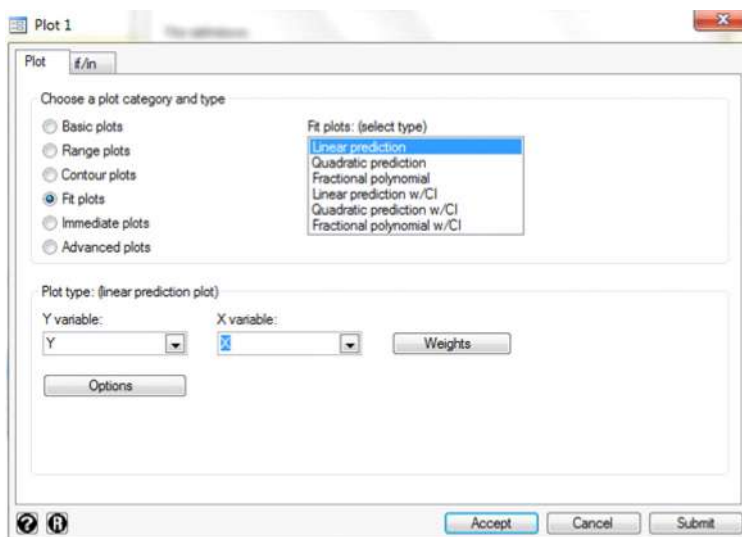
Bir nuqtada elastiklikni hisoblash o'rniga, y va x ning har bir qiymatida moslashuvchanlikni topishimiz va keyinchalik bu miqdorni barcha kuzatishlar bo'yicha o'rtacha hisoblashimiz mumkin. Bunga o'rtacha marginal ta'sir (**Average marginal effect**) deyiladi, bu holda o'rtacha elastiklik quyidagicha:

$$AME = \bar{\varepsilon} = \frac{1}{N} \sum_{i=1}^N b_2 \frac{x_i}{y_i}$$

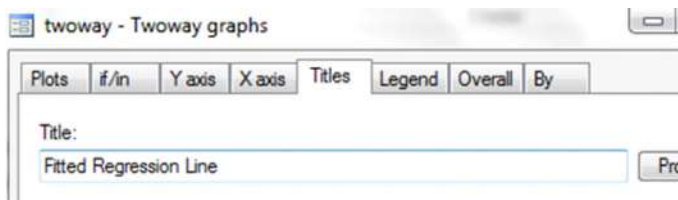
Stata dasturi quyidagicha hisoblanadi:

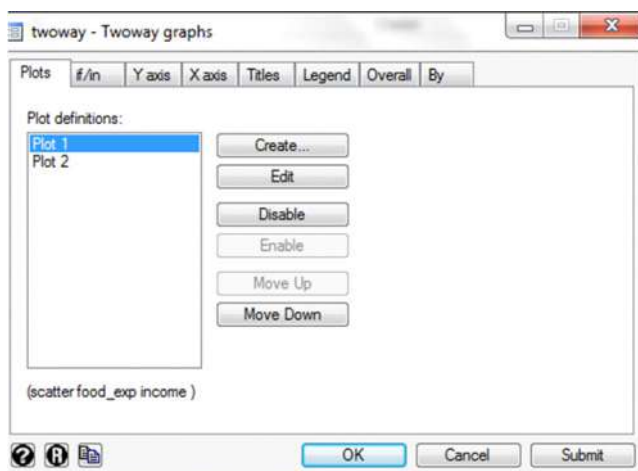
$$\widehat{AME} = \frac{1}{N} \sum_{i=1}^N b_2 \frac{x_i}{\hat{y}_i} = \frac{1}{N} \sum_{i=1}^N b_2 \left(\frac{x_i}{b_1 + b_2 x_i} \right)$$

(scatter, line, etc.) tanlang. Agar siz bu mashgʻulotni davom ettirmoqchi boʻlsangiz va nuqtali grafikni 2.3-boʻlimga kiritgan boʻlsangiz, u holda 1-grafikda nuqtali diagrammasi allaqachon mavjud. Biz chiziqli prognoz uchun yangi grafik yaratamiz. **Create** tugmasini bosing. **Plot 2** dialog oynasida **Fit plots** tanlang va **Linear prediction** bilan birgalikda **Y** oʻzgaruvchi (**consumer exp**) va **X** oʻzgaruvchi (**food exp**)sini kiriting.

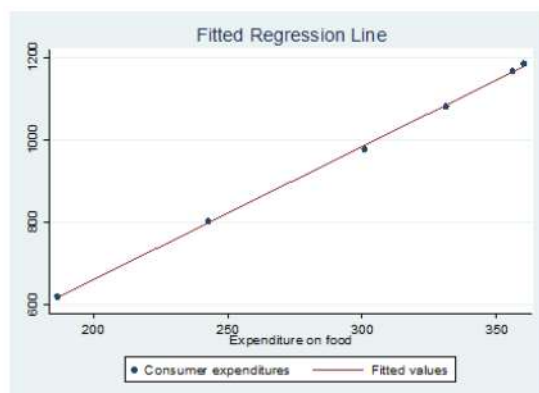


Accept tugmasini bosing. **Titles** soʻzini ustiga bosib, **Plot-2** ni nomni oʻzgartiring, yaʼni yangi nom kiriting. **Plots** menu siga qayting. **Submit** tugmasini bosish orqali grafikani yaratadi va grafikni koʻrsatadi.





Qachonki bir nechta grafiklar mavjud va xotirada bo'lsa, Stata bir grafini boshqasiga joylashtiradi. Tayyorlangan grafik:



Buyruqlar oynasida biz quyidagi buyruqni kiritsa ham bo'ladi:

```
twoway (scatter Y X)          ///      /* basic plot control */
      (lfit Y X),             ///      /* add linear fit */
      ylabel(600(100)1200)    ///      /* label Y axis */
      xlabel(150(50)375)      ///      /* label X axis */
      title(Fitted Regression Line) /*      graph title */
```

2.4.4. Variatsiya va kovarvatsiyani hisoblash

Regressiya koeffitsiyentlaridan tashqari yana bir muhim parametr – bu variatsion xatolik atamasi:

$$\text{var}(e_i) = \sigma^2 = E[e_i - E(e_i)]^2 = E(e_i^2)$$

Ushbu parametрни kiritish quyidagicha:

$$\hat{\sigma}^2 = \frac{\sum \hat{e}_i^2}{N - 2}$$

Yuqoridagi $\hat{e}_i = y_i - \hat{y}_i = y_i - b_1 - b_2 x_i$ bu soʻnngi kvadrat qoldiqlari. "2" soni – bu β_1 va β_2 regression tenglamaning parametrlari soni. Stataning regressiya buyrugʻidagi natijaning yuqori qismi, yaʼni quyidagi bu jadval **Analysis of Variance** berilgan:

Source	SS	df	MS
Model	248002.086	1	248002.086
Residual	186.620714	4	46.6551785
Total	248188.707	5	49637.7414
<i>Kvadrat qoldiqlar summasi</i>		<i>N-2</i>	<i>Xatolikning variatsion hisobi</i>

Residual qatoriga eʼtibor bering. **SS** bilan belgilangan ustunda kvadrat yigʻindisi mavjud. 186.620714 qiymati kvadratlar qoldiqlarining yigʻindisidir:

$$SSE = \sum_{i=1}^N \hat{e}_i^2 = SS \text{ Residual}$$

df bilan belgilangan ustun bu (residual) erkinlik darajasi boʻlib, u bu holda $N-2 = 4$ boʻladi. Modelning erkinlik darajalari 1, qaysiki biri oʻzgaruvchi soni boʻlsa, boshqa biri bu parametrlar soni, yaʼni β_2 . Ustun boshi, yaʼni **MS** oʻrtacha kvadratni anglatadi. Kvadrat qoldiqlarining oʻrtacha qiymati taxminiy xato variatsiyasidir:

$$\hat{\sigma}^2 = \frac{\sum \hat{e}_i^2}{N - 2} = MS \text{ Residual} = 46.6551785$$

Jadvalning boshqa qismlarini keyingi boblarda tushuntiriladi.

Regressiya tenglamasini tuzgandan so'ng, bu regressiyaga bog'liq bir necha statistik hisob-kitoblar bor, lekin Stata dasturi ularni hisoblab ko'rsatmaydi. Oziq-ovqat mahsulotlarini sarf qilish modeli uchun eng kichkina kvadratlar variatsiyasi va kovarvatsiya:

$$\widehat{var}(b_1) = \hat{\sigma}^2 \left[\frac{\sum x_i^2}{N \sum (x_i - \bar{x})^2} \right] = 180.35254$$

$$\widehat{var}(b_2) = \frac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2} = 0.00196526$$

$$\widehat{cov}(b_1, b_2) = \hat{\sigma}^2 \left[\frac{-\bar{x}}{\sum (x_i - \bar{x})^2} \right] = -0.58237347$$

Bularni **estat** buyrug'i yordamida ko'rish ham mumkin. Buyruqlar oynasiga quyidagini kiriting: **help estat**

Variatsiyani va kovarvatsiani hisoblash uchun quyidagini kiritish kerak: **estat vce**

Natijada, regressiya koeffitsiyentlarining hisoblangan variatsiyasi diagonaliga joylashgan va regressiya koeffitsiyentlarining hisoblangan kovarvatsiyasi "diagonal bo'lmagan" mavjud jadval ko'rsatadi. Stata hisoblagan matritsasining yuqori chap burchakdagi b_2 parametrining variatsiyasidir.

$$\text{Stata covariance matrix} = \begin{bmatrix} \widehat{var}(b_2) & \widehat{cov}(b_1, b_2) \\ \widehat{cov}(b_1, b_2) & \widehat{var}(b_1) \end{bmatrix}$$

```
. estat vce
```

Covariance matrix of coefficients of **regress** model

	e (V)	X	_cons
X		.00196526	
_cons		-.58237347	180.35254

Variatsiyaning kvadrat ildizlar koeffitsiyentlari standart xatolikning hisoblangan koeffitsiyentidir:

$$se(b_1) = \sqrt{\widehat{var}(b_1)} = 13.42954$$

$$se(b_2) = \sqrt{\widehat{var}(b_2)} = 0.0443313$$

Ular regressiya natijasida ko'rsatiladigan, ya'ni Stata dasturi tomonidan avtomatik ravishda chiqariladigan **Std.Err** deb izohlanadi:

Y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
X	3.232126	.0443313	72.91	0.000	3.109042	3.355209
_cons	13.69666	13.42954	1.02	0.365	-23.58972	50.98304

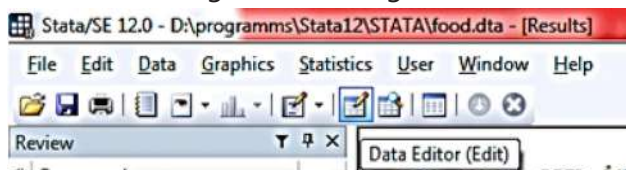
2.5. Regression modelning prognoz qiymatlarini olishda Stata dasturidan foydalanish

Iste'mol va oziq-ovqat sarf-xarajatlari modeliga asoslanib, oziq-ovqat xarajatlari uchun yilda 20 dollar xarajat keltiradigan taxminiy qiymatini olaylik:

$$\hat{y}_i = 13.696 + 3.21x_i = 13.696 + 3.21(20) = 77.896$$

Ma'lumotlar fayliga to'liq bo'lmagan kuzatuv qo'shish orqali

biz Stata dasturini ishlashda davom ettiramiz. Stata uskunalari panelidan **Data Editor** tugmasini bosib.



Bu ma'lumotlar bazasi kabi elektron jadval ko'rinishini ochib beradi, xuddi **Data Browser** ga o'xshash, lekin bu **Data Editor** farq qiladi, unda ma'lumotlarni o'zgartira olamiz. 6-, ya'ni oxirgi kuzatish natijasi uchun pastga tushing, 7-qatorda daromad uchun yachyekani tanlang. **"Formula bar"** oynasida 20 qiymatini kiriting va **Enter** tugmasini bosib.

	years	Y	X	yhat	
1	2013	976.5	301	986.5666	-10
2	2014	1080.1	331.2	1084.177	-4.
3	2015	1166.2	356	1164.333	1
4	2016	1185	360.2	1177.908	7.
5	2017	801.9	242.8	798.4568	3.
6	2018	619.2	186.8	617.4578	1.
7	.	.	20	.	.

Data Editor endi daromadning 7-kuzatish natijasi uchun 20 qiymatga ega ekanligini ko'rsatadi, ammo boshqa qiymatlar "." ekanini ko'rsatadi. Stata da "." belgisi bir davr raqamli ma'lumotlar qiymati yo'qligini bildiradi. **Data Editor** oynasini yopish uchun x tugmasini bosib. Shu bilan Stata buyruqlar oynasiga quyidagi buyruqlar orqali amalga oshirsa ham bo'ladi:

edit

set obs 7

replace X = 20 in 7

Ushbu set buyrug'i ko'plab tizim parametrlarini boshqaradi, kuzatishlar sonini qo'shadigan yordamchi buyruq ham hisoblanadi.

Buyruqning joyi **Data > Create or change data > Change contents of variable** orqali ham amalga oshirish mumkin. **Replace** buyrug`ining sintaksisi: **replace oldvar = exp [if] [in] [, nopromote]**

X ning 7-qatoriga qiymati 20-sonli qiymatni joylashtirdik. Bu buyruqlar haqida ko`proq ma`lumot olish uchun **help set** va **help replace** orqali ega bo`lishingiz mumkin.

Bashoratli yoki hisoblangan qiymatlarni allaqachon **yhat** deb topib olgandik, shuning uchun biz "**yhat0**" deb nomlab, qiymatni topamiz. **Predict** buyrug`idagi **xb** parametrini tushirib qoldirib topishimiz mumkin: **predict yhat0**

Kuzatishdagi 7-X va yhat0 ma`lumotlarini ro`yxatlash: **list X yhat0 in 7**

Natijalar oynasida ko`rishingiz mumkin:

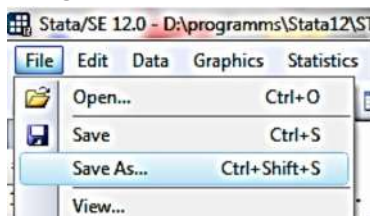
```
. list x yhat0 in 7
```

	x	yhat0
7.	20	78.33918

Har yilda \$ 20 oziq-ovqat mahsulotlar xarajati uchun kutilgan yillik iste`mol xarajatlari **yhat0** qiymatida hisoblab ko`rsatilgan.

Stata dastur ma`lumoti faylini saqlash

Stata ma`lumot faylini saqlash uchun **File > Save as** ketma-ketligini amalga oshiring:



Ma`lumotlar faylini **chap02.dta** kabi yangi nom bilan saqlang. Stata dasturida quyidagi buyruq orqali amalga oshiriladi: **save chap02.dta**

Oxirgi qadam sifatida **log** faylini yopasiz: **log close**

2.6. Chiziqsiz regression modellarni yaratish usullari

Chiziqli regression modeli chiziqsiz, egri chiziqli, munosabatlarni kiritish uchun foydalanish mumkin. Bu muammo emas, chunki chiziqli regressiya parametrlarni chiziqsiz bo'lmagan tarzda kiritilmaydigan modelga ishora qiladi. Chiziqli regressiya $y = \beta_1 + \beta_2 x + e$ Y va x o'zgaruvchilar boshqa o'zgaruvchilardan iborat bo'lishi mumkin. Chiziqsiz regressiya $y = \exp(\beta_1 + \beta_2 x) + e$, chunki parametrlar chiziqsiz parametr sifatida ishtirok etadi, xuddiki eksponensial shaklda.

Ikkita o'zgaruvchi polinomiya va logarifmik atamalar sifatida o'zgarib ishlatiladi. Ushbu bo'limda kvadratik va log-linear modellarni ko'rib chiqamiz.

2.6.1. Kvadratik model

Biz haqiqiy estate data ma'lumotlarini qo'llash orqali biz kvadratik modelni hisoblaymiz, ya'ni $y = \beta_1 + \beta_2 x^2 + e$. Yangi log faylini ishga tushiring va **br.dta** ma'lumotni oching.

log using chap02_quad, replace text

use "C:\Users\User\Documents\stata\br.dta", clear

describe

summarize

Ma'lumotlar tasviri va qisman xulosa statistikasi:

variable name	storage type	display format	value label	variable label
price	float	%9.0g		sale price, dollars
sqft	float	%9.0g		total square feet

Variable	Obs	Mean	Std. Dev.	Min	Max
price	1080	154863.2	122912.8	22000	1580000
sqft	1080	2325.938	1008.098	662	7897

Yangi o'zgaruvchi yaratishning eng sodda usuli – o'zgaruvch sqft kvadratiga teng bo'lgan yangi **sqft2** o'zgaruvchini yaratishdir:

generate sqft2=sqft^2

Uy narxi bilan uy hajmining kvadrati bilan tuzilgan regressiya va hisoblangan qiymatlarni olish, priceq.

regress price sqft2

predict priceq, xb

. regress price sqft2

Source	SS	df	MS
Model	1.1286e+13	1	1.1286e+13
Residual	5.0150e+12	1078	4.6522e+09
Total	1.6301e+13	1079	1.5108e+10

Number of obs = 1080
F(1, 1078) = 2425.98
Prob > F = 0.0000
R-squared = 0.6923
Adj R-squared = 0.6921
Root MSE = 68207

price	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
sqft2	.0154213	.0003131	49.25	0.000	.014807 .0160356
_cons	55776.56	2890.441	19.30	0.000	50105.04 61448.09

Hisoblangan chiziqni tuzish uchun biz twoway graph buyrug'ining bir necha yangi imkoniyatlarini ko'rsatamiz. Birinchidan, biz mavjud ma'lumotlar bilan grafik tuzamiz. Ikkinchidan, ikkita ham nuqtali grafik, ham hisoblangan to'g'ri chiziqli grafikni o'rnatamiz. Boshqa grafik turlari uchun help twoway ga qarang. Sort imkoniyati qo'shilishi kerak, shuning uchun o'rnatilgan chiziq kuzatishlar muntazam ravishda, har xil o'lchamdagi uylar bir-birini ta'qib qilgan holda uyg'unlashtirilgan natija o'rniga silliq va uzluksiz bo'ladi. O'rnatilgan chiziqni asl holatidan ancha qalinroq qilish uchun, lwidth (medthick) so'zini qo'shing.

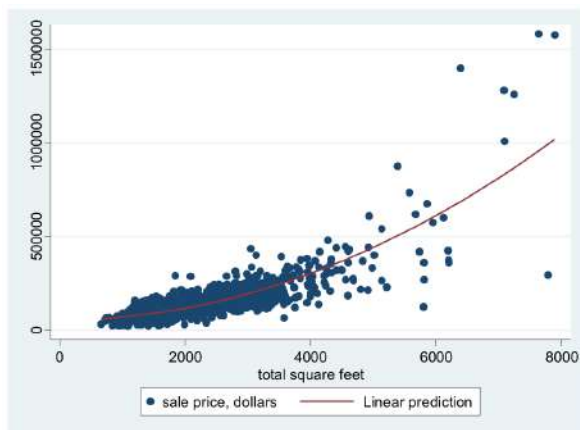
twoway (scatter price sqft) /// /* basic plot */

(line priceq sqft, /// /* 2nd plot: line is continuous */

sort lwidth(medthick)) /* sort & change line thickness */

graph save br_quad, replace

Natijada grafik, hisoblangan kvadrat funksiyani egri chiziqli shaklini ko'rsatadi. Nuqtalar barqaror emas va ular katta uylar qo'shimcha kattalikdagi narxning ko'payishini ko'rsatib turibdi. Bu funksiya tobora yuqoriga o'sib bormoqda. Nuqtalar va moslashuvchanlik hisob-kitoblari biz ilgari ko'rib chiqqan "to'g'ri chiziq"dan yoki chiziqli munosabatlarning funksional shaklidan o'zgacha.



Hisoblangan kvadrat regressiya funksiyasi $y = b_1 + b_2 x^2$ burchagi ($\frac{dy}{dx} = 2b_2 x$) ni tashkil qiladi. $x = \text{sqft}$ ning turli xil qiymatlarida nuqtalarni hisoblash uchun biz oddiy display (ma'lumot uchun `help display`) buyrug'idan foydalanishimiz mumkin, bu di deb qisqartirilishi mumkin. Ushbu bobning avvalida mavzusida bo'lgani kabi, biz regressiya koeffitsiyentiga `_b[sqft2]` ko'rinishida foydalanilgandik. Nuqtalarni `sqft = 2000, 4000` va `6000` qiymatida hisoblashimiz mumkin:

di "slope at 2000 = " `2*_b[sqft2]*2000`

di "slope at 4000 = " `2*_b[sqft2]*4000`

di "slope at 6000 = " `2*_b[sqft2]*6000`

Nuqta va elastiklik qiymati:

```
. di "slope at 2000 = " 2*_b[sqft2]*2000
slope at 2000 = 61.685207
```

```
.
. di "slope at 4000 = " 2*_b[sqft2]*4000
slope at 4000 = 123.37041
```

```
.
. di "slope at 6000 = " 2*_b[sqft2]*6000
slope at 6000 = 185.05562
```

Xuddi shunday buyruqdan foydalanib, biz regressiyadan hisoblangan qiymatlarni ham topa olamiz:

di "predicted price at 2000 = " `_b[_cons]+_b[sqft2]*2000^2`

di "predicted price at 4000 = " _b[_cons]+_b[sqft2]*4000^2
 di "predicted price at 6000 = " _b[_cons]+_b[sqft2]*6000^2

```
. di "predicted price at 2000 = " _b[_cons]+_b[sqft2]*2000^2
predicted price at 2000 = 117461.77
```

```
. di "predicted price at 4000 = " _b[_cons]+_b[sqft2]*4000^2
predicted price at 4000 = 302517.39
```

```
. di "predicted price at 6000 = " _b[_cons]+_b[sqft2]*6000^2
predicted price at 6000 = 610943.42
```

Hisoblangan funksiya uchun elastiklik $\varepsilon = ((dy/dx) \times (xy)) = 2b_2 x^2/y$ dir. Ushbu qiymatlardan foydalanib, koeffitsiyentlarni topamiz

di"elasticityat2000="2*_b[sqft2]*2000^2/(_b[_cons]+_b[sqft2]*2000^2)
 di"elasticityat4000="2*_b[sqft2]*4000^2/(_b[_cons]+_b[sqft2]*4000^2)
 di"elasticityat6000="2*_b[sqft2]*6000^2/(_b[_cons]+_b[sqft2]*6000^2)

```
. di "elasticity at 2000 = " 2*_b[sqft2]*2000^2/(_b[_cons]+_b[sqft2]*2000^2)
elasticity at 2000 = 1.0503027
```

```
. di "elasticity at 4000 = " 2*_b[sqft2]*4000^2/(_b[_cons]+_b[sqft2]*4000^2)
elasticity at 4000 = 1.6312505
```

```
. di "elasticity at 6000 = " 2*_b[sqft2]*6000^2/(_b[_cons]+_b[sqft2]*6000^2)
elasticity at 6000 = 1.8174084
```

Keyinchalik zamonaviy va samarali yondashuv faktor parametrlarini ishlatishdan iborat. Agar biz $x^2 = x \times x$ dan foydalansak, o'zgaruvchan $x = \text{sqft}$ o'zgaruvchining funksiyali o'zgaruvchan belgisi $c.\text{sqft}$ bilan ifodalanadigan bo'lsa, biz yangi o'zgaruvchini yaratmasdan turib, kvadrat funksiyasini to'g'ridan to'g'ri baholashimiz mumkin.

regress price c.sqft#c.sqft

```
. regress price c.sqft#c.sqft
```

Source	SS	df	MS	Number of obs = 1080		
Model	1.1286e+13	1	1.1286e+13	F(1, 1078) = 2425.98		
Residual	5.0150e+12	1078	4.6522e+09	Prob > F = 0.0000		
Total	1.6301e+13	1079	1.5108e+10	R-squared = 0.6923		
				Adj R-squared = 0.6921		
				Root MSE = 68207		

price	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
c.sqft#						
c.sqft	.0154213	.0003131	49.25	0.000	.014807	.0160356
_cons	55776.57	2890.441	19.30	0.000	50105.04	61448.09

Ushbu natija bo'yicha oldingi hisoblangan regressiya bilan bir xil, chunki siz ma'lumotni **browsing** orqali tasdiqlashingiz mumkin: **predict price2**

Faktor belgilaridan foydalanishning katta afzalligi mavjud, Stata **margins** buyrug'i orqali nuqtalarni aniq hisoblashi va elastikni to'g'ri hisoblashga yordam beradi. Birinchidan, nuqtalar **dydx (*)** bilan chekka joylarni ishlatadi, * bu holatda faqatgina **sqft** bo'lgan barcha model o'zgaruvchilari uchun nuqtalar uchun so'rov ko'rsatiladi. Qulaylik bilan biz bir nechta qiymatlarni hisoblash uchun **at** so'zi bilan nuqtalarni kiritishimiz mumkin:

margins, dydx(*) at(sqft=(2000 4000 6000))

```
. margins, dydx(*) at(sqft=(2000 4000 6000))
```

Conditional marginal effects Number of obs = 1080
 Model VCE : OLS
 Expression : Linear prediction, predict()
 dy/dx w.r.t. : sqft

1._at : sqft = 2000
 2._at : sqft = 4000
 3._at : sqft = 6000

		Delta-method		z	P> z	[95% Conf. Interval]	
		dy/dx	Std. Err.				
sqft							
	_at						
	1	61.68521	1.252385	49.25	0.000	59.23058	64.13963
	2	123.3704	2.504769	49.25	0.000	118.4612	128.2797
	3	185.0556	3.757154	49.25	0.000	177.6917	192.4195

Nuqtalar faqat to'g'ri hisoblanmaydi, shuningdek bizga standart xatolik va intervalli kiritish ham taqdim etiladi. Elastikliklarni topish uchun **eyex (*)** buyrug'i ishlatiladi:

margins, eyex(*) at(sqft=(2000 4000 6000))

```
. margins, eyex(*) at(sqft=(2000 4000 6000))
```

Conditional marginal effects
Model VCE : OLS

Number of obs = 1080

Expression : Linear prediction, predict()
ey/ex w.r.t. : sqft

1._at : sqft = 2000
2._at : sqft = 4000
3._at : sqft = 6000

		Delta-method		z	P> z	[95% Conf. Interval]	
		ey/ex	Std. Err.				
sqft							
	_at						
	1	1.050303	.0336868	31.18	0.000	.9842778	1.116328
	2	1.631251	.0203148	80.30	0.000	1.591434	1.671067
	3	1.817408	.0112071	162.17	0.000	1.795443	1.839374

Yuqorida qayd etilgan nuqtalar va elastikliklar **shartli** hisoblanadi, chunki ular ma'lum qiymatlarda hisoblangan. O'rtacha marginal effektlar yoki o'rtacha elastikliklarni hisoblash uchun **at** buyrug'i parametridan foydalanmang.

margins, eyex(*)

```
. margins, eyex(*)
```

```
Average marginal effects          Number of obs   =       1080
Model VCE      : OLS

Expression   : Linear prediction, predict()
ey/ex w.r.t. : sqft
```

	Delta-method					[95% Conf. Interval]	
	ey/ex	Std. Err.	z	P> z			
sqft	1.102401	.0292176	37.73	0.000		1.045135	1.159666

Ushbu o'rtacha marginal effektlar kutilgan natijalar ekanligiga ishonch hosil qilish uchun biz ularni to'g'ridan to'g'ri hisoblash uchun oldin hisoblagan regression koeffitsiyentlarni hisoblashimiz mumkin. Birinchidan, alohida parametr koeffitsiyentlarga qanday murojaat qilishni bilish qiyin bo'lishi mumkin. Bu regressiya parametrini **coeflegend** buyrug'i orqali hal qilinishi mumkin: **regress, coeflegend**

Bu ANOVA jadvaliga va koeffitsiyentlarga qayta ifodalaniladi:

price	Coef.	Legend
c.sqft# c.sqft	.0154213	_b[c.sqft#c.sqft]
_cons	55776.57	_b[_cons]

Faktor modeli spetsifikatsiyasi uchun **_b [varname] - _b [c.sqft # c.sqft]**.

```
generate elas2 = 2*_b[c.sqft#c.sqft]*(sqft^2)/price2
summarize elas2
```

```
. generate elas2 = 2*_b[c.sqft#c.sqft]*(sqft^2)/price2
```

```
. summarize elas2
```

Variable	Obs	Mean	Std. Dev.	Min	Max
elas2	1080	1.102401	.3528353	.2161448	1.890364

E'tibor bering, o'rtacha elastiklik aslida qanday **margins** hisoblangan bo'lsa-da, yana bir marta summarize **standart** egiluvchanliklarni hisoblab chiqsangiz va **margins** Delta uslubidagi standart xatoni hisoblab chiqadi. Endi ishchi faylini yopishingiz mumkin: **log close**

2.6.2. Chiziqli logorifmik model

Xuddi shu ma'lumotlardan foydalanib biz log-linear modelni taxmin qilamiz. Hisoblangan chiziqli model quyidagicha bo'ladi: $(\ln y) = b_1 + b_2 x$

Y ning hisoblangan qiymatini olish uchun eng to'g'ri yo'l – bu antilogni hisoblashdir.

$$y = \exp((\ln(y))) = \exp(b_1 + b_2 x)$$

Hisoblangan log-chiziqlini egri chizig'i $(dy/dx) = b_2 y$ dir va elastiklik $\epsilon = ((dy/dx) \times (x/y)) = b_2 x$ dir. Yangi faylni oching va br.dta dan foydalaning. O'zgaruvchi price bo'yicha batafsil xulosa statistikasi va uning histogrammini o'ng tomondan uzun egri chiziq bilan ko'rsatilgan taqsimot borligini ko'rasiz.

log using chap02_llin, replace text

use br, clear

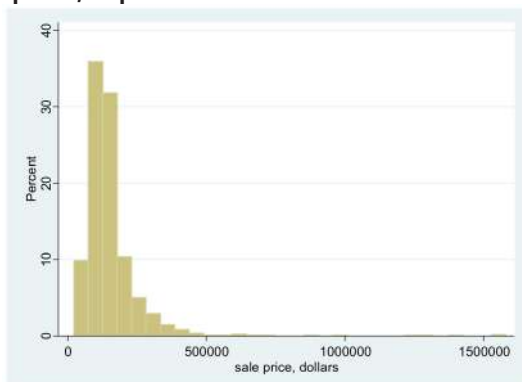
summarize price, detail

```
. summarize price, detail
```

sale price, dollars				
Percentiles		Smallest		
1%	31000	22000		
5%	59897.5	22000		
10%	74450	22654	Obs	1080
25%	99000	23000	Sum of Wgt.	1080
50%	130000		Mean	154863.2
		Largest	Std. Dev.	122912.8
75%	170325	1280000		
90%	244200	1400000	Variance	1.51e+10
95%	315000	1575000	Skewness	6.291909
99%	610000	1580000	Kurtosis	60.94976

Shuni ta'kidlash kerakki, natijaga asosan o'rtacha qiymatdan (50 foiz) juda katta qiymatga ega bo'lib, 1,5 million dollarni tashkil etadi. Skewness koeffitsiyenti noldan ko'ra ijobiydir, biz kutgan edik nosimmetrik tarqalishi xuddi normal taqsimotga o'xshashini:

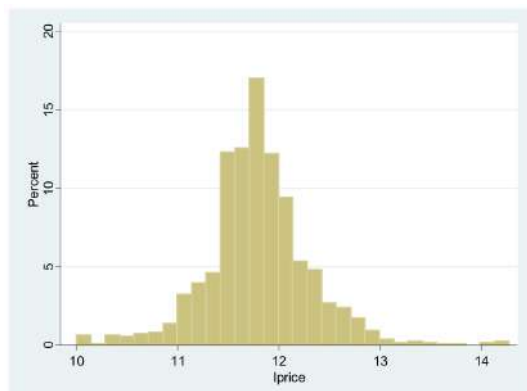
```
histogram price, percent  
graph save price, replace
```



Endi narxni logarifimlangan qiymatini yaratamiz va uning histogrammasini tuzamiz:

```
generate lprice = ln(price)  
histogram lprice, percent  
graph save lprice, replace
```

Quyidagi histogramma simmetrikdir, lekin qo'ng'iroq shakilday emas, ya'ni normal taqsimlanmagan:



Log-linear regressiya modeli quyidagi buyruq orqali ifodalaniladi: **reg lprice sqft**

```
. reg lprice sqft
```

Source	SS	df	MS	Number of obs = 1080		
Model	185.472091	1	185.472091	F(1, 1078) = 1794.78		
Residual	111.400275	1078	.103339773	Prob > F = 0.0000		
Total	296.872366	1079	.275136577	R-squared = 0.6248		
				Adj R-squared = 0.6244		
				Root MSE = .32147		

lprice	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
sqft	.0004113	9.71e-06	42.36	0.000	.0003922	.0004303
_cons	10.8386	.0246075	440.46	0.000	10.79031	10.88688

So'ngra hisoblangan qiymatlarni topamiz:

```
predict lpricef, xb
```

```
generate pricef = exp(lpricef)
```

O'zgaruvchi **pricef** hisoblangan (yoki prognoz qilingan) narxdir.

Hisoblangan egri chiziq grafigi:

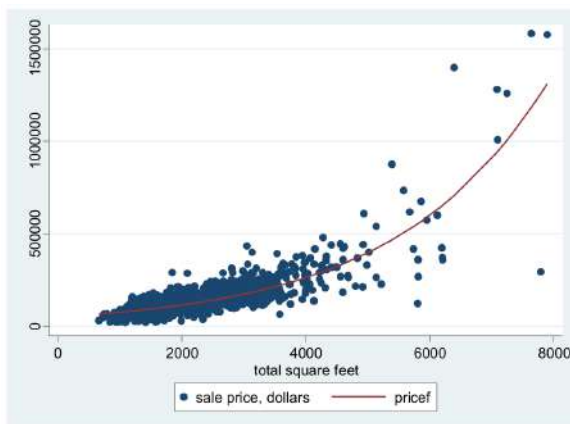
```
reg lprice sqft
```

```
predict lpricef, xb
```

```
generate pricef = exp(lpricef)
```

```
twoway (scatter price sqft) ///
```

(line pricef sqft, sort lwidth(medthick))
graph save br_loglin, replace



Hisoblangan egri chiziqli munosabat emas, chunki nuqtalar va elastiklik aniq nuqtalarda hisoblanishi kerak. Shuni ta'kidlash kerakki, bu yerda qiymat mavjud emas, **margins** hisoblari amalga oshirilmaydi, chunki Stata logaritmik transformatsiyalangan ta'sir etuvchi o'zgaruvchini o'qiy olmaydi. Shu sababdan biz nuqtalarni va elastiklikni aniq tarzda hisoblashimiz kerak. E'tibor beraylikki, nuqtalar bir uyning kattaligi bo'yicha prognoz baholar bilan emas, balki ma'lum bir uy narxini hisobga olgan holda hisoblab chiqilgan:

```
di "slope at 100000 = " _b[sqft]*100000
di "slope at 500000 = " _b[sqft]*500000
di "elasticity at 2000 = " _b[sqft]*2000
di "elasticity at 4000 = " _b[sqft]*4000
```

```

. di "slope at 100000 = " _b[sqft]*100000
slope at 100000 = 41.126885

.
. di "slope at 500000 = " _b[sqft]*500000
slope at 500000 = 205.63442

.
. di "elasticity at 2000 = " _b[sqft]*2000
elasticity at 2000 = .82253769

.
. di "elasticity at 4000 = " _b[sqft]*4000
elasticity at 4000 = 1.6450754

```

Shuningdek, hisoblangan har bir uyning narxi bo'yicha o'rtacha marginal ta'sirlarni hisoblashimiz mumkin:

```

generate me = _b[sqft]*pricef
summarize me

```

```

. generate me = _b[sqft]*pricef

```

```

. summarize me

```

Variable	Obs	Mean	Std. Dev.	Min	Max
me	1080	61.00072	42.91725	27.51118	539.2198

Xuddi shunday o'rtacha elastiklik ham topa olamiz:

```

generate elas = _b[sqft]*sqft
summarize elas

```

```

. generate elas = _b[sqft]*sqft

```

```

. summarize elas

```

Variable	Obs	Mean	Std. Dev.	Min	Max
elas	1080	.9565858	.4145993	.27226	3.24779

Ushbu ishchi faylini quyidagi buyruq orqali yopishingiz mumkin:
log close

2.7. Regression jarayonda ko'rsatkich o'zgaruvchilarni ishlatish

Ko'rsatkichli o'zgaruvchilar odatda ikki tomonlama 0-1 o'zgaruvchilardir. Ular ko'chmas mulk modelidagi joy kabi sifat omillarini ko'rsatish uchun regressiyada foydalanish mumkin. Yangi ishchi faylni va ma'lumotlar **utown.dta** faylini oching. Ma'lumotlarni **describe** va **summarize** qiling.

log using chap02_indicator, replace text

use utown, clear

describe

summarize

Ma'lumotlar shuni tasvirlaydiki, **utown** o'zgaruvchisi 1 soni bilan agarda uy universitetga yaqin bo'lsa va boshqa hollarda 0 soni bilan ifodalangan.

variable name	storage type	display format	value label	variable label
price	double	%10.0g		house price, in \$1000
sqft	double	%10.0g		square feet of living area, in 100s
age	byte	%8.0g		house age, in years
utown	byte	%8.0g		=1 if close to university
pool	byte	%8.0g		=1 if house has pool
fplace	byte	%8.0g		=1 if house has fireplace

Xulosa statistikasi shuni ko'rsatadiki, 1000ta kuzatuv, taxminan 52% i universitetga yaqin mahallada joylashgan.

. summarize

Variable	Obs	Mean	Std. Dev.	Min	Max
price	1000	247.6557	42.19273	134.316	345.197
sqft	1000	25.20965	2.91848	20.03	30
age	1000	9.392	9.426728	0	60
utown	1000	.519	.4998889	0	1
pool	1000	.204	.4031706	0	1
fplace	1000	.518	.4999259	0	1

Umumiy grafikda umumiy bahoga ega bo'lgan uy narxlarining histogrammasini yaratish ularni osonlik bilan taqqoslash imkonini

beradi. Avval \$ 12000 `[width(12)]` kenglikdagi va \$ 130000 dan boshlanadigan `[start (130)]`, misol minimalidan past bo'lgan qutilari yordamida ikki alohida histogramma yarating va saqlang. \$ 12000 miqyosdagi yirik konteynerlardan foydalanish sinov va xatolarga asoslangan edi. `xlabel` tasdiqlash indikatori 130 dan 350 gacha 24 kenglikda bo'lgan ustunlar orqali ifodalanishi kerakligini ko'rsatib turibdi. **Histogram** buyrug'i mantiqiy operatoridan foydalaning, masalan, `ifutown == 0` bo'lsa, `utown` o'zgaruvchi tomonidan ikkita belgilangan ma'lumotlarni tanlang:

```
histogram price if utown==0, width(12) start(130) percent ///
xtitle(House prices ($1000) in Golden Oaks) ///
```

```
xlabel(130(24)350) legend(off)
```

```
graph save utown_0, replace
```

```
histogram price if utown==1, width(12) start(130) percent ///
xtitle(House prices ($1000) in University Town) ///
```

```
xlabel(130(24)350) legend(off)
```

```
graph save utown_1, replace
```

Saqlab olingan grafiklarni **graph combine** buyrug'i yordamida birlashtiriladi. Ularni bir-biriga biriktirish uchun ularni bitta ustunga `[col(1)]` ga kiritishni tanlaymiz. `help graph combine` buyrug'idan foydalanib, "... iscale (1) da **matn** va **belgilar** aslida bir xil o'lchamlarda bo'lishi kerakligini anglatadi" deb izoh qoldirilgan. Grafik nomlari ularni * `.gph` fayllari sifatida identifikatsiya qilish uchun kodirovkalarga qo'yiladi:

```
graph combine "utown_0" "utown_1", col(1) iscale(1)
```

```
graph save combined, replace
```



Grafika shuni ko'rsatadiki, universitet mahallasidagi narxlar boshqa mahallalardagi uylarga qaraganda yuqori qiymatga ega.

Ikki grafika yaratib, ularni birlashtirmasdan, Stataning `by` buyrug'ini ma'lumotlar to'plamlari orqali takrorlaydigan juda muhim xususiyatdan foydalanishimiz mumkin. Undan foydalanishda izohlardan foydalanish muhim ahamiyatga ega. Keyingi ikkita bayonnomada avval biz "Golden Oaks" uchun 0, "university town" uchun "1" deb nomlangan izohli belgilarni yaratamiz. Ikkinchidan, biz izohni o'zgaruvchi **UTOWN** ga qo'llaymiz:

```
label define utownlabel 0 "Golden Oaks" 1 "University Town"
```

```
label value utown utownlabel
```

Histogramma ma'lumotlarning pastki qismlari **UTOWN** tomonidan aniqlanganligini bildiruvchi parametrdan foydalaning:

```
histogram price, by(utown, cols(1)) ///
```

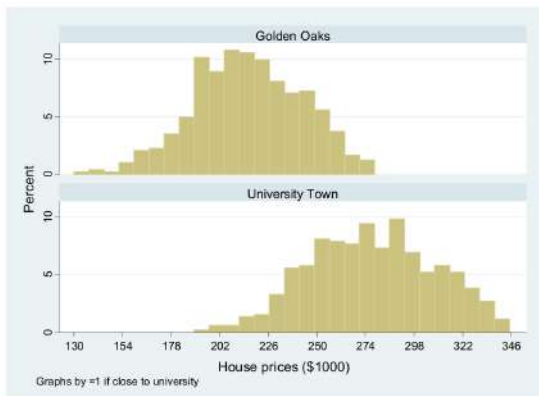
```
start(130) percent ///
```

```
xtitle(House prices ($1000)) ///
```

```
xlabel(130(24)350) legend(off)
```

```
graph save combined2, replace
```

Olingan grafik narx uchun bir o'lchov va `by` opsiyaga ko'ra izoh ko'rsatilgan.



Bundan tashqari, alohida mahallalarda narxlar bahosi statistikasini ham o'rganishi mumkin:

```
summarize price if utown==0
```

```
summarize price if utown==1
```

```
. summarize price if utown=0
```

Variable	Obs	Mean	Std. Dev.	Min	Max
price	481	215.7325	26.73736	134.316	276.977

```
. summarize price if utown=1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
price	519	277.2416	30.78208	191.57	345.197

Shunga qaramay, **by** opsiyasidan foydalanish oson va samaraliroq. Buyruq bo'yicha ishlatilayotganda ma'lumotlarni kuzatishning pastki qismlarini belgilaydigan o'zgarmaydigan qiymatiga qarab tartiblash kerak. Tartiblash quyidagi buyruqlar yordamida amalga oshirilishi mumkin.

by utown, sort: summarize price

```
. by utown, sort: summarize price
```

```
-> utown = Golden Oaks
```

Variable	Obs	Mean	Std. Dev.	Min	Max
price	481	215.7325	26.73736	134.316	276.977

```
-> utown = University Town
```

Variable	Obs	Mean	Std. Dev.	Min	Max
price	519	277.2416	30.78208	191.57	345.197

Yuqoridagidan farqli o'laroq, buyruqni qisqaroq qilish ham mumkin.

bysort utown: summarize price

```
. bysort utown: summarize price
```

```
-> utown = Golden Oaks
```

Variable	Obs	Mean	Std. Dev.	Min	Max
price	481	215.7325	26.73736	134.316	276.977

```
-> utown = University Town
```

Variable	Obs	Mean	Std. Dev.	Min	Max
price	519	277.2416	30.78208	191.57	345.197

0

Taqqoslama o'zgaruvchisi sifatida indikator o'zgaruvchisi bilan regressiya odatdagidek bir xil sintaksisga ega, lekin biz i.utown buyrug'idan foydalangan bo'lardik.

regress price utown

```
. regress price utown
```

Source	SS	df	MS	Number of obs =	1000
Model	944476.744	1	944476.744	F(1, 998) =	1130.24
Residual	833969.397	998	835.640678	Prob > F =	0.0000
				R-squared =	0.5311
				Adj R-squared =	0.5306
Total	1778446.14	999	1780.22637	Root MSE =	28.907

price	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
utown	61.50911	1.829589	33.62	0.000	57.91882 65.09939
_cons	215.7325	1.318066	163.67	0.000	213.146 218.319

E'tibor bering, taxmin qilingan o'zgarmas atama o'rtacha narx bo'lsa, **utown = 0** ga teng bo'ladi va **utown** ning koeffitsienti namuna vositalarining orasidagi farqdir.

Ikkala o'zgaruvchini statistika kursida teng yoki tengmasligini tekshirishingiz mumkin. Stata bu tekshirishni amalga oshirish uchun **ttest** buyrug'ini o'z ichiga oladi.

ttest price, by(utown)

Statistik natijalarni olingandan so'ng ularni buyruqlarini do-file ga saqlab ishchi faylni yopishingiz mumkin.

log close

II bob mavzularini mustahkamlash uchun savollar

1. Klassik chiziqli modelni yaratish uchun ma'lumotlarini yuklash ketma ketligi qanday?
2. Kundalik log faylni ishga tushirish va yopish uchun qanday buyruq amalga oshiriladi?
3. Stata xotirasini bir vaqtning o'zida yangi ma'lumotlar faylini ochish hamda tozalash zarur bo'lsa, qanday buyruq kiritiladi?
4. Stata dasturida internet saytidan ma'lumotlarni yuklash uchun qanday amal bajarish lozim?

5. Xulosa statistika buyrug'i sintaksisi burug'i qanday beriladi?
6. Oddiy regressiya modelida ma'lumotlar qiymatlarini qanday diagramma orqali chizish mumkin?
7. Stata buyruqlar oynasida Scatter diagrammasini yaratish qaysi quyruq orqali amalga oshiriladi?
8. Oddiy chiziqli regressiya modeli qanday formula orqali ifodalanadi?
9. Stata dasturida elastiklikni avtomatik ravishda hisoblash qanday ketma ketlik asosida amalga oshiriladi?
10. marginal effect deganda nimani tushunasiz?
11. Linear (to'g'ri chiziqli) regressiya modelida elastiklik qanday ifodalanadi?
12. Hisoblangan regressiya chizig'ini tuzish qanday amallar orqali bajariladi?

Ikkinchi bob mavzularida qo'llanilgan buyruqlar va atamalar
atmeans

by	graph combine	name
bysort	graph save	nlcom
cd	graph save	postestimation
clear	histogram	predict
data browser	label define	program
data editor	label value	quadratic model
delta method	lfit list	random number
di	list in	range/delta
display	log	reg
dydx	log close	regress
elasticity	log-linear model	replace
estat	lwidth	residuals
estat vce	marginal effects	return list
eyex	margins	return scalar
factor variable	mean	rnormal
gen	model, linear	save
generate	model, log-linear	saving
global	model, quadratic	scalar
graph	msymbol	scatter

seed,	random	std. dev.	twoway scatter
number		std. err.	use
set more off		summarize, detail	variance
set obs		title	xb
set seed		ttest	xlabel
simulate		twoway	ylabel
standard deviation		twoway lfit	
standard error		twoway line	

Ikkinchi bob mavzularini mustahkamlash uchun masalalarning do-file lari

Ma'lumotlar buyrug'i: `cd c:\data\poe4stata`

1-masala. **log file ni ochish va datalardan foydalanish**

`log using chap02_food, replace text`

`use food, clear`

`* examine data`

`describe`

`* browse`

`list`

`list in 1/5`

`list food_exp in 1/5`

`list food_exp if income < 10`

`* compute summary statistics`

`summarize`

`summarize food_exp, detail`

`* simple plot data`

`twoway (scatter food_exp income)`

`graph save food1, replace // open for editing with: graph use food1`

`twoway (scatter food_exp income), saving(food1, replace)`

`twoway (scatter food_exp income), name(food1, replace)`

`* compute least squares regression`

`regress food_exp income`

`predict yhat, xb`

`predict ehat, residuals`

`* compute elasticity at means`

```

margins, eyex(income) atmeans
margins, eyex(income)
generate elas = _b[income]*income/yhat
summarize elas
* examine variances and covariances
estat vce
* add observation to data file
edit
set obs 41
replace income=20 in 41
* obtain prediction
predict yhat0
list income yhat0 in 41
log close
2-masala. log file ni ochish.
log using chap02_quad, replace text
* open br data and examine
use br, clear
describe
summarize
* create new variable
generate sqft2=sqft^2
* regression
regress price sqft2
predict priceq, xb
* plot fitted line
tway (scatter price sqft)
(line priceq sqft,
sort lwidth(medthick))
graph save br_quad, replace ///
/// /* basic plot */
/* 2nd plot: line is continuous */
/* sort & change line thickness */
* slope and elasticity calculations
di "slope at 2000 = " 2*_b[sqft2]*2000

```

```

di "slope at 4000 = " 2*_b[sqft2]*4000
di "slope at 6000 = " 2*_b[sqft2]*6000
di "predicted price at 2000 = " _b[_cons]+_b[sqft2]*2000^2
    di "predicted price at 4000 = " _b[_cons]+_b[sqft2]*4000^2
    di "predicted price at 6000 = " _b[_cons]+_b[sqft2]*6000^2
    di "elasticity at 2000 = " 2*_b[sqft2]*2000^2/(_b[_cons]+_b[sqft2]*2000^2)
    di "elasticity at 4000 = " 2*_b[sqft2]*4000^2/(_b[_cons]+_b[sqft2]*4000^2)
    di "elasticity at 6000 = " 2*_b[sqft2]*6000^2/(_b[_cons]+_b[sqft2]*6000^2)
* using factor variables
regress price c.sqft#c.sqft
predict price2
margins, dydx(*) at(sqft=(2000 4000 6000))
margins, eyex(*) at(sqft=(2000 4000 6000))
margins, eyex(*)
regress, coeflegend
generate elas2 = 2*_b[c.sqft#c.sqft]*(sqft^2)/price2
summarize elas2
log close
3-masala. log file ni ochish va datalardan foydalanish
log using chap02_llin, replace text
use br, clear
* distribution of prices
summarize price, detail
histogram price, percent
graph save price, replace
* distribution of log(price)
generate lprice = ln(price)
histogram lprice, percent
graph save lprice, replace
* log-linear regression
reg lprice sqft

```

```

predict lpricef, xb
* price prediction using anti-log
generate pricef = exp(lpricef)
twoway (scatter price sqft) (line pricef sqft, sort lwidth(medthick))
graph save br_loglin, replace
* slope and elasticity calculations
di "slope at 100000 = " _b[sqft]*100000
di "slope at 500000 = " _b[sqft]*500000
di "elasticity at 2000 = " _b[sqft]*2000
di "elasticity at 4000 = " _b[sqft]*4000
* average marginal effects
generate me = _b[sqft]*pricef
summarize me
generate elas = _b[sqft]*sqft
summarize elas
log close
4-masala. log file ni ochish
log using chap02_indicator, replace text
* open utown data and examine
use utown, clear
describe
summarize
* histograms of utown data by neighborhood
histogram price if utown==0, width(12) start(130) percent ///
xtitle(House prices ($1000) in Golden Oaks) xlabel(130(24)350)
legend(off)
graph save utown_0, replace
histogram price if utown==1, width(12) start(130) percent ///
xtitle(House prices ($1000) in University Town) xlabel(130(24)350)
legend(off)
graph save utown_1, replace
graph combine "utown_0" "utown_1", col(1) iscale(1)
graph save combined, replace
* using by option
label define utownlabel 0 "Golden Oaks" 1 "University Town"

```

```
label value utown utownlabel
histogram price, by(utown, cols(1)) start(130) percent
xtitle(House prices ($1000)) /// xlabel(130(24)350) legend(off)
graph save combined2, replace
* summary stats
summarize price if utown==0
summarize price if utown==1
* summary stats using by
by utown, sort: summarize price
* summary stats using bysort
bysort utown: summarize price
* regression
regress price utown
* test of two means
ttest price, by(utown)
log close
```

III BOB. INTERVALLI BAHOLASH VA GIPOTEZA TESTI

Intervalli baholashlar ishonch oraliq'i sifatida ham tanilgan. Stata regressiya tahlilini amalga oshirganda, uning standart chiqishi, ya'ni bir qismi har bir koeffitsiyent uchun intervalning 95%ni baholab ko'rsatishidir. Yangi Stata faylini yarating va ish katalogiga o'ting. Oldingi bobning 2.4 bo'limida ko'rsatilgandek, jurnal faylini oching va oziq-ovqat xarajatlari modelini baholang.

log using chap03, replace text

use "D:\programms\Stata12\STATA\food.dta", clear

reg food_exp income

Regressiya tahlilida **[95% Conf. Interval]** ni o'z ichiga oladi, tegishli koeffitsiyentlar uchun intervalli hisob-kitoblarning pastki va yuqori chegaralari berilgan.

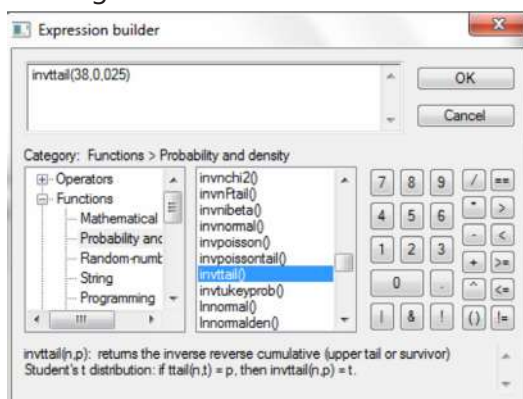
food_exp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
income	10.20964	2.093264	4.88	0.000	5.972052	14.44723
_cons	83.416	43.41016	1.92	0.062	-4.463279	171.2953

Intervalli baholashlar **Coef. $\pm$ t-critical*Std. Err.** sifatida hisoblanadi. Koeffitsiyentlar qiymatlari va standart xatolar berilgan. Qolgan tarkibiy qismi, ya'ni t muhim ahamiyatga ega. Buni Principles of Econometrics kitobining 2-jadvalida yoki biz hozirda ko'rsatganimizdek, Stata yordamida topish ham mumkin.

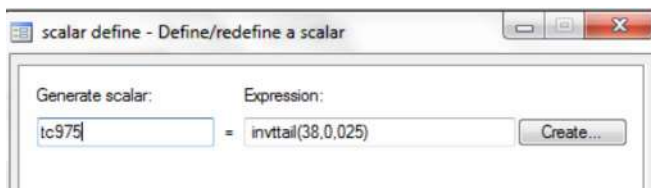
3.1. T taqsimotning kritik qiymatlarini hisoblashda Stata dasturining imkoniyati (invttail)

Ko'pgina ehtimolli taqsimotlarning kritik qiymatlarini hisoblash uchun biz Stata dan foydalanishimiz mumkin, bu ko'plab kontekstlarda juda qulaydir. Kritik qiymatlar Statada skalyar shaklida yaratilgan va ular "inverse" funksiyalarni ko'rsatuvchi umumiy prefiksi **inv** buyrug'ini o'z ichiga oladi. Muayyan skalyar qiymatni chaqirish uchun quyidagi buyruqni kiriting: **help scalar**

Agar dialog oynasidan foydalanmoqchi bo'lsangiz, **Viewer** oynasidan **Dialog** menusedagi define tugmasini bosing. Shunda scalar **define** oynasi paydo bo'ladi, so'ngra **Create** tugmasini bosing. **Expressionbuilder** oynasidan **Category** qismidan **functions** ni oching va uning ichidan **probabilitiy** and **density** dan foydalanib, **invttail()** ni ikki marta bosing va **N-2 = 38** erkinlik darajalari va taqsimotning yuqori qismidagi ehtimollik miqdorini to'ldiring. 95% oraliq baholash: yuqoridagi buyruqlardan so'ng keyingi oynada ehtimolning 2,5% t taqsimotining 97,5 foizini belgilaydi. So'ngra **OK** tugmasini bosing.



So'ngra bizda mavjud bo'lgan skalyar belgilash oynasi paydo bo'ladi:



Ushbu oynaning pastki qismidagi **"Submit"**ni bosing. Natijalar oynasida (va ko'rib chiqish oynasida) Stata buyrug'i quyidagicha ko'rsatilgan bo'ladi: **scalar define tc975 = invttail(38,0,025)**

Ixtiyoriy aniqlash shart emas, shuning uchun buyruqni soddalashtirish mumkin: **scalar tc975 = invttail(38,0,025)**

Ushbu skalarlarning qiymatini ko'rish uchun, biz uni natijasini

ko'rsatishimiz uchun display buyrug'idan foydalanamiz: `di "t critical value 97.5 percentile = " tc975`

Natijada quyidagi natijaga erishamiz:

```
. di "t critical value 97.5 percentile = " tc975
t critical value 97.5 percentile = 2.0243942
```

t – kritik qiymatlarning boshqa misollari:

```
di "t(30) 95th percentile = " invttail(30,0.05)

. di "t(30) 95th percentile = " invttail(30,0.05)
t(30) 95th percentile = 1.6972609

di "t(20) 5th percentile = " invttail(20,0.95)

. di "t(20) 5th percentile = " invttail(20,0.95)
t(20) 5th percentile = -1.7247182

di "t(30) 2.5th percentile = " invttail(30,0.975)

. di "t(30) 2.5th percentile = " invttail(30,0.975)
t(30) 2.5th percentile = -2.0422725
```

Oraliq interval tuzish

Regressiya natijalari va biz hozirgina 95% intervalgacha hisoblagan t – kritik qiymatdan kalkulator yordamida ham hisoblash mumkin. Interval hisobini olish uchun Stata xotirasida saqlangan natijalardan ham foydalanishingiz mumkin. Yuqorida ta'kidlab o'tilgandek, agar regressiya aniqlansa, ma'lum natijalar saqlanib qoladi va keyinchalik ulardan foydalanish mumkin bo'ladi. Hisoblangan koeffitsiyentlar va standart xatolar navbati bilan `_b [varname]` va `_se [varname]` sifatida saqlanadi. Oziq-ovqat xarajatlari tenglamasini hisoblab chiqqandan so'ng, nuqta koeffitsiyenti `_b [income]` deb nomlanadi va taxmin qilinadigan nuqta `_b [_cons]`. Ularning standart xatolari `_se [income]` va `_se [_cons]` dir. Qo'shimcha ma'lumot olish uchun buyruqlar oynasida `help _variables` buyrug'ini kiriting. 95% oraliq hisob-kitoblar `b_k ± t_c se(b_k)` formula orqali hisoblanadi. Nuqta uchun intervalni hisoblashning yuqori chegarasi va pastki chegarasi quyidagicha topiladi:

```
scalar ub2 = _b[income] + tc975*_se[income]
scalar lb2 = _b[income] - tc975*_se[income]
```

Bularning natijalarini quyidagi buyruq orqali namoyish etilishi mumkin:

```
di "beta 2 95% interval estimate is " lb2 " , " ub2  
va buning natijasi:
```

```
. di "beta 2 95% interval estimate is " lb2 " , " ub2  
beta 2 95% interval estimate is 5.9720525 , 14.447233
```

3.2. Gipoteza testlarini hisoblash va tahlil qilish

Parametrlar to'g'risidagi gipoteza testlari uchun ishlatiladigan t statistikasi regressiya chiqishidagi kalkulator va statistik jadvalning t kritik qiymati yordamida hisoblash mumkin. Ammo ushbu bo'limda biz Stata yordamida testning statistik qiymatlari, tanqidiy qiymatlari va p qiymatlarini hisoblab chiqamiz. Misol sifatida biz oziq-ovqat xarajatlari regressiyasining modelini davom ettiramiz:

$H_0: \beta_2 = 0$ nol gipotezani muqobil $H_0: \beta_2 > 0$ gipotezaga qarshi tekshiramiz. t statistik qiymatni va kritik qiymatni ishlatib qurishimiz va namoyish etishimiz mumkin:

```
scalar tstat0 = _b[income]/_se[income]  
di "t statistic for Ho: beta2=0 = " tstat0  
di "t(38) 95th percentile = " invttail(38,0.05)
```

```
. scalar tstat0 = _b[income]/_se[income]  
  
. di "t statistic for Ho: beta2=0 = " tstat0  
t statistic for Ho: beta2=0 = 4.8773806  
  
. di "t(38) 95th percentile = " invttail(38,0.05)  
t(38) 95th percentile = 1.6859545
```

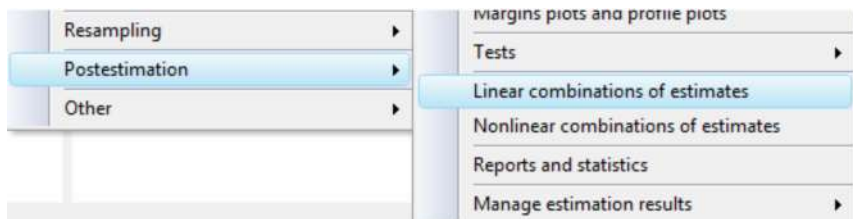
E'tibor bering, kritik qiymat t taqsimotning o'ng tomonida keladi va biz kritik qiymatni topish uchun invttail buyrug'idan foydalanamiz. Koeffitsiyentlar nolga teng bo'lganda gipotezaning t statistik qiymatlari Stata tomonidan avtomatik ravishda ishlab chiqariladi, ya'ni regressiya modelida "t" deb ko'rsatilgan ustunda hisoblangandi.

food_exp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
income	10.20964	2.093264	4.88	0.000	5.972052	14.44723
_cons	83.416	43.41016	1.92	0.062	-4.463279	171.2953

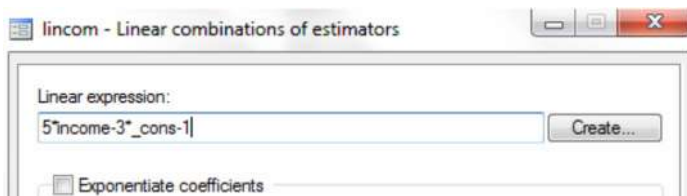
Koeffitsientlar nolga teng bo'lganda
gipotezaning t-statistik qiymatlari

Ba'zida koeffitsiyentlar haqida yanada murakkab farazlarni sinab ko'rishni istaymiz va ularni baholashda keyingi buyruq bu **lincom** yordamida amalga oshirish mumkin. Stata ochiladigan menyusidan ko'rsatilgan ketma-ketlikni tanlang:

Statistics > Postestimation > Linear combinations of estimates.



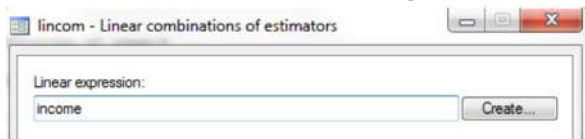
Muloqot oynasida siz taxmin qilmoqchi bo'lgan gipotezani kiriting. Bu esa $5b_2 - 3b_1 = 1$ kabi b_1 va b_2 koeffitsiyentlarini o'z ichiga olgan chiziqli kombinatsiya deb ataladigan har qanday chiziqli ifoda bo'lishi ham mumkin. Albatta, Stata b_1 va b_2 ga to'g'ridan to'g'ri murojaat qilmaydi, aksincha $_b$ [income] va $_b$ [_cons] orqali. Agar buyruq **lincom** bo'lsa, $5 * \text{income} - 3 * \text{_cons} - 1$ kabi o'zgaruvchan nomlarga murojaat qilishimiz kerak. Buyruq ifoda qiymatini va uning standart xatosini hisoblab chiqadi va t statistik va interval bahosini hisoblab ko'rsatadi.



```
. lincom 5*income-3*_cons-1
( 1) 5*income - 3*_cons = 1
```

food_exp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
(1)	-200.1998	140.1664	-1.43	0.161	-483.9518	83.55225

Boshqa misol sifatida shunchaki kiriting:



OK tugmasini bosing. So'ngra natijalar oynasida biz kiritgan Stata buyrug'i va regressiya qiymatlariga o'xshash quyidagi natija paydo bo'ladi:

```
. lincom income
( 1) income = 0
```

food_exp	Coef.	Std. Err.	t	P> t	[95% conf. Interval]	
(1)	10.20964	2.093264	4.88	0.000	5.972052	14.44723

3.2.1. Iqtisodiy gipotezaning right-tail testi

$H_0: \beta_2 \leq 5.5$ muqobil gipotezaga qarshi $H_1: \beta_2 > 5.5$ sinovini o'tkazish uchun biz yana test statistikasi va 0.01 o'ng tomonning kritik qiymatini hisoblab chiqamiz:

scalar tstat1 = (_b[income]-5.5)/_se[income]

E'tibor bering, biz ish tartibini boshqarish uchun qavslardan foydalanamiz:

di "t-statistic for Ho: beta2 = 5.5 is " tstat1

di "t(38) 99th percentile = " invttail(38,0.01)

Bu buyruqlar quyidagi natijalarni hisoblab ko'rsatadi:

```
. di "t-statistic for Ho: beta2 = 5.5 is " tstat1
t-statistic for Ho: beta2 = 5.5 is 2.2499045
```

```
. di "t(38) 99th percentile = " invttail(38,0.01)
t(38) 99th percentile = 2.4285676
```

So'ngra **lincom** buyrug'idan foydalaning: **lincom income-5.5**

Natija shuni ko'rsatadiki, $b_2-5.5$ qiymati hisoblangan [Coef.] va t statistik [t] hisoblab chiqilgan, shuningdek 95% interval bilan hisoblangan.

```
. lincom income-5.5
(1) income = 5.5
```

food_exp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
(1)	4.709643	2.093264	2.25	0.030	.4720525	8.947233

↑ Estimate minus 5.5
↑ t-statistic value

3.2.2. Iqtisodiy gipotezani left-tail testi

Left-tail testini tekshirish uchun $H_1: \beta_2 < 15$ muqobil faraz bilan $H_0: \beta_2 \geq 15$ nol gipoteza bilan hisoblab chiqamiz. Buyruqlar ketma-ketligi quyidagicha:

scalar tstat2 = (_b[income]-15)/_se[income]

di "t-statistic for Ho: beta2 = 15 is " tstat2

0.05 kritik qiymatini hisoblash uchun yana **invttail** buyrug'idan foydalanamiz. Left-tail testini tekshirib ko'rishimizga qaramay, bizga bir vaqtning o'zida right-tail qismidagi 95%li ehtimollik kerak bo'ladi. Shu asnoda quyidagi buyruqga yuzlanamiz:

di "t(38) 5th percentile = " invttail(38,0.95)

Quyidagi natijaga erishamiz:

```
. di "t-statistic for Ho: beta2 = 15 is " tstat2
t-statistic for Ho: beta2 = 15 is -2.2884634
```

```
. di "t(38) 5th percentile = " invttail(38,0.95)
t(38) 5th percentile = -1.6859545
```

So'ngra **lincom** buyrug'idan foydalanamiz: **lincom income-15**

```
. lincom income-15
(1) income = 15
```

food_exp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
(1)	-4.790357	2.093264	-2.29	0.028	-9.027948	-.5527666

3.2.3. Iqtisodiy gipotezani ikki tomonlama tekshirish

Ikki tomonlama sinov kritik qiymatlarni hisoblashdan tashqari bitta tomon sinov bilan bir xil natijaga ega. 0,05 ahamiyatlilik darajasidagi sinov uchun t taqsimotning 2,5 va 97,5 foizi bo'lishi kerak. $H_0: \beta_2 = 7.5$ muqobil gipotezaga qarshi $H_1: \beta_2 \neq 7.5$ sinovini o'tkazish uchun quyidagi buyruqlardan foydalaning:

```
scalar tstat3 = (_b[income]-7.5)/_se[income]
```

```
di "t-statistic for Ho: beta2 = 7.5 is " tstat3
```

```
di "t(38) 97.5th percentile = " invttail(38,0.025)
```

```
di "t(38) 2.5th percentile = " invttail(38,0.975)
```

Natija quyidagicha:

```
. di "t-statistic for Ho: beta2 = 7.5 is " tstat3
t-statistic for Ho: beta2 = 7.5 is 1.2944586

.
. di "t(38) 97.5th percentile = " invttail(38,0.025)
t(38) 97.5th percentile = 2.0243942

.
. di "t(38) 2.5th percentile = " invttail(38,0.975)
t(38) 2.5th percentile = -2.0243942
```

Stata **invttail** buyrug'i t taqsimotning yuqori qismi bilan ishlaganligi sababli, pastki qismi uchun kritik qiymatlarni hisoblash chalkash bo'lishi mumkin. Lekin har doim funksiyani ishlashni davom ettiring. T taqsimot nosimmetrik va 90, 95, 97.5 va 99 foizlar ijobiy qiymatlar, 1, 2.5, 5 va 10 foizlar esa salbiydir. **Lincom** buyrug'idan foydalaning: **lincom income-7.5**

```
lincom income-7.5
```

```
(1) income = 7.5
```

food_exp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
(1)	2.709643	2.093264	1.29	0.203	-1.527948	6.947233

3.3. p qiymatlar (p-values)ni hisoblash usullari

p qiymatlarni osonlikcha hisoblash qobiliyati Stataning kuchli xususiyatidir. Eslatib o'tamiz:

agar $H_1:\beta_k > c$, $p = t$ ning o'ng tomon ehtimolligi

agar $H_1:\beta_k < c$, $p = t$ ning chap tomon ehtimolligi

agar $H_1:\beta_k \neq c$, $p =$ o'ng tomon $|t|$ va chap tomon $-|t|$ ehtimolligining summasi.

Gipotezalarni sinash uchun p qiymat qoidasi: Nol gipotenuzani rad etiladi qachonki p qiymat past yoki unga teng bo'lganda ya'ni α darajadagi qiymatga. Ya'ni, agar $p \leq \alpha$ bo'lsa, H_0 ni rad etadi. Agar $p > \alpha$ bo'lsa, H_0 ni rad qilinmaydi.

T taqsimotining kritik qiymatlarini jadvalda ko'rish yoki invttail funksiyasi yordamida hisoblash mumkin. Ammo kompyuter yordamida p qiymatlarni hisoblash kerak. **Stata** **ttail** funksiyasidan foydalanadi. Buyruqning sintaksisi va uning ta'rifi buyruqlar oynasida **help ttail** buyrug'ini kiritish orqali ko'rish mumkin:

ttail(n,t) "returns the reverse cumulative (upper-tail) Student's t distribution; it returns the probability $T > t$."

Bu yerda **n** erkinlik darajalarining soni va **t** – **t** statistikaning qiymati. **ttail** funksiyasi yana bir marta yuqori darajadagi ehtimollik qiymatini hisoblab ko'rsatadi.

3.3.1. p-qiymatning o'ng tomon qiymatini tekshirish

Yuqoridagi 3.2.2-bo'limda biz $H_0:\beta_2 \leq 5.5$ nol gipotezaga qarshi muqobil gipoteza $H_1:\beta_2 > 5.5$ sinov bajargandik. T statistik qiymatdan foydalanib hisoblagandik:

scalar tstat1 = (_b[income]-5.5)/_se[income]

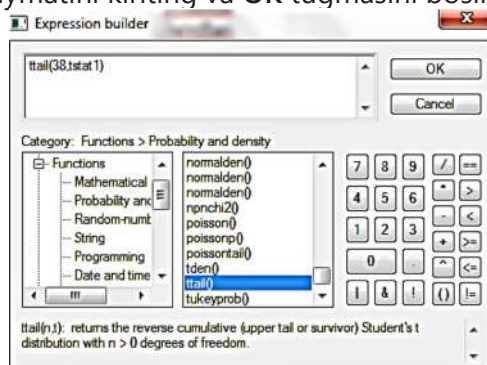
p qiymatidan foydalanib hisoblaymiz va ko'rsatib o'tamiz:

di "p value right tail test ho:beta2 = 5.5 is " ttail(38,tstat1)

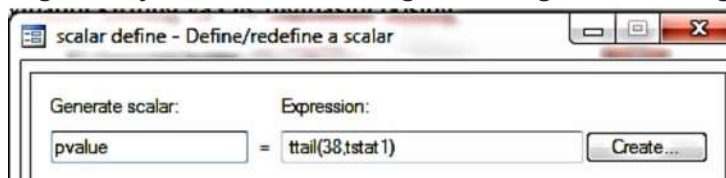
Natija esa quyidagicha:

. di "p value right tail test ho:beta2 = 5.5 is " ttail(38,tstat1)
p value right tail test ho:beta2 = 5.5 is .01516329

Eslatib o'tamiz, agar siz sintaksisini unutib qo'ysangiz, **help scalar** buyrug'ini kiritib, undan **Dialog** menyusidan **scalar define dialog** oynasini ochishingiz mumkin. **Create** tugmasini va keyin **Expression builder** muloqot oynasida **Probability functions**ni tanlang va o'ng tomonga o'ting. Vazifalarning ta'riflari qutining pastki qismida ko'rinadi, bu juda yaxshi yordam beradi. Keyin funksiya nomini ikki marta bosib va **Expression builder** oynasida n erkinlik darajalari va t t statistik qiymatini kiriting va **OK** tugmasini bosib.



Olingan natija uchun nom kiriting va OK tugmasini bosib



3.3.2. p qiymatning chap tomon qiymatini tekshirish

Chap tomonning sinov qiymatini ko'rsatish uchun yuqoridagi 3.2.3-bo'limdagi misoldan foydalaning. $H_0: \beta_2 \geq 15$ nol gipoteza $H_1: \beta_2 < 15$ muqobil faraz bilan bo'lsin! . Hisoblaymiz:

scalar tstat2 = (_b[income]-15)/_se[income]

Chap tomon test p qiymati t taqsimotning chap yoki pastki qismida joylashgan. Buyruqni kiriting:

di "p value left tail test ho:beta2 = 15 is " 1-ttail(38,tstat2)

Biz 1-ttail (38,tstat2)dan foydalanishimiz kerak, chunki tstat2 ning chap tomonining hududini topishni xohlaymiz, qaysiki ttail (38, tstat2) tstat2 ning o'ng tomonini hisoblab chiqadi. Natijada

```
. di "p-value for left-tail test ho:beta2 = 15 is " 1-ttail(38,tstat2)
p-value for left-tail test ho:beta2 = 15 is .01388071
```

3.3.3. p qiymatning ikki tomonli testini tekshirish

Yuqoridagi 3.2.4-bo'limda biz $H_1: \beta_2 \neq 7.5$ muqobil gipotezaga qarshi nol gipoteza $H_0: \beta_2 = 7.5$ ga tekshirgan edik. t statistik dan foydalanib hisoblab chiqilgan:

scalar tstat3 = (_b[income]-7.5)/_se[income]

Ikki tomonli test uchun p qiymati t ning o'ng va chap tomonidagi maydonlar yig'indisidir. Quyidagidan foydalanamiz:

scalar phalf = ttail(38,abs(tstat3))

Ushbu buyruq t statistikaning mutlaq qiymati (abs funksiyasi) o'ng tomonidagi t taqsimotning yuqori qismidagi qismini, ya'ni p qiymatining yarmini hisoblaydi. Shu sababdan ushbu qiymatni 2 ga ko'paytiring va natijani ko'ring:

scalar p3 = 2*phalf

di "p value for two tail test ho:beta2 = 7.5 is " p3

Natija esa:

```
. di "p value for two tail test ho:beta2 = 7.5 is " p3
p value for two tail test ho:beta2 = 7.5 is .20331828
```

Albatta, alohida hisob-kitoblar talab qilinmaydi. Shuning uchun bitta buyruq sifatida kiritib topsa ham bo'ladi.

di "p value for ho:beta2 = 7.5 is " 2*ttail(38,abs(tstat3))

```
. di "p value for ho:beta2 = 7.5 is " 2*ttail(38,abs(tstat3))
p value for ho:beta2 = 7.5 is .20331828
```

3.3.4. STATA tomonidan p qiymatlarni topish

Qachonki regressiya aniqlansa va baholashdan keyingi buyruq `lincom` dan foydalansa, p qiymatlari ko'rsatib o'tiladi. Masalan, regressiya natijasi:

food_exp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
income	10.20964	2.093264	4.88	0.000	5.972052	14.44723
_cons	83.416	43.41016	1.92	0.062	-4.463279	171.2953

koeffitsient nol bo'lganda ikki tomonlama testing p-qiymati

$P>|t|$ deb belgilangan ustun bu koeffitsiyent nolga teng ekanligi haqidagi nol gipoteza uchun ikki tomonli test p qiymatidir. Belgining o'zi P ehtimolligi t statistik qiymatining $|t|$ qiymatidan kattaroq o'rtacha qiymatiga olinadi. Bu, haqiqatan ham t ning musbat qiymatidan katta va t ning manfiy qiymatidan kichikroq bo'lgan ikkita iboraga aylanadi. Agar ahamiyatlilikning bir tomonli testi talab qilinsa, p qiymat ikki tomonli testning p qiymati, modomiki taxmin alternativ gipoteza qondiradi, chunki u taqsimotning bitta tomonida sodir bo'ladi.

`Lincom` ishlatilganda ham bir xil natijalar paydo bo'ladi.

. <code>lincom income-15</code>						
(1) <code>income = 15</code>						
food_exp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
(1)	-4.790357	2.093264	-2.29	0.028	-9.027948	-.5527666

gipoteza $\beta_2 - 15 = 0$ bo'lgandagi ikki tomonlama p-qiyamat

$\beta_2 - 15$ bo'lgandagi hisoblangan qiymat

$\beta_2 - 15$ bo'lgandagi standart xatolik

gipoteza $\beta_2 - 15 = 0$ bo'lgandagi t-statistik qiymat

$\beta_2 - 15 = 0$ ning 95% li interval qiymati

3.3.5. Chiziqli funksiyalarning parametrlarini tekshirish va baholash

Ko'pincha umumiy chiziqli gipoteza ikkala parametрни ham o'z ichiga oladi va shunday aytilishi mumkin:

$$H_0: c_1 \beta_1 + c_2 \beta_2 = c_0$$

Bu erda c_0 , c_1 va c_2 konstantalar ko'rsatilgan. Ushbu gipotezani tekshirishda t statistikadan foydalaniladi:

$$t = ((c_1 b_1 + c_2 b_2) - c_0) / \text{se}(c_1 b_1 + c_2 b_2)$$

Bir va ikki tomonli alternatalarni (i) - (iii) rad etish hududlari 3.3-bo'limda tavsiflanganlar bilan bir xil, xulosalar ham xuddi shunday izohlanadi. t statistikaning aniqlangan xatosi bu ildizning kvadratidir:

$$\text{var}[c_1 b_1 + c_2 b_2] = c_1^2 \text{var}(b_1) + c_2^2 \text{var}(b_2) + 2c_1 c_2 \text{cov}(b_1, b_2)$$

Ushbu qiymatni hisoblash uchun estat vce yordamida hisob-kitobdan keyingi olingan eng kam kvadratlarning hisoblangan kovariyalik matritsasidan foydalaning.

```
. estat vce
Covariance matrix of coefficients of regress model
```

e(V)	income	_cons
income	4.3817522	
_cons	-85.903157	1884.4423

Lincom yordamida biz $c_1 \beta_1 + c_2 \beta_2$ kabi chiziqli funktsiyani taxmin qilishimiz va chiziqli farazning umumiy shaklini sinab ko'rishimiz mumkin. Masalan, Agar $c_1 = 1$ va $c_2 = 20$ bo'lsa, u holda: **lincom _cons + income*20**

```
. lincom _cons + income*20
```

```
( 1) 20*income + _cons = 0
```

food_exp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
(1)	287.6089	14.17804	20.29	0.000	258.9069	316.3108

Ushbu chiziqli funktsiya 250 ga teng bo'ladi, nol gipotezani tekshirganda va undan foydalaning

lincom _cons + income*20 - 250

```
. lincom _cons + income*20 - 250
```

```
( 1) 20*income + _cons = 250
```

food_exp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
(1)	37.60886	14.17804	2.65	0.012	8.906915	66.31081

Uchinchi bob mavzularini mustahkamlash uchun savollar:

1. Kritik qiymatni topish uchun qanday buyruqdan foydalanish kerak?
2. 0,05 ahamiyatlilik darajasidagi sinov uchun t taqsimotning foiz ko'rsatkichi qanday bo'lishi lozim?
3. Statada skalyar shaklida yaratilgan va ular "inverse" funksiyalarni ko'rsatuvchi umumiy prefiksi qaysi buyruqni o'z ichiga oladi?
4. Gipotezalarni sinash uchun p-qiymat qoidasi nimani ifodalaydi?
5. p qiymatning o'ng tomon qiymatni tekshirish qanday amallar orqali bajariladi?
6. p qiymatning chap tomon qiymatni tekshirish ketma-ketligini ifodalang.
7. p qiymatning ikki tomonli testini tekshirish buyrug'ini ifodalang.
8. Iqtisodiy gipotezani left-tail testi deganda nimani tushunasiz?
9. Iqtisodiy gipotezaning right-tail testi va left-tail testi qanday farqlanadi?
10. Umumiy chiziqli gipoteza qanday ifodalanadi?

Uchinchi bob mavzularida qo'llanilgan buyruqlar va atamalar

b[_cons]	builder	recast(area)
_b[varname]	interval estimate	right-tail test
_se[_cons]	invttail	scalar
_se[varname]	left-tail test	set obs
confidence interval	lincom	simulate
coverage probability	linear combinations	tdden
define	of estimates	ttail
estat vce	postestimation	two-tail test
expression	p-value	twoway

Uchinchi bob mavzularini mustahkamlash uchun masalalarning do-file lari

Ma'lumotlar buyrug'i: cd c:\data\poe4stata

1-masala. log file ni ochish va datalardan foydalanish.

log using chap03, replace text

use food, clear

* estimate regression

reg food_exp income

* compute t-critical value

scalar tc975 = invttail(38,.025)

di "t critical value 97.5 percentile = " tc975

* calculating 95% interval estimate

scalar ub2 = _b[income] + tc975*_se[income]

scalar lb2 = _b[income] - tc975*_se[income]

di "beta 2 95% interval estimate is " lb2 " , " ub2

* examples of computing t-critical values

di "t(30) 95th percentile = " invttail(30,0.05)

di "t(20) 95th percentile = " invttail(20,0.05)

di "t(20) 5th percentile = " invttail(20,0.95)

di "t(30) 97.5th percentile = " invttail(30,0.025)

di "t(30) 2.5th percentile = " invttail(30,0.975)

* right-tail test $h_0: \beta_2 = 0$

scalar tstat0 = _b[income]/_se[income]

di "t statistic for $H_0: \beta_2=0$ = " tstat0

di "t(38) 95th percentile = " invttail(38,0.05)

* using lincom

lincom income

* right-tail test $h_0: \beta_2 = 5.5$

scalar tstat1 = (_b[income]-5.5)/_se[income]

di "t-statistic for $H_0: \beta_2 = 5.5$ is " tstat1

di "t(38) 99th percentile = " invttail(38,0.01)

* using lincom for calculation

lincom income-5.5

* left-tail test $h_0: \beta_2 = 15$

```

scalar tstat2 = (_b[income]-15)/_se[income]
di "t-statistic for Ho: beta2 = 15 is " tstat2
di "t(38) 5th percentile = " invttail(38,0.95)
lincom income-15
* two-tail test ho:beta2 = 7.5
scalar tstat3 = (_b[income]-7.5)/_se[income]
di "t-statistic for Ho: beta2 = 7.5 is " tstat3
di "t(38) 97.5th percentile = " invttail(38,0.025)
di "t(38) 2.5th percentile = " invttail(38,0.975)
lincom income-7.5
* two-tail test ho:beta1 = 0
lincom _cons
* p-value for right-tail test
scalar tstat1 = (_b[income]-5.5)/_se[income]
di "p-value for right-tail test ho:beta2 = 5.5 is " ttail(38,tstat1)
* p-value for left-tail test
scalar tstat2 = (_b[income]-15)/_se[income]
di "p-value for left-tail test ho:beta2 = 15 is " 1-ttail(38,tstat2)
* p-value for a two-tail test
scalar tstat3 = (_b[income]-7.5)/_se[income]
scalar phalf = ttail(38,abs(tstat3))
scalar p3 = 2*phalf
di "p-value for two-tail test ho:beta2 = 7.5 is " p3
di "p-value for ho:beta2 = 7.5 is " 2*ttail(38,abs(tstat3))
* linear combinations of parameters
* estimating a linear combination
estat vce
lincom _cons + income*20
* testing a linear combination
lincom _cons + income*20 - 250
log close

```

IV BOB. CHIZIQSIZ REGRESSIYALAR VA MODELLASHTIRISH MASALALARI

4.1. Eng kichik kvadratlarini hisoblash

Biz II bobda $\hat{Y}$ hisoblangan masalaga to'xtalgandik. Endi biz prognozning aniq xatoligini yoki bashorat qilishning aniqligini o'lchash sifatida prognozning standart xatosini topamiz. Ishchi katalogingizga o'ting, bo'lim uchun ishchi faylni oching va oziq-ovqat xarajatlari ma'lumotlarini joylashtiring.

version 11.1

capture log close

set more off

log using chap04_food, replace text

use food, clear

Keling, haftalik daromadi 2000 dollar bo'lgan uy xo'jaliklari uchun uy-ro'zg'or xarajatlarining taxminiy qiymatini topaylik.

4.1.1. Ma'lumotlarni tahrirlash

41-kuzatuvda $\text{income}=20$ ni kiritib, ma'lumotlarni tahrir qilamiz. Bosqichlar ushbu qo'llanmaning 2.5-bo'limida to'liq tushuntirilgan.

edit

set obs 41

replace income = 20 in 41

4.1.2. Regressiyani baholash va keyingi baholash natijalarini olish

Endi biz quyidagi buyruq orqali oziq-ovqat xarajatlarining regressiyasini hisoblaymiz:

quietly regress food_exp income

Hisoblashdan keyingi **predict** buyrug'idan foydalanib,

hisoblangan qiymatlarni (yhat), eng kichik kvadratlar qoldiqlarini (ehat) topamiz. Bularni hisoblashni biz 2.4.1-bo'limda to'liq to'xtalib o'tgandik.

predict yhat

predict ehat, residuals

Biz qo'shadigan yangi buyruq - **stdf**, bu prognozning standart xatosini hisoblab chiqadi. Biz prognoz qilmoqchi bo'lgan yangi kuzatuv topamiz: $y_0 = \beta_1 + \beta_2 x_0 + e_0$

Bu yerda e_0 – tasodifiy xato. $E(y_0) = \beta_1 + \beta_2 x_0$ va $E(e_0) = 0$ deb faraz qilamiz. Shuningdek, biz e_0 ning regressiya xatoliklari bilan bir xil dispersiyaga ega deb faraz qilamiz, $var(e_0) = \sigma^2$ va e_0 namuna ma'lumotlarining bir qismi bo'lgan tasodifiy xatolar bilan bog'liq emas, shuning uchun $cov(e_0, e_i) = 0$, $i = 1, 2, \dots, N$. y_0 ning eng kichik kvadratlar orqali hisoblanib, mos keladigan regressiya chizig'i topiladi: $y_0 = b_1 + b_2 x_0$

Ya'ni, hisoblangan qiymat y_0 eng kichik kvadratlar orqali hisoblangan chiziqning $x = x_0$ nuqtasida berilgan. Ushbu bashoratchi qanchalik yaxshi ishlashini baholash uchun biz eng kichik kvadratlar qoldig'iga o'xshash prognoz xatosini aniqlaymiz:

$$f = y_0 - y_0 = (\beta_1 + \beta_2 x_0 + e_0) - (b_1 + b_2 x_0)$$

Ushbu prognoz xatosining taxminiy farqi

$$var(\widehat{f}) = \hat{\sigma}^2 \left[1 + \frac{1}{N} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right]$$

Bu taxminiy dispersiyaning kvadrat ildizi prognozning standart xatosidir:

$$se(f) = \sqrt{var(\widehat{f})}$$

Quyidagi buyruq yordamida hisoblangan miqdor topiladi.

predict sef, stdf

Ma'lumotlarni ko'rib chiqamiz:

	food_exp	income	yhat	ehat	sef
37	438.29	27.16	360.7099	77.58009	91.99861
38	587.66	28.62	375.616	212.044	92.57295
39	257.95	29.4	383.5795	-125.6295	92.91955
40	375.73	33.4	424.4181	-48.68808	95.11842
41	.	20	287.6089	.	90.63284

yhat tuzilgan modelning hisoblangan qiymati va sef buyrug'i orqali prognozning standart xatosi qiymatlari hisoblanadi. Food_exp bog'liq o'zgaruvchi va eng kichik kvadrat qoldiqlari 41-kuzatuv uchun qiymatlar topilmagan.

4.1.3. $\hat{Y}$ hisoblangan qiymat oralig'ini yaratish

t_c kritik qiymatini t taqsimotidan $100(1 - \alpha/2)$ protsentil deb belgilab, $100(1 - \alpha)\%$ prognozlash oralig'ini quyidagicha olishimiz mumkin: $\hat{y}_0 \pm t_c \text{ se}(f)$

Biz t -kritik qiymatni $\text{invttail}(n, p)$ yordamida yaratamiz:

scalar tc = $\text{invttail}(38, .025)$

di "t critical value 97.5 percentile = " tc

. di "t critical value 97.5 percentile = " tc
t critical value 97.5 percentile = 2.0243942

Hisoblangan qiymat – 2.0243942ga teng. So'ngra biz bu qiymatni yhat va sef bilan birgalikda bashorat qilish oralig'ining pastki va yuqori chegarasi bo'lgan yangi o'zgaruvchilarni yaratish uchun ishlatamiz. Yaratish buyrug'i gen ga qisqartirilib topish mumkin:

gen lb = yhat - tc*sef

gen ub = yhat + tc*sef

Biz bu o'zgaruvchilar qiymatlarini 41-kuzatuvda ro'yxat yordamida ko'rishimiz ham mumkin. Biz list buyrug'i orqali 41-kuzatuvda income, lb, yhat va ub o'zgaruvchilarining miqdorini ifodalaymiz:

list income lb yhat ub in 41

```
. list income lb yhat ub in 41
```

	income	lb	yhat	ub
41.	20	104.1323	287.6089	471.0854

Biz 41-kuzatuvni endi ishlatmaymiz, shuning uchun uni ma'lumotlar faylidan olib tashlaymiz.

drop in 41

drop buyrug'i bir nechta funksiyalarga ega. Ma'lumotlar faylidan, biz yuqorida aytganimizdek, ba'zi kuzatuvlarni o'chirish uchun ishlatilishimiz mumkin va undan o'zgaruvchilarni tushirish uchun foydalanish mumkin. Sintaksisni ko'rish uchun help drop buyrug'iga to'xtalib, qo'shimcha ma'lumotlarni olishingiz mumkin.

4.2. Moslikni o'lchash.

Muvofiqlik o'lchovi R^2 kvadratlarning parchalanish yig'indisi bilan izohlanadi:

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum \hat{e}_i^2$$

Bu "kvadratlar yig'indisi":

$\sum (y_i - \bar{y})^2$ = kvadratlarning umumiy yig'indisi = SST: o'rtacha qiymat bo'yicha y ning umumiy o'zgarish o'lchovi.

$\sum (\hat{y}_i - \bar{y})^2$ = hisoblangan regressiya orqali kvadratlar yig'indisi = SSR: hisoblangan y qiymatining umumiy y ning o'rtacha qiymatidan ayirmasining kvadratining summasi izohlanadi. Shuningdek, "hisoblangan kvadratlar yig'indisi" deb ham ataladi.

$\sum \hat{e}_i^2$ = xatolikning kvadratlar yig'indisi = SSE: regressiyada izohlanmagan, y ning umumiy o'rtacha o'zgaruvchanligi. Kvadratlarning tushuntirilmagan yig'indisi, kvadrat qoldiqlar yig'indisi yoki kvadrat xatolar yig'indisi sifatida ham tanilgan.

Ushbu qisqartmalar yordamida kvadratlarning parchalanishi yig'indisi hosil bo'ladi: **SST=SSR+SSE**

U holda moslik o'lchovi R^2 bo'ladi:

$$R^2 = SSR/SST = 1 - SSE/SST$$

Regressiya baholanganda, bu miqdorlarning barchasi regressiya koeffitsiyentlarining yuqori qismida Analysis of Variance Table paydo bo'ladi. Quydagi rasmdagi ma'lumotlar (Model, Residual, Total) kvadratlarning parchalanish yig'indisi bilan bir xil bo'linishni anglatadi. SS deb belgilangan ustun kvadratlar yig'indisiga ishora qiladi. df ustuni erkinlik darajalari uchun. "Erkinlik darajalari" modeli 1 ga teng, chunki modelda bitta tushuntirish o'zgaruvchisi mavjud. Erkinlikning qoldiq darajalari $N-2$ ni tashkil qiladi, bu kuzatishlar sonidan model parametrlari soni ayirmasidir. MS ustuni Mean Square hisoblanadi. Ushbu ustunda SS ustunining df ustuniga bo'lingan nisbati mavjud. Shunday qilib, Residual MS – bu taxminiy xato dispersiyamizdir.

SSR ifodalangan *x larning soni*

```
. regress food_exp income
```

Source	SS	df	MS
Model	190626.984	1	190626.984
Residual	304505.176	38	8013.2941
Total	495132.16	39	12695.6964

SSE ifodalangan *SST summasi* *N-1* *N-2* *y ning umumiy variatsiyasi* *hisoblangan dispersiya* *a'rtacha kvadratik farq*

R^2 *N*

Number of obs = 40
 F(1, 38) = 23.79
 Prob > F = 0.0000
 R-squared = 0.3850
 Adj R-squared = 0.3688
 Root MSE = 89.517

4.2.1. Korrelatsiya va determinatsiya R^2

Shuningdek, Principle of Econometrics adabiyotida qayd etilishicha, oddiy regressiya modelida R^2 - y va x o'zgaruvchilar orasidagi korrelatsion bog'lanishning kvadrati. Bundan tashqari, umuman olganda, R^2 - y va uning hisoblangan qiymati $\hat{y}$ o'rtasidagi kvadratik korrelatsiyasi ham deb ataladi. Stataning menyusidan foydalanib, quyidagi ketma-ketlik asosida topiladi:

Statistics > Summaries, tables, and tests > Summary and descriptive

statistics > Pairwise correlations

Pairwise correlations muloqot oynasida ochiladigan ro'yxat yordamida o'zgaruvchilar nomlarini kiriting.

Ok tugmasini bosing. So'ngra Stata buyrug'i quyidagicha namoyon bo'ladi:

`pwcorr food_exp income yhat`

Shuningdek, natijalar oynasida Pirson korrelatsiya koeffitsiyenti namoyon bo'ladi:

```
pwcorr food_exp income yhat
```

	food_exp	income	yhat
food_exp	1.0000		
income	0.6205	1.0000	
yhat	0.6205	1.0000	1.0000

So'ngra food_exp va income o'rtasidagi kvadratik bog'liqlikni aniqlang:

$$r_{xy}^2 = 0,62^2 = 0,385 = R^2$$

4.3. Ma'lumotlarni kattalashtirish va o'zgartirish ta'sirlari

Regressiya modelidagi ma'lumotlar hech qanday haqiqiy ma'noni yo'qotmasdan kattalashtirish mumkin. O'zgaruvchilarning miqyosi o'ta katta yoki juda kichik sonlarsiz oqilona o'lchov birliklariga olib kelishi kerak. O'lchovli va o'zgartirilgan o'zgaruvchilar **generate** buyrug'i yordamida osongina yaratiladi.

Oziq-ovqat xarajatlari to'g'risidagi ma'lumotlar uchun o'zgaruvchan daromad 100 AQSH dollari birliklari bilan o'lchanadi, shuning uchun daromad = 20 xonadonning oylik daromadi 2000 AQSh dollarini tashkil etadi. Daromad o'zgaruvchisining miqdorini dollarga o'zgartirish uchun buyruqlar oynasiga quyidagini kiriting:

gen inc_dollar = income*100

Ushbu yangi daromad o'zgaruvchisi bilan oziq-ovqat xarajatlari regressiyasini hisoblang. E'tibor bering, biz regress buyrug'ini reg buyrug'iga qisqartirdik va quyidagi natijalarga erishdik:

```
reg food_exp income
```

food_exp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
income	10.20964	2.093264	4.88	0.000	5.972052 14.44723
_cons	83.416	43.41016	1.92	0.062	-4.463279 171.2953

```
reg food_exp inc_dollar
```

food_exp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
inc_dollar	.1020964	.0209326	4.88	0.000	.0597205 .1444723
_cons	83.416	43.41016	1.92	0.062	-4.463279 171.2953

Regressiya natijalarida daromad koeffitsiyentining o'zgarishi va uning standart xatosi, shu bilan ishonch oralig'i qayd etiladi:

log close

4.3.1. Liner-log funksiyasining shakli

Chiziqli regressiya modeli o'zgaruvchilarning oddiy o'zgarishi yordamida chiziqli bo'lmagan o'zgaruvchilar o'rtasidagi munosabatni ifodalashi mumkin. Oziq-ovqat xarajatlari misoli uchun chiziqli-log modelini quyidagicha ifodalash mumkin:

$$\text{FOOD_EXP} = \beta_1 + \beta_2 \ln(\text{INCOME}) + e$$

Ushbu modelni baholash uchun biz $\ln(\text{income})$ bo'lgan yangi o'zgaruvchini yaratamiz va keyin eng kichik kvadrat regressiyasini qo'llaymiz:

log using chap04_linlog, replace text

use food, clear

gen lnincome = ln(income)

reg food_exp lnincome

. reg food_exp lnincome

Source	SS	df	MS
Model	176519.828	1	176519.828
Residual	318612.333	38	8384.53507
Total	495132.16	39	12695.6964

Number of obs = 40
F(1, 38) = 21.05
Prob > F = 0.0000
R-squared = 0.3565
Adj R-squared = 0.3396
Root MSE = 91.567

food_exp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lnincome	132.1659	28.80461	4.59	0.000	73.85397 190.4777
_cons	-97.18645	84.23744	-1.15	0.256	-267.7162 73.34333

4.3.2-rejada biz foydalanadigan loglinear model regressidan mos qiymatlar va qoldiqlarni olish uchun taxmin qilishdan keyingi buyruq predict dan foydalanamiz:

predict lyhat

predict lehat, resid

Chiziqli logli model uchun qiyalik β_2/INCOME . Buni hisoblash uchun Stata dasturidan qanday foydalanishimiz mumkin? Avval daromad uchun qiymatni tanlashimiz kerak. Namunaviy o'rtacha qiymat, boshqa aniq qiymat qiziqtirmasa, yaxshi tanlov bo'ladi. Daromadlarning umumiy statistikasini hisoblaymiz:

summarize income

Stata hisoblangan bir necha natijalarni saqlaydi. Buni quyidagi buyruqni kiritish orqali ko'rishingiz mumkin:

return list

```
. summarize income
```

variable	Obs	Mean	Std. Dev.	Min	Max
income	40	19.60475	6.847773	3.69	33.4

```
. return list
```

scalars:

```

r(N) = 40
r(sum_w) = 40
r(mean) = 19.60475
r(var) = 46.89198967948718
r(sd) = 6.847772607168492
r(min) = 3.69
r(max) = 33.4
r(sum) = 784.1899999999999

```

Yuqoridagi natijalarni ko'ramiz, unda bir qancha skalarlar ifodalangan, shu jumladan o'rtacha qiymat ham. Uni skalyar sifatida saqlashda quyidagi buyruqni kiriting:

scalar xbar = r(mean)

Endi slop formulasini hisoblash uchun lincom buyrug'dan foydalaning.

lincom lincome/xbar

```
. lincom lincome/xbar
```

(1) .051008*lincome = 0

food_exp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
(1)	6.741522	1.469267	4.59	0.000	3.767147	9.715897

O'rtacha daromad darajasida har haftalik uy xo'jaliklari daromadi 100 dollarga oshishi, oziq-ovqat xarajatlarini 6,74 dollarga ko'payishiga olib keladi. Hisoblangan chiziqli log modelning parametri boshqa qiymatlarda ham hisoblanishi mumkin.

lincom lincome/10

lincom lincome/20

lincom lincome/30

```
. lincom lincome/10
(1) .1*lincome = 0
```

food_exp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
(1)	13.21659	2.880461	4.59	0.000	7.385397 19.04777

```
. lincom lincome/20
(1) .05*lincome = 0
```

food_exp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
(1)	6.608293	1.440231	4.59	0.000	3.692698 9.523887

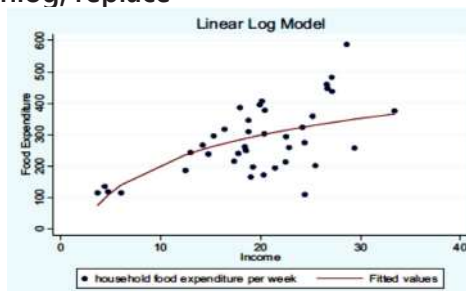
```
. lincom lincome/30
(1) .0333333*lincome = 0
```

food_exp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
(1)	4.405528	.9601537	4.59	0.000	2.461799 6.349258

4.3.2. Hisoblangan liner-log modeli grafigini chizish

Oldingi bo'limda hisoblangan qiymatlar va qoldiqlardan foydalanib, biz twoway grafiklardan foydalangan holda ikkita chizmani topamiz, birinchisi nuqtali chizma, ikkinchisi esa chiziqli chizma. Raqamlar chiroyli tarzda ifodalandi, chunki ma'lumotlar INCOME bo'yicha tartiblangan.

```
twoway (scatter food_exp income, sort) ///
(line lyhat income, sort lwidth(medthick)), ///
xtitle(Income) ytitle(Food Expenditure) ylabel(0(100)600) ///
title(Linear Log Model)
graph save linlog, replace
```



Bu munosabatlar to'g'ri chiziqli emas. Ma'lumotlar markaziga

yaqin joyda o'rnatilgan chiziq juda tik emas. Koeffitsiyent 6,74, oddiy chiziqli regressiyadan kichikroq. Buni aniqroq ko'rish uchun keling, uchta chizmanini ustiga qo'yamiz. Natijani topish va bashorat qilingan qiymatlar va qoldiqlarni olish uchun quietly variantni ishlatib, chiziqli munosabatlarni baholaymiz:

```
quietly reg food_exp income
```

```
predict yhat
```

```
predict ehat, resid
```

Mos keladigan chiziqni boshqa yangi buyruq bilan qo'shing, lpattern(dash).

```
twoway (scatter food_exp income, sort) ///
```

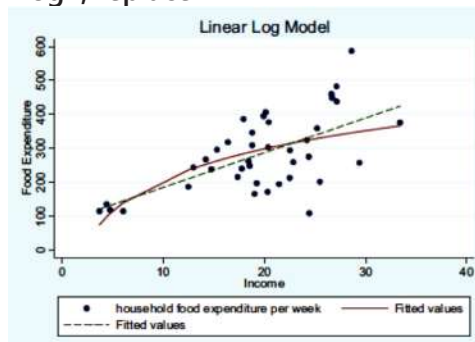
```
(line lyhat income, sort lwidth(medthick)) ///
```

```
(line yhat income, sort lpattern(dash) lwidth(medthick)), ///
```

```
xtitle(Income) ytitle(Food Expenditure) ylabel(0(100)600) ///
```

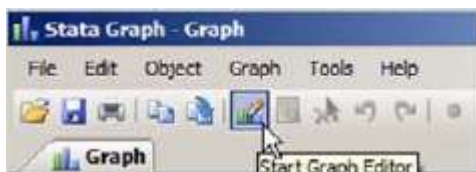
```
title(Linear Log Model)
```

```
graph save linlog2, replace
```

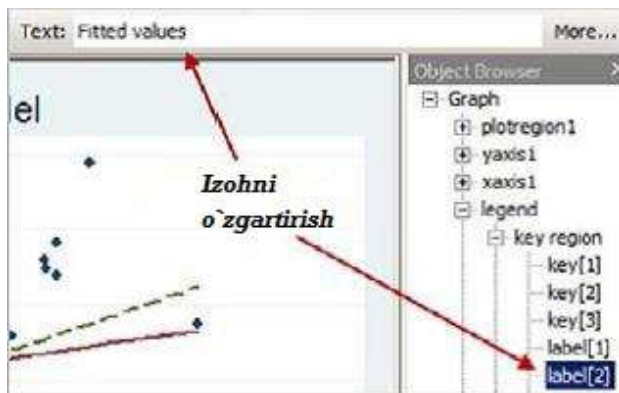


4.3.3. Grafiklarni tahrirlash

Graph Editor yordamida grafikning aspektlarini interaktiv tarzda o'zgartirish mumkin.



O'ng tomonda batafsilroq ro'yxatga ochiladigan elementlarni bildiruvchi "+" bilan grafik elementlarning ro'yxati mavjud. Elementlarni bosish ular bilan bog'langan ajratilgan maydonlarni ko'rsatadi. Chiziqli log modeli uchun "Hisoblangan qiymatlar" izohini o'zgartirish uchun label[2] buyrug'idan foydalaning.



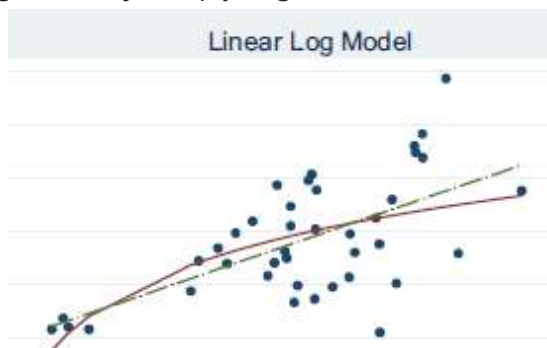
label [2] ni tanlash chiziqni rangi, o'lchami va matni uchun yuqori tahrirlash imkoniyatlarini ko'rsatadi. Faqat kerakli matnni kiriting va Enter tugmasini bosib xolos.

Plotregion va **plot3**ni tanlang. **Plot3** tanlanganda, chiziqli o'rnatilgan chiziq tanlanganligini ko'rsatadigan kesishishlarni ko'rsatadi. Chiziq shaklini to'g'ri chiziqdan boshqasiga o'zgartirish uchun Stataning menyusidan foydalaning.

O'zgartirishlar kiritish uchun **Apply** ni tanlang. Keyin o'zgartirishlaringizni **Save** orqali saqlang va **Stop** orqali plot muharririni to'xtating.



Bizning grafik natijada quyidagicha ifodalaniladi:



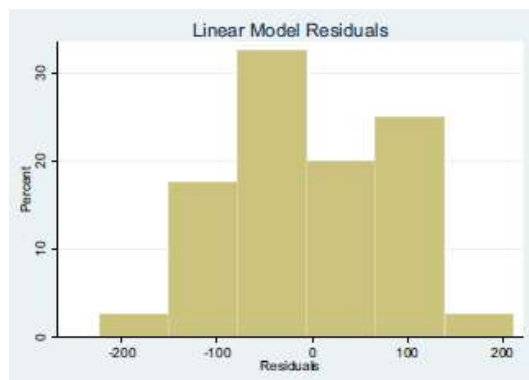
Graph Editor da grafikingizning deyarli barcha jihatlarini boshqarish mumkin. Siz kiritgan tahrirlarni **graph recorder** yordamida yozib olish va boshqa grafiklarda qayta qo'llash mumkin. Ushbu xususiyat haqida ko'proq ma'lumot olish uchun **help graph recorder** ga kirib o'rganishingiz mumkin.

4.4. Qoldiqlarni tahlil qilish

Tanlangan funksional shaklning adekvatligini aniqlashning asosiy elementlaridan biri – bu model qoldiqlarini tahlil qilishdir. Yuqoridagi 4.1.2-rejada chiziqli model uchun eng kichik kvadratlar qoldiqlarini hisoblab chiqdik. Ular xotiradagi ma'lumotlar to'plamida saqlab qoyilgandi. Agar yo'q bo'lsa, oziq-ovqat xarajatlarining chiziqli modelini qayta hisoblang va biz **ehat** deb ataydigan qoldiqlarni hisoblang. Agar sizda kengaytirilgan natijalar bo'lsa, keyinroq foydalanish uchun kerak bo'ladigan o'zgaruvchilarni saqlashni unutmang.

Birinchi qadam sifatida eng kichik kvadratlar qoldiqlarining gistogrammasini tuzing. Biz har doim natija normal taqsimotni eslatib, qo'ng'iroq shaklida bo'lishiga umid qilamiz.

```
histogram ehat, percent title(Linear Model Residuals)  
graph save olsehat_hist, replace
```



Ikkinchidan, biz qoldiqlarni batafsil umumlashtirishimiz mumkin: summarize ehat, detail

. summarize ehat, detail

Residuals					
	Percentiles	Smallest			
1%	-223.0255	-223.0255			
5%	-133.9407	-142.2519			
10%	-115.6628	-125.6295	Obs		40
25%	-52.94326	-119.058	Sum of wgt.		40
50%	-6.324473		Mean		4.77e-08
		Largest	Std. Dev.		88.3619
75%	68.72928	117.4039	Variance		7807.825
90%	112.8848	120.0951	Skewness		-.0973187
95%	121.0697	122.0443	Kurtosis		2.989034
99%	212.044	212.044			

4.4.1. Jarque-Bera testi

Qoldiqlarni normal taqsimlanganini batafsil tahlil qilish uchun Jarque-Bera testidan foydalaniladi, bunda uning tarkibiy qismlari bo'lgan Skewness va Kurtosis koeffitsiyentlarini o'z ichiga oladi. Y o'zgaruvchisi uchun namuna jarayonlarini quyidagicha aniqlang:

$$\tilde{\mu}_2 = \sum (Y_i - \bar{Y})^2 / N = \tilde{\sigma}^2$$

$$\tilde{\mu}_3 = \sum (Y_i - \bar{Y})^3 / N$$

$$\tilde{\mu}_4 = \sum (Y_i - \bar{Y})^4 / N$$

E'tibor bering, ushbu hisob-kitoblarni biz N ga bo'lamiz. Markaziy momentlarning ushbu namunaviy bahalaridan foydalanib, biz skewness koeffitsiyenti (S) va kurtosis koeffitsiyenti (K) ning taxminiy qiymatini olishimiz mumkin:

$$\widehat{skewness} = S = \frac{\tilde{\mu}_3}{\tilde{\sigma}^3}$$

$$\widehat{kurtosis} = K = \frac{\tilde{\mu}_4}{\tilde{\sigma}^4}$$

Yuqorida ta'kidlab o'tilganidek, umumlashtirishdan foydalanilganda, bir qator elementlar saqlanadi. Quyidagi buyruqni kiriting:

return list

Ba'zi elementlar takrorlanadi:

. return list

scalars:

```

r(N) = 40
r(sum_w) = 40
r(mean) = 4.76837158203e-08
r(Var) = 7807.824984910715
r(sd) = 88.36189781184373
r(skewness) = -.0973187365457582
r(kurtosis) = 2.989033831200444

```

Qaytarilgan N, skewness va kurtosis qiymatlaridan foydalanib, Jarque-Bera statistikasini hisoblash uchun Stata (yoki kalkulator) dan foydalanishimiz mumkin.

$$JB = \frac{N}{6} \left(S^2 + \frac{(K - 3)^2}{4} \right)$$

Statistikani yaratish va uni ko'rsatish uchun buyruqlar:

scalar jb = (r(N)/6)*((r(skewness)^2) + ((r(kurtosis)-3)^2)/4)

di "Jarque-Bera Statistic = " jb

Natijalar oynasida quyidagi ko'rinishda natija chiqadi:

```
. di "Jarque-Bera Statistic = " jb
Jarque-Bera Statistic = .06334
```

4.4.2. χ^2 kvadrat taqsimotining kritik qiymatlari

Ushbu taqsimot uchun kritik qiymat 2 daraja erkinlik bilan χ^2 -kvadrat taqsimotidan kelib chiqadi. Kritik qiymatlar **inv** tipidagi funksiyalar yordamida hisoblanadi. To'g'ri funksiya turini topish uchun **help scalar**, so'ngra biz oldin bir necha marta qilganimizdek, belgilash tugmasini bosamiz. Yana bir tezkor alternativ – kirish usulidir:

db scalar

"db" "muloqot oynasi" degan ma'noni anglatadi. Stata buyrug'i:

scalar chic = invchi2tail(2,.05)

di "Chi-square(2) 95th percentile = " chic

Natijalar oynasida quyidagi hisoblangan qiymat:

```
. scalar chic = invchi2tail(2,.05)
```

```
. di "Chi-square(2) 95th percentile = " chic
Chi-square(2) 95th percentile = 5.9914645
```

Hisoblangan Jarque-Bera statistikasining qiymati = 0,06334 testning kritik qiymati = 5,99 dan ancha past, shuning uchun biz regressiya xatolari normal taqsimlangan degan gipotezani rad eta olmaymiz.

4.4.3. χ^2 kvadrat taqsimotining p qiymatlari

χ^2 kvadrat kritik qiymatini olish o'rniga, biz Jarque-Bera test statistik qiymati uchun qiymatni olish orqali testni amalga oshirishimiz mumkin. To'g'ri funksiya turini topish uchun db skaler dan foydalanamiz. Stata buyrug'i esa:

scalar pvalue = chi2tail(2,jb)

di "Jarque-Bera p-value = " pvalue

Olingan p qiymati

```
. scalar pvalue = chi2tail(2,jb)  
  
. di "Jarque-Bera p-value = " pvalue  
Jarque-Bera p-value = .96882624
```

p qiymati 0,9688 ga teng, bu esa I turdagi xatolik $\alpha=0,05$ ehtimolidan katta bo'lganligi sababli, biz avval aytib o'tganimizdek, normallikni rad eta olmaymiz.

log close

4.5. Polinomial modellar

Polinom modeliga misol sifatida biz G'arbiy Avstraliyaning ba'zi fermalarida vaqt o'tishi bilan bug'doy hosili orasidagi bog'liqlikni ko'rib chiqamiz. Ma'lumotlar **wa_wheat.dta** deb nomlangan fayldan foydalanamiz. Stata xotirasini tozalang va ushbu faylni ishga tushiring. O'zgaruvchilar tavsifi va umumiy statistikasi natijasini oling.

log using chap04_wheat, replace text

use wa_wheat, clear

describe

summarize

Variable	Obs	Mean	Std. Dev.	Min	Max
northampton	48	1.168654	.4250324	.3024	2.3161
chapman	48	1.072385	.3328069	.4167	2.0244
mullawa	48	.9840625	.3352854	.3965	1.7992
greenough	48	1.15306	.3653873	.4369	2.2353
time	48	24.5	14	1	48

Xulosa statistik ma'lumotlar shuni ko'rsatadiki, 4 xil fermada $t=48$ kuzatuv mavjud. Shunday qilib, bizning terminologiyamiz Principle of Econometrics, 4e nashirdagi adabiyot bilan parallel bo'lishi uchun, Greenough femasidagi ishlab chiqarishni o'rganish uchun YIELD deb nomlangan yangi o'zgaruvchini yaratamiz:

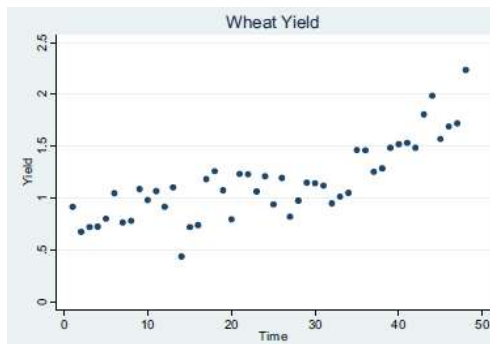
gen yield = greenough

label variable yield "wheat yield greenough shire"

YIELD va TIME o'rtasidagi munosabatni ko'rsatadigan nuqtali grafikni tuzamiz.

twoway (scatter yield time, sort) , ///

```
xtitle(Time) ylabel(0(.5)2.5) ytitle(Yield) ///
title(Wheat Yield)
graph save wawheat, replace
```



4.5.1. Chiziqli munosabatni baholash va tekshirish

Bug'doy hosildorligining o'sish sur'ati, ayniqsa, davr oxiriga kelib ortgan. YIELD va TIME o'rtasidagi chiziqli munosabat mos kelmaydi, chunki u o'zgaruvchan nuqtalarga mos kelmaydi. Ko'rishimiz uchun YIELD va TIME o'rtasidagi chiziqli regressiyani hisoblaymiz va hisoblangan qiymat va eng kichik kvadratlar qoldiqlarini hisoblab olamiz:

reg yield time

predict yhat

predict ehat, residuals

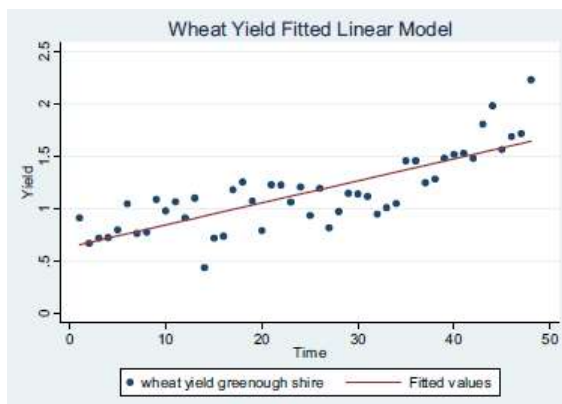
Baholash natijalari quyidagicha bo'ladi:

SS	df	MS				
4.07486005	1	4.07486005	Number of obs = 46			
2.20000947	46	.047826293	F(1, 46) = 85.20			
			Prob > F = 0.0000			
			R-squared = 0.6494			
			Adj R-squared = 0.6411			
6.27486952	47	.133507862	Root MSE = .2186			

Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
.0210319	.0022785	9.23	0.000	.0164455	.0256184
-.6377778	.0641305	-9.94	0.000	-.7668698	-.5086858

Hisoblangan eng kichik kvadratlar chizig'i va ma'lumotlarga mos chiziq bilan bir grafikda ifodalaymiz:

```
twoway (scatter yield time, sort) ///
(line yhat time, sort lwidth(medthick)) , ///
xtitle(Time) ytitle(Yield) ylabel(0(.5)2.5) ///
title(Wheat Yield Fitted Linear Model)
graph save wheat_fit, replace
```

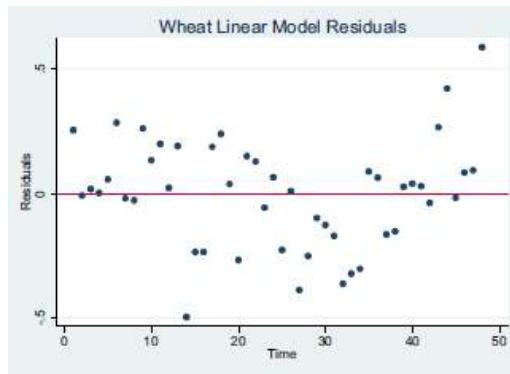


Izoh: Uzoq qoldiqli o'zgaruvchilar grafik ostidagi matnni o'qib bo'lmasligiga olib kelishi mumkin. Grafik muharriri yordamida grafikni tahrirlang va yangi belgini kiriting.

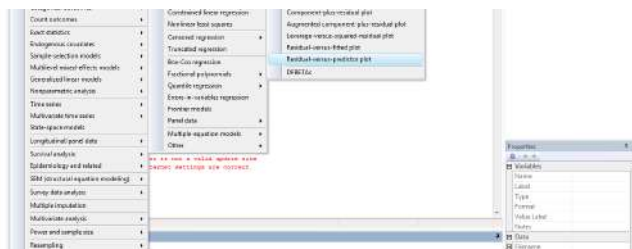
Amaliyot orqali siz vaqt davrining boshida va oxirida ijobiy qoldiqlar klasterini va o'rtadagi salbiy qoldiqlar klasterini aniqlay olasiz. Buni osonroq ko'rish uchun biz qoldiqlarni vaqtga nisbatan chizishimiz mumkin. Oddiy grafikdan foydalanib olish mumkin:

```
twoway (scatter ehat time, sort) , ///
xtitle(Time) ytitle(Residuals) yline(0) ///
title(Wheat Linear Model Residuals)
graph save wheat_ehat, replace
```

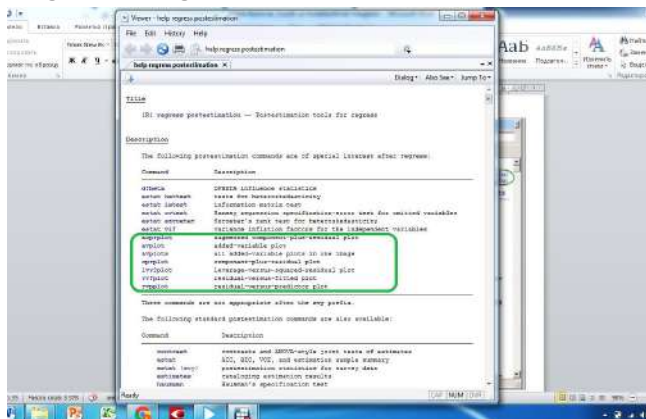
`yline(0)` varianti gorizontall mos yozuvlar chizig'ini nolga tenglashtiradi.



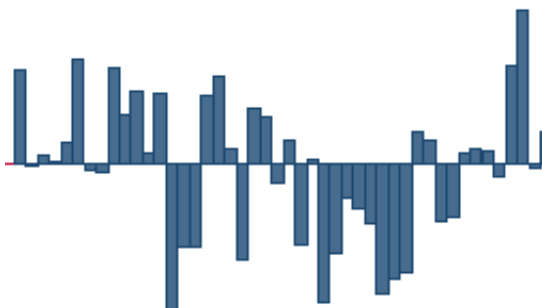
Shu bilan bir qatorda, Stata ba'zi o'rnatilgan qoldiq diagnostika grafiklariga ega. Stata menyusida quyidagilarni tanlang:



help regress postestimation yordamida biz dialoglar va mavjud grafiklarning tavsiflariga havolalarni topa olamiz:



Shuningdek, nuqtali grafikni o'rniga shtrixli grafikni olish uchun `recast (bar)` buyrug'idan foydalanish mumkin,
`rvpplot time, recast(bar) yline(0)`



Ushbu grafik vaqt davrining ikki oxiridagi ijobiy qoldiqlar to'plamini va markazdagi salbiy qoldiqlarning katta klasterini aniq ko'rsatadi. Qoldiqlardagi nuqtalar istalmagan va yaxshiroq funksional shaklni topish zarurligini ko'rsatishi mumkin.

4.5.2. Kubik tenglamani baholash va tekshirish

`time0=time/100` o'zgaruvchisini yarating. Bu o'lchov kub o'zgaruvchining juda katta bo'lishiga yo'l qo'ymaydi va taxminiy koeffitsiyentni nisbatan kattaroq qiladi. Kub tenglamani faktor o'zgaruvchilari yozuvidan foydalanib baholang va o'rnatilgan qiymatlar va qoldiqlarni oling, ularni **yhat3** va **ehat3** deb nomlang, chunki yhat va ehat oldingi regressiyada mavjud edi:

`generate time0=time/100`

`list yield time0 in 1/5`

	yield	time0
1.	.9141	.01
2.	.6721	.02
3.	.7191	.03
4.	.7258	.04
5.	.7998	.05

summarize time0

Variable	Obs	Mean	Std. Dev.	Min	Max
time0	48	.245	.14	.01	.48

reg yield c.time0#c.time0#c.time0

```
. reg yield c.time0#c.time0#c.time0
```

Source	SS	df	MS			
Model	4.71126527	1	4.71126527	Number of obs =	48	
Residual	1.56360425	46	.033991397	F(1, 46) =	138.60	
Total	6.27486952	47	.133507862	Prob > F =	0.0000	
				R-squared =	0.7508	
				Adj R-squared =	0.7454	
				Root MSE =	.18437	

yield	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
c.time0# c.time0# c.time0#	9.681516	.8223546	11.77	0.000	8.026202	11.33683
_cons	.8741166	.0356307	24.53	0.000	.8023958	.9458374

Eng kichik kvadratlar qoldiqlari va taxmin qilingan qiymatlarni hisoblaymiz:

predict yhat3

predict ehat3, residuals

Faktor belgilarining afzalligi shundaki, chekkalar biz uchun qiyalikni to'g'ri hisoblab chiqadi:

margins, dydx(*) at(time=(0.15 0.30 0.45))

	Delta-method dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
time0						
-at						
1	.6535024	.0555089	11.77	0.000	.5447069	.762297
2	2.61401	.2220357	11.77	0.000	2.178827	3.04919
3	5.881521	.4995804	11.77	0.000	4.902361	6.8606

Ma'lumotlarning grafigi va o'rnatilgan chiziq kubik funksiyaning egriligi davr boshida va oxirida ma'lumotlarga yaxshiroq mos kelishini ko'rsatadi:

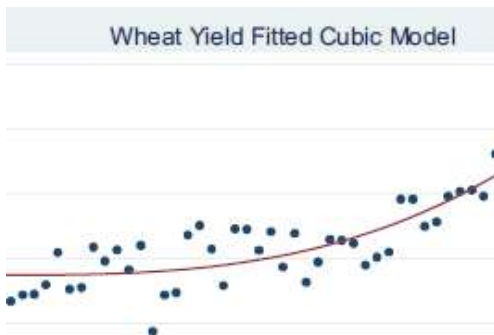
twoway (scatter yield time, sort) ///

(line yhat3 time, sort lwidth(medthick)) , ///

xtitle(Time) ytitle(Yield) ylabel(0(.5)2.5) ///

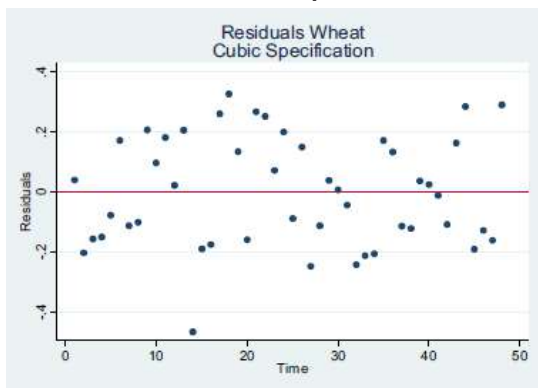
title(Wheat Yield Fitted Cubic Model)

graph save wheat_cubic_fit, replace



Qoldiq diagnostika grafikdan foydalanib, biz ijobiy va salbiy qoldiqlarning qisqarishini aniqlaymiz:

```
twoway (scatter ehat3 time, sort) , ///  
xtitle(Time) ytitle(Residuals) yline(0) ///  
title("Residuals Wheat" "Cubic Specification")  
graph save wheat_cube_ehat, replace
```



4.5.3. Hosildorlikning log-liner o'sish modelini baholash

Oldingi bo'limning kubik modeliga muqobil ravishda log chiziqli o'sish modelidan foydalanish ham mumkin:

```
gen lyield = ln(yield)
reg lyield time
```

```
. reg lyield time
```

Source	SS	df	MS			
Model	2.93313558	1	2.93313558	Number of obs =	48	
Residual	1.82466561	46	.039666644	F(1, 46) =	73.94	
Total	4.75780119	47	.101229813	Prob > F =	0.0000	
				R-squared =	0.6165	
				Adj R-squared =	0.6082	
				Root MSE =	.19916	

lyield	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
time	.0178439	.0020751	8.60	0.000	.0136669	.0220208
_cons	-.3433665	.0584042	-5.88	0.000	-.460928	-.2258049

Biz yillik o'sish sur'atini 1,78% deb hisoblaymiz. log chiziqli funksional shakl haqida keyingi bo'limda yanada chuqurroq to'xtalib o'tamiz.

log close

4.6. Log-liner oylik maosh tenglamasini hisoblash

Yangi faylda ishlashni boshlang, bunda `cps4_small.dta` ma'lumotlar faylidan foydalaning. Shuningdek, o'zgaruvchilar tavsifini ifodalang va tasviriy statistikaning toping. **EDUC** qiymatlarini jadvalda ifodalang:

```
log using chap04_lwage, replace text
```

```
use cps4_small, clear
```

```
describe
```

```
summarize
```

```
tabulate educ
```

Biz foydalanadigan o'zgaruvchilar quyidagilar:

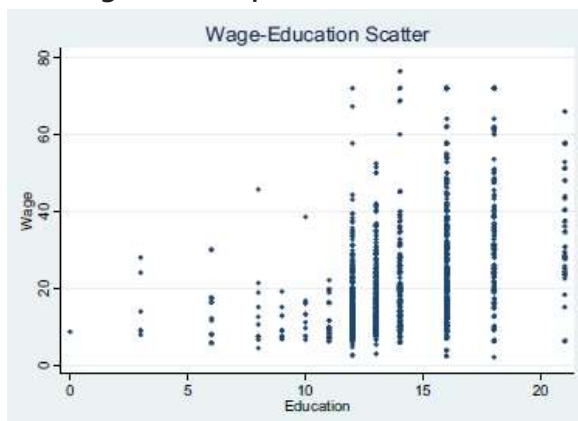
variable name	storage type	display format	value label	variable label
wage	double	%10.0g		earnings per hour
educ	byte	%8.0g		years of education

`msize(small)` varianti bilan kichik belgilardan foydalangan holda ish haqining ta'lim yillariga nisbatan nuqtali grafisini tuzamiz.

```

twoway (scatter wage educ, msize(small)) , ///
xtitle(Education) ytitle(Wage) ///
title(Wage-Education Scatter)
graph save wage_educ, replace

```



Ushbu ma'lumotlardan foydalanib, biz ish haqi va ta'lim yillari o'rtasidagi bog'liqlikni o'rganamiz. Ish haqining ta'limga qarshi grafigi ta'limning yuqori darajalarida ba'zi ish haqining tarqoqlik tendensiyasini ko'rsatadi. Bu turdagi maoshlar, ish haqi va daromad ma'lumotlariga xosdir.

4.6.1. Log-liner modeli

Bu log-chiziqli model, ya'ni ta'sir etuvchi o'zgaruvchini o'zgartirish orqali amalga oshiriladi.

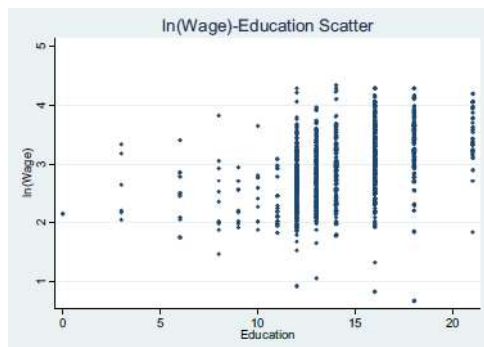
gen lwage = ln(wage)

Endi ta'limga qarshi **ln(wage)** nuqtali grafigni tayyorlaymiz. O'zgartirilgan ma'lumotlar sezilarli darajada kamroq tarqoqlikka ega.

```

twoway (scatter lwage educ, msize(small)), ///
xtitle(Education) ytitle(ln(Wage)) ///
title(ln(Wage)-Education Scatter)
graph save lwage_educ, replace

```



Ushbu misolda biz 12 yillik ta'limdagi ma'lumotga ega bo'lgan shaxslarning ish haqini taxminiy hisoblaymiz. Shunday qilib, 1001 kuzatuviga **EDUC = 12** qiymatini qo'shish orqali ma'lumotlar faylini tahrirlaymiz:

edit

set obs 1001

replace educ=12 in 1001

Log chiziqli regressiyani hisoblaymiz va hisoblangan qiymatlarni, eng kichik kvadratlar qoldiqlarini va prognozning standart xatosini hisoblaymiz:

reg lwage educ

. reg lwage educ

Source	SS	df	MS
Model	60.015841	1	60.015841
Residual	276.76489	998	.27731953
Total	336.780731	999	.337117849

Number of obs = 1000
F(1, 998) = 216.41
Prob > F = 0.0000
R-squared = 0.1782
Adj R-squared = 0.1774
Root MSE = .52661

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
educ	.0904082	.0061456	14.71	0.000	.0783484 .1024681
_cons	1.609444	.0864229	18.62	0.000	1.439853 1.779036

Hisoblangan qiymatlarni, qoldiqlarni va kelajakda foydalanish uchun prognozning standart xatosini hisoblaymiz:

predict lwagehat

predict ehat, residuals

predict sef, stdf

Regressiya natijalariga ko'ra, biz har bir qo'shimcha ta'lim yili ish haqining taxminan 9% oshishiga olib kelishini taxmin qilishimiz mumkin.

Quyida biz hisoblashda xato atamasining taxminiy dispersiyasidan foydalanamiz, shuning uchun regressiya natijalari Stata xotirasida bo'lsa, biz undan foydalanamiz. Regressiyadan keyin Stata tomonidan saqlangan elementlar ro'yxatini ko'rish uchun **ereturn list** buyrug'idan foydalanamiz:

ereturn list

```
. ereturn list
```

```
scalars:
```

```

e(N) = 1000
e(df_m) = 1
e(df_r) = 998
e(F) = 216.4140443756974
e(r2) = .1782044973688944
e(rmse) = .5266113647668089
e(mss) = 60.01584096379838
e(rss) = 276.764890442558

```

Qaytarilgan skalarlar orasida erkinlik darajalari (**df_r**) va kvadrat qoldiqlar yig'indisi (**ssr**) mavjud. Xato atamasining taxminiy dispersiyasi bo'lgan skarlarni hisoblaymiz:

```
scalar sig2 = e(rss)/e(df_r)
```

```
di "sigma-hat squared = " sig2
```

```

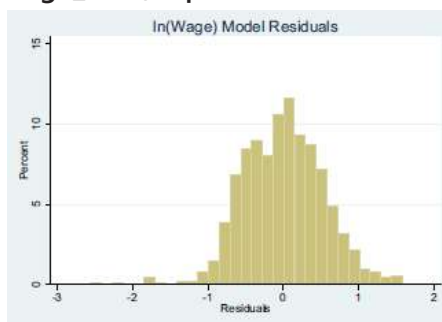
. di "sigma-hat squared = " sig2
sigma-hat squared = .27731953

```

Endi oldingi kabi buyruqlardan foydalanib, gistogramma tuzamiz va Jarque-Bera testini normallik uchun tekshiramiz:

```
histogram ehat, percent title(log(Wage) Model Residuals)
```

```
graph save lwage_ehat, replace
```



summarize ehat, detail

scalar jb = (r(N)/6)*((r(skewness)^2) + ((r(kurtosis)-3)^2)/4)

di "Jarque-Bera Statistic = " jb

scalar chic = invchi2tail(2,.05)

di "Chi-square(2) 95th percentile = " chic

scalar pvalue = chi2tail(2,jb)

di "Jarque-Bera p-value = " pvalue

```
. di "Jarque-Bera Statistic = " jb  
Jarque-Bera Statistic = 27.528329
```

```
. scalar chic = invchi2tail(2,.05)
```

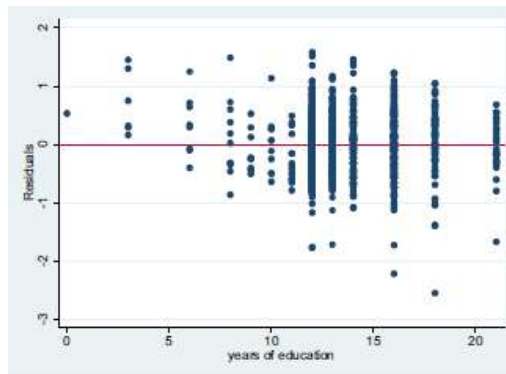
```
. di "Chi-square(2) 95th percentile = " chic  
Chi-square(2) 95th percentile = 5.9914645
```

```
. scalar pvalue = chi2tail(2,jb)
```

```
. di "Jarque-Bera p-value = " pvalue  
Jarque-Bera p-value = 1.053e-06
```

Natijada qoldiqlarning normalligi rad etiladi. So'ngra qoldiqlarning taqsimlanishi yordamida grafikni olamiz:

rvplot educ, yline(0)



Qoldiqlar ham pasayish tendensiyasini ifodalayapti, bu esa ta'lim oshgani sayin ish haqi oshishini ko'rsatadi.

4.6.2. Hisoblangan ish haqlarini hisoblash

Log chiziqli modeldan hisoblangan qiymat ($\ln(y) = b_1 + b_2 x$)[^] dir. y bog'liq o'zgaruvchining bashoratini olish uchun biz "natural logorifimlangan" qiymatini olgan holda logarifimlangan qiymatini olamiz:

$$\hat{y}_n = \exp((\ln(y))) = \exp(b_1 + b_2 x)$$

Stataning `exp()` funksiyasi eksponensial funksiyani ifodalaydi. Hisoblangan $\ln(\text{wage})$ dan foydalanish uchun logorifimlangan o'zgaruvchisini yaratamiz:

$$\text{gen yhatn} = \exp(\text{lwagehat})$$

Ko'pchilik namunalarda yaxshiroq hisoblangan qiymatga erishish uchun "to'g'rilangan" deb nomlanuvchi hisoblangan qiymatlar hisoblanadi:

$$\hat{y}_c = \widehat{E(y)} = \exp\left(b_1 + b_2 x + \frac{\hat{\sigma}^2}{2}\right) = \hat{y}_n e^{\hat{\sigma}^2/2}$$

Eslatib o'tamiz, biz hisoblangan xato dispersiyasi σ^2 ni `sig2` deb hisoblab chiqdik. Ushbu koeffitsiyentni "tuzatish koeffitsiyenti" hisoblangan qiymatlar yordamida hosil qilish mumkin:

```
di "correction factor = " exp(sig2/2)
```

```
. di "correction factor = " exp(sig2/2)
```

```
correction factor = 1.1487332
```

```
gen yhatc = yhatn*exp(sig2/2)
```

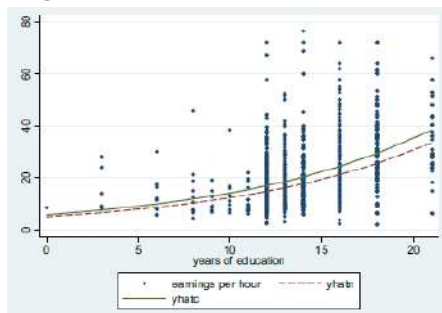
Tuzatish koeffitsiyenti 1,1487 ni tashkil qildi va shuning uchun "tuzatilgan" hisoblangan qiymatlar "tabiiy" qiymatlardan kattaroqdir. Bu har doim shunday bo'ladi, chunki har qanday $a > 0$ qiymati uchun tuzatish koeffitsiyenti $e^a > 1$ va taxminiy dispersiya σ^2 har doim noldan katta.

4.6.3. Ish haqi grafiklarini tuzish

Ikki hisoblangan qiymatlarni taqqoslaydigan grafikda ko'rish yanada yaxshi xulosalar chiqarish mumkin:

```
twoway (scatter wage educ, sort msize(small)) ///
```

```
(line yhatn educ, sort ///
lwidth(medthick) lpattern(dash)) ///
(line yhatc educ, sort lwidth(medthick) lpattern(solid))
graph save lwage_predict, replace
```



Grafik **yhatc** har doim **yhatn** dan katta ekanligini ko'rsatadi. Tabiiy va hisoblangan qiymatlarni ko'rish uchun biz 1001 kuzatuvni qiymatini ko'rib o'tamiz:

list educ yhatn yhatc in 1001

	educ	yhatn	yhatc
1001.	12	14.7958	16.99643

Bu qiymatlar mantiqiyimi? Bizning namunamizda 12 yillik ta'lim ma'lumotga ega bo'lganlar uchun ish haqi bo'yicha umumiy statistikani hisoblash orqali biz taxminiy tekshirishimiz mumkin:

summarize wage if educ==12 in 1/1000

. summarize wage if educ==12 in 1/1000					
Variable	Obs	Mean	Std. Dev.	Min	Max
wage	328	15.99329	8.843706	2.5	72.13

Biz prognozlar 12 yillik ta'limga ega bo'lgan shaxslarning haqiqiy o'rtacha ish haqiga mos kelishini ko'rishimiz mumkin.

4.6.4. Determinatsiya koeffitsiyenti R^2

Determinatsiya koeffitsiyenti R^2 o'lchovi "eng yaxshi" hisoblangan qiymatni baholovchi **yhatc** va o'zgaruvchan **income**

o'rtasidagi korrelyatsiya kvadratidir. Bu holda **yhatc** va **yhatn** faqat multiplikativ konstanta bilan farqlanadi, shuning uchun ularning **income** bilan bog'liqligi bir xil bo'ladi. Korrelyatsiyalarni hisoblash uchun quyidagi buyruqdan foydalaning:

```
correlate wage yhatn yhatc
```

```
di "r2g = " r(rho)^2
```

```
. correlate wage yhatn yhatc  
(obs=1000)
```

	wage	yhatn	yhatc
wage	1.0000		
yhatn	0.4312	1.0000	
yhatc	0.4312	1.0000	1.0000

```
. di "r2g = " r(rho)^2  
r2g = .18593072
```

4.6.5. Log-liner modeldagi bashorat oraliqlari

Log-chiziqli modeldagi bashorat oralig'i sifatida quyidagi formula tuzilgan.

$$\left[\exp\left(\widehat{\ln(y)} - t_c se(f)\right), \exp\left(\widehat{\ln(y)} + t_c se(f)\right) \right]$$

U "tabiiy" hisoblangan qiymatga va yuqoridagi 4.6.1-bo'limda tuzilgan prognozning standart xatosiga asoslanadi. Birinchidan, 998 daraja erkinlik bilan t taqsimotidan 97,5 foizli ehtimollikni hisoblaymiz:

```
scalar tc = invttail(998,.025)
```

ln(income) ning pastki va yuqori prognoz oraliqlarini topamiz:

```
gen lb_lwage = lwagehat - tc*sef
```

```
gen ub_lwage = lwagehat + tc*sef
```

Ekspontensial funksiyadan foydalanib, logorifim qiymatlarini topamiz:

```
gen lb_wage = exp(lb_lwage)
```

```
gen ub_wage = exp(ub_lwage)
```

1001-kuzatish uchun bashorat oralig'ining qiymatlarini ko'rib o'tamiz:

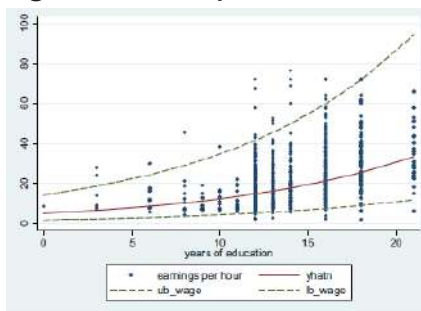
list lb_wage ub_wage in 1001

```
. list lb_wage ub_wage in 1001
```

	lb_wage	ub_wage
1001.	5.260397	41.61581

Ish haqi ma'lumotlari, tabiiy hisoblangan qiymat va ta'limga qarshi intervalli hisoblangan qiymat bilan grafik yaratamiz:

```
twoway (scatter wage educ, sort msize(small)) ///  
(line yhatn educ, sort lwidth(medthick) lpattern(solid)) ///  
(line ub_wage educ, sort lcolor(forest_green) lwidth(medthick) ///  
lpattern(dash)) ///  
(line lb_wage educ, sort lcolor(forest_green) lwidth(medthick) ///  
lpattern(dash))  
graph save lwage_interval, replace
```



Natijalarga erishilgandan so'ng ishchi faylni yopish tavsiya qilinadi va buni quyidagi buyruq orqali amalga oshiring:

```
log close
```

4.7. Log-log modeli

Log-log funksiyasi $\ln(y) = \beta_1 + \beta_2 \ln(x)$ talab tenglamalari va ishlab chiqarish funksiyalarini hisoblash uchun keng qo'llaniladi. "Log-log" nomi logarifm tenglamaning har ikki tomonida paydo bo'lishidan kelib chiqadi. Ushbu modeldan foydalanish uchun y va x

ning barcha qiymatlari ijobiy bo'lishi kerak. Ushbu egri chiziqlarning qiyaliklari har bir nuqtada o'zgaradi, lekin elastiklik doimiy va β_2 ga teng. Log-log funksiyasi haqida fikr yuritishning foydali usuli uning elastikligi $\frac{dy}{dx} = \beta_2 \left(\frac{y}{x}\right)$ ni yaqinroq tekshirishdan kelib chiqadi.

Buni $\beta_2 = \left(\frac{dy}{y}\right) / \left(\frac{dx}{x}\right)$ ga o'zgartiring. Shunday qilib, log-log funksiyasining elastikligi doimiy nisbiy o'zgarishlarni ko'rsatadi, chiziqli funksiya esa doimiy mutlaq o'zgarishlarni ko'rsatadi. Yangi ishchi faylini oching va **newbroiler.dta** ma'lumotlaridan foydalaning:

```
log using chap04_loglog, replace text
use newbroiler, clear
describe
summarize
```

Bunda sifat o'zgaruvchilarni ham o'z ichiga olgan:

variable name	storage type	display format	value label	variable label
year	float	%9.0g		year
q	float	%9.0g		per capita consumption of boneless chicken, pounds
y	float	%9.0g		per capita real disposable income, 1996 = 100
p	float	%9.0g		real price (index) of fresh chicken

Miqdor va narxning logarifmini tuzing va log-log modelini baholang. Bu yerda logga ekvivalent bo'lgan ln log funksiyasidan foydalanamiz:

```
gen lq = ln(q)
gen lp = ln(p)
reg lq lp
```

Source	SS	df	MS
Model	7.36410139	1	7.36410139
Residual	.69608941	50	.013921788
Total	8.0601908	51	.158042957

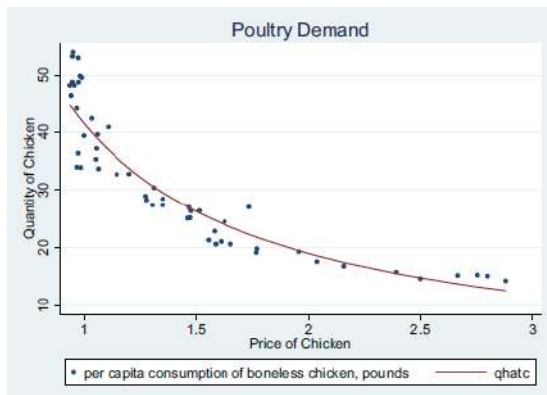
```
Number of obs = 52
F( 1, 50) = 528.96
Prob > F = 0.0000
R-squared = 0.9136
Adj R-squared = 0.9119
Root MSE = .11799
```

lq	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lp	-1.121358	.0487564	-23.00	0.000	-1.219288 -1.023428
_cons	3.716944	.0223594	166.24	0.000	3.672034 3.761854

Hisoblangan elastiklik -1,12 ga teng hisoblanadi.

Hisoblangan qiymatlar va grafikni tuzamiz. Plot buyruqlari biroz uzun, shuning uchun biz ularni bu yerda ikkinchi qatorga ifodaladik, lekin uni do-faylda ifodalamaymiz:

```
predict lqhat
scalar sig2 = e(rss)/e(df_r)
gen qhatc = exp(lqhat)*exp(sig2/2)
twoway (scatter q p, sort msize(small) lwidth(medthick) ///
lpattern(solid)) ///
(line qhatc p, sort lwidth(medthick)), ///
xtitle(Price of Chicken) ytitle(Quantity of Chicken) ///
title(Poultry Demand)
```



Determinatsiya koeffitsiyenti R^2 ni hisoblaymiz:

```
correlate q qhatc
```

```
di "r2g = " r(rho)^2
```

```
log close
```

```
. correlate q qhatc
(obs=52)
```

	q	qhatc
q	1.0000	
qhatc	0.9390	1.0000

```
. di "r2g = " r(rho)^2
r2g = .88177576
```

To'rtinchi bob mavzularini mustahkamlash uchun savollar:

1. Regressiyani baholash uchun qanday amallar ketma-ketligi bajarilishi kerak?
2. $Y^{\wedge}$ hisoblangan qiymat oralig'ini yaratish buyrug'i qanday?
3. drop buyrug'i funksiyalarini sanang.
4. Muvofiqlik o'lchovini ifodalang.
5. Oddiy regressiya modelida R^2 nimani ifodalaydi?
6. O'lchovli va o'zgartirilgan o'zgaruvchilar qanday buyruq orqali yaratiladi?
7. Hisoblangan liner-log modeli grafigini chizish ketma-ketligini yozing.
8. Grafik yaratilganidan so'ng uni qayta tahrirlash qanday amalga oshiriladi?
9. Jarque-Bera testi vazifasi nima?
10. Kub tenglamani faktor o'zgaruvchilari yozuvidan foydalanib baholash qanday ketma-ketlik asosida amalga oshiriladi?
11. Hisoblangan ish haqi grafiklarini tuzishni amalda bajaring.
12. Log-log funksiyasi qay hollarda qo'llaniladi?

To'rtinchi bob mavzularida qo'llanilgan buyruqlar va atamalar

analysis of variance	dydx	invchi2
table	edit	invchi2tail
chi2	ereturn list	Jarque-Bera test
chi2tail	exponential function	kurtotsis
connected	expression builder	lcolor
correlate	factor variables	lfit
correlations	gen	lincom
create	generate	linear-log model
data editor	goodness of fit	list
db scalar	graph editor	ln
drop	histogram	log

log-linear model	prediction interval	set obs
log-log model	regress	skewness
lpattern	replace	sort on x variable
lwidth	residual diagnostic	standard error of
margins	plots	forecast
msymbol	return list	stdf
natural log	R-squared	summarize
plot definition	rvpplot	twoway
predict	scalar	yline
prediction	scatter	

To'rtinchi bob mavzularini mustahkamlash uchun masalalarning do-file lari

Ma'lumotlar buyrug'i: `cd c:\data\poe4stata`

1-masala. log file ni ochish va datalardan foydalanish.

`log using chap04_food, replace text`

`use food, clear`

`* add observation`

`edit`

`set obs 41`

`replace income=20 in 41`

`* estimate regression`

`quietly regress food_exp income`

`predict yhat`

`predict ehat, residuals`

`predict sef, stdf`

`* compute t-critical value`

`scalar define tc = invttail(38,.025)`

`di "t critical value 97.5 percentile = " tc`

`gen lb = yhat - tc*sef`

`gen ub = yhat + tc*sef`

`list income lb yhat ub in 41`

`drop in 41`

`* R2`

```

pwcrr food_exp income yhat
* effect of scaling
* create $ income and regress
gen inc_dollar = income*100
reg food_exp income
reg food_exp inc_dollar
log close

```

2-masala. log file ni ochish

```

log using chap04_linlog, replace text
* open data
use food, clear
* log of income
gen lincome = ln(income)
* linear-log regression
reg food_exp lincome
predict lyhat
predict lehat, resid
* slope = beta2/x
summarize income
return list
scalar xbar = r(mean)
lincom lincome/xbar
lincom lincome/10
lincom lincome/20
lincom lincome/30
* fitted value plot
twoway (scatter food_exp income, sort) ///
(line lyhat income, sort lwidth(medthick)), ///
xtitle(Income) ytitle(Food Expenditure) ylabel(0(100)600) ///
title(Linear Log Model)
graph save linlog, replace
* linear relationship
quietly reg food_exp income
predict yhat
predict ehat, resid

```

```

* linear and linear-log fitted lines
tway (scatter food_exp income, sort) ///
(line lyhat income, sort lwidth(medthick)) ///
(line yhat income, sort lpattern(dash) lwidth(medthick)), ///
xtitle(Income) ytitle(Food Expenditure) ylabel(0(100)600) ///
title(Linear Log Model)
graph save linlog2, replace
* plot linear-log model residuals
tway (scatter lehat income, sort) , ///
xtitle(Income) ytitle(Residuals) ///
title(Linear Log Model Residuals)
graph save linlog_residual, replace
* analyze residuals from original equation
histogram ehat, percent title(Linear Model Residuals)
graph save olsehat_hist, replace
* Jarque-Bera test of error normality
summarize ehat, detail
return list
scalar jb = (r(N)/6)*( (r(skewness)^2) + ((r(kurtosis)-3)^2)/4 )
di "Jarque-Bera Statistic = " jb
scalar chic = invchi2tail(2,.05)
di "Chi-square(2) 95th percentile = " chic
scalar pvalue = chi2tail(2,jb)
di "Jarque-Bera p-value = " pvalue
log close
3-masala. log file ni ochish va datalardan foydalanish.
log using chap04_wheat, replace text
* open data and examine
use wa_wheat, clear
describe
summarize
gen yield = greenough
label variable yield "wheat yield greenough shire"
* plot data
tway (scatter yield time, sort msymbol(circle)) ,

```

```

xtitle(Time) ylabel(0(.5)2.5) ytitle(Yield) ///
title(Wheat Yield) ///
graph save wawheat, replace
* regression
reg yield time
predict yhat
predict ehat, residuals
* plot fitted lines and data
twoway (scatter yield time, sort) ///
(line yhat time, sort lwidth(medthick)) , ///
xtitle(Time) ytitle(Yield) ylabel(0(.5)2.5) ///
title(Wheat Yield Fitted Linear Model)
graph save wheat_fit, replace
* plot residuals
twoway (scatter ehat time, sort) ,
xtitle(Time) ytitle(Residuals) yline(0) ///
title(Wheat Linear Model Residuals) ///
graph save wheat_ehat, replace
rvpplot time, recast(bar) yline(0)
graph save wheat_ehat_bar, replace
* Cubic equation for yield
* create scaled cubic variable
generate time0=time/100
list yield time0 in 1/5

summarize time0
* cubic regression
reg yield c.time0#c.time0#c.time0
predict yhat3
predict ehat3, residuals
* slopes
margins, dydx(*) at(time=(0.15 0.30 0.45))
* plot fitted lines and data
twoway (scatter yield time, sort) ///
(line yhat3 time, sort lwidth(medthick)) , ///
xtitle(Time) ytitle(Yield) ylabel(0(.5)2.5) ///

```

```

title(Wheat Yield Fitted Cubic Model)
graph save wheat_cubic_fit, replace
* plot residuals
twoway (scatter ehat3 time, sort) ,
xtitle(Time) ytitle(Residuals) yline(0) ///
title("Residuals Wheat" "Cubic Specification") ///
graph save wheat_cube_ehat, replace
* Log-linear Models
* Wheat growth model
gen lyield = ln(yield)

reg lyield time
log close
* Wage Equation
* open new log file
log using chap04_lwage, replace text
* open cps4_small data
use cps4_small, clear
* summarize and plot
describe
summarize
tabulate educ
twoway (scatter wage educ, msize(small)) ,
xtitle(Education) ytitle(Wage) ///
title(Wage-Education Scatter)///
graph save wage_educ, replace
* create log(wage) and plot
gen lwage = ln(wage)
twoway (scatter lwage educ, msize(small)),
xtitle(Education) ytitle(ln(Wage)) ///
title(ln(Wage)-Education Scatter) ///
graph save lwage_educ, replace
* log-linear regression
* add one observation
edit

```

```

set obs 1001
replace educ=12 in 1001
reg lwage educ
predict lwagehat
predict ehat, residuals
predict sef, stdf
* calculate sigma-hat^2
ereturn list
scalar sig2 = e(rss)/e(df_r)
di "sigma-hat squared = " sig2
* Analyze residuals
histogram ehat, percent title(ln(Wage) Model Residuals)
graph save lwage_ehat, replace
summarize ehat, detail
scalar jb = (r(N)/6)*( (r(skewness)^2) + ((r(kurtosis)-3)^2)/4 )
di "Jarque-Bera Statistic = " jb
scalar chic = invchi2tail(2,.05)
di "Chi-square(2) 95th percentile = " chic
scalar pvalue = chi2tail(2,jb)
di "Jarque-Bera p-value = " pvalue
rvpplot educ, yline(0)
* compute natural and corrected predictor and plot
gen yhatn = exp(lwagehat)
di "correction factor = " exp(sig2/2)
gen yhatc = yhatn*exp(sig2/2)
twoway (scatter wage educ, sort msize(small)) ///
(line yhatn educ, sort ///
lwidth(medthick) lpattern(dash))///
(line yhatc educ, sort lwidth(medthick) lpattern(solid))
graph save lwage_predict, replace
* list predicted values
list educ yhatn yhatc in 1001
summarize wage if educ==12 in 1/1000
* R^2
correlate wage yhatn yhatc

```

```

di "r2g = " r(rho)^2
* prediction interval
scalar tc = invttail(998,.025)
gen lb_lwage = lwagehat - tc*sef
gen ub_lwage = lwagehat + tc*sef
gen lb_wage = exp(lb_lwage)
gen ub_wage = exp(ub_lwage)
* list and plot
list lb_wage ub_wage in 1001
twoway (scatter wage educ, sort msize(small)) ///
(line yhatn educ, sort lwidth(medthick) lpattern(solid)) ///
(line ub_wage educ, sort lcolor(forest_green) lwidth(medthick) ///
lpattern(dash)) ///
(line lb_wage educ, sort lcolor(forest_green) lwidth(medthick) ///
lpattern(dash))
graph save lwage_interval, replace
log close

```

4-masala. log file ni ochish va datalardan foydalanish.

```

* A log-log model example
log using chap04_loglog, replace text
use newbroiler, clear
describe
summarize
gen lq = ln(q)
gen lp = ln(p)
reg lq lp
predict lqhat
scalar sig2 = e(rss)/e(df_r)
gen qhatc = exp(lqhat)*exp(sig2/2)
twoway (scatter q p, sort msize(small) lwidth(medthick) ///
lpattern(solid)) (line qhatc p, sort lwidth(medthick)), ///
xtitle(Price of Chicken) ytitle(Quantity of Chicken) ///
title(Poultry Demand)
correlate q qhatc
di "r2g = " r(rho)^2
log close

```

V BOB. KO'P OMILLI REGRESSION MODELLAR

5.1. Big andy's burger barn ma'lumotlari asosida model tuzish

Oddiy chiziqli regressiyada erksiz o'zgaruvchining o'rtacha qiymati doimiy va bitta erkli o'zgaruvchining chiziqli funksiyasi sifatida modellashtiriladi. Ko'p chiziqli regressiya modelida erkli o'zgaruvchilar sonini kengaytiradi. Shunday qilib, bu oddiy, ammo muhim kengaytma bo'lib, chiziqli regressiyani juda kuchli qiladi.

Ushbu bobda foydalanilgan misol Big Andy's Burger Barn uchun savdo modeli tuziladi. Big Andy gamburgerining sotuvi uning narx va reklama darajasiga bog'liqligini ifodalaniladi. Shunday qilib, model ikkita ta'sir etuvchi omill va o'zgarmasni (constant)ni o'z ichiga oladi:

$$\text{SALES} = \beta_1 + \beta_2 \text{ PRICE} + \beta_3 \text{ ADVERT} + e$$

Bu yerda SALES ma'lum bir shahardagi oylik sotuvlar hajmi (1000 AQSH dollari miqdorida), PRICE – gamburgerning narxi (dollarda) va ADVERT – reklama xarajatlari (ming dollarlarda) shuningdek, $i = 1, 2, \dots, N$.

Birinchidan, Stata ni ishga tushiring va buyruq satridan ishchi katalogni ma'lumotlar fayllaringiz joylashgan joyga o'zgartiring yoki Internet manzilida saqlangan fayllardan foydalanmoqchi bo'lsangiz, Stata jurnali faylini yozmoqchi bo'lgan joy. Barcha misollarimizda bo'lgani kabi, `cd c:\data\poe4stata`

Biz ochmoqchi bo'lgan faylni toping, bu yerda andy.dta ni tanlaymiz va Open tugmasini bosing. Har qanday modelni taxmin qilishdan oldin, ma'lumotlarning dasturiy ta'minotigizga to'g'ri yuklanganligini tekshirish yaxshi amaliyotdir. Bunday holda, umumiy statistikani tekshiring va dastlabki bir nechta kuzatuvlarni sanab o'ting. Ular qo'llanmaning 5.1-jadvalidagilarga mos kelishi kerak. Asosan, siz ular mantiqiy yoki mantiqsizligini bilish uchun umumiy statistikaga qarashimiz kerak. Sizda kerakli kuzatuvlar soni

bormi? O'zgaruvchilar diapazoni o'rtacha ko'rinadimi – bu holda savdo, narxlar va reklama xarajatlari ijobiymi?

Eslatib o'tamiz, summarize buyrug'i umumiy statistik ma'lumotlarning asosiy to'plamini va list ma'lumotlar to'plamidagi o'zgaruvchilar hamda kuzatishlarni ifodalaydi. Ro'yxatga qo'shilgan in 1/5 qismi Stataga chop etishni 1 dan 5 gacha bo'lgan kuzatishlar bilan cheklashni aytadi. Natijalar:

```
. summarize
```

Variable	obs	Mean	Std. Dev.	Min	Max
sales	75	77.37467	6.488537	62.4	91.2
price	75	5.6872	.518432	4.83	6.49
advert	75	1.844	.8316769	.5	3.1

```
. list in 1/5
```

	sales	price	advert
1.	73.2	5.69	1.3
2.	71.8	6.49	2.9
3.	62.4	5.63	.8
4.	67.4	6.22	.7
5.	89.3	5.02	1.5

Odatda vaqtni tejaydigan buyruq satridan regressiyani baholashni tanlashingiz mumkin: **regress sales price advert**

Agar sintaksisni eslay olmasangiz, ochiladigan menyulardan foydalaning. **Statistics > Linear models and related > Linear regression**-ni tanlang. Bu 2.4-bo'limda ko'rsatilgan regress dialog oynasini ochadi. **Sales** ni bog'liq o'zgaruvchi sifatida kiriting, **price** va **advert**ni mustaqil o'zgaruvchi sifatida kiriting hamda "yuborish" tugmasini bosing. Esda tutingki, Stata modelga kesishuvni joylashtiradigan doimiy (constant) o'zgaruvchini o'z ichiga oladi. Agar sizda modeldagi kesishuvni olib tashlash uchun jiddiy sabab bo'lmasa, har doim bittasini kiritishingiz kerak. Demak, bu **Stata** uchun standart hisoblanadi. Natija quyida keltirilgan:

Source	SS	df	MS			
Model	1396.53893	2	698.269465	Number of obs =	75	
Residual	1718.94294	72	23.8742075	F(2, 72) =	29.25	
				Prob > F =	0.0000	
				R-squared =	0.4483	
				Adj R-squared =	0.4329	
				Root MSE =	4.8861	
Total	3115.48187	74	42.1011063			

sales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
price	-7.907854	1.095993	-7.22	0.000	-10.09268	-5.723032
advert	1.862584	.6831955	2.73	0.008	.500659	3.22451
_cons	118.9136	6.351638	18.72	0.000	106.2519	131.5754

Kesish uchun parametr **_cons** va β_2 va β_3 qiyaliklar uchun mos ravishda price va advert bilan belgilanadi. Dispersiya-kovariatsiya

matritsasi eng kichik kvadratlar usuli sizning modelingiz parametrlarini o'lchashga qodir bo'lgan aniqlikni ifodalaydi. Eng kichik kvadratlarning aniqligi bir qancha narsalarga, jumladan, ma'lumotlaringizning o'zgaruvchanligiga (σ^2), namunangizning o'lchamiga va mustaqil o'zgaruvchilarning raqamli qiymatlari nazarda tutilgan "tajriba" dizayniga bog'liq. Ushbu ma'lumot kesma, har bir qiyalik va ular orasidagi har qanday kovariatsiyaning o'lchovini o'z ichiga olgan taxminiy variatsiya-kovariatsiya matritsasi bilan umumlashtiriladi.

$$\text{cov}(b_1, b_2, b_3) = \begin{bmatrix} \text{var}(b_1) & \text{cov}(b_1, b_2) & \text{cov}(b_1, b_3) \\ \text{cov}(b_1, b_2) & \text{var}(b_2) & \text{cov}(b_2, b_3) \\ \text{cov}(b_1, b_3) & \text{cov}(b_2, b_3) & \text{var}(b_3) \end{bmatrix}$$

Eng kichik kvadratlar hisoblagichining dispersiyalari diagonalga to'g'ri keladi va har bir juftlik orasidagi kovariatsiya pastki uchburchakda bo'ladi.

Regressiyadan keyin dispersiya-kovariatsiya matritsasi taxminini topish uchun quyidagi buyruqdan foydalaning va natijani oling:

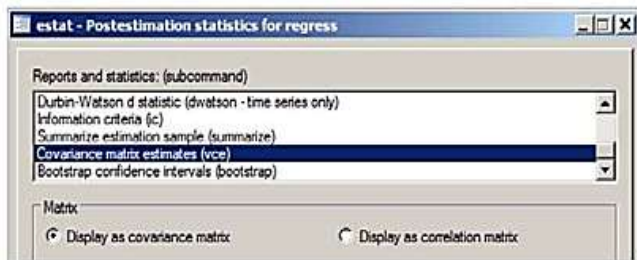
```
. estat vce
```

Covariance matrix of coefficients of regress model

e(V)	price	advert	_cons
price	1.2012007		
advert	-.01974215	.46675606	
_cons	-6.7950641	-.7484206	40.343299

Shunday qilib, b_2 ning taxminiy dispersiyasi 1,20 ga, b_3 bilan taxminiy kovariatsiyasi esa 0,0197 ga teng. Diagonal elementlarning kvadrat ildizlarini olish eng kichik kvadratlar standart xatolarni keltirib chiqaradi va buni osongina tekshirish mumkin, yuqoridagi regressiya jadvalidagi natijalarga mos keladi.

Statistics > Linear Models and related > Regression diagnostics > Specification tests, etc., tanlab, ochiladigan menyular yordamida dialog oynasini chaqirish mumkin, bu esa regressiya uchun taxminiy statistik ma'lumotlar dialog oynasi ochiladi. Quyida ko'rsatilganidek, Kovarians matritsasini baholash (vce) buyrug'idan foydalanishingiz mumkin:



5.2. Eng kichik kvadratlar bashorati

Hisoblangan qiymatlarni topishni II bobda ko'rib chiqqandik va bu yerda qayta ko'rib chiqamiz. Aytaylik, Big Andy 5,50 dollarlik narxda va 1200 dollarlik reklama xarajatlari uchun savdo daromadini hisoblangan qiymatlarni topib olamiz. Ushbu hisoblangan qiymatni quyidagi buyruq orqali topib olamiz:

di _b[_cons] + _b[price]*5.50 + _b[advert]*1.2

Results oynasida quyidagi natijalar aks ettiriladi:

```
. di _b[_cons] + _b[price]*5.50 + _b[advert]*1.2
77.655513
```

Ushbu misolda, berilgan narx va reklama xarajatlari uchun hisoblangan qiymat sotish 77,66 ni tashkil qiladi.

Shartli prognozlarni yaratish uchun ma'lumotlar to'plamiga kuzatishlar va hisoblangan qiymat buyrug'ini qo'shadigan IV bobda o'rganilgan buyruqdan ham foydalanish mumkin. Ma'lumotlar muharriri **Data>Data editor** ni oching va yangi ma'lumotlar qatoriga mustaqil o'zgaruvchilarning kerakli qiymatlarini qo'shing (bu yerdagi kuzatish 76). Buni **Command** oynasidan ham ifodalash mumkin. Avval 76-kuzatishni ifodalash uchun **set obs 76** buyrug'idan foydalaning. So'ngra **replace** buyrug'idan foydalanib, yangi kuzatish uchun mustaqil o'zgaruvchilarning kerakli qiymatlarini kiriting. Bu quyidagi dastlabki uchta qatorda amalga oshiriladi:

set obs 76

replace price = 5.50 in 76

```
replace advert = 1.2 in 76
```

```
predict yhat
```

```
list yhat in 76
```

price va **advert** o'zgaruvchilarga kerakli qiymatlari kiritilgandan so'ng, **Stata**ning **predict** buyrug'i orqali hisoblangan qiymatni yaratish uchun foydalanish mumkin. Bashorat qilish buyrug'i **Big Andy** regressiyasi oxirgi hisoblangan deb taxmin qiladi va u regressiyadan bashorat qilingan qiymatlarni olish uchun standart variant, **xb** dan foydalanadi. Yana bir bor, ushbu **list yhat** ma'lumotlarning 76-kuzatuvini ro'yxatga olish uchun **in 76** shartidan foydalaning:

```
. list yhat in 76
```

	yhat
76.	77.65551

5.3. Tasviriy statistika ma'lumotlarini olish aniqligi

Ushbu va boshqa statistik ma'lumotlarni qo'lda hisoblash uchun zarur bo'lgan ko'plab natijalar **Stata** tomonidan ichki hisob-kitoblar natijalarida saqlanadi. Baholashdan foydalanish natijasida olingan natijalar mazmunini ko'rish: **ereturn list**

Qisman natijalar ifodalanadi:

```
scalars:
```

```
e(N) = 75  
e(df_m) = 2  
e(df_r) = 72  
e(F) = 29.24785947967357  
e(r2) = .4482577622149436  
e(rmse) = 4.886123970679952  
e(mss) = 1396.538929773235  
e(rss) = 1718.942936893432
```

Tenglamaning error (xato) dispersiyasini baholash uchun tenglamadan foydalaning:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^N \hat{e}_i^2}{N - K}$$

Bu yerda $\hat{e}_i^2$ eng kichik kvadratlarning kvadrati qoldiqlaridir. Asosan, numerator faqat erkinlikning qoldiq darajalariga bo'lingan regressiyadan olingan kvadrat xatolar yig'indisidir. Erkinlikning qoldiq darajalari $e(df_r)$ da, kvadratik xatolar yig'indisi esa $e(rss)$ da ifodalanadi. Shunday qilib,

`scalar sighat2 = e(rss)/e(df_r)`

`scalar list sighat2`

Natijasi:

```
. scalar list sighat2
      sighat2 = 23.874207
```

Ushbu ma'lumotlarning barchasini Stata tomonidan ishlab chiqarilgan dispersiya jadvali tahlilidan olish mumkin, u quyida ko'rsatilgan:

`. regress sales price advert`

	SS	df	MS	
Model	1396.53921	2	698.269603	
Residual	1718.94261	72	23.8742057	
Total	3115.48202	74	42.1011083	

Number of obs = 75
 F(2, 72) = 29.25
 Prob > F = 0.0000
 R-squared = 0.4483
 Adj R-squared = 0.4329
 Root MSE = 4.8861

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
price	-7.907856	1.095993	-7.22	0.000	-10.09268 -5.723034
advert	1.862504	.6831955	2.73	0.008	.5006587 3.224509
_cons	118.9136	6.351638	18.72	0.000	106.2519 131.5754

Ko'rib turganingizdek, taxminiy farq 23,874 ni tashkil qiladi. Ildiz MSE bu raqamning kvadrat ildizidir va sariq rang bilan belgilangan jadvalning o'ng tomonidagi chop etishda ko'rinadi.

5.4. Ishonch intervallari

Regressiya jadvali sizga eng kichik kvadratlar hisobini va taxminiy standart xatolarni beradi. Berilgan standart xato eng kichik kvadratlar qiziqish parametrini qanchalik aniq o'lchashga qodirligini baholashdir. Ishonch oralig'i xuddi shunday maqsadga

xizmat qiladi, garchi uni izohlash ancha sodda bo'lsa-da, chunki u sizga yuqori va pastki chegaralarni beradi, ular orasida noma'lum parametr berilgan ehtimollik bilan yotadi. Faraz qilaylik, kimdir Big Andy's Burger Barn'da ma'lum darajadagi reklama uchun narx o'zgarishiga sotishdan tushgan tushumning javobi β_2 uchun 95% intervalli taxminni baholaydi. Bu jarayon III bobdagi tahlil bilan bir xil. Bu safar kritik qiymatlarni olish uchun foydalaniladigan taqsimoti $N-K=75-3=72$ erkinlik darajasiga ega bo'ladi.

95% ishonch oralig'i matndagi (5.13) tenglamaga asoslanadi:

$$P(-t_c < t_{((72))} < t_c) = 0.95$$

Bu tenglamani qanoatlantiradigan $t_{((72))}$ taqsimotdan, uni t_c deb ataydigan kritik qiymat kerak. Doimiy t_c - bu t distributiondan $\alpha/2$ kritik qiymat va α "rad etish" maydoni (ishonch oralig'idan tashqaridagi maydon) bilan bog'liq bo'lgan jami kerakli ehtimollikdir. Kritik qiymat t_c ni statistik jadvalda topish mumkin yoki Statadagi $\text{invttail}(df, \alpha/2)$ funksiyasidan foydalanib, $t_{((72))}$ taqsimotdan 0,025 kritik qiymat 1,993 ekanligini aniqlash mumkin. Biroz algebra natijasida β_2 ning 95% intervalli baholovchisi, uning eng kichik kvadratlar hisoblagichi b_2 va hisoblangan standart xatosi, $se(b_2)$:

$$[b_2 - 1.993 * se(b_2), b_2 + 1.993 * se(b_2)]$$

Quyida siz taxminiy ishonch oralig'i uchun pastki va yuqori so'nggi nuqtalarni yaratish uchun Stata buyruqlaridan foydalanamiz:

```
scalar lb = _b[price] - invttail(e(df_r),.025) * _se[price]
```

```
scalar ub = _b[price] + invttail(e(df_r),.025) * _se[price]
```

```
scalar list lb ub
```

```
. scalar list bL bU
      bL = -10.092676
      bU = -5.7230322
```

Bunday holda, biz Big Andy regressiya tenglamasining bir nechta saqlangan natijalaridan foydalandik. Bularga narx bo'yicha koeffitsiyentning eng kichik kvadratlari bahosi, $_b[\text{price}]$, uning taxminiy standart xatosi, $_se[\text{price}]$ va regressiyadan ozod bo'lishning qoldiq darajalari ($N-K$), $e(df_r)$ kiradi.

Har safar chiziqli regressiya modelini taxmin qilganingizda 95% ishonch oralig'i sukut bo'yicha hisoblanadi. Interval regressiya natijalarining oxirgi ikki ustunida ko'rinadi. Regressiyadan so'ng level() opsiyasidan foydalanib, ehtimolliklarni 90%ga o'zgartirishingiz mumkin. Misol uchun, regressiya bilan 90% intervallarni olish uchun quyidagi buyruqdan foydalaning:

regress sales price advert, level(90)

Intervalni menyu tizimi orqali ham o'zgartirish mumkin. "Statistics > Linear models and related > Linear regression"ni ochish uchun ochiladigan menyudan "Chiziqli regressiya" dialog oynasini tanlang – "Reporting" yorlig'ini tanlang va bu yerda ko'rsatilganidek, ishonch darajasini o'zingiz xohlagan darajaga o'zgartiring:



5.4.1. Parametrlarning chiziqli birikmasi uchun ishonch oralig'i

Big Andy reklama xarajatlarini 800 dollarga oshirishni va narxni 40 sentga tushirishni rejalashtirmoqda. Kutilayotgan savdo hajmining o'zgarishi: $\lambda = -0.4\beta_2 + 0.8\beta_3$

parametrlarning chiziqli birikmasidir. Stata parametrlarning chiziqli kombinatsiyasi uchun turli statistik ma'lumotlarni hisoblaydigan o'rnatilgan lincom buyruqlarini o'z ichiga oladi; lincom har qanday baholashdan so'ng koeffitsiyentlarning chiziqli birikmalari uchun nuqta baholarini, standart xatolarni, t yoki z statistikasini, p qiymatlarini va ishonch oraliqlarini hisoblaydi. Big Andy sotuvining eng kichik kvadratlari λ yordamida taxmin qilish uchun: **lincom -0.4*price+0.8*advert, level(90)**

Bunday holda, ishonch oralig'i uchun qamrov ehtimolini 90% ga o'rnatish uchun ixtiyoriy level(90) buyrug'i kiritilgan. Natija:

```
. lincm -0.4*price+0.8*advert, level(90)
(1) - .4*price + .8*advert = 0
```

sales	Coef.	Std. Err.	t	P> t	[90% Conf. Interval]	
(1)	4.653209	.7096133	6.56	0.000	3.470785	5.835633

Stata buyruqni ekranga aks ettiradi va chiziqli kombinatsiyani takrorlaydi. λ ning taxminiy qiymati 4,65, uning taxminiy standart xatosi 0,7096 va 90% ishonch oralig'i (3,471, 5,835).

Xuddi shu narsani bajarish uchun algebra va arifmetikadan foydalanish mumkin. Stata hisoblash ishlarini osonlashtirsa-da, foydalanuvchi algebrani taqdim etishi kerak. Bu yerda λ ning standart xatosini hisoblash uchun eng kichik baholovchining taxminiy dispersiyalari va kovariatsiyalari qo'llaniladi.

```
matrix cov=e(V)
```

```
scalar lambda = -0.4*_b[price]+0.8*_b[advert]
```

```
scalar var_lambda = (-0.4)^2*cov[1,1]+(0.8)^2*cov[2,2]+2* ///
(-0.4)*(0.8)*cov[1,2]
```

matrix cov=e(V) bayonoti **e(V)** da saqlangan dispersiya kovariatsiyasini cov etiketli matritsaga yozadi. Uchinchi va to'rtinchi qatorlarda Stata matritsasini indekslash imkoniyatlari kovariatsiya matritsasidan λ dispersiyasini hisoblashda foydalanish uchun kerakli elementlarni tanlash uchun ishlatiladi. Masalan, **cov[1,1]** cov ning birinchi qatori va birinchi ustunidagi elementga ishora qiladi, ya'ni β_2 ning dispersiyasi. Esda tutingki, Stata avtomatik ravishda tenglama oxiriga doimiy qo'yadi, shuning uchun β_2 aslida koeffitsiyent vektorining birinchi elementida joylashgan.

Dispersiya hisoblangandan so'ng, uning kvadrat ildizini oling va standart xatoni olish uchun λ bahosini natijaga bo'ling. U yerdan ishonch oralig'i odatdagi usulda hisoblanadi.

```
scalar se = sqrt(var_lambda)
```

```
scalar t = lambda/se
```

```
scalar lb = lambda-invttail(e(df_r),.05)*se
```

```
scalar ub = lambda+invttail(e(df_r),.05)*se
```

Natija olishda **invttail(e(df_r),.05)** **t_((72))** -taqsimotidan 5%

kritik qiymatni olish uchun ishlatiladi. Natijalar lincom natijalariga deyarli mos keladi:

```
. scalar list lambda var_lambda se t lb ub
      lambda = 4.6532091
      var_lambda = .50355097
      se = .70961325
      t = 6.5573876
      lb = 3.4707851
      ub = 5.8356332
```

5.5 Gipoteza testlari

5.5.1. Ikki tomonlama testlar

Ikki tomonlama muqobilga nisbatan $\beta_2=0$ bo'lgan testning t nisbati $\beta_2 \neq 0$

$$t = \frac{b_2 - 0}{se(b_2)} \sim t_{(N-K)}$$

$\beta_2=0$ to'g'ri degan nol gipoteza mavjud bo'lsa. Hisoblangan qiymat:

$$t = \frac{-7.908}{1.096} = -7.215$$

Stata buni buyruq satridan osongina hisoblaydi. **p** qiymati **tail (df,stat)** buyrug'i yordamida olinadi, bu yerda **df** erkinlik darajalari va **stat** - **test** statistikasing qiymati. Bu funksiya **df** erkinlik darajasi bilan **t** distribution uchun **stat** ning o'ng tomonidagi quyruq maydonini o'lchaydi. Ikki tomonlama test uchun **t** ning ijobiy qiymatining quyruq maydonini ikki barobarga oshirish orqali hisoblangan **statistik (stat)** ning chap tomonidagi maydonni o'z ichiga oladi. Demak, **2*ttail(72,abs(t1))** dan foydalaning.

To'liq Stata buyrug'i quyidagicha:

```
scalar t1 = (_b[price]-0)/_se[price]
```

```
scalar p1 = 2*ttail(72,abs(t1))
```

```
scalar list t1 p1
```

$\beta_3=0$ bo'lgan ikki tomonlama test xuddi shunday amalga oshiriladi va to'liq natijalar ushbu bobning oxirida topiladi.

5.5.2. Bir tomonlama testlar

Muqobilga qarshi $\beta_2 \geq 0$ degan nol gipotezani sinab ko'rish uchun! $\beta_2 < 0$, xuddi shu statistikadan foydalaning:

$$t = \frac{b_2 - 0}{se(b_2)} \sim t_{(N-K)}$$

Hisoblash bir xil bo'lsa-da, kritik qiymat endi t distributionning chap dumida:

```
scalar t1 = (_b[price]-0)/_se[price]
```

```
scalar crit = -invttail(e(df_r),.05)
```

```
scalar pval = 1-ttail(e(df_r),t1)
```

ttail(e(df_r),t1) ni tasdiqlashi **df_r** erkinlik darajasiga ega t tasodifiy o'zgaruvchining t1 dan katta bo'lish ehtimolini hisoblaydi. t distribution simmetrik bo'lgani uchun va biz t1 dan kichik bo'lish ehtimolini xohlaymiz, biz 1 dan **ttail(e(df_r),t1)** ni ayiramiz.

Agar p qiymati testning 5% ahamiyatlilik darajasidan kam bo'lsa, u holda statistik ahamiyatga ega bo'lgan ushbu darajadagi nol gipotezani rad eting. Aks holda, ushbu ahamiyat darajasida nolli gipotezani rad etib bo'lmaydi.

Sinovni o'tkazishning boshqa usuli – statistik ma'lumotlarni 5% kritik qiymat bilan solishtirish. Stata kritik qiymatni **invttail(e(df_r),.05)** funksiyasi yordamida ishlab chiqaradi. Ushbu hisob-kitoblarning natijalari:

```
. scalar list t1 crit pval  
      t1 = -7.2152415  
      crit = -1.6662937  
      pval = 2.212e-10
```

Reklama samaradorligi testi $\beta_3 \leq 1$ muqobilga qarshi $\beta_3 > 1$ degan nol gipotezani hisobga oladi. Buyruqlar quyidagicha:

```
scalar t2 = (_b[advert]-1)/_se[advert]
```

```
scalar crit = invttail(e(df_r),.05)
```

```
scalar pval = ttail(e(df_r),t2)
```

Muqobil variant t-distributionning o'ng tomonida joylashganligi sababli buyruq biroz shaffofroq. Yana **invttail** t distribution uchun 5%

o'ng dum kritik qiymatini va ttailt2 ning o'ng tomonidagi maydonni hisoblaydi. Ushbu buyruqlar quyidagi natijalarni ifodalaydi:

```
. scalar list t2 crit pval
      t2 = 1.2625732
      crit = 1.6662937
      pval = .10540831
```

5.5.3. Chiziqli parametrlarni tekshirish

Big Andy kompaniyasining marketing bo'yicha maslahatchisi-ning ta'kidlashicha, narxni 20 sentga tushirish reklama xarajatlarini 500 dollarga oshirishdan ko'ra savdo daromadini oshirish uchun samaraliroq bo'ladi. Model nuqtayi nazaridan $-0.2\beta_2 > 0.5\beta_3$ deb hisoblaydi. Nol gipoteza $-0.2\beta_2 - 0.5\beta_3 \leq 0$ va muqobil $-0.2\beta_2 - 0.5\beta_3 > 0$ ga teng.

Yana bir bor, lincom buyrug'i foydalidir, chunki u chiziqli kombinatsiyani va uning standart xatosini baholaydi. Bir tomonlama test sifatida, lincomning topilgan natijasiga tayangandan ko'ra, to'g'ridan to'g'ri tanqidiy qiymatni yoki p qiymatini hisoblash yaxshidir, bu sizga ikki tomonlama t testi uchun faqat p qiymatini beradi. Birinchidan, lincomdan foydalaning:

lincom -0.2*price-0.5*advert

natija esa:

```
( 1)  - .2*price - .5*advert = 0
```

sales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
(1)	.6502787	.4009846	1.62	0.109	-.1490694 1.449627

Hisoblangan koeffitsiyent 0,650 ni standart xatolik 0,400 ga bo'lish natijasida t nisbati 1,62 ga teng bo'ladi. Bir tomonlama p qiymati ikki tomonlama qiymatning yarmini tashkil etadi, agar taxminiy chiziqli kombinatsiya kutilgan belgiga ega bo'lsa. Chalkashmaslik uchun biz quyida ko'rsatilganidek, bir tomonlama testlar uchun mos kritik qiymat yoki p qiymatini hisoblashni tavsiya qilamiz:

```

scalar t = r(estimate)/r(se)
scalar crit = invttail(r(df),.05)
scalar pval = ttail(r(df),t)
scalar list crit t pval

```

```

. scalar list crit t pval
    crit = 1.6662937
      t = 1.6217052
    pval = .05461891

```

Yangi element **r(estimate)**, **r(df)** va **r(se)** ni o'z ichiga oladi. Bu lincom tomonidan hisoblangan taxminiy chiziqli parametr va standart xato. Hisoblashdan keyin mavjud bo'lgan **e()** buyruqlari singari, boshqa ko'plab buyruqlar ham (masalan, **test** va **lincom**) keyingi hisoblash uchun natijalarni saqlaydi. Buyruqlar:

return list

natija:

```

scalars:
      r(df) = 72
      r(se) = .4009845529554849
r(estimate) = .65027873012098

```

Bu shuni ko'rsatadiki, erkinlik darajalari (N-K), baholash va uning standart xatosi lincom buyrug'idan oldingi natijalar qismida saqlangan. Natijalarni takror ko'rish uchun, saqlangan natijalar to'plamida nima borligini tekshirish uchun **ereturn list** yoki **return list** buyruqlaridan foydalanishingiz mumkin.

5.6. Polinomial tenglamalar

Mustaqil va bog'liq o'zgaruvchilar o'rtasidagi chiziqli bo'lmagan munosabatlarga ruxsat berishning usullaridan biri modelga regressorlarning polinomlarini kiritishdir. Ushbu misolda qo'shimcha 1000 dollarlik reklamanning marjinal ta'siri ko'proq reklama ishlatilganligi sababli kamayishi kutilmoqda. Model quyidagicha hosil bo'ladi:

$$\text{SALES} = \beta_1 + \beta_2 \text{ PRICE} + \beta_3 \text{ ADVERT} + \beta_4 \text{ ADVERT}^2 + e$$

Buni baholashning ikki yo'li mavjud. Birinchidan, yangi

ADVERT^2 o'zgaruvchisini yaratish, uni modelga qo'shish va parametrlarni baholash mumkin. Bu quyidagicha:

generate a2 = advert*advert

reg sales price advert a2

Buyruq natijasida quyidagi natijalar ifodalanadi:

. reg sales price advert a2

Source	SS	df	MS			
Model	1583.39744	3	527.799145	Number of obs	= 75	
Residual	1532.08443	71	21.5786539	F(3, 71)	= 24.46	
Total	3115.48187	74	42.1011063	Prob > F	= 0.0000	
				R-squared	= 0.5082	
				Adj R-squared	= 0.4875	
				Root MSE	= 4.6453	

sales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
price	-7.64	1.045939	-7.30	0.000	-9.725543	-5.554457
advert	12.15124	3.556164	3.42	0.001	5.060447	19.24203
a2	-2.767964	.9406241	-2.94	0.004	-4.643515	-.8924123
_cons	109.719	6.799045	16.14	0.000	96.16212	123.2759

advert*advert ko'paytirish orqali yaratilgan a2 o'zgaruvchisi ba'zan o'zaro ta'sir o'zgaruvchisi deb ataladigan oddiy misoldir. O'zaro ta'sir o'zgaruvchisi haqida fikr yuritishning eng oddiy usuli shundaki, uning bog'liq o'zgaruvchiga ta'siri boshqa o'zgaruvchiga bog'liq – ikkita o'zgaruvchiga bog'liq o'zgaruvchining o'rtacha qiymatini aniqlash uchun o'zaro ta'sir qiladi. Bunday holda, reklamaning o'rtacha savdoga ta'siri reklamaning o'z darajasiga bog'liq.

O'zgaruvchilar o'zaro ta'sir qilganda, boshqa reklama birligining marjinal ta'siri (\$1000) hisob-kitob asosida qo'lda hisoblanishi kerak. Reklama daromadiga nisbatan o'rtacha savdoning qisman hosilasini olish mumkin:

$$\frac{\partial E(\text{SALES})}{\partial \text{ADVERT}} = \beta_3 + 2\beta_4 \text{ADVERT}$$

Marjinal ta'sirning kattaligi parametrlarga, shuningdek, reklama darajasiga bog'liq. Misolda u ikki nuqtada baholanadi, ADVERT=0.5 va ADVERT=2. Buyruq quyidagicha:

scalar me1 = _b[advert]+2*(.5)*_b[a2]

scalar me2 = _b[advert]+2*(2)*_b[a2]

va natijalar:

```
. scalar list me1 me2
      me1 = 9.3832736
      me2 = 1.079383
```

5.6.1. Optimal reklama: parametrlarning chiziqli bo'lmagan kombinatsiyasi

Ushbu misolda reklamaning optimal darajasi aniq sof sotishni maksimal darajada oshiradi. Endi yana bir dollar sarf-xarajatlar bir dollar savdoni qo'shadigan nuqtaga qadar reklama qiladi, va bundan ortiq emas. Bu nuqtada marjinal effekt birga teng:

$$\beta_3 + 2\beta_4 \text{ ADVERT} = 1$$

ADVERTni parametrlar bo'yicha yechish:

$$\text{ADVERT}_0 = \frac{1 - \beta_3}{2\beta_4}$$

Model parametrlarida chiziqli bo'lmagan. Reklamaning optimal darajasining izchil bahosini eng kichik kvadratlar baholarini parametrlar uchun almashtirish orqali olish mumkin. Delta usuli orqali standart xatoni baholash ba'zi hisob-kitoblarni talab qiladi, ammo Stata nlcom buyrug'ini o'z ichiga oladi, u baholashdan keyin parametrlarning chiziqli bo'lmagan birikmalarini, shuningdek standart xato, t nisbat va ishonch oralig'ini hisoblaydi:

$$\text{nlcom } (1 - \text{b[advert]}) / (2 * \text{b[a2]})$$

```
. nlcom (1-_b[advert])/(2*_b[a2])
      _nl_1:  (1-_b[advert])/(2*_b[a2])
```

sales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
_nl_1	2.01434	.128723	15.65	0.000	1.757673	2.271006

Ushbu hisob-kitoblarga ko'ra, reklamaning optimal darajasi 2014 dollarni tashkil qiladi.

5.6.2. O'zaro ta'sirlar uchun omil o'zgaruvchilardan foydalanish

O'zaro ta'sirga ega modellarni baholash va marjinal ta'sirlarni baholashning eng yaxshi usuli – bu Stata omil o'zgaruvchilari va o'rnatilgan **margins** buyrug'idan foydalanishdir. Statada o'zgaruvchilar bilan o'zaro ta'sir qilish juda oson, ammo siz nimani nazarda tutayotganingizni taxmin qilish uchun biroz ehtiyot bo'lishingiz kerak. Stata omil o'zgaruvchilari toifali o'zgaruvchilardan indikator o'zgaruvchilarni, toifali o'zgaruvchilar ko'rsatkichlarining o'zaro ta'sirini, toifali va uzluksiz o'zgaruvchilarning o'zaro ta'sirini va uzluksiz o'zgaruvchilarning (polinomiya) o'zaro ta'sirini yaratadi. Ularga bir nechta boshqa buyruqlar bilan bir qatorda ko'p baholash va baholash buyruqlari bilan ruxsat beriladi.

Statada faktor o'zgaruvchilar o'z operatorlariga ega:

Operator	Description
i.	unary operator to specify indicators
c.	unary operator to treat as continuous
#	binary operator to specify interactions
##	binary operator to specify factorial interactions

Unary operator (birlik operator) – bu faqat bitta raqam yoki o'zgaruvchiga ta'sir qiladigan matematik operatsiya uchun math-speak – operand deb ataladi. Misol sifatida raqamdan oldin paydo bo'lgan manfiy belgi bo'lishi mumkin; shunday foydalanilganda u raqamni oladi va uni manfiy qiladi. Demak, -2 ifodasida manfiy ishora 2 raqamida ishlaydigan unar operator bo'lib, uni manfiy qiladi. Ikkilik operatorlar ikkita o'zgaruvchida ishlaydi. Masalan, 1-2 ifodadagi - 1 dan 2 ni ayirib, 1 va 2 da ishlaydi.

Stata bir nechta unar operatorlarni o'z ichiga oladi, ular orasida i. va c. O'zgaruvchi nomidan oldin foydalanilganda, bu operatorlar Stataga o'zgaruvchiga uzluksiz (masalan, c.price) yoki diskret (i.price) sifatida qarash kerakligini ko'rsatadi. Ikkinchi holda, i.price PRICE ning turli qiymati uchun indikator o'zgaruvchisini

yaratadi. O'zgaruvchi uzluksiz va ma'lumotlar to'plamidagi har bir qiymat noyob bo'lganligi sababli, Stata modelda foydalanish uchun N ko'rsatkich o'zgaruvchilari to'plamini yaratadi, ehtimol siz odatdagidek qilmoqchi bo'lgan narsa emas. Boshqa tomondan, ma'lumotlaringizdagi diskret o'zgaruvchilar uchun indikator o'zgaruvchilari to'plamlarini yaratishda foydali bo'ladi. Bunga ba'zi misollar ushbu qo'llanmaning boshqa joylarida keltirilgan.

Hozirgi holatda ADVERT uzluksiz (c.advert ga ekvivalent) va biz uni o'zi bilan o'zaro aloqada bo'lishni xohlaymiz (c.advert). Ikkilik operator # Statada ikkita omilli o'zgaruvchilarni ko'paytirish uchun ishlatiladi. Shu tarzda u ikki o'zgaruvchining o'zaro ta'sirini yaratadi. Shunday qilib, ADVERB^2 uzluksiz o'zgaruvchisi c.advert#c.advert yordamida yaratiladi.

O'zgaruvchilarni shu tarzda belgilashning kamida ikkita sababi bor. Birinchidan, ma'lumotlar to'plamiga kiritish uchun o'zaro ta'sir o'zgaruvchisini maxsus yaratishga hojat yo'q; c.advert#c.advert to'g'ridan to'g'ri modelga uni regress buyrug'ining o'zgaruvchilar ro'yxatiga kiritish orqali qo'shilishi mumkin. Ikkinchidan, shu tarzda yaratilgan o'zgaruvchilar marjinal effektlarni hisoblash uchun Stata-ning o'rnatilgan margins buyrug'idan foydalanishi mumkin. Oddiy ko'paytirish * o'rniga omil o'zgaruvchilarni ko'paytirish uchun # dan foydalanish Stataga uni ko'paytma qilgan ikkita o'zgaruvchining o'zaro ta'sir qilganligini bildiradi. Bu sizga hech qanday hisob-kitob qilmasdan marjinal effektlarni kuzatib borish imkonini beradi. Asosiysi, ushbu omil o'zgaruvchilari yordamida o'zaro ta'sirlarni yaratish va chekkalarda to'g'ri variantlardan foydalanish mumkin.

Regressiyani baholash uchun buyruq

regress sales price advert c.advert#c.advert va natijalari quyidagilardan iborat:

regress sales price advert c.advert#c.advert						Number of obs = 75	
Source	SS	df	MS			F(3, 71) =	24.46
Model	1583.39741	3	527.799136			Prob > F =	0.0000
Residual	1532.08446	71	21.5786543			R-squared =	0.5082
Total	3115.48187	74	42.1011063			Adj R-squared =	0.4875
						Root MSE =	4.6453

sales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
price	-7.64	1.045939	-7.30	0.000	-9.725543	-5.554457
advert	12.15124	3.556164	3.42	0.001	5.060446	19.24203
c.advert# c.advert	-2.767963	.940624	-2.94	0.004	-4.643514	-.892412
_cons	109.719	6.799045	16.14	0.000	96.16212	123.2759

E'tibor bering, bu regressiya natijalari a2 qo'lda yaratilgan va modelga kiritilgan natijalar bilan bir xil. O'zaro ta'sir muddati bo'yicha koeffitsient endi c.advert#c.advert sifatida aniqlanadi. To'lov marjinal effektni hisoblash uchun marjlardan foydalanish bilan birga keladi. 0,5 va 2 reklama darajalarida marjinal ta'sirni baholash olinadi:

```
. margins, dydx(advert) at(advert=(.5 2))
```

Average marginal effects
Model VCE : OLS
Expression : Linear prediction, predict()
dy/dx w.r.t. : advert

1._at	: advert	=	.5
2._at	: advert	=	2

	dy/dx	Delta-method Std. Err.	z	P> z	[95% Conf. Interval]	
advert						
_at						
1	9.383273	2.636965	3.56	0.000	4.214916	14.55163
2	1.079383	.7019353	1.54	0.124	-.2963846	2.455151

Chegaralarning sintaksisi ba'zi tushuntirishlarni talab qiladi. Birinchi variant - dydx (advert). Bu qavs ichidagi o'zgaruvchining (advert) bir birlik o'sishining qaram o'zgaruvchining o'rtacha qiymatiga marjinal ta'sirini so'raydi. Ikkinchi variant Stata-ga advert ning qaysi qiymatlarida marjinal effektni baholashini aytadi. Har holda, marjlar birinchi usul bilan bir xil natijani berdi va buni deyarli hech qanday dasturlashsiz va matematikasiz qilinadi. Bonus sifatida margins standart xatolik, t-nisbati, ikki tomonlama p-qiymati va ishonch oralig'i taxminini ham ifodalaniladi.

5.7. O'zgaruvchilar orasidagi o'zaro ta'sirlar

O'zgaruvchilar orasidagi o'zaro ta'sir ko'p nomli atamalar yaratish uchun oldingi bo'limda tanishtirilgandi. Tushuncha juda umumiy bo'lib, bir o'zgaruvchining o'zgarishining bog'liq o'zgaruvchining o'rtacha qiymatiga ta'siri boshqa o'zgaruvchiga bog'liq bo'lgan har qanday vaziyatga nisbatan qo'llanilishi mumkin. Ko'rib chiqilgan asosiy model quyidagicha:

$$PIZZA = \beta_1 + \beta_2 AGE + \beta_3 INCOME + e$$

Inson yoshi ulg'aygan sari uning pitsaga sarflashga bo'lgan cheklov moyilligi pasayadi – bu koeffitsiyentning inson yoshiga bog'liqligini anglatadi:

$$\beta_3 = \beta_4 + \beta_5 AGE$$

Buni modelga almashtirish orqali ifodalaymiz:

$$PIZZA = \beta_1 + \beta_2 AGE + \beta_4 INCOME + \beta_5 (INCOME * AGE) + e$$

Bu o'zaro ta'sir o'zgaruvchisi bo'lgan $INCOME * AGE$ yangi o'zgaruvchini taqdim etadi. Ushbu modeldagi AGE ning birlik o'sishining marjinal ta'siri $INCOME$ ga bog'liq va $INCOME$ ning o'sishining marjinal ta'siri AGE ga bog'liq.

O'zaro ta'sir Statada yaratish generate buyrug'i yordamida yaratilishi mumkin, ammo omil o'zgaruvchilardan foydalanish yaxshiroq tanlovdir. Bu o'rnatilgan margins buyrug'i yordamida marjinal effektlarni baholashga imkon beradi:

```
use pizza4, clear
```

```
regress pizza age income c.age#c.income
```

```
margins, dxdy(age) at(income=(25 90))
```

Regressiyadan olingan natija:

```
. regress pizza age income c.age#c.income
```

Source	SS	df	MS			
Model	367043.25	3	122347.75			
Residual	580608.65	36	16128.0181			
Total	947651.9	39	24298.7667			

Number of obs =	40
F(3, 36) =	7.59
Prob > F =	0.0005
R-squared =	0.3873
Adj R-squared =	0.3363
Root MSE =	127

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
pizza					
age	-2.977423	3.352101	-0.89	0.380	-9.775799 3.820952
income	6.979905	2.822768	2.47	0.018	1.255067 12.70474
c.age#					
c.income	-.1232394	.0667187	-1.85	0.073	-.2585512 .0120725
_cons	161.4654	120.6634	1.34	0.189	-83.25131 406.1822

va margins buyrug'idan esa:

```
. margins, dydx(age) at(income=(25 90))
```

```
Average marginal effects
Model VCE : OLS
Expression : Linear prediction, predict()
dy/dx w.r.t. : age
1._at : income = 25
2._at : income = 90
```

		Delta-method				
		dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]
age						
	_at					
	1	-6.058407	2.390502	-2.53	0.011	-10.74371 -1.373109
	2	-14.06896	4.171058	-3.37	0.001	-22.24409 -5.893842

Yana bir misol, cps_small.dta ma'lumotlar to'plamidan foydalanadigan IV bobda ko'rib chiqilgan ish haqi modeliga ta'lim va tajriba o'rtasidagi o'zaro ta'sirni ko'rib chiqamiz:

$\ln(\text{WAGE}) = \beta_1 + \beta_2 \text{ EDUC} + \beta_3 \text{ EXPER} + \beta_4 (\text{EDUC} * \text{EXPER}) + e$

use cps4_small, clear

gen lwage = ln(wage)

regress lwage educ exper c.educ#c.exper

```
. regress lwage educ exper c.educ#c.exper
```

Source	SS	df	MS			
Model	65.5449479	3	21.848316			
Residual	271.235783	996	.272325084			
Total	336.780731	999	.337117849			

Number of obs =	1000
F(3, 996) =	80.23
Prob > F =	0.0000
R-squared =	0.1946
Adj R-squared =	0.1922
Root MSE =	.52185

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lwage					
educ	.0949385	.0146246	6.49	0.000	.06624 .123637
exper	.0063295	.0066985	0.94	0.345	-.0068153 .0194743
c.educ#					
c.exper	-.0000364	.0004838	-0.08	0.940	-.0009858 .0009129
_cons	1.392318	.2066447	6.74	0.000	.986809 1.797827

va - regress lwage educ exper c.educ#c.exper c.exper#c.exper

. regress lwage educ exper c.educ#c.exper c.exper#c.exper					
Source	SS	df	MS		
Model	82.3591698	4	20.5897924	Number of obs =	1009
Residual	254.421562	995	.255709062	F(4, 995) =	80.52
Total	336.780732	999	.337117849	Prob > F =	0.0000
				R-squared =	0.2445
				Adj R-squared =	0.2415
				Root MSE =	.50567

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
educ	.1271953	.0147188	8.64	0.000	.0983118 .1560789
exper	.0629807	.0095361	6.60	0.000	.0441676 .0816938
c.educ# c.exper	-.0013224	.0004949	-2.67	0.008	-.0022935 -.0003513
c.exper# c.exper	-.0007139	.000088	-8.11	0.000	-.0008867 -.0005412
_cons	.5296774	.2267415	2.34	0.020	.084731 .9746237

Yana bir bor ta'kidlaymizki, faktor o'zgaruvchilari uchun Stata unary operatorlaridan foydalanish bu regressiyani baholashni oson qiladi.

5.8. Modellar orasidagi moslikni baholash

Biz IV bobda ko'rganimizdek, regressiyaning mosligi dispersiya jadvali tahlilidan olingan statistik ma'lumotlarga asoslanadi. R^2 regressiya hisobiga bog'liq bo'lgan bog'liq o'zgaruvchidagi namunaviy o'zgarishlar ulushini o'lchaydi. Xuddi shu narsa ko'p regressiya modeli uchun ham amal qiladi. Yagona farq shundaki, endi modelda bir nechta tushuntiruvchi o'zgaruvchilar mavjud. Shunday qilib, IV bobdagi munosabatlardan foydalanish mumkin. Erksiz o'zgaruvchi (SST) kvadratlarining umumiy yig'indisini (SSR) ichida ifodalaniladi va tushuntirilmagan (SSE) o'zgarishlarga ajrating: $SST=SSR+SSE$

Keyin moslik yaxshiligi o'lchovi R^2 bo'ladi:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

Principle of Econometrics adabiyotida ishlab chiqilgan va bu yerda qo'llaniladigan yozuvlar bilan Stataning dispersiya jadvalining tahlilidan foydalanganda ehtiyot bo'lishingiz kerak. Stata kvadratlarining namunaviy yig'indisi sifatida tushuntirilgan o'zgarishlarga ishora qiladi. Matn buni kvadratlar yig'indisi

regressiya yoki SSR deb ataydi. Bog'liq o'zgaruvchining tushunarsiz o'zgarishi eng kichik kvadrat qoldiqlarida olinadi. Sizning darsligingizda bu kvadrat xatolar yig'indisi SSE deb ataladi. Stata bu tushunarsiz o'zgarishlarni kvadratlarning qoldiq yig'indisi deb ataydi.

Concept	POE	Stata
Unexplained variation	SSE (sum of squared errors)	Residual SS
Explained variation	SSR (sum of squared regression)	Model SS
Total variation	SST	Total SS

Bu yo'ldan chiqib, siz R^2 ni bir qancha usullar bilan olishingiz mumkin. Birinchisi, dispersiya jadvalining tahlilidan foydalanish va uni qo'lda hisoblash. Ikkinchisi, uni regressiya natijasidan to'g'ridan to'g'ri olish mumkin, chunki u chiziqli regressiya uchun hisoblangan natijalarda o'z ifodasini topadi. Uchinchisi, oldingi regressiyadan saqlangan natijalar yordamida hisoblash uchun buyruq satridan foydalanish. Albatta, bu variant birinchisidan osonroq, lekin ikkinchisidan qiyinroq! Shunday qilib, Stata `e(mss)` da saqlaydigan kvadratlarning model yig'indisi va `e(rss)` da saqlangan kvadratlarning qoldiq yig'indisidan foydalanib, biz R^2 ni hisoblaymiz va ko'rsatamiz. E'tibor bering, buyruq bo'yicha hisoblangan R^2 ham Stata xotirasida saqlanadi va uni `e(r2)` yordamida qayta ifodalashimiz mumkin:

use andy, clear

reg sales price advert

`. reg sales price advert`

Source	SS	df	MS			
Model	1396.53893	2	698.269465		Number of obs =	75
Residual	1718.94294	72	23.8742075		F(2, 72) =	29.25
Total	3115.48187	74	42.1011063		Prob > F =	0.0000
					R-squared =	0.4483
					Adj R-squared =	0.4329
					Root MSE =	4.8861

sales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
price	-7.907854	1.095993	-7.22	0.000	-10.09268 -5.723032
advert	1.862584	.6831955	2.73	0.008	.500659 3.22451
_cons	118.9136	6.351638	18.72	0.000	106.2519 131.5754

di "R-square " $e(mss)/(e(mss)+e(rss))$

di "R-square " $1-e(rss)/(e(mss)+e(rss))$

```
. di "R-square " e(mss)/(e(mss)+e(rss))
R-square .44825776

. di "R-square " 1-e(rss)/(e(mss)+e(rss))
R-square .44825776
```

R^2 ning moslik mezoni sifatidagi kamchiliklari shundaki, modelga regressorlarni qo'shish har doim moslikni yaxshilaydi degan tushunchadan yiroqdir. Tegishli moslik o'lchovi regressorlarni qo'shish uchun kichik o'zgarishni qo'llaydi, modelga aloqador bo'lmagan mustaqil o'zgaruvchilar qo'shilganligi sababli qayta hisoblangan R^2 kichikroq bo'lishi mumkin. Hisoblangan R^2

$$\bar{R}^2 = 1 - \frac{SSE/(N - K)}{SST/(N - 1)}$$

Ushbu statistika **Stata regress** buyrug'i tomonidan buyruq bo'yicha ham xabar qilinadi. E'tibor bering, sozlangan **R squared R** kvadratidan kichikroq. Aslida, **R** kvadrat hech qachon ko'p chiziqli regressiyada qayta hisoblangan **R squared** dan kichik bo'lmaydi, chunki taxminiy moslik tushuntirish o'zgaruvchilarni qo'shish uchun ifodaladi.

Beshinchi bob mavzularini mustahkamlash uchun savollar:

1. Ko'p chiziqli regressiya va oddiy chiziqli regressiyaning farqli jihatlarini aniqlang.
2. Tenglamaning error (xato) dispersiyasini baholash uchun qanday formuladan foydalaniladi?
3. Taxminiy ishonch oralig'i uchun pastki va yuqori so'nggi nuqtalarni yaratish uchun qanday buyruq amalga oshiriladi?
4. Ikki tomonlama muqobilga nisbatan $\beta_2=0$ bo'lgan testning t nisbati qanday bo'ladi?

5. Muqobilga qarshi $\beta_2 \geq 0$ degan nol gipotezani sinab ko'rish uchun qanday formuladan foydalanish kerak?

6. Agar p qiymati testning 5% ahamiyatlilik darajasidan kam bo'lsa qanday amal bajarish talab etiladi?

7. Chiziqli parametrlarni tekshirishda lincom buyrug'i qanday amal bajaradi?

8. Parametrlarning chiziqli bo'lmagan kombinatsiyasi haqida nimalarni bilib oldingiz?

9. O'zaro ta'sirlar uchun omil o'zgaruvchilardan foydalanish qanday buyruqlar orqali bajariladi?

10. O'zgaruvchilarni c.advert#c.advert tarzida belgilashning sabablarini tushuntiring.

11. O'zaro ta'sir Statada yaratish qaysi buyruq orqali amalga oshiriladi?

12. Modellar orasidagi moslikni baholashni tusuntirib bering.

Beshinchi bob mavzularida qo'llanilgan buyruqlar va atamalar

#

adjusted R2	index	regress
binary operators	interaction variable	replace
c.variable	invtail(df,alpha)	return list
confidence interval	level(90)	Root MSE
di	lincom	scalar
dydx	list	scalar list
e(df_r)	marginal effect	set obs
e(mss)	margins	test
e(rss)	matrix	t-ratio
e(V)	multiple	linear
ereturn list	regression	ttail(df,tstat)
estat vce	nlcom	unary operators
factor variables	prediction	variance
i.variable	p-value	R2
		variance covariance matrix

Beshinchi bob mavzularini mustahkamlash uchun masalalarning do-file lari

Ma'lumotlar buyrug'i: `cd c:\data\poe4stata`

1-masala. log file ni ochish va datalardan foydalanish.

`log using chap05_food, replace text`

`use andy, clear`

* Summary Statistics

`summarize`

* List subset of observations

`list in 1/5`

* Least squares regression with covariance matrix

`regress sales price advert`

`estat vce`

* Predict sales when price is 5.50 and adv is 1200

`di _b[_cons] + _b[price]*5.50 + _b[advert]*1.2`

* Using the data editor to predict

`set obs 76`

`replace price = 5.50 in 76`

`replace advert = 1.2 in 76`

`predict yhat`

`list yhat in 76`

* Calculate sigma-hat square

`ereturn list`

`scalar sighat2 = e(rss)/e(df_r)`

`scalar list sighat2`

* Standard error of the regression

`di sqrt(sighat2)`

* Confidence Intervals

`scalar bL = _b[price] - invttail(e(df_r),.025) * _se[price]`

`scalar bU = _b[price] + invttail(e(df_r),.025) * _se[price]`

`scalar list bL bU`

* Using the level() command to change size of default intervals

`regress sales price advert, level(90)`

* Interval for a linear combination

```

* Easy way
lincom -0.4*price+0.8*advert, level(90)

* Hard way
matrix cov=e(V)
scalar lambda = -0.4*_b[price]+0.8*_b[advert]
scalar var_lambda = (-0.4)^2*cov[1,1]+(0.8)^2*cov[2,2]+2*(-
0.4)*(0.8)*cov[1,2]
scalar se = sqrt(var_lambda)
scalar t = lambda/se
scalar lb = lambda-invttail(e(df_r),.05)*se
scalar ub = lambda+invttail(e(df_r),.05)*se
scalar list lambda var_lambda se t lb ub

* t-ratios
scalar t1 = (_b[price]-0)/_se[price]
scalar t2 = (_b[advert]-0)/_se[advert]
scalar list t1 t2

* pvalues
scalar p1 = 2*ttail(72,abs(t1))
scalar p2 = ttail(72,abs(t2))
scalar list p1 p2

* One sided significance test
scalar t1 = (_b[price]-0)/_se[price]
scalar crit = -invttail(e(df_r),.05)
scalar pval = 1-ttail(e(df_r),t1)
scalar list t1 crit pval

* One sided test of Advertising effectiveness
scalar t2 = (_b[advert]-1)/_se[advert]
scalar crit = invttail(e(df_r),.05)
scalar pval = ttail(e(df_r),t2)
scalar list t2 crit pval

* Linear combination
lincom -0.2*price-0.5*advert
scalar t = r(estimate)/r(se)
scalar crit = invttail(e(df_r),.05)
scalar pval = ttail(e(df_r),t)

```

```

scalar list crit t pval
return list
* Polynomial
generate a2 = advert*advert
reg sales price advert a2
scalar me1 = _b[advert]+2*(.5)*_b[a2]
scalar me2 = _b[advert]+2*(2)*_b[a2]
scalar list me1 me2
* Nonlinear combinations of variables
scalar advertt0 = (1-_b[advert])/(2*_b[a2])
scalar list advertt0
nlcom (1-_b[advert])/(2*_b[a2])
* Polynomial using factor variables
regress sales price advert c.advert#c.advert
margins, dydx(advert) at(advert=(.5 2))
* Interactions
use pizza4, clear
regress pizza age income c.age#c.income
margins, dydx(age) at(income=(25 90))
use cps4_small, clear
gen lwage = ln(wage)
regress lwage educ exper c.educ#c.exper
regress lwage educ exper c.educ#c.exper c.exper#c.exper
use andy, clear
reg sales price advert
di "R-square " e(mss)/(e(mss)+e(rss))
di "R-square " 1-e(rss)/(e(mss)+e(rss))
log close

```

VI BOB. KO'P OMILLI REGRESSIYA MODELLARINI BAHOLASH

6.1. F testni hisoblash

Ushbu bobda V bobda ko'rib chiqilgan savdo modeli yoki Big Andy's Burger Barn modelidan foydalanamiz. Model uchta mustaqil o'zgaruvchini va natijaviy omilni o'z ichiga oladi:

$$\text{SALES}_i = \beta_1 + \beta_2 \text{ PRICE}_i + \beta_3 \text{ ADVERT}_i + \beta_4 \text{ ADVERT}_i^2 + e_i$$

Bu yerda SALES_i ma'lum bir shahardagi oylik sotuvlar va u 1000 AQSH dollari miqdorida o'lchanadi, PRICE_i – gamburgerning dollarda o'lchanadigan narxi, ADVERT_i – ming dollarda o'lchanadigan reklama xarajatlari va bularda $i=1,2,\dots,N$ ni tashki etadi.

Nol gipoteza shundan iboratki, reklama o'rtacha savdoga ta'sir qilmaydi. Reklamaning barcha qiymatlari uchun marjinal effekt nolga teng bo'lishi uchun $\beta_3=0$ va $\beta_4=0$ bo'lishi kerak. Muqobil variant $\beta_3 \neq 0$ yoki $\beta_4 \neq 0$. Modelning nol gipotezasi ostidagi parametrlari nolga teng, muqobil ostidagi parametrlar esa cheklanmagan.

F testi cheklanmagan modeldan cheklangan modelga kvadrat xatolar yig'indisi solishtiradi. Katta farq cheklovlarning noto'g'ri ekanligiga dalil sifatida qabul qilinadi. Nol gipotezani (cheklovlarni) tekshirish uchun foydalaniladigan statistika.

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N - K)}$$

Cheklovlar to'g'ri bo'lganda, J numerator va $N-K$ maxraj erkinlik darajalari bilan F taqsimotiga ega.

Statistik ma'lumotlar ikkita regressiyani bajarish orqali hisoblanadi. Birinchisi cheklanmagan; ikkinchisida cheklovlar mavjud. Har bir regressiyadan kvadratik xatolar yig'indisini, cheklanmagan regressiyadan erkinlik darajalarini $(N-K)$ va o'rnatilgan mustaqil cheklovlar sonini (J) ifodalang. Keyin, quyidagilarni hisoblang:

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N - K)} = \frac{(1896,91 - 1532,084)/2}{1532,084/(75 - 4)} = 8,44$$

Ushbu modelni taxmin qilish uchun **andy.dta** ma'lumotlar faylidan foydalaning: **use andy, clear**

Stata o'zgaruvchilar oynasida siz ma'lumotlar uchta o'zgaruvchidan iborat ekanligini ko'rasiz: **sales**, **price**, va **advert**. Ular cheklanmagan modelni regress baholash funksiyasi bilan ishlatiladi.

regress sales price advert c.advert#c.advert

regress sales price advert c.advert#c.advert

. regress sales price advert c.advert#c.advert

Source	SS	df	MS			
Model	1583.39741	3	527.799136			
Residual	1532.08446	71	21.5786543			
Total	3115.48187	74	42.1011063			

Number of obs =	75
F(3, 71) =	24.46
Prob > F =	0.0000
R-squared =	0.5082
Adj R-squared =	0.4875
Root MSE =	4.6453

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
sales						
price	-7.64	1.045939	-7.30	0.000	-9.725543	-5.554457
advert	12.15124	3.556164	3.42	0.001	5.060446	19.24203
c.advert# c.advert	-2.767963	.940624	-2.94	0.004	-4.643514	-.892412
_cons	109.719	6.799045	16.14	0.000	96.16212	123.2759

Kvadrat xatolar yig'indisini **e(ssr)** yordamida **sseu** deb nomlangan yangi skalyarga va dispersiya jadvali tahlilidan qolgan erkinlik darajalarini **e(df_r)** yordamida **df_unrest** deb nomlangan o'zgaruvchiga ifodalang.

scalar sseu = e(ssr) scalar df_unrest=e(df_r)

Keyinchalik, modelga cheklov qo'ying va uni eng kichik kvadratlar yordamida qayta baholang. Shunga qaramay, kvadrat xatolar yig'indisini va qoldiq erkinlik darajalarini saqlang.

regress sales price

. regress sales price

Source	SS	df	MS
Model	1219.09103	1	1219.09103
Residual	1896.39084	73	25.9779567
Total	3115.48187	74	42.1011063

Number of obs = 75
 F(1, 73) = 46.93
 Prob > F = 0.0000
 R-squared = 0.3913
 Adj R-squared = 0.3830
 Root MSE = 5.0969

sales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
price	-7.829074	1.142865	-6.85	0.000	-10.1068 -5.551348
_cons	121.9002	6.526291	18.68	0.000	108.8933 134.9071

scalar sser = e(ssr)

scalardf_rest=e(df_r)

Cheklangan modeldan saqlangan qoldiq erkinlik darajasidan kiritilgan cheklovlar sonini olish uchun foydalanish mumkin. Chiziqli modeldagi har bir noyob cheklash modeldagi parametrlar sonini bittaga kamaytiradi. Shunday qilib, uchta parametrlilik cheklanmagan modelga (masalan, Big Andy's) bitta cheklov qo'yish cheklangan modeldagi parametrlar sonini ikkitaga qisqartiradi. Cheklangan modeldagi regressorlar soni Kr, cheklanmagan modeldagi esa Ku soni bo'lsin. Cheklanmagan modeldagi erkinlik darajalarini (N-Ku) cheklangan modeldagidan (N-Kr) ayirish siz kiritgan cheklovlar sonini beradi, ya'ni (N-Kr)-(N-Ku) = (Ku-Kr)=J. Statada:

scalarJ=df_rest-df_unrest

Keyin F statistikasi hisoblash mumkin:

scalarfstat=((sser-sseu)/J)/(sseu/(df_unrest))

F(J,N-K) taqsimotidan kritik qiymat va hisoblangan statistika uchun p qiymati odatdagi usulda hisoblanishi mumkin. Bu holda invFtail(J,N-K,α) erkinlik darajalari J numerator va N-K maxraj bilan F tarqatishdan α darajadagi kritik qiymat hosil qiladi. Ftail(J,N-K,fstat) funksiyasi hisoblangan statistika uchun p qiymatini qaytarish uchun xuddi shunday ishlaydi, fstat.

scalarcrit1=invFtail(J,df_unrest,.05)

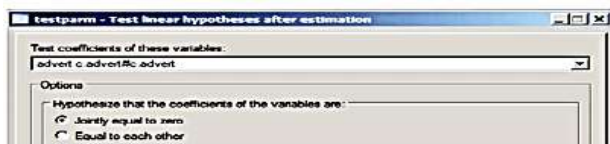
scalarpvalue=Ftail(J,df_unrest,fstat)

scalarlistsseusserJdf_unrestfstatpvaluecrit1

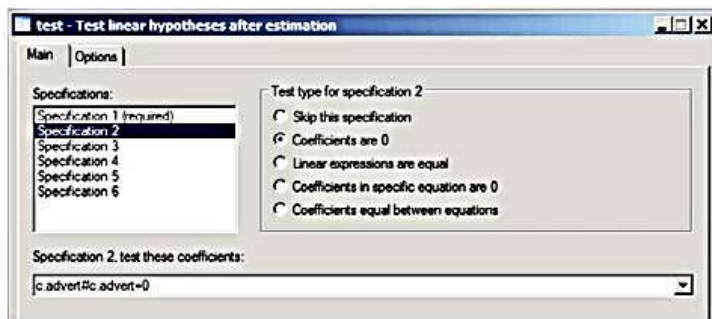
Natijada:

```
. scalar list sseu sser J df_unrest fstat pvalue crit1
      sseu = 1532.0845
      sser = 1896.3908
      J = 2
df_unrest = 71
      fstat = 8.44136
      pvalue = .00051416
      crit1 = 3.1257642
```

Muloqot oynalari model parametrlari bo'yicha cheklovlarni sinab ko'rish uchun ham ishlatilishi mumkin. Birinchi qadam regress yordamida modelni baholashdir. Bu xuddi yuqoridagi 5.1-bo'limda bo'lgani kabi davom etadi. Stataning ochiladigan menyusidan Select Statistics>Linear models and related>Linear regressionni tanlang. Bu regress dialog oynasini ochadi. Regress–Linear regression dialog oynasidagi qaram o'zgaruvchi sifatida sales dan va price, advert hamda o'zaro ta'sir c.advert#c.advertrt dan mustaqil o'zgaruvchilar sifatida foydalanib, OK tugmasini bosish orqali regressiyani hisoblang. Regressiya taxmin qilingandan so'ng, gipotezani tekshirish uchun baholashdan keyingi buyruqlar qo'llaniladi. Ochilgan menyudan Statistics>Postestimation>Tests>Test parametersni tanlang, bu testparm dialog oynasini ochadi:



Shuningdek, test dialog oynasidan Statistics>Postestimation>Tests>Test linear hypothesesni tanlash orqali ham foydalanish mumkin. Test dialogidan foydalanish qiyinroq. Har bir chiziqli gipoteza spetsifikatsiya sifatida kiritilishi kerak. Spetsifikatsiya 1 (majburiy) uchun advert=0 kiriting va koeffitsiyentlar nolga teng yoki chiziqli ifodalar teng radio tugma tanlanganligiga ishonch hosil qiling. Keyin Spetsifikatsiya 2 ni belgilang va c.advert#c.advert=0 kiriting va Submit tugmasini bosing. Ushbu bosqich uchun dialog oynasi quyida ko'rsatilgan:



Ikkala holatda ham Command oynasidan foydalanish ancha oson. Testparm bayonoti parametrlar bo'yicha nol cheklovlarni sinab ko'rish uchun foydalanish eng oddiy hisoblanadi. Sintaksis quyidagicha: **testparmvarlist**

Bu shuni anglatadiki, null ostida nol koeffitsiyentga ega bo'lgan o'zgaruvchilarni oddiygina ro'yxatga olish mumkin. Bundan tashqari, teng variantdan foydalangan holda koeffitsiyentlar bir-biriga teng ekanligini tekshirishga ko'maklashish mumkin.

Test buyrug'i Wald testi yordamida eng so'nggi moslashtirilgan modelning parametrlari haqidagi qo'shma gipotezalarni tekshirish uchun ishlatilishi mumkin. Gipotezalarni aniqlashning bir necha xil usullari mavjud va ulardan bir nechtasi bu qo'llanmada o'rganiladi. Umumiy sintaksis: **test(hypothesis1)(hypothesis2)**

Qo'shma gipotezalarning har biri qavslar to'plamiga kiritilgan. Chiziqli modelda koeffitsiyentlarni o'zgaruvchilar nomlari bilan aniqlash mumkin, chunki ularning ma'nosi aniq. Umuman olganda, agar avval belgilangan bo'lsa, parametr nomidan ham, chiziqli modelda ham foydalanish mumkin `_b[variable name]` sintaksisi. Bu yerda qo'shma nullni tekshirishning uchta ekvivalent usuli mavjud.

```
regress sales price advert c.advert#c.advert
```

```
testparm advert c.advert#c.advert
```

```
test (advert=0)(c.advert#c.advert=0)
```

```
test (_b[advert]=0)(_b[c.advert#c.advert]=0)
```

```

. testparm advert c.advert#c.advert

( 1)  advert = 0
( 2)  c.advert#c.advert = 0

      F( 2,    71) =    8.44
      Prob > F =    0.0005

. test (advert=0)(c.advert#c.advert=0)

( 1)  advert = 0
( 2)  c.advert#c.advert = 0

      F( 2,    71) =    8.44
      Prob > F =    0.0005

. test (_b[advert]=0)(_b[c.advert#c.advert]=0)

( 1)  advert = 0
( 2)  c.advert#c.advert = 0

      F( 2,    71) =    8.44
      Prob > F =    0.0005

```

6.1.1. Modelning ahamiyatini tekshirish

F testining ushbu formula natijasida siz modelingiz kerakli statistik ahamiyatlik darajasi muhim yoki yo'qligini aniqlaymiz. K regressorlari bilan umumiy chiziqli modelni ko'rib chiqamiz:

$$y_i = \beta_1 + x_{i2} \beta_2 + x_{i3} \beta_3 + \dots + x_{iK} \beta_K + e_i$$

Agar erkli o'zgaruvchilar y ning o'rtacha qiymatiga ta'sir qilmasa, unda har bir qiyalik nolga teng bo'lib, nol va muqobil gipotezalarga olib keladi:

$$H_0: \beta_2 = 0, \beta_3 = 0, \dots, \beta_K = 0$$

H_1 : $k=2,3,\dots,K$ uchun β_k ning kamida bittasi nolga teng.

Bu $J=K-1$ cheklovlariga teng. Shunga qaramay, modelni har biri uchun cheklanmagan va cheklangan erkinlik darajalarini baholang. Keyin, test statistikasini hisoblash uchun yuqoridagi Stata buyrug'idan foydalaning:

$$F = \frac{(SST - SSE)/(K - 1)}{SSE/(N - K)} = \frac{(3115,485 - 1532,08)/3}{1532,084/(75 - 4)} = 24,459$$

Stata buyrug'i:

* Unrestricted Model (all variables)

```
regress sales price advert c.advert#c.advert
```

```
scalar sseu = e(rss)
```

```
scalar df_unrest = e(df_r)
```

* Restricted Model (no explanatory variables)

```
regress sales
```

```
scalar sser = e(rss)
```

```
scalar df_rest = e(df_r)
```

```
scalar J = df_rest - df_unrest
```

* F-statistic, critical value, pvalue

```
scalar fstat = ((sser - sseu)/J)/(sseu/(df_unrest))
```

```
scalar crit2 = invFtail(J,df_unrest,.05)
```

```
scalar pvalue = Ftail(J,df_unrest,fstat)
```

```
scalar list sseu sser J df_unrest fstat pvalue crit2
```

Buyruq natijasi:

```
. scalar list sseu sser J df_unrest fstat pvalue crit2
      sseu = 1532.0845
      sser = 3115.4819
        J =      3
df_unrest =     71
      fstat = 24.459316
      pvalue = 5.600e-11
      crit2 = 2.7336472
```

Regressiya ahamiyatining ushbu maxsus testi Stata yordamida hisoblangan har bir chiziqli regressiyaning standart natijasida paydo bo'ladigan darajada muhim. Quyidagi natijada ushbu test uchun F statistika 24,4595 va uning p qiymati 5% dan ancha past. Shuning uchun biz modelning besh foiz darajasida ahamiyatsizligi haqidagi nol gipotezani rad qilamiz.

```
. regress sales price advert c.advert#c.advert
```

Source	SS	df	MS		Number of obs =	75
Model	1583.39741	3	527.799136		F(3, 71) =	24.46
Residual	1532.08446	71	21.5786543		Prob > F =	0.0000
Total	3115.48187	74	42.1011063		R-squared =	0.5082
					Adj R-squared =	0.4875
					Root MSE =	4.6453

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
price	-7.64	1.045939	-7.30	0.000	-9.725543 -5.554457
advert	12.15124	3.556164	3.42	0.001	5.060446 19.24203
c.advert#c.advert	-2.767963	.940624	-2.94	0.004	-4.643514 -.892412
_cons	109.719	6.799045	16.14	0.000	96.16212 123.2759

6.1.2. t va F testlari o'rtasidagi bog'liqlik

Ushbu misolda muhimlik uchun t testi va F testinging ekvivalentligi ko'rsatilgan. Asosiy model:

$$\text{SALES}_i = \beta_1 + \beta_2 \text{PRICE}_i + \beta_3 \text{ADVERT}_i + \beta_4 \text{ADVERT}_i^2 + e_i$$

β_2 uchun t nisbati -7,30 ga teng (6.1.2-bo'lim oxiridagi natijada ifodalanilgan). F testi nol emas degan ikki tomonlama muqobilga nisbatan $\beta_2=0$ degan gipotezani tekshirish uchun ishlatilishi mumkin. Cheklangan model esa:

$$\text{SALES}_i = \beta_1 + \beta_3 \text{ADVERT}_i + \beta_4 \text{ADVERT}_i^2 + e_i$$

Cheklanmagan modelni va cheklanmagan modelni baholash hamda Statada F statistikani hisoblash uchun buyruqlar ketma-ketligi:

* Unrestricted Regression

```
regress sales price advert c.advert#c.advert
```

```
scalar sseu = e(rss)
```

```
scalar df_unrest = e(df_r)
```

```
scalar tratio = _b[price]/_se[price]
```

```
scalar t_sq = tratio^2
```

* Restricted Regression

```
regress sales advert c.advert#c.advert
```

```
scalar sser = e(rss)
```

```
scalar df_rest = e(df_r)
```

```
scalar J = df_rest - df_unrest
```

* F-statistic, critical value, pvalue

```
scalar fstat = ((sser - sseu)/J)/(sseu/(df_unrest))
```

```
scalar crit = invFtail(J,df_unrest,.05)
```

```
scalar pvalue = Ftail(J,df_unrest,fstat)
```

```
scalar list sseu sser J df_unrest fstat pvalue crit tratio t_sq
```

Buyruqlar natijasida quyidagi hisob-kitoblarni ifodalaydi:

```
. scalar list sseu sser J df_unrest fstat pvalue crit tratio t_sq
      sseu = 1532.0845
      sser = 2683.4109
        J = 1
df_unrest = 71
      fstat = 53.354875
     pvalue = 3.236e-10
      crit = 3.9758102
     tratio = -7.3044421
      t_sq = 53.354875
```

F statistika 53,35 ni tashkil qiladi. t nisbatining kvadrati F ga teng ekanligi bejiz emas: $-7,304^2 = 53,35$. Buning sababi t va F tarqatishlar o'rtasidagi aniq bog'liqlikdir.

Erkinlik darajasi df bo'lgan t tasodifiy o'zgaruvchining kvadrati F tasodifiy o'zgaruvchi bo'lib, hisoblagichda erkinlik darajasi 1 va maxrajda erkinlik darajasi df bo'ladi.

6.1.3. Qo'shimcha umumiy F testlari

F testi regressiya koeffitsiyentlari bo'yicha nol cheklovlarni o'z ichiga olgan gipotezalarga qaraganda umumiyroq gipotezalarni tekshirish uchun ham ishlatilishi mumkin. Teng belgilarga ega chiziqli gipotezalarni o'z ichiga olgan K gacha farazlarni tekshirish mumkin. Test xuddi shu tarzda kvadrat xatolarning cheklangan yig'indisini uning cheklanmagan qiymatiga solishtirish orqali amalga oshiriladi. Buning uchun foydalanuvchi tomonidan ba'zi algebraik amallar talab qilinadi.

Ko'rib chiqilgan misol avval V bobda ko'rib chiqilgan reklamaning optimal darajasiga asoslangan. Agar reklamaga qaytishlar kamaysa, reklamaga sarflangan keyingi dollar sotishning faqat yana bir dollarini hosil qilganda reklamaning optimal darajasi yuzaga keladi. Boshqa (ming) dollarning 1 ga teng savdodagi marjinal ta'sirini belgilash:

$$\beta_3 + 2\beta_4 A_0 = 1$$

va A_0 uchun yechish natijasida $A_0 = (1 - \beta_3) / 2\beta_4$ hosil bo'ladi, bunda β_3 va β_4 eng kichik kvadratlar orqali olingan natijalardir. Hisoblangan model natijalarini kiritish reklamaning taxminiy optimal darajasini 2,014 (\$2014) tashkil qiladi. Aytaylik, Andy reklamaning optimal darajasi 1900 dollar degan taxminni sinab ko'rmoqchi 1.9 ni almashtirish (esda tutingki, ma'lumotlardagi reklama 1000 dollar bilan o'lchanadi) nol va muqobil farazlarga olib keladi:

$$H_0: \beta_3 + 3.8\beta_4 = 1$$

$$H_1: \beta_3 + 3.8\beta_4 \neq 1$$

Stata bu taxminning qiymatini nol gipoteza va uning standart xatosi ostida hisoblashni quyidagi buyruq orqali amalga oshiradi:

lincom _b[advert]+3.8*_b[c.advert#c.advert]-1

Oldingi boblardan lincomm buyrug'i o'zidan oldingi regressiya asosida parametrlarning chiziqli birikmalarini topgandik va shu buyruq asosida natijalarni olamiz. **Lincom** dan olingan natija va t nisbatini hisoblash:

```
( 1)  advert + 3.8*c.advert#c.advert = 1
```

sales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
(1)	.6329759	.6541901	0.97	0.337	-.6714421	1.937394

Regressiya chiziqli bo'lgani uchun oddiyroq sintaksisdan bir xil natijalarni olish uchun ham foydalanish mumkin:

lincom advert+3.8*c.advert#c.advert-1

Ikkala holatda ham taxminiy va standart xato hosil bo'ladi va bu miqdorlar mos ravishda $r(\text{estimate})$ va $r(\text{se})$ da saqlanadi. Shunday qilib, siz ularni eslab qolishingiz va t nisbatini qo'lda hisoblash uchun skalyar buyrug'idan foydalanishingiz mumkin. Buni amalga oshirish uchun quyidagi buyruqlardan foydalanamiz:

scalar t = r(estimate)/r(se)

scalar pvalue2tail = 2*ttail(e(df_r),t)

scalar pvalue1tail = ttail(e(df_r),t)

scalar list t pvalue2tail pvalue1tail

Hisoblangan t nisbat uchun bir tomonlama p qiymatini olish uchun `ttail()` buyrug'idan foydalaniladi. Unda `e(df_r)` dan foydalaniladi, bu uni ishlatishdan oldingi savdo regressiyasidan qolgan erkinlik darajasini ifodalaydi.

Natijada:

```
. scalar list t pvalue2tail pvalue1tail
      t = .96757186
pvalue2tail = .33654267
pvalue1tail = .16827134
```

Istalgan cheklovni o'zida mujassam etgan yangi parametr bo'yicha modelni qayta tartiblash imkonini beradigan algebraik formulalarni ishlatish mumkin. Bu **lincom** buyrug'i kabi biror narsani o'z ichiga olmaydigan dasturiy ta'minotdan foydalanilganda foydali bo'ladi.

$$\text{SALES}_i - \text{ADVERT}_i = \beta_1 + \beta_2 \text{PRICE}_i + \theta \text{ADVERT}_i + \beta_4 (\text{ADVERT}_i^2 - 3.8 \text{ADVERT}_i) + e_i$$

Reklama koeffitsiyenti to'liq cheklovni o'z ichiga oladi. Cheklovning to'g'ri yoki yo'qligini tekshirish uchun uning koeffitsiyenti bo'yicha t nisbatidan foydalaning. Statada ikkita yangi o'zgaruvchi kiritamiz:

$x_i = \text{ADVERT}_i^2 - 3.8 \text{ADVERT}_i$ and $y_i = \text{SALES}_i - \text{ADVERT}_i$

gen xstar = c.advert#c.advert-3.8*advert

gen ystar = sales - advert

Kiritilgan yangi o'zgaruvchilardan regressiyada foydalaniladi.

regress ystar price advert xstar

O'zgaruvchan **advert** bo'yicha t nisbati kerakli statistik qiymatdir. Uning ikki tomonlama p qiymati natijada berilgan. Agar buni qo'lda hisoblamoqchi bo'lsangiz, quyidagi buyruq orqali amalga oshirish mumkin:

scalar t = (_b[advert])/_se[advert]

scalar pvalue = ttail(e(df_r),t)

scalar list t pvalue

Yuqoridagi buyruqlar uchun quyidagicha natijaga erishiladi:

. regress ystar price advert xstar

Source	SS	df	MS	Number of obs =	75
Model	1457.21493	3	485.738311	F(3, 71) =	22.51
Residual	1532.08447	71	21.5786545	Prob > F =	0.0000
				R-squared =	0.4875
				Adj R-squared =	0.4658
Total	2989.2994	74	40.3959379	Root MSE =	4.6453

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
price	-7.64	1.045939	-7.30	0.000	-9.725542 -5.554457
advert	.632976	.6541901	0.97	0.337	-.671442 1.937394
xstar	-2.767964	.9406241	-2.94	0.004	-4.643515 -.8924125
_cons	109.719	6.799046	16.14	0.000	96.16212 123.2759

. scalar t = (_b[advert])/_se[advert]

. scalar pvalue = ttail(e(df_r),t)

. scalar list t pvalue
t = .96757201
pvalue = .1682713

Regressiya jadvalidagi advert o'zgaruvchining t nisbati 0,97 ga va ikki tomonlama p qiymati 0,337 ga teng. Skalyar buyrug'i yordamida hisoblangan t nisbat bir xil (ko'proq raqamlarga o'tkazilgan bo'lsa ham) va uning bir tomonlama p qiymati

jadvaldagi ikki tomonlama nisbatning yarmiga teng. Natijalar bevosita mos keladi.

Ushbu bo'lim Big Andyning ikkita taxminini birgalikda sinovdan o'tkazish bilan yakunlanadi. Oylik reklama xarajatlarning optimal darajasi 1900 AQSH dollari tashkil etishini taklif qilishdan tashqari, Big Andy PRICE \$6 narxi va ADVERT 1.9 reklama xarajatlari o'rtacha hisobda sotishdan tushgan daromad keltiradi degan faraz asosida xodimlarni ta'minlash va 80 000 AQSH dollar sotuvda olish rejalashtirmoqda. Qo'shma nol gipoteza:

$$H_0: \beta_3 + 3.8\beta_4 = 1 \quad \text{va} \quad \beta_1 + 6\beta_2 + 1.9\beta_3 + 3.61\beta_4 = 80$$

Bunga qarshi alternativdan farazlarning bittasi to'g'ri emas deb xulosa qilinadi. Qo'shmcha test uchun Stata quyidagi buyruqdan foydalanamiz:

```
regress sales price advert c.advert#c.advert
```

```
test (_b[advert]+3.8*_b[c.advert#c.advert]=1) (_b[_cons] + ///
6*_b[price] + 1.9*_b[advert]+3.61*_b[c.advert#c.advert]= 80)
```

Bu misolda test buyrug'i qo'llaniladi, undan so'ng har ikkala cheklov qo'yiladi, ularning har biri alohida qavslar to'plamida joylashgan. E'tibor bering, test oldingi regressiyadan saqlangan `_b[varname]` koeffitsiyentlari ishlatiladi. Yana bir bor, buni faqat o'zgaruvchilar nomlaridan foydalangan holda chiziqli regressiyada soddalashtirish mumkin:

```
test (advert+3.8*c.advert#c.advert=1) (_cons + 6*price + ///
1.9*advert+3.61*c.advert#c.advert= 80)
```

Natijalar:

```
. test (_b[advert]+3.8*_b[c.advert#c.advert]=1) ///
> (_b[_cons]+6*_b[price]+1.9*_b[advert]+3.61*_b[c.advert#c.advert]= 80)

( 1)  advert + 3.8*c.advert#c.advert = 1
( 2)  6*price + 1.9*advert + 3.61*c.advert#c.advert + _cons = 80

F( 2, 71) = 5.74
Prob > F = 0.0049
```

p qiymati 0,0049 va 5%dan kam bo'lganligi sababli, ushbu muhimlik darajasida nol (qo'shma) gipoteza rad etiladi.

6.2. Namuna bo'lmagan ma'lumotlar orqali model tuzish

Ba'zan sizda modelni baholashda foydalanmoqchi bo'lgan aniq namuna bo'lmagan ma'lumotlar mavjud. Namuna bo'lmagan ma'lumotlardan foydalanish qolgan parametrlarni baholashning aniqligini oshiradi. Princeple of Econometrics-4th nashrdagi adabiyotdan olingan ushbu misolda mualliflar pivo sotish modelini pivo narxlari, spirtli ichimliklar narxi, boshqa tovarlar narxi va daromadning funksiyasi sifatida ko'rib chiqishgan. Shu asnoda ushbu ma'lumotlar asosida modelni tuzib olamiz. O'zgaruvchilar tabiiy logarifmlarida ifodalaymiz:

$$\ln(Q_t) = \beta_1 + \beta_2 \ln(PB_t) + \beta_3 \ln(PL_t) + \beta_4 \ln(PR_t) + \beta_5 \ln(I_t) + e$$

Modelga asosan iqtisodiy nazariya shuni ifodalaydi:

$$\beta_2 + \beta_3 + \beta_4 + \beta_5 = 0$$

Modelni baholash uchun beer.dta ma'lumotlar faylidan foydalanamiz. Ma'lumotlar faylini quyidagi buyruq orqali ochamiz:

use beer, clear

Keyin ma'lumotlar to'plami uchun har bir o'zgaruvchining natural logarifmlarini tuzing. Stataning buyrug'i `log(variable)` o'zgaruvchining natural logarifmini olish uchun ishlatiladi. Shunday qilib, har bir o'zgaruvchining natural logarifimli o'zgaruvchilarini yaratish uchun foydalaning:

use beer, clear

gen lq = ln(q)

gen lpb = ln(pb)

gen lpl = ln(pl)

gen lpr = ln(pr)

gen li = ln(i)

Chiziqli cheklovlarni o'rnatish uchun siz Stata constrained regression deb ataladigan buyruqdan foydalanasiz. Stata cheklashni constraint deb ataladi va chiziqli regressiya modeliga ushbu cheklovlarni qo'yish uchun foydalanadigan buyruq `cnsreg`

hisoblanadi. Sintaksi quyidagicha ifodalanadi:

constraint 1

constraint 2

cnsreg depvar indepvars [if] [in] [weight] , constraints(1 2)

Cheklovlarning (constraints) har biri birinchi navbatda sanab o'tiladi va ularga noyob raqam beriladi. Natijalar olinganidan keyin cnsreg buyrug'i regress kabi ishlatiladi; regressiya modeliga vergul bilan amal qiling va cheklash raqamlari ro'yxati constraints (1 2 ...) va Stata sanab o'tilgan cheklovlarni qo'yadi hamda qolgan parametrlarni baholash uchun eng kichik kvadratlardan foydalanadi. Constraints buyrug'i quyida ko'rsatilgandek c(1 2) qisqartirilishi mumkin. Pivo misoli asosida quyidagi sintaksi orqali amalga oshiriladi:

constraint 1 lpb+lp1+lpr+li=0

cnsreg lq lpb lp1 lpr li, c(1)

Natijada:

```
. constraint 1 lpb+lp1+lpr+li=0
. cnsreg lq lpb lp1 lpr li, c(1)

constrained linear regression
```

	Number of obs = 30
	F(3, 26) = 36.46
	Prob > F = 0.0000
	ROOT MSE = 0.0617

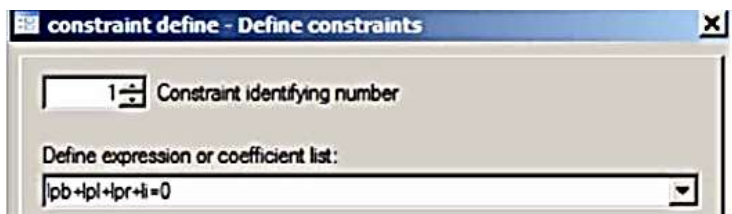

```
( 1) lpb + lp1 + lpr + li = 0
```

	lq	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lpb		-1.299387	.1657377	-7.84	0.000	-1.640065 - .958708
lp1		.1868161	.2843833	0.66	0.517	-.3977422 .7713744
lpr		.1667424	.0770752	2.16	0.040	-.0083121 .3251727
li		.9458282	.4270468	2.21	0.036	.0680209 1.823635
_cons		-4.797793	3.713905	-1.29	0.208	-12.43183 2.836247

Ko'proq harakat talab qilsa ham, ushbu natijalarni olish uchun ochiladigan menyulardan ham foydalanish mumkin. Birinchidan, cheklovni aniqlash kerak. Ketma-ketlik: Statistics > Other > Manage Constraints



Raqamlash va cheklovlarni aniqlash uchun ishlatiladigan dialog oynasini ochish uchun Create tugmasini bosing.



Cheklovraqamini tanlang va Define expression maydoniga kerakli cheklovni kiring. Cheklovni qabul qilish va qutini yopish uchun OK tugmasini bosing. Cheklovlarni qo'shish uchun constraint—Manage constraints maydonidagi Create tugmasini yana bir marta bosing. Ish tugagach, oynani yopish uchun Close tugmasini bosing. Cheklangan modelni baholash uchun ko'rsatilganidek, ochiladigan menyudan Statistics > Linear models and related > Constrained linear regressionni tanlang:



Cheklangan modelni baholash uchun OK yoki Submit tugmasini bosing.

6.3. Model xususiyatlari

Modelni tanlashning uchta asosiy xususiyati quyidagilardan iborat: (1) funksional shaklni tanlash, (2) modelga kiritiladigan erkli o'zgaruvchilarni (regressorlar) tanlash va (3) V bobda sanab o'tilgan ko'plik regressiya modelining MR1-MR6 farazlarini tanlash. Ushbu bo'lim ularning dastlabki ikkitasini o'rganadi.

6.3.1. Yo'qolgan o'zgaruvchilar

Agar siz modelingizdagi tegishli o'zgaruvchilarni o'tkazib yuborsangiz, eng kichik kvadratlar usuli noto'g'ri bo'ladi. Omitted variable problem bilan bog'liq ko'rsatish uchun biz ishlaydigan er va xotin juftliklar namunasini ko'rib chiqamiz. Ma'lumotlar edu_inc.dta faylida saqlanadi.

Ma'lumotlar faylini oching va Stata xotirasidan avval saqlangan ma'lumotlarni o'chiring: **use edu_inc, clear**

Birinchi regressiya oila daromadini bog'liq o'zgaruvchi sifatida (faminc) va tushuntirish o'zgaruvchisi sifatida erning (he) va xotinning ma'lumotini (we) o'z ichiga oladi. Buyruqlar qatoridan:

regress faminc he we

Natijada:

. regress faminc he we

Source	SS	df	MS			
Model	1.3405e+11	2	6.7027e+10	Number of obs =	428	
Residual	6.9703e+11	425	1.6401e+09	F(2, 425) =	40.87	
Total	8.3109e+11	427	1.9463e+09	Prob > F =	0.0000	
				R-squared =	0.1613	
				Adj R-squared =	0.1574	
				Root MSE =	40498	

faminc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
he	3131.509	802.908	3.90	0.000	1553.344	4709.674
we	4522.641	1066.327	4.24	0.000	2426.711	6618.572
_cons	-5533.629	11229.53	-0.49	0.622	-27605.97	16538.71

Xotinning ma'lumotsizligi (we) quyidagilarni beradi:

Source	SS	df	MS			
Model	1.0455e+11	1	1.0455e+11	Nun		
Residual	7.2654e+11	426	1.7055e+09	F(		
Total	8.3109e+11	427	1.9463e+09	Pr		
				R-s		
				Adj		
				R		

faminc	Coef.	Std. Err.	t	P> t
he	5155.483	658.4574	7.83	0.000

Oddiy korrelatsiya tahlili shuni ko'rsatadiki, er va xotinning ta'lim darajasi ijobiy bog'liqdir. Matnda taklif qilinganidek, bu we modeldan chiqarib tashlashhe koeffitsiyentida ijobiy natija olib kelishi mumkinligini anglatadi. Bu baholash modellar tomonidan qo'llab-quvvatlanadi.

```
. correlate
(obs=428)
```

	faminc	he	we	kl6	xtra_x5	xtra_x6
faminc	1.0000					
he	0.3547	1.0000				
we	0.3623	0.5943	1.0000			
kl6	-0.0720	0.1049	0.1293	1.0000		
xtra_x5	0.2898	0.8362	0.5178	0.1487	1.0000	
xtra_x6	0.3514	0.8206	0.7993	0.1595	0.9002	1.0000

Xotinning ma'lumoti va maktabgacha yoshdagi bolalar soni (kl6) hisobga olinsa, quyidagi natijalar hosil bo'ladi:

```
. regress faminc he we kl6
```

Source	SS	df	MS			
Model	1.4725e+11	3	4.9082e+10	Number of obs =	428	
Residual	6.8384e+11	424	1.6128e+09	F(3, 424) =	30.43	
Total	8.3109e+11	427	1.9463e+09	Prob > F =	0.0000	
				R-squared =	0.1772	
				Adj R-squared =	0.1714	
				Root MSE =	40160	

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
he	3211.526	796.7026	4.03	0.000	1645.547 4777.504
we	4776.907	1061.164	4.50	0.000	2691.111 6862.704
kl6	-14310.92	5003.928	-2.86	0.004	-24146.52 -4475.326
_cons	-7755.33	11162.93	-0.69	0.488	-29696.91 14186.25

E'tibor bering, we va he uchun koeffitsiyentlarni baholash avvalgi regressiyaga nisbatan unchalik o'zgarmadi. Buning sababi shundaki, kl6 hech qanday shakllanish o'zgaruvchilari bilan kuchli bog'liq emas. Bu shuni anglatadiki, tegishli o'zgaruvchi ko'rsatilmagan taqdirda ham foydali natijalarga erishish mumkin. O'tkazib yuborilgan o'zgaruvchining ushbu misolda ta'lim o'zgaruvchilari bo'lgan qiziqish o'zgaruvchilari bilan bog'liq bo'lmasligi talab qilinadi. Bunday holda, tegishli o'zgaruvchini o'tkazib yuborish testlarning haqiqiyiligi va we yoki he ishtirok etadigan ishonch oralig'iga ta'sir qilmaydi.

6.3.2. Ahamiyatsiz o'zgaruvchilar

Modelga ahamiyatsiz o'zgaruvchilarni kiritish eng kichik kvadrlarni baholashning aniqligini pasaytiradi. Eng kichik kvadrlar usuli xolis, ammo koeffitsiyentlarning standart xatolari zarur bo'lganidan kattaroq bo'ladi. Ushbu misol modelga ikkita ahamiyatsiz o'zgaruvchini qo'shadi (xtra_x5 va xtra_x6). Bu

o'zgaruvchilar he va we bilan bog'liq, lekin oila daromadining o'rtacha qiymati bilan bog'liq emas. Quyidagilarni olish uchun chiziqli regressiya yordamida modelni baholang:

. regress faminc he we k16 xtra_x5 xtra_x6

Source	SS	df	MS		Number of obs =	428
Model	1.4776e+11	5	2.9553e+10		F(5, 422) =	18.25
Residual	6.8332e+11	422	1.6192e+09		Prob > F =	0.0000
					R-squared =	0.1778
Total	8.3109e+11	427	1.9463e+09		Adj R-squared =	0.1681
					Root MSE =	40240

faminc	coef.	Std. Err.	t	P> t	[95% Conf. Interval]
he	3339.792	1250.039	2.67	0.008	882.7131 5796.871
we	5868.677	2278.067	2.58	0.010	1390.905 10346.45
k16	-14200.18	5043.72	-2.82	0.005	-24114.13 -4286.242
xtra_x5	888.8426	2242.491	0.40	0.692	-3519.001 5296.686
xtra_x6	-1067.186	1981.685	-0.54	0.590	-4962.389 2828.018
_cons	-7558.613	11195.41	-0.68	0.500	-29564.33 14447.1

Hisoblangan standart xatolar oldingi regressiya xatolariga nisbatan qanchalik ko'payganiga e'tibor bering. Agar ular he va we bilan bog'lanmagan bo'lsa, biz ularning standart xatolariga juda kam ta'sir ko'rsatishini kutgan bo'lardik.

6.3.3. Model tanlash

To'g'ri o'zgaruvchilar to'plamini va to'g'ri funksional shaklni tanlash nafaqat fan, balki san'at hamdir. Ideal holda, siz qiziqtirgan parametrlarni izchil va samarali baholash imkonini beruvchi ma'lumotlar va o'zgaruvchilarning tegishli xususiyatlarini qamrab oladigan funksional shaklni tanlashni xohlaysiz.

Ushbu bo'limda ko'pincha o'zgaruvchilarni vaqtincha tanlash va funksional shaklning muvofiqligini tekshirish uchun foydalaniladigan statistik ma'lumotlar muhokama qilinadi.

Modelni tanlash mezonlari

Modelni tanlash uchun uchta mezon ko'rib chiqiladi: hisoblangan R^2 , AIC va SC (BIC). Ushbu statistika muqobil modellarni tanlashda foydali bo'lishi mumkin, garchi ulardan foydalanish munozaralidir. Qanday bo'lmasin, ular faqat model spetsifikatsiyasining barcha boshqa manbalari tugaganidan keyin ishlatilishi kerak. Ya'ni, iloji boricha nazariya va aql-idrokdan foydalaning va model

tanlash qoidalariga faqat muqobil modellarning nisbiy afzalliklari haqida ko'proq ma'lumot olish uchun murojaat qiling. Ushbu ogohlantirishga e'tibor bermay, davom etaylik.

Beshinchi bobda hisoblangan R^2 eng kichik kvadratlarga moslashishning muqobil o'lchovi sifatida taqdim etilgan bo'lib, u oddiy R^2 bilan mashhur muammoni yengib chiqadi, ya'ni modelga regressorlar qo'shilganda u hech qachon kichraymaydi. Hisoblangan R^2 regressor qo'shilganligi sababli tegishli salbiy natijani belgilaydi. Agar moslashuvning yaxshilanishi salbiy natija bilan solishtirganda juda kichik bo'lsa, modelga ahamiyatsiz regressor qo'shilganda hisoblangan R^2 kichikroq bo'lishi mumkin. Hisoblangan R^2 quyidagicha:

$$\bar{R}^2 = 1 - \frac{SSE/(N - K)}{SST/(N - 1)}$$

Ushbu statistika ma'lumot bo'yicha Regress Stata buyrug'i yordamida qayta hisoblaydi. Ko'rib chiqilgan boshqa model tanlash qoidalari Akaike information criterion (AIC) tomonidan berilgan:

$$AIC = \ln\left(\frac{SSE}{N}\right) + 2\frac{K}{N}$$

va Bayesian information criterion(SC) quyidagi formula bilan berilgan.

$$SC = \ln\left(\frac{SSE}{N}\right) + 2\frac{K \ln(N)}{N}$$

Bu ikki statistika juda o'xshash va ikkita atamadan iborat. Birinchisi, muvofiqlik o'lchovidir; moslik qanchalik yaxshi bo'lsa, SSE shunchalik past bo'ladi va uning natural logarifmi shunchalik past bo'ladi. Regressor qo'shish bu formulaning hajmini oshirib bo'lmaydi. Ikkinchisi, regressorni qo'shish mezoniga nisbatan qo'llaniladigan salbiy natija. K qanchalik katta bo'lsa, salbiy natija shunchalik katta bo'ladi. Modellar orasida AIC yoki SC ni minimallashtiradigan modelni tanlash g'oyasidir. Ular faqat salbiy

natija hajmida farqlanadi, SC esa biroz kattaroqdir.

Ushbu mezonlar Stata da mavjud, ammo ular boshqacha hisoblanadi. Stata versiyalari bu yerda muhokama qilinganidan ko'ra kengroq ma'lumotlarni ishlab chiqarish jarayonlarida foydalanish uchun mo'ljallangan, shuning uchun agar kerak bo'lsa, ulardan foydalanishni unutmang¹.

Ushbu mezonlar "Principles of Econometrics", 4-nashr adabiyotida qayta-qayta qo'llaniladi va ushbu qo'llanmaning maqsadlaridan biri ularning natijalarini takrorlashdir. Shuning uchun, uchta model tanlash qoidalarini hisoblash va ko'rsatish uchun dastur yozish tavsiya etiladi. Yozilgandan so'ng, dastur modelning turli xususiyatlarini solishtirish uchun bir necha marta ishga tushirilishi mumkin. IV bobda model tanlash dasturi qayta ko'rib chiqilgan va dasturlash sikllarida qo'llanilgan.

Statada dastur bu shunchaki dastur nomini kiritish orqali kod bloklarini bajarish imkonini beruvchi tuzilmadir. Quyidagi misolda modelsel deb nomlangan dastur yaratilgan. Buyruqlar oynasida modelsel har gal kiritilganda dastur ichidagi kod satrlari bajariladi. Bunday holda, dastur AIC, SC ni hisoblab chiqadi va sozlangan R^2 qiymatini chop etadi, barchasi avval amalga oshirilgan regressiyaga asoslanadi.

Statada dasturlash quyidagicha ishlaydi. Dasturda dastur buyrug'ini berish va unga nom berishdan boshlanadi, masalan, progname. Keyin dastur har safar ishga tushirilganda bajariladigan Stata buyruqlar bloki yoziladi. Dastur oxirida yopiladi. Quyidagi asosiy tuzilma hosil bo'ladi:

```
program progname  
Stata commands  
end
```

Dasturni yozgandan so'ng, uni kompilatsiya qilishingiz kerak. Agar dastur alohida do faylida bo'lsa, oddiy usulda do faylini ishga tushiring. Agar dastur do faylidagi boshqa kod bilan birga bo'lsa, kodni ajratib ko'rsating va ma'lumotni odatdagi tarzda ishga

¹ Darhaqiqat, Stataning baholashdan keyingi buyrug'i estat ic va dan foydalanadi, bu yerda L – model xatolari normal taqsimlanganda maksimal ehtimollik funksiyasining qiymati.

tushiring. Dasturni faqat bir marta kompilatsiya qilish kerak. Dastur Stata so'roviga nuqta bilan dastur nomini, dastur nomini kiritish orqali boshlanadi.

Modelsel dasturi:

program modelsel

scalar aic = $\ln(e(rss)/e(N)) + 2 * e(rank)/e(N)$

scalar bic = $\ln(e(rss)/e(N)) + e(rank) * \ln(e(N))/e(N)$

di "r-square = "e(r2) " and adjusted r_square " e(r2_a)

scalar list aic bic

end

Stata jarayonni tugatmaguningizcha yoki Stataga dasturni xotiradan olib tashlashni aytmaguningizcha dastur xotirada qoladi. Bu ikki yo'l bilan amalga oshiriladi. Dastur nomini tushirish dasturi avval berilgan dasturni (ya'ni, dastur nomini) xotiradan olib tashlaydi. Yana bir usul – drop_all programm yordamida barcha dasturlarni xotiradan olib tashlashdir. Ushbu usuldan faqat Stata xotirasidan barcha foydalanuvchi dasturlarini olib tashlamoqchi bo'lsangiz foydalaning.

Ushbu maxsus dastur Stata regressiyadan so'ng yaratadigan va saqlaydigan natijalardan foydalanadi. Ulardan ba'zilar sizga allaqachon tanish bo'ladi. $e(rss)$ kvadrat xatolar yig'indisini o'z ichiga oladi va $e(N)$ – namuna hajmi. Ishlatilgan yangi natija $e(rank)$ bo'lib, u asosan modelda qancha erkli o'zgaruvchilar borligini o'lchaydi, boshqalar bilan to'liq mos keladiganlar bundan mustasno. Aniqlangan regressiya modelida bu odatda modeldagi koeffitsiyentlar sonini o'lchaydi, K.

Dasturning jarayonida aic va bic skayarlari (ba'zan SC - Shvarts mezonini deb ataladi) hisoblab chiqiladi va modeldagi hisoblangan R^2 qiymatini chop etish uchun display buyrug'i beriladi. Nihoyat, hisoblangan scalar listni chop etish uchun skalyar ro'yxat buyrug'i beriladi.

Modelni baholash va undan olingan modelni tanlash qoidalarini hisoblash uchun, agar siz hali buni qilmagan bo'lsangiz, modelsel dasturini ishga tushiring. Keyin regressiyani baholang va sel

modelini kiriting. Masalan:

```
quietly regress faminc he  
di "Model 1 (he) "  
modelsel  
estimates store Model1
```

Natijada:

```
. di "Model 1 (he) "  
Model 1 (he)  
  
. modelsel  
r-square = .12580103 and adjusted r-square .12374892  
aic = 21.261776  
bic = 21.280744
```

Model tanlash qoidalaridan foydalanish uchun har bir modeldan keyin modelselni ishga tushiring va eng katta hisoblangan R^2 (odatda yomon fikr) yoki eng kichik AIC yoki BIC (yaxshiroq, lekin ajoyib g'oya emas) bo'lganini tanlang:

```
quietly regress faminc he we  
di "Model 2 (he, we) "  
modelsel  
estimates store Model2  
quietly regress faminc he we kl6  
di "Model 3 (he, we, kl6) "  
modelsel  
estimates store Model3  
quietly regress faminc he we kl6 xtra_x5 xtra_x6  
di "Model 4 (he, we, kl6. x5, x6) "  
modelsel  
estimates store Model4
```

Natijada:

```

. di "Model 2 (he, we) "
Model 2 (he, we)

. modelsel
r-square = .16130044 and adjusted r_square .15735362
aic = 21.224993
bic = 21.253445

. di "Model 3 (he, we, k16) "
Model 3 (he, we, k16)

. modelsel
r-square = .17717332 and adjusted r_square .17135143
aic = 21.210559
bic = 21.248495

. di "Model 4 (he, we, k16. x5, x6) "
Model 4 (he, we, k16. x5, x6)

. modelsel
r-square = .17779646 and adjusted r_square .16805472
aic = 21.219148
bic = 21.276051

```

Bu misolda har bir modeldan keyin Statani estimates store buyrug'i chiqariladi va natijalar estimates table buyrug'i yordamida to'planadi.

estimates table Model1 Model2 Model3 Model4, b(%9.3f)
 stfmt(%9.3f) ///
 se stats(N r2 r2_a aic bic)

Variable	Model1	Model2	Model3	Model4
he	5155.483 658.457	3131.509 802.908	3211.526 796.703	3339.792 1250.039
we		4522.641 1066.327	4776.907 1061.164	5868.677 2278.067
k16			-1.43e+04 5003.928	-1.42e+04 5043.720
xtra_x5				888.843 2242.491
xtra_x6				-1067.186 1981.685
_cons	26191.270 8541.108	-5533.629 11229.533	-7755.330 11162.935	-7558.613 11195.411
N	428	428	428	428
r2	0.126	0.161	0.177	0.178
r2_a	0.124	0.157	0.171	0.168
aic	10314.652	10298.909	10292.731	10296.407
bic	10322.770	10311.086	10308.967	10320.761

Legend: b/se

Stata tomonidan yaratilgan ushbu jadval har bir regressiya uchun hisoblangan aic va bic statistikasing Stata versiyalaridan foydalanadi. Shubhasiz, Stata boshqa hisoblashdan foydalanadi. Xavotir olmang, ikkalasi ham hisoblash to'plamlari haqiqiy va bir xil

xulosaga olib keladi. Eng katta R^2 , eng kichik aic va bic statistikasi kabi 3-modeldan olingan. Ushbu misolda 3-model afzal ko'rilgan spetsifikatsiya ekanligi tushuniladi.

Funksional shakl

Funksional shaklni tanlashda nazariy mulohazalar asosiy qo'llanma bo'lishi kerak bo'lsa-da, ko'p hollarda iqtisodiy nazariya yoki sog'lom fikr yetarli emas. Bu yerda RESET testi yordam beradi. RESET funksional shaklni belgilashda aniq xatoga yo'l qo'yganingizni aniqlash uchun qo'pol tekshirish sifatida ishlatilishi mumkin. Bu aslida yetishmayotgan o'zgaruvchilar uchun test EMAS; Buning o'rniga, bu sizning funksional shaklingizning mosligini tekshiramiz.

Sinov oddiy. Nol gipoteza shundaki, sizning funksional shaklingiz adekvat; muqobil bunday emas. Funksional shaklni to'g'ri deb hisoblab, regressiyani baholang va taxmin qilingan qiymatlarni oling. Ularni kvadratga aylantiring va kubga aylantiring, ularni yana modelga qo'shing, regressiyani qayta baholang va y^2 va y^3 ning ahamiyatini birgalikda sinab ko'ring.

Aslida, ushbu testning bir nechta variantlari mavjud. Birinchisi modelga faqat y^2 qo'shadi va uning ahamiyatini F testi yoki unga tenglashtirilgan t testi yordamida tekshiradi. Ikkinchisi y^2 va y^3 ni qo'shadi va keyin ularning ahamiyatini birgalikda tekshirishni amalga oshiradi. Biz ularni mos ravishda RESET (1) va RESET (2) deb ataymiz.

Misol yana oila daromadining regressiyasiga asoslangan. Modelni eng kichik kvadratlar yordamida baholang va chiziqli hisoblangan qiymatlarni regressiyada saqlash uchun predict buyrug'idan foydalaning:

regress faminc he we kl6

predict yhat

Eslatib o'tamiz, regressiya namunasida predict qiymatlarni olish sintaksisi, y_i , yhat, xb ni taxmin qiladi. Ushbu buyruqda yhat siz tayinlayotgan nomdir. Biz xb parametrini xavfsiz o'tkazib yuborishimiz mumkin, chunki bu standart Stata sozlamasidir. Endi y_i yordamida kvadrat va kublarni hosil qiling:

```
gen yhat2 = yhat^2
```

```
gen yhat3 = yhat^3
```

Modelga qo'shilgan yhat2 bilan dastlabki regressiyani taxmin qiling. Yhat2 ning ahamiyatini t-test yoki F-test yordamida tekshiring. Ikkinchisi uchun ko'rsatilgandek Stata test buyrug'idan foydalaning.

```
regress faminc he we kl6 yhat2
```

```
test yhat2
```

Test natijasi:

```
. test yhat2

( 1)  yhat2 = 0
      Constraint 1 dropped

      F( 0, 423) =          .
      Prob > F =          .
```

Shubhasiz, bu iborada muammo bor. Stata cheklov olib tashlanganligini va tekshirish uchun hech narsa qoldirmagani haqida bizga xabar beradi! Muammo shundaki, ma'lumotlar noto'g'ri. Kompyuter arifmetik amallarni bajarishi uchun ma'lumotlar to'plamidagi o'zgaruvchilar bir xil qiymatga ega bo'lishi kerak. Modeldagi o'zgaruvchilar uchun umumiy statistik ma'lumotlarni ko'rib chiqing.

```
. summarize faminc he we kl6
```

variable	obs	Mean	Std. Dev.	Min	Max
faminc	428	91213	44117.35	9072	344146.3
he	428	12.61215	3.035163	4	17
we	428	12.65888	2.285376	5	17
kl6	428	.1401869	.3919231	0	2

faminc darajasi boshqa o'zgaruvchilarga qaraganda 1000 baravar ko'p. Chiziqli regressiya bashoratlari ham xuddi shunday masshtabga ega bo'ladi. Eslatib o'tamiz, chiziqli regressiyada bog'liq va mustaqil o'zgaruvchilar shkalasining o'zgarishi faqat koeffitsiyentlarning qiymatlariga ta'sir qiladi, ammo sezilarli regressiya natijalari emas. Shunday qilib, ma'lumotlardan yomon shartli bashoratlarni olib tashlang va faminc ko'lamini 10 000 ga bo'lish orqali o'zgartiring:

drop yhat yhat2 yhat3
 gen faminc_sc = faminc/10000

Endi modelni baholaymiz, hisoblangan qiymatlarni saqlang va kvadrat hamda kublarni yarating:

regress faminc_sc he we kl6

. regress faminc_sc he we kl6

Source	SS	df	MS	Number of obs =	428
Model	1472.46499	3	490.821663	F(3, 424) =	30.43
Residual	6838.40844	424	16.1283218	Prob > F =	0.0000
				R-squared =	0.1772
Total	8310.87343	427	19.4634038	Adj R-squared =	0.1714
				Root MSE =	4.016

faminc_sc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
he	.3211526	.0796703	4.03	0.000	.1645547 .4777504
we	.4776908	.1061164	4.50	0.000	.2691111 .6862704
kl6	-1.431092	.5003928	-2.86	0.004	-2.414652 -.4475326
_cons	-.775533	1.116293	-0.69	0.488	-2.969691 1.418625

predict yhat

gen yhat2 = yhat^2

gen yhat3 = yhat^3

RESET(1) uchun modelga yhat2 qo'shing va uning t testi yoki F-testi yordamida uning ahamiyatini tekshiring.

. regress faminc_sc he we kl6 yhat2

Source	SS	df	MS	Number of obs =	428
Model	1567.85524	4	391.963811	F(4, 423) =	24.59
Residual	6743.01819	423	15.9409413	Prob > F =	0.0000
Total	8310.87343	427	19.4634038	R-squared =	0.1887
				Adj R-squared =	0.1810
				Root MSE =	3.9926

faminc_sc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
he	-.2381465	.2419692	-0.98	0.326	-.7137582 .2374653
we	-.4235108	.383214	-1.11	0.270	-1.176752 .32973
kl6	1.088733	1.143928	0.95	0.342	-1.159757 3.337224
yhat2	.099368	.0406211	2.45	0.015	.0195237 .1792123
_cons	8.724297	4.03894	2.16	0.031	.785406 16.66319

. test yhat2

(1) yhat2 = 0

F(1, 423) = 5.98
 Prob > F = 0.0148

Yana bir bor, kvadrat t-nisbati F statistikaga teng va ular bir xil p qiymatiga ega. RESET(2) uchun yhat3 ni qo'shing va kvadrat hamda kubik bashoratlarning umumiy ahamiyatini tekshiring:

```
. regress faminc_sc he we k16 yhat2 yhat3
```

Source	SS	df	MS
Model	1572.19036	5	314.438072
Residual	6738.68307	422	15.9684433
Total	8310.87343	427	19.4634038

```
Number of obs = 428
F( 5, 422) = 19.69
Prob > F = 0.0000
R-squared = 0.1892
Adj R-squared = 0.1796
Root MSE = 3.9961
```

faminc_sc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
he	-.8451478	1.189891	-0.71	0.478	-3.184 1.493704
we	-1.301625	1.72841	-0.75	0.452	-4.698991 2.095741
k16	3.741007	5.217535	0.72	0.474	-6.514588 13.9966
yhat2	.3234728	.4320297	0.75	0.454	-.5257254 1.172671
yhat3	-.0085692	.0164465	-0.52	0.603	-.0408965 .023758
_cons	15.01857	12.73868	1.18	0.239	-10.0206 40.05774

```
. test yhat2 yhat3
```

```
( 1) yhat2 = 0
( 2) yhat3 = 0
```

```
F( 2, 422) = 3.12
Prob > F = 0.0451
```

RESET(1) va RESET(2) ikkalasi ham 5% darajasida muhim ahamiyatga ega va asl chiziqli funksional shakl ushbu munosabatlarni modellashtirish uchun mos emas degan xulosaga kelish mumkin.

Stata regressiyadan so'ng RESET(3) testini bajaradigan keyingi baholash buyrug'ini o'z ichiga oladi. Sintaksi quyidagicha:

```
regress faminc he we k16
```

```
estat ovtest
```

```
. estat ovtest
```

```
Ramsey RESET test using powers of the fitted values of faminc
Ho: model has no omitted variables
F(3, 421) = 2.15
Prob > F = 0.0931
```

RESET ning ushbu versiyasi modelga y^2 , y^3 va y^4 qo'shadi va ularning umumiy ahamiyatini tekshiradi. Texnik jihatdan, buning hech qanday yomon joyi yo'q. Biroq, juda ko'p y^2 kuchlarini kiritish ko'pincha tavsiya etilmaydi, chunki RESET y^2 kuchlari qo'shilganda statistik quvvatni tezda yo'qotadi.

6.4. Ma'lumotlar yetishmaslik, kolinearlik va ahamiyatsizlik

Oldingi bo'limda biz Stata hisob-kitoblaridan biri ma'lumotlarning yomon ishlashi tufayli muvaffaqiyatsizlikka uchraganini aytib o'tgan edik. Bu kollinearlik regressiya bilan

nima qilishiga o'xshaydi. Kollinearlik parametrlarni baholash va boshqa turli statistik ma'lumotlarni katta aniqlik bilan hisoblashni qiyinlashtiradi yoki imkonsiz qiladi. Statistik modelda kollinearlik noto'g'ri eksperimental dizayndan kelib chiqadi yoki bizning holatimizda parametrlarni aniq o'lchash imkonini beradigan ma'lumotlar yetarli darajada farq qilmaydi. Afsuski, buning oddiy davosi yo'q; ma'lumotlarning o'lchamini o'zgartirish uning tarkibidagi chiziqli munosabatlarga ta'sir qilmaydi.

Bu misol cars.dta dan foydalanadi. Oldingi barcha ma'lumotlarni xotiradan o'chirish orqali avtomobil ma'lumotlarini yuklab oling:

use cars, clear

Xulosa statistik ma'lumotlarga nazar tashlash (summarize) ma'lumotlardagi asosli farqlarni ko'rsatadi.

. summarize

variable	Obs	Mean	Std. Dev.	Min	Max
mpg	392	23.44592	7.805007	9	46.6
cyl	392	5.471939	1.705783	3	8
eng	392	194.412	104.644	68	455
wgt	392	2977.584	849.4026	1613	5140

O'zgaruvchilarning har biri ularning diapazoni va standart og'ishlari bilan o'lchanadigan o'zgarishlarni o'z ichiga oladi. Oddiy korrelatsiyalar (corr) potensial muammoni ochib beradi.

**. corr
(obs=392)**

	mpg	cyl	eng	wgt
mpg	1.0000			
cyl	-0.7776	1.0000		
eng	-0.8051	0.9508	1.0000	
wgt	-0.8322	0.8975	0.9330	1.0000

E'tibor bering, potensial tushuntirish o'zgaruvchilari (cyl, eng, wgt) orasida korrelatsiyalar juda yuqori; eng kichik qiymat cyl va wgt o'rtasida sodir bo'ladi va deyarli 0,9 ni tashkil qiladi. Ushbu o'zgaruvchilarning har birining gallon uchun milga mustaqil ta'sirini baholash qiyin bo'ladi.

Birinchidan, dvigateldagi silindrlar (mpg) soniga bog'liq bo'lgan

gallon boshiga milning oddiy modelini taxmin qiling:

regress mpg cyl

. regress mpg cyl

Source	SS	df	MS
Model	14403.0831	1	14403.0831
Residual	9415.91039	390	24.14336
Total	23818.9935	391	60.9181419

Number of obs = 392
 F(1, 390) = 596.56
 Prob > F = 0.0000
 R-squared = 0.6047
 Adj R-squared = 0.6037
 Root MSE = 4.9136

mpg	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
cyl	-3.558078	.1456755	-24.42	0.000	-3.844486 -3.271671
_cons	42.91551	.8348668	51.40	0.000	41.2741 44.55691

Avtomobil dvigatelining hajmini kub dyuymdagi modelning og'irligiga (wgt) qo'shing.

regress mpg cyl eng wgt

. regress mpg cyl eng wgt

Source	SS	df	MS
Model	16656.4443	3	5552.1481
Residual	7162.54916	388	18.4601782
Total	23818.9935	391	60.9181419

Number of obs = 392
 F(3, 388) = 300.76
 Prob > F = 0.0000
 R-squared = 0.6993
 Adj R-squared = 0.6970
 Root MSE = 4.2965

mpg	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
cyl	-.2677968	.4130673	-0.65	0.517	-1.079927 .5443336
eng	-.012674	.0082501	-1.54	0.125	-.0288944 .0035465
wgt	-.0057079	.0007139	-8.00	0.000	-.0071115 -.0043043
_cons	44.37096	1.480685	29.97	0.000	41.45979 47.28213

Endi bir qator farazlarni sinab ko'ring. Birinchisi cyl qiymati uchun, ikkinchisi eng qiymati uchun, uchinchi esa ularning birlashgan qiymati uchun:

test cyl

test eng

test cyl eng

Natijalar:

```

. test cyl
( 1)  cyl = 0
      F( 1, 388) = 0.42
      Prob > F = 0.5172

. test eng
( 1)  eng = 0
      F( 1, 388) = 2.36
      Prob > F = 0.1253

. test eng cyl
( 1)  eng = 0
( 2)  cyl = 0
      F( 2, 388) = 4.30
      Prob > F = 0.0142

```

Aslida, o'zgaruvchilarning hech biri alohida ahamiyatga ega emas, lekin ular birgalikda 5% darajasida muhim ahamiyatga ega. Bu ularning individual ta'sirini aniq o'lchay olmaganligingiz bilan bog'liq bo'lishi mumkin. Oddiy korrelatsiyalar cyl, eng va wgt mustaqil o'zgaruvchilari bir-biri bilan yuqori korrelatsiya qilishini ko'rsatdi. Bu mustaqil o'zgaruvchilarning har biri boshqalar uchun regressga uchragan bir nechta yordamchi regressiyalarni baholash orqali tasdiqlanishi mumkin:

regress cyl eng wgt

scalar r1 = e(r2)

regress eng wgt cyl

scalar r2 = e(r2)

regress wgt eng cyl

scalar r3 = e(r2)

0,8 dan yuqori bo'lgan R^2 qiymati kuchli kollinearlikni ko'rsatadi, bu barcha o'zgaruvchilarni o'z ichiga olgan model parametrlarini baholashning aniqligiga salbiy ta'sir ko'rsatishi mumkin. Ushbu misolda R^2 0,93, 0,90 va 0,87 bo'lib, 0,8 chegaradan ancha yuqori. Bu avtomobilning gaz yurishiga siljishning individual hissi va silindrlar sonini farqlash qiyin bo'lishining yana bir tasdig'idir.

```
. scalar list r1 r2 r3
      r1 = .90490236
      r2 = .93665456
      r3 = .87160914
```

Kollinearlikni aniqlash uchun oddiy korrelatsiyalar o'rniga yordamchi regressiyalardan foydalanishning afzalligi ushbu misolda unchalik aniq emas. Regressiyada ko'plab o'zgaruvchilar mavjud bo'lganda, korrelatsiyalar bilan kollinearlikni aniqlash qiyin bo'lishi mumkin. Ikkita o'zgaruvchining yuqori korrelatsiyasi bo'lmasa-da, ba'zi o'zgaruvchilar aniq bo'lmagan chiziqli munosabatlarga ega bo'lishi mumkin. Bunday vaziyatlarda R^2 ni yordamchi ko'p regressiyadan ko'rib chiqish foydaliroqdir.

Oltinchi bob mavzularini mustahkamlash uchun savollar:

1. F testi nimaga xizmat qiladi?
2. Nol gipotezani (cheklovlarni) tekshirish uchun qanday formuladan foydalaniladi?
3. F testining ahamiyatini tekshirish qanday ifodalanadi?
4. t va F testlari o'rtasidagi bog'liqlikni tushuntirib bering.
5. Namuna bo'lmagan ma'lumotlar orqali model tuzish ketma-ketligini tushuntiring.
6. O'zgaruvchilar tabiiy logarifmlarida ifodalash formulasini ifodalang.
7. Cheklangan modelni baholash uchun qanday buyruqlar ketma ketligi bajariladi?
8. Yo'qolgan o'zgaruvchilar ma'lumotlari qaysi faylda saqlanadi?
9. Akaike information criterion (AIC) qanday amal bajarishga xizmat qiladi?
10. Bayesian information criterion(SC) qanday formula orqali ifodalanadi?
11. RESET testi qanday vazifa bajaradi?

Oltinchi bob mavzularida qo'llanilgan buyruqlar va atamalar

adjusted R2	F-statistic	progname
AIC	functional form	program drop _all
BIC	invFtail(J,N_K,alpha)	regress
cnsreg	invttail(df,alpha)	RESET
collinearity	irrelevant variables	restricted regression
constraint	joint significance	restricted sum of
e(df_r)	test	squares
e(r2)	lincom	Schwartz criterion
e(r2_a)	Manage constraints	test (hypoth 1)
e(rank)	model selection	(hypoth 2)
e(rss)	omitted variables	testparm varlist
estat ovtest	overall F-test	t-ratio
estimates store	predict, xb	ttail(df,tstat)
estimates table	program	unrestricted sum of
Ftail(J,N-K,fstat)	program drop	squares

Oltinchi bob mavzularini mustahkamlash uchun masalalarning do-file lari

Ma'lumotlar buyrug'i: `cd c:\data\poe4stata`

1-masala. log file ni ochish va datalardan foydalanish.

`log using chap06, replace text`

`use andy, clear`

* -----

* The following block estimates Andy's sales

* and uses the difference in SSE to test

* a hypothesis using an F-statistic

* -----

* Unrestricted Model

`regress sales price advert c.advert#c.advert`

`scalar sseu = e(rss)`

`scalar df_unrest = e(df_r)`

```

* Restricted Model
regress sales price
scalar sser = e(rss)
scalar df_rest = e(df_r)
scalar J = df_rest - df_unrest
* F-statistic, critical value, pvalue
scalar fstat = ((sser - sseu)/J)/(sseu/(df_unrest))
scalar crit1 = invFtail(J,df_unrest,.05)
scalar pvalue = Ftail(J,df_unrest,fstat)
scalar list sseu sser J df_unrest fstat pvalue crit1
* -----
* Here, we use Stata's test statement
* to test hypothesis using an F-statistic
* Note: Three versions of the syntax
* -----
regress sales price advert c.advert#c.advert
testparm advert c.advert#c.advert
test (advert=0)(c.advert#c.advert=0)
test (_b[advert]=0)(_b[c.advert#c.advert]=0)
* -----
* Overall Significance of the Model
* Uses same Unrestricted Model as above
* -----
* Unrestricted Model (all variables)
regress sales price advert c.advert#c.advert
scalar sseu = e(rss)
scalar df_unrest = e(df_r)
* Restricted Model (no explanatory variables)
regress sales
scalar sser = e(rss)
scalar df_rest = e(df_r)
scalar J = df_rest - df_unrest
* F-statistic, critical value, pvalue
scalar fstat = ((sser - sseu)/J)/(sseu/(df_unrest))
scalar crit2 = invFtail(J,df_unrest,.05)

```

```

scalar pvalue = Ftail(J,df_unrest,fstat)
scalar list sseu sser J df_unrest fstat pvalue crit2
* -----
* Relationship between t and F
* -----
* Unrestricted Regression
regress sales price advert c.advert#c.advert
scalar sseu = e(rss)
scalar df_unrest = e(df_r)
scalar tratio = _b[price]/_se[price]
scalar t_sq = tratio^2
* Restricted Regression
regress sales advert c.advert#c.advert
scalar sser = e(rss)
scalar df_rest = e(df_r)
scalar J = df_rest - df_unrest
* F-statistic, critical value, pvalue
scalar fstat = ((sser -sseu)/J)/(sseu/(df_unrest))
scalar crit = invFtail(J,df_unrest,.05)
scalar pvalue = Ftail(J,df_unrest,fstat)
scalar list sseu sser J df_unrest fstat pvalue crit tratio t_sq
* -----
* Optimal Advertising
* Uses both sets of syntax for test
* -----
* Equivalent to Two sided t-test
regress sales price advert c.advert#c.advert
test _b[advert]+3.8*_b[c.advert#c.advert]=1
test advert+3.8*c.advert#c.advert=1
* t stat for Optimal Advertising (use lincom)
lincom _b[advert]+3.8*_b[c.advert#c.advert]-1
lincom advert+3.8*c.advert#c.advert-1
scalar t = r(estimate)/r(se)
scalar pvalue2tail = 2*ttail(e(df_r),t)
scalar pvalue1tail = ttail(e(df_r),t)

```

```

scalar list t pvalue2tail pvalue1tail
* t stat for Optimal Advertising (alternate method)
gen xstar = c.advert#c.advert-3.8*advert
gen ystar = sales - advert
regress ystar price advert xstar
scalar t = (_b[advert])/_se[advert]
scalar pvalue = ttail(e(df_r),t)
scalar list t pvalue
* One-sided t-test
regress sales price advert c.advert#c.advert
lincom advert+3.8*c.advert#c.advert-1
scalar tratio = r(estimate)/r(se)
scalar pval = ttail(e(df_r),tratio)
scalar crit = invttail(e(df_r),.05)
scalar list tratio pval crit
* Joint Test
regress sales price advert c.advert#c.advert
test (_b[advert]+3.8*_b[c.advert#c.advert]=1) ///
(_b[_cons]+6*_b[price]+1.9*_b[advert]+3.61*_b[c.advert#c.
advert]= 80)

```

* -----

2-masala. Noodatiy ma'lumotlar asosida ishlash.

* -----

```

use beer, clear
gen lq = ln(q)
gen lpb = ln(pb)
gen lpl = ln(pl)
gen lpr = ln(pr)
gen li = ln(i)
constraint 1 lpb+lpl+lpr+li=0
cnsreg lq lpb lpl lpr li, c(1)

```

* -----

3-masala. MROZ misollari bilan ishlash.

* -----

```

use edu_inc, clear

```

```

regress faminc he we
regress faminc he
* correlations among regressors
correlate
regress faminc he we kl6
* Irrelevant variables
regress faminc he we kl6 xtra_x5 xtra_x6
* Model selection
program modelsel
scalar aic = ln(e(rss)/e(N))+2*e(rank)/e(N)
scalar bic = ln(e(rss)/e(N))+e(rank)*ln(e(N))/e(N)
di "r-square = "e(r2) " and adjusted r_square " e(r2_a)
scalar list aic bic
end
quietly regress faminc he
di "Model 1 (he) "
modelsel
estimates store Model1
quietly regress faminc he we
di "Model 2 (he, we) "
modelsel
estimates store Model2
quietly regress faminc he we kl6
di "Model 3 (he, we, kl6) "
modelsel
estimates store Model3
quietly regress faminc he we kl6 xtra_x5 xtra_x6
di "Model 4 (he, we, kl6. x5, x6) "
modelsel
estimates store Model4
estimates table Model1 Model2 Model3 Model4, b(%9.3f)
stfmt(%9.3f) se stats(N r2 r2_a aic bic)
* RESET
regress faminc he we kl6
predict yhat

```

```
gen yhat2=yhat^2
gen yhat3=yhat^3
summarize faminc he we kl6
```

* -----

4-masala. Mashinalar ma'lumotlari asosida ishlash.

* -----

```
use cars, clear
summarize
corr
regress mpg cyl
regress mpg cyl eng wgt
test cyl
test eng
test eng cyl
* Auxiliary regressions for collinearity
* Check: r2 > .8 means severe collinearity
regress cyl eng wgt
scalar r1 = e(r2)
regress eng wgt cyl
scalar r2 = e(r2)
regress wgt eng cyl
scalar r3 = e(r2)
scalar list r1 r2 r3
log close
program drop modelsel
```

VII BOB. O'ZGARUVCHILAR INDIKATORIDAN FOYDALANISH

7.1. O'zgaruvchilar indikatori

Indikator yoki dummy model, o'zgaruvchilar ikkilik sanoq sistemasi 0/1 o'zgaruvchilari bo'lib, a ning mavjud yoki yoqligini xarakterlaydi. Ushbu bo'limda biz ko'chmas mulk misolida o'zgaruvchi indikatorlaridan foydalanishni o'rganamiz. Yangi log faylini oching va **utown.dta** data faylini oching.

log using chap07_utown, replace text

use utown, clear

describe

summarize

Ma'lumotlarni va birinchi oltita kuzatuvni umumlashtiring:

list in 1/6

list in 501/506

. list in 1/6

	price	sqft	age	utown	pool	fplace
1.	205.452	23.46	6	0	0	1
2.	185.328	20.03	5	0	0	1
3.	248.422	27.77	6	0	0	0
4.	154.69	20.17	1	0	0	0
5.	221.801	26.45	0	0	0	1
6.	199.119	21.56	6	0	0	1

. list in 501/506

	price	sqft	age	utown	pool	fplace
501.	314.65	29.28	24	1	1	0
502.	288.556	24.48	4	1	0	1
503.	302.834	27.02	1	1	0	1
504.	247.82	21.26	2	1	0	1
505.	269.971	22.76	4	1	0	0
506.	292.926	26	17	1	0	1

7.1.1. O'zgaruvchilar indikatorlarini yaratish

Stata dasturida yangi indikatorli o'zgaruvchilarni yaratish uchun ko'plab xususiyatlar mavjud bo'lib, boshqa indikatorli o'zgaruvchilarning qiymatlariga asoslangan holda yangi indikatorli o'zgaruvchilarni hosil qilish uchun **"generate"** (yoki **"gen"**) buyrug'idan foydalaniladi.

Yakuniy statistik xulosaning har bir detallarini hisoblash uchun **price** va **sqft** buyruqlaridan foydalaniladi:

summarize price sqft, detail

Katta uylar sifatidagi o'zgaruvching o'zgaruvchilar indikatorini yarating, ya'ni o'lchami 2500 kvadrat futdan katta bo'lgan uylar uchun. Buni amalga oshirishda **"generate"** buyrug'i bilan birgalikda quyidagicha shart qo'yamiz ($SQFT > 25$):

gen large = (sqft > 25)

Agar qo'yilgan shart bo'yicha uylarning maydoni ($SQFT > 25$) 2500 dan katta bo'lsa, natija **"to'g'ri"** va **"generate"** funksiyasi natijaga 1 qiymat beradi. Agar aksincha bo'lsa, natija **"notog'ri"** yoki qiymati 0 ga teng bo'ladi.

Eslatma: Biz taqdim etgan darslik ma'lumotlarida yetishmayotgan qiymatlar yo'q. Agar ma'lumotlar yetishmayotgan qiymatlarga ega bo'lsa, yuqorida tavsiflangan "mantiqiy operatorlar"dan foydalanish xavflidir. Misol uchun, agar sqft qiymatlari yetishmayotgan bo'lsa, ular "katta" deb tasniflanadi, bu siz xohlagan natija bo'lmasligi mumkin. Ushbu avtomatik buyruqlar bilan ehtiyot bo'ling.

O'zgaruvchi indikatorlarni yaratilayotganda "o'rta narxli" uylar uchun 1 qiymatga teng o'zgaruvchi foydalaniladi.

gen midprice = (215 < price) & (price < 275)

"&" – bu mantiqiy amal bo'lib, ikki shart kiritilganda ularning har ikkisi shartga muvofiqligi yoki aksinchaligini ifodalash uchun ishlatiladi. Ya'ni kiritilgan ikki shart **PRICE > 215** va **PRICE < 275** lar

har ikkisi ham to'g'ri bo'lsa, **MIDPRICE** o'zgaruvchisining qiymati 1 ga teng bo'ladi. Quyida keltirilgan bir necha kuzatuvlarning natijalarini ko'rishingiz mumkin: **list sqft price large midprice in 1/5**

7.1.2. O'zgaruvchilar indikatorining regressiyasini baholash

Aslida o'zgaruvchili indikatorli modellarni aniqlash (baholash) boshqa reggresiyalardan farq qilmaydi. Quyidagi modelni ko'rib chiqamiz:

$$\text{PRICE} = \beta_1 - \delta_1 \text{UTOWN} + \delta_2 \text{SQFT} + \gamma(\text{SQFT} * \text{UTOWN}) + \beta_3 \text{AGE} + \delta_2 \text{POOL} + \delta_3 \text{FPLACE} + e$$

Faktor o'zgaruvchisi yozuvidan foydalanish modeli

reg price i.utown sqft i.utown#c.sqft age i.pool i.fplace

Regressiya buyrug'i, ya'ni regress qisqartirib reg shaklida qo'llaniladi. Bu Stata tez-tez ishlatiladigan buyruqlarni qisqartma shaklini qabul qilishiga yaqqol misol (masalan, **generate** buyrug'ini gen shaklda qo'llaniladi). Uzlüksiz uchun omil o'zgaruvchisi **c. faqatgina** o'zgaruvchilarning o'zaro ta'sir etivchi holatlarda qo'llaniladi.

i.utown#c.sqft kiritmasi **UTOWN** va **SQFT** orasidagi o'zaro ta'sirni ifodalaydi. Biz **Stata A,B** va **A#B** larni qabul qilishi uchun "**A##B**" dan foydalanishimizning o'zi kifoya.

reg price i.utown##c.sqft age i.pool i.fplace

Hosil bo'ladigan natijada o'zgaruvchi indikator **UTOWN 1.utown** kabi va o'zaro ta'sir koeffitsiyenti iborasi **SQFT *UTOWN utown#c.sqft.** kabi namoyon bo'ladi.

```
. reg price i.utown#c.sqft age i.pool i.fpplace
```

Source	SS	df	MS			
Model	1548261.71	6	258043.619	Number of obs = 1000		
Residual	230184.426	993	231.807076	F(6, 993) = 1113.18		
Total	1778446.14	999	1780.22637	Prob > F = 0.0000		
				R-squared = 0.8706		
				Adj R-squared = 0.8698		
				Root MSE = 15.225		

price	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
1.utown sqft	27.45295 7.612177	8.422582 .2451765	3.26 31.05	0.001 0.000	10.92485 7.131053	43.98106 8.0933
utown#c.sqft 1	1.299405	.3320478	3.91	0.000	.6478091	1.951001
age	-.1900864	.0512046	-3.71	0.000	-.2905681	-.0896048
1.pool	4.377163	1.196692	3.66	0.000	2.028828	6.725498
1.fpplace	1.649176	.9719568	1.70	0.090	-.2581495	3.556501
_cons	24.49998	6.191721	3.96	0.000	12.34962	36.65035

7.1.3. O'zgaruvchilar indiktorining ahamiyatlilik testi

O'zgaruvchili indikator koeffitsiyenti haqidagi gipotezalarini tekshirish boshqa koeffitsiyentlar haqidagi gipotezalarni tekshirishdan farq qilmaydi. Universitet shaharchasi joylashuvining ahamiyatlilik testini o'tkazish uchun biz boshqa nolli bo'lmagan gipotezalarga qarshi $H_0: \delta_1 = 0, \gamma = 0$ nolli gipotezasini qo'yamiz.

Bu gipotezani F testiga tekshirish uchun post-estimation buyrug'i yordamida amalga oshiriladi. Stata menyusining yuqorisidagi Statistics>Postestimation >Tests>Test linear hypotheses ketma-ketlikni amalga oshiring yoki regressiyadan so'ng test dialog oynasini ochish uchun db testni kiriting. Stata buyrug'i esa quydagicha:

```
test 1.utown = 0 1.utown#c.sqft
```

Natija:

```
. test 1.utown 1.utown#c.sqft
```

```
( 1) 1.utown = 0
( 2) 1.utown#c.sqft = 0
```

```
F( 2, 993) = 1954.83
Prob > F = 0.0000
```

Sinov natijasi asosida, p qiymati 0.0000 bilan, ahamiyatlilik

darajasi $\alpha=0.05$ yoki hatto $\alpha=0.001$ bo'lmagan sababli biz nolli gipotezimizni inkor etamiz.

7.1.4. Qo'shimcha hisob-kitoblar

Uylarning universitet oldida joylashuvi funksiyasining regressiyasi quyidagicha hisoblanadi:

$$(\text{PRICE}) = (24.5 + 27.453) + (7.6122 + 1.2994)\text{SQFT} - .1901\text{AGE} + 4.3772\text{POOL} + 1.6492\text{FPLACE} = 51.953 + 89116\text{SQFT} - .1901\text{AGE} + 4.3772\text{POOL} + 1.6492\text{FPLACE}$$

Universitet yaqinida joylashgan uylar regressiya og'ishi va kesishishini aniqlash uchun lincom buyrug'idan foydalanamiz:

lincom _cons + 1 .utown

lincom c.sqft + 1 .utown#c.sqft

Natijalar nafaqat ehtimollikni ko'rsatadi, balki ularning 95% ehtimollik orlig'ida ekanini ham ko'rsatadi.

```
. lincom _cons + 1.utown
( 1) 1.utown + _cons = 0
```

price	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
(1)	51.95294	5.767235	9.01	0.000	40.63557	63.2703

```
. lincom c.sqft + 1.utown#c.sqft
( 1) sqft + 1.utown#c.sqft = 0
```

price	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
(1)	8.911581	.2247944	39.64	0.000	8.470455	9.352708

7.1.5. O'rtacha marjinal ta'sirni hisoblash

Faktorli o'zgaruvchilarni yozma ko'rinishda foydalanishning yana bir qulaylik tomoni shundaki, marjinal effektlarni hisoblayotganda margins buyrug'idan to'g'ridan to'g'ri foydalaniladi:

marjins, dydx(*)

Natijada:

```
. margins, dydx(*)
```

```
Average marginal effects  
Model VCE      : OLS
```

```
Number of obs   =      1000
```

```
Expression      : Linear prediction, predict()  
dy/dx w.r.t.    : 1.utown sqft age 1.pool 1.fplace
```

	dy/dx	Delta-method Std. Err.	z	P> z	[95% Conf. Interval]	
1.utown	60.21049	.9646176	62.42	0.000	58.31988	62.10111
sqft	8.286568	.1661803	49.86	0.000	7.96086	8.612275
age	-.1900864	.0512046	-3.71	0.000	-.2904456	-.0897272
1.pool	4.377163	1.196692	3.66	0.000	2.031691	6.722636
1.fplace	1.649176	.9719568	1.70	0.090	-.2558247	3.554176

Note: dy/dx for factor levels is the discrete change from the base level.

Yuqoridagi natija bu faqat Stata tomonidan chiqarilayotgan o'rtacha marjinal effektlarning qiymatidir. AGE, POOL va FPLACE uchun buni hisoblash mushkul emas. AGE uchun bu AGE ning PRICE ga marjinal ta'siridir, taxminiy koeffitsiyentidir. POOL va FPLACE indikatorli o'zgaruvchilar bo'lganligi uchun ularning marjinal effekti hosil bo'lmaydi, balki birinchi navbatda hovuzga ega bo'lish yoki yo'qligi va ikkinchidan kaminga ega bo'lish yoki yo'qligi o'rtasidagi PRICening diskret o'zgarishidir. Shunday bo'lishiga qaramasdan 1.utown va sqft uchun unchalik oddiy emas.

Ikki qo'shnilar o'rtasidagi kutilayotgan taklif etilayotgan PRICE farqlari quyidagicha:

$$(E(PRICE)|UTOWN=1) - (E(PRICE)|UTOWN=0) = \delta_1 + \gamma SQFT$$

Statada quyidagicha hisoblaydi:

$$AME(UTOWN) = 1/N \sum_{i=1}^N (\delta_1 + \gamma SQFT_i) = \delta_1 + (\gamma SQFT)^-$$

Hisoblangan marjinal effektning dispersiyasi quyidagicha:

$$\begin{aligned} \text{var}((AME)^-(UTOWN)) &= \text{var}(\delta_1^- + \gamma^-(SQFT)) = \text{var}(\delta_1^-) + (\gamma^-(SQFT))^2 \\ &+ 2 \text{var}(\gamma) + 2 SQFT \text{cov}(\delta_1^-, \gamma) \end{aligned}$$

Bu to'g'riligini ko'rish uchun quyidagicha ishlarni amlaga oshiramiz:

```
quietly summarize sqft
```

```
scalar asqft = r(mean)
```

```
lincom 1.utown+c.sqft#1.utown*asqft
```

Natija esa:

```
. quietly summarize sqft
. scalar asqft = r(mean)
. lincom 1.utown+c.sqft#1.utown*asqft
( 1) 1.utown + 25.20965*1.utown#c.sqft = 0
```

	price	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
(1)		60.21049	.9646176	62.42	0.000	58.31757	62.10342

Xuddi shunday SQFT ning marjinal effektining ta'siri:

$$\frac{\partial E(PRICE)}{\partial SQFT} = \beta_2 + \gamma UTOWN$$

Stata quyidagicha hisoblaydi:

$$AME(SQFT) = \frac{1}{N} \sum_{i=1}^N \beta_2 + \gamma UTOWN_i = \beta_2 + \gamma \overline{UTOWN}$$

Hisoblash uchun quyidagi amallarni bajaring:

quietly summarize utown

scalar autown = r(mean)

lincom c.sqft+c.sqft#1.utown*autown

```
. quietly summarize utown
. scalar autown = r(mean)
. lincom c.sqft+c.sqft#1.utown*autown
( 1) sqft + .519*1.utown#c.sqft = 0
```

	price	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
(1)		8.286568	.1661803	49.86	0.000	7.960463	8.612673

log close

7.2. O'zgaruvchilar indikatorini qo'llash

Ushbu bo'limda biz indikator o'zgaruvchilarning turli xil iloalarini ko'rib chiqamiz. cps4_small.dta data faylini oching. Yangi jurnalni ishga tushiring va ma'lumotlarni tekshiring.

log using chap07_cps4, replace text
use cps4_small, clear
describe

```
Contains data from cps4_small.dta
obs:      1,000
vars:      12
size:      23,000 (99.9% of memory free)
```

variable name	storage type	display format	value label	variable label
wage	double	%10.0g		earnings per hour
educ	byte	%8.0g		years of education
exper	byte	%8.0g		post education years experience
hrswk	byte	%8.0g		usual hours worked per week
married	byte	%8.0g		= 1 if married
female	byte	%8.0g		= 1 if female
metro	byte	%8.0g		= 1 if lives in metropolitan area
midwest	byte	%8.0g		= 1 if lives in midwest
south	byte	%8.0g		= 1 if lives in south
west	byte	%8.0g		= 1 if lives in west
black	byte	%8.0g		= 1 if black
asian	byte	%8.0g		= 1 if asian

summarize

Natijada:

```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
wage	1000	20.61566	12.83472	1.97	76.39
educ	1000	13.799	2.711079	0	21
exper	1000	26.508	12.85446	2	65
hrswk	1000	39.952	10.3353	0	90
married	1000	.581	.4936423	0	1
female	1000	.514	.5000541	0	1
metro	1000	.78	.4144536	0	1
midwest	1000	.24	.4272968	0	1
south	1000	.296	.4567194	0	1
west	1000	.24	.4272968	0	1
black	1000	.112	.3155243	0	1
asian	1000	.043	.2029586	0	1

7.2.1. Sifat omillarining o'zaro ta'siri

Birinchidan, biz modeldagi ikkita ko'rsatkich o'zgaruvchisi, qora va ayol o'rtasidagi o'zaro ta'sirni ko'rib chiqamiz:

$$WAGE = \beta_1 + \beta_2 EDUC + \delta_1 EDUC + \delta_2 BLACK + \delta_3 FEMALE + \gamma(BLACK \times FEMALE) + e$$

Faktorli o'zgaruvchilar operatoriga "##" yordamida regressiya hisoblanadi.

reg wage educ i.black##i.female

```
. reg wage educ i.black##i.female
```

Source	SS	df	MS	Number of obs =	1000
Model	34370.7606	4	8592.69016	F(4, 995) =	65.67
Residual	130194.667	995	130.848912	Prob > F =	0.0000
				R-squared =	0.2089
				Adj R-squared =	0.2057
				Root MSE =	11.439
Total	164565.428	999	164.730158		

wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
educ	2.070391	.1348781	15.35	0.000	1.805712 2.335069
1.black	-4.169077	1.774714	-2.35	0.019	-7.651689 -.6864656
1.female	-4.784607	.7734139	-6.19	0.000	-6.302317 -3.266898
black#female					
1 1	3.844294	2.327653	1.65	0.099	-.7233779 8.411966
_cons	-5.281159	1.900468	-2.78	0.006	-9.010544 -1.551774

Biz oq tanli erkaklar va qora tanli ayollar o'rtasidagi ish haqi farqini hisoblaymiz:

lincom 1.black + 1.female + 1.black#1.female

Natijada:

```
. lincom 1.black + 1.female + 1.black#1.female
```

```
( 1) 1.black + 1.female + 1.black#1.female = 0
```

wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
(1)	-5.10939	1.510567	-3.38	0.001	-8.073652 -2.145128

FEMALE va BLACK orasidagi o'zaro ta'sirni tekshirish uchun F test o'kazing:

test 1.female 1.black 1.black#1.female

Natija quyidagicha:

```
. test 1.female 1.black 1.black#1.female
```

```
( 1) 1.female = 0
```

```
( 2) 1.black = 0
```

```
( 3) 1.black#1.female = 0
```

```
F( 3, 995) = 14.21
```

```
Prob > F = 0.0000
```

Marjinal effektlarni **margins** buyrig'i yordamida hisoblanadi.

margins, dydx(*)

Natija:

```
. margins, dydx(*)
Average marginal effects          Number of obs   =       1000
Model VCE      : OLS
Expression     : Linear prediction, predict()
dy/dx w.r.t.   : educ 1.black 1.female
```

	dy/dx	Delta-method Std. Err.	z	P> z	[95% Conf. Interval]	
educ	2.070391	.1348781	15.35	0.000	1.806034	2.334747
1.black	-2.19311	1.160919	-1.89	0.059	-4.468469	.0822488
1.female	-4.354046	.7313539	-5.95	0.000	-5.787474	-2.920619

Note: dy/dx for factor levels is the discrete change from the base level.

Eslatib o'tamiz, model quyidagicha:

$$WAGE = \beta_1 + \beta_2 EDUC + \beta_1 BLACK + \beta_1 FEMALE + \gamma (BLACK \times FEMALE) + e$$

Ya'ni:

$$E(WAGE|FEMALE=1) = \beta_1 + \beta_2 EDUC + \beta_1 BLACK + \beta_2 + \gamma BLACK$$

$$= (\beta_1 + \beta_2) + \beta_2 EDUC + (\beta_1 + \gamma) BLACK$$

va $E(WAGE|FEMALE=1) = \beta_1 + \beta_2 EDUC + \beta_1 BLACK$. Farqi esa quyidagicha:

$$E(WAGE|FEMALE=1) - E(WAGE|FEMALE=0) = \beta_2 + \gamma BLACK$$

Statada quyidagi natija topiladi:

$$AME(FEMALE) = \frac{1}{N} \sum_{i=1}^N (\hat{\delta}_2 + \hat{\gamma} BLACK_i) = \hat{\delta}_2 + \hat{\gamma} \overline{BLACK}$$

Ayolning o'rtacha marjinal ta'sirini hisoblash quyidagicha tekshiriladi:

quietly summarize black

scalar ablack = r(mean)

lincom 1.female + 1.black#1.female*ablack

```
. lincom 1.female + 1.black#1.female*ablack
```

```
(1) 1.female + .112*1.black#1.female = 0
```

wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
(1)	-4.354046	.7313539	-5.95	0.000	-5.789219	-2.918873

7.2.2. Mintaqaviy o'zgaruvchilarni qo'shish

Keyin bir nechta toifali ko'rsatkich o'zgaruvchilari, mintaqaviy o'zgaruvchilar qo'shing. Modeli quyidagicha:

$$WAGE = \beta_1 + \beta_2 EDUC + \beta_3 BLACK + \beta_4 FEMALE + \beta_5 (BLACK \times FEMALE) + \delta_1 SOUTH + \delta_2 MIDWEST + \delta_3 WEST + e$$

Mintaqaviy indikator o'zgaruvchilari ma'lumotlar faylida allaqachon aniqlanganligi sababli, biz ularni regressiya modeliga shunchaki qo'shamiz:

reg wage educ i.black##i.female i.south i.midwest i.west
Natijaning bir qismi:

wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	2.071231	.1344687	15.40	0.000	1.807355	2.335106
1.black	-3.905465	1.786258	-2.19	0.029	-7.410743	- .4001873
1.female	-4.744129	.7698381	-6.16	0.000	-6.254827	-3.233431
black#female						
1 1	3.625021	2.318375	1.56	0.118	-.9244618	8.174504
1.south	-.4499056	1.025024	-0.44	0.661	-2.46137	1.561558
1.midwest	-2.608406	1.059644	-2.46	0.014	-4.687807	-.5290049
1.west	.9866332	1.059815	0.93	0.352	-1.093104	3.06637
_cons	-4.80621	2.028691	-2.37	0.018	-8.787229	-.8251912

Ushbu mintaqaviy o'zgaruvchining umumiy ahamiyatini tekshirish uchun **test** buyrug'idan foydalanimiz:

test 1.south 1.midwest 1.west

. test 1.south 1.midwest 1.west

```
( 1) 1.south = 0
( 2) 1.midwest = 0
( 3) 1.west = 0
```

```
F( 3, 992) = 4.25
Prob > F = 0.0054
```

3 ta numerator va 992 bilan F statistikadan kritik qiymatlarni hisoblang va maxraj erkinlik darajalarini ko'rsating:

di "F(3,992,.95) = " invFtail(3,992,.05)

di "F(3,992,.90) = " invFtail(3,992,.10)

Natija:

```
. di "F(3,992,.95) = " invFtail(3,992,.05)
F(3,992,.95) = 2.6138755
```

```
. di "F(3,992,.90) = " invFtail(3,992,.10)
F(3,992,.90) = 2.0893205
```

7.2.3. Ikki regressiya uchun tenglik testini o'tkazish

Janubiy mintaqa uchun ish haqi tenglamalarining qolgan qismiga nisbatan ekvivalentligini tekshirish mamlakat (ya'ni, janubiy emas) biz regressiya modelidagi har bir o'zgaruvchi uchun janubiy ko'rsatkich o'zgaruvchisi bilan o'zaro ta'sir o'zgaruvchisini yaratamiz. Biz sinab ko'rmoqchi bo'lgan tenglama quyidagicha:

$$WAGE = \beta_1 + \beta_2 EDUC + \beta_1 BLACK + \beta_2 FEMALE + \gamma(BLACK \times FEMALE) + e$$

Indikator o'zgaruvchan o'zaro ta'sirga ega model quyidagicha ko'rinishda:

$$WAGE = \beta_1 + \beta_2 EDUC + \beta_1 BLACK + \beta_2 FEMALE + \gamma(BLACK \times FEMALE) + \theta_1 SOUTH + \theta_2 (EDUC \times SOUTH) + \theta_3 (BLACK \times SOUTH) + \theta_4 (FEMALE \times SOUTH) + \theta_5 (BLACK \times FEMALE \times SOUTH) + e$$

Biz SOUTH bilan regressiya modelidagi har bir o'zgaruvchi bilan, kesishuvchi o'zaro ta'sir bo'yicha ish olib bordik. ## Stata operatoridan foydalanib, barcha o'zaro ta'sir etuvchilarni yarating. Avvalo, BLACK, FEMALE va ularning o'zaro ta'sir etuvchilarini yaratishda biz i.black##i.female. dan foydalanamiz. So'ng to'liq o'zaro ta'sir modelini yaratamiz:

reg wage i.south##(c.educ i.black##i.female)

Natija esa:

. reg wage i.south##(c.educ i.black##i.female)					
Source	SS	df	MS		
Model	34581.0189	9	3842.33543		
Residual	129984.409	990	131.297383		
Total	164565.428	999	164.730158		
				Number of obs =	1000
				F(9, 990) =	29.26
				Prob > F =	0.0000
				R-squared =	0.2101
				Adj R-squared =	0.2030
				Root MSE =	11.459
wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
1.south	3.94391	4.048453	0.97	0.330	-4.000625 11.88845
educ	2.172554	.1664639	13.05	0.000	1.845891 2.499216
1.black	-5.08936	2.64306	-1.93	0.054	-10.276 .0972837
1.female	-5.005078	.8990074	-5.57	0.000	-6.769257 -3.240899
black#female					
1 1	5.305574	3.497267	1.52	0.130	-1.557333 12.16848
south#c.educ					
1	-.308541	.2857343	-1.08	0.280	-.8692554 .2521734
south#black					
1 1	1.704396	3.633327	0.47	0.639	-5.42551 8.834302
south#female					
1 1	.9011198	1.772665	0.51	0.611	-2.577492 4.379732
south#black#female					
1 1 1	-2.935834	4.787647	-0.61	0.540	-12.33094 6.459268
_cons	-6.605572	2.336628	-2.83	0.005	-11.19088 -2.02026

Janubiy va qolgan qismi uchun model o'rtasida "hech qanday farq" yo'qligi haqidagi gipotezani sinab ko'rish uchun mamlakat $H_0: \theta_1 = \theta_2 = \theta_3 = \theta_4 = \theta_5 = 0$ qo'shma gipotezasini test bayonoti yordamida tekshiramiz.

**test 1.south 1.south#c.educ 1.south#1.black 1.south#1.female
/// 1.south#1.black#1.female**

Natija:

```
. test 1.south 1.south#c.educ 1.south#1.black 1.south#1.female ///
> 1.south#1.black#1.female

( 1) 1.south = 0
( 2) 1.south#c.educ = 0
( 3) 1.south#1.black = 0
( 4) 1.south#1.female = 0
( 5) 1.south#1.black#1.female = 0

F( 5, 990) = 0.32
Prob > F = 0.9009
```

To'liq o'zaro ta'sirlar modelidan biz QORA va JANUBIY o'zgaruvchilar yordamida birgalikdagi effektini olishimiz mumkin.

lincom 1.black + 1.black#1.south

Natija quyidagicha:

```
. lincom 1.black + 1.black#1.south
( 1) 1.black + 1.south#1.black = 0
```

	wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
(1)		-3.384964	2.49305	-1.36	0.175	-8.277233 1.507305

Xuddi shunday, AYOL va SOUTHning birgalikdagi ta'siri yordamida:

lincom 1.female + 1.female#1.south

Natija:

```
. lincom 1.female + 1.female#1.south
( 1) 1.female + 1.south#1.female = 0
```

	wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
(1)		-4.103958	1.527785	-2.69	0.007	-7.102027 -1.105889

7.2.4. Alohida regressiyalarni baholash

Oldingi bo'limda to'liq o'zaro ta'sirlangan model yondashuvidan foydalanish o'rniga, F statistik bo'lishi mumkin, kvadrat qoldiqlarning

cheklangan va cheklanmagan yig'indisi yordamida hisoblangan. Kvadrat yig'indisi to'liq model uchun qoldiqlar ikkita alohida regressiyadan olingan SSE yig'indisidir:

$$SSE_{full} = SSE_{non-south} + SSE_{south} = 89088.5 + 40895.9 = 129984.4$$

Ikki mintaqaga bo'yicha hisob-kitoblar standart regress yordamida `bysort` buyrug'i bilan samarali amalga oshirilishi mumkin, bu Stata buyruqlarini ma'lumotlarning kichik to'plamlarida takrorlash imkonini beradi. `help bysort` ni kiring. `By` prefiksdan foydalanish uchun ma'lumotlarni guruhlash o'zgaruvchisi yoki o'zgaruvchilari bo'yicha tartiblash kerak. Agar bu hali bajarmagan bo'lsa, `by` va `sort` buyruqlarini `bysort` shaklida kiring. Sintaksis buyruqlari quyidagicha:

by varlist: stata_cmd

bysort varlist: stata_cmd

Ikki mintaqaga uchun regressiyalarni amalga oshirish uchun bizda mavjud:

bysort south: reg wage educ i.black##i.female

Ya'ni, biz avval `SOUTH` qiymatlari bo'yicha tartiblaymiz va keyin har bir kuzatish guruhi uchun regressiyani amalga oshiramiz.

```
. bysort south: reg wage educ i.black##i.female
```

-> south = 0						
Source	SS	df	MS		Number of obs = 704	
Model	25346.0083	4	6336.50209		F(4, 699) = 49.72	
Residual	89088.4615	699	127.451304		Prob > F = 0.0000	
					R-squared = 0.2215	
					Adj R-squared = 0.2170	
Total	114434.47	703	162.780185		Root MSE = 11.289	
	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wage						
educ	2.172554	.1640077	13.25	0.000	1.850547	2.49456
1.black	-5.08936	2.604061	-1.95	0.051	-10.20208	.0233585
1.female	-5.005078	.8857423	-5.65	0.000	-6.744112	-3.266044
black#female						
1 1	5.305574	3.445664	1.54	0.124	-1.459516	12.07066
_cons	-6.605572	2.30215	-2.87	0.004	-11.12553	-2.085615

-> south = 1						
Source	SS	df	MS		Number of obs = 296	
Model	9234.26014	4	2308.56503		F(4, 291) = 16.43	
Residual	40895.9474	291	140.535902		Prob > F = 0.0000	
					R-squared = 0.1842	
					Adj R-squared = 0.1730	
Total	50130.2075	295	169.932907		Root MSE = 11.855	
	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wage						
educ	1.864013	.2402682	7.76	0.000	1.391129	2.336896
1.black	-3.384964	2.579268	-1.31	0.190	-8.46135	1.691422
1.female	-4.103958	1.580621	-2.60	0.010	-7.214857	-.993059
black#female						
1 1	2.36974	3.382739	0.70	0.484	-4.287995	9.027476
_cons	-2.661662	3.420413	-0.78	0.437	-9.393547	4.070223

Ikki SSE alohida regressiyalardan dispersiya jadvalining tahlilidan kelib chiqadi.

7.2.5. Log chiziqli modellariga o'zgaruvchilar indikatorining qo'llanilishi

Log chiziqli modeldagi indikator o'zgaruvchisining aniq ta'sirini hisoblash murakkab ko'rinadi, ammo Stataning `nlcom` buyrug'i buni ancha osonlashtiradi.

ln(wage) tuzing va quyidagi tenglamani hisoblang:

$$\ln(WAGE) = \beta_1 + \beta_2 EDUC + \delta FEMALE$$

gen lwage = ln(wage)

reglwageeduci.female

Natija quyidagicha:

`. reg lwage educ i.female`

Source	SS	df	MS			
Model	74.5420655	2	37.2710328		Number of obs =	1000
Residual	262.238666	997	.263027749		F(2, 997) =	141.70
Total	336.780731	999	.337117849		Prob > F =	0.0000
					R-squared =	0.2213
					Adj R-squared =	0.2198
					Root MSE =	.51286

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
educ	.0962484	.0060365	15.94	0.000	.0844026 .1080942
1.female	-.243214	.0327275	-7.43	0.000	-.3074367 -.1789913
_cons	1.653868	.0843786	19.60	0.000	1.488288 1.819448

Ko'rsatkichning aniq ta'siri va indikatorli o'zgaruvchisi:

$$100(e^{\delta}-1)\%$$

Bu `nlcom`.dan foydalanishni talab qiluvchi parametrlarning chiziqli bo'lmagan funksiyasi:

$$\text{nlcom.} 100 * (\exp_b[1.female]) - 1) \quad 100 * (\exp_b[1.female]) - 1)$$

Buning natijasi:

$$(_b[1.female]) - 1)$$

$$100 * (\exp_b[1.female]) - 1)$$

Ya'ni, qolgan hamma narsani o'zgarmagan holda, biz ayol ishchilar erkaklarnikiga qaraganda 21,6% kamroq maosh olishlarini taxmin qilamiz.

Xuddi shunday, biz boshqa chiziqli bo'lmagan marjinal effektlarni hisoblashimiz mumkin. Modelni ko'rib chiqamiz.

$$\ln(\text{WAGE}) = \beta_1 + \beta_2 \text{EDUC} + \beta_3 \text{EXPER} + \gamma(\text{EDUC} * \text{EXPER})$$

Taxminiy marjinal ta'sir

$$100(\beta_3 + \gamma \text{EDUC}) \%$$

Ta'lim va tajribaning o'zaro ta'sirini yarating va uni regressiya modeliga qo'shing:

reg lwage c.educ##c.exper

Baholash natijalari:

. reg lwage c.educ##c.exper

Source	SS	df	MS
Model	65.5449479	3	21.848316
Residual	271.235783	996	.272325084
Total	336.780731	999	.337117849

Number of obs = 1000
F(3, 996) = 80.23
Prob > F = 0.0000
R-squared = 0.1946
Adj R-squared = 0.1922
Root MSE = .52185

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
educ	.0949385	.0146246	6.49	0.000	.06624 .123637
exper	.0063295	.0066985	0.94	0.345	-.0068153 .0194743
c.educ#c.exper	-.0000364	.0004838	-0.08	0.940	-.0009858 .0009129
_cons	1.392318	.2066447	6.74	0.000	.986809 1.797827

Lincom yoki **nlcom**dan foydalanish uchun **_b** [o'zgaruvchi] ko'rinishidagi koeffitsiyent nomlari ba'zan faktor belgilaridan foydalanganda darhol aniqlanmaydi. Buni koeffitsiyentlar jadvalini emas, balki **coeflegend** afsonasini ko'rsatadigan variant koeflegendasi bilan **regress** buyrug'ini ko'rsatish orqali aniqlash mumkin.

reg, coeflegend

. reg, coeflegend

Source	SS	df	MS
Model	65.5449479	3	21.848316
Residual	271.235783	996	.272325084
Total	336.780731	999	.337117849

Number of obs = 1000
F(3, 996) = 80.23
Prob > F = 0.0000
R-squared = 0.1946
Adj R-squared = 0.1922
Root MSE = .52185

lwage	Coef.	Legend
educ	.0949385	_b[educ]
exper	.0063295	_b[exper]
c.educ#c.exper	-.0000364	_b[c.educ#c.exper]
_cons	1.392318	_b[_cons]

Taxminiy va aniq effektlar yordamida hisoblab chiqiladi

lincom 100*(exper+ c.educ#c.exper*16)

```
. lincom 100*(exper+ c.educ#c.exper*16)
( 1) 100*exper + 1600*c.educ#c.exper = 0
```

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
(1)	.574639	.174402	3.29	0.001	.2324014	.9168765

nlcom 100*(exp(_b[exper]+_b[c.educ#c.exper]*16) - 1)

```
. nlcom 100*(exp(_b[exper]+_b[c.educ#c.exper]*16) - 1)
      _nl_1: 100*(exp(_b[exper]+_b[c.educ#c.exper]*16) - 1)
```

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
_nl_1	.5762932	.1754071	3.29	0.001	.2320833	.920503

log close

7.3. Chiziqli empirik model

Ikki muqobil o'rtasida tanlovni modellashirishda indikator o'zgaruvchisi bog'liq bo'ladi, regressiya modelida mustaqil o'zgaruvchi emas, balki erksiz o'zgaruvchi. Faraz qilaylik:

$$Y = \begin{cases} 1 & \text{birinchi muqobil tanlansa} \\ 0 & \text{agar ikkinchi variant tanlansa} \end{cases}$$

Agar p birinchi muqobil tanlanish ehtimoli bo'lsa, u holda $P[y=1]=p$ bo'lsa, y ning kutilayotgan qiymati $E(y) = p$, dispersiyasi esa $var(y)=p(p-1)$ bo'ladi.

Biz chiziqli regressiya funksiyasidan yoki shu nuqtayi nazardan chiziqli ehtimollik modelidan foydalangan holda p ehtimolligiga ta'sir etuvchi omillarni aniqlashdan manfaatdormiz:

$$E(y) = p = \beta_1 + \beta_2 x_2 + \dots + \beta_K x_K$$

Chiziqli ehtimollik regressiya modeli:

$$y = \beta_1 + \beta_2 x_2 + \dots + \beta_K x_K + e$$

Xato atamasining dispersiyasi e ga teng:

$$var(e) = (\beta_1 + \beta_2 x_2 + \dots + \beta_K x_K)(1 - \beta_1 - \beta_2 x_2 - \dots - \beta_K x_K)$$

Bu xato homosedastik emas va buni VIII bobda to'liq ko'rib chiqamiz.

Misol sifatida Cola va Pepsi o'rtasidagi tanlovni ko'rib chiqing. Coke.dta ni oching va uning mazmunini tekshiring:

log using chap07_coke, replace text
use coke, clear
describe
summarize

variable name	storage type	display format	value label	variable label
coke	byte	%8.0g		=1 if coke chosen, =0 if pepsi chosen
pr_pepsi	double	%10.0g		price of 2 liter bottle of pepsi
pr_coke	double	%10.0g		price of 2 liter bottle of coke
disp_pepsi	byte	%8.0g		= 1 if pepsi is displayed at time of purchase, otherwise = 0
disp_coke	byte	%8.0g		= 1 if coke is displayed at time of purchase, otherwise = 0
pratio	double	%10.0g		price of coke relative to price of pepsi

Eng kichik kvadratlar regressiyasidan foydalanib, cola tanlash uchun chiziqli ehtimollik modelini taxmin qilamiz.

reg coke pratio disp_coke disp_pepsi

. reg coke pratio disp_coke disp_pepsi

Source	SS	df	MS		Number of obs =
Model	33.8378078	3	11.2792693		1140
Residual	248.004297	1136	.218313642		F(3, 1136) = 51.67
Total	281.842105	1139	.247446976		Prob > F = 0.0000
					R-squared = 0.1201
					Adj R-squared = 0.1177
					Root MSE = .46724

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
pratio	-.4008614	.0613494	-6.53	0.000	-.5212324 -.2804904
disp_coke	-.0771745	.0343919	2.24	0.025	-.0096956 .1446533
disp_pepsi	-.1656637	.0355997	-4.65	0.000	-.2355122 -.0958152
_cons	.8902151	.0654849	13.59	0.000	.7617301 1.0187

Chiziqli regressiya yondashuvi bilan bog'liq tashvish shundaki, bashorat qilingan ehtimolliklar [0, 1] oraliqdan tashqarida bo'lishi mumkin. Bu holda ehtimollik bo'lgan bashorat qilingan qiymatlarni oling va umumlashtiring:

predict phat
summarize phat

. summarize phat

Variable	Obs	Mean	Std. Dev.	Min	Max
phat	1140	.4473684	.1723611	-.2073211	.7680784

Biz minimal qiymat noldan kichik ekanligini ko'ramiz, lekin bashorat qilingan ehtimollarning hech biri birdan katta emas.

Qancha bashorat qilingan ehtimollar salbiy ekanligini ko'rish uchun ulardan foydalanib, ularni umumlashtiring:

summarize phat if phat<=0

```
. summarize phat if phat<=0
```

Variable	Obs	Mean	Std. Dev.	Min	Max
phat	16	-.0183585	.0523201	-.2073211	-.0002385

1140 ta kuzatuvdan 16 tasi salbiy bashorat qilingan ehtimollarga ega ekanligini ko'ramiz: **log close**

Yettinchi bob mavzularini mustahkamlash uchun savollar:

1. O'zgaruvchilar indikatori deganda nimani tushunasiz?
2. O'zgaruvchilarning qiymatlariga asoslangan holda yangi indikatorli o'zgaruvchilarni hosil qilish uchun qanday buyruqdan foydalaniladi?
3. Yakuniy satatistik xulosaning har bir detallarini hisoblash uchun qaysi buyruqlardan foydalanish lozim?
4. O'zgaruvchili indikator koffitsiyenti gipotezalarini tekshirish uchun qanday buyruqlarni bajarish talab etiladi?
5. O'rtacha marjinal ta'sirni hisoblash buyrug'ini ifodalang.
6. Indikator o'zgaruvchilarini ifodalash qanday fayl orqali amalga oshiriladi?
7. Mintaqaviy indikator o'zgaruvchilari ma'lumotlarini regressiya modeliga kiritish uchun qanday amallar bajarilishi lozim?
8. Ikki mintaqa bo'yicha hisob-kitoblar qaysi buyruq yordamida amalga oshiriladi?
9. Chiziqli empirik model tuzish qanday amallarni bajarishni talab etadi?
10. Chiziqli ehtimollik regressiya modelini ifodalang.

Yettinchi bob mavzularida qo'llanilgan buyruqlar va atamalar

!missing	esttab factor	margins, dydx
areg	variables	missing value
areg varlist,	fixed effects	nlcom
absorb(var)	F-test	not equal to !=
average marginal effect	F-test critical value	pairwise correlations
by	generate	pwcorr
bysort	global	quietly
Chow test	if	regional indicator variables
delta method	indicator variables	regress, coeflegend
DID	interaction variables	tabulate
difference estimator	invFtail	tabul varname, gen()
differences-in-differences	lincom	test
dummy variables	linear probability model	treatment effects
estimates store	log-linear model	
estimates table	marginal effect	
	margins	

Yettinchi bob mavzularini mustahkamlash uchun masalalarning do-file lari

Ma'lumotlar buyrug'i: `cd c:\data\poe4stata`

1-masala. log file ni ochish va datalardan foydalanish.

`log using chap07_utowm, replace text`

`use utowm, clear`

`* summarize and examine`

`describe`

`summarize`

`list in 1/6`

`list in 501/506`

`* examples creating indicator variables`

`summarize price sqft, detail`

`gen large = (sqft > 25)`

`gen midprice = (215 < price) & (price < 275)`

```

list sqft price large midprice in 1/5
* estimate dummy variable regression
reg price i.utown sqft i.utown#c.sqft age i.pool i.fplace
reg price i.utown##c.sqft age i.pool i.fplace
* test significance of utown
test 1.utown 1.utown#c.sqft
* use lincom for utown slope and intercept
lincom _cons + 1.utown
lincom c.sqft + 1.utown#c.sqft
* ame
margins, dydx(*)
* ame for utown
quietly summarize sqft
scalar asqft = r(mean)
lincom 1.utown+c.sqft#1.utown*asqft
* ame for sqft
quietly summarize utown
scalar autown = r(mean)
lincom c.sqft+c.sqft#1.utown*autown
matrix list e(b)
matrix list e(V)
matrix vbols = e(V)
* -----
* for utown
* -----
* extract variances and covariance
scalar vb2=vbols[2,2]
scalar vb5=vbols[5,5]
scalar cov52 = vbols[5,2]
* mean of _cons and sqft
quietly summarize sqft
scalar asqft = r(mean)
scalar aconst = 1
* delta method for ame of utown
scalar vame=(aconst^2)*vb2+(asqft^2)*vb5+2*asqft*aconst*cov52
scalar seame = sqrt(vame)

```

```

di "Delta-method standard error for utown " seame
*-----
* for sqft
*-----
* delta method se for sqft ame
quietly summarize utown
scalar autown = r(mean)
scalar vb3=vbols[3,3]
scalar cov53 = vbols[5,3]
* delta method
scalar vame=(aconst^2)*vb3+(autown^2)*vb5+2*autown*aco
nst*cov53
scalar seame = sqrt(vame)
di "Delta-method standard error for sqft " seame
log close
2-masala. Applying indicator variables
* open new log
log using chap07_cps4, replace text
* open data
use cps4_small, clear
describe
summarize
* estimate model with black-female interaction
reg wage educ i.black##i.female
* estimate wage difference between black-female and white-male
lincom 1.black + 1.female + 1.black#1.female
* F-test of joint significance
test 1.female 1.black 1.black#1.female
* Average marginal effects
margins, dydx(*)
quietly summarize black
scalar ablack = r(mean)
lincom 1.female + 1.black#1.female*ablack
3-masala. Qo'shimcha indikatorlarni qo'shish.
reg wage educ i.black##i.female i.south i.midwest i.west
test 1.south 1.midwest 1.west

```

```

di "F(3,992,.95) = " invFtail(3,992,.05)
di "F(3,992,.90) = " invFtail(3,992,.10)
reg wage i.south##(c.educ i.black##i.female)
test 1.south 1.south#c.educ 1.south#1.black 1.south#1.female
1.south#1.black#1.female
* constructing estimates in separate regressions from fully
interacted model
lincom 1.black + 1.black#1.south
lincom 1.female + 1.female#1.south
* Estimate separate regressions
bysort south: reg wage educ i.black##i.female
4-masala. Log-linear modellarini tuzish.
gen lwage = ln(wage)
* estimate regression
reg lwage educ i.female
* use nlcom to obtain exact effect of dummy variable
nlcom 100*(exp(_b[1.female]) - 1)
* using nlcom with interaction variables
reg lwage c.educ##c.exper
reg, coeflegend
lincom 100*(exper+ c.educ#c.exper*16)
nlcom 100*(exp( _b[exper]+_b[c.educ#c.exper]*16) - 1)
log close
5-masala. Chiziqli empirik model tuzish.
* open new log
log using chap07_coke, replace text
* open data and examine
use coke, clear
describe
summarize
* estimate regression
reg coke pratio disp_coke disp_pepsi
predict phat
summarize phat
summarize phat if phat<=0
log close

```

VIII BOB. GETEROSKIDASTLIK

8.1. Geteroskidastlik ahamiyati

Uchinchi va to'rtinchi boblardagi oddiy chiziqli regressiya modellari va ko'p regressiya modeli V va VI boblarni bir necha jihatdan umumlashtirish mumkin. Masalan, ushbu modellarning tasodifiy o'zgaruvchilari (yi yoki ei) 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 mustaqil o'zgaruvchi bo'yicha kuzatish, e_i tasodifiy xato va β_1 hamda β_2 siz hisoblamog'chi 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:

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

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 geteroskidastlik bo'lgan parametrlar o'rganiladi.

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

$MR3, \text{var}(y_i) = \text{var}(e_i) = \sigma^2$ 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
```

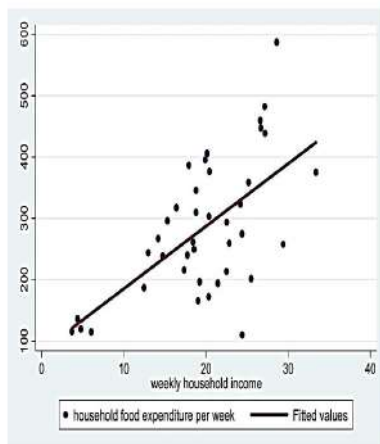
Ishni eng kichik kvadratlar yordamida oziq-ovqat xarajatlari modelini baholashdan boshlang:

```
regress food_exp income
```

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

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

bir xil grafikda ikkita tarqoq chizma hosil qiladi. Har bir sujetni 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-grafikka juda o'xshaydi.

8.2. Geteroskedastiklikni aniqlash

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

8.2.1. Qoldiq maydonlari

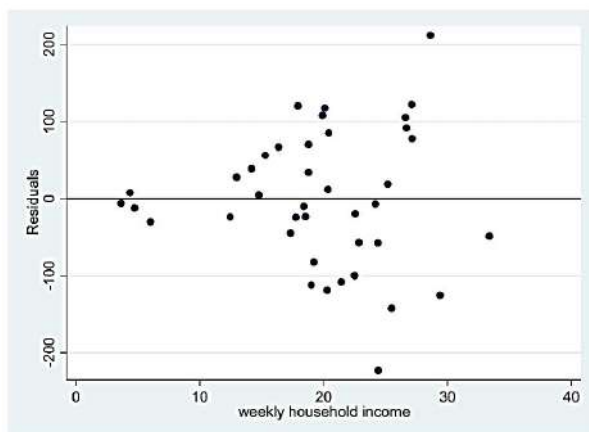
Xatolarning geteroskedastikligini aniqlash usullaridan biri – ularni mustaqil o'zgaruvchining tartiblangan qiymatiga nisbatan chizish. Oldingi bo'limda bir nechta misollar keltirilgan edi. O'zaro bog'liqlikni hisoblashning yana bir usuli – modelni baholash, qoldiqlarni aniqlashdir. Ularni tuzish uchun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

```
regress food_exp income
```

```
predict ehat, res
```

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

quyidagicha grafik hosil bo'ladi:

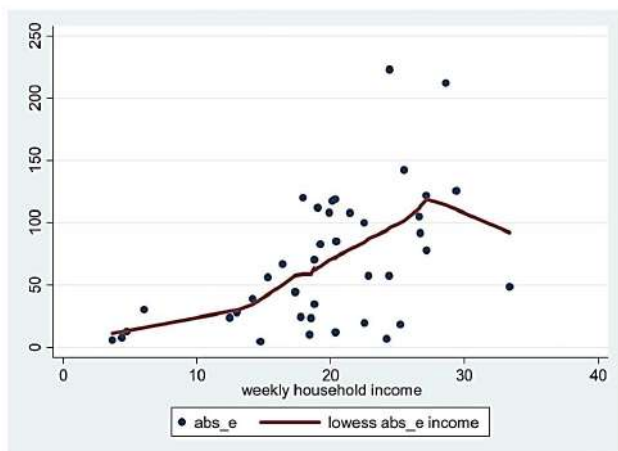


Qizil gorizontaal chiziqni nolga tenglashtirish uchun buyruqqa `yline(0)` birikma qo'shildi. Grafikdan ko'rinib turibdiki, daromadning kattaroq qiymatlari uchun qoldiqlar kattaroqdir. Buni quyidagi testlardan biri yoki bir nechta yordamida statistik tasdiqlash mumkin.

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 `lowfess` 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 yetadi. 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

Boshqajoylarda 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 $\alpha_2, z_{i2}, \dots, \alpha_s$, o'zgaruvchilarga bog'liq bo'lgan tarzda geteroskedastikdir. $h()$ funksiyasi esa belgilanmagan. Bu uning argumentiga, ya'ni z dagi o'zgaruvchilarning chiziqli funksiyasiga bog'liq bo'lgan har qanday narsa bo'lishi mumkin.

Quyidagi qadamlar orqali amalga oshiriladi:

regressiya modelini baholang;

qoldiqlarni hisoblang;

qoldiqlarni kvadratga aylantiring;

$z_{i2}, z_{i3}, \dots, \alpha_s$ dagi kvadrat qoldiqlarni regressiyaga aylantiring.

Ushbu regressiyadan NR^2 ni hisoblang va uni $x^2 (S-1)$ va taqsimotidagi α darajadagi kritik qiymat bilan solishtiring.

Statada 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 $chi2tailLM$ 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:

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

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. Statada daromadning kvadrat qiymatini yarating va daromad hamda uning kvadrati bo'yicha modeldan kvadrat qoldiqlarni regressga aylantiring. Ushbu regressiyadan NR^2 ni hisoblang va uni $x^2 (S-1)$ taqsimotidagi α darajasidagi kritik qiymat bilan solishtiring. Ushbu kitobda ko'rib chiqilgan barcha LM testlarda 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))
```

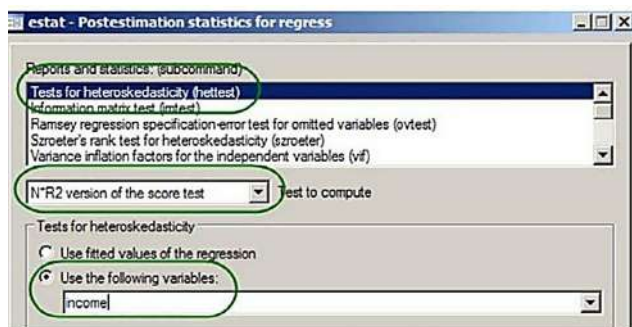
Natija esa quyidagicha:

NR2 = 7.5550786

5% critical value = 5.9914645

P-value = .02287892

Stata ushbu ikkala test statistikasi va ularning p-qiymatlarini hisoblash uchun funksiyalarni 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 (hettest) ni tanlash uchun sichqonchani o'ng tomonidagi tugmadan foydalaning. Keyingi ochiladigan menyuda N*R2 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:

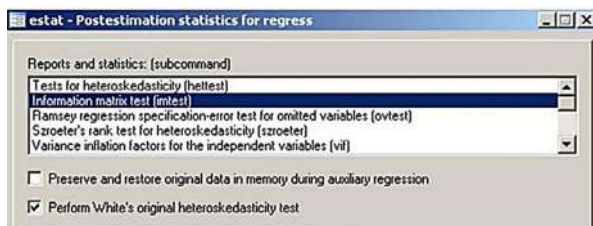
```
. estat hettest income, iid

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 H0: 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 funksiyasi sifatida baholanadigan misoldan foydalanamiz. Bundan tashqari, agar odam metropolitik hududida yashasa, unga teng bo'lgan soxta o'zgaruvchi kiritilgan. Bu "to'siq" o'zgaruvchisi, ya'ni metro hududlarida yashovchi odamlar ta'lim va tajribadagi o'zgarishlarga (bir xil qiymatlarda) 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.

Birinchidan, 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 = Stataning tayinlash operatori. U o'zining o'ng tomonida joylashgan qiymatni chap tomonidagi nomga belgilaydi; x=2 assigns2 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 σ 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_expincome 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 heteroskedastisity 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, koeffitsiyent 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 geteroskedastistiklik 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 mezoni 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 i ning o'zgaruvchan qismini bilsangiz, Statada buni qilish ancha oson.

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

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

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 Statada analytic weights yordamida amalga oshiriladi. Analitik og'irliklar kuzatishning variance sig'a inversely proporsionaldir. Shunday qilib, oziq-ovqat xarajatlari modelini qayta tortish uchun ishlatiladigan sintaksis quyidagicha:

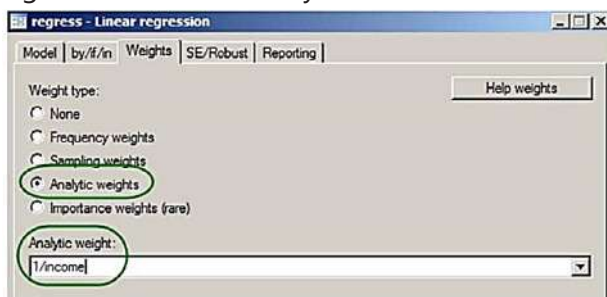
regress food_exp income [aweight=1/income]

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 bosib va quyida ko'rsatilganidek, kerakli analitik vaznni maydonchaga kiriting. Bunday holda, biz income ning o'zaro nisbatidan foydalanamiz.



OK tugmasini bosib 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	
Total	501235.248	39	12852.1858	Prob > F =	0.0000	
				R-squared =	0.5994	
				Adj R-squared =	0.5889	
				Root MSE =	72.688	

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 namunalarini 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
Residual	26284.1489	996	26.3897077
Total	36081.2155	999	36.1173329

```
Number of obs = 1000
F( 3, 996) = 123.75
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 koeffitsiyentlari 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 kvadratlarga (FGLS) aylantiradi.

Birinchi qadam ba'zi mustaqil o'zgaruvchilarning funksiyasi 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 funksional munosabatni belgilashingiz kerak bo'ladi. Dispersiyaning umumiy modeli eksponensial funksiyadan foydalanadi:

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

Buyerda 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:

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

Bu yerda $\hat{\sigma}_i^2$ sizning asl geteroskedastik regressiya modelingizning eng kichik kvadratlar bahosidan olingan. α_1 ni hisoblash uchun $z_i = \ln(\text{income})$ 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, Insig2 regressiyasidan chiziqli taxminlarni oling va eksponensial funksiya exp (Insig2) yordamida og'irliklarni hosil qiling:

```
predict Insig2, xb
gen wt = exp(Insig2)
```

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-bo'limda 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 tanlangan bo'lsa, 1-p ehtimollik bilan nol qiymatini oladi.

i^{th} kuzatish uchun regressiya funksiyasi:

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

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 nechta (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
	0.0613	0.0604	0.0444	0.0527
disp_coke	0.0772	0.0772	0.0940	0.0760
	0.0344	0.0339	0.0399	0.0353
disp_pepsi	-0.1657	-0.1657	-0.1314	-0.1587
	0.0356	0.0344	0.0354	0.0360
_cons	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.

Sakkizinchi 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)` (Ifit `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.

Sakkiznchi bob mavzularida qo'llanilgan buyruqlar va atamalar

<code>analytic weights</code>	<code>groupwise</code>	<code>lowess</code>
<code>aweight</code>	<code>heteroskedastic.</code>	<code>replace</code>
Breusch-Pagan test	HCCME	<code>residual plots</code>
<code>chi2tail(df,stat)</code>	<code>heteroskedasticity</code>	<code>robust standard</code>
<code>drop</code>	<code>imtest, white</code>	<code>errors</code>
<code>e(df_r)</code>	<code>invchi2tail(df,alpha)</code>	<code>subsample</code>
<code>e(rmse)</code>	<code>invFtail(J,N-K,alpha)</code>	<code>twoway</code>
<code>estat hettest</code>	Lagrange multiplier	<code>vce(robust)</code>
<code>Ftail(J,N-K,fstat)</code>	Ifit	Weighted Least
Generalized Least	linear probability	Squares
Squares	model	White's standard
Goldfeld-Quandt	LM test	<code>errors</code>
test <code>graph</code>		White's test
<code>twoway</code>		<code>ylines(0)</code>

VIII bob mavzularini mustahkamlash uchun masalalarning do-file lari

Ma'lumotlar buyrug'i: `cd c:\data\poe4stata`

1-masala. log file ni ochish va datalardan foydalanish.

log using chap08, replace text

* -----

* food expenditure example

* OLS, OLS with White's std errors, GLS

* -----

use food, clear

regress food_exp income

predict ehat, res

graph twoway (scatter food_exp income) (lfit food_exp income,
lw(thick))

* -----

* Graph relationship between size of errors and income

* -----

generate abs_e = abs(ehat)

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

* -----

* Graph relationship between errors and income

* -----

graph twoway scatter ehat income, yline(0)

drop ehat

* -----

* Breusch-Pagan and White tests

* -----

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))

```
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))
quietly regress food_exp income
estat hettest income, iid
estat imtest, white
```

```
* -----
```

2-masala. Goldfeld Quandt testi.

```
* -----
```

```
use cps2, clear
regress wage educ exper metro
```

```
* -----
```

```
* Rural subsample regression
```

```
* -----
```

```
regress wage educ exper if metro == 0
scalar rmse_r = e(rmse)
scalar df_r = e(df_r)
```

```
* -----
```

```
* Urban subsample regression
```

```
* -----
```

```
regress wage educ exper if metro == 1
scalar rmse_m = e(rmse)
scalar df_m = e(df_r)
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
```

```
* -----
```

3-masala. Goldfeld Quandt testini oziq-ovqat xarajatlari ma'lumotlarida foydalanish.

```
* -----
```

```
use food, clear
sort income
regress food_exp income in 1/20
```

```

scalar s_small = e(rmse)^2
scalar df_small = e(df_r)
regress food_exp income in 21/40
scalar s_large = e(rmse)^2
scalar df_large = e(df_r)
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
* -----

```

4-masala. HCCME testi

```

* -----
use food, clear
quietly reg food_exp income
estimates store Usual
scalar bL = _b[income] - invttail(e(df_r),.025) * _se[income]
scalar bU = _b[income] + invttail(e(df_r),.025) * _se[income]
scalar list bL bU
quietly reg food_exp income, vce(robust)
estimates store White
scalar bL = _b[income] - invttail(e(df_r),.025) * _se[income]
scalar bU = _b[income] + invttail(e(df_r),.025) * _se[income]
scalar list bL bU
estimates table Usual White, b(%7.4f) se(%7.3f) stats(F)
reg food_exp income, vce(robust) level(90)
* -----
* GLS
* -----
regress food_exp income [aweight = 1/income]
scalar bL = _b[income] - invttail(e(df_r),.025) * _se[income]
scalar bU = _b[income] + invttail(e(df_r),.025) * _se[income]
scalar list bL bU
* -----
* cps example
* -----

```

```

use cps2, clear
regress wage educ exper
* -----
* Groupwise heteroskedastic regression using FGLS
* -----
gen rural = 1 - metro
gen wt=(rmse_r^2*rural) + (rmse_m^2*metro)
regress wage educ exper metro [aweight = 1/wt]
* -----
* subsample regressions using dummy variables
* for weights
* -----
regress wage educ exper [aweight = rural]
scalar sr = e(rmse)^2
regress wage educ exper [aweight = metro]
scalar sm = e(rmse)^2
scalar df_r = e(df_r)
* -----
* Groupwise heteroskedastic regression using FGLS
* -----
gen wtall=(sr*rural) + (sm*metro)
regress wage educ exper metro [aweight = 1/wtall]
regress wage educ exper metro
predict ehat, residual
twoway (scatter ehat metro)
more
twoway (scatter ehat wage)
more
* -----

```

5-masala. Geteroskedastiklik regressiyada FGLS ni qo'llash.

```

* -----
use food, clear
gen z = ln(income)
reg food_exp income
predict ehat, residual

```

```

gen ln_ehat_sq = ln(ehat^2)
reg ln_ehat_sq z
predict sighat, xb
gen wt = exp(sighat)
regress food_exp income [aweight=(1/wt)]
* -----
* FGLS with Linear Probability Model
* -----

use coke, clear
summarize
* OLS with inconsistent std errors
quietly regress coke pratio disp_coke disp_pepsi
estimates store LS
predict p, xb
gen var = p*(1-p)
summarize p var
predict ehat, res
gen ehat2=ehat^2
* White's test
quietly imtest
scalar NR2 = r(chi2_h)
scalar crit05 = invchi2tail(r(df_h),.05)
scalar pval = chi2tail(r(df_h),r(chi2_h))
scalar list NR2 crit05 pval
* White's test manually
quietly regress ehat2 pratio disp_coke disp_pepsi i.disp_coke#i.
disp_pepsi
i.disp_coke#c.pratio i.disp_pepsi#c.pratio c.pratio#c.pratio
di "NR2 = " e(N)*e(r2)
* OLS with HCCME std errors
quietly reg coke pratio disp_coke disp_pepsi, vce(robust)
estimates store Robust
* OLS, omitting observations with negative variances
quietly reg coke pratio disp_coke disp_pepsi [aweight=1/var] if
p > 0

```



```
estimates store Omit
* OLS, where all  $p < .01$  are truncated to be equal .01
replace p = .01 if p < .01
replace var =  $p*(1-p)$ 
quietly reg coke pratio disp_coke disp_pepsi [aweight=1/var]
estimates store Trunc
estimates table LS Robust Trunc Omit, b(%7.4f) se(%7.4f) stats(F N)
* Test for heteroskedasticity
quietly regress coke pratio disp_coke disp_pepsi
imtest, white
log close
```

IX BOB. DINAMIK QATORLAR REGRESSIYASI

9.1. Asosiy tushunchalar

Dinamikaning regressiya munosabatlariga kirishi mumkin bo'lgan uchta usul ko'rib chiqamiz: tushuntirish o'zgaruvchining kechikish qiymatlari, bog'liq o'zgaruvchining kechikish qiymatlari va ulardagi xatoning kechikish qiymati.

Vaqtli qatorlarning regressiyasida odatiy ekonometrik jarayonlar tegishli statistik xususiyatlarga ega bo'lishi uchun ma'lumotlar statsionar bo'lishi kerak. Bu, asosan, vaqtli qatorlar ma'lumotlarining o'rtacha dispersiyasi va kovariatsiyalari ular kuzatilgan vaqt davriga bog'liq bo'lmashligi talab qiladi. Masalan, 1973-yilning uchinchi choragida YaIM ning o'rtacha va dispersiyasi 2006-yilning 4-choragidan farq qilishi mumkin emas. Ushbu muammoni hal qilish usullari so'nggi yillarda ekonometriklar uchun boy tadqiqot sohasini taqdim etdi va bu usullarning bir nechitasi XII bobda batafsil to'xtalib o'tamiz.

Qo'llaniladigan birinchi tahlil vositalaridan biribuma'lumotlarining oddiy vaqtli qatorlarni grafigidir. Vaqtli qatorlar grafigi ma'lumotlar bilan yuzaga kelishi mumkin bo'lgan muammolarni ochib beradi va statistik jihatdan davom etish yo'llarini taklif qiladi. Oldingi boblarda ko'rib chiqqanimizdek, Statada vaqtli qatorlar grafiklarini yaratish oson va quyida bir nechta yangi grafiklar ko'rib chiqiladi.

Nihoyat, bu bobda vaqtli ketma-ket kuzatuvlar bilan bog'liq bo'lgani sababli kuzatuvlarning odatiy soni, N , ko'proq qo'llaniladigan T bilan almashtiriladi. Keyingi boblarda ham vaqt seriyasi, ham kesma ma'lumotlar qo'llaniladi, N va T dan foydalaniladi.

9.1.1. Statada vaqtli qatorlar grafigini ifodalash

Vaqtli qatorlar ma'lumotlarini tahlil qilish uchun Stata-ning ko'plab o'rnatilgan funksiyalaridan foydalanish uchun, bir to'plamdagi ma'lumotlarni vaqtli qatorlar deb e'lon qilishi kerak. Vaqtli qatorlarni o'z vaqtida tartiblanganligi sababli boshqa kuzatuvlarga nisbatan pozitsiyasi saqlanishi kerak. Axir bu turdagi ma'lumotlarni tahlil qilish, kesma tahlilidan farq qiladigan narsa ularning vaqtinchalik munosabatlaridir.

Agarsizda mavjud bo'lgan ma'lumotlarkuzatuvto'planganvaqttni aniqlash uchun tegishli sanaga ega bo'lmasa, uni qo'shish yaxshi fikrdir. Bu tarixni aniqlash imkonini beradi, davrlarni osonlashtiradi va grafiklarning axborot mazmunini sezilarli darajada oshiradi. Kitobda ifodalangan ma'lumotlar to'plamlari vaqtli qatorlar deb e'lon qilinmagan va ularning aksariyati o'zgaruvchilar to'plamida tegishli sanalarni o'z ichiga olmaydi. Shunday qilib, biznesning birinchi tartibi bu ma'lumotni ma'lumotlar to'plamiga qo'shishdir va keyin kuzatuvlarni aniqlash uchun sanalardan foydalanishdir (masalan kunlik, oylik, choraklik, yillik). Ma'lumotlarning vaqtga bog'liqligini tahlil qilishda, bu juda muhim ma'lumot bo'lib, quyida tushintiriladi.

Matndan aniq misollarga o'tishdan oldin, Stata sana va vaqtlarni qanday boshqarishi haqida qo'shimcha ma'lumot kiritish kerak. Butun son kelishilgan asosdan (ularni qanday belgilashingizdan qat'i nazar) o'tgan vaqt birliklari sonini yozadi, Stata uchun bu 1960.

Masalan, 1961-yilda boshlangan 100 ta choraklik ma'lumotlarni kuzatish uchun biz Stata sanalarini yaratishimiz mumkin:

set obs 100

generate date = tq(1961q1) + _n-1

tq(1961q1) psevdofunksiya deb ataladi. Ular psevdofunksiyalar deb atalishi, ular yozganlaringizni butun son ekvivalentlariga tarjima qiladi. 1961q1 ning butun ekvivalenti 4 ga teng, ya'ni 1960-yil birinchi chorakdan necha chorak o'tdi. Ikkinchi chorak 5 va shunga o'xshash tarzda o'rnatiladi. _n-1 qo'shilishi kuzatuvlarni 1 ga oshirish uchun amalga oshiriladi. Ro'yxatning birinchi 5 ta kuzatuvi sanani ko'rib chiqamiz:

```
. list date in 1/5
```

	date
1.	4
2.	5
3.	6
4.	7
5.	8

Bu biz kutgan narsadir. Buni odamlar uchun mazmunli qilish uchun, ularni amalga oshirish uchun satrlar sifatida formatlash kerak 1960-yildan 20 chorak qaysi sana ekanligini aytish biz uchun oson. Bu format buyrug'i yordamida amalga oshiriladi:

format %tq date

Format buyrug'i butun sanalarni ko'rsatish usulini o'zgartiradi;

```
. list date in 1/5
```

	date
1.	1961q1
2.	1961q2
3.	1961q3
4.	1961q4
5.	1962q1

Ko'rib turganingizdek, format %tq date Stataga 4, 5, 6 va 7-sonlarni ko'rsatishni buyuradi va o'zgaruvchan date 1961q1, 1961q2 va hokazo. Nihoyat, kuzatishlar vaqtli qatorlar deb e'lon qilinadi, vaqt o'zgaruvchisini identifikatsiya qiluvchi o'zgaruvchi nomidan keyin tsset buyrug'i yordamida amalga oshiriladi:

tsset date

```
. tsset date
```

```
time variable: date, 1961q1 to 1985q4  
delta: 1 quarter
```

Ma'lumotlar vaqtli qatorlar deb e'lon qilingandan so'ng,

Stata qamrab olingan davr va o'lchov oralig'i haqidagi muhim ma'lumotlarni chop etadi. U vaqt o'zgaruvchisining nomini, u qamrab olgan sanalarni va delta yoki kuzatishlar orasidagi vaqtni aniqlaydi. Yaratilgan sanalar kerakli narsaga mos kelishiga ishonch hosil qilish uchun har doim sanalarni yaratishda ehtiyotkorlik bilan buni tekshiring. Stata haftalik (tw), oylik (tm) yillik (ty) va boshqalar ni aniqlash uchun boshqa funksiyalar va psevdofunksiyalarni o'z ichiga oladi. Shunga qaramay, ular 1960q1 yildan buyon o'tgan davrlarini ko'rsatadigan butun sonlar to'plamini yaratadi. Butun sonlarni sana sifatida ko'rsatish uchun mos formatlar (%tw), (%tm) va (%ty) mos ravishda. Boshqa variantlarni ko'rish va ularning turi qanday ishlashi haqida ko'proq ma'lumot olish uchun quyidagi buyruqdan foydalanamiz:

help dates and times

Command oynasida va Stata viewer oynasini ochadi hamda sizni tegishli ma'lumotlarga olib boradi. Sanalar yaratilgandan va ma'lumotlar to'plami vaqtli qatorlar deb e'lon qilingandan so'ng, ma'lumotlar to'plamini saqlang. shuning uchun bu jarayon ushbu ma'lumotlar uchun takrorlanishi shart emas. Stata yangi o'zgaruvchini, kerakli display formati va ma'lumotlar to'plami bilan birga vaqtli qatorlar ma'lumotlarini saqlaydi:

save new.dta, replace

O'zgartirish birikmasi Stata bir xil nomdagi mavjud ma'lumotlar to'plamini qayta yozishga yordam beradi, shuning uchun ushbu parametrdan ehtiyot bo'ling.

Okun ma'lumotlar to'plami

Qilish kerak bo'lgan birinchi narsa – katalogni ma'lumotlaringizni o'z ichiga olgan katalogga o'zgartirish va ma'lumotni yuklash. Shu mashqda biz okun.dta ma'lumotlaridan foydalanamiz:

use okun, clear

Ushbu ma'lumotlar to'plamida foiz bo'yicha choraklik kuzatuvlar bo'lgan g va u ikkita o'zgaruvchi mavjud, bular mos ravishda 1985q2- 2009q3 yillarda AQSHda Groos Domestik va ishsizlik darajasining har choraklik kuzatuvlaridir. Ma'lumotlar

yuklangandan so'ng, yaratish buyrug'i yordamida sana tayinlanadi. Stata sanalarni yaratish uchun maxsus funksiyalarni o'z ichiga oladi, ular Stata sanalarga qanday munosabatda bo'lishini (butun sonlar) va odamlarning ishlash usullari (kunlar, oylar, yillar va boshqalar) tarjima qiladi.

Har choraklik ma'lumotlar 1985-yilning ikkinchi choragida boshlanadi. Sanalarni belgilash va barcha o'zgaruvchilarni vaqtli qatorlardan foydalanishga aylantirish.

```
generate date = tq(1985q2) + _n-1
list date in 1
format %tq date
list date in 1
tsset date
```

Natija:

```
. generate date = tq(1985q2) + _n-1
```

```
. list date in 1
```

	date
1.	101

```
. format %tq date
```

```
. list date in 1
```

	date
1.	1985q2

```
. tsset date
      time variable:  date, 1985q2 to 2009q3
              delta:  1 quarter
```

Stata nima qilayotganini ko'rsatish uchun ikkita list in 1 buyruqlar qo'shiladi, ular amaliyotda kerak emas. Shunga qaramay, ular 1985q2 1960q1dan 101 chorak oldinda ekanligini aniqladilar. Format buyrug'i Stataga 101 sanasini 1985q2 sifatida ko'rsatishni ifodalaydi.

9.1.2. Vaqtli qatorlar grafiklari

Ma'lumotlar yuklangandan so'ng, vaqt o'zgaruvchisi yaratilgan, formatlangan va o'zgaruvchilar vaqtli qatorlar sifatida e'lon qilinganidan so'ng, siz tahlilning dastlabki bosqichlarini boshlashga tayyorsiz. Vaqt ketma- ketligi bilan, vaqtga nisbatan o'zgaruvchilarni chizishdan ko'ra boshlash uchun yaxshiroq joy yo'q. Bu ma'lumotlar muhim xususiyatlarni ochib beradi (masalan, statsionarlik, tendensiyalar, tizimli tanaffuslar va boshqalar). Ishsizlik darajasi va YaIM ning o'sish sur'atlarini aniqlash uchun tsline ishlatiladi. Ikkala uchastkaning teglarini bir xil grafikda olish uchun teglar label var buyruqlari yordamida qisqartiriladi. Keyin tsline, ya'ni graph twoway tsline qisqartmasi bo'lib, ikkala qatorni bir xil grafikda ifodalanadi:

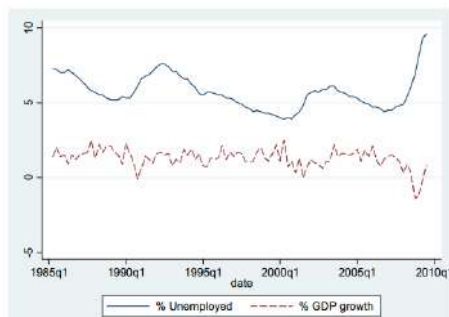
```
label var u "% Unemployed"
```

```
label var g "% GDP growth"
```

```
tsline u g, lpattern(solid dash)
```

Ikki vaqtli qatorlar grafigi ustiga qo'yiladi, chunki grafik chiziladigan qatorlarning har biri qavs ichiga olinadi. Boshqa variantlardan foydalanish mumkin, ammo biz buni hozircha soddalashtiramiz.

Stata grafiklari quyidagicha ifodalanadi:



Ishsizlik grafigi (qalin chiziq) YaIM o'sishiga qaraganda keng diapazondagi o'zgarishlarni ko'rsatadi, lekin bir vaqt oralig'idan ikkinchisiga kamroq farq qiladi. O'zgaruvchilarning har biri statsionar emasligini ko'rsatadigan aniq tendensiyalar, tanaffuslar yoki boshqa

xususiyatlar mavjud emas. Shuning uchun bu o'zgaruvchilar, ehtimol, ushbu bobda muhokama qilingan an'anaviy regressiya usullari uchun juda mos keladi. Bobda grafiklarning mumkin bo'lgan nostatsionarligini o'rganish uchun yana 12 ta rasmiy test ishlab chiqilgan. Hozirgacha ular statsionar ekanligi taxmin qilinadi.

9.1.3. Stataning lagli va farqli o'zgaruvchilari

Ilgari ko'rilganidek, list buyrug'i ma'lumotlar to'plamidan ekranga o'zgaruvchilarni chop etish uchun foydalaniladi. Bunda kuzatuvlarni cheklash uchun in 1/5 va 96/98 oralig'ida qo'llaniladi. Chop etilgan o'zgaruvchilar birinchi bo'lib V bobda o'rganilgan Stata unary operatorlarining boshqa namunasidan foydalaniladi.

Stata maxsus unary operatorlarini o'z ichiga oladi, ular vaqtli qatorlardagi ma'lumotlarning laglari va farqlarini juda oson hamda samarali qilish uchun foydalanilishi mumkin. Bu o'zgaruvchilar Stata User's Manual qo'llanmasida Time-series varlists sarlavhasi ostida to'liq ifodalangan. Bu yerda operatorlarning qisman ro'yxati va ularning ma'nolari:

Operator	Meaning
L.	lag (x_{t-1})
L2.	2-period lag (x_{t-2})
...	
D.	difference ($x_t - x_{t-1}$)
D2.	difference of difference ($x_t - 2x_{t-1} + x_{t-2}$)

Bu (unary) operatorlar davrdan keyingi o'zgaruvchisida ishlaydi. Masalan, L.u o'zgaruvchini oladi u va uni bir davrga kechiktiradi (laglanadi). Xuddi shunday D.u bir davr vaqt farqini $u_t - u_{t-1}$ oladi.

Kechikish (laglash) va farq operatorlari chiziqli bo'lib, har qanday tartibda birgalikda ishlatilishi mumkin. Masalan, u (i.e., $ldut = u_t - u_{t-2}$)dagi kuzatuvlar orasidagi kechikkan farqni bir marta olish uchun L.D.u dan foydalanish mumkin. Bu o'ngdan chapga ishlaydi: u farqini oling va keyin uni bir davrga kechiktiring(laglang). Operatsiyalarda chiziqchilik bu D.L.u ekvivalentligini bildiradi, kechikish u bir davrda va keyin farq qiladi. Bu shuningdek to'g'ri $L.L = L2$. O'zgaruvchini ikki davrga kechiktirish uchun L.L.u yoki soddaroq qilib aytganda,

L2.u dan foydalaning. Raqam quyidagi o'zgaruvchining o'tmishda qancha davr ortda qolganligini ko'rsatadi. Shunday qilib L2.u, u ikki davrga ortda qoladi (i.e., = ut-2).

Yetakchi (F) va mavsumiy farqlar (S) ni yaratadigan qo'shimcha vaqtli qatorlar operatorlari mavjud. Faktor o'zgaruvchilari uchun unary operatorlar misolida bo'lgani kabi, bu vaqtli qator operatorlari ham modelga kiritish uchun alohida o'zgaruvchilar yaratish zaruriyatidan xalos qiladi. Quyida muhokama qilinadigan bir qancha boshqa yorliqlar mavjud.

Ushbu operatorlardan foydalanishni ko'rsatish uchun o'zgaruvchilar, kechikishlar va farqli ma'lumotlar to'plamining boshida hamda oxirida kuzatuvlar uchun quyida keltirilgan. Umuman olganda, ketma-ketlik mazmuni ma'noli bo'lishini va vaqt davrlari to'g'ri o'zgaruvchilarga tayinlanganligini ta'minlash uchun bir nechta kuzatuvlarni chop etish yaxshi amaliyotdir. Sana u ostida u,g, o'zgarishi va bir nechta kechikishlar vaqtli qatorlar operator-lari yordamida chop etiladi.

list date u L.u D.u g L1.g L2.g L3.g in 1/5

list date u L.u D.u g L1.g L2.g L3.g in 96/98

. list date u L.u D.u g L1.g L2.g L3.g in 1/5

	date	u	L. u	D. u	g	L. g	L2. g	L3. g
1.	1985q2	7.3	.	.	1.4	.	.	.
2.	1985q3	7.2	7.3	-.1	2	1.4	.	.
3.	1985q4	7	7.2	-.2	1.4	2	1.4	.
4.	1986q1	7	7	0	1.5	1.4	2	1.4
5.	1986q2	7.2	7	.2	.9	1.5	1.4	2

. list date u L.u D.u g L1.g L2.g L3.g in 96/98

	date	u	L. u	D. u	g	L. g	L2. g	L3. g
96.	2009q1	8.1	6.9	1.2	-1.2	-1.4	.3	.9
97.	2009q2	9.3	8.1	1.2	-.2	-1.2	-1.4	.3
98.	2009q3	9.6	9.3	.3	.8	-.2	-1.2	-1.4

Vaqtli qatorlar o'zgaruvchilari ulardan foydalanishni

osonlashtiradigan yana bir xususiyatga ega. Statada operator(numlist) buyrug'idir.

A *numlist* is a list of numbers with blanks or commas in between. There are a number of shorthand conventions to reduce the amount of typing necessary. For instance:

2	just one number
1 2 3	three numbers
3 2 1	three numbers in reverse order
5 1 1.5	three different numbers
1 3 -2.17 5.12	four numbers in jumbled order
1/3	three numbers: 1, 2, 3
3/1	the same three numbers in reverse order
5/8	four numbers: 5, 6, 7, 8
-8/-5	four numbers: -8, -7, -6, -5
-5/-8	four numbers: -5, -6, -7, -8
-1/2	four numbers: -1, 0, 1, 2
1 2 to 4	four numbers: 1, 2, 3, 4
4 3 to 1	four numbers: 4, 3, 2, 1
10 15 to 30	five numbers: 10, 15, 20, 25, 30
1 2:4	same as 1 2 to 4
4 3:1	same as 4 3 to 1
10 15:30	same as 10 15 to 30
1(1)3	three numbers: 1, 2, 3
1(2)9	five numbers: 1, 3, 5, 7, 9
1(2)10	the same five numbers: 1, 3, 5, 7, 9
9(-2)1	five numbers: 9, 7, 5, 3, and 1
-1(.5)2.5	the numbers: -1, -.5, 0, .5, 1, 1.5, 2, 2.5
1[1]3	same as 1(1)3
1[2]9	same as 1(2)9
1[2]10	same as 1(2)10
9[-2]1	same as 9(-2)1
-1[.5]2.5	same as -1(.5)2.5
1 2 3/5 8(2)12	eight numbers: 1, 2, 3, 4, 5, 8, 10, 12
1,2,3/5,8(2)12	the same eight numbers
1 2 3/5 8 10 to 12	the same eight numbers
1,2,3/5,8,10 to 12	the same eight numbers
1 2 3/5 8 10:12	the same eight numbers

Ko'rib turganingizdek, numlist juda moslashuvchan. U diapazonlarni, ketma-ketliklarni, shuningdek, aniq raqamlar ro'yxatini belgilash imkonini beradi. Bular manfiy raqamlarni o'z ichiga olishi mumkin va ularning tartibi osongina o'zgartirilishi mumkin. Ushbu sintaksisdan foydalanib, list buyruqlarini qisqartirish mumkin.

list date L(0/1).u D.u L(0/3).g in 1/5

list date L(0/1).u D.u L(0/3).g in 96/98

L(0/1).u buyrug'i u L.u va L(0/3).g ga ekvivalent g L.g L2.g L3.g bilan bir xildir.

9.2. Cheklangan taqsimlangan kechikishlar (laglar)

Cheklangan taqsimlangan kechikish modellari mustaqil o'zgaruvchilarni va ularning kechikishlarini o'zgaruvchilar sifatida o'z ichiga oladi:

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \dots + \beta_q x_{t-q} + e_t, \quad t = q + 1, \dots, T$$

$$\beta_q x_{t-q} + e_t, \quad t = q + 1, \dots, T$$

Bu yerda ko'rib chiqilgan maxsus misol Okun qonunini tekshirishdir. Bu modelda ishsizlikning bir davrdan ikkinchi davrgacha o'zgarishi iqtisodiyotdagi ishlab chiqarishning o'sishi sur'atiga bog'liqligini ifodalaydi:

$$U_t - U_{t-1} = -\gamma (G_t - G_N)$$

Bu yerda U_t – ishsizlik darajasi, G_t – YaIM o'sishi, G_N – YaIM o'sishning normal sur'ati. Regressiya modeli:

$$\Delta U_t = \alpha + \beta_0 G + e_t$$

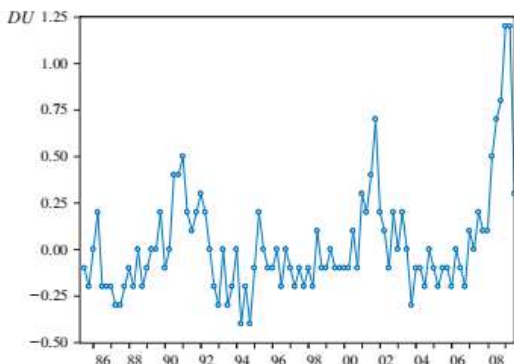
Bu yerda Δ farq operatori, $\alpha = \gamma G_N$, $\beta_0 = -\gamma$ va xato atamasi modelga qo'shilgan. Ishlab chiqarish hajmidagi o'zgarishlar ishsizlikka taqsimlangan kechikish ta'siriga ega bo'lishi mumkinligini tan olgan holda barcha ta'sirlar bir zumda sodir bo'lmaydi, ishlab chiqarish uchun modelga kechikishlar(laglar) qo'shiladi:

$$\Delta U_t = \alpha + \beta_0 G_t + \beta_1 G_{t-1} + \beta_2 G_{t-2} + \dots + \beta_q G_{t-q} + e_t, \quad t = q + 1, \dots, T$$

Ikki vaqtli ketma-ketlikdan foydalanib, grafik hosil qilish mumkin:

tsline D.u g

Quyidagi ko'rinishdagi grafik hosil bo'ladi:



Statada cheklangan taqsimlangan kechikish modelini hisoblash vaqtli qatorlar operatorlari yordamida juda oddiy amalga oshiriladi.

q=3 bo'lsin va

regress D.u L(0/3).g

Natija:

. regress D.u L(0/3).g

Source	SS	df	MS			
Model	5.13367789	4	1.28341947			
Residual	2.73516422	90	.030390714			
Total	7.86884211	94	.083711086			

Number of obs =	95
F(4, 90) =	42.23
Prob > F =	0.0000
R-squared =	0.6524
Adj R-squared =	0.6370
Root MSE =	.17433

D.u	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
g						
--	-.2020526	.0330131	-6.12	0.000	-.267639	-.1364663
L1.	-.1645352	.0358175	-4.59	0.000	-.2356929	-.0933774
L2.	-.071556	.0353043	-2.03	0.046	-.1416941	-.0014179
L3.	.003303	.0362603	0.09	0.928	-.0687345	.0753405
_cons	.5809746	.0538893	10.78	0.000	.4739142	.688035

Yana bir bor L(numlist) sintaksisi g ning bir vaqtda va 3 ta kechikkan qiymatlarini joylashtirish uchun modelda ishlatiladi.

Ikki o'zgaruvchining kechikish qiymatidan foydalangan holda modelni qayta baholaymiz.

. regress D.u L(0/2).g

Source	SS	df	MS			
Model	5.17925206	3	1.72641735			
Residual	2.74074794	92	.029790739			
Total	7.92	95	.083368421			

Number of obs =	96
F(3, 92) =	57.95
Prob > F =	0.0000
R-squared =	0.6539
Adj R-squared =	0.6427
Root MSE =	.1726

D.u	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
g						
--	-.2020216	.0323832	-6.24	0.000	-.2663374	-.1377059
L1.	-.1653269	.0335368	-4.93	0.000	-.2319339	-.0987198
L2.	-.0700135	.0331	-2.12	0.037	-.1357529	-.0042741
_cons	.5835561	.0472119	12.36	0.000	.4897892	.6773231

g dagi statistik jihatdan ahamiyatsiz uchinchi kechikishning pasayishi natijasida model mosligida deyarli hech qanday o'zgarish yo'q.

9.3. Ketma-ket korrelatsiya

Vaqtli qatorlar regressiyadagi yana bir murakkablik, regressiya modelining xatolari bir-biri bilan bog'langan bo'lganda yuzaga keladi. Bu Gauss Markov teoremasining asosiy taxminlaridan

birini buzadi va parametrlarning eng kichik kvadratlarini baholash xususiyatlariga sezilarli ta'sir ko'rsatadi.

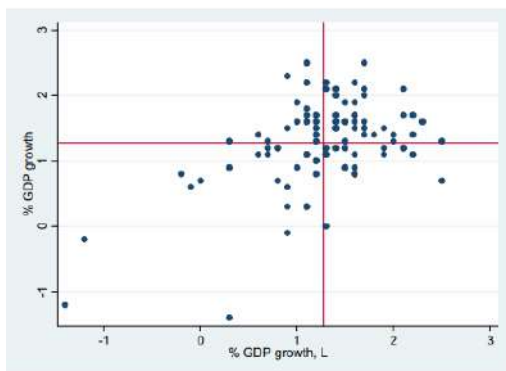
Iqtisodiyotda ketma-ket korrelatsiya iqtisodiy tebranishlar davomiyligi ma'lumotlarning tanlab olish chastotasidan oshib ketganda sodir bo'ladi. Bu tebranish vaqt oralig'ida yo'qotishlarni olib keladi, bu esa xatolarning ijobiy bog'lanishiga olib keladi. Ko'p hollarda bu regressiyaning vaqt strukturasi to'g'ri modellashning muvaffaqiyatsizligini anglatadi yoki kiritilgan o'zgaruvchilar bilan bog'liq bo'lgan kechikkan o'zgaruvchilar o'tkazib yuboriladi yoki to'g'ri modellashtirilmagan bog'liq o'zgaruvchida ba'zi bir barqarorlik mavjud bo'ladi. Yechim regressiya funksiyasini to'g'ri belgilashdir $E(e_t | \text{barcha o'zgaruvchilar}) = 0$. Bu eng kichik kvadratlarining izchil bo'lishi kesishma va qiyaliklar uchun zarur shartni qondiradi.

Eng kichik kvadratlar qoldiqlarida avtokorrelatsiyani aniqlash muhim, chunki eng kichik kvadratlar bu holatda nomuvofiq bo'lishi mumkin. Birinchi vosita g va $L.g$ ning tarqalish grafigini ishlab chiqarish uchun ishlatiladi. Gorizontaal va vertikal chiziqlar taxminan o'rtaga joylashtiriladi.

summarize g

scatter g L.g, xline('r(mean)') yline('r(mean)')

Natija esa:



Grafikdan so'ng summarize buyrug'ini amalga oshirish zarur, chunki YalM o'sishi o'rtacha grafikda ko'rsatilgan chiziqlarni chizish

uchun hisoblash kerak. Saqlanganlar orasida g ning o'rtachasi natijalari, ularni odatdagi return list buyrug'i yordamida ko'rish mumkin:

```
. summarize g
Variable | Obs      Mean      Std. Dev.      Min      Max
-----+-----+-----+-----+-----+-----
g        | 98      1.276531    .6469279      -1.4      2.5

. return list
scalars:
      r(N) = 98
      r(sum_w) = 98
      r(mean) = 1.276530612244898
      r(Var) = .4185156743109615
      r(sd) = .6469278741180978
      r(min) = -1.4
      r(max) = 2.5
      r(sum) = 125.1
```

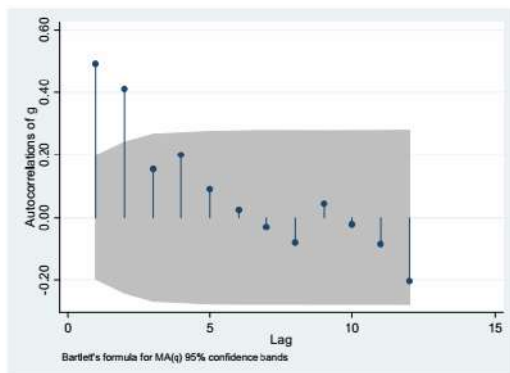
O'rtachaga kirish uchun uning katta nomi, r(mean), bittali qo'shtirnoq ichiga olinishi kerak, ya'ni `r(mean)'. Birinchi qo'shtirnoq chap bitta qo'shtirnoq ('-- klaviaturalarning yuqori chap tomonida), ikkinchisi esa o'ngdagi yagona qo'shtirnoq ('--located under the double quote " ko'pgina klaviaturalarda ostida joylashgan).

Raqamli yondashuv hisoblangan namunadagi avtokorrelatsiyalarni ko'rib chiqishdir. Bular quyidagi buyruq orqali topiladi:

ac g, lags(12) generate(ac_g)

list ac_g in 1/12

ac buyrug'i keyingi (g) o'zgaruvchisi uchinchi namunaviy avtokorrelatsiyalarni hisoblaydi va lags(12) birikmasi Stataga 12-davrgacha bo'lgan g oraliqda kompyuter avtokorrelatsiyalarini bildiradi. Natijalar grafikda ifodalaniladi, garchi avtokorrelyatsiyalar ac_g nomli o'zgaruvchidagi generate birikmasi yordamida saqlanadi.



Soyali maydonda 95% ishonch bandi paydo bo'ladi. E'tibor bering, faqat birinchi ikkitasi avtokorrelatsiyalar 5% darajasida noldan sezilarli darajada farq qiladi.

Taxminan 95% ishonch diapazonlari $\sqrt{Tr_k^a} \sim N(0,1)$ faktidan foydalanib hisoblanadi. Gen dan $z = \sqrt{e(N)} * ac_g$ chegaraga hosil qilish uchun foydalaning. Agar raqamlardan biri 1,96 dan kichik yoki 1,96 dan katta bo'lsa, u taxminan 95% ishonch oralig'idan tashqarida yotadi va 5% darajasida statistik ahamiyatga ega bo'ladi. Stataning ishlash funksiyasi boshqa (Bartlett's) metoddan foydalanadi va natijalar ushbu oddiy taxminga asoslanganidan farq qilishi mumkin.

ac_g da saqlangan qiymatlari va z chegaralari:

. list ac_g z in 1/12

	ac_g	z
1.	.49425676	4.842708
2.	.4107073	4.024093
3.	.1544205	1.513006
4.	.20043788	1.963882
5.	.09038538	.8855922
6.	.02447111	.239767
7.	-.03008434	-.2947652
8.	-.08231978	-.8065658
9.	.04410661	.4321548
10.	-.02128483	-.2085479
11.	-.08683463	-.8508022
12.	-.20404326	-1.999207

Phillips egri chizig'i

Ikkinchi misol Fillips egri chizig'iga asoslangan bo'lib, inflatsiya va ishsizlik orasidagi munosabatni ifodalaydi.

Inflyatsiya va ishsizlikning o'zgarishi bilan bog'liq oddiy regressiya:

$$INF_t = \beta_1 + \beta_2 DU_t + e_t$$

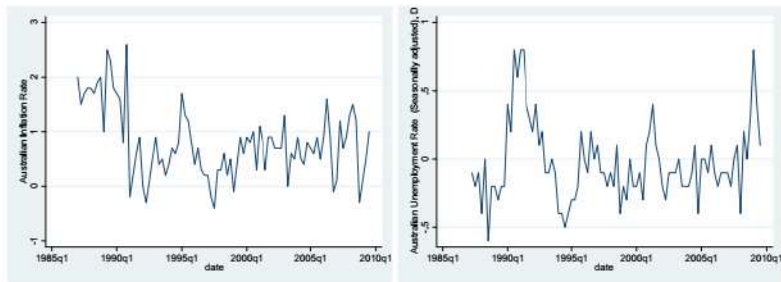
Model Avstraliyaning inflatsiya darajasi va ishsizlik darajasini

1987-yil 1-chorakni o'z ichiga olgan phillips_au.dta ma'lumotlari yordamida baholanadi. Ma'lumotlarni yuklang, sanani satrga formatlang va ma'lumotlar to'plamini vaqtli qatorlar sifatida o'rnatg:

```
use phillips_au, clear
generate date = tq(1987q1) + _n-1
format %tq date
tsset date
```

Birinchidan, inflatsiya darajasi va ishsizlikning o'zgarishini grafiklarda ifodalang.

```
tsline inf
tsline D.u
```



Keyin eng kichik kvadratlar yordamida modelni baholang va qoldiqni toping:

```
reg inf D.u
predict ehat, res
. reg inf D.u
```

Source	SS	df	MS
Model	2.04834633	1	2.04834633
Residual	34.0445426	88	.386869802
Total	36.0928889	89	.405538077

Number of obs = 90
F(1, 88) = 5.29
Prob > F = 0.0238
R-squared = 0.0568
Adj R-squared = 0.0460
Root MSE = .62199

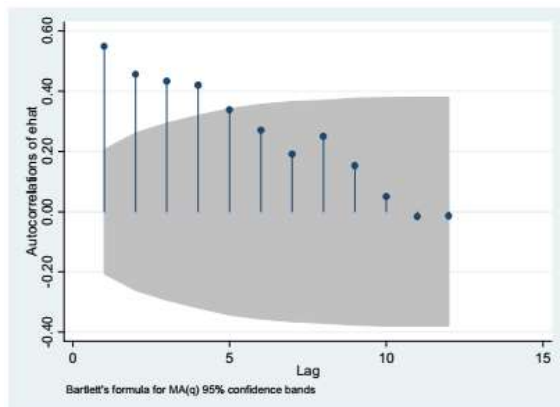
	inf	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
D1.		-.5278638	.2294049	-2.30	0.024	-.9837578 -.0719699
_cons		.7776213	.0658249	11.81	0.000	.646808 .9084345

Qoldiqlar residual correlogram yordamida avtokorrelatsiya uchun tekshiriladi. Qoldiq korrelogramma e_t va $e_{(t-1)}$ o'rtasidagi avtokorrelatsiyalar qatorini kuzatishlar orasidagi vaqt oralig'iga

nisbatan $j=1,2,\dots,m$ ga teng chiziqli grafik ifodalanadi. Namunadagi avtokorrelatsiyalar rk deb nomlangan o'zgaruvchida saqlanadi, birinchi beshtasi chop etiladi va keyin ma'lumotlar to'plamidan o'chiriladi, chunki ular endi kerak emas.

ac ehat, lags(12) generate(rk)

list rk in 1/5



. list rk in 1/5

	rk
1.	.54865864
2.	.45573248
3.	.43321579
4.	.42049358
5.	.33903419

Ko'rinib turibdiki, bir qator muhim avtokorrelatsiyalar mavjud va ular nisbatan katta. Stata corrgram funksiyasini o'z ichiga oladi, uning kichik to'plami ac buyrug'i hisoblanadi. Corrgram avtokorrelatsiyalar jadvalini ishlab chiqaradi (shuningdek, qisman avtokorrelatsiyalar va Portmanto (Q) statistikasi). Shuningdek, u avtokorrelatsiyalarning xarakterga asoslangan sujetini ko'rsatadi. Corrgramning yana bir xususiyati bu statistikaning har biri $r()$ sifatida saqlanadi. Birinchi beshlik avtokarrelatsiyani saqlash va

natijani olish uchun corrgramdan foydalaniladi.

corrgram ehat, lags(5)

. corrgram ehat, lags(5)

LAG	AC	PAC	Q	Prob>Q	-1 [Autocorrelation]	0	1 [Partial Autocor]	0	1
1	0.5487	0.5498	28.006	0.0000					
2	0.4557	0.2297	47.548	0.0000					
3	0.4332	0.1926	65.409	0.0000					
4	0.4205	0.1637	82.433	0.0000					
5	0.3390	0.0234	93.63	0.0000					

Birinchi uchta avtokorrelatsiya namunasini natijani olish:

```
. di "rho1 = " r(ac1) " rho2 = " r(ac2) " rho3 = " r(ac3)
rho1 = .54865864 rho2 = .45573248 rho3 = .43321579
```

corrgram bajargandan so'ng saqlanadigan boshqa statistik ma'lumotlarni ko'rish uchun return list kiriting.

9.4. Ketma-ket korellatsiya uchun boshqa testlar

Ketma-ket korrelatsiya uchun ikkinchi test VIII bobda geteroskedastiklik kontekstida muhokama qilingan Langrage multiplikator testi prinsipiga asoslanadi. Sinov statistikasi TR^2 ga asoslangan yordamchi regressiya. Avtokorrelatsiya uchun bu test yordamchi regressiyaga asoslanadi, bunda siz eng kichik kvadratlar qoldiqlari va asl regressorlar bo'yicha eng kichik qoldiqlarni regressiya qilasiz. Agar TR^2 ni o'z ichiga olgan regressorlar e $\hat{e}_{t-1}$ dagi yetarli o'zgarishlarni tushuntirsa, u holda $\hat{e}_{t-1}$ tufayli avtokorrelatsiya bo'lishi kerak. Regressiya modeli uchun:

$$y_t = \beta_1 + \beta_2 x_t + e_t$$

eng kichik kvadratlar yordamida parametrlarni baholang va qoldiqlarni saqlang $\hat{e}_t$. $\hat{e}_{t-1}$ olish uchun Lag qoldiqlaridan foydalanamiz. Keyin, ikkinchi "yordamchi" regressiyani $\hat{e}_t$ bog'liq o'zgaruvchi sifatida va kechikish qiymatini $\hat{e}_{t-1}$ mustaqil

o'zgaruvchi sifatida baholang. Asl regressiyadan boshqa barcha mustaqil o'zgaruvchilarni ham o'z ichiga oladi. Oddiy chiziqli regressiya uchun yordamchi regressiya hisoblanadi.

$$\hat{e}_t = \gamma_t + \gamma_2 x_t + \rho \hat{e}_{t-1} + \text{residual}$$

Agar avtokorrelatsiyaning yo'qligi haqidagi nol gipoteza to'g'ri bo'lsa, regressiyadan TR^2 , x^2 taqsimotga ega bo'ladi, bu yerda T yordamchi regressiyadagi kuzatishlar soni. Muhim avtokorrelatsiya mavjudligi haqidagi xulosa rad etishga olib keladi. Fillips egri chizig'i misoli uchun, ehat asl regressiyadan saqlanib qolgan deb faraz qilsak:

quietly regress ehat D.u L. ehat

di "Observations = " e(N) " and TR2 = " e(N)*e(r2)

Natija:

```
. di "Observations = " e(N) " and TR2 = " e(N)*e(r2)
Observations = 89 and TR2 = 27.608808
```

Principles of Econometrics, 4-nashr adabiyotida bu statistika barcha mavjud kuzatishlar yordamida hisoblab chiqilgan. Modelda kechikkan (lagli) qiymat e_{t-1} paydo bo'lganligi sababli, yordamchi regressiyani baholashda odatda kuzatuv kamayadi. Bu holda, birinchi uchun yetishmayotgan qiymat qoldiq nolga almashtirilishi mumkin. Bu joriy kontekstda ruxsat etiladi, chunki uning kutilgan qiymati shunday (ya'ni, e_1). To'g'ri test statistikasini olish uchun texnik jihatdan $e_1 = 0$ o'rnatish kerak emas; ammo, natijani matnda takrorlash uchun bu yerda amalga oshiriladi. Keyin,

replace ehat = 0 in 1

quietly regress inf D.u L. ehat

di "Observations = " e(N) " and TR2 = " e(N)*e(r2)

Natijada esa:

```
. di "Observations = " e(N) " and TR2 = " e(N)*e(r2)
Observations = 90 and TR2 = 27.592347
```

Bu matndagi natijaga mos keladi. Yuqori darajadagi avtokorrelatsiyani tekshirish oddiy. AR(4) ni hisoblash uchun eng kichik kvadratlar qoldiqlarini regressorlar sifatida kiriting va TR^2 ni hisoblang. Chi kvadrat uchun erkinlik darajalari muqobil ostidagi avtokorrelatsiya tartibiga teng (bu holda, 4).

Buni amalga oshirish uchun buyruq:

quietly regress ehat D.u L(1/4).ehat

Natijada:

```
. di "Observations = " e(N) " and TR2 = " e(N)*e(r2)
Observations = 86 and TR2 = 33.385269
```

Kechikishlarni qabul qilishda yuzaga keladigan ehat yetishmayotgan qiymatlari nolga o'rnatiladi (ya'ni, $e_3=0$, $e_2=0$, $e_1=0$, $e_0=0$). Bu butun namunadan foydalanish imkonini beradi. Ma'lum bo'lishicha, bu Statada dasturlash juda oson emas, shuning uchun biz bu yerda muhokamani o'tkazib yuboramiz biroq, buni amalga oshirish uchun buyruqni ushbu bobning oxiridagi do faylda topish mumkin.

Natijalarni **estat bgodfrey** o'rnatilgan post baholash buyrug'i yordamida osongina takrorlash mumkin.

regress inf D.u

estat bgodfrey, lags(1)

estat bgodfrey, lags(4)

Buyruq modeldagi regressorlar sifatida qancha kechikkan qoldiqlarni kiritish kerakligini ko'rsatadigan variantdan foydalanadi. AR(1) misolida, muqobil gipoteza, modelning xatolari birinchi tartibli avtokorrelatsiyaga ega; kechikishlar (1) ishlatiladi. Natijada:

```
. estat bgodfrey, lags(1)
```

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
1	27.592	1	0.0000

H0: no serial correlation

AR(4) muqobil uchun:

```
. estat bgodfrey, lags(4)
```

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
4	36.672	4	0.0000

H0: no serial correlation

Harbirholatdaoddiyregressiyaqoldiqlaridaavtokorrelatsiyaning aniq natijalari mavjud.

9.5. Ketma-ket bo'g'liq xatolar bilan baholash

Regressiya modeli regressorlar sifatida qaram o'zgaruvchining kechikishlarini o'z ichiga olmasa, eng kichik kvadratlar xatolar AR(q) modeliga amal qilgan taqdirda ham izchil bo'ladi. $t \neq s$ uchun $MR4$, $cove(e_t, e_s) = 0$ eng kichik kvadratlar farazi buzilganda u endi samarali emas (asimptotik). Afsuski, odatiy standart xatolar endi to'g'ri emas, bu statistik jihatdan noto'g'ri gipoteza testlari va ishonch oralig'i olib keladi.

9.5.1. Eng kichik kvadratlar va HAC standart xatolar

Odatdagidek eng kichik kvadratlar standart xatolari, xuddi biz geteroskedastikmodellarda bo'lgani kabi, Newey va West tomonidan baholovchi yordamida taklif qilingan. Newey-West standart xatolar (HAC deb ham ataladi – geteroskedastiklik va avtokorrelatsiya izchil standart xatolar) VIII bobda kiritilgan geteroskedastik izchil standart xatolarga o'xshaydi. Ular avtokorrelatsiya qilingan xatolar uchun AR(1) bo'lishi shart bo'lmagan va pastroq dispersiyaga ega bo'lgan baholovchini olish uchun zarur bo'lgan dinamik xato modelining

o'ziga xosligini talab qilmaydigan izchil bo'lish afzalligiga ega. HAC VIII bobdagi geteroskedastika mustahkam standart xato hisoblagichi kabi avtomatik ishlatilmaydi. Avtokorrelatsiyaga nisbatan ishonchli bo'lishi uchun qoldiq avtokorrelatsiya vaqt o'tishi bilan qanchalik muhim bo'lishi mumkinligini aniqlash kerak. Asosan, tanlangan vaqt oynasidagi avtokorrelatsiya qilingan xatolar HAC hisoblashda o'rtacha hisoblanadi; o'rtacha hisoblab chiqiladigan davrlar soni va har bir qoldiq uchun qancha og'irlik berilishi foydalanuvchi tomonidan belgilanishi kerak.

O'rtacha og'irlik kernel deb ataladigan narsa yordamida amalga oshiriladi va tortish sxemasi (yadrosi) yordamida o'rtacha xatoliklar soni bandwidth deb ataladi. Rostini aytsam, bu atamalar o'rtacha foydalanuvchi haqida ozgina ma'lumot beradi. Yadroni o'rtacha og'irlikdagi boshqa nom va tarmoqli kengligini o'rtacha atamalar soni uchun atama sifatida o'ylab ko'ring. Stata kernel tanlashning hech qanday usulini taklif qilmaydi; yagona Bartlett mavjud. Biroq, bandwidth kengligi tanlanishi kerak.

Mos bandwidth kengligi tanlashga yordam beradigan bir necha usullar mavjud va ikkitasi bu yerda berilgan. T, ikkala holatda ham bandwidth kengligi namuna hajmiga bog'liq. Birinchisi $B = 0,75T$

$T^{1/3}$) dan foydalanadi. Boshqa mashhur tanlov- $B = 4(T/100)^{2/9}$. Bu EViews kabi boshqa dasturlarda standart bo'lib ko'rinadi va bu yerda matndagi natijalarni olish uchun foydalaniladi.

Bilvosita ko'rib chiqilishi kerak bo'lgan kelishuv mavjud. Kattaroq bandwidth kengligi noto'g'rilikni (yaxshi) va aniqlikni (yomon) kamaytiradi. Kichikroq bandwidth kengligi ko'proq mos keladigan avtokorrelatsiyalarni istisno qiladi (shuning uchun ko'proq tarafdashlik qiladi), lekin kichikroq farqqa ega. Umumiy tamoyil eng katta avtokorrelatsiyalarni o'z ichiga oladigan darajada katta bo'lgan tarmoqli (bandwidth) kengligini tanlashdir.

O'tkazish bandwidthni hisoblash uchun quyidagi buyruqdan foydalaning:

scalar B = round(4*(e(N)/100)^(2/9))

scalar list B

Bu philips _ aus.dta ma'lumotlar to'plamidagi 4 qiymatini

qaytaradi. Natija yaxlitlanadi, chunki Stata HAC ning hisoblashdagi kechikishlar sonini belgilash uchun butun sondan foydalanishni talab qiladi.

Statada mavjud bo'lgan yagona kernel bu Bartlettidir. Bu Newey va West tomonidan ushbu masala bo'yicha tadqiqotlarida foydalanilgan. Shuning uchun Stata HACni newey sifatida hisoblaydigan jarayonga ishora qiladi. U asosan regress o'rnini bosadi va u tarmoqli kengligi spetsifikatsiyasini talab qiladi. Keyin modelni Newey-West standart xatoliklari va tarmoqli kengligi 4 bo'lgan eng kichik kvadratlar bo'yicha baholash uchun quyidagi buyruqdan foydalaning:

newey inf D.u, lag(4)

Misolda model odatdagi eng kichik kvadratlar standart xatolari va HAC standart xatolari bilan eng kichik kvadratlar yordamida baholanadi. Natijalar quyida ko'rinadi, HAC standart xatolar o'ng tomondagi ustundagi taxminlar ostida ko'rinadi:

```
esttab Wrong_SE HAC_4, compress se(%12.3f) b(%12.5f) gaps ///
scalars(r2_a rss aic) title("Dependent Variable: inf") ///
mtitles("LS" "HAC(4)")
```

Dependent variable: inf		
	(1) LS	(2) HAC(4)
D.u	-0.52786* (0.229)	-0.52786 (0.318)
_cons	0.77762*** (0.066)	0.77762*** (0.112)
N	90	90
r2_a	0.04603	
rss	34.04454	
aic	171.91634	.
Standard errors in parentheses		
* p<0.05, ** p<0.01, *** p<0.001		

Compress opsiyasi qatorlar orasidagi vertikal bo'shliqni

qisqartirish uchun ishlatiladi, gaps opsiyasi koeffitsiyentlar orasiga bo'sh qatorlarni (yoki umuman olganda, qo'shimcha vertikal bo'shliq) qo'shadi va scalars opsiyasi turli statistik ma'lumotlarni ifodalash imkonini beradi, regressiya natijalari bilan birga saqlanadi.

Bu holda, sozlangan R^2 , kvadratlar xatosining regressiya yig'indisi va Stata tomonidan AIC mezonini hisoblash. Bundan tashqari, mtitle variantlari har bir ustunga mazmunli nom berish uchun ishlatiladi; bu parametr qo'llanilganda, qaram o'zgaruvchining nomi bo'lgan standart ustun nomi har bir qo'shtirnoq to'plamiga qo'ygan narsangiz bilan almashtiriladi. Title variant o'quvchilarga har bir holatda ishlatiladigan qaram o'zgaruvchi inf ekanligini bilish uchun ishlatiladi. Bu misolda, HAC standart xatolar odatdagidan (mos kelmaydigan) sezilarli darajada kattaroqdir.

9.5.2. Nochiziqli eng kichik kvadratlar

Ko'rib turganingizdek, HAC standart xatolar kamida ikkita kamchilikka ega: 1) ular avtomatik emas, chunki ular tarmoqli kengligi spetsifikatsiyasini talab qiladi; 2) ular oddiy chiziqli regressiyaga qaraganda samaraliroq bo'lgan ko'proq baholovchilarning standart xatolaridan kattaroqdir. Ushbu bo'limda AR(1) modelining parametrlarini samarali baholash uchun chiziqli bo'lmagan eng kichik kvadratlardan foydalaniladi.

Sizning darsligingizda mualliflar AR(1) regressiya modelidan boshlaydilar va bir ozalgebradan foydalanib, bu natijalarga erishiladi:

$$y_t = \beta_1(1 - \rho) + \beta_2 x_t + \rho y_{t-1} - \rho \beta_2 x_{t-1} + v_t$$

Bu model parametrlarda chiziqli emas, lekin qo'shimcha oq shovqin xatosiga ega. Ushbu xususiyatlar modelni chiziqli bo'lmagan eng kichik kvadratlarni baholash uchun mos keladi. Chiziqli bo'lmagan eng kichik kvadratlar kvadrat xatolar yig'indisini minimallashtiradigan parametrlarning qiymatlarini topish uchun raqamli usullardan foydalanadi. Modelni baholash uchun Stataning umumiy chiziqli bo'lmagan eng kichik kvadratlar buyrug'idan foydalaning, nl:

**nl (inf={b1}*(1-{rho}) + {b2}*D.u + {rho}*L.inf - {rho}*{b2}*(L.D.u)), ///
 variabels(inf D.u L.inf L.D.u)**

Sintaksis juda oddiy, ammo ba'zi tushuntirishlarni talab qiladi. Asosiy sintaksis quyidagicha:

nl (depvar=) [if] [in] [weight] [, options]

Modelning tizimli qismi birinchi qavslar to'plamiga kiritilgan. Parametrlar qavslar {} ichiga olinishi kerak. If, in and weight iborolari chiziqli regressiyadagi kabi ishlatiladi. Biroq, kechikkan o'zgaruvchilar tufayli, ma'lumotlar to'plamidagi kechikkan o'zgaruvchilar bo'yicha birinchi kuzatish uchun yetishmayotgan qiymatlar yaratiladi. Buning ishlashi uchun namuna faqat to'liq kuzatuvlar bilan cheklanishi kerak. Buning ikkita usuli bor – bu: birinchidan, siz (depvar= <sepx>) 2/34 da foydalanishingiz mumkin. yoki o'zgaruvchilarni bajarilganidek ro'yxatlashingiz mumkin, bu yerda variant o'zgaruvchilari (inf D.u L.inf L.D.u) yordamida topiladi.

Baholash natijalari:

urce	SS	df	MS	
odel	12.3860433	2	6.19302165	Number of obs =
dual	23.1986758	86	.269752044	R-squared =
				Adj R-squared =
				Root MSE =
otal	35.5847191	88	.404371808	Res. dev. =

inf	Coef.	Std. Err.	t	P> t	[95% Conf. Int.]
/b1	.7608716	.1245311	6.11	0.000	.513312 1.008431
/rho	.5573922	.0901546	6.18	0.000	.3781709 .7366135

Koeffitsiyent baholari matndagilarga mos keladi. Xuddi shu parameter baholarida funksiya kvadratlari yig'indisining minimaliga erishiladi. Hisoblangan standart xatolarda ba'zi kichik farqlar mavjud. Bu chiziqsiz modellarda ularni izchil ravishda baholashning turli usullari mavjudligi sababli sodir bo'ladi; ushbu misoldagi kabi kichik namunalarda bu farqlar bo'rttirilgan bo'lishi mumkin. Kattaroq namunalarda farqlar odatda kichik bo'ladi va namuna hajmi o'sishi bilan nazariyaga ko'ra yo'qoladi. Parametr ρ bo'yicha t nisbati 6.18 ga teng, bu p qimati 0,001 dan kam. Bu

shuni anglatadiki, har qanday o'rtacha darajadagi ahamiyatga ega (masalan 5%) qoldiqlar o'rtasida birinchi tartibli avtokorrelatsiya dalillari mavjud. Modelni baholagandan so'ng, keyingi bo'limda foydalanish uchun bir nechta skalyarlar hisoblab chiqiladi:

$$\text{scalar delta} = _b[b1_cons]*(1_b[rho_cons])$$

$$\text{scalar delta1} = -_b[rho_cons]*_b[b2_cons]$$

delta deb nomlangan skalyar $\hat{\beta}_1(1 - \hat{\rho})$ va delta1- $\hat{\beta}_1(1 - \hat{\rho})$

Bularning sabablari keyingi bobda muhokama qilinadi. Biroq e'tibor bering, taxminlar chiziqli regressiyaga qaraganda biroz boshqacha ko'rsatilgan. Chiziqli modellarda ishlatiladigan `_b[varname]` konvensiyasi `_b[paramname:_cons]` bilan almashtiriladi. Parametrlar uchun tegishli nomlarni topish uchun `nl buyrug'idan` keyin `coeflegend` variantidan foydalanish mumkin. Parametrlarni to'g'ri aniqlaganligingizni tekshirib ko'ring, `coeflegend` buyrug'i yordamida chiziqli bo'lmagan eng kichik kvadratlar regressiyasini qayta ishga tushiring. Bu odatda kerakli natijaning ko'p qismini bostiradi, lekin u har bir parametr uchun Stata nomlarini aniqlaydigan afsonani yaratadi. Bobning oxiridagi do faylda buning misoli mavjud.

9.5.3. Odatda ko'proq ishlatiladigan model

Modelning umumiyroq shaklini ko'rib chiqamiz:

$$y_t = \alpha + \delta_0 x_t + \delta_1 x_{t-1} + \theta_1 y_{t-1} + v_t$$

Parametrlarda chiziqli va chiziqli regressiya bilan baholanishi mumkin. Bu model oldingi modelga munosabatlar orqali bog'langan:

$$\alpha = \beta_1(1 - \hat{\rho}) \quad \delta_0 = \beta_2 \quad \delta_1 = -\rho\beta_2 \quad \theta_1 = \hat{\rho}$$

Chiziqli modelni (chiziqli) eng kichik kvadratlar bilan baholash mumkin va nazarda tutilgan cheklovning gipoteza testini o'tkazish mumkin. Cheklovlar nazarda tutuvchi nol gipoteza $H_0: \delta_1 = \theta_1 \delta_0$ teng emas degan muqobilga qarshi gipotezadir. Birinchi qadam eng kichik kvadratlar yordamida modelni baholashdir:

regress inf L.inf D.u L.D.u

```
. regress inf L.inf D.u L.D.u
```

Source	SS	df	MS	Number of obs =	89
Model	12.4166337	3	4.13887791	F(3, 85) =	15.18
Residual	23.1680854	85	.27256571	Prob > F =	0.000
				R-squared =	0.3489
				Adj R-squared =	0.3268
				Root MSE =	.52208
Total	35.5847191	88	.404371808		

	inf	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
inf						
L1.		.5592676	.0907962	6.16	0.000	.3787403 .7397948
D1.		-.6881852	.2498704	-2.75	0.007	-1.184994 -.191376
LD.		.3199526	.257504	1.24	0.217	-.1920343 .8319396
_cons		.3336325	.0899028	3.71	0.000	.1548817 .5123834

Oldingi bo'lim oxirida hisoblangan skalyar δ va δ_1 ga to'g'ri keladi. Hisoblangan qiymatlar:

```
. scalar list delta delta1
      delta = .33676767
      delta1 = .38704645
```

Chiziqli regressiyadan olingan umumiy baholar $\delta=0,334$ va $\delta_1=0,320$. Ikkala qiymat ham nl tomonidan baholangan cheklovchi modelda nazarda tutilgan qiymatlarga juda yaqin.

Chiziqsiz gipotezani haqiqatda sinab ko'rish uchun $H_0: \delta_1 = -\delta_0$ parametrlarning chiziqli bo'lmagan funksiyasini sinab ko'rish uchun Stataning o'rnatilgan funksiyasidan foydalaning, testnl.

```
testnl _b[L.D.u]=-_b[L.inf]*_b[D.u]
```

```
. testnl _b[L.D.u]=-_b[L.inf]*_b[D.u]
```

```
(1) _b[L.D.u] = -_b[L.inf]*_b[D.u]
```

```
      F(1, 85) =      0.11
      Prob > F =      0.7384
```

Katta p qiymati 0,74 chiziqli bo'lmagan eng kichik kvadratlar bilan baholangan AR(1) modeli haddan tashqari cheklovchi emasligini ko'rsatadi.

esttab buyrug'i yordamida ko'rib chiqilayotgan modellarning turli chiziqli xususiyatlari taqqoslanadi.

```
. esttab General No_LDu Original, compress se(%12.3f) b(%12.5f) ///
> gaps scalars(r2_a rss aic)
```

	(1) inf	(2) inf	(3) inf
L.inf	0.55927*** (0.091)	0.52825*** (0.085)	
D.u	-0.68819** (0.250)	-0.49086* (0.192)	-0.52786* (0.229)
LD.u	0.31995 (0.258)		
_cons	0.33363*** (0.090)	0.35480*** (0.088)	0.77762*** (0.066)
N	89	90	90
r2_a	0.32595	0.33137	0.04603
rss	23.16809	23.59054	34.04454
aic	140.78946	140.90217	171.91634

Standard errors in parentheses

* p<0.05, ** p<0.01, *** p<0.001

compress buyrug'i qatorlar orasidagi vertikal bo'shliqni qisqartirish uchun ishlatiladi, gaps opsiyasi koeffitsiyentlar orasiga bo'sh qatorlarni (yoki umuman olganda, qo'shimcha vertikal bo'shliqni) qo'shadi va scalars opsiyasi saqlanadigan turli statistik ma'lumotlarni chop etish imkonini regressiya natijalari bilan birga beradi. Bu holda, sozlangan R^2 kvadratlar xatosining regressiya yig'indisi va Stata tomonidan AIC mezonini hisoblash kerak.

9.6. Avtoregressive taqsimlangan kechikish (lagli) modellari

Cheklangan taqsimlangan kechikishlarni birlashtirgan va avtoregressiv model ko'rib chiqiladi. Bu avtoregressiv taqsimlangan kechikish modeli (ARDL) deb ataladi. ARDL (p,q) modeli umumiy shaklga ega:

$$y_t = \alpha + \theta_1 y_{t-1} + \dots + \theta_p y_{t-p} + \delta_0 x_t + \delta_1 x_{t-1} + \dots + \delta_q x_{t-q} + v_t$$

Regressor sifatida u p bog'liq o'zgaruvchining y_t va q mustaqil o'zgaruvchining x_t kechikishiga ega. Inflatsiyaning ARDL (1,1) va ARDL (1,0) modellari eng kichik kvadratlar yordamida baholaymiz.

Taxminlar saqlanadi va quyidagi jadvalda natijalar ifodalanadi:

regress inf L.inf L(0/1).D.u

estimates store AR1_DL1
regress inf L.inf D.u
estimates store AR1_DL0

```
. esttab AR1_DL1 AR1_DL0, compress se(%12.3f) b(%12.5f) ///  
> gaps scalars(r2_a rss aic)
```

	(1) inf	(2) inf
L.inf	0.55927*** (0.091)	0.52825*** (0.085)
D.u	-0.68819** (0.250)	-0.49086* (0.192)
LD.u	0.31995 (0.258)	
_cons	0.33363*** (0.090)	0.35480*** (0.088)
N	89	90
r2_a	0.32595	0.33137
rss	23.16809	23.59054
aic	140.78946	140.90217

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Ushbu modellar orasidan tanlash bir necha usul bilan amalga oshirilishi mumkin. Birinchidan, agar DU_{t-1} dagi t-ratio ahamiyatsiz bo'lsa unda dalillar shuni ko'rsatadiki, uni o'tkazib yuborish cheklangan modelning eng kam chayqaladigan baholovchisi xususiyatlariga salbiy ta'sir ko'rsatmasligi mumkin.

Yana bir imkoniyat VI bobda muhokama qilingan model tanlash qoidalaridan birini qo'llash mumkin. Eslaymizki, biz AIC va SC modelini tanlash qoidalarini hisoblaydigan **modelsel** deb nomlangan dasturni yozgan edik. Bu yerda dastur hisoblangan R^2 displeyini qoldirib, modeldagi kuzatuvlar sonini chop etish orqali biroz o'zgartiriladi. Statada dastur tuzilishi haqida batafsil ma'lumot olish uchun VI bobga qarang. AIC yoki SC yordamida ARDL (1,1) va ARDL (1,0) o'rtasida tanlash uchun modelsel deb nomlangan quyidagi dasturni yarating va ishga tushiring:

program modelsel

scalar aic = ln(e(rss)/e(N))+2*e(rank)/e(N)

scalar sc = ln(e(rss)/e(N))+e(rank)*ln(e(N))/e(N)

scalar obs = e(N)

```
scalar list aic sc obs
```

```
end
```

Endi quyida ko'rsatilgan tanlov mezonlarini tekshirib, har bir modelni baholang. Bu quyidagi natijalarni chiqaradi:

```
. quietly regress inf L.inf L(0/1).D.u
```

```
. modelsel  
    aic = -1.255973  
    sc = -1.1441242  
    obs =      89
```

```
. quietly regress inf L.inf L.D.u
```

```
. modelsel  
    aic = -1.1929642  
    sc = -1.1090776  
    obs =      89
```

ARDL (1,0) ham AIC, ham SC ni kamaytiradi va afzal qilingan modeldir. Ushbu tahlil bilan bog'liq muammolardan biri shundaki, qoldiqlar hali ham avtokorrelatsiyalangan bo'lishi mumkin yoki bu yerda ko'rib chiqilganlardan uzoqroq kechikishlar o'tkazib yuborilgan. Keyingi bo'limda bu batafsilroq ko'rib chiqiladi.

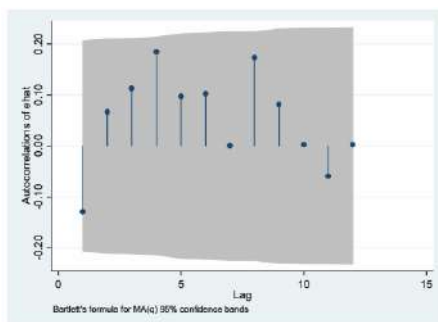
9.6.1. Fillips egri chizig'i

Birinchidan, ARDL (1,0) xatolari avtokorrelatsiya uchun tekshirilishi kerak. Buni korrelogrammaga qarash orqali amalga oshirish mumkin:

```
quietly regress inf L.inf D.u
```

```
predict ehat, res
```

```
ac ehat, lags(12)
```



yoki LM (Breusch-Godfrey) testi bo'yicha, bu holda 1-5 kechikishlar uchun statistik ma'lumotlarni o'z ichiga oladi.

```
. estat bgodfrey, lags(1 2 3 4 5)
```

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
1	4.130	1	0.0421
2	5.123	2	0.0772
3	5.221	3	0.1563
4	9.554	4	0.0486
5	12.485	5	0.0287

H0: no serial correlation

Ikki jarayon natijalari ARDL (1,0) qoldiqlarida avtokorrelatsiya mavjudligi haqida turli xil taassurotlarni beradi. Korrelogrammdagi avtokorrelatsiyalarning hech biri 95% ishonch oralig'idan tashqarida emas, lekin 1,4 yoki 5 kechikishlarga ega modellar uchun LM statistikasi 5% darajasida statistik ahamiyatga ega. Ikkinchidan model tanlash qoidalarini ko'proq avtokorrelyatsiya shartlarini o'z ichiga olgan kengroq modellarga qo'llash kerakligini taklif qiladi. Bu o'n ikkita modelni baholashni anglatadi; AR shartlari 1 dan 6 gacha, DL esa 0 dan 1 gacha, har bir taxminiy kombinatsiya bilan farqlanadi. AIC va SC statistikasi har bir model uchun hisoblab chiqiladi va umumiy minimallar qidiriladi. 1988q3 dan keyingi davrlar uchun ushbu modellarning har birini baholash uchun Stata buyrug'i ushbu bobning oxirida do faylda keltirilgan. Quyida buyruq parchasi berilgan va uning sintaksisi tushuntirilgan.

Quyidagi buyruq 1983-yilning uchinchi choragida boshlangan ma'lumotlar uchun ARDL (1,1) modelini baholaydi. Regressiya haqiqiy regressiya natijalarini bostirish uchun Stata quyida qisqartirilgan quietly buyrug'i yordamida baholanadi; bizning qiziqishimiz faqat shu nuqtada model tanlash qoidalarining qadriyatlariga bog'liq. Namunani ma'lum sanalar bilan cheklash uchun psevdofunksiya tq(1988q3) ishlatiladi. Eslatib o'tamiz, ushbu bobning boshida ushbu psevdofunksiya 1988q3 sanasini Stata tushunadigan raqamga aylantirilgandi. Choraklik ma'lumotlar uchun bu to'g'ri sintaksisdir:

```
qui reg L(0/1).inf L(0/1).D.u if date>= tq(1988q3)
di "p=1 q=1"
modelsel
```

Kechikish operatorlari bitta bayonotda L(0/1).inf. qaram o'zgaruvchi va avtoregressiv mustaqil o'zgaruvchilarni ko'rsatish orqali o'zlarining to'liq afzalliklaridan foydalanilmoqda. Ushbu bayonotda birinchi o'zgaruvchi inflatsiyaning nol kechikishi L(0).inf bo'lib, u shunchaki inf. Demak, L(0/1).inf iborasi inf L.inf ga ekvivalentdir, chunki inf regress dan keyin birinchi bo'lib paydo bo'ladi, Stata uni bog'liq o'zgaruvchi sifatida tan oladi.

Bu buyruq qismidan olingan natijalar:

```
. di "p=1 q=0"
p=1 q=0

. modelsel
      aic = -1.2466292
      sc  = -1.160418
      obs =           85
```

Halqalardan foydalanish model tanlashni ancha osonlashtirishi mumkin. Misol sifatida, keling, p=1,2,3,4,5 va 6 q=0 va 1 uchun barcha mumkin bo'lgan modellarni qidiramiz. Ketma-ket qiymatlar ustida aylanadigan forvalues buyrug'i yordamida ichki halqa hosil qilish mumkin.

Asosiy tuzilma quyidagicha bo'ladi:

```
forvalues q=0/1 {
  forvalues p=1/6 {
    [statements to compute and print]
  }
}
```

q va p ning hisoblangan qiymatlari berilgan diapazonda bo'lsa, sikillar bajariladi (masalan, q uchun 0 va 1 p uchun 1dan 6 gacha). Bu shaklda p va q qiymatlari 1 bosqichga ortadi. Qavslar forvalues bilan belgilanishi kerak va:

- 1). ochiq qavs forvalues bilan bir qatorda paydo bo'lishi kerak;
- 2). ochiq qavsga, albatta, izohlardan tashqari hech narsa amal

qilolmaydi; bajariladigan birinchi buyruq yangi satrda paydo bo'lishi kerak;

3). yaqin qavs o'z-o'zidan chiziqda paydo bo'lishi kerak.

Okun qonunining ARDL (p,q) uchun buyrug'i quyidagicha ko'rinadi:

```
for values q=0/1 {  
  for values p=1/6 {  
    quietly regress L(0/`p').inf L(0/`q').D.u if date >= tq(1988q3)  
    display "p=`p' q=`q'"  
    modelsel  
  }  
}
```

Halqalar ichida hisoblanayotgan bayonotlar haqida bir nechta narsaga e'tibor bering. Birinchidan, p va q endi makro nomlari bilan ataladi. Bu shuni anglatadiki, ular haqida gap ketganda, ular bitta qo'shtirnoq ichiga olinishi kerak (yuqorida qilganimizdek, chap va o'ng). Ikkinchidan bog'liq o'zgaruvchi va avtoregressiv mustaqil o'zgaruvchi yana bitta $L(0/p').inf$ bayonotiga kiritilgan. p soni 1dan 6 gacha oshib, kechikishlar qo'shiladi va p va q ning ekranga chop etilgandan so'ng modelsel dastur bajariladi. Shunday qilib, bir nechta qisqa bayonotlarda ko'plab modellarni ko'rib chiqish mumkin va taqsimlangan kechikishlar tartibini osongina o'zgartirish mumkin. sikllar shu tarzda o'rnatilganda, q sikli noldan boshlanadi, keyin p sikli 1 dan 6 gacha takrorlanadi. Birida p sikli tugaydi, q tsikili 1 ga ortadi va p sikli qaytadan boshlanadi. Agar xohlasangiz, ularning tartibini o'zgartirishingiz mumkin.

9.6.2. Okun qonuni

Okun qonuni vaqtli qatorlar modelining adekvat spetsifikatsiyasini izlash uchun yana bir imkon beradi. okun.dta ma'lumotlarini yuklang, 1985q2 dan boshlangan sanalarni yarating, ularni satrlar sifatida chop etish uchun formatlang va ma'lumotlarni vaqtli qatorlar deb e'lon qiling:

```
use okun, clear  
generate date = tq(1985q2) + _n-1
```

format %tq date

tsset date

9.2-bo'limda taxmin qilingan model ARDL (0,2) bo'lib, ishsizlik darajasining o'zgarishini YaIM o'sishi bilan bog'laydi. Model ostida eng kichik kvadratlar bilan baholanadi, korrelogramm olinadi va 5 tagacha avtokorrelatsiya qilingan qoldiqlarni o'z ichiga olgan modellar uchun LM statistikasi ishlab chiqariladi:

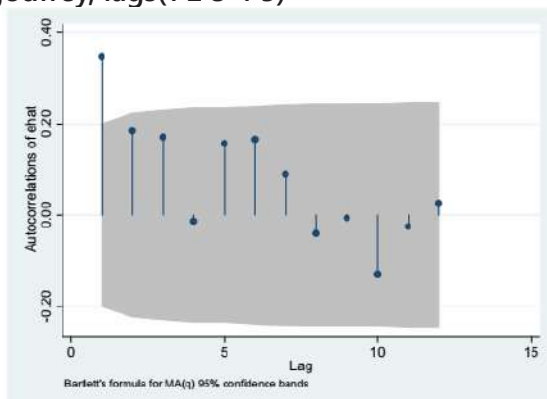
reg D.u g L(1/2).g L.D.u

predict ehat, res

ac ehat, lags(12)

drop ehat

estat bgodfrey, lags(1 2 3 4 5)



. estat bgodfrey, lags(1 2 3 4 5)

Breusch-Godfrey LM test for autocorrelation

lags(<i>p</i>)	chi2	df	Prob > chi2
1	12.364	1	0.0004
2	12.894	2	0.0016
3	13.754	3	0.0033
4	15.228	4	0.0043
5	19.648	5	0.0015

H0: no serial correlation

Korrelogramm bitta muhim avtokorrelatsiyaga ega va LM statistikasining har biri 5% darajasida ahamiyatlidir. Bu, ARDL (0,2) noto'g'ri ko'rsatilganligini ko'rsatadi. Model oxiridagi do faylda Okun ma'lumotlar to'plamidan foydalangan holda bir qator modellarni baholash uchun buyruq berilgan. Namuna oldingi misoldagidek

cheklangan, bu safar 1986-yilning birinchi choragida boshlangan kuzatishlar uchun.

Avtokorrelatsiyalar 0 dan 2 gacha va taqsimlangan kechikishlar 1 dan 3 gacha o'zgaradi; bu taxmin qilish uchun 9 ta modelga olib keladi:

```
for values q=1/3 {
  for values p=0/2 {
    quietly regress L(0/'p').D.u L(0/'q').g if date >= tq(1986q1)
    display "p='p' q='q'"
    modelsel
  }
}
```

Bu quyidagini hosil qiladi:

```
obs = 95
p=2 q=1
a ic = -3.5693074
sc = -3.4348928
obs = 95
p=0 q=2
a ic = -3.4633827
sc = -3.355851
obs = 95
p=1 q=2
a ic = -3.5675498
sc = -3.4331352
obs = 95
p=2 q=2
a ic = -3.5483196
sc = -3.3870221
obs = 95
p=0 q=3
a ic = -3.4424223
sc = -3.3080077
```

aic = -3.5490965

sc = -3.3609161

obs = 95

AIC va SC ni minimallashtiradigan model ARDL (1,1). Ushbu model butun namunadan foydalangan holda baholanadi va xatolar LM statistikasi yordamida qolgan avtokorrelatsiya uchun tekshiriladi.

Quyidagi buyruqlar yordamida amalga oshiriladi:

reg D.u L.D.u L(0/1).g

estat bgodfrey

Natija:

Source	SS	df	MS	Number of obs =	96
Model	5.49727601	3	1.83242534	F(3, 92) =	69.58
Residual	2.42272399	92	.026333956	Prob > F =	0.0000
Total	7.92	95	.083368421	R-squared =	0.6941
				Adj R-squared =	0.6841
				Root MSE =	.16228

D.u	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
u					
LD.	.3501158	.084573	4.14	0.000	.1821466 .518085
g					
--	-.1840843	.0306984	-6.00	0.000	-.245054 -.1231146
L1.	-.0991552	.0368244	-2.69	0.008	-.1722917 -.0260187
_cons	.3780104	.0578398	6.54	0.000	.2631356 .4928853

. estat bgodfrey

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
1	0.170	1	0.6804

H0: no serial correlation

Model qiymatlarida avtokorrelatsiya qolmagan ko'rinadi (p-qiymat=0,68), bu ARDL (1,1) modelining adekvat ekanligini ko'rsatadi.

9.6.3. Avtoregressiv modellar

Avtoregressiv modellar ARDL (p,q) ning maxsus holatlari bo'lishi mumkin. Asosan, AR (p) modeli ARDL (p,0)ga ekvivalentdir. okun.dta da topilgan AQSHning YalM o'sishi haqidagi ma'lumotlar

9.3-bo'limda avtokorrelatsiya uchun tekshiriladi. `g` ning korrelogrammasida vaqt qatori kuzatuvlari o'rtasida korrelatsiya mavjudligini tasdiqlovchi dalillar mavjud edi.

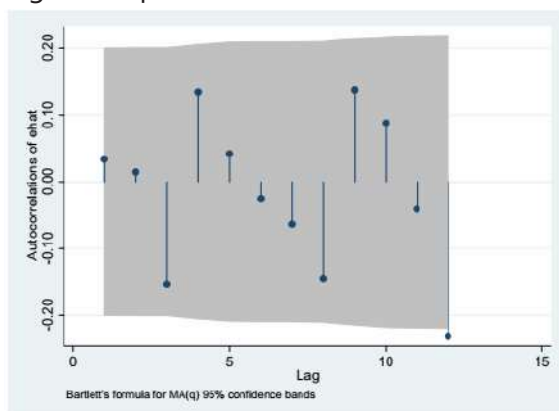
Buni batafsil o'rganish uchun YaLM o'sishi uchun AR(2) modelini taxmin qilamiz va qoldiqlarning korrelogrammasi tuziladi. Dastlabki 12 kechikish uchun avtoregressiya, saqlangan qoldiqlar va avtokorrelatsiyalar taxmin qilinadi:

`reg g L(1/2).g`

`predict ehat, res`

`ac ehat, lags(12)`

Bu quyidagi hosil qiladi:



AR(2) model adekvat ko'rinadi, chunki qoldiq avtokorrelatsiyalarning aksariyati kichik va ahamiyatsizdir. Buni batafsil o'rganish uchun avtoregressiya tartibi 1 dan 5 gacha o'zgarib turadi va model tanlash qoidalari afzal modelni tanlash uchun ishlatiladi, yana model AIC yoki SC ning eng kichik qiymatini ishlab chiqaradi:

`forvalues p=1/5 {`

`qui reg L(0/'p').g if date> tq(1986q2)`

`display "p='p'`

`modelsel`

`} Bu quyidagini hosil qiladi:`

```

p=1
    aic = -1.0935183
    sc = -1.0390538
    obs =          93

p=2
    aic = -1.130582
    sc = -1.0488852
    obs =          93

p=3
    aic = -1.1242025
    sc = -1.0152735
    obs =          93

p=4
    aic = -1.1331587
    sc = -.99699743
    obs =          93

p=5
    aic = -1.1116622
    sc = -.94826871
    obs =          93

```

AIC ning eng kichik qiymatini ishlab chiqaruvchi model AR(4) dir, bunda $AIC = -1,133$ ga teng. Eng kichik SC ishlab chiqaruvchi model AR (2) dir, bunda $SC = -1,049$ ga teng. Bu ikkita model tanlash qoidalari o'rtasidagi muhim farqni ko'rsatadi: SC mezoni regressorni qo'shish uchun katta jarima soladi va AIC ga qaraganda kichikroq modellarni tanlashga intiladi. Bu amaliyotchilar o'rtasida yaxshi ma'lum va tushunarli.

9.7. Prognozlash

Ushbu bo'limda biz 3 xil modellar, AR modeli, ARDL modeli va eksponensial tekislash modelidan foydalangan holda prognozlashni ko'rib chiqamiz. Misollar odatda kelajakda 3 davrgacha bo'lgan qisqa muddatli prognozlashga qaratiladi.

9.7.1. AR modeli bilan prognoz qilish

Aytaylik, bu 2009-yilning 3-choragi va 2009-yilning 3-choragigacha bo'lgan ma'lumotlardan foydalangan holda YaIM o'sishining AR(2) modelini taxmin qildik. Ushbu bo'limda keyingi davrlarni bashorat qilish uchun AR(2) modelidan foydalanish muhokama qilinadi va keyingi uch davrni prognoz qilish muhokama qilinadi hamda prognozning ishonch intervallari yaratiladi.

Uning noma'lum koeffitsiyentlari bo'yicha AR(2) modeli:

$$G_t = \bar{a} + \theta_1 G_{t-1} + \theta_2 G_{t-2} + v_t$$

Oxirgi namunali kuzatishni G_T deb belgilab, vazifa G_{T+1} , G_{T+2} va G_{T+3} ni bashorat qilishadi. Mavjud namunadan keyingi kuzatuvning qiymati:

$$G_{T+1} = \bar{a} + \theta_1 G_T + \theta_2 G_{T-1} + v_{T+1}$$

So'nggi 2 chorak uchun o'sish sur'atlari $G_T = G_{2009q3} = 0,8$ va $G_{T-2} = G_{2009q2} = -0,2$ bo'lib, bu parametrlarning taxminiy qiymatlari bilan $G_{T+1} = G_{2009q4}$ prognozlarini yaratish uchun foydalaniladi.

$$\hat{G}_{T+1} = \hat{a} + \hat{\theta}_1 G_T + \hat{\theta}_2 G_{T-1}$$

$$= 0,46573 + 0,37700 * 0,8 + 0,24624 * (-0,2)$$

$$= 0,7181$$

Model Statada taxmin qilinganidan keyin uni hisoblash oson. AR(2) modelini baholang:

reg g L(1/2).

Keyin yordamida skalyar prognozini hisoblang:

$$\text{scalar ghat1} = _b[\text{cons}] + _b[L1.g]*g[98] + _b[L2.g]*g[97]$$

Stataning indekslash qobiliyati ma'lumotlar to'plamidagi G dagi oxirgi ikkita kuzatishni olish uchun ishlatiladi. Ma'lumotlar 2009q3

da tugaydigan 98 ta kuzatuvdan iborat bo'lgani uchun G_{2009q3}

bo'yicha 98-kuzatuvga ishora qiladi. Xuddi shunday, $g[97]$ 2009q2 dan G dagi kuzatuvga ishora qiladi.

Keyingi prognoz:

$$\hat{G}_{T+2} = \hat{\delta} + \hat{\theta}_1 \hat{G}_{T+1} + \hat{\theta}_2 G_T$$

Yuqoridagi yordamida:

$$\text{scàlâr ghàt2} = _b[_{cons}] + _b[L1.g]*ghàt1 + _b[L2.g]*g[98]$$

$\hat{G}_{T+1}$ ni baholash uchun ghat1 prognozidan foydalanilganiga e'tibor bering. G_T haqiqatda kuzatiladi va 98-kuzatuvdagi ma'lumotlar to'plamida joylashgan.

$$\text{scalar ghat2} = _b[_{cons}] + _b[L1.g]*ghat1 + _b[L2.g]*g[98]$$

Nihoyat, oxirgi prognoz:

$$\hat{G}_{T+3} = \hat{\delta} + \theta_1 \hat{G}_{T+2} + \hat{\theta}_2 \hat{G}_{T+1}$$

Natijada:

$$\text{scalar ghat3} = _b[_{cons}] + _b[L1.g]*ghat2 + _b[L2.g]*ghat1$$

Prognoz $ghat2$ G_{T+2} ni baholash uchun ishlatiladi va bu prognoz butunlay oldingi prognozlarga bog'liq.

Shu tarzda yaratilgan prognozlarning to'liq to'plami:

```
. scalar list ghat1 ghat2 ghat3
      ghat1 = .71807948
      ghat2 = .93343472
      ghat3 = .99445191
```

Prognoz xatosi farqlari:

$$\sigma_1^2 = \text{var}(u_1) = \sigma_v^2$$

$$\sigma_2^2 = \text{var}(u_2) = \sigma_v^2 (1 + \theta_1^2)$$

$$\sigma_3^2 = \text{var}(u_3) = \sigma_v^2 ((\theta_1^2 + \theta_2)^2 + \theta_1^2 + 1)$$

AR(2) modelidagi baholarni almashtirish orqali baholanadi.

$$\text{scalar var} = e(\text{rmse})^2$$

```

scalar se1 = sqrt(var)
scalar se2 = sqrt(var*(1+(_b[L1.g])^2))
scalar se3 = sqrt(var*(_b[L1.g]^2+_b[L2.g]^2+1+_b[L1.g]^2))
scalar list se1 se2 se3

```

```

. scalar list se1 se2 se3
      se1 =  .55268751
      se2 =  .59065984
      se3 =  .62845236

```

95% prognozli ishonch oraliqlari odatiy tarzda tuzilgan. Ular prognoz markazida joylashgan va har ikki yo'nalishda taxminan 2 ta standart og'ishlarni kengaytiradi. Aniqrog'i, ular t distribution 2,5% kritik qiymatidan foydalangan holda hisoblanadi va yuqorida hisoblangan prognoz standart xatolaridan foydalanadi:

```

scalar f1L = ghat1 - invttail(e(df_r),.025)*se1
scalar f1U = ghat1 + invttail(e(df_r),.025)*se1

```

```

scalar f2L = ghat2 - invttail(e(df_r),.025)*se2
scalar f2U = ghat2 + invttail(e(df_r),.025)*se2

```

```

scalar f3L = ghat3 - invttail(e(df_r),.025)*se3
scalar f3U = ghat3 + invttail(e(df_r),.025)*se3

```

```

scalar list f1L f1U f2L f2U f3L f3U

```

Statada aniq kritik qiymatni hisoblash invttail funksiyasidan foydalanadi. Natijalar:

```

. scalar list f1L f1U f2L f2U f3L f3U
      f1L = -.37944839
      f1U =  1.8156073
      f2L = -.23949866
      f2U =  2.1063681
      f3L = -.25352994
      f3U =  2.2424338

```

9.7.2. Ekspontensial tekislash

O'zgaruvchining kelajakdagi qiymatini uning tarixiga asoslangan

holda bashorat qilish uchun ishlatiladigan yana bir mashhur model eksponensial tekislashdir. AR modeli bilan prognoz qilish kabi, eksponensial tekislash yordamida bashorat qilish boshqa o'zgaruvchilardan ma'lumotdan foydalanmaydi.

Asosiy g'oya shundan iboratki, keyingi davr uchun prognoz joriy davr uchun prognozning o'rtacha og'irligi va joriy davrda real realizatsiya qilingan qiymatdir:

$$\hat{y}_{T+1} = \alpha y_T + (1-\alpha) \hat{y}_T$$

Ekponensial tekislash usuli ko'p qirrali prognozlash vositasidir, ammo $\hat{y}_{T+1}$ ni yaratish uchun tekislash parametri α va $\hat{y}_T$ uchun qiymat kerak bo'ladi. α qiymati joriy ma'lumotlarning nisbiy og'irligi haqidagi fikrni aks ettirishi mumkin; muqobil ravishda within-sample forecasts olish orqali tarixiy ma'lumotlardan taxmin qilish mumkin:

$$\hat{y}_t = \alpha y_{t-1} + (1-\alpha) \hat{y}_{t-1} \quad t=2,3,\dots,T$$

va kvadratlar yig'indisini minimallashtiradigan α qiymatini tanlash one-step forecast errors:

$$v_t = y_t - \hat{y}_t = y_t - (\alpha y_{t-1} + (1-\alpha) \hat{y}_{t-1})$$

α ning kichikroq qiymatlari prognozning yanada tekislanishiga olib keladi. Stata dasturida tssmooth vaqtli qatorlar uchun tekislashning turli shakllarini amalga oshiradigan tartib mavjud. tssmooth yangi o'zgaruvchini newvar ko'rinishida yaratadi va o'zgaruvchini so'ralgan silliqlash vositasidan o'tkazish orqali uni to'ldiradi. Eksponensialni o'z ichiga olgan bir nechta silliqlash vositalari mavjud. Bir marta tekislash parametrining kerakli qiymatini belgilashi mumkin, α yoki uning qiymati POE4 adabiyotda muhokama qilingandek, namunadagi kvadratlarining yig'indisini bashorat qilish xatolarini minimallashtirish uchun avtomatik ravishda tanlanishi mumkin.

Quyida, okun.dta ma'lumotlari YaIM o'sishining eksponensial tekislangan prognoz qiymatlarini olish uchun ishlatiladi. Avval

ma'lumotlar ochiladi, yaratilgan sanalar qayta formatlanadi va o'zgaruvchilar vaqtli qatorlar sifatida o'rnatiladi:

use okun, clear

generate date = tq(1985q2) + _n-1

format %tq date

tsset date

Tekislashdan g oldin qilinadigan birinchi narsa vaqt qatoriga (bo'sh) kuzatuvni qo'shishdir. Buni tekislashdan oldin qilish Stataga ushbu kuzatuvni bir qadam oldinga bashorat bilan to'ldirishga imkon beradi:

tsappend, add(1)

g qatorni eksponent tekislash buyrug'i:

tssmooth exponential sm1=g, parms(.38)

Quyida keltirilgan tssmooth sintaksisi biroz tushuntirishga loyiqdir:

Syntax

tssmooth *smoother* [*type*] *newvar* = *exp* [*if*] [*in*] [, ...]

Smoother category	<i>smoother</i>
Moving average	
with uniform weights	<u>ma</u>
with specified weights	<u>ma</u>
Recursive	
exponential	<u>exponential</u>
double exponential	<u>dexponential</u>
nonseasonal Holt-winters	<u>hwinters</u>
seasonal Holt-winters	<u>shwinters</u>
Nonlinear filter	<u>nl</u>

Ko'rsatish kerak bo'lgan birinchi narsalar kerakli tekislash turi hisoblanadi. Bu yerda biz eksponensialni tanlaymiz. Keyin yangi o'zgaruvchi nomi yaratilishi va tekislanadigan qatorga teng bo'lishi kerak (sm1=g). Bundan keyin ba'zi variantlar keladi. Birinchi, parm(.38), tekislash parametrlarining qiymatini o'rnatadi. Agar bu parametr ko'rsatilmagan bo'lsa, u holda tssmooth kvadrat xatolar yig'indisini minimallashtiradiganini tanlaydi.

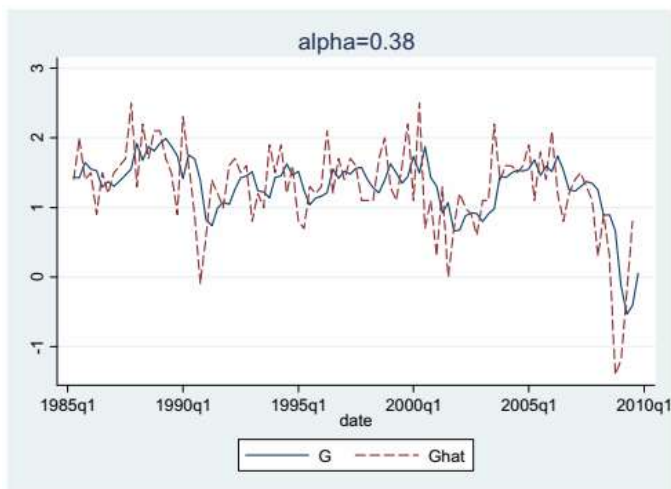
Bu quyidagi natijalarni chiqaradi:

```
. tssmooth exponential sm1=g, parms(.38)
```

```
exponential coefficient =      0.3800
sum-of-squared residuals =    31.122
root mean squared error =    .56354
```

Yangi o'zgaruvchi sm1 eksponent tekislangan qatorni o'z ichiga oladi va u ma'lumotlar to'plamiga qo'shiladi. Tekislangan qator hosil bo'lgach, uni vaqtli qatorlardagi tekislanmagan versiya bilan solishtirish mumkin. Keyingi satrda ikkita seriya chiziladi va rivoyatning nomi o'zgartiriladi, shunda hamma narsa grafikaga biroz yaxshiroq mos tushadi:

```
tsline sm1 g, legend(lab(1 "G") lab(2 "Ghat")) title(alpha=.38)
\\ pattern(solid dash)
```



Keyingi davr uchun prognoz tuziladi va ro'yxatga olinadi. Shuningdek, ma'lumot tssmooth tomonidan 1 davr uchun avtomatik ravishda hisoblanadi. Agar ko'proq so'ralsa, Stata buning uchun variantlarni taklif qiladi. Qo'lda yaratilgan va Statadan avtomatik prognoz mos keladi:

```
scalar f1 = .38*g[98]+(1-.38)*sm1[98]
scalar list f1
```

list sm1 in 99

```
. scalar list f1  
      f1 = .05356533  
  
. list sm1 in 99
```

	sm1
99.	.0535653

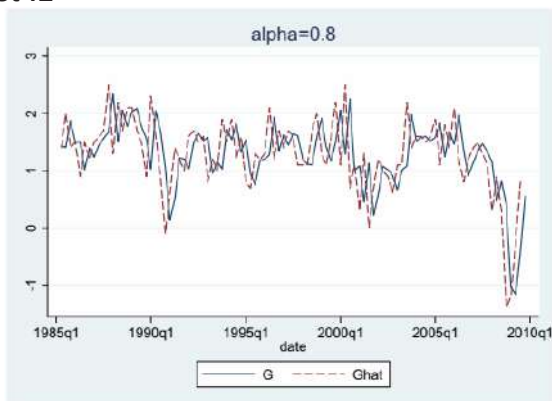
$\alpha = 0,8$ tekislash parametri uchun mashq takrorlanadi. Buyruq va natijalar quyida keltirilgan.

```
tssmooth exponential sm2=g, parms(.8)
```

```
tsline sm2 g, legend(lab(1 "G") lab(2 "Ghat")) title(alpha=.8) \\\  
lpattern(solid dash)
```

```
scalar f2 = .8*g[98]+(1-.8)*sm2[98]
```

```
scalar list f2
```



$\alpha = 0,8$ ning kattaroq qiymati kamroq silliqlikka olib keladi; eksponensial tekislangan qator $\alpha = 0,38$ bo'lganda asl nusxaga ancha yaqinroq. Prognoz , f2, 0,56128444 bu kattaroq tekislash parametri bilan yaratilganidan ancha katta.

Nihoyat, agar yumshatuvchi parametrning o'ziga xos qiymati variant sifatida o'tkazib yuborilsa, Stata namunadagi kvadratlar yig'indisini bashorat qilish xatolarini minimallashtiradigan qiymatni

tanlaydi. Bu Stata yordamida baholanadi:

```
tsmooth exponential sm3=g  
scalar f3 = r(alpha)*g[98]+(1-r(alpha))*sm3[98]  
scalar list f3
```

Natija:

```
computing optimal exponential coefficient (0,1)  
  
optimal exponential coefficient =      0.3803  
sum-of-squared residuals       =     31.122043  
root mean squared error       =     .56353515  
  
. scalar f3 = r(alpha)*g[98]+(1-r(alpha))*sm3[98]  
  
. scalar list f3  
      f3 = .05367152
```

Bu hisob-kitoblarga ko'ra, 0,38 ning belgilangan tanlovi ma'lumotli edi. Tekislash parametrining qiymati tekislashdan keyin $r(\alpha)$ sifatida saqlanadi va u prognozlarni xuddi belgilangan qiymat bilan osonlik bilan yaratish uchun ishlatilishi mumkinligini unutmang.

9.8. Multiplikator tahlili

Multiplikator tahlili bir o'zgaruvchidagi o'zgarishning boshqa o'zgaruvchining natijasiga ta'siri va ta'sir qilish vaqtini bildiradi. Multiplikator tahlilining eng oddiy shakli cheklangan taqsimlangan kechikish modeliga asoslanadi:

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \dots + \beta_q x_{t-q} + e_t$$

Ushbu modeldan hisoblangan koeffitsiyentlar impact, delay va interim multiplikatorlarni ishlab chiqarish uchun ishlatilishi mumkin.

Ta'sir ko'paytmasi x_t ning bir birlikka o'zgarishining y_t o'rtacha qiymatiga ta'siri. x va y bir xil vaqt oralig'ida bo'lgani uchun ta'sir bir vaqtda va shuning uchun o'zgarishning dastlabki ta'siriga teng. S davr delay multiplier hisoblanadi.

$$\frac{\partial E(y_t)}{\partial x_{t-s}} = \beta_s$$

O'tmishdagi x s davrlardagi o'zgarishning joriy davrdagi qaram o'zgaruvchining o'rtacha qiymatiga ta'siri. Agar x_t birlikka oshirilsa va keyingi davrlarda (t+1), (t+2),, yangi darajasida saqlansa, oraliq multiplikatorni hisoblash mumkin. Interim multiplier to'plangan effektni o'lchash uchun darhol ta'sirni (ta'sir multiplikatori) β_0 ni keyingi kechikish ko'paytirgichlariga qo'shadi. Shuning uchun t+1 davrida oraliq effekt $\beta_0 + \beta_1$ ga teng. t+2 davrida $\beta_0 + \beta_1 + \beta_2$ bo'ladi va hokazo. total multiplier q yoki undan ortiq davrlar o'tgandan keyin barqaror o'sishning y ga yakuniy ta'siri; u $\sum_{s=0}^q \beta_s$ bilan berilgan.

ARDL modeli AR modeliga qaram o'zgaruvchining kechikkan qiymatlarini qo'shadi:

$$y_t = \alpha + \theta_1 y_{t-1} + \dots + \theta_p y_{t-p} + \delta_0 x_t + \delta_1 x_{t-1} + \dots + \delta_q x_{t-q} + v_t$$

va bu multiplikator tahlilini biroz qiyinlashtiradi. Asosan, bu kechikish operatori L xususiyatlaridan foydalangan holda cheksiz taqsimlangan kechikish modeliga aylantirilishi kerak, bu xuddi unga asoslangan Stata buyruqlari kabi ishlaydi. Ya'ni, $L^i x_t = x_{t-i}$. Bu modelni tanish AR shakliga kiritadi va multiplikatorlarning odatiy ta'riflari qo'llanilishi mumkin.

Okun qonunini tasvirlash uchun foydalaniladigan ARDL (1,1) modeli uchun bizda mavjud:

$$DU_t = \alpha + \theta_1 DU_{t-1} + \delta_0 G_t + \delta_1 G_{t-1} + v_t$$

yoki kechikish operatori bilan L yozilgan:

$$(1 - \theta_1 L) DU_t = \alpha + (\delta_0 + \delta_1 L) G_t + v_t$$

$$DU_t = (1 - \theta_1 L)^{-1} \alpha + (1 - \theta_1 L)^{-1} (\delta_0 + \delta_1 L) G_t + (1 - \theta_1 L)^{-1} v_t$$

$$DU_t = \alpha + \beta_0 G_t + \beta_1 G_{t-1} + \beta_2 G_{t-2} + \beta_3 G_{t-3} + \dots + e_t$$

$$= \alpha + (\beta_0 + \beta_1 L + \beta_2 L^2 + \beta_3 L^3 + \dots) G_t + e_t$$

Bu shunchaki cheksiz taqsimlangan kechikish modeli.

Multiplikatorlar uchun koeffitsiyentlar β_s ni o'z ichiga oladi, ular ARDL ning taxminiy parametrlari nuqtayi nazaridan hal qilinishi kerak.

Berilgan yechimlar:

$$\beta_0 = \delta_0$$

$$\beta_1 = \delta_1 + \beta_0 \theta_1$$

$$\beta_j = \beta_{j-1} \theta_1 \quad \text{bu uchun} \quad j \geq 2$$

Okun modeli uchun ARDL (1,1) ga asoslangan ta'sirini va dastlabki bir necha oraliq ko'paytirgichlarni baholash uchun Stata buyrug'i:

```
regress D.u L.D.u L(0/1).g
scalar b0 = _b[g]
scalar b1 = _b[L1.D.u]*b0+_b[L1.g]
scalar b2 = b1*_b[L1.D.u]
scalar b3 = b2*_b[L1.D.u]
va boshqalar.
```

Stata ularni grafik qilish uchun ma'lumotlar to'plamiga kiritishning silliq usulini taqdim etadi. Regressiyadan so'ng mult deb nomlangan yangi o'zgaruvchini yarating va β_0 hisoblangan koeffitsiyentini birinchi kuzatuvga qo'ying:

```
gen mult = _b[g] in 1
```

Ikkinchi kuzatuv uchun $\beta_1 = \delta_1 + \beta_0 \theta_1$ hisoblangan qiymatni ikkinchi kuzatishga kiritish uchun almashtirish buyrug'idan foydalaning:

```
replace mult = L.mult*_b[L1.D.u]+_b[L1.g] in 2
```

β_0 ni baholash uchun L.mult ishlatilganiga e'tibor bering.

Qolgan ko'paytirgichlar $\beta_j = \beta_{j-1} \theta_1$ asosida hisoblanadi, uni bitta chiziqli yordamida hisoblash mumkin:

```
replace mult = L.mult*_b[L1.D.u] in 3/8
list mult in 1/8
```

Bu holda L.mult ko'paytirgichlarni o'z ichiga olgan mult ning kechikish qiymati. Agar xohlasangiz ularni T gacha osongina hisoblashingiz mumkin, biz faqat sakkiztasini tanladik. Nihoyat, kechikish og'irliklari (1 dan 8 gacha) sifatida ishlatiladigan butun sonlarni o'z ichiga olgan lag deb nomlangan yangi o'zgaruvchini yarating. Nihoyat, siz ularni chizishingiz mumkin:

gen lag = _n-1 in 1/8

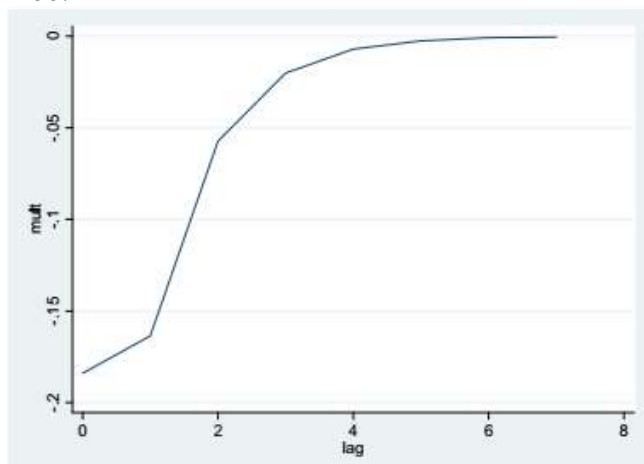
line mult lag in 1/8

Ko'paytiruvchilar quyidagilardir:

`. list mult in 1/8`

	mult
1.	-.1840843
2.	-.163606
3.	-.057281
4.	-.020055
5.	-.0070216
6.	-.0024584
7.	-.0008607
8.	-.0003013

va grafikda:



Dastlabki ta'sir salbiy ko'rinadi, lekin vaqt o'tishi bilan nolga yaqinlashadi. 6-davrga kelib YaIM o'sishining bir birlik o'zgarishining ishsizlikka ta'siri deyarli nolga teng.

9.9. QO'SHIMCHA TESTLAR

9.9.1. Durbin-Watson testi

Durbin-Watson statistikasi regressiyadan keyin estat dwatson yordamida osongina natijalarni chiqaramiz. Phillips ma'lumotlar to'plami uchun DW statistikasi quyidagi buyruq yordamida topiladi:

*** Durbin Watson test**

use phillips_aus, clear

generate date = tq(1987q1) + _n-1

format %tq date

tsset date

regress inf D.u

estat dwatson

Natija esa:

. estat dwatson

Durbin-Watson d-statistic(2, 90) = .8872891

E'tibor bering, Statadagi dwatson testi jadvaldagi yuqori va pastki chegaralarni ko'rib chiqishni talab qiladi. DW ning taqsimlash funksiyasini integratsiyalash natijasida olingan aniq p qiymati hozirgi vaqtda bajarilmaydi.

9.9.2. Prais-Winsten FGLS TESTI

AR(1) modelining amalga oshirilishi mumkin bo'lgan GLS baholovchisi prais deb nomlangan Stata protsedurasi yordamida baholanishi mumkin. Prais buyrug'i regress kabi ishlaydi va shunga o'xshash sintaksisdan foydalanadi. Agar qiziqsangiz o'rganishga arziydigan bir nechta qo'shimcha variantlar mavjud. Praisning eng katta cheklovi shundaki, u faqat birinchi darajali avtokorrelatsiyaga ega modellarni baholaydi. Murakkabroq modellar uchun maksimal ehtimollik yordamida umumiyroq modellarni baholaydigan arima

buyrug'iga qarang. AR(1) xatoliklari bilan chiziqli regressiya modelining ikki bosqichli FGLS bahosi taxmin qilingan.

* Prais-Winsten FGLS estimator

prais inf D.u, twostep

Agar twostep variant berilmasa, hisoblagich barqaror yechim topilguncha takrorlanadi. Ikkala baholovchi ham bir xil asimptotik xususiyarlarga ega, shuning uchun takrorlashning hojati yo'q.

prais inf D.u

Bu quyidagi natijani hosil qiladi:

```
. esttab _2step Iterate, compress se(%12.3f) b(%12.5f) gaps scalars(rss rho) //
> /
      mtitle("2-step" "Iterated") title("Dependent Variable: inf")

Dependent Variable: inf
```

	(1) 2-step	(2) Iterated
D.u	-0.69943** (0.243)	-0.70236** (0.243)
_cons	0.78584*** (0.120)	0.78619*** (0.122)
N	90	90
rss	23.50157	23.49538
rho	0.54988	0.55825

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Birinchi ustun ikki bosqichli FGLS natijalarini, ikkinchisi esa iteratsiya natijalarini o'z ichiga oladi. Natijalar juda o'xshash.

Arima buyrug'i orqali olingan maksimal ehtimollik taxminchisi yordamida baholanadi:

arima inf D.u, ar(1)

Natija:

ARIMA regression						
Sample: 1987q2 - 2009q3			Number of obs		= 90	
Log likelihood = -67.4559			Wald chi2(2)		= 44.96	
			Prob > chi2		= 0.0000	
	inf	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
inf						
u						
o1.		-.7025681	.3167053	-2.22	0.027	-1.323299 -.0818371
_cons		.7861493	.1398032	5.62	0.000	.51214 1.060159
ARMA						
ar						
l1.		.5588218	.0873961	6.39	0.000	.3875285 .7301151
/sigma		.5109273	.0277513	18.41	0.000	.4565358 .5653188

Birinchi blok regressiya taxminlarini o'z ichiga oladi. Kesishish 0,786 ishsizlikning o'zgarishi bo'yicha qiyalik 0,7023 deb baholanmoqda. Avtokorrelatsiya parametri ARMA etiketli qutida

0,559 deb baholanadi. Bu natijalar prais yordamida olingan FGLS baholariga juda o'xshash.

To'qqizinchi bob mavzularini mustahkamlash uchun savollar:

1. Vaqtli qatorlar ma'lumotlarini tahlil qilish deganda nimani tushunasiz?
2. Psevdofunksiya deb nimaga aytiladi?
3. Vaqtli qatorlar bilan ishlashda format buyrug'i qanday vazifani bajaradi?
4. tsset buyrug'i vaqtli qatorlarda qaysi buyruqdan so'ng amalga oshiriladi?
5. okun.dta ma'lumotlari nima uchun xizmat qiladi?
6. Vaqtli qatorlar o'zgaruvchilaridan foydalanishni osonlashtirish uchun xizmat qiluvchi buyruq qaysi?
7. Okun qonuning mazmuni qanday?
8. Ishsizlik darajasi formulasi qanday ifodalanadi?
9. return list buyrug'i qanday vazifani bajaradi?
10. Fillips egri chizig'i nimani ifodalaydi?
11. O'rtacha og'irlik qanday buyruq asosida amalga oshiriladi?
12. Qaysi buyruq orqali ko'rib chiqilayotgan modellarning turli chiziqli xususiyatlari taqqoslanadi?
13. Ekponentsial tekislash usuli nima?
14. Multiplikator tahlili nimani anglatadi?

To'qqizinchi bob mavzularida qo'llanilgan buyruqlar va atamalar

%tq	autocorrelation	correlogram
ac	autoregressive	corrgram
AIC criterion	autoregressive error	D. operator
AR(1) error	autoregressive	delay multiplier
AR(p) model	model	difference operator
ARDL(p,q) model	bandwidth	distributed lags
arima	BIC criterion	e(df_r)

e(N)	interim multiplier	Prais-Winsten
e(r2)	kernel	prais
estat bgodfrey	L(0/4).varname	program
estat dwatson	L. operator	pseudofunctions
exponential	lag length	replace
smoothing	lag operator	sample
finite distributed lag	lagged dependent	autocorrelations
forecast error	variable	SC criterion
forecast standard	LM test	serial correlation
error	macro	TR2 form of LM test
forecasting	multiplier analysis	total multiplier
format	newey	tsline
forvalues	Newey-West	tsset
HAC standard errors	standard errors	tssmooth
impact multiplier	nl	tsvarlist
infinite distributed	nonlinear least	within-sample
lag	squares	forecasts

To'qqizinchi bob mavzularini mustahkamlash uchun masalalarning do-file lari

Ma'lumotlar buyrug'i: `cd c:\data\poe4stata`

1-masala. **log file ni ochish va datalardan foydalanish**

`log using chap09, replace text`

`use okun, clear`

`generate date = tq(1985q2) + _n-1`

`list date in 1`

`format %tq date`

`list date in 1`

`tsset date`

`label var u "% Unemployed"`

`label var g "% GDP growth"`

`tsline u g, lpattern(solid dash)`

`list date u L.u D.u g L1.g L2.g L3.g in 1/5`

`list date u L.u D.u g L1.g L2.g L3.g in 96/98`

```

regress D.u L(0/3).g
regress D.u L(0/2).g
summarize g
return list
scatter g L.g, xline(`r(mean)') yline(`r(mean)')
ac g, lags(12) generate(ac_g)
* approximate z scores
gen z=sqrt(e(N))*ac_g
list ac_g z in 1/12
use phillips_aus, clear
generate date = tq(1987q1) + _n-1
format %tq date
tsset date
tsline inf
tsline D.u
reg inf D.u
predict ehat, res
ac ehat, lags(12) generate(rk)
list rk in 1/5
* -----
* Corrgram
* -----
corrgram ehat, lags(5)
di "rho1 = " r(ac1) " rho2 = " r(ac2) " rho3 = " r(ac3)
drop rk ehat
* LM tests for AR(1) and AR(4) alternatives
reg inf D.u
predict ehat, res
regress inf D.u L.ehat
test L.ehat
* LM test for AR(1)
quietly regress ehat D.u L.ehat
di "Observations = " e(N) " and TR2 = " e(N)*e(r2)
* LM test for AR(4)
quietly regress ehat D.u L(1/4).ehat

```

```

di "Observations = " e(N) " and TR2 = " e(N)*e(r2)
drop ehat
* Using the built-in bgodfrey command to test the
* AR(1) and AR(4) alternatives
regress inf D.u
predict ehat, res
estat bgodfrey, lags(1)
estat bgodfrey, lags(4)
* Replacing ehat(1) with zero and computing LM
replace ehat = 0 in 1
regress inf D.u L.ehat
test L.ehat
quietly regress ehat D.u L.ehat
di "Observations = " e(N) " and TR2 = " e(N)*e(r2)
drop ehat
* Getting Stata to use 90 observations for the LM test
reg inf D.u
predict ehat, res
* Using all observations for bgodfrey test
set obs 94 // add 3 observations to data
gsort -date // moves missing observations to end
replace date = date[_n-1] - 1 if missing(date) // creates dates
for missing obs
replace ehat = 0 if missing(ehat) // puts zeros in for missing
ehats
sort date // re-sort data into ascending order
regress ehat D.u L(1/4).ehat
di "Observations = " e(N) " and TR2 = " e(N)*e(r2)
2-masala. Fillips egri chizig`i.
use phillips_aus, clear
generate date = tq(1987q1) + _n-1
format %tq date
tsset date
scalar B = round(4*(e(N)/100)^(2/9))
scalar list B

```

```

regress inf D.u
estimates store Wrong_SE
newey inf D.u, lag(4)
estimates store HAC_4
esttab Wrong_SE HAC_4, compress se(%12.3f) b(%12.5f) gaps
///
scalars(r2_a rss aic) title("Dependent Variable: inf") ///
mtitles("LS" "HAC(4)")
* -----
* Nonlinear least squares of AR(1) regression model
* -----
nl (inf = {b1}*(1-{rho}) + {b2}*D.u + {rho}*L.inf - {rho}*{b2}*(L.D.u)),
///
variables(inf D.u L.inf L.D.u)
* To see the coefficient legend use coeflegend option
nl (inf = {b1}*(1-{rho}) + {b2}*D.u + {rho}*L.inf - {rho}*{b2}*(L.D.u)),
///
variables(inf D.u L.inf L.D.u) coeflegend
scalar delta = _b[b1:_cons]*(1-_b[rho:_cons])
scalar delta1 = -_b[rho:_cons]*_b[b2:_cons]
* -----
* More general model
* -----
regress inf L.inf D.u L.D.u
estimates store General
scalar list delta delta1
testnl _b[L.D.u] = -_b[L.inf]*_b[D.u]
regress inf L.inf D.u
estimates store No_LD_u
regress inf D.u
estimates store Original
esttab General No_LD_u Original, compress se(%12.3f) b(%12.5f)
///
gaps scalars(r2_a rss aic)
* ARDL

```

```

regress inf L.inf L(0/1).D.u
estimates store AR1_DL1
regress inf L.inf D.u
estimates store AR1_DL0
esttab AR1_DL1 AR1_DL0, compress se(%12.3f) b(%12.5f) ///
gaps scalars(r2_a rss aic)
* Model selection program computes aic and sc
* To remove it from memory use:
* program drop modelsel
capture program drop modelsel
program modelsel
scalar aic = ln(e(rss)/e(N))+2*e(rank)/e(N)
scalar sc = ln(e(rss)/e(N))+e(rank)*ln(e(N))/e(N)
scalar obs = e(N)
scalar list aic sc obs
end
quietly regress inf L.inf L(0/1).D.u
modelsel
quietly regress inf L.inf L.D.u
modelsel
* Residual correlogram and graph
quietly regress inf L.inf D.u
predict ehat, res
corrgram ehat, lags(12)
ac ehat, lags(12)
estat bgodfrey, lags(1 2 3 4 5)
drop ehat
* Table 9.4 AIC and SC Values for Phillips Curve ARDL model
* Note that regress can be abbreviated to reg and quietly to qui
quietly reg L(0/1).inf D.u if date>= tq(1988q3)
di "p=1 q=0"
modelsel
quietly regress L(0/2).inf D.u if date>= tq(1988q3)
di "p=2 q=0"
modelsel

```

```

quietly regress L(0/3).inf D.u if date>= tq(1988q3)
di "p=3 q=0"
modelsel
quietly regress L(0/4).inf D.u if date>= tq(1988q3)
di "p=4 q=0"
modelsel
quietly regress L(0/5).inf D.u if date>= tq(1988q3)
di "p=5 q=0"
modelsel
quietly regress L(0/6).inf D.u if date>= tq(1988q3)
di "p=6 q=0"
modelsel
qui reg L(0/1).inf L(0/1).D.u if date>= tq(1988q3)
di "p=1 q=1"
modelsel
qui reg L(0/2).inf L(0/1).D.u if date>= tq(1988q3)
di "p=2 q=1"
modelsel
qui reg L(0/3).inf L(0/1).D.u if date>= tq(1988q3)
di "p=3 q=1"
modelsel
qui reg L(0/4).inf L(0/1).D.u if date>= tq(1988q3)
di "p=4 q=1"
modelsel
qui reg L(0/5).inf L(0/1).D.u if date>= tq(1988q3)
di "p=5 q=1"
modelsel
qui reg L(0/6).inf L(0/1).D.u if date>= tq(1988q3)
di "p=6 q=1"
modelsel
* Table 9.4 AIC and SC Values for Phillips Curve ARDL model
* Here is the entire thing again, using nested loops
forvalues q=0/1 {
forvalues p=1/6 {
quietly regress L(0/`p').inf L(0/`q').D.u if date >= tq(1988q3)

```

```
display "p=`p' q=`q'"
```

```
modelsel
```

```
}
```

```
}
```

```
* Using var to estimate ARDL
```

```
* Disadvantage: No estat after the procedure
```

```
var inf in 7/91, lags(1/3) exog(L(0/1).D.u)
```

```
3-masala. ARDL modellari.
```

```
use okun, clear
```

```
generate date = tq(1985q2) + _n-1
```

```
format %tq date
```

```
tsset date
```

```
* Estimate the ARDL(0,2)
```

```
* Generate the correlogram and test for autocorrelation
```

```
reg D.u L(0/2).g
```

```
predict ehat, res
```

```
ac ehat, lags(12)
```

```
drop ehat
```

```
estat bgodfrey, lags(1 2 3 4 5)
```

```
* Model Selection for Okun's Law model
```

```
forvalues q=1/3 {
```

```
forvalues p=0/2 {
```

```
quietly regress L(0/`p').D.u L(0/`q').g if date >= tq(1986q1)
```

```
display "p=`p' q=`q'"
```

```
modelsel
```

```
}
```

```
}
```

```
reg D.u L.D.u L(0/1).g
```

```
estat bgodfrey
```

```
* Figure 9.11
```

```
reg g L(1/2).g
```

```
predict ehat, res
```

```
ac ehat, lags(12)
```

```
* Table 9.6
```

```
forvalues p=1/5 {
```

```

qui reg L(0/\`p').g if date> tq(1986q2)
display "p=\`p'
modelsel
}
* Forecasting using -arima- instead of -regress-
* which, of course, yields different predictions
arima g, ar(1/2)
tsappend, add(3)
predict ghat, y // for the point estimates
predict ghatse, mse // for the standard error of prediction
* Forecasting with an AR model
reg g L(1/2).g
scalar ghat1 = _b[_cons]+_b[L1.g]*g[98]+ _b[L2.g]*g[97]
scalar ghat2 = _b[_cons]+_b[L1.g]*ghat1+ _b[L2.g]*g[98]
scalar ghat3 = _b[_cons]+_b[L1.g]*ghat2+ _b[L2.g]*ghat1
scalar list ghat1 ghat2 ghat3
scalar var = e(rmse)^2
scalar se1 = sqrt(var)
scalar se2 = sqrt(var*(1+(_b[L1.g])^2))
scalar se3 = sqrt(var*((_b[L1.g]^2+_b[L2.g])^2+1+_b[L1.g]^2))
scalar list se1 se2 se3
scalar f1L = ghat1 - invttail(e(df_r),.025)*se1
scalar f1U = ghat1 + invttail(e(df_r),.025)*se1
scalar f2L = ghat2 - invttail(e(df_r),.025)*se2
scalar f2U = ghat2 + invttail(e(df_r),.025)*se2
scalar f3L = ghat3 - invttail(e(df_r),.025)*se3
scalar f3U = ghat3 + invttail(e(df_r),.025)*se3
scalar list f1L f1U f2L f2U f3L f3U
4-masala. Exponensial tekislash.
use okun, clear
generate date = tq(1985q2) + _n-1
format %tq date
tsset date
tsappend, add(1)
tssmooth exponential sm1=g, parms(.38)

```

```

    tsline sm1 g, legend(lab (1 "G") lab(2 "Ghat")) title(alpha=0.38)
lpattern(solid dash)
    scalar f1 = .38*g[98]+(1-.38)*sm1[98]
    scalar list f1
    list sm1 in 99
    tssmooth exponential sm2=g, parms(.8)
    tsline sm2 g, legend(lab (1 "G") lab(2 "Ghat")) title(alpha=0.8)
lpattern(solid dash)
    scalar f2 = .8*g[98]+(1-.8)*sm2[98]
    scalar list f2
    tssmooth exponential sm3=g
    scalar f3 = r(alpha)*g[98]+(1-r(alpha))*sm3[98]
    scalar list f3
    list sm3 in 99
    program drop modelsel
    drop sm1 sm2 sm3
5-masala. Durbin Watson testi.
    use phillips_aus, clear
    generate date = tq(1987q1) + _n-1
    format %tq date
    tsset date
    regress inf D.u
    estat dwatson
    * Prais-Winsten FGLS estimator
    prais inf D.u, twostep
    estimates store _2step
    prais inf D.u
    estimates store lterate
    esttab _2step lterate, compress se(%12.3f) b(%12.5f) gaps
    scalars(rss rho) ///
    mtitle("2-step" "Iterated") title("Dependent Variable: inf")
    * AR(1) using arima
    arima inf D.u, ar(1)
    log close

```

X BOB. TASODIFIY REGRESSORLAR VA MOMENTGA ASOSLANGAN BAHOLASH

10.1. Ish haqi tenglamasining eng kichik kvadratlarini baholash

Principle of Econometrics 4-nashr adabiyotida X 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 if lfp==0  
summarize wage educ exper
```

```
. summarize wage educ exper
```

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

In (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		

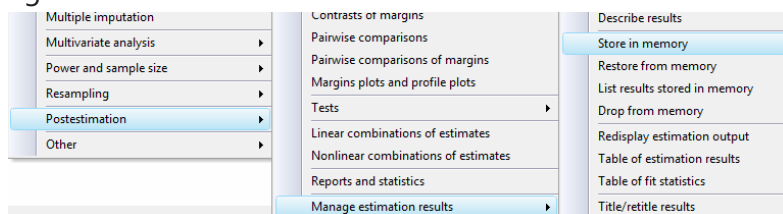
lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	.1074896	.0141465	7.60	0.000	.0796837	.1352956
exper	.0415665	.0131752	3.15	0.002	.0156697	.0674633
exper2	-.0008112	.0003932	-2.06	0.040	-.0015841	-.0000382
_cons	-.5220406	.1986321	-2.63	0.009	-.9124667	-.1316144

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

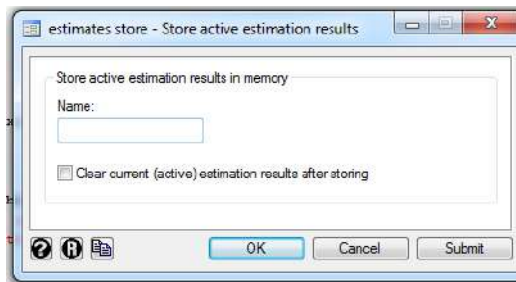
Ushbu post-estimation buyrug'i natijalarni Stata xotirasida saqlaydi va keyinroq foydalanish uchun chiqarilishi mumkin. Hisob-kitoblarni "eng kichik kvadratlar" uchun "ls" nomi ostida saqlang. Buyruq quyidagicha:

```
estimates store ls
```

Buyruqni unutib qoldirgan bo'lsangiz Statistics > Postestimation ni tanlang. So'ngra Manage estimation results > Store in memoryga 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
Residual	1889.65843	424	4.45674158
Total	2230.19626	427	5.22294206

Number of obs = 428
 F(3, 424) = 25.47
 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 testining 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 $\ln(\text{wage})$ tenglamada educ o'rniga tushuntiruvchi o'zgaruvchi sifatida foydalaning:

predict educ_hat

reg lwage educ_hat exper exper2

Olingan koeffitsiyent 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.

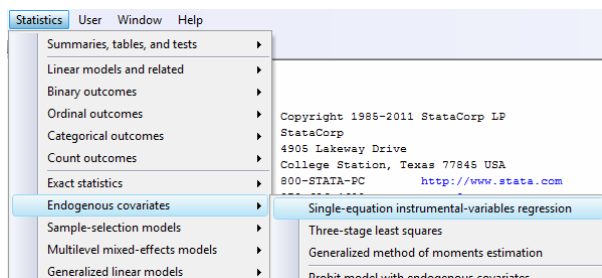
. reg lwage educ_hat exper exper2

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	

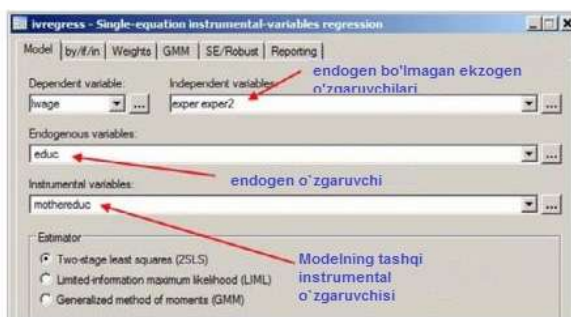
lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ_hat	.0492629	.0390562	1.26	0.208	-.0275049	.1260308
exper	.0448558	.0141644	3.17	0.002	.0170147	.072697
exper2	-.0009221	.000424	-2.17	0.030	-.0017554	-.0000887
_cons	.1981861	.4933427	0.40	0.688	-.7715157	1.167888

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**)

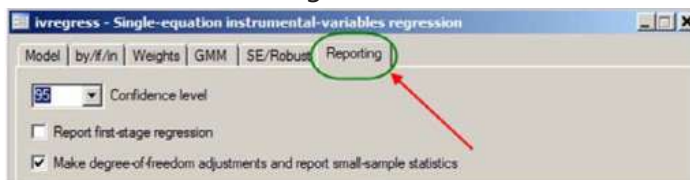
Bu yerda:

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.

Koeffitsiyent baholari IV baholar bo'lib, standart xatolar to'g'ri hisoblab chiqilgan. Xabar qilingan test statistikasi "z" bilan belgilanadi, chunki IV baholovchilar asimptotik xususiyatga 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'rileydigan, 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

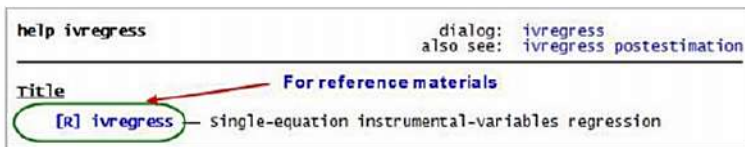
E'tibor bering, bu buyruqda biz erksiz o'zgaruvchidan keyin (**educ=mothereduc**) qo'yganmiz. Ta'kidlanganidek, bu joylashtirish tadqiqotchining ixtiyorida.

Instrumental variables (2SLS) regression

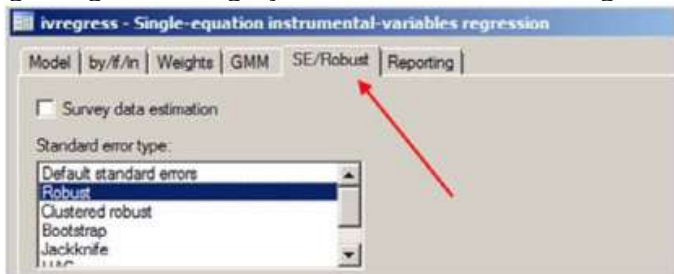
Source	SS	df	MS			
Model	27.4983827	3	9.16612758	Number of obs =	428	
Residual	195.829059	424	.461860988	F(3, 424) =	7.35	
Total	223.327442	427	.523015086	Prob > F =	0.0001	
				R-squared =	0.1231	
				Adj R-squared =	0.1169	
				Root MSE =	.6796	

	Twage	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

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 ilq'or va matritsa algebrasidan foydalanadi.



Mroz ma'lumotlari kabi kesa ma'lumotlar uchun, biz, shuningdek, ivregress ga vce(robust) variantini qo'shish orqali Oq geteroskedastlikni tuzatishdan foydalangan holda, geteroskedastlik 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
Residual	1838.7191	424	4.33660166
Total	2230.19626	427	5.22294206

Number of obs = 428
 $F(3, 424) = 30.09$
 $\text{Prob} > F = 0.0000$
 $R\text{-squared} = 0.1755$
 $\text{Adj } R\text{-squared} = 0.1697$
 $\text{Root MSE} = 2.0825$

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
educ					
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
```

$F(2, 423) = 55.40$
 $\text{Prob} > F = 0.0000$

yoki bu testni geteroskedastikaga chidamli qilish uchun
reg educ exper exper2 mothereduc fathereduc, vce(robust)
test mothereduc fathereduc

Natija:

```
. test mothereduc fathereduc
```

```
( 1) mothereduc = 0
( 2) fathereduc = 0
```

$F(2, 423) = 49.53$
 $\text{Prob} > F = 0.0000$

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 asboblارimiz 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

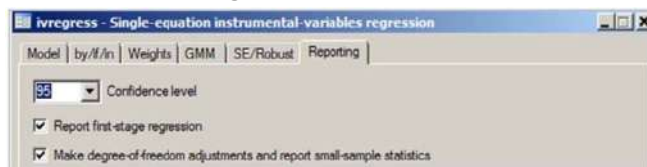
Source	SS	df	MS	
Model	30.3074259	3	10.1024753	
Residual	193.020016	424	.455235886	
Total	223.327442	427	.523015086	

Number of obs = 428
F(3, 424) = 8.14
Prob > F = 0.0000
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

at firststage

st-stage regression summary statistics

Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	F(2,423)	Prob >
educ	0.2115	0.2040	0.2076	55.4003	0.000

imum eigenvalue statistic = 55.4003

tical values	# of endogenous regressors: 1			
Instruments are weak	# of excluded instruments: 2			
s relative bias	5%	10%	20%	30%
	(not available)			
	10%	15%	20%	25%

F qiymatiga Minimum eigenvalue statistic nomi beriladi. Ushbu terminologiya va statistik ma'lumotlarning quyida keltirilgan kritik qiymatlarning foydaliligi ushbu bobning 10.6-bo'limida to'liq tushuntiriladi. IV bahosi yordamida geteroskedastikaga 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

	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 geteroskedastika ostida haqiqiy emas.

10.3.1. Qisman korrelatsiyalarni tasvirlash

O'zgaruvchi kuchini muhokama qilishda eshtilgan iboralardan biri – bu "qisman korrelatsiyalar". 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 korrelatsiyasi 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 korrelatsiya 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 korrelatsiya 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 korrelatsiya 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	Number of obs =	428
Model	329.557956	1	329.557956	F(1, 427) =	74.47
Residual	1889.65843	427	4.42542958	Prob > F =	0.0000
Total	2219.21639	428	5.18508502	R-squared =	0.1485
				Adj R-squared =	0.1465
				Root MSE =	2.1037

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

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

Korrelatsiyalar va kovarians o'rtasidagi bog'liqlik yuqoridagi regressiya koeffitsiyentini 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 koeffitsiyenti va korrelatsiyani hisoblashimiz mumkin.

di "partial LS coefficient = " $r(\text{cov_12})/r(\text{Var_2})$

di "partial correlation = " $r(\text{cov_12})/\sqrt{r(\text{Var_2}) * r(\text{Var_1})}$

```
. 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, $y = \beta_1 + \beta_2 x + \delta v + e$. Ushbu sun'iy regressiyani 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		
				Prob > F = 0.0000		
				R-squared = 0.1624		
				Adj R-squared = 0.1544		
Total	223.327442	427	.523015086	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 geteroskedastiklikka 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
```

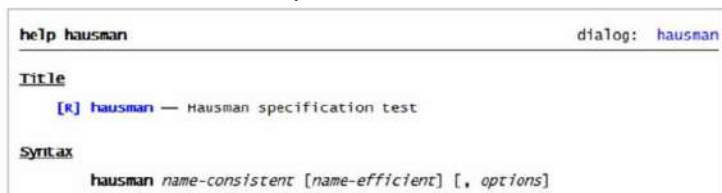
```
Number of obs = 428
F( 4, 423) = 21.52
Prob > F = 0.0000
R-squared = 0.1624
Root MSE = .66502
```

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

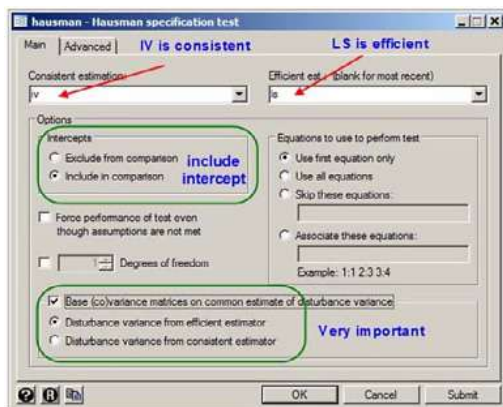
Ushbu testlar educ ning taxminan 10% ahamiyatga ega bo'lgan

darajada endogen ekanligini ko'rsatadi.

Biz aksariyat hollarda regressiyaga asoslangan sinovni afzal ko'ramiz. Stataning "avtomatik" Hausman testini amalga oshirish uchun biz avval saqlangan instrumental o'zgaruvchilar (1v) va eng kichik kvadratlar (1s) 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, u eng yaxshi chiziqli xolis 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 geteroskedastlik sharoitida haqiqiy emas, chunki test eng kichik kvadratlar hisoblagichining

samarali bo'lishiga asoslanadi. Agar geteroskedastlik mavjud bo'lsa, eng kichik kvadratlar samarali emas, chunki Gauss-Markov teoremasi bajarilmaydi. Bu geteroskedastik 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 "1s" 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 σ^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 nuqtayi nazaringizdan 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.

```
chi2(1) = (b-B)'[(v_b-v_B)^(-1)](b-B)
          = 2.78
Prob>chi2 = 0.0954
```

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 haqiqiylik 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)  
expres exper2,
```

```
small
```

```
predict ehat, residuals
```

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

```
reg ehat exper exper2 mothereduc fathereduc
```

Qaysi elementlar regressiyadan saqlanganligini eslash uchun ereturn list dan foydalaning. Keyin NR^2 ni hisoblang:

```
ereturn list
```

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

Invchi2 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^2 from artificial regression = " e(r2)

di "NR^2 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 "R^2 from artificial regression = " e(r2)  
R^2 from artificial regression = .00088334
```

```
. di "NR^2 test of overidentifying restriction = " nr2  
NR^2 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
```

Statadan 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, **NR²**, 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)  
Basmann 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 bizda 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 korrelatsiyalar - bu ikki o'zgaruvchi o'rtasidagi korrelatsiya haqidagi odatiy tushunchaning umumlashtirilishi va ikkita o'zgaruvchi to'plami o'rtasidagi bog'lanishni tavsiflashga urinishdir. Kanonik korrelatsiyalarni 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 korrelatsiya 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 korrelatsiya, $r_1 \geq r_2 \geq \dots \geq r_B$, r_B minimal kanonik korrelatsiya. Zaif identifikatsiya uchun test, o'zgaruvchilar endogen regressorlar bilan korrelatsiya qilinganda yuzaga keladigan vaziyat, lekin faqat zaif, Cragg-Donald F test statistikasiga asoslanadi:

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

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 koeffitsiyentlarini 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 koeffitsiyentlari 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'lmagimiz 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 Stataga kiritilgan. ivregressdan keyin estat firststage ishlatilsa, ushbu bobning

10.3-bo'limida ko'rsatilganidek, ushbu muhim qiymatlar xabar qilinadi. Qadamlar quyidagilardir:

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.

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:

$$\hat{y} = \beta_1 + \beta_2 MTR + \beta_3 EDUC + \beta_4 KIDSL6 + \beta_5 NWIFE1$$

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 marjinal soliq stavkasi, shu jumladan ijtimoiy sug'urta soliqlari. Bu tenglamada MTR, KIDSL6 va NWIFEINCda koeffitsiyent belgilari manfiy va EDUC koeffitsiyent noaniq belgiga ega.

Marjinal soliq stavkasi MTR va ta'lim EDUCni endogen deb hisoblang, shuning uchun B=2. L=2 bo'lishi uchun ona va otaning ta'limini, MOTHEREDUC va FATHEREDUC ni o'zgaruvchilar sifatida ishlatib. 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) kids16 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
kids16	-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: kids16 nwifeinc mothereduc fathereduc

Birinchi bosqich natijalari esa:

```
. estat firststage
```

Shea's partial R-squared

Variable	Shea's Partial R-sq.	Adj. Shea's Partial R-sq.
mtr	0.0005	-0.0065
educ	0.0024	-0.0046

Minimum eigenvalue statistic = .100568

Critical Values

Ho: Instruments are weak

of endogenous regressors: 2
of excluded instruments: 2

2SLS relative bias	5%	10%	20%	30%
	(not available)			
2SLS Size of nominal 5% Wald test	10%	15%	20%	25%
	7.03	4.58	3.95	3.63
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 XI bobda tushuntiriladi. Shuningdek, POE4 adabiyotidagi 10E.1-jadvalga qarang. Cragg-Donald F testining 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 parametrlar 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 X bob do-faylida 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)
kidl6
```

```
    nwifeinc, small
```

```
    ereturn list
```

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

Kidl 6 va nwifeincning endogen o'zgaruvchilardan va o'zgaruvchilardan ta'sirini yuqoridagi 10.3.1-bo'limda ko'rsatilgan jarayondan foydalanib, qisman olib tashlang:

```
reg mtr kidl6 nwifeinc
```

```
predict mtrr, r
```

```
reg educ kidl6 nwifeinc
```

```
predict educr, r
```

```
reg mothereduc kidl6 nwifeinc
```

```
predict mothereducr, r
```

```
reg fathereduc kidl6 nwifeinc
```

```
predict fathereducr, r
```

Kanonik korrelatsiyalar 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 bosing 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 korrelatsiyalarni topish buyrug'i:

canon (mtrr educr) (mothereducr fathereducr)

Boshqa natijalar mavjud bo'lsa-da, kanonik korrelatsiyalarda ifodalaniladi:

Canonical correlations:
0.4356 0.0218

Saqlangan natijalar va minimal kanonik korrelatsiya 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

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

10.8. Simulatsiya tajribasi

Principles of Econometrics 4-nashri darsligada, 10F.2-ilovada IV/2SLS baholovchisining namuna olish xususiyatlarini ko'rsatish uchun simulatsiya tajribasi o'tkaziladi. Simulatsiyada 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. π parametri instrumental o'zgaruvchi kuchini nazorat qiladi. Agar $\pi = 0$ bo'lsa, instrumental o'zgaruvchilar x bilan korrelatsiya qilinmaydi va instrumental o'zgaruvchilarni baholash muvaffaqiyatsiz bo'ladi. π qanchalik katta bo'lsa, asboblari shunchalik kuchli bo'ladi. Nihoyat, biz x ning endogenligini boshqaradigan p korrelatsiyasi bilan standart normal taqsimotlarga ega bo'lish uchun tasodifiy e va v xatolarini yaratamiz. Agar $p=0$ bo'lsa, x endogen emas. p qanchalik katta bo'lsa, endogenlik shunchalik kuchli bo'ladi. Biz $N=100$ o'lchamdagi 10 000 ta namunani yaratamiz va keyin bir nechta ssenariylar 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 $p = 0$ (x ekzogen) va $p = 0,8$ (x yuqori endogen)ga ruxsat beramiz.

Simulatsiya barcha xotirani tozalash va simulatsiyani 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 qiymatlarni tushuntirish uchun POE4, B ilovasiga qarang. Simulatsiya tajribasining asosiy komponenti e va v xato atamaları o'rtasidagi bog'liqlikdir. Korrelatsiya qilingan tasodifiy sonlarni

```
help drawnorm                                dialog: drawnorm
_____
title
  [D] drawnorm — Draw sample from multivariate normal distribution

Syntax
  drawnorm newvarlist [, options]
```

n(#)	# of observations to be generated
cov(matrix vector)	covariance matrix

Biz kovarian matritsiasidan foydalanamiz, ya'ni:

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

```
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 z2 = rnormal()
```

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

Yuqorida qayd etilgan ma'lumotlarni yaratish jarayonidan foydalanib, biz x va y ni yaratamiz. Xato shartlari σ_{ρ} korrelatsiyasi bilan o'zaro boq'liq:

generate $x = \pi z_1 + \pi z_2 + \pi z_3 + v$

generate $y = x + e$

correlate x e

Biz topgan 100 ta tasodifiy qiymatlar namunaviy korrelatsiyaga ega.

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

	x	e
x	1.0000	
e	0.7960	1.0000

Simulatsiya 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		

	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

Koeffitsiyentning 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		

	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 koeffitsiyentning taxmini ham haqiqiy qiymatga yaqin emasligini ko'ramiz.

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

Natija esa:

. reg y x

Source	SS	df	MS
Model	369.721081	1	369.721081
Residual	40.7364595	98	.415678159
Total	410.45754	99	4.14603576

Number of obs = 100
 F(1, 98) = 889.44
 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

Biz simulatsiya 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
```

Koeffitsiyent 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-bo'limidagi 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. Koeffitsiyent 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. Simulatsiya aslida Stata simulatsiyasi 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. Simulatsiyadan 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
```

W

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. Mseolsning o'rtacha qiymati baholashning "o'rtacha kvadrat xatosi" hisoblanadi. Bu empirik analogidir.

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

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% vaqtini 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 π ni 0,5 ga o'rnatish orqali instrumental o'zgaruvchi uzunligini oshirsak, eng kichik kvadratlarni baholashda biz hech qanday yaxshilanishni topmaymiz.

. tabstat bolshr seolshr rolshr mseolshr hausr, stat(mean sd)

stats	bolshr	seolshr	rolshr	mseolshr	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 b2slsr se2slsr r2slsr mse2sls, stat(mean sd)

stats	b2slsr	se2slsr	r2slsr	mse2sls
mean	1.011116	.11695	.0636	.0139068
sd	.1174081	.0274133	.2440512	.0227001

O'ninchi bob mavzularini mustahkamlash uchun savollar:

1. Regressiya natijalarini saqlash qaysi buyruq orqali amalga oshiriladi?
2. Agar estimates store ls buyrug'i unutilsa qanday amal bajarilishi lozim?

3. Ivregress buyrug'ning imkoniyatlarini to'liq tavsiflash uchun qayday buyruqqa 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 qanday 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?

O'ninchi bob mavzularida qo'llanilgan buyruqlar va atamalar

2SLS	first stage	minimum
canon	regression	eigenvalue statistic
canonical	F-test	MSE
correlations	global	option, sigmamore
chi2tail(n,x)	Hausman test	option, small
correlate	Hausman test,	overidentifying
Cragg-Donald	regression based	instruments
F-test	heteroskedasticity	partial correlation
drawnorm	robust	quietly
drop	instrumental	random regressors
endogenous	variables estimation	rejection rate
covariates	invchi2tail(n,p)	criterion
ereturn list	ivregress	relative bias
estat firststage	LM test	criterion
estat overid	matrix	return list
estimates store	mean squared error	rnormal()

robust standard errors	strong instruments	vce(robust)
Sargan statistic	surplus instruments	Wald chi-square test
seed	tabstat	test
Stock-Yogo critical values	two-stage least squares	weak instruments
	valid instruments	

O'ninchi bob mavzularini mustahkamlash uchun masalalarning do-file lari

Ma'lumotlar buyrug'i: `cd c:\data\poe4stata`

1-masala. log file ni ochish.

`log using chap10_wage, replace text`

* open data and examine

`use mroz, clear`

`describe`

`summarize`

* drop nonworking women and summarize

`drop if lfp==0`

`summarize wage educ exper`

* create variables

`gen lwage = ln(wage)`

`gen exper2 = exper^2`

* Least squares estimation

`reg lwage educ exper exper2`

`estimates store ls`

* IV estimation of wage equation

* using only mothereduc as IV

* first stage regression

`reg educ exper exper2 mothereduc`

* test IV strength

`test mothereduc`

* obtain predicted values

`predict educhat`

* 2sls using 2-stages

`reg lwage educhat exper exper2`

* IV estimation using automatic command

```

ivregress 2sls lwage (educ=mothereduc) exper exper2
ivregress 2sls lwage (educ=mothereduc) exper exper2, small
ivregress 2sls lwage (educ=mothereduc) exper exper2,
vce(robust) small
*Add fathereduc as an IV
* Test fathereduc alone
reg educ exper exper2 fathereduc
* joint first stage regression F-test for weak instruments
reg educ exper exper2 mothereduc fathereduc
test mothereduc fathereduc
reg educ exper exper2 mothereduc fathereduc, vce(robust)
test mothereduc fathereduc
* IV estimation with surplus instruments
ivregress 2sls lwage (educ=mothereduc fathereduc) exper
exper2, small
estimates store iv
* Testing for weak instruments using estat
estat firststage
* IV estimation with robust standard errors
ivregress 2sls lwage (educ=mothereduc fathereduc) exper
exper2, vce(robust) small
estat firststage
* Illustrate partial correlation
ivregress 2sls lwage (educ=mothereduc) exper exper2, small
estat firststage
* partial out exper and exper^2
reg educ exper exper2
predict v1, r
reg mothereduc exper exper2
predict v2, r
* partial correlation
correlate v1 v2
return list
di "partial correlation = "r(rho)^2
* effect of mothereduc on educ controlling for exper and
exper^2
reg v1 v2, noconstant

```

```

* partial correlation
correlate v1 v2, covariance
return list
* calculate partial least squares regression coefficient
di "partial LS coefficient = " r(cov_12)/r(Var_2)
* calculate partial correlation
di "partial correlation = " r(cov_12)/sqrt(r(Var_2)*r(Var_1))
* Hausman test
* reduced form
reg educ exper exper2 mothereduc fathereduc
predict vhat, residuals
* augment wage equation with reduced form residuals
reg lwage exper exper2 educ vhat
reg lwage exper exper2 educ vhat, vce(robust)
* Hausman test automatic
hausman iv ls, constant sigmamore
* Testing surplus moment conditions
* obtain 2sls residuals
quietly ivregress 2sls lwage (educ=mothereduc fathereduc)
exper exper2, small
predict ehat, residuals
* regress 2sls residuals on all IV
reg ehat exper exper2 mothereduc fathereduc
return list
* NR^2 test
scalar nr2 = e(N)*e(r2)
scalar chic = invchi2tail(1,.05)
scalar pvalue = chi2tail(1,nr2)
di "R^2 from artificial regression = " e(r2)
di "NR^2 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
* Using estat
quietly ivregress 2sls lwage (educ=mothereduc fathereduc)
exper exper2, small
estat overid
log close

```

XI BOB. SINXRON TENGLAMALI MODELLAR

11.1. TRUFFLE kamponiyasining taklif va talab modeli

Truffles kamponiyasi uchun talab va taklif modelini ko'rib chiqamiz:

$$\text{Talab: } Q = \alpha_1 + \alpha_2 P + \alpha_3 PS + \alpha_4 DI + e_d$$

$$\text{Taklif: } Q = \beta_1 + \beta_2 P + \beta_3 PF + e_x$$

Q talab tenglamasida ma'lum bir fransuz bozorida sotiladigan truffle miqdori, P – truffelning bozor narxi, PS – haqiqiy truffle o'rnini bosuvchi mahsulotning bozor narxi (boshqa mahsulot juda kam baholanadi), va DI - mahalliy aholining jon boshiga oylik ixtiyoridagi daromadidir. Taklif tenglamasi bozor narxini va taklif qilingan miqdorni o'z ichiga oladi. Shuningdek, u ishlab chiqarish omilining narxini o'z ichiga oladi, PF, bu holda qidiruv jarayonida ishlatiladigan trufflening soatlik ijara narxi. Ushbu modelda biz P va Q endogen o'zgaruvchilar deb faraz qilamiz. Ekzogen o'zgaruvchilar PS, DI, PF va kesishuvchi o'zgaruvchilardir.

Ushbu misol uchun truffles.dta ma'lumotlari faylidan foydalanamiz. Odatiy boshlash buyruqlarini bajaring, jurnalni ishga tushiring va ma'lumotlar faylini oching:

use truffles, clear

describe

Dastlabki 5 ta kuzatuvni sanab o'tish va umumiy statistikani hisoblash orqali ma'lumotlarni tekshiring.

list in 1/5

Natijalar:

. list in 1/5

	p	q	ps	di	pf
1.	29.64	19.89	19.97	2.103	10.52
2.	40.23	13.04	18.04	2.043	19.67
3.	34.71	19.61	22.36	1.87	13.74
4.	41.43	17.13	20.87	1.525	17.95
5.	53.37	22.55	19.79	2.709	13.71

summarize

Variable	Obs	Mean	Std. Dev.	Min	Max
p	30	62.724	18.72346	29.64	105.45
q	30	18.45833	4.613088	6.37	26.27
ps	30	22.022	4.077237	15.21	28.98
di	30	3.526967	1.040803	1.525	5.125
pf	30	22.75333	5.329654	10.52	34.01

11.2. Qisqartirilgan shakldagi tenglamalarni baholash

Qisqartirilgan shakldagi tenglamalar har bir endogen o'zgaruvchini P va Q ni ifodalaydi, ekzogen o'zgaruvchilar PS, DI, PF va kesishuvchi o'zgaruvchi, ortiqcha xato atamasidir. Ular:

$$Q = \pi_{11} + \pi_{21}PS + \pi_{31}PS + \pi_{41}PF + v_1$$

$$P = \pi_{12} + \pi_{22}PS + \pi_{32}PF + v_2$$

Biz bu tenglamalarni eng kichik kvadratlar bilan baholashimiz mumkin, chunki o'ng tomondagi o'zgaruvchilar ekzogen va tasodifiy xatolar bilan bog'liq bo'lmagan. QUANTITY uchun qisqartirilgan shakl yordamida olinadi.

reg q ps di pf

Source	SS	df	MS	Number of obs = 30		
Model	430.382604	3	143.460868	F(3, 26) = 19.97		
Residual	186.754213	26	7.18285434	Prob > F = 0.0000		
Total	617.136817	29	21.2805799	R-squared = 0.6974		
				Adj R-squared = 0.6625		
				Root MSE = 2.6801		

q	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ps	.6564021	.1425376	4.61	0.000	.3634118	.9493923
di	2.167156	.7004738	3.09	0.005	.727311	3.607
pf	-.5069823	.1212617	-4.18	0.000	-.7562392	-.2577254
_cons	7.895099	3.243422	2.43	0.022	1.228151	14.56205

Narx uchun qisqartirilgan shakl esa:

reg p ps di pf

Source	SS	df	MS			
Model	9034.77551	3	3011.59184	Number of obs = 30		
Residual	1131.69721	26	43.5268157	F(3, 26) = 69.19		
				Prob > F = 0.0000		
				R-squared = 0.8887		
				Adj R-squared = 0.8758		
Total	10166.4727	29	350.568025	Root MSE = 6.5975		

p	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ps	1.708147	.3508806	4.87	0.000	.9869017	2.429393
di	7.602491	1.724336	4.41	0.000	4.058068	11.14691
pf	1.353906	.2985062	4.54	0.000	.7403175	1.967494
_cons	-32.51242	7.984235	-4.07	0.000	-48.92425	-16.10059

Keyinchalik foydalanish uchun biz taxmindan keyingi taxmin buyrug'i yordamida narxning o'rnatilgan yoki predict qiymatlarini olamiz. $P^{\wedge}$ ni eslatish uchun phat o'zgaruvchisini nomlang.

predict phat

11.3. 2SLS truffle talabini baholash

Ikki bosqichli eng kichik kvadratlar (2SLS) hisoblarini strukturaviy tenglamalarning o'ng tomonidagi endogen o'zgaruvchini qisqartirilgan shakldagi o'rnatilgan qiymatga almashtirish va keyin eng kichik kvadratlarni qo'llash orqali olish mumkin. Ushbu yondashuv yordamida olingan talab tenglamasining ikki bosqichli eng kichik kvadrat natijalari:

reg q phat ps di

Source	SS	df	MS			
Model	430.382596	3	143.460865	Number of obs = 30		
Residual	186.754221	26	7.18285466	F(3, 26) = 19.97		
				Prob > F = 0.0000		
				R-squared = 0.6974		
				Adj R-squared = 0.6625		
Total	617.136817	29	21.2805799	Root MSE = 2.6801		

q	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
phat	-.374459	.0895643	-4.18	0.000	-.5585611	-.1903569
ps	1.296033	.1930944	6.71	0.000	.8991219	1.692944
di	5.013976	1.241414	4.04	0.000	2.462213	7.56574
_cons	-4.27947	3.013834	-1.42	0.168	-10.47449	1.915554

Ushbu chiqishdagi standart xatolar, t statistik ma'lumotlar va 95% ishonch oraliqlari noto'g'ri, chunki xato dispersiyasi eng kichik

kvadratlar qoldiqlariga asoslangan. 2SLS uchun dasturiy ta'minot buyruqlaridan foydalanish har doim yaxshiroqdir.

Statada ivregress "Instrumental o'zgaruvchilar regressiyasi" deb nomlangan 2SLS baholash uchun o'rnatilgan buyruq mavjud. 2SLS nima uchun chaqirilganini to'liq tushuntirish uchun instrumental o'zgaruvchilar baholovchisini Principle of Econometrics 4-nashr adabiyotining X bobida to'liq ifodalangan. Stata yordami uchun help ivregress buyrug'ni kiriting.

```

help ivregress                                dialog: ivregress
                                              also see: ivregress postestimation

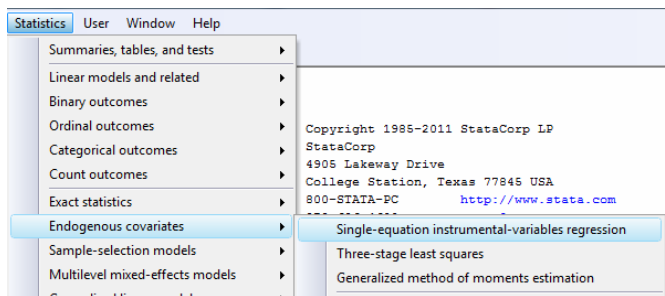
Title
[R] ivregress — single-equation instrumental-variables regression

Syntax
ivregress estimator depvar [varlist1] (varlist2 = varlist_iv) [if] [
in] [weight] [, options]

estimator      description
-----
2sls            two-stage least squares (2SLS)
liml            limited-information maximum likelihood (LIML)
gmm             generalized method of moments (GMM)
  
```

Uni ochiladigan menyular orqali ifodalash ham mumkin. Quyidagi ketma-ketlikni bajaring:

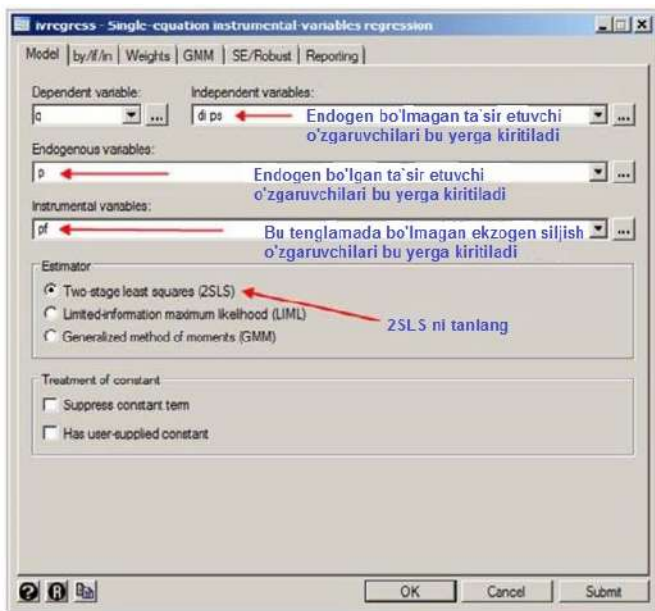
Statistics > Endogenous covariates > Single-equation instrumental-variables regression



Statadagi nomlash konvensiyalari statistik amaliyotchilarning keng segmentiga murojaat qiladi. "Covariates" – bu regressiyani tushuntiruvchi o'zgaruvchilar uchun ishlatiladigan yana bir atama. "Endogenous covariates" o'ng tomondagi tushuntirish

o'zgaruvchilari endogen ekanligini va ular bilan bog'liqligini anglatadigan xato atama. Ushbu turkumda bir nechta tanlov mavjud, ammo biz bitta tenglama bilan ishlayapmiz va tanlov aniq bo'lishi uchun instrumental o'zgaruvchilarni baholashdan foydalanmoqchimiz.

Buyruqlar qatoriga db ivregress ni kiritish yoki yordam oynasi dialog oynasini bosish orqali muloqot oynasiga kirish mumkin: ivregress. Muloqot oynasidagi (keyingi sahifadagi) "instrumental o'zgaruvchilari" talab tenglamasida bo'lmagan ekzogen o'zgaruvchilardir. Bunday holda, bu o'zgaruvchi pf bo'lib, u talab tenglamasida emas, balki taklif tenglamasida ko'rinadigan ishlab chiqarish omilining narxidir. Muloqot oynasidagi "mustaqil o'zgaruvchilar" endogen bo'lmagan o'ng tomondagi o'zgaruvchilardir. Biz 2SLS variantini tanlaymiz va OK tugmasini bosing.



Natijalar oynasida ko'rsatilgan Stata buyrug'i quyidagicha:
ivregress 2sls q di ps (p = pf)

```
. ivregress 2sls q di ps (p = pf)
```

Instrumental variables (2SLS) regression

Number of obs = 30
 Wald chi2(3) = 20.43
 Prob > chi2 = 0.0001
 R-squared = .
 Root MSE = 4.5895

q	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
p	-.3744591	.1533755	-2.44	0.015	-.6750695	-.0738486
di	5.013977	2.125875	2.36	0.018	.847339	9.180615
ps	1.296033	.3306669	3.92	0.000	.6479381	1.944128
_cons	-4.279471	5.161076	-0.83	0.407	-14.39499	5.836052

Instrumented: p
 Instruments: di ps pf

lvregress buyrug'idan so'ng biz qaysi hisoblagichni xohlayotganimizni, bu holda qaysi ekanligini ko'rsatishimiz kerak 2sls. Regressiya modeli spetsifikatsiyasi standart bo'lib, birinchi navbatda qaram o'zgaruvchi q keladi. Endogen bo'lgan har qanday tushuntirish o'zgaruvchisi uchun bizda qavs ichidagi bayonot mavjud, ya'ni:

(varlist2 = varlist_iv)

Bunda

varlist2 o'ng tomondagi barcha endogen o'zgaruvchilar ro'yxati.

varlist_iv modelda mavjud bo'lmagan barcha ekzogen o'zgaruvchilar ro'yxati.

Endogen bo'lmagan izohli o'zgaruvchilar qavs ichidagi ifodadan oldin yoki keyin sanab o'tiladi. Shuningdek, olingan natijalardagi mustaqil o'zgaruvchilarning tartibi ularning dialog oynasiga kiritilish tartibiga bog'liq.

IV regressiya chiqishi z qiymatlari haqida xabar beradi, chunki IV regressiya katta qiymatlar xususiyatiga ega bo'lgan va biz bilamizki, katta qiymatlarda t distribution standart normal N (0,1) taqsimotiga yaqinlashadi. Shunday qilib, t statistika z statistikaga aylanadi. Katta qiymatlarda testlar uchun kritik qiymatlar u yoki bu taqsimotdan kelib chiqishi muhim emas, lekin kichikroq namunalarda bu muhim bo'lishi mumkin. Xulosalarni har doim t taqsimotiga asoslashni afzal ko'ramiz. Bunga Stata-dasmallopsiyasi yordamida erishiladi. Quyidagi buyruqda, shuningdek, endogen o'zgaruvchini (p=pf) bog'liq o'zgaruvchidan keyin qo'yganimizni,

uning oxirida ko'rsatilishi shart emasligini ta'kidlashni unutmang.

ivregress 2sls q (p=pf) ps di, sma11

Natija:

Instrumental variables (2SLS) regression

Source	SS	df	MS
Model	-14.780326	3	-4.92677534
Residual	631.917143	26	24.3045055
Total	617.136817	29	21.2805799

Number of obs = 30
F(3, 26) = 5.90
Prob > F = 0.0033
R-squared = .
Adj R-squared = .
Root MSE = 4.93

q	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
p	-.3744591	.1647517	-2.27	0.032	-.713111 - .0358071
ps	1.296033	.3551932	3.65	0.001	.5659232 2.026143
di	5.013977	2.283556	2.20	0.037	.3200608 9.707893
_cons	-4.279471	5.543884	-0.77	0.447	-15.67509 7.116147

Instrumented: p
Instruments: ps di pf

Natija ikki jihatdan farq qiladi. Oldingi baholashda model ahamiyatining umumiy testi Wald chi kvadrat testiga asoslangan edi. Small varianti bilan umumiy test F testi sifatida xabar qilinadi. Ikkinchi farq shundaki, t qiymatlari xabar qilinadi va p qiymatlari va intervalli taxminlar t taqsimotiga asoslanadi.

Stata shuningdek, ikki bosqichning birinchi bosqichini (qisqartirilgan shakl) ko'rsatadigan variantni o'z ichiga oladi – eng kichik kvadratlar. Bu variantni nomlanishi first bilan.

ivregress 2sls q (p=pf) ps di, small first

First-stage regressions

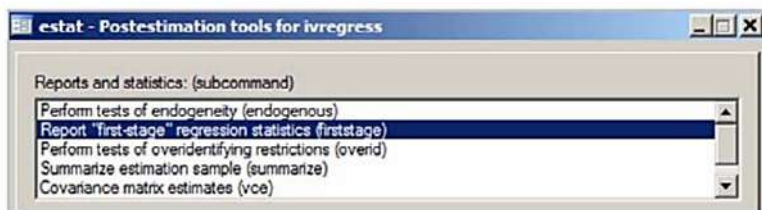
Number of obs = 30
F(3, 26) = 69.19
Prob > F = 0.0000
R-squared = 0.8887
Adj R-squared = 0.8758
Root MSE = 6.5975

p	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
ps	1.708147	.3508806	4.87	0.000	.9869017 2.429393
di	7.602491	1.724336	4.41	0.000	4.058068 11.14691
pf	1.353906	.2985062	4.54	0.000	.7403175 1.967494
_cons	-32.51242	7.984235	-4.07	0.000	-48.92425 -16.10059

Qo'shimcha natija birinchi bosqichdagi regressiya hisoblanadi. Bunday holda, bitta o'ng tomonda endogen o'zgaruvchi mavjud

bo'lganligi sababli, PRICE(p), Stata uning qisqartirilgan shaklini ifodalaydi. Buning sababi shundaki, bu qisqartirilgan shaklda o'zgaruvchi, pf, aslida muhim tushuntirish o'zgaruvchisi ekanligiga dalil bo'lishi kerak. Uning t statistikasi 4,54 ekanligini ko'ramiz. Ushbu adabiyotda ikki bosqichli eng kichik kvadratlarni baholash ishonchli bo'lishi uchun t statistika taxminan 3,3 dan katta bo'lishi yoki o'zgaruvchilarni sinash uchun F qiymati 10 dan katta bo'lishi kerak bo'lgan umumiy qoida mavjud. Ushbu masala bo'yicha batafsilroq ma'lumot olish uchun Principle of Econometrics, 4-nashri darsligining X bobi, E ilovasida va 10.6-bo'limda to'liqroq ifodalangan.

O'garuvchining haqiqiyligini tekshirishni amalga oshiradigan estat firststage deb nomlangan "post baholash" buyrug'i mavjud. U ochiladigan Stata menyusida Statistics > Postestimation > Reports and statistics ketma-ketlikni ham bajarish mumkin. Olingan ro'yxatdan birinchi elementni tanlang.



Natija juda ko'p va siz ko'pchiligini tushunmaysiz, lekin asosiy element qiymati F statistikadir. Bu qiymat bitta tashqi vosita bo'lganligi sababli, 20,5717 qisqartirilgan shakldan t statistikaning kvadratidir.

```
. estat firststage
```

First-stage regression summary statistics

Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	F(1,26)	Prob > F
p	0.8887	0.8758	0.4417	20.5717	0.0001

Keyingi sahifada ko'rsatilgan chiqishning pastki qismida, birinchi bosqichdan boshlab, o'zgaruvchilar kuchli yoki kuchli emasligini tekshirish uchun Stock-Yogo kritik qiymatlari mavjud. Birinchi

bosqich F testi uchun umumiy qiymat qoidasi 10 ga teng, ammo bu aniqlangan. Ushbu muhim qiymatlardan qanday foydalanish haqida tushuntirish uchun ushbu qo'llanmaning 10.6-bo'limida ifodalangan.

Minimum eigenvalue statistic = 20.5717

Critical values	# of endogenous regressors:				1
Ho: Instruments are weak	# of excluded instruments:				1
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	16.38	8.96	6.66	5.53	
	16.38	8.96	6.66	5.53	

Stata buyrug'i:
estat firststage

11.4. 2SLS orqali truffle kompaniyasining taklif modelini baholash

Taklif tenglamasining ikki bosqichli eng kichik kvadratlari xuddi shunday tarzda olinadi. E'tibor bering, bu baho uchun ikkita ekzogen siljish o'zgaruvchilari mavjud, ular ps va di talab tenglamasida o'zgaruvchilar edi:

ivregress 2sls q (p=ps di) pf, small first

Birinchi bosqichni baholash natijalari quyida ko'rsatilgan. Ushbu baholashda tashqi o'zgaruvchilardan kamida bittasi, ps va di siljish o'zgaruvchilari muhim bo'lishi kerak.

First-stage regressions

						Number of obs	=	30
						F(3, 26)	=	69.19
						Prob > F	=	0.0000
						R-squared	=	0.8887
						Adj R-squared	=	0.8758
						Root MSE	=	6.5975
p	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]			
pf	1.353906	.2985062	4.54	0.000	.7403175	1.967494		
ps	1.708147	.3508806	4.87	0.000	.9869017	2.429393		
di	7.602491	1.724336	4.41	0.000	4.058068	11.14691		
_cons	-32.51242	7.984235	-4.07	0.000	-48.92425	-16.10059		

Taklif tenglamasining 2SLS natijalari:

Instrumental variables (2SLS) regression					
Source	SS	df	MS	Number of obs = 30	
Model	556.582251	2	278.291126	F(2, 27) = 95.26	
Residual	60.5545652	27	2.24276167	Prob > F = 0.0000	
Total	617.136817	29	21.2805799	R-squared = 0.9019	
				Adj R-squared = 0.8946	
				Root MSE = 1.4976	

q	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
p	.3379816	.0249196	13.56	0.000	.2868509	.3891123
pf	-1.000909	.0825279	-12.13	0.000	-1.170243	-.831576
_cons	20.0328	1.223115	16.38	0.000	17.52318	22.54243

Instrumented: p
Instruments: pf ps di

Shift o'zgaruvchilarning umumiy ahamiyatini tekshirish uchun biz quyidagi buyruqdan foydalanamiz:
estat firststage

First-stage regression summary statistics

Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	F(2,26)	Prob > F
p	0.8887	0.8758	0.7614	41.4873	0.0000

Asosiy ko'rsatkich F statistikadir. 41,4873 qiymati 10 ga teng bo'lgan chegara qoidasidan kattaroqdir, bu bizni ps va di o'zgaruvchan o'zgaruvchilardan kamida bittasining birinchi bosqich koeffitsiyenti noldan sezilarli darajada farq qilishiga ishonтиради.

11.5. Baliq mahsulotiga bo'lgan taklif va talab modellari

Ikkinchi misolimiz bir vaqtda tenglamalar Nyu-York shahridagi Fulton baliq bozoridan olingan ma'lumotlar bilan modellarni tuzamiz. Bu bozor uchun talab tenglamasini quyidagicha belgilaymiz:

$$\ln(\text{QUAN}_t) = \alpha_1 + \alpha_2 \ln(\text{PRICE}_t) + \alpha_3 \text{MON}_t + \alpha_4 \text{TUE}_t + \alpha_5 \text{WED}_t + \alpha_6 \text{THH}_t + e_t^x$$

Bu yerda QUANT – funtda sotilgan mahsulot miqdori va PRICEt – funt uchun oʻrtacha kunlik narx. Eʼtibor bering, biz maʼlumotlarning vaqtli qatorlar xususiyatiga ega boʻlganligi sababli, biz ushbu munosabat boʻyicha kuzatuvlarni indekslash uchun "t" pastki belgisidan foydalanmoqdamiz. Qolgan oʻzgaruvchilar haftaning kunlari uchun koʻrsatkich oʻzgaruvchilari boʻlib, juma oʻtkazib yuborilgan. α_2 koeffitsiyenti talabning narx egiluvchanligi boʻlib, biz uni manfiy boʻlishini kutgandik. Kundalik koʻrsatkich oʻzgaruvchilari talabning kundan kunga oʻzgarishini aks ettiradi. Taklif tenglamasi:

$$\ln(QUANT_t) = \beta_1 + \beta_2 \ln(PRICE_t) + \beta_3 STORMY_t + e_t^x$$

β_2 koeffitsiyenti taklifning narx egiluvchanligi. STORMY oʻzgaruvchisi oldingi uch kun ichida boʻronli ob-havoni koʻrsatadigan indikator oʻzgaruvchisidir. Bu oʻzgaruvchi taʼminot tenglamasida muhim ahamiyatga ega, chunki boʻronli ob-havo baliq ovlashni qiyinlashtiradi, bozorga olib kelingan baliq taʼminotini kamaytiradi.

Yangi ishchi faylini oching va **fultonfish.dta** maʼlumotlar faylini oching hamda tekshiramiz:

**use fultonfish, clear
describe**

Tizimdagi oʻzgaruvchilar uchun dastlabki 5 ta kuzatuvni roʻyxatga olish orqali maʼlumotlarni tekshiramiz: **list lquan lprice mon tue wed thu stormy in 1/5**

	lquan	lprice	mon	tue	wed	thu	stormy
1.	8.994421	-.4307829	1	0	0	0	1
2.	7.707063	0	0	1	0	0	1
3.	8.350194	.0723207	0	0	1	0	0
4.	8.656955	.247139	0	0	0	1	1
5.	7.844241	.6643268	0	0	0	0	1

Endi ushbu oʻzgaruvchilar uchun umumiy statistik maʼlumotlarni olamiz:

summarize lquan lprice mon tue wed thu stormy
Natijasi:

Variable	Obs	Mean	Std. Dev.	Min	Max
lquan	111	8.52343	.741672	6.194406	9.981374
lprice	111	-.1936811	.3819346	-1.107745	.6643268
mon	111	.1891892	.3934351	0	1
tue	111	.2072072	.4071434	0	1
wed	111	.1891892	.3934351	0	1
thu	111	.2072072	.4071434	0	1
stormy	111	.2882883	.4550202	0	1

11.6. Baliq narxi va miqdori uchun qisqartirilgan shakllar

Bir vaqtning o'zida tenglamalar tizimida har bir endogen o'zgaruvchi uchun qisqartirilgan shakldagi tenglamalarni baholash juda muhimdir. Qisqartirilgan shakldagi tenglamalarni eng kichik kvadratlar bilan baholash mumkin, chunki barcha o'ng tomondagi o'zgaruvchilar ekzogendir. In (QUAN) uchun qisqartirilgan shakl yordamida tenglamani tuzib olamiz:

reg lquan mon tue wed thu stormy

In(PRICE) o'ng tomondagi ta'sir etuvchi o'zgaruvchisi bo'lgani uchun, keling, uning qisqartirilgan shakli tenglamasini batafsilroq ko'rib chiqamiz:

reg lprice mon tue wed thu stormy

Source	SS	df	MS	Number of obs = 111		
Model	2.87047878	5	.574095757	F(5, 105) = 4.58		
Residual	13.1756621	105	.125482496	Prob > F = 0.0008		
Total	16.0461409	110	.145874008	R-squared = 0.1789		
				Adj R-squared = 0.1398		
				Root MSE = .35424		

lprice	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
mon	-.1129225	.1072918	-1.05	0.295	-.3256623	.0998174
tue	-.0411493	.1045087	-0.39	0.695	-.2483707	.1660721
wed	-.011825	.1069299	-0.11	0.912	-.2238473	.2001973
thu	.0496456	.1044582	0.48	0.636	-.1574758	.256767
stormy	.3464055	.0746776	4.64	0.000	.1983337	.4944774
_cons	-.2717054	.076389	-3.56	0.001	-.4231706	-.1202402

Identifikatsiya tushunchasi POE4 adabiyotida to'liqroq muhokama qilingan. 2SLS dan foydalanish uchun tenglamadan chiqarib tashlangan M-1 (M – tenglamalar soni) ekzogon o'zgaruvchilar bo'lishi kerak – bular instrumental o'zgaruvchilardir. Biroq, ular nafaqat ko'rib chiqilayotgan tenglamadan chiqarib

tashlanishi kerak, balki ular qisqartirilgan shaklda statistik ahamiyatga ega bo'lishi kerak.

Talab tenglamasida o'zgaruvchan stormy kiritilmagan, chunki bo'ronlar talabga emas, balki taklifga ta'sir qiladi. In(PRICE) uchun qisqartirilgan shaklda o'zgaruvchan bo'ron 2SLS yaxshi ishlashi uchun juda muhim bo'lishi kerak. E'tibor bering, bo'ron uchun t statistik 4,64 va p qiymati juda kichik. Bu juda yaxshi. Shift o'zgaruvchisining t statistik qiymati uchun asosiy "asosiy qoidasi" chegarasi 3.3 ni tashkil qiladi. Agar t statistika ushbu qiymatdan past bo'lsa, 2SLS juda yaxshi ishlaymasligi mumkin.

Taklif tenglamasida o'zgarmaydiganlar haftaning kunlari: dushanba, seshanba, chorshanba va payshanba kunlari hisoblanadi. Taklif tenglamasini baholash uchun 2SLS dan foydalanish uchun ushbu o'zgaruvchilardan kamida bittasi (juda) muhim bo'lishi kerak. t qiymatlari kichik. Ushbu o'zgaruvchilarning barchasi hech qanday ta'sir ko'rsatmasligi haqidagi qo'shma nol gipotezaning F statistikasi yordamida olinadi.

test mon tue wed thu

Ushbu maxsus sintaksis VI bobdagi test bayonotini muhokama qilishda qo'llanganning yana bir soddalashtirilganidir. Har bir koeffitsiyent nol gipoteza bo'yicha nolga teng bo'lganligi sababli, testdan so'ng oddiygina o'zgaruvchilarni ro'yxatga olishingiz mumkin. Natija esa:

```
. test mon tue wed thu
```

```
( 1) mon = 0  
( 2) tue = 0  
( 3) wed = 0  
( 4) thu = 0
```

```
F( 4, 105) = 0.62  
Prob > F = 0.6501
```

Taklif tenglamasi uchun 2SLS dan foydalanish uchun biz F qiymati 10 dan yuqori bo'lgan juda muhim sinov natijasini qidiramiz. Bu shunday emasligi aniq. Shunday qilib, amaliy jihatdan ta'minot tenglamasi aniqlanmagan va shuning uchun biz ushbu tenglama uchun 2SLS taxminlariga tayanmasligimiz kerak.

11.7. 2SLS orqali baliq mahsulotiga bo'lgan talablarni baholash

Talab tenglamasining 2SLS bahosini olish uchun biz yana Stataning ivregress buyrug'idan foydalanamiz. Ekzogen siljish o'zgaruvchisi bu stormy dir. Qavslar ichida o'ng tomondagi endogen o'zgaruvchi lprice bilan birga ko'rinadi va boshqa ta'sir etuvchi o'zgaruvchilari oldin yoki keyin sanab o'tilgan. Biz z statistik emas, balki t statistik ma'lumotlar ko'rsatilishi uchun kichik variantdan foydalanamiz va biz birinchi bosqich regressiyasini olish uchun birinchi variantdan foydalanamiz, ya'ni narx uchun pasaytirilgan shakl:

ivregress 2sls lquan (lprice=stormy) mon tue wed thu, small first

First-stage regressions

Number of obs = 111
F(5, 105) = 4.58
Prob > F = 0.0008
R-squared = 0.1789
Adj R-squared = 0.1398
Root MSE = 0.3542

lprice	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
mon	-.1129225	.1072918	-1.05	0.295	-.3256623	.0998174
tue	-.0411493	.1045087	-0.39	0.695	-.2483707	.1660721
wed	-.011825	.1069299	-0.11	0.912	-.2238473	.2001973
thu	.0496456	.1044582	0.48	0.636	-.1574758	.256767
stormy	.3464055	.0746776	4.64	0.000	.1983337	.4944774
_cons	-.2717054	.076389	-3.56	0.001	-.4231706	-.1202402

Esda tutingki, stormy t qiymati 4,64 bo'lganida ahamiyatli bo'lib, bu umumiy qiymat 3,3 qoidasidan kattaroqdir. Shuni ham ta'kidlash kerakki, kichik variant p qiymatlari va ishonch oraliqlarini hisoblashni o'zgartiradi, ular ham t distribution ga asoslangan. Ikki bosqichli eng kichik kvadratlar, instrumental o'zgaruvchilar, hisob-kitoblari:

Instrumental variables (2SLS) regression

Source	SS	df	MS
Model	8.41819623	5	1.68363925
Residual	52.0903208	105	.496098293
Total	60.508517	110	.550077427

Number of obs = 111
 F(5, 105) = 4.72
 Prob > F = 0.0006
 R-squared = 0.1391
 Adj R-squared = 0.0981
 Root MSE = .70434

lquan	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lprice	-1.119417	.428645	-2.61	0.010	-1.969341 - .269493
mon	-.0254022	.2147742	-0.12	0.906	-.4512596 .4004553
tue	-.5307694	.2080001	-2.55	0.012	-.9431951 -.1183437
wed	-.5663511	.2127549	-2.66	0.009	-.9882047 -.1444975
thu	-.1092673	.2087866	0.52	0.602	-.3047179 .5232525
_cons	8.505911	.1661669	51.19	0.000	8.176433 8.83539

Instrumented: lprice
 Instruments: mon tue wed thu stormy

Hisoblashdan keyingi estat firststage bosqichi buyrug'ining haqiqiyligini tekshirish uchun ham instrumental o'zgaruvchi stormy dan foydalanish mumkin:

estat firststage

First-stage regression summary statistics

Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	F(1,105)	Prob > F
lprice	0.1789	0.1398	0.1701	21.5174	0.0000

Minimum eigenvalue statistic = 21.5174

Critical values	# of endogenous regressors: 1			
Ho: Instruments are weak	# of excluded instruments: 1			
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	16.38	8.96	6.66	5.53
	16.38	8.96	6.66	5.53

Yana bir bor shuni ta'kidlaymizki, mahsulotning pastki qismida birinchi bosqich F testi uchun Stok-Yogo kritik qiymatlari mavjud va bu muhim qiymatlardan foydalanish POE4 darsligida, X bob, E ilovasi va 10.6-bo'limda tushuntirilgan.

11.8. 2SLS ning alternativlari

Standart IV/2SLS hisoblagichiga muqobillarga har doim katta qiziqish bo'lgan. Principles of Econometrics, 4-nashr adabiyotining, E ilovasida to'liq ifodalangan. Cheklangan ma'lumotlarning maksimal ehtimolligi (LIML) asoschisi birinchi marta 1949-yilda²

Anderson va Rubin tomonidan yaratilgan. Kuchsiz instrumental o'zgaruvchilar mavjud bo'lganda LIMLga ehtiyoj tug'iladi. LIMLning bir qancha modifikatsiyalari Fuller (1977) va boshqalar tomonidan taklif qilingan. Bu baholovchilar k-sinf baholovchilar g'oyasidan foydalangan holda 2SLS bilan birgalikda umumiy tizimda birlashtirilgan. LIML 2SLS hisoblagichiga qaraganda test o'lchamidagi aberatsiyalardan kamroq aziyat chekadi, Fuller modifikatsiyasi esa tarafkashlikdan kamroq aziyat chekadi.

M bir vaqtda tenglamalar tizimida endogen o'zgaruvchilar bo'lsin $y_1, y_2, \dots, y_M$. K ekzogen o'zgaruvchilar bo'lsin $x_1, x_2, \dots, x_K$. Faraz qilaylik, ushbu tizimdagi birinchi struktura tenglamasi:

$$y_1 = \alpha_2 y_2 + \beta_1 x_1 + \beta_2 x_2 + e_1$$

Endogen o'zgaruvchi y_2 qisqartirilgan shaklga ega:

$$y_2 = \pi_{12} x_1 + \pi_{22} x_2 + \dots + \pi_{K2} x_K + v_2 = E(y_2) + v_2.$$

Qisqartirilgan shakl tenglamasining parametrlari doimiy ravishda eng kichik kvadratlar bilan baholanadi, shuning uchun uni quyidagicha ifodalaymiz:

$$\widehat{E(y_2)} = \hat{\pi}_{12} x_1 + \hat{\pi}_{22} x_2 + \dots + \hat{\pi}_{k2} x_k$$

Qisqartirilgan shakl qoldiqlari esa:

$$\hat{v}_2 = y_2 - \widehat{E(y_2)}$$

Ikki bosqichli eng kichik kvadratlar hisoblagichi IV baholovchidan foydalanadi $\widehat{E(y_2)}$ instrumental o'zgaruvchi sifatida. K sinf baholovchisi instrumental o'zgaruvchidan foydalangan holda IV baholovchi hisoblanadi $y_2 - kv_2$. LIML hisoblagichi foydalanadi $k = \hat{e}$ qayerdagi $\hat{e}$ ikki regressiyadan olingan kvadrat qoldiqlar yig'indisining minimal nisbati. To'liq tushunish uchun POE4 qo'llanmaning 468-469-betlarida ma'lumotlar keltirilgan. Ueyn

² Anderson, T.W. and H. Rubin (1949) "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," Annals of Mathematical Statistics, 21, pp. 46-63

Fuller (1977)³ tomonidan taklif qilingan modifikatsiya k sinf qiymatidan foydalanadi:

$$k = \hat{e} - \frac{a}{N - K}$$

Bu yerda K – instrumental o'zgaruvchilarning umumiy soni (qo'shilgan va chiqarib tashlangan ekzogen o'zgaruvchilar) va N - tanlov hajmi. a qiymati doimiy, odatda 1 yoki 4 ga teng.

Mroz ma'lumotlari bilan biz HOURS taklif tenglamasini taxmin qilamiz:

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

Ushbu misol ushbu qo'llanmaning 10.6-bo'limida ishlatilgan. Biz ko'rib chiqayotgan misolda educ va mtr hamda IV mothereduc va fathereduc hamda experience, exper endogen o'zgaruvchilari mavjud.

Yangi ishchi faylni oching va Mroz ma'lumotlarini qayta oching, xotirani tozalang. Misolda ishlatiladigan o'zgaruvchilarni yarating.

use mroz, clear

drop if lfp==0

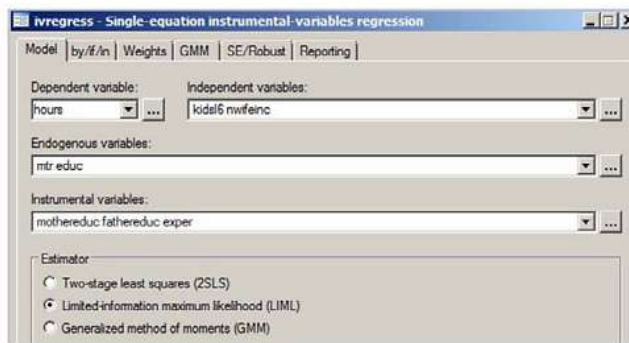
gen lwage=ln(wage)

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

gen exper2 = exper^2

LIML baholari ivregress yordamida olinadi. db ivregress dialog oynasidan foydalanib, cheklangan ma'lumotlarning maksimal ehtimoli (LIML) tugmasini bosing.

³ 2 "Some Properties of a Modification of the Limited Information Estimator," *Econometrica*, 45, pp.939-953.



Stata buyrug'i esa:
ivregress liml hours kidsl6 nwifeinc (mtr educ = mothereduc fathereduc exper)

```
. ivregress liml hours kidsl6 nwifeinc (mtr educ = mothereduc fathereduc exper)
```

Instrumental variables (LIML) regression

Number of obs = 428
Wald chi2(4) = 36.52
Prob > chi2 = 0.0000
R-squared = .
Root MSE = 852.35

hours	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
mtr	-19196.52	3980.227	-4.82	0.000	-26997.62	-11395.42
educ	-197.2591	64.24267	-3.07	0.002	-323.1724	-71.34579
kidsl6	207.5531	162.2957	1.28	0.201	-110.5406	525.6469
nwifeinc	-104.9415	20.56548	-5.10	0.000	-145.2491	-64.63395
_cons	18587.91	3662.026	5.08	0.000	11410.47	25765.35

Instrumented: mtr educ

Instruments: kidsl6 nwifeinc mothereduc fathereduc exper

Kuchsiz instrumental o'zgaruvchilar uchun test yordamida
tekshiriladi.

estat firststage

. estat firststage

Shea's partial R-squared

Variable	Shea's Partial R-sq.	Adj. Shea's Partial R-sq.
mtr	0.0618	0.0529
educ	0.1042	0.0957

Minimum eigenvalue statistic = 8.60138

Critical values # of endogenous regressors: 2
Ho: Instruments are weak # of excluded instruments: 3

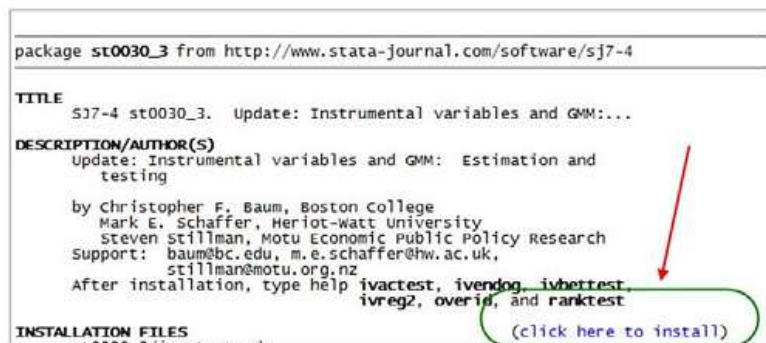
2SLS relative bias	5%	10%	20%	30%
	(not available)			
2SLS Size of nominal 5% Wald test	10%	15%	20%	25%
LIML Size of nominal 5% Wald test	13.43	8.18	6.40	5.45
	5.44	3.81	3.32	3.09

10% maksimal hajmidan foydalanib, endogen o'zgaruvchining koeffitsiyentiga oid gipotezaning nominal 5% sinovi uchun Stata chiqishida kritik qiymat 5,44 sifatida berilgan. Biz Cragg-Donald F test statistikasi (Stata tomonidan Minimal xos qiymat statistikasi deb ataladi) 8,60 dan foydalangan holda instrumental o'zgaruvchilar kuchsizligi haqidagi nol gipotezani rad qilamiz. Agar biz 2SLS/IV hisoblagichidan foydalansak, instrumental o'zgaruvchilar kuchsiz degan gipotezani rad etmagan bo'lardik, chunki kritik qiymat 13,43.

Stata ivregress buyrug'ida Fullerning o'zgartirilgan k sinfi uchun imkoniyat yo'q baholovchi. Biroq, foydalanish uchun foydalanuvchi tomonidan yozilgan tartib mavjud. Stata buyrug'i oynasida findit ivreg2 ni kiriting. Ma'lumotni yordam oynasidan topishingiz mumkin.

```
SJ-7-4 st0030_3 . . . . . Enhanced routines for IV/GMM estimation and testing
. . . . . C. F. Baum, M. E. Schaffer, and S. Stillman
(help ivactest, ivendog, ivhettest, ivreg2, ivreset,
overid, ranktest if installed)
Q4/07 SJ 7(4):465--506
extension of IV and GMM estimation addressing hetero-
skedasticity- and autocorrelation-consistent standard
errors, weak instruments, LIML and k-class estimation,
tests for endogeneity and Ramsey's regression
specification-error test, and autocorrelation tests
for IV estimates and panel-data IV estimates
```

st0030_3 tugmasini bosing. Olingan natijada siz ushbu paketni kompyuterda ma'muriy imtiyozlarga ega bo'lsangiz o'rnatishingiz mumkin.



Buyruqning sintaksisi **ivregress** ga o'xshaydi.

ivreg2 hours (mtr educ = mothereduc fathereduc exper)
kidsl6 nwifeinc,
fuller(1) small

Fullerr(1) variantida 1 a = 1 konstantasini o'rnatishni bildiradi. Natijaning yuqori qismida biz parametr baholarini va baholashda ishlatiladigan k qiymatini, topamiz.

LIML estimation

k=1.00051

lambda=1.00288

Fuller parameter=1

Estimates efficient for homoskedasticity only

Statistics consistent for homoskedasticity only

Total (centered) SS = 257311019.9

Total (uncentered) SS = 983895094

Residual SS = 302240888.2

Number of obs = 428

F(4, 423) = 9.22

Prob > F = 0.0000

Centered R2 = -0.1746

Uncentered R2 = 0.6928

Root MSE = 845.3

hours	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
mtr	-18730.16	3870.958	-4.84	0.000	-26338.87	-11121.45
educ	-191.1248	62.73944	-3.05	0.002	-314.4446	-67.80487
kidsl6	193.2295	159.1413	1.21	0.225	-119.5767	506.0358
nwifeinc	-102.629	20.03279	-5.12	0.000	-142.0052	-63.25276
_cons	18156.78	3560.13	5.10	0.000	11159.04	25154.53

Hisob-kitoblarga qo'shimcha ravishda bizga avtomatik ravishda ko'plab diagnostika natijalari beriladi. Joriy maqsadlar uchun biz faqat kuchsiz instrumental o'zgaruvchilar testi natijalarini xabar

qilamiz. Cragg-Donald F statistik ma'lumotlari nisbiy tarfakashlikka asoslangan mezonlar uchun kritik qiymatlar bilan birga xabar qilinadi.

<u>Weak identification test</u> (Cragg-Donald Wald F statistic):		8.601
Stock-Yogo weak ID test critical values:		
5% maximal Fuller rel. bias		10.83
10% maximal Fuller rel. bias		8.96
20% maximal Fuller rel. bias		7.18
30% maximal Fuller rel. bias		6.15
5% Fuller maximum bias		10.00
10% Fuller maximum bias		8.39
20% Fuller maximum bias		6.79
30% Fuller maximum bias		5.88

NB: Critical values based on Fuller parameter=1
Source: Stock-Yogo (2005). Reproduced by permission.

11.9. Monte Karlo simulatsiya natijalari

Bu rejada biz LIML va k sinf baholovchilarining aspektlarini qo'shib, xuddi shu tajribadan foydalanamiz. Buyruq ushbu bobning oxiridagi XI bobda to'liq berilgan. Tuzilma ushbu qo'llanmaning 10.8-bo'limda tushuntirilgan misol bilan davom ettiramiz. Birinchi qism global boshqaruv parametrlari va ma'lumotlarni yaratish jarayoni o'zgarmagan.

```
clear all
set more off
global numobs 100
global pi 0.5                // reduced form parameter controls
IV strength
global rho 0.8               // rho controls endogeneity
set seed 1234567            // random number seed
set obs $numobs
matrix sig = (1, $rho \ $rho, 1)    // corr(e1,v2)
drawnorm e v, n($numobs) corr(sig)  // e1 & v2 values
generate z1 = rnormal()
generate z2 = rnormal()
generate z3 = rnormal()
generate x = $pi*z1 + $pi*z2 + $pi*z3 + v    // reduced
form
generate y = x + e
correlate x e
```

regress x z1 z2 z3

regress y x

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

Birinchi yangi element simulatsiya qilingan ma'lumotlarga LIML hisoblagichini qo'llashdir. Buyruq va chiqish quyida keltirilgan. E'tibor bering, LIML baholari haqiqiy qiymatlarga yaqin, chunki biz instrumental o'zgaruvchilar kuchli bo'lishi uchun global o'zgaruvchisini $\pi = 0,5$ o'rnatdik:

ivregress liml y (x=z1 z2 z3), small

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

Instrumental variables (LIML) regression

Source	SS	df	MS	Number of obs =	100
Model	356.810445	1	356.810445	F(1, 98) =	51.15
Residual	112.041511	98	1.14328073	Prob > F =	0.0000
Total	468.851956	99	4.73587834	R-squared =	0.7610
				Adj R-squared =	0.7586
				Root MSE =	1.0692

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x	.9951923	.1391497	7.15	0.000	.7190542 1.27133
_cons	.1290802	.1074206	1.20	0.232	-.0840924 .3422528

Instrumented: x

Instruments: z1 z2 z3

Keyinchalik, $a = 1$ va $a = 4$ bo'lgan Fuller tomonidan o'zgartirilgan k sinf baholovchisini kiritamiz:

ivreg2 y (x=z1 z2 z3), small fuller(1)

```
. ivreg2 y (x=z1 z2 z3), small fuller(1)
```

LIML estimation

k = 0.99954
lambda = 1.00996
Fuller parameter = 1

Estimates efficient for homoskedasticity only
Statistics consistent for homoskedasticity only

Total (centered) SS	=	468.8519559	Number of obs =	100
Total (uncentered) SS	=	469.1579693	F(1, 98) =	55.25
Residual SS	=	109.2043273	Prob > F =	0.0000
			Centered R2 =	0.7671
			Uncentered R2 =	0.7672
			Root MSE =	1.056

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x	1.010894	.1359953	7.43	0.000	.7410159 1.280772
_cons	.130244	.106042	1.23	0.222	-.0801929 .3406808

ivreg2 y (x=z1 z2 z3), small fuller(4)

```
. ivreg2 y (x=z1 z2 z3), small fuller(4)
```

LIML estimation

```
k           =0.96829
lambda      =1.00996
Fuller parameter=4
```

Estimates efficient for homoskedasticity only
Statistics consistent for homoskedasticity only

Total (centered) SS	=	468.8519559	Number of obs =	100
Total (uncentered) SS	=	469.1579693	F(1, 98) =	68.37
Residual SS	=	101.7893518	Prob > F =	0.0000
			Centered R2 =	0.7829
			Uncentered R2 =	0.7830
			Root MSE =	1.019

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x	1.054443	.1275256	8.27	0.000	.801373	1.307514
_cons	.1334717	.1023524	1.30	0.195	-.0696432	.3365867

Simulatsiya uchun foydalaniladigan dastur tuzilishi jihatidan ushbu qo'llanmaning 10.8-rejasida qo'llangan dasturga o'xshaydi.

program ch11sim, rclass

version 11.1

drop _all

set obs \$numobs

```
matrix sig = (1, $rho \ $rho, 1)           // cov(e1,v2)
drawnorm e v, n($numobs) corr(sig)         // e1 & v2 values
generate z1 = rnormal()
generate z2 = rnormal()
generate z3 = rnormal()
generate x = $pi*z1 + $pi*z2 + $pi*z3 + v
generate y = x + e
```

```
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(r
eturn(t2sls))>invttail($numobs-2,.025)
```

Biz LIML va Fuller tomonidan o'zgartirilgan k-sinfni 2SLS bilan bir xil daromad bilan tanishtirdik.

```
ivregress liml y (x=z1 z2 z3), small  
return scalar bliml = _b[x]  
return scalar seliml = _se[x]  
return scalar tliml = (_b[x]-1)/_se[x]  
return scalar rliml = abs(return(tliml))>invttail($numo  
bs-2,.025)
```

```
ivreg2 y (x=z1 z2 z3), small fuller(1)  
return scalar bfull = _b[x]  
return scalar sefull = _se[x]  
return scalar tfull = (_b[x]-1)/_se[x]  
return scalar rfull = abs(return(tfull))>invttail($numo  
bs-2,.025)
```

```
ivreg2 y (x=z1 z2 z3), small fuller(4)  
return scalar bfull4 = _b[x]  
return scalar sefull4 = _se[x]  
return scalar tfull4 = (_b[x]-1)/_se[x]  
return scalar rfull4 = abs(return(tfull4))>invttail($numo  
bs-2,.025)  
end
```

Simulyatsiya buyrug'i ko'proq elementlarga ega, ammo ushbu qo'llanmaning 10.8-rejasidagi kabi tuzilishga ega.

```
simulate b2slsr=r(b2sls) se2slsr=r(se2sls) t2slsr=r(t2sls) ///  
r2slsr=r(r2sls) blimlr=r(bliml) selimlr=r(seliml) ///  
tlimlr=r(tliml) rlimlr=r(rliml) bfullr=r(bfull) ///  
sefullr=r(sefull) tfullr=r(tfull) rfullr=r(rfull) ///  
bfull4r=r(bfull4) sefull4r=r(sefull4) tfull4r=r(tfull4)///  
rfull4r=r(rfull4), reps(10000) nodots nolegend seed(1234567):  
ch11sim
```

2SLS uchun birinchi displey va natijalar ham ushbu qo'llanmaning 10.8-bo'limidagi bilan bir xil.

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 = .5

di " 2sls"
gen mse2sls = (b2slsr-1)^2
tabstat b2slsr se2slsr r2slsr mse2sls, stat(mean sd)

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

```

stats	b2slsr	se2slsr	r2slsr	mse2sls
mean	1.011116	.11695	.0636	.0139068
sd	.1174081	.0274133	.2440512	.0227001

Yuqorida aytib o'tilganlarga e'tibor bering, o'rtacha 10 000 dan ortiq simulatsiya qilingan ikki bosqichli eng kichik kvadratlarni baholash 1-haqiqiy koeffitsiyent parametr qiymatiga juda yaqin va haqiqiy nol gipotezaning t testi kerak bo'lganidek, vaqtning taxminan 5%ni rad etadi.

```

di " liml"
gen mseliml = (blimlr-1)^2
tabstat blimlr selimlr rlimlr mseliml, stat(mean sd)

```

```
. tabstat blimlr selimlr rlimlr mseliml, stat(mean sd)
```

stats	blimlr	selimlr	rlimlr	mseliml
mean	.9881047	.1210493	.0509	.0153914
sd	.1234965	.029831	.2198045	.0276647

Yuqoridagi LIML natijalari ushbu kuchli IV misolda o'xshash.
di " fuller(1)"

```
gen msefull = (bfullr-1)^2
```

```
tabstat bfullr sefullr rfullr msefull, stat(mean sd)
```

```
. tabstat bfullr sefullr rfullr msefull, stat(mean sd)
```

stats	bfullr	sefullr	rfullr	msefull
mean	.999965	.1189061	.0553	.0141615
sd	.1190081	.0283846	.2285763	.0239569

Fuller tomonidan o'zgartirilgan $a = 1$ bo'lgan k sinf baholovchisi deyarli xolis bo'lgan baholovchini ishlab chiqarish uchun mo'ljallangan va yuqoridagi natijalar ushbu maqsadga mos keladi.

di " fuller(4)"

```
gen msefull4 = (bfull4r-1)^2
```

```
tabstat bfull4r sefull4r rfull4r msefull4, stat(mean sd)
```

```
. tabstat bfull4r sefull4r rfull4r msefull4, stat(mean sd)
```

stats	bfull4r	sefull4r	rfull4r	msefull4
mean	1.033343	.1130738	.0812	.0126647
sd	.1074901	.0247115	.2731557	.0177391

Fuller tomonidan o'zgartirilgan $a=4$ bo'lgan k sinf baholovchisi kichik o'rtacha kvadrat xatoga ega bo'lishi uchun mo'ljallangan va yuqoridagi natijalar uning boshqa baholovchilarga qaraganda past MSEga ega ekanligini ko'rsatadi. Eslatib o'tamiz, har bir baholovchi uchun o'rtacha kvadrat xatosi hisoblanadi:

$$mse(\beta_2) = \frac{\sum_{m=1}^{10000} (\beta_{2m} - \beta_2)^2}{10000}$$

O'n birinchi bob mavzularini mustahkamlash uchun savollar:

1. Truffles kompaniyasi uchun talab va taklif modelini ifodalang.
2. Ikki bosqichli eng kichik kvadratlar (2SLS) hisoblaridan nima uchun foydalaniladi?
3. Covariates nima?
4. Qaysi buyruq o'zgaruvchining haqiqiylikini tekshirishni amalga oshiradi?
5. In (QUAN) uchun qisqartirilgan shakl yordamida tenglama qanday tuziladi?
6. 2SLS orqali talabni baholash qanday amalga oshiriladi?
7. 2SLS ning alternativlari haqida nimalarni bilasiz?
8. Stata tomonidan minimal xos qiymat statistikasi deb nimaga aytiladi?
9. Monte Karlo simulatsiya natijalari nima uchun qo'llaniladi?
10. Nima sababdan baholovchi uchun o'rtacha kvadrat xatosi hisoblanadi?

O'n birinchi bob mavzularida qo'llanilgan buyruqlar va atamalar

2sls	ivreg2	rnormal()
correlate	ivregress	seed
demand equation	k-class estimator	simulate
drawnorm	LIML	simultaneous
endogenous	list	equations
variables	matrix	Stock-Yogo critical
esttab	mean squared error	value
exogenous variables	Monte Carlo	summarize
findit	simulation	supply equation
F-test	predict	test
global	reduced form	two-stage least
identification	equations	squares
instrument variables	regress	Wald test

O'n birinchi bob mavzularini mustahkamlash uchun masalalarning do-file lari

Ma'lumotlar buyrug'i: `cd c:\data\poe4stata`

1-masala. log file ni ochish.

`log using chap11_truffles, replace text`

`* open data`

`use truffles, clear`

`* examine data`

`describe`

`list in 1/5`

`summarize`

`* reduced form equations`

`reg q ps di pf`

`reg p ps di pf`

`predict phat`

`* 2sls of demand`

`reg q phat ps di`

`* IV/2sls of demand equation`

`ivregress 2sls q (p=pf) ps di`

`ivregress 2sls q (p=pf) ps di, small`

`ivregress 2sls q (p=pf) ps di, small first`

`estat firststage`

`* 2sls of supply using least squares`

`reg q phat pf`

`* IV/2sls of supply equation`

`ivregress 2sls q (p=ps di) pf, small first`

`estat firststage`

`***** 2sls using REG3`

`***** This is not discussed in the chapter.`

`***** Enter help reg3`

`reg3 (q p ps di) (q p pf), endog(q p) 2sls`

`log close`

2-masala. Fulton Fish Market ma'lumotlari asosida.

`* open log`

log using chap11_fish, replace text

* open data

use fultonfish, clear

* examine data

describe

list lquan lprice mon tue wed thu stormy in 1/5

summarize lquan lprice mon tue wed thu stormy

* estimate reduced forms

reg lquan mon tue wed thu stormy

reg lprice mon tue wed thu stormy

test mon tue wed thu

* IV/2sls

ivregress 2sls lquan (lprice=stormy) mon tue wed thu, small first

estat firststage

log close

3-masala. Log file ni oching.

log using chap11_liml, replace text

use mroz, clear

drop if lfp==0

gen lwage=ln(wage)

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

gen exper2 = exper^2

* B=1, L=1

ivregress liml hours (mtr = exper) educ kidsl6 nwifeinc, small

estat firststage

estimates store m11

* B=1, L=2

ivregress liml hours (mtr = exper exper2) educ kidsl6 nwifeinc,
small

estat firststage

estimates store m12

*View LIML as IV estimator

* save liml k-value

scalar kvalue=e(kappa)

* reduced form residuals

```

reg mtr exper exper2 educ kidsl6 nwifeinc
predict vhat, r
* create purged endogenous variable
gen emtr = mtr - kvalue*vhat
* apply 2sls with IV = purged endogenous variable
ivregress 2sls hours (mtr = emtr) educ kidsl6 nwifeinc, small
* B=1, L=3
ivregress liml hours (mtr = exper exper2 largacity) educ kidsl6
nwifeinc, small
estat firststage
estimates store m13
* B=1, L=4
ivregress liml hours (mtr = exper exper2 largacity unemployment)
educ kidsl6 nwifeinc,
small
estat firststage
estimates store m14
* B=2, L=2
ivregress liml hours (mtr educ = mothereduc fathereduc) kidsl6
nwifeinc, small
estat firststage
estimates store m22
* B=2, L=3
ivregress liml hours (mtr educ = mothereduc fathereduc exper)
kidsl6 nwifeinc, small
estat firststage
estimates store m23
* B=2, L=4
ivregress liml hours (mtr educ = mothereduc fathereduc exper
exper2) kidsl6 nwifeinc,
small
estat firststage
estimates store m24

```

XII BOB. VAQT KETMA-KETLIGI MA'LUMOTLARI BILAN REGRESSIYA: STATSIONAR BO'LMAGAN O'ZGARUVCHILAR

12.1. Statsionnar va nostatsionnar ma'lumotlar

Ushbu bobning asosiy maqsadi sizning ma'lumotlaringizda vaqt seriyalari xususiyatlarini o'rganish uchun Statadan qanday foydalanishni ko'rsatishdir. Ekonometrikaning asosiy tamoyillaridan biri shundan iboratki, baholovchilarning statistik xossalari va ularning tadqiqot uchun foydaliligi natijada ma'lumotlarning qanday harakat qilishiga bog'liq. Masalan, xatoliklar regressorlar bilan bog'langan chiziqli regressiya modelida eng kichik kvadratlar baholovchisi izchil emasligi sababli uni baholash yoki keyingi sinov uchun ishlatmaslik kerak.

Vaqtli qatorlarining regressiyasida ma'lumotlar odatiy ekonometrik bo'lishi uchun statsionar bo'lishi, tegishli statistik xususiyatlarga ega bo'lish uchun jarayonlarga ega bo'lishi kerak. Asosan, bu vaqt ketma-ketliklari ma'lumotlarining vositalari, dispersiyalari va kovariatsiyalari ular kuzatilgan vaqt davriga bog'liq bo'lmasligini talab qiladi. Masalan, 1973-yilning uchinchi choragidagi YalMning o'rtacha va dispersiyasi 2006-yilning 4-choragidagidan farq qilishi mumkin emas. Ushbu muammoni hal qilish usullari so'nggi yillarda ekonometriklar uchun boy tadqiqotlar maydonini bo'lib xizmat qildi va bu usullarning bir nechta bu yerda o'rganildi.

Birinchi qo'llaniladigan tashxis vositalaridan biri – bu ma'lumotlarning oddiy vaqt seriyasi maydonidir. Vaqt ketma-ketligining maydoni ma'lumotlar bilan bog'liq bo'lgan muammolarni ochib beradi va statistik jihatdan davom etish yo'llarini taklif qiladi. Oldingi boblarda ko'rganimizdek, Statada vaqt maydonlarini yaratish oson va quyida bir nechta yangi manbalar ko'rib chiqamiz.

Birinchi narsa, katalogni ma'lumotlarni o'z ichiga olgan katalogga o'zgartirish va uni xotiraga yuklashdir. Ushbu mashqda

biz usa.dta ma'lumotlaridan foydalanamiz.

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

```
use usa, clear
```

Ushbu ma'lumotlar to'plami to'rtta o'zgaruvchini (gdp, inf, f, and b) o'z ichiga oladi, lekin vaqt o'zgaruvchilari yo'q. Stataning o'rnatilgan vaqt seriyasi funksiyalaridan foydalanish uchun biz vaqt o'zgaruvchisini yaratishimiz va keyin ma'lumotlarni tsset yordamida vaqt seriyasi deb e'lon qilishimiz kerak. Grafiklarni yanada mazmunli qilish uchun, haqiqiy ma'lumotlar to'plamidagi sanalarga mos keladigan sanalar to'plamini yaratish uchun qo'shimcha vazifalarni bajarish kerak. Ma'lumotlar bilan taqsimlangan ta'rif fayllari ma'lumotlarning har chorakda bo'lishini ko'rsatadi, 1984-yilning birinchi choragida (1984q1) boshlanadi va 2009-yilning uchinchi choragida (2009q3) bilan tugaydi. Statada tegishli sanalarni yaratish to'g'risida to'liqroq muhokama IX bobda berilgan agar siz buni haliham ko'rib chiqmagan bo'lsangiz, ushbu materialni hozir ko'rib chiqishingiz tavsiya etiladi.

12.1.1. Statada sanalarni yaratishni ko'rib chiqish

Aslida, birinchi narsa, kerakli sanalarni belgilaydigan bir qator butun sonlarni kiritishdir. IX bobni esga olaylik, Stata vaqt o'tishini asosiy sanadan (1960) o'tadigan vaqt birliklari soni sifatida qayd etadi. Shuning uchun sana yaratish vaqt birligi va boshlanish sanasini ko'rsatadigan funksiyani o'z ichiga olishi kerak. Shunday qilib, q(1984q1) o'sish choraklik ekanligini bildiradi va butun sonlar qatori 1984-yilning birinchi choragida boshlanadi. Buni tekshirish uchun quyidagi buyruqni yozing:

```
display q(1984q1)
```

Qaysiki ochib beradi:

```
. di q(1984q1)
```

```
96
```

Bu shuni ko'rsatadiki, 1984-yilning 1-choragi 1960q1 yilga nisbatan 96 chorakni tashkil etadi. Raqamlarni ma'lumotlar to'plamining qatorlari bo'yicha oshirish uchun ushbu raqamga _n-1

qo'shing. $_n$ – Stataning kuzatuv raqamini aniqlash usuli. Shunday qilib, 1-kuzatuvda bu nolga teng. Birinchi 1984q1 kuzatuv 96 ga teng. Kuzatish raqami 2 uchun ($_n=2$) va sana 97 ga teng bo'ladi va hokazo. Bu o'zgaruvchi generate buyrug'i yordamida yoziladi.

gen date = q(1984q1) + $_n$ – 1

Keyinchalik format buyrug'i display formati %tq yordamida butun sonlarni satrlarga aylantirish uchun ishlatiladi. Ya'ni, 96 raqami 1984q1 sifatida ko'rsatiladi, bu esa kimgadir sana aslida nima ekanligini aniqlashni osonlashtiradi. Nihoyat, yangi o'zgaruvchi tsset buyrug'i yordamida vaqt seriyasi deb ko'rsatiladi.

format %tq date

tsset date

12.1.2. Sanalarni formatlash

Ayrim yil va chorak o'zgaruvchilari foydali bo'lishi mumkin bo'lgan holatlar mavjud. Vaqtli qatorlar yaratilgan, formatlangan va ko'rsatilgandan so'ng, buni qilish juda oson. usa.dta dagi chorak va yil ma'lumotlarini olish uchun bir necha bosqichlar kerak. Birinchidan, sanaga yangi format berilishi kerak. Bizning holatda, %tq tarzida formatlangan sanalarni %td formatiga o'zgartirish kerak. %td formati Statadagi ona formatdir va bu sana haqidagi ma'lumotlardan oy, kun, yil va hokazolarni olish mumkin bo'lgan yagona format hisoblanadi. %tq choraklik ma'lumotlarni %td formatiga aylantirish funksiyasi dofq(). Bu "har chorakda bir kunlik" deb o'qiladi. Argumentda sanalar choraklik formatda bo'lishi kerak. Format o'zgartirilgandan so'ng, quyida ko'rsatilganidek, yil va chorak juda aniq sintaksis yordamida chiqarilishi mumkin:

gen newdate = dofq(newdate)

gen y = year(newdate)

gen q = quarter(newdate)

%tm ni %td ga aylantirish uchun dofmm() dan foydalaniladi. Boshqalar uchun help dates Command oynasida sanalarini kiriting va Sana va vaqt qiymatlarini aylantirish va chiqarish bo'limlariga qarang.

Ma'lumotlar to'plamida sana ma'lumotlari to'g'ri bo'lsa,

ma'lumotlaringizni saqlab qo'ygan ma'qul. Bu yangi sana ma'lumotlarini doimiy ravishda ma'lumotlar to'plamiga joylashtiradi va har safar foydalanmoqchi bo'lganingizda uni qo'lqa kiritish muammosidan xalos qiladi. File > Save yoki klaviatura buyrug'i Ctrl+S joriy ma'lumotlarni usa.dta ga saqlaydi.

12.1.3. Ma'lumotlarning grafigini tuzish

Gdp seriyasining grafigini tuzish uchun graph buyrug'idan foydalaning. Grafik vaqt qatorlarining to'liq sintaksisi:

graph twoway (tsline gdp)

Birinchi argument twoway undan keyin tsline (vaqt qatori chizig'ini anglatadi) va u o'zgaruvchi sana o'rniga x o'qi bo'yicha gdp o'zgaruvchisining vaqt o'lchamini o'lchash uchun ishlatilishi mumkin. Birinchi farqlarni grafik qilish uchun Stataning o'rnatilgan farqlilik operatori D. (yoki D1) dan foydalaning. Lag operatori L kabi farq operatori ld qo'shimcha sifatida ishlatiladi va ma'lumotlarni juda oson farq qiladi. Lag operatori L kabi farqli operatori old qo'shimcha sifatida ishlatiladi va ma'lumotlarni juda oson farqlaydi. Boshqa vaqt seriyalari operatorlari haqida qo'shimcha ma'lumot olish uchun onlayn yordamdan tsvarlist so'zini qidiring yoki kuzatuvchi orqali.

Bu quydagicha:

tsline gdp

yuqoridagi graph twoway ga qisqartirish mumkin. Bu biz quyida amal qiladigan konvensiya.

tsline D.gdp

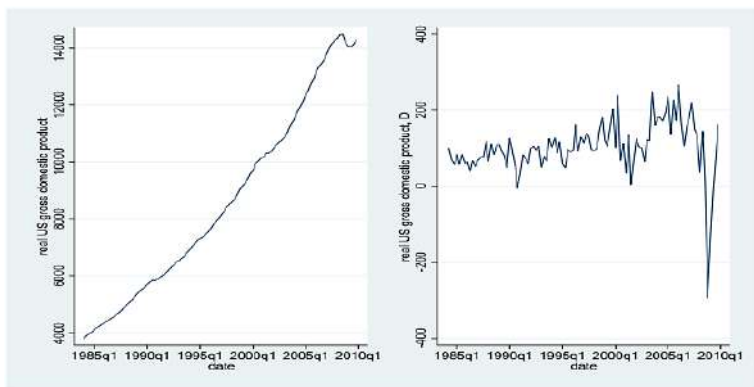
yordamida farqlangan qatorlarning grafigi olinadi.

Grafikni birlashtirishning oson yo'li – ularni quyidagi name(graph_name) opsiyasidan foydalanib nomlash. gdp va uning farqlari grafiklari tuziladi, nomlar beriladi hamda keyin graph combine buyrug'i yordamida birlashtiriladi. Stata buyrug'i:

qui tsline gdp, name (gdp, replace)

qui tsline D.gdp, name (dgdg, replace)

graph combine gdp dgdg



Chapdagi panel AQSH YaIM darajasi va choraklik o'zgarishlar o'ng tomonda. Yaratilgan va nomlangan grafiklarni xotiradan olib tashlash uchun `graph drop graph_name` buyrug'ini bering.

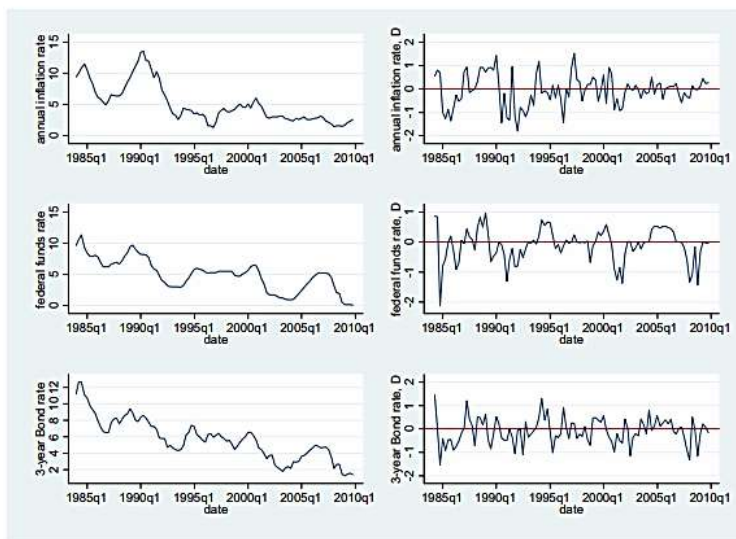
Agar siz grafikni saqlamoqchi bo'lsangiz, `saving(graph_name)` buyrug'idan foydalaning. Grafiklarni boshqa dasturga joylashtirish uchun saqlamoqchi bo'lsangiz, bu qulay imkoniyat yaratadi. `saving(graph_name)` buyrug'i yordamida yaratilgan va saqlangan grafiklarni o'chirish uchun operatsion tizimingiz uchun buyruq qatoriga to'g'ridan to'g'ri kirish uchun Stata shell buyrug'idan foydalaning. Mening joriy tizimim Microsoftning Windows XP-tizimida ishlaydi. Shell dan foydalanish uchun Windows XP buyruqlar oynasini ochish talab etiladi. U yerdan kompyuter uchun standart operatsion tizim buyruqlari berilishi mumkin.

Keyinchalik, inflatsiya darajasi, Fed fondlari stavkasi va 3 yillik obligatsiyalar stavkasi uchun grafiklar to'plami ularning o'zgarishlari bilan birlashtiriladi:

```

qui tsline inf, name (inf, replace)
qui tsline D.inf, name (dinf, replace) yline(0)
qui tsline f, name (f, replace)
qui tsline D.f, name (df, replace) yline(0)
qui tsline b, name (b, replace)
qui tsline D.b, name (db, replace) yline(0)
graph combine inf dinf f df b db, cols (2)

```



O'zgarishlarda o'lchangan seriyalarning har biri uchun yline(0) opsiyasi yordamida gorizontalar chiziqlar nolga joylashtirildi. Bu salbiy va ijobiy qiymatlarni osongina aniqlash uchun foydali hisoblanadi. Shuningdek, graph combine buyrug'i cols(2) opsiyasidan foydalaniladi, shunda har bir seriya va farq uchun grafiklar yonma-yon paydo bo'ladi.

Keyin, umumiy statistik ma'lumotlar to'plami yaratiladi. Bunday holda, Principles of Econometrics, 4th Edition (POE4) sizga ma'lumotlarning kichik namunalari uchun umumlashtirilgan statistik ma'lumotlarni taqdim etadi. Birinchi kichik namuna 1984q1 dan 1996q4gacha bo'lgan 52 ta kuzatuvdan iborat. Ikkinchisi 52 ta kuzatuvni o'z ichiga oladi va 1996q4 dan 2009q4gacha davom etadi. Summarize buyrug'i shartli if iborasi yordamida umumlashtirilgan statistik ma'lumotlarni olish uchun ishlatiladi. Qulay usul quyi namunani kerakli davrga moslashtirishdan iborat bo'ladi.

Buni amalga oshirishning bir necha yo'li mavjud. Birinchidan, standart if tasdig'idan foydalanib, namuna sanalarini cheklashimiz mumkin. Kichik namunali ma'lumotlar to'plamining boshidan boshlanadi va 1996q4 da tugaydi. Tuzilishi esa:

summarize if date>=q(1996q4)

Bu (1996q4) gacha bo'lgan barcha ma'lumotlarni umumlashtiradi. Sanalar o'zgaruvchan date, q(1996q4) saqlanadi, Stataga sana har chorakda q() dan foydalanishini va tegishli sana 1996q4 ekanligini aytadi. <= operatori matematik jihatdan "kichik yoki teng"ga ekvivalentdir. Ikkinchi kichik namuna 1997-yilning birinchi choragida boshlanadi va ma'lumotlar to'plamining oxirigacha davom etadi.

Tuzilishi esa:

summarize if date>=q(1997q1)

Agar summarize buyrug'idan keyin o'zgaruvchilar nomlari ro'yxatga kiritilmagan bo'lsa, u holda ma'lumotlar to'plamidagi barcha o'zgaruvchilar uchun umumiy statistika yoziladi. Albatta, siz avval ularni yaratib, alohida generate buyruqlari yordamida ma'lumotlar to'plamiga joylashtirmasangiz, unda hech qanday o'zgarish yuzaga kelmaydi. Ushbu misolda farqlar ilgari yaratilmagan va ular umumlashtirish buyrug'idan keyin maxsus ro'yxatga olinishi kerak.

Namunani cheklashning boshqa usuli - o'rnatilgan tin() funksiyasidan foydalanish, ammo bu faqat vaqt seriyasi uchun ishlaydi. Onlayn yordamdan tin() tuzilishi:

Time-series function

tin(d1, d2)

Domain d1: data or time literals recorded in units of t previously tsset

Domain d2: data or time literals recorded in units of t previously tsset

Range: 0 and 1, 1 means true

Description: true if $d1 < t < d2$, where t is the time variable previously tsset.

You must have previously tsset the data to use tin(). When you tsset the data, you specify a time variable t, and the format on t states how it is recorded. You type d1 and d2 according to that format.

"Tin" nomi "t in" tarzida o'qiladi, bu uning nima qilishini

anglatadi. Bu mohiyatan mantiqiy funksiya bo'lib, uning d1 va d2 (shu jumladan) argumentlari o'rtasida belgilangan vaqt oynasida kuzatuv mavjudligini tekshiradi. Argumentlar sanalar yoki butun sonlar bo'lishi mumkin, lekin har ikkala holatda ham bu funksiya ishlashi uchun ma'lumotlar to'plami vaqt seriyasi sifatida tsset bo'lishi kerak.

Agar siz namunaning namuna boshidan boshlanishini istasangiz, d1 argumentini bo'sh qoldiring. Kuzatishlarni oxirigacha kiritish uchun d2 ni o'tkazib yuboring. Bu esa tin() yordamida tuzilish:

summarize if tin(1996q4)

summarize if tin(1997q1,)

POE4 adabiyotning 12.1-jadvalidagi raqamlarni haqiqatda takrorlash uchun d1 ko'rsatilishi kerak, chunki birinchi kuzatuvdagi ma'lumotlar o'chirilgan bo'ladi. Jadvalni hisoblashda ishlatiladigan birinchi kuzatuv 1984q2 dan boshlanadi, shuning uchun buyruq:

summarize gdp inf b f D.gdp D.inf D.b D.f if tin(1984q2,1996q4)

Birinchi kichik namuna uchun natija

Variable	Obs	Mean	Std. Dev.	Min	Max
gdp	51	5813.02	1204.604	3906.3	8023
inf	51	6.903725	3.337811	1.28	13.55
b	51	7.343137	1.939775	4.32	12.64
f	51	6.417255	2.130539	2.99	11.39
gdp d1.	51	82.65882	29.33348	-4.6	161.8
inf d1.	51	-.1605882	.8320058	-1.8	1.43
b d1.	51	-.1029412	.6312822	-1.54	1.45
f d1.	51	-.0864706	.5860711	-2.12	.97

va

ikkinchisi

summarize gdp inf b f D.gdp D.inf D.b D.f if tin(1997q1,)

Variable	Obs	Mean	Std. Dev.	Min	Max
gdp	52	11458.19	2052.135	8137	14484.9
inf	52	3.219423	1.116619	1.45	6.04
b	52	3.977115	1.564322	1.27	6.56
f	52	3.4875	2.025269	.12	6.52
gdp					
dl.	52	120.275	92.91987	-293.7	267.9
inf					
dl.	52	.0251923	.4617422	-.93	1.52
b					
dl.	52	-.0875	.4788502	-1.33	.81
f					
dl.	52	-.0992308	.5142893	-1.43	.59

12.2. Soxta regressiyalar

Regressiyani taxmin qilish va statistik jihatdan ahamiyatli munosabatni topish mumkin, hatto ular mavjud bo'lmasa ham. Vaqt ketma-ketligini tahlil qilishda bu ma'lumotlar statsionar bo'lmaganda holatda tez-tez uchraydigan hodisadir. Ushbu misolda mustaqil tasodifiy yurishlar sifatida yaratilgan ikkita ma'lumotlar seriyasi, rw1 va rw2 ishlatiladi:

$$rw_1: y_t = y_{t-1} + v_{1t}$$

$$rw_1: y_t = y_{t-1} + v_{1t}$$

Xatolar psevdotasodifiy raqamlar generatori yordamida yaratilgan mustaqil standart normal tasodifiy og'ishlardir. Ko'rib turganingizdek, xt va yt hech qanday tarzda bog'liq emas. Ushbu bog'liq bo'lmagan qatorlar orasidagi empirik munosabatni o'rganish uchun soxta.dta ma'lumotlarini yuklang, vaqt o'zgaruvchisini yarating va ma'lumotlarni vaqt seriyasi deb ko'rsating:

`use spurious,clear`

`gen time = _n`

`tsset time`

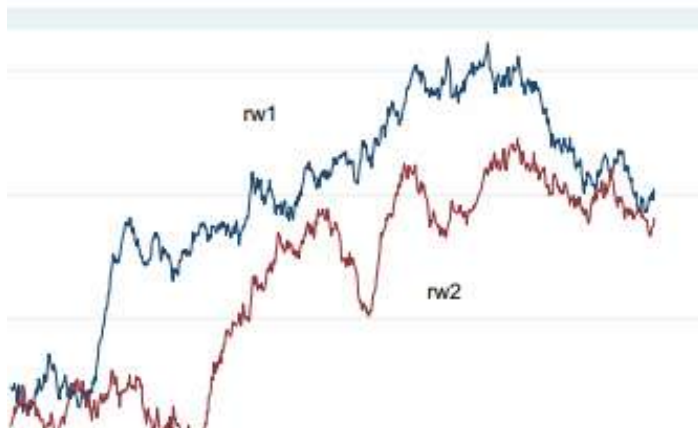
Ma'lumotlar yasama bo'lgani uchun, haqiqiy sanalarni yaratishga vaqt ajratishning hojati yo'q. Bunday holda oddiy davr hisoblagichi kifoya qiladi va _n yordamida kuzatish soniga teng bo'lgan biri yaratiladi. Vaqt o'zgaruvchisini yaratishning ushbu oddiy usuli muntazam oraliqlarda qayd etilgan har qanday seriya

uchun ishlatilishi mumkin (ya'ni, kuzatishlar orasidagi o'tgan vaqt teng).

Birinchi narsa, ma'lumotlarni vaqtli ketma-ketlik grafigi yordamida chizishdir. Oddiy tarzda foydalanish

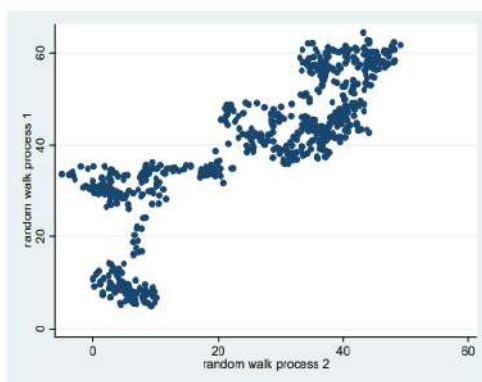
```
tsline rw1 rw2, name(g1, replace)
```

quyidagi sujetni ishlab chiqarish uchun:



Tarqalish sxemasi o'zgaruvchilar o'rtasidagi potensial soxta munosabatni ochib beradi:

```
scatter rw1 rw2, name(g2, replace)
```



name() opsiyasi shart emas, lekin bu buyruqlarni ommaviy ish

faylida ishga tushirishda qulay. Bu har bir grafikka nom beradi va ularning barchasi alohida oynalarda ochiladi.

Chiziqli regressiya bu ikki bog'liq bo'lmagan vaqt qatorlari o'rtasidagi chiziqli munosabatlarning ko'rinishini tasdiqlaydi:

regress rw1 rw2

natija:

. regress rw1 rw2

Source	SS	df	MS	Number of obs =	700
Model	122116.557	1	122116.557	F(1, 698) =	1667.65
Residual	51112.3314	698	73.2268359	Prob > F =	0.0000
Total	173228.888	699	247.823874	R-squared =	0.7049
				Adj R-squared =	0.7045
				Root MSE =	8.5573

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
rw2	.8420412	.0206196	40.84	0.000	.8015572 .8825251
_cons	17.81804	.6204776	28.72	0.000	16.59981 19.03627

rw2 bo'yicha koeffitsiyent ijobiy (.842) va muhim ($t = 40,84 > 1,96$). Biroq, bu o'zgaruvchilar aslida bog'liq emas. Kuzatilgan munosabatlar soxta. Soxta natijaning sababi ikki qatorning statsionar emasligidir. Shuning uchun regressiyada vaqt seriyasidan foydalanganda ma'lumotlaringizning statsionarligini tekshirishingiz kerak bo'ladi.

Qoldiqlarni avtokorrelatsiya qilish imkoniyatini tezkor tekshirish LM testi yordamida amalga oshirilishi mumkin (IX bobga qarang).

. estat bgodfrey

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
1	682.958	1	0.0000

H0: no serial correlation

p qiymati kichik va juda kichik va modeldagi noto'g'ri tavsifning dalili hisoblanadi. Keyinchalik tekshiruv o'tkazilishi shart bo'ladi.

12.3. Statsionarlik uchun unit root testlari

Ma'lumotlaringizning statsionarligini tekshirish uchun (kengaytirilgan) Dikki-Fuller testlaridan foydalanishimiz mumkin bo'ladi. Ushbu testni amalga oshirish uchun vaqt seriyasiga oid bir nechta qarorlar qabul qilinishi kerak. Ba'zan bu tanlovlar

vaqt seriyalarining chizmalarini vizual tekshirish asosida amalga oshirilishi ham mumkin. Chiziqlarni tekshirish orqali siz vaqt seriyalarining nolga teng o'rtachaga ega yoki chiziqli yoki kvadratik tendensiya ega ekanligini aniqlashga harakat qiling. Agar seriyadagi tendensiya kvadratik bo'lsa, seriyaning farqlangan versiyasi ularda chiziqli tendensiya ega bo'ladi, ya'ni chiziqli yo'nalishga. Yuqoridagi Fed yig'indilarining grafiklarida siz Fpastga qarab borayotganini va uning farqi (D.f) qandaydir doimiy miqdor atrofida aylanib yurganini ko'rishingiz mumkin. Bular orasidagi aloqalar xuddi shunday harakat qiladi. Bu shuni ko'rsatadiki, har bir seriya uchun kengaytirilgan Dikki Fuller (ADF) test regressiyalari doimiylikni o'z ichiga olishi kerak, ammo vaqt tendensiyasini o'z ichiga olmasligi lozim.

YalM seriyasi vaqt bo'yicha biroz kvadratik ko'rinishga ega bo'ladi. Quyida paydo bo'ladigan seriyaning farqli versiyasida biroz yuqoriga siljish kuzatiladi va shuning uchun siz doimiy va vaqt tendensiyasini o'z ichiga olgan ADF testini tanlashingiz lozim. Aytganingizdek, mulohaza yuritish talab etiladi va undan oqilona foydalanishning imkoni mavjud. Bizning maqsadimiz imkon qadar rasmiy testlar yordamida noaniqliklarning bir qismini kamaytirishdir, ammo ekonometrik xulosadan kelib chiqqan holda tegishli testing o'ziga xos xususiyatlarini aniqlash ba'zi bir jarayonlarni talab qilishini yodda tuting.

Keyingi qaror, ADF regressiyalariga qo'shiladigan kechikishlar sonini tanlashdir. Qoida bo'yicha, ADF regressiyasining qoldiqlari avtokorrelatsiya qilinmaydi, qoldiqlarning tebranish chastotasi bir xil intensivlikda bo'lishini ta'minlash uchun yetarli darajada chetlanishli farqlarni qamrab olishi kerak

Ushbu bo'limda Dikki-Fuller testlarining statsionarligini baholashning ikkita usuli ko'rib chiqiladi. Birini qo'llanma sifatida ko'ramiz. Bunday holda siz eng kichik kvadratlar yordamida mos modelni baholaysiz, test uchun t nisbatini topasiz va uni matningizdagi jadval qiymatlari bilan solishtirasiz. Eslatib o'tamiz, sizning seriyangizning kechikish qiymati bo'yicha t nisbati t taqsimotiga ega emas. To'g'ri taqsimlash murakkab va biz sinov

uchun to'g'ri kritik qiymatlarni olish uchun belgilangan jadvallar yoki ko'rsatkichlarga tayanishimiz kerak.

Ikkinchi usulda Stataning o'rnatilgan funksiyalaridan biridan foydalaniladi. Bu yerda afzallik shundaki, Stata test uchun to'g'ri kritik qiymatlarni yaratadi va ularni olish uchun tashqi manbaga (ya'ni, jadvalga) murojaat qilishingiz shart emas. Stata sizga taxminiy p qiymatini ham beradi.

Birinchiidan, bu yerda Dikki-Fuller regressiyalarining asosiy klassifikatsiyasi

Series Characteristics	Regression Model
No Constant and No Trend	$\Delta y_t = \gamma y_{t-1} + v_t$
Constant, but No Trend	$\Delta y_t = \alpha + \gamma y_{t-1} + v_t$
Constant and Trend	$\Delta y_t = \alpha + \gamma y_{t-1} + \lambda t + v_t$

Har bir holatda nol va muqobil gipotezalar $H_0: \gamma = 0 = 0$ va $H_1: \gamma < 0$ lardir. Asosan, regressiyalar γ bo'yicha t nisbati hisoblab chiqiladi va matndagi jadvaldagi kritik qiymat bilan baholanadi yoki, yaxshiroq, Stata tomonidan taqdim etilgan qiymat bilan taqqoslanadi.

Dikki-Fuller testining kengaytirilgan versiyasi modelga kechiktirilgan farqlarni qo'shadi. Doimiy va trendsiz model uchun bu quyidagicha bo'ladi:

$$\Delta y_t = \alpha + \gamma y_{t-1} + \sum_{s=1}^m a_s \Delta y_{t-s} + v_t$$

Qo'shish uchun kechikishlar sonini tanlashingiz kerak. Asosan, qoldiqlar o'zaro bog'liq emasligini ta'minlash uchun Δy_{t-s} ning yetarlicha kechikishlarini kiritish kerak. Bunga misol ushbu qo'llanmada keyinroq keltirilgan. Kechiktirilgan atamalar sonini v_t qoldiqlarining avtokorrelatsiya funksiyasini (ACF) yoki hisoblangan kechikish koeffitsiyentlarining ahamiyatini a_s tekshirish orqali ham aniqlash mumkin.

Misolda, Federal jamg'armalar stavkasi (f) va 3 yillik obligatsiyalar stavkasi (b) ko'rib chiqiladi. Seriya syujetlari ma'lumotlarni

jamlab bo'lmashligini ko'rsatadi, bu esa ularni nostatsionar bo'lishi mumkinligini ko'rsatadi. Dikki-Fuller testlarini o'tkazish uchun avvalo doimiy va/yoki trenddan foydalaning. Tebranishi nolga teng bo'lmagan o'zgarishlar uchun doimiyini kiritamiz. Ko'rinib turibdiki, chiziqli yoki kvadratik tendensiya yo'q, shuning uchun biz doimiy, trend formulasini qabul qilmaymiz. Keyin tenglamaning o'ng tomoniga qancha kechikkan farqli shartlarni kiritish kerakligini aniqlang. IX bobda tasvirlangan modelni tanlash qoidalaridan foydalanib, biz ikkala holatda ham qoldiqlardagi avtokorrelatsiyani bartaraf etish uchun bitta kechikkan farq atamasini kiritish yetarli ekanligini aniqlaymiz.

usa.dta ma'lumotlarini qayta yuklang, Stata xotirasidagi oldingi ma'lumotlarni o'chiring.

use usa, clear

Agar **usa.dta** ma'lumotlari, vaqt o'zgaruvchilari yaratilgach va ma'lumotlarni birinchi misolda e'lon qilgandan keyin saqlanmasa, **date** o'zgaruvchisini qayta yaratishingiz kerak:

use usa, clear

gen date = q(1984q1) + _n - 1

format %tq date

tsset date

Statadagi regressiyalar

regress D.f L.f L.D.f

regress D.b L.b L.D.b

Farq operatori tenglamaning chap tomonidagi **f** va **b** dagi o'zgarishlarni yaratish uchun ishlatiladi. Kechikish operatori **L** operator **f** va **b** ning birinchi kechikish darajasini olish uchun ishlatiladi. Oxirgi o'zgaruvchi har ikkala **L.D.f** va **L.D.b** operatorlaridan ham foydalanadi. Bu operatorlarning chiziqliligi ularni har qanday tartibda (kommutativ) birlashtirish va ishlatish imkonini beradi. Demak **L.D.fD.f** ning kechikkan qiymatini oladi, bu esa o'z navbatida **f** ning birinchi farqlangan qiymati hisoblanadi. Kommutativ xususiyat, shuningdek, **D.L.f** yordamida tartibni o'zgartirishimiz va bir xil natijaga erishishimiz mumkinligini anglatadi.

. regress D.F L.F L.D.f

Source	SS	df	MS
Model	10.0957158	2	5.0478579
Residual	19.6353195	99	.198336561
Total	29.7310353	101	.294366686

Number of obs = 102
F(2, 99) = 25.45
Prob > F = 0.0000
R-squared = 0.3396
Adj R-squared = 0.3262
Root MSE = .44535

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
f						
L1.	-.0446213	.0178142	-2.50	0.014	-.0799685	-.0092741
L.D.	.5610582	.0809827	6.93	0.000	.4003708	.7217455
_cons	.1725221	.1002333	1.72	0.088	-.0263625	.3714067

. regress D.b L.b L.D.b

Source	SS	df	MS
Model	4.20542707	2	2.10271354
Residual	25.0098641	99	.25262489
Total	29.2152912	101	.289260309

Number of obs = 102
F(2, 99) = 8.32
Prob > F = 0.0005
R-squared = 0.1439
Adj R-squared = 0.1267
Root MSE = .50262

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
b						
L1.	-.0562412	.0208081	-2.70	0.008	-.097529	-.0149534
L.D.	.2903078	.0896069	3.24	0.002	.1125084	.4681072
_cons	.236873	.1291731	1.83	0.070	-.0194345	.4931804

f va b ning kechikkan qiymatlari bo'yicha t-nisbatlari yuqoridagi rasmda aylantirilgan. Bular Dikki-Fuller testini o'tkazish uchun tegishli qiymatlardir.

Kechikish operatorlari numlistni ham qo'llab-quvvatlaydi va bu matn tuzilishini yanada qisqartirishi mumkin.

reg L(0/1).D.f L.f

reg L(0/1).D.b L.b

Bu holda regressreg sifatida qisqartiriladi va L(0/1).D.fD.f L.D.f o'rnini egallaydi. Ikkinchisi aslida bu misolda hech qanday belgilarni saqlamaydi, lekin f dagi farqning ko'proq kechikishi bo'lsa modelga qo'shildi. Quyida ko'rsatilgandek, model kechikish tanlash uchun halqalardan foydalanish ham foydalidir.

Statada Dikki-Fuller regressiyalarini bajarish uchun o'rnatilgan buyruqlar mavjud. Buyruq dfuller va onlayn yordam sifatida gap tuzilishi

Syntax

dfuller *varname* [*if*] [*in*] [, *options*]

<i>options</i>	<i>description</i>
Main	
noconstant	suppress constant term in regression
trend	include trend term in regression
drift	include drift term in regression
regress	display regression table
lags(<i>#</i>)	include <i>#</i> lagged differences

You must **tsset** your data before using **dfuller**; see **[*ts*] tsset**.
varname may contain time-series operators; see **tsvarlist**.

Variantlar uchun siz doimiy, trend, drift (trend kvadrati) qoʻshilishi yoki kechikishlar sonini belgilashingiz mumkin. Agar siz regress opsiyasidan foydalansangiz, toʻliq regressiya natijalari aks etadi. Taqqoslash uchun biz quyida ushbu variantdan foydalanamiz:

dfuller f, regress lags(1)

Regress variantidan foydalanilganda Dikki-Fuller testi statistikasini, kritik qiymatlarni va regressiya natijalarini ishlab chiqaradi. Sinov uchun taxminiy p qiymati ham berilgan, bu testni oʻtkazishni osonlashtiradi. Bunday holda, p qiymati 0,10 dan katta boʻladi va birlik ildizning nol gipotezasini muhimlik darajasida rad etib boʻlmaydi.

```
. dfuller f, regress lags(1)
```

Augmented Dickey-Fuller test for unit root Number of obs = 102

	Test Statistic	1% Critical Value	Interpolated Dickey-Fuller		10% Critical Value
			5% Critical Value		
Z(t)	-2.505	-3.509	-2.890		-2.580

Mackinnon approximate p-value for Z(t) = 0.1143

D.f	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
f						
L1.	-.0446213	.0178142	-2.50	0.014	-.0799685	-.0092741
LD.	.5610582	.0809827	6.93	0.000	.4003708	.7217455
_cons	.1725221	.1002333	1.72	0.088	-.0263625	.3714067

Obligatsiyalar rentabelligi seriyasi uchun sinov:

```
. dfuller b, regress lags(1)
Augmented Dickey-Fuller test for unit root      Number of obs   =      102

               Test Statistic      Interpolated Dickey-Fuller
               1% Critical Value      5% Critical Value      10% Critical Value
Z(t)          -2.703                -3.509                -2.890                -2.580

Mackinnon approximate p-value for Z(t) = 0.0735
```

	D.b	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
b						
LL		-.0562412	.0208081	-2.70	0.008	-.097529 - .0149534
LD		.2903078	.0896069	3.24	0.002	.1125084 .4681072
_cons		.236873	.1291731	1.83	0.070	-.0194345 .4931804

Ushbu test uchun taxminiy p qiymati 5%dan yuqori va statsionar bo'lmagan nol gipoteza bu darajada rad etilmaydi.

Dikki-Fuller statistikasi berilganiga e'tibor bering (Z(t) deb nomlanadi) va test uchun 1%, 5% va 10% kritik qiymatlar ilova qilingan jadvalda berilgan. Stata siz uchun qilgan qo'shimcha kiritish tufayli, bu raqamlar matningizdagi raqamlardan biroz farq qiladi. Qanday bo'lmasin, test statistikasi rad etish hududiga to'g'ri kelmaydi va biz ma'lumotlarning darajalarini kamida 5% ahamiyatlilik darajasida statsionar emas degan xulosaga keldik.

Dikki-Fuller testlari farqli seriyalar uchun takror foydalaniladi. Farqlangan qatorlarda trend ham, doimiy ham mavjud emas. Hech qanday kechikish kerak bo'lmaydi, shuning uchun kod quyidagicha bo'ladi:

dfuller D.f, noconstant lags(0)

dfuller D.b, noconstant lags(0)

Bu safar regressiya natijalari olib tashlandi va biz ega bo'lamiz:

```
. dfuller D.f, noconstant lags(0)
Dickey-Fuller test for unit root      Number of obs   =      102

               Test Statistic      Interpolated Dickey-Fuller
               1% Critical Value      5% Critical Value      10% Critical Value
Z(t)          -5.487                -2.600                -1.950                -1.610

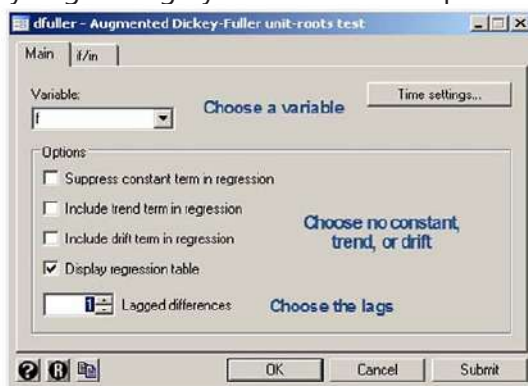
. dfuller D.b, noconstant lags(0)
Dickey-Fuller test for unit root      Number of obs   =      102

               Test Statistic      Interpolated Dickey-Fuller
               1% Critical Value      5% Critical Value      10% Critical Value
Z(t)          -7.562                -2.600                -1.950                -1.610
```

Bunday holda, biz statsionar nol gipotezani rad qilamiz va ikkala qator ham o'zlarining farqlarida statsionar (1-tartibdagi

integratsiya, ya'ni I (1)) degan xulosaga kelamiz.

Dikki-Fuller testi uchun dialog oynalari Statistics > Time series > Tests > Augmented Dickey-Fuller unit root tests ni tanlashda topiladi. Bu quyidagi dialog oynasini keltirib chiqaradi:



Yana bir bor, dfuller dialog oynasi yuqorida muhokama qilingan buyruqlarni yaratadi. Sinov qilmoqchi bo'lgan o'zgaruvchini tanlang, variantlar ro'yxatidan noconstant, trend yoki drift ni tanlang va Dikki-Fuller regressiyasini oshirish uchun kechikishlar sonini tanlang. Shuningdek, bu yerda bajarilgan regressiya natijalarini ko'rsatish uchun katakchani belgilashingiz mumkin.

Statada statsionarlik uchun boshqa testlar mavjud bo'lib, ular sizga foydali bo'lishi mumkin. Birinchisi DFGLS testidir. dfpls vaqt seriyasida birlik ildizi uchun testlar. U Elliott, Rothenberg va Stock (1996) tomonidan taklif qilingan o'zgartirilgan Dikki-Fuller t testini (DF-GLS testi sifatida tanilgan) amalga oshiradi. Aslini olganda, test Stataning dfuller buyrug'i bilan bajarilgan testga o'xshash kengaytirilgan Dikki-Fuller testidir, bundan tashqari vaqt seriyasi testni o'tkazishdan oldin umumlashtirilgan eng kichik kvadratlar (GLS) regressiyasi orqali o'zgartiriladi. Elliott, Rothenberg va Stok va keyingi tadqiqotlari shuni ko'rsatdiki, bu test kengaytirilgan Dikki-Fuller testining oldingi versiyalariga qaraganda sezilarli darajada kuchliroqdir. Shunday qilib, odatdagi kengaytirilgan Dikki-Fuller testi bo'lmasa, ushbu testning nostatsionarlikni shunchaki rad etib bo'lmaydi.

dfgls DF-GLS testini o'z ichiga birinchi farqlangan, o'zgaras o'zgaruvchining 1 dan k gacha kechikishlarini o'z ichiga oladi, bunda k foydalanuvchi tomonidan yoki Schwert (1989)da tasvirlangan usul bilan o'rnatilishi mumkin. Yuqorida va POE4 da muhokama qilinganidek, kengaytirilgan Dikki-Fuller testi shaklining regressiyasini moslashtirishni:

$$\Delta y_t = \beta y_{t-1} + \delta t + \gamma_1 \Delta y_{t-1} + \dots + \gamma_k \Delta y_{t-k} + u_t$$

ushbu formula orqali oladi va keyin $H_0: \beta = 0$ nol gipotezasini sinab ko'ring. DF-GLS testi shunga o'xshash tarzda amalga oshiriladi, lekin GLSdan ajratilgan ma'lumotlarga asoslangan. Sinovning nol gipotezasi shundan iboratki, y_t ehtimollik yo'nalish bilan tasodifiy yurish ikkita muqobil gipoteza mavjud: $y_t(t)$ chiziqli vaqt tendensiyasi bo'yicha statsionar yoki y_t nolga teng bo'lmagan o'rtacha bilan statsionar, lekin chiziqli vaqt tendensiyasiz. Bajara olmagan holda esa avvalgisidan foydalanish kerak. Oxirgi muqobilni belgilash uchun notrend opsiyasidan foydalaning.

Fed fondlari darajalari va 3 yillik obligatsiyalar stavkalari uchun:

dfgls f

dfgls b

Natijalar esa:

```
. dfgls f
```

DF-GLS for f					Number of obs = 91	
Maxlag = 12 chosen by Schwert criterion						
[lags]	DF-GLS tau Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value		
12	-3.035	-3.575	-2.753	-2.479		
11	-3.025	-3.575	-2.783	-2.508		
10	-2.905	-3.575	-2.813	-2.537		
9	-3.107	-3.575	-2.842	-2.565		
8	-3.097	-3.575	-2.870	-2.591		
7	-3.602	-3.575	-2.898	-2.617		
6	-3.732	-3.575	-2.924	-2.641		
5	-3.436	-3.575	-2.949	-2.664		
4	-3.290	-3.575	-2.972	-2.685		
3	-3.793	-3.575	-2.994	-2.706		
2	-3.499	-3.575	-3.014	-2.723		
1	-3.278	-3.575	-3.031	-2.739		

Opt Lag (Ng-Perron seq t) =	1 with RMSE	.3535111
Min SC = -1.980541 at lag	1 with RMSE	.3535111
Min MAIC = -1.808821 at lag	1 with RMSE	.3535111

```
. dfgls b
```

```
DF-GLS for b
```

```
Number of obs = 91
```

```
Maxlag = 12 chosen by Schwert criterion
```

[lags]	DF-GLS tau Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
12	-2.556	-3.575	-2.753	-2.479
11	-2.571	-3.575	-2.783	-2.508
10	-2.265	-3.575	-2.813	-2.537
9	-2.787	-3.575	-2.842	-2.565
8	-3.270	-3.575	-2.870	-2.591
7	-2.896	-3.575	-2.898	-2.617
6	-3.185	-3.575	-2.924	-2.641
5	-3.206	-3.575	-2.949	-2.664
4	-2.995	-3.575	-2.972	-2.686
3	-3.392	-3.575	-2.994	-2.706
2	-2.924	-3.575	-3.014	-2.723
1	-3.095	-3.575	-3.031	-2.739

```
Opt Lag (Ng-Perron seq t) = 3 with RMSE .4644163
```

```
Min SC = -1.371452 at lag 1 with RMSE .4793635
```

```
Min MAIC = -1.216796 at lag 1 with RMSE .4793635
```

dfgls dan foydalanishning afzalliklaridan biri ko'rinib turibdi. Schwert mezonini maksimal kechikish uzunligini 12 ga o'rnatadi va modellar 1 dan 12 gacha kechikishlar izlanadi. SC ni minimallashtiradigan model, MAIC va Ng-Perron statistikasi berilgan. f uchun kechikish tanlovi bir ma'noda 1 ga teng va statsionarlik gipotezasi 5% darajasida rad etiladi. b uchun ikkita mezon 1 dan kechikishni tanlaydi va ular ham 5%da muhim.

GLS hisoblagichining yuqori samaradorligi bir qator ikki ma'noli natijalarga olib keldi. Seriya darajalari esa aslida statsionar bo'lishi mumkin. Ba'zida yana bir sinov tavsiya etiladi. Phillips Perron (1988) testi, shuningdek, vaqt seriyasining statsionar bo'lgan muqobiliga nisbatan statsionar emasligi haqidagi nol gipotezaga ega. pperron ketma-ket korrelatsiyani hisobga olish uchun IX bobda muhokama qilingan Newey-West standart xatolaridan foydalanadi, dfuller da qo'llaniladigan kengaytirilgan Dikki-Fuller testi esa birinchi farqlangan o'zgaruvchining qo'shimcha kechikishlaridan foydalanadi. Ushbu testning afzalligi shundaki, regressiyani qanday oshirish kerakligini hal qilish uchun model tanlash jarayonini ko'rib chiqishning hojati yo'q. Shunday qilib, regressiya opsiyasi regressiya natijalarini chop etish uchun ishlatilishi mumkin. Kamchilik shundaki, natijalar tarmoqli kengligi tanloviga bog'liq. IX bobda muhokama qilingan yangi protseduradan farqli o'laroq, pperron testi avtomatik ravishda u yerda tavsiya etilgan variantlardan birini tanlaydi, ya'ni

$4(T/100)^{2/9}$. Sizda trend qo'shish yoki doimiyni olib tashlash imkoniyati mavjud. Ushbu misolda, vaqt seriyasidagi grafiklar pasayish tendensiyasini ko'rsatganligi sababli, tendensiya kiritilgan.

Ikki vaqt seriyasining darajalari uchun bizda mavjud

pperron b, regress trend

pperron f, regress trend

Natijalar

. pperron f, regress trend

Phillips-Perron test for unit root				Number of obs =	103
				Newey-West lags =	4
	Test Statistic	1% Critical Value	Interpolated 5% Critical Value	Interpolated 10% Critical Value	
Z(rho)	-13.209	-27.420	-20.712	-17.510	
Z(t)	-2.560	-4.039	-3.450	-3.150	
Mackinnon approximate p-value for Z(t) = 0.2985					

f	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
f					
_l1.	.9460524	.0326592	28.97	0.000	.8812574 .1.010847
_trend	-.0035557	.0027713	-1.28	0.202	-.0090539 .0019426
_cons	.3633958	.2941354	1.24	0.220	-.2201604 .9469521

va

. pperron b, regress trend

Phillips-Perron test for unit root				Number of obs =	103
				Newey-West lags =	4
	Test Statistic	1% Critical Value	Interpolated 5% Critical Value	Interpolated 10% Critical Value	
Z(rho)	-16.361	-27.420	-20.712	-17.510	
Z(t)	-2.978	-4.039	-3.450	-3.150	
Mackinnon approximate p-value for Z(t) = 0.1382					

b	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
b					
_l1.	.8924541	.0429605	20.77	0.000	.8072218 .9776865
_trend	-.0069282	.0035347	-1.96	0.053	-.013941 .0000845
_cons	.8823205	.4187536	2.11	0.038	.0515253 1.713116

Ushbu natijalar to'plami odatdagi kengaytirilgan Dikki-Fuller regressiyalarini qo'llab-quvvatlaydi. z(rho) statistikasi ikkala testning rad etish sohasida emas, bu ham b va f ning statsionar emasligini qo'llab-quvvatlaydi. f dagi trend atamasi ahamiyatli emas va uni olib tashlash mumkin. Bu boshqa turkumlarda muhim ahamiyatga ega bo'lib, uning kiritilganligini tasdiqlaydi.

12.4. Integratsion va kointegratsion

Ikki statsionar bo'lmagan vaqt seriyalari, agar ular vaqt o'tishi bilan birga harakat qilsalar, kointegratsiyalashgan bo'ladi. Masalan, biz Fed jamg'armasi stavkasi va 3 yillik obligatsiya stavkasi darajalari statsionar emasligini, ularning farqlari esa statsionar ekanligini aniqladik. Vaqt seriyalari adabiyotida qo'llaniladigan noaniq terminologiyada har bir seriya "integrated of order 1" yoki I(1) deyiladi. Agar ikkita statsionar bo'lmagan qator vaqt o'tishi bilan birga harakat qilsa, biz ularni "kointegratsiyalashgan" deb aytamiz. Iqtisodiy nazariya ularni arbitraj orqali bir-biriga bog'lash kerakligini taklif qiladi, ammo bu kafolat emas, shuning uchun biz rasmiy statistik testni o'tkazamiz.

Sinov jarayoni juda oddiy. Eng kichik kvadratlar yordamida bitta I(1) o'zgaruvchini boshqasiga regresslash. Keyin (kengaytirilgan) Dikki-Fuller testi yordamida qoldiqlarni statsionarlik uchun sinab ko'ring. Agar seriyalar kointegratsiyalangan bo'lsa, Dikki-Fuller testi statistikasi statistik ahamiyatga ega bo'ladi. Nol gipoteza shundan iboratki, qoldiqlar statsionar emas. Buni rad etish qoldiqlar statsionar va qatorlar kointegratsiyalashgan degan xulosaga keladi.

12.4.1. Engle-Granger testi

Oldingi bo'limda tasvirlangan test odatda Engle-Granger testi deb ataladi. b ni f va konstantada regressiya qiling, qoldiqlarni saqlang va ularni Dikki-Fullerning kengaytirilgan regressiyasida ishlatib. Bu qo'lda amalga oshiriladi:

regress b f

. regress b f					
Source	SS	df	MS		
Model	568.17396	1	568.17396	Number of obs =	104
Residual	66.9519745	102	.656391907	FC (1, 102) =	865.60
Total	635.125935	103	6.16627121	Prob > F =	0.0000
				R-squared =	0.8946
				Adj R-squared =	0.8936
				Root MSE =	.81018
b	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
_f	.9144114	.0310801	29.42	0.000	-.8527641 .9760587
_cons	1.13983	.1740833	6.55	0.000	.7945362 1.485123

predict ehat, residual

regress D.ehat L.ehat L.D.ehat, noconstant

. regress D.ehat L.ehat L.D.ehat, noconstant

Source	SS	df	MS			
Model	3.53907328	2	1.76953664	Number of obs =	102	
Residual	17.4123657	100	.174123657	F(2, 100) =	10.16	
Total	20.951439	102	.205406265	Prob > F =	0.0001	
				R-squared =	0.1689	
				Adj R-squared =	0.1523	
				Root MSE =	.41728	

D.ehat	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ehat						
L1.	-.2245093	.0535039	-4.20	0.000	-.3306595	-.1183592
L.D.	.2540448	.0937006	2.71	0.008	.0681454	.4399442

Uning kechikkan qiymati bo'yicha t nisbati $-4,20$ ga teng. Kritik qiymat tegishli jadvaldan olinishi kerak, masalan, POE4 dagi 12.4-jadval. Kesishmani o'z ichiga olgan kointegratsiya regressiyasi uchun 5% kritik qiymat $-3,37$ va t nisbati bundan kichik bo'ladi. Kointegratsiya yo'qligi haqidagi nol gipoteza $t \leq t_e$ bo'lganda rad etiladi, $t > t_e$ bo'lganda esa rad etilmaydi. Bu holda t statistika $-4,196 < -3,37$ va eng kichik kvadratlar qoldiqlari statsionar emasligi haqidagi nol gipoteza rad etiladi; qoldiqlar statsionardir. Bu shuni anglatadiki, obligatsiya stavkasi va federal mablag'lar stavkasi kointegratsiyalashgan.

Bir marta t nisbatini olish uchun o'rnatilgan df fuller buyrug'idan foydalanish mumkin, ammo Stata tomonidan chop etilgan kritik qiymatlar to'g'ri bo'lmaydi. Bular hali ham kointegratsiya tenglamasi uchun mos qiymatlar jadvalidan kelib chiqishi kerak. Bu an'anaviy Dikki-Fuller regressiyasi bilan bir xil bo'lmaydi.

. dfuller ehat, noconstant lags(1)

Augmented Dickey-Fuller test for unit root			Number of obs	=	102
Test Statistic	1% Critical Value	Interpolated Dickey-Fuller	5% Critical Value	10% Critical Value	
z(t)	-4.196	-2.600	-1.950	-1.610	

Yana bir bor, qoldiqlarning statsionarligini tekshirishda ushbu jadvaldagi kritik qiymatlarni ishlatmaslikdan ehtiyot bo'ling; berilgan kritik qiymatlar sinovdan o'tkazilayotgan vaqt qatori baholanmagan degan faraz ostida hisoblanadi.

12.4.2. Xatolarni tuzatish modeli

Kointegratsiya ikki statsionar bo'lmagan $I(1)$ o'zgaruvchilar o'rtasidagi munosabatdir. Ushbu o'zgaruvchilar umumiy tendensiyaga ega va uzoq vaqt mobaynida birgalikda harakat qilishadi. Ushbu bo'limda qisqa muddatli xatolarni tuzatish modeli deb nomlanuvchi kointegratsiya munosabatlarini o'zida mujassam etgan $I(0)$ o'zgaruvchilar orasidagi dinamik munosabat ko'rib chiqiladi.

ARDL (1,1) bilan boshlang

$$y_t = \delta + \theta_1 y_{t-1} + \delta_0 x_t + \delta_1 x_{t-1} + v_t$$

Ba'zi ta'sir o'tkazishdan so'ng (batafsil ma'lumot uchun POE4 ga qarang):

$$= -(1 - \theta_1)(y_{t-1} - \beta_1 - \beta_2 x_{t-1}) + \delta_0 \Delta x_t + \delta_1 \Delta x_{t-1}$$

Qavslarning ikkinchi to'plamidagi atama kointegratsion munosabatdir. y va x darajalari chiziqli bog'liqdir. $\alpha = (1 - \theta_1)$ bo'lsin va tenglamaning parametrlari chiziqli bo'lmagan eng kichik kvadratlar orqali baholanishi mumkin:

gen Db=D.b

nl (Db = -{alpha}*(L.b-{beta1}-beta2)*L.f) + ///

{delta0}*D.f+{delta1}*D.L.f), variables(L.b L.f D.L.f)

Bu yerda birgina qulaylik shundan iboratki, vaqt qatori operatori D. tobe o'zgaruvchini hosil qilishda qo'llanilmaydi. Uni nl da ishlatishdan oldin uni alohida yaratish kerak. Kechikish va farq operatorlari, agar o'zgaruvchilar parametrida keltirilgan bo'lsa, tenglamaning o'ng tomonida ishlatilishi mumkin. Natijalar:

Source	SS	df	MS			
Model	15.0345905	4	3.75864763	Number of obs =	102	
Residual	14.1807005	97	.146192788	R-squared =	0.5146	
Total	29.215291	101	.289260307	Adj R-squared =	0.4946	
				Root MSE =	.3823517	
				Res. dev. =	88.20819	

db	coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
/alpha	.1418774	.0496561	2.86	0.005	.0433237	.240431
/beta1	1.429188	.6246253	2.29	0.024	.18948	2.668897
/beta2	.7765569	.1224753	6.34	0.000	.5334773	1.019637
/delta0	.8424631	.0897482	9.39	0.000	.6643378	1.020588
/delta1	-.3268445	.0847928	-3.85	0.000	-.4951347	-.1585544

Parameter beta1 taken as constant term in model & ANOVA table

θ_1 ni baholash

```
scalar theta1 = 1-_b[alpha_cons]
```

```
scalar list theta1
```

```
. scalar list theta1  
theta1 = .85812265
```

Avvalgidek, kuchga ega bo'lmagan (y, x) kointegratsiyalanmagan. Kointegratsiya munosabatlari doimiy terminni o'z ichiga olganligi sababli, kritik qiymat -3,37 ga teng. Hisoblangan qiymatni (-3,912) kritik qiymat bilan taqqoslab, biz nol gipotezani rad etamiz va (y, x) kointegratsiyalashgan degan xulosaga kelimiz.

O'n ikkinchi bob mavzularini mustahkamlash uchun savollar:

1. Statsion va nostatsionar ma'lumotlarni o'rganishning asosiy maqsadi nima?
2. Statada sanalarni yaratish qanday amalga oshiriladi?
3. Sanalarni formatlash qanday amallar ketma-ketligi bajarilishi kerak?
4. Gdp seriyasining grafigini tuzish uchun qanday buyruqdan foydalaniladi?
5. graph combine buyrug'i qanday maqsadda qo'llaniladi?
6. Agar summarize buyrug'idan keyin o'zgaruvchilar nomlari ro'yxatga kiritilmagan bo'lsa qanday amal bajarilishi lozim?
7. Soxta regressiyalar haqida nimalarni bilib oldingiz?
8. Dikki-Fuller testlaridan nima maqsadda foydalanish mumkin?
9. Integratsion va kointegratsion tushunchalariga ta'rif bering.
10. Engle-Granger testi nima uchun qo'llaniladi?

O'n ikkinchi bob mavzularida qo'llanilgan buyruqlar va atamalar

augmented DF test	dofm	q(1996q4)
cointegration	drift	name(graph,
combine graph	Engle-Granger test	replace)
date functions	estat bgodfrey	nonstationary
dfuller	format %tm	shell stationary
Dickey-Fuller (DF)	format %tq	time series plots
test	integration	tin(d1,d2)
difference operator,	lag operator, L.	trend
D.	nl	tsline
dofq	noconstant	tsvarlist

O'n ikkinchi bob mavzularini mustahkamlash uchun masalalarning do-file lari

Ma'lumotlar buyrug'i: `cd c:\data\poe4stata`

1-masala. log file ni ochish va datalardan foydalanish.

`log using chap12, replace text`

`use usa, clear`

* -----

* Create dates and declare time-series

* -----

`generate date = q(1984q1) + _n-1`

`format date %tq`

`tsset date`

* -----

* Extract dates with year and quarter

* -----

`gen double newdate = dofq(date)`

`gen y = year(newdate)`

`gen q = quarter(newdate)`

`list date y q in 1/9`

* -----

* Graph time-series

- * Graphs are named with replace option
- * and combined.

```
* -----
qui tsline gdp, name(gdp, replace)
qui tsline D.gdp, name(dgdp, replace)
graph combine gdp dgdp
qui tsline inf, name(inf, replace)
qui tsline D.inf, name(dinf, replace) yline(0)
qui tsline f, name(f, replace)
qui tsline D.f, name(df, replace) yline(0)
qui tsline b, name(b, replace)
qui tsline D.b, name(db, replace) yline(0)
graph combine inf dinf f df b db, cols(2)
```

- * Two ways to limit dates

```
summarize if date<=q(1996q4)
```

```
summarize if date>=q(1997q1)
```

```
summarize if tin(,1996q4)
```

```
summarize if tin(1997q1,)
```

- * To get summary stats for all variables and differences without generate

```
summarize gdp inf b f D.gdp D.inf D.b D.f if tin(1984q2,1996q4)
```

```
summarize gdp inf b f D.gdp D.inf D.b D.f if tin(1997q1,)
```

```
summarize
```

```
* -----
```

2-masala. **Spurious Regression tenglamalar tuzish.**

```
* -----
```

```
use spurious, clear
```

```
gen time = _n
```

```
tsset time
```

```
regress rw1 rw2
```

```
estat bgodfrey
```

```
tsline rw1 rw2, name(g1, replace)
```

```
scatter rw1 rw2, name(g2, replace)
```

```
regress rw1 rw2
```

```
estat bgodfrey
```

```
* -----
```

3-masala. **Unit root tests va cointegration amallarini bajarish.**

```

* -----
use usa, clear
gen date = q(1984q1) + _n - 1
format %tq date
tsset date
* Augmented Dickey Fuller Regressions
regress D.f L.f L.D.f
regress D.b L.b L.D.b
* Augmented Dickey Fuller Regressions with built in functions
dfuller f, regress lags(1)
dfuller b, regress lags(1)
* ADF on differences
dfuller D.f, noconstant lags(0)
dfuller D.b, noconstant lags(0)
* DF-GLS tests
dfglsl f
dfglsl b
* Phillips-Perron tests
pperron f, regress trend
pperron b, regress trend
* Engle Granger cointegrations test
regress b f
predict ehat, residual
regress D.ehat L.ehat L.D.ehat, noconstant
* Using the built-in Stata commands
dfuller ehat, noconstant lags(1)
drop ehat
gen Db=D.b
nl(Db=-{alpha}*(L.b-{beta1}-{beta2}*L.f)+{delta0}*D.f+{delta1}*D.L.f), ///
variables(L.b L.f D.L.f)
scalar theta1 = 1-_b[alpha:_cons]
scalar list theta1
gen ehat = L.b - _b[beta1:_cons]-_b[beta2:_cons]*L.f
qui reg D.ehat L.ehat L.D.ehat, noconst
di _b[L.ehat]/_se[L.ehat]
log close

```

XIII BOB. VEKTOR XATOLARINI TUZATISH VA VEKTOR AVTOREGRESSIV MODELLARI

13.1. VEC va VAR modellari

Vektor avtoregressiv (VAR) modeli statsionar o'zgaruvchilar o'rtasidagi dinamik o'zaro bog'liqlikni tavsiflash uchun ishlatiladigan umumiy asosdir. Shunday qilib, vaqt seriyasini tahlil qilishda birinchi qadam ma'lumotlar darajalari statsionar yoki yo'qligini aniqlash bo'lishi kerak. Agar yo'q bo'lsa, seriyaning birinchi farqlarini oling va qayta urinib ko'ring. Odatda, vaqt seriyangizning darajalari (yoki log darajalari) statsionar bo'lmasa, birinchi farqlar bo'ladi.

Agar vaqt seriyasi statsionar bo'lmasa, seriyalar o'rtasidagi munosabatlarning izchil bahosini ta'minlash uchun VAR tuzilishini o'zgartirish kerak. Vektor xatosini tuzatish (VEC) modeli farqlarida statsionar bo'lgan o'zgaruvchilar uchun VARning maxsus holatidir (ya'ni, I(1)). VEC shuningdek, o'zgaruvchilar o'rtasidagi har qanday kointegratsiya munosabatlarini hisobga olishi mumkin.

Ikki vaqt seriyali o'zgaruvchilarni ko'rib chiqing, yt va xt. IX bobdagi dinamik munosabatlar haqidagi munozarani ushbu ikki o'zaro bog'liq o'zgaruvchilarga umumlashtirish tenglamalar tizimini beradi:

$$y_t = \beta_{10} + \beta_{11}y_{t-1} + \beta_{12}x_{t-1} + v_t^y$$

$$x_t = \beta_{20} + \beta_{21}y_{t-1} + \beta_{22}x_{t-1} + v_t^x$$

Tenglamalar har bir o'zgaruvchining o'ziga xos kechikishi va tizimdagi boshqa o'zgaruvchining kechikish funksiyasi bo'lgan tizimni tavsiflaydi. Bunday holda, tizim ikkita y va x o'zgaruvchini o'z ichiga oladi. Tenglamalar birgalikda vektor avtoregressiyasi (VAR) deb nomlanuvchi tizimni tashkil qiladi. Ushbu misolda maksimal kechikish birinchi darajali bo'lgani uchun bizda VAR (1) mavjud.

Agar y va x statsionar bo'lsa, tizimni har bir tenglamaga qo'llaniladigan eng kichik kvadratlar yordamida baholash mumkin. Agar y va x o'z darajalarida statsionar bo'lmasa, lekin farqlarda

statsionar bo'lsa (ya'ni, I (1)), u holda farqlarni oling va baholang:

$$\Delta y_t = \beta_{11}\Delta y_{t-1} + \beta_{12}\Delta x_{t-1} + v_t^{\Delta y}$$

$$\Delta x_t = \beta_{21}\Delta y_{t-1} + \beta_{22}\Delta x_{t-1} + v_t^{\Delta x} \text{ eng}$$

kichik kvadratlardan foydalanish.

Agar y va x I (1) ga teng bo'lsa va kointegratsiyalashgan bo'lsa, u holda tenglamalar tizimini I (1) o'zgaruvchilar orasidagi kointegratsiya munosabatini hisobga olgan holda o'zgartirish mumkin. Kointegratsiya munosabatlarini joriy etish vektor xatosini tuzatish (VEC) modeli deb nomlanuvchi modelga olib keladi.

13.2. VEC modelini baholash

Birinchi misolda Avstraliya va AQSH yalpi ichki mahsuloti ma'lumotlari VEC modelini baholash uchun ishlatiladi. Biz vektor xatosini tuzatish modelidan foydalanishga qaror qildik, chunki (1) vaqt seriyalari o'z darajalarida statsionar emas, balki ularning farqlarida (2) o'zgaruvchilar kointegratsiyalashgan. Bizning dastlabki taassurotlarimiz ikki seriyaning sujetlarini ko'rib chiqishdan olingan. Boshlash uchun katalogni ma'lumotlaringiz bo'lgan katalogga o'zgartiring va ma'lumotlarni yuklang. Ushbu mashqda biz gdp.dta ma'lumotlaridan foydalanamiz:

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

```
use gdp, clear
```

Ma'lumotlar ikki choraklik vaqt seriyasini o'z ichiga oladi: Avstraliya va AQSHning 1970q1 dan 2004q4 yillardagi yalpi ichki mahsuloti. Xuddi XIII bobda qilganingizdek, choraklik sanalar ketma-ketligini yarating:

```
gen date = q(1970q1) + _n - 1
```

```
format %tq date
```

```
tsset date
```

Ikki YalM seriyasining darajalari va farqlarini chizish ma'lumotlar darajalari bo'yicha statsionar emas, lekin farqlar bo'yicha statsionar ekanligini ko'rsatadi. Ushbu misolda biz tsline buyrug'ini ixtiyoriy sxema bilan ishlatdik. Sxemada keyinchalik foydalanish uchun saqlangan grafik afzalliklari mavjud. Siz o'zingizni yaratishingiz

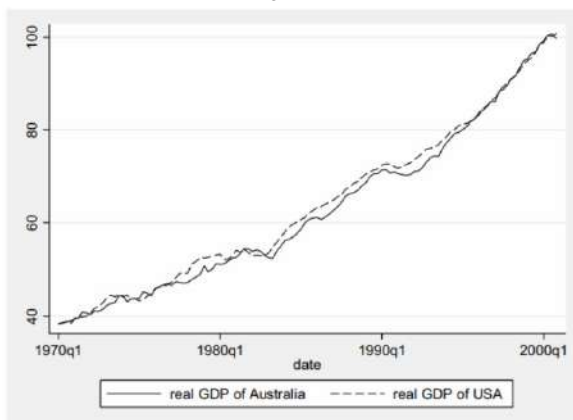
yoki Stata bilan o'rnatilganlardan birini ishlatishingiz mumkin. Buyruqlar satrida siz yozish orqali kompyuteringizda qaysi sxemalar o'rnatilganligini aniqlashdan foydalanishingiz mumkin:

graph query, schemes

Ushbu misolda men Stata Journal degan ma'noni anglatuvchi sj deb nomlangan sxemadan foydalandim. Bu xuddi shu yerda chop etilganlarga o'xshash grafiklarni ishlab chiqaradi. Bunday holda, u ikki xil chiziqli ta'rifga ega kulrang chiziqli diagramma hosil qiladi: aus uchun qattiq chiziqlar va usa uchun kesilgan chiziqlar. Ixtiyoriy sxemaga ega grafiklar uchun to'liq sintaksis:

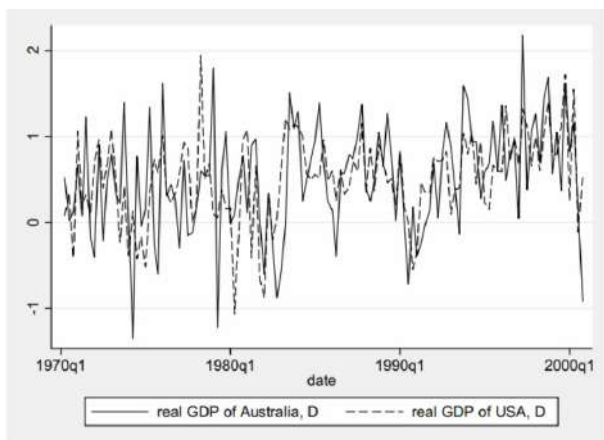
tsline aus usa, scheme(sj)

tsline D.aus D.usa, scheme(sj)



Ikkala seriya ham o'z darajalarida statsionar ko'rinmaydi. Ular umumiy tendensiyaga ega bo'lib ko'rinadi, bu ular kointegratsiyalangan bo'lishi mumkinligidan dalolat beradi.

Birlik ildiz testlari XII bobda muhokama qilingan protseduralar yordamida amalga oshiriladi. Kengaytirilgan Dikki-Fuller regressiyalari spetsifikatsiya haqida ba'zi mulohazalarni talab qiladi. Foydalanuvchi doimiy, trend yoki drift va doimiy Dikki-Fuller regressiyalarini oshiradigan farqlar uchun kechikish uzunliklarini kiritish haqida qaror qabul qilishi kerak. Farqlar grafikda ko'rsatilgan va bu spetsifikatsiya haqida ba'zi maslahatlar beradi. Quyidagi grafikda tendensiya yoki driftning ozgina dalillari ko'rsatilgan.



Kechikish uzunligini model tanlash qoidalari yoki maksimal kechikish uzunligidan, masalan, 4 dan boshlash va oxirgi kechikishdagi t nisbati ahamiyatli bo'lgunga qadar kechikishlarni birma-bir yo'q qilish orqali tanlash mumkin.

dfuller aus, regress lags(1)

dfuller usa, regress lags(3)

Yo'q qilish jarayoni orqali konstantani (keraksiz ko'rinsa ham) kiritish va aus uchun 1 lag va usa seriyasi uchun 3 lagni kiritish to'g'risida qaror qabul qilinadi. Men baholagan har qanday ADF regressiyasida deyarli hech bir ADF statistikasi 5% darajasida ahamiyatli emas edi. Seriyalar darajalari bo'yicha statsionar emasligidan qoniqish hosil qilib, ularning kointegratsiyasi o'rganiladi.

```
. dfuller aus, regress lags(1)
Augmented Dickey-Fuller test for unit root      Number of obs   =      122

```

	Test Statistic	1% Critical Value	Interpolated Dickey-Fuller 5% Critical Value	10% Critical Value
Z(t)	2.658	-3.503	-2.889	-2.579

MacKinnon approximate p-value for Z(t) = 0.9991

```
. dfuller usa, regress lags(3)
Augmented Dickey-Fuller test for unit root      Number of obs   =      120

```

	Test Statistic	1% Critical Value	Interpolated Dickey-Fuller 5% Critical Value	10% Critical Value
Z(t)	1.691	-3.503	-2.889	-2.579

MacKinnon approximate p-value for Z(t) = 0.9981

Har bir holatda, nostatsionarlikning nol gipotezasini har qanday oqilona darajadagi ahamiyatlilik darajasida rad etib bo'lmaydi. E'tibor bering, ikkala kechikish farqlari AQSH tenglamasida muhim va Avstraliya tenglamasida 3-lag muhim; muhim kechikish koeffitsiyentlari kechikkan o'zgaruvchilarni kiritish kerakligi haqida ba'zi dalillarni beradi.

Keyinchalik, eng kichik kvadratlar yordamida kointegratsiya tenglamasini baholang. E'tibor bering, kointegratsiya aloqasi doimiyni o'z ichiga olmaydi.

regress aus usa, noconst

. reg aus usa, noconst

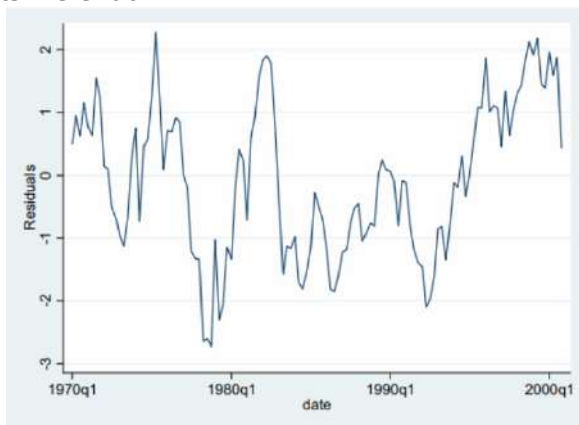
Source	SS	df	MS
Model	526014.204	1	526014.204
Residual	182.885542	123	1.48687433
Total	526197.09	124	4243.52492

Number of obs = 124
 F(1, 123) = .
 Prob > F = 0.0000
 R-squared = 0.9997
 Adj R-squared = 0.9996
 Root MSE = 1.2194

	aus	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
	usa	.9853495	.0016566	594.79	0.000	.9820703 .9886288

Qoldiqlar kointegratsiyaning Engle-Granjer testini o'tkazish uchun saqlanadi va chiziladi.

predict ehat, residual
tsline ehat



Qoldiqlar nolga teng bo'lib, tendensiyanı ko'rsatmaydi. Nihoyat, saqlangan qoldiqlar yordamchi regressiyada ishlatiladi.

$$\Delta \hat{e}_t = 0 \hat{e}_{t-1} + v_t$$

Stata buyrug'i:

regress D.ehat L.ehat, noconstant

. reg D.ehat L.ehat, noconst

Source	SS	df	MS	Number of obs =	123
Model	2.99032657	1	2.99032657	F(1, 122) =	8.35
Residual	43.7006336	122	.358201914	Prob > F =	0.0046
				R-squared =	0.0640
				Adj R-squared =	0.0564
Total	46.6909601	123	.379601302	Root MSE =	.5985

D.ehat	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
ehat L1.	-.1279366	.0442792	-2.89	0.005	-.2155916 -.0402816

T nisbati -2,89 ga teng. Kesishsiz kointegratsiya aloqasi uchun 5% kritik qiymat -2,76 ni tashkil qiladi va shuning uchun bu testning rad etish hududiga to'g'ri keladi. Kointegratsiya yo'qligi haqidagi nol gipoteza 5% ahamiyatlilik darajasida rad etiladi.

Iqtisodiy shoklarga real YaIMning chorak javobini o'lchash uchun vektor xatolarni tuzatish modelining parametrlarini eng kichik kvadratlar bo'yicha baholaymiz:

regress D.aus L1.ehat

regress D.usa L1.ehat

Xatolarni tuzatish modeli Avstraliya yalpi ichki mahsulotining natijalari:

Source	SS	df	MS	Number of obs =	123
Model	1.77229686	1	1.77229686	F(1, 121) =	4.32
Residual	49.697821	121	.410725793	Prob > F =	0.0399
				R-squared =	0.0344
				Adj R-squared =	0.0265
Total	51.4701178	122	.421886212	Root MSE =	.64088

D. aus	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
ehat L1.	-.0987029	.0475158	-2.08	0.040	-.1927729 -.0046329
_cons	.4917059	.0579095	8.49	0.000	.3770587 .606353

$\hat{e}_{t-1}$ dagi muhim salbiy koeffitsiyent Avstraliya YaIM AQSH va Avstraliya o'rtasidagi vaqtinchalik muvozanatsizlikka javob berishini ko'rsatadi. AQSH uchun:

Source	SS	df	MS			
Model	.166467786	1	.166467786			
Residual	32.2879333	121	.266842424			
Total	32.4544011	122	.266019681			

Number of obs =	123
F(1, 121) =	0.62
Prob > F =	0.4312
R-squared =	0.0051
Adj R-squared =	-0.0031
Root MSE =	.51657

D.usa	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ehat						
L1.	.0302501	.0382992	0.79	0.431	-.0455732	.1060734
_cons	.5098843	.0466768	10.92	0.000	.4174752	.6022934

Qo'shma Shtatlar ikki iqtisod o'rtasidagi muvozanatsizlikka javob bermayapti; t -1 dagi t nisbati ahamiyatsiz. Ushbu natijalar Avstraliyadagi iqtisodiy sharoitlar AQSHdagi sharoitlardan ko'ra ko'proq AQSHdagilarga bog'liq degan fikrni qo'llab-quvvatlaydi. Ikki iqtisodiy savdoning oddiy modelida AQSH yirik yopiq iqtisodiyotdir va Avstraliya kichik ochiq iqtisodiyotdir.

13.3. VAR modelini baholash

Vektor avtoregressiv modeli (VAR) VEC modelidan ko'ra baholash uchun aslida osonroqdir. U o'zgaruvchilar o'rtasida kointegratsiya bo'lmaganda qo'llaniladi va ularning statsionar qiymatlariga aylantirilgan vaqt seriyalari yordamida baholanadi. POE4 misolida bizda 1960:1 dan 2009:4 gacha bo'lgan davrda AQSH iqtisodiyoti uchun mavjud bo'lgan real shaxsiy daromadlar (Y bilan belgilanadi) va real shaxsiy iste'mol xarajatlari jurnali (C bilan belgilanadi) fred.dta ma'lumotlar to'plamida makroiqtisodiy ma'lumotlar jurnali mavjud. Oldingi misolda bo'lgani kabi, birinchi qadam o'zgaruvchilar doimiy yoki yo'qligini aniqlashdir. Agar ular bo'lmasa, ularni farqlang va farqlar mos kelishini tekshiring (ya'ni birlashtirilganligiga). Keyinchalik, kointegratsiyani sinab ko'ring. Agar ular kointegratsiyalashgan bo'lsa, VEC modelini baholang. Agar yo'q bo'lsa, VAR modelini baholash uchun farqlar va kechiktirilgan farqlardan foydalaning. Birinchidan, katalogni ma'lumotlaringizni o'z ichiga olgan katalogga o'zgartiring va ma'lumotlarni yuklang. Ushbu mashqda biz fred.dta ma'lumotlaridan foydalanamiz:

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

```
use fred, clear
```

Ma'lumotlar choraklik bo'lib, 1960q1da boshlanadi va

2009q4 gacha cho'ziladi. Xuddi XII bobda va yuqoridagi misolda qilganimizdek, chorak sanalari ketma-ketligi:

```
gen date = q(1960q1) + _n - 1
```

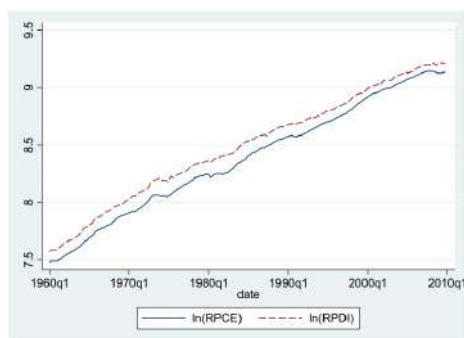
```
format %tq date
```

```
tsset date
```

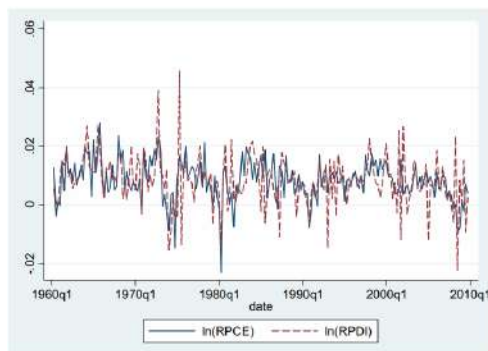
Birinchi qadam statsionarlik testlariga konstantalar yoki tendensiyalarni kiritish kerakligini aniqlash uchun qatorni chizishdir. Darajalar ham, farqlar ham chizilgan.

```
tsline c y, legend(lab (1 "ln(Consumption)" ) lab(2 "ln(PDI)"))
```

```
tsline D.c D.y, legend(lab (1 "D.ln(Consumption)" ) lab(2 "D.ln(PDI)"))
```



Darajalar qatori birga ketayotganga o'xshaydi. Farqlar aniq tendensiyani ko'rsatmaydi, lekin ketma-ket o'rtacha noldan katta bo'lib ko'rinadi, bu ADF regressiyalariga doimiy kiritilganligini ko'rsatadi.



Qabul qilinishi kerak bo'lgan yana bir qaror – bu Diki Fullerning ortib borayotgan regressiyalariga qancha kechikish o'zgarishlarini kiritishdir. ADF gradiyenti qoldiqlari avtomatik ravishda birlashtirilmasligi uchun amal qilish kerak bo'lgan prinsip faqat yetarli miqdorda qo'shilishi kerak. Shunday qilib, kechikishlarsiz asosiy regressiyadan boshlang, DF regressiyasini baholang, keyin qoldiqlarning avtokorrelatsiyalanganligini aniqlash uchun IX bobda muhokama qilingan LM testidan foydalaning. Qolgan qiymatlar orasidagi avtokorrelyatsiyani olib tashlash uchun yetarli kechikishlarni qo'shing. Agar bu strategiya Statada qo'llanilsa, ADF regressiyalari aniq baholanishi kerak bo'ladi; estat bgodfrey buyrug'i, agar dfuller dan keyin chiqarilgan bo'lsa, to'g'ri regressiyaga asoslanmaydi.

ADF testlari uchun regressiyalar:

qui reg L(0/1).D.c L.c

estat bgodfrey, lags(1 2 3)

qui reg L(0/2).D.c L.c

estat bgodfrey, lags(1 2 3)

qui reg L(0/3).D.c L.c

estat bgodfrey, lags(1 2 3)

Oxirgi ikki regressiya uchun test natijalari quyida ko'rinadi.

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
1	2.077	1	0.1495
2	2.539	2	0.2810
3	2.542	3	0.4677

H0: no serial correlation

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
1	0.157	1	0.6916
2	1.271	2	0.5297
3	2.098	3	0.5523

H0: no serial correlation

ADF (2) qoldiqlari avtomatik ravishda bog'langani va ADF (3) qoldiqlari bog'lanmaganligi aniq. Olingan ADF statistikasi quyidagilar yordamida olinadi:

dfuller c, lags(3)

Bu yerda ko'rsatilgan kechikishlar soni qo'llaniladi.

```
. dfuller c, lags(3)
```

Augmented Dickey-Fuller test for unit root Number of obs = 196

	Test Statistic	1% Critical Value	Interpolated Dickey-Fuller 5% Critical Value	10% Critical Value
Z(t)	-1.995	-3.478	-2.884	-2.574

Mackinnon approximate p-value for Z(t) = 0.2886

Shuni ham yodda tutingki, bu regressiya doimiylikni o'z ichiga oladi va test statistikasi -1.995 bo'ladi. Birlik ildiz gipotezasi 5% darajasida rad etilmaydi.

Stata sikli y uchun bir xil amalni bajarish uchun ishlatiladi.

for values p = 1/3 {

qui reg L(0/'p').D.y L.y

di "Lags =" 'p'

estat bgodfrey, lags(1 2 3)

}

Loop makro p ni 1 dan 3 gacha oshiradi. Chiqishni to'xtatish uchun quietly bekor qilish operatsiyasi ishlatiladi. Chiziqli regressiya uchun qisqartirilgan kodlash qo'llaniladi, bu D.y bog'liq o'zgaruvchini kechikkan regressorlar bilan bitta bayonotda birlashtiradi, ya'ni $\text{regL}(0/'p').D.y$. $p=1$ uchun bu reg D.y L.D.y ga teng. LM test natijalarini chop etishdan oldin Stataga joriy p qiymatini chop etishni buyurdim. Natija:

Lags =1

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
1	0.208	1	0.6487
2	2.853	2	0.2401
3	2.880	3	0.4105

H0: no serial correlation

Lags =2

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
1	2.077	1	0.1495
2	2.539	2	0.2810
3	2.542	3	0.4677

H0: no serial correlation

Lags =3

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
1	0.157	1	0.6916
2	1.271	2	0.5297
3	2.098	3	0.5523

H0: no serial correlation

y dagi kechikuvchi tafovutlarni regressiya omillari (ya'ni, oddiy Dikki-Fuller regressiyasi) sifatida kiritish kerakligi haqida hech qanday dalil yo'q. Dikki Fuller testi natijalari:

. dfuller y, lags(0)

Dickey-Fuller test for unit root

Number of obs = 199

Test Statistic	Interpolated Dickey-Fuller		
	1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-2.741	-3.477	-2.883

MacKinnon approximate p-value for Z(t) = 0.0673

Eslatib o'tamiz, kointegratsiya munosabatlarini eng kichik kvadratlar yordamida baholash mumkin:

$$C_t = \beta_1 + \beta_2 Y_t + v_t$$

Ushbu regressiyadan qoldiqlar olinadi va ularning o'zgarishlari kechikish qiymati bo'yicha regressiya qilinadi:

$$\Delta \hat{e}_t = \gamma \hat{e}_{t-1} + \delta \Delta \hat{e}_{t-1} + v_t$$

Ushbu harakat uchun Stata buyrug'i:

reg c y

predict ehat, res
 reg D.ehat L.ehat D.L.ehat, noconst
 di _b[L.ehat]/_se[L.ehat]
 va natijalar:

```
. di _b[L.ehat]/_se[L.ehat]
-2.8728997
```

E'tibor bering, bu yerda ixtiyoriy daromaddan mustaqil bo'lgan (logarifmik) iste'mol komponentini olish uchun kesishish atamasi kiritilgan. 5% kritik qiymat kointegratsiya qoldiq barqarorligi testi uchun -3.37. -2.873 ning birlik ildiz t qiymati -3.37 dan katta bo'lganligi sababli, bu xatolar statsionar emasligini ko'rsatadi va demak, C (ya'ni, $\ln(RPCE)$) va Y (ya'ni, $\ln(RPDI)$) o'rtasida soxta bog'liqlik mavjud, ya'ni bizda kointegratsiya yo'q. Bunday holda, variatsiyalarda VAR yordamida model koeffitsiyentlarini baholang.

Statada VARni hisoblash oson. Eng oson yo'l – varbasic buyrug'idan foydalanish.

varbasic asosiy vektor avtoregressiv (VAR) modeliga mos keladi va impuls javob funksiyalari (IRF) va prognoz xato dispersiyalarining (FEVDs) grafiklarini tuzadi. Tenglamalarning o'zgarishidagi statik VAR ning asosiy tuzilishi quyida keltirilgan:

$$\Delta y_t = \beta_{11}\Delta y_{t-1} + \beta_{12}\Delta x_{t-1} + v_t^{\Delta y}$$

$$\Delta x_t = \beta_{21}\Delta y_{t-1} + \beta_{22}\Delta x_{t-1} + v_t^{\Delta x}$$

x_t va y_t o'zgaruvchilari statsionar emas, lekin farqlar statsionardir. Har bir farq tizimdagi boshqa o'zgaruvchilarning o'ziga xos kechikkan farqlari va kechikkan farqlarining chiziqli funksiyasidir. Tenglamalar chiziqli bo'lib, parametrlarni baholash uchun eng kichik kvadratlardan foydalanish mumkin. Varbasic buyrug'i buni soddalashtiradi. Tizimdagi o'zgaruvchilarni (Δy_t va Δx_t) va modelning o'ng tomoniga kiritish uchun kechikishlar sonini ko'rsatishingiz kerak. Bizning misolimizda faqat 1 kechikish kiritilgan va VARni baholash sintaksisi:

varbasic D.c D.y, lags(1/1) step(12) nograph

Lags (1/1) sintaksisi Stataga birinchi raqamdan oxirgi

raqamgacha kechikishlarni kiritishni aytadi, bu holda bu 1 dan kechikish 1 ga teng. Agar VAR bir kechikishdan uzoqroq bo'lsa, u buni o'zgartiradi. Step (12) varianti ham qo'shildi. Ushbu parametr impulsli javoblarni (IRF) hisoblash va xato diskompozitsiyasini (FEVD) prognoz qilish uchun kechiktirilgan davrlar sonini cheklash uchun ishlatiladi – foydalanilganda grafiklarni sharhlashni osonlashtiradi va Stata yaratishi mumkin bo'lgan jadvallarda bo'sh joyni tejaydi. Nihoyat, nograph varianti IRF va FEVD grafiklarini bostirish uchun ishlatiladi. Bularni keyinroq quyida aytilgandek alohida iboralarda chaqirish mumkin. Bundan chiqadigan narsa:

```
. varbasic D.c D.y, lags(1/1) step(12) nograph
Vector autoregression

Sample: 1960q3 - 2009q4
Log likelihood = 1400.444
FPE = 2.62e-09
Det(Sigma_ml) = 2.46e-09

No. of obs = 198
AIC = -14.0853
HQIC = -14.0496
SBIC = -13.98565
```

Equation	Parms	RMSE	R-sq	chi2	P>chi2
D_c	3	.006575	0.1205	27.12459	0.0000
D_y	3	.008562	0.1118	24.92656	0.0000

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
D_c					
C					
LD.	.2156068	.0741801	2.91	0.004	.0702164 .3609972
Y	.1493798	.0572953	2.61	0.009	.0370832 .2616765
_cons	.0052776	.0007516	7.02	0.000	.0038046 .0067507
D_y					
C					
LD.	.4754276	.0965863	4.92	0.000	.286122 .6647332
Y	-.2171679	.0746013	-2.91	0.004	-.3633839 -.070952
_cons	.0060367	.0009786	6.17	0.000	.0041187 .0079547

Dikki-Fuller regressida uzoqroq kechikishlar ishlatilganligi sababli, VAR ham uzoqroq kechikishlarga ega bo'lishi mumkin. Amalda, VAR qoldiqlarini avtokorrelatsiya uchun sinab ko'rish yaxshi fikr bo'lar edi. Varbasic dan keyin chiqarilgan Stata varlmar buyrug'i avtokorrelyatsiya uchun bajarilganlarga o'xshash qoldiqlarning LM testini amalga oshiradi.

```
. varlmar
```

Lagrange-multiplier test

lag	chi2	df	Prob > chi2
1	9.5086	4	0.04957
2	5.6784	4	0.22449

H0: no autocorrelation at lag order

Qoldiqlarda avtokorrelatsiya haqida dalillar mavjud, chunki kechikish 1 da p qiymati 5%dan kam. Kechikish uzunligini 3 ga uzaytirish noaniqlikni yo'q qiladi.

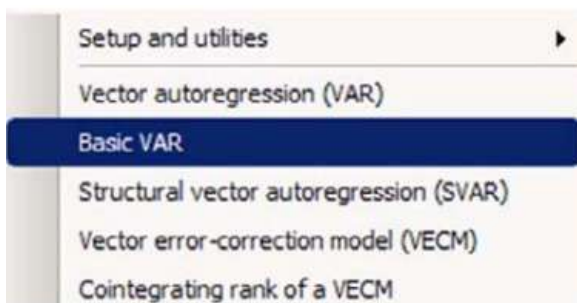
Stata yana bir protsedurani o'z ichiga oladi, bu VAR modellarida kechikish uzunligini tanlashni juda osonlashtiradi. Varsoc buyrug'i yakuniy bashorat qilish xatosi (FPE), Akaike ma'lumot mezoni (AIC), Shvarsning Bayes ma'lumot mezoni (SC) va bir qator vektor avtoregressiyalari uchun Xannan va Quinn axborot mezonlari (HQIC) kechikish tartibini tanlash statistikasi haqida xabar beradi. Bu noma'lum tartibdagi VAR yoki VEC modellarini uchun kechikish uzunliklarini topish uchun ishlatilishi mumkin. Yuqoridagi misol uchun Stata hosil qiladi:

```
. varsoc D.c D.y, maxlag(4)
```

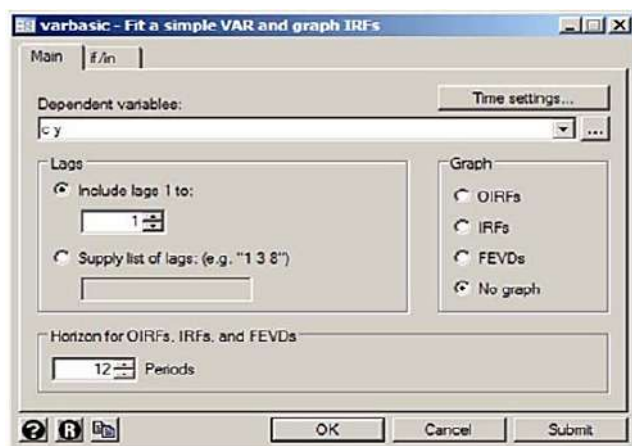
Selection-order criteria					Number of obs		=		195
Sample: 1961q2 - 2009q4									
lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC	
0	1355.02				3.2e-09	-13.8772	-13.8636	-13.8436	
1	1379.09	48.129	4	0.000	2.6e-09	-14.083	-14.0422*	-13.9823*	
2	1383.92	9.6655*	4	0.046	2.6e-09	-14.0915	-14.0235	-13.9237	
3	1388.24	8.6379	4	0.071	2.6e-09*	-14.0948*	-13.9996	-13.8598	
4	1391.6	6.7149	4	0.152	2.6e-09	-14.0882	-13.9659	-13.7861	

AIC 3 ta kechikish tartibini tanlaydi, SC esa (Stata, SBIC tomonidan belgilangan) 1 ni tanlaydi.

Varbasic uchun dialog oynasi Statistika>Multivariate time series>Basic VAR ni topadi.



Bu dialog oynasi



Ikkala c va y ni bog'liq o'zgaruvchilar sifatida ro'yxatlang va Include lags 1to etiketli radio tugmani bosib; POE4 da natijalarni chiqarish uchun ushbu tugma ostidagi maydondagi raqam 1 ga o'rnatilishi kerak, ammo biz ushbu misolda uni uchtaga qo'yganmiz. Qutining pastki chap qismida biz impulsli javoblar va prognoz xatosi dispersiyalari uchun 12 davrli horizon ni tanlab. OK tugmasini bosib.

13.4. Impuls javoblari va variatsiyalarni ayirishlari

Impuls javob funksiyalari o'zgaruvchilarni sozlash yo'liga zarba ta'sirini ko'rsatadi. Prognoz xatosi tafovutining parchalanishi har bir zarba turining prognoz xatosi farqiga qo'shgan hissasini o'lchaydi. Ikkala hisoblash ham iqtisodiy o'zgaruvchilarga zarbalar tizim orqali qanday aks etishini baholashda foydalidir.

Impuls-javob funksiyalari (IRF) va prognoz xatoliklarning parchalanishi (FEVD) varbasic buyrug'idan foydalangandan so'ng ishlab chiqarilishi mumkin. Natijalar jadval yoki grafik shaklida taqdim etilishi mumkin. Ushbu misolda biz ikkalasini ham tasvirlaymiz. Varbasic buyrug'idan so'ng biz IRF va FEVDlarni yaratish uchun irf table buyrug'idan foydalanamiz:

irf table irf

irf table fevd

irf table sintaksisi:

irf table [stat] [, options]

Bu yerda stat quyidagilardan biri bo'lishi mumkin:

irf impulse-response function

oirf orthogonalized impulse-response function
multiplier function

cirf cumulative impulse-response function

coirf cumulative orthogonalized impulse-response function

cdm cumulative dynamic-multiplier function

fevd Cholesky forecast-error variance decomposition

sirf structural impulse-response function

sfevd structural forecast-error variance decomposition

IRF jadvalidagi natijalar:

step	(1) irf	(1) Lower	(1) Upper	(2) irf	(2) Lower	(2) Upper
0	1	1	1	0	0	0
1	.215607	.070216	.360997	.475428	.286122	.664733
2	.117506	.042463	.192549	-.000742	-.088665	.087181
3	.025224	-.014994	.065442	.056027	.007234	.104819
4	.013808	-.003806	.031422	-.000175	-.020898	.020548
5	.002951	-.004713	.010615	.006603	-.002593	.015799
6	.001623	-.001478	.004723	-.000031	-.003694	.003632
7	.000345	-.000905	.001595	.000778	-.000719	.002275
8	.000191	-.000295	.000676	-4.9e-06	-.000581	.000571
9	.00004	-.000148	.000228	.000092	-.000134	.000317
10	.000022	-.000049	.000094	-7.2e-07	-.000086	.000084
11	4.7e-06	-.000022	.000032	.000011	-.000022	.000043
12	2.6e-06	-7.4e-06	.000013	-1.0e-07	-.000012	.000012

step	(3) irf	(3) Lower	(3) Upper	(4) irf	(4) Lower	(4) Upper
0	0	0	0	1	1	1
1	.14938	.037083	.261676	-.217168	-.363384	-.070952
2	-.000233	-.027858	.027392	.118181	.043031	.193332
3	.017604	-.000946	.036153	-.025776	-.066356	.014804
4	-.000055	-.006566	.006456	.013967	-.003818	.031752
5	.002075	-.001028	.005177	-.003059	-.010845	.004726
6	-9.7e-06	-.001161	.001141	.001651	-.001506	.004807
7	.000244	-.000241	.00073	-.000363	-.001642	.000916
8	-1.5e-06	-.000182	.000179	.000195	-.000303	.000693
9	.000029	-.000043	.000101	-.000043	-.000237	.000151
10	-2.2e-07	-.000027	.000026	.000023	-.000051	.000097
11	3.4e-06	-6.9e-06	.000014	-5.1e-06	-.000033	.000023
12	-3.2e-08	-3.8e-06	3.7e-06	2.7e-06	-7.7e-06	.000013

95% lower and upper bounds reported

(1) irfname = varbasic, impulse = D.c, and response = D.c

(2) irfname = varbasic, impulse = D.c, and response = D.y

(3) irfname = varbasic, impulse = D.y, and response = D.c

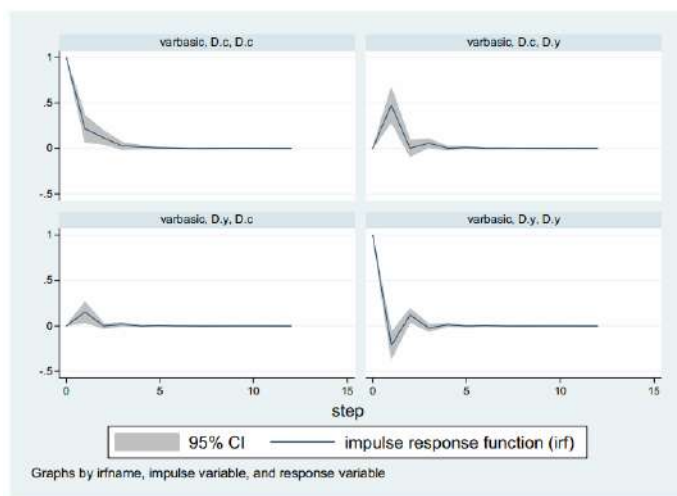
(4) irfname = varbasic, impulse = D.y, and response = D.y

Natijalar to'rtta kvadrantda ko'rinadi va jadvalning pastki qismida kalit berilgan. Birinchi kvadrantda (shimoli-g'arbiy) In (RCPE) da o'z-o'zidan zarbaga javob. Esingizda bo'lsin, bu seriya statsionar va shuning uchun zarbalar doimiy emas; ularning ta'siri

oxir-oqibat o'radi. $\ln(RPDI)$ ga zarbalar ham barqaror emas. Qizig'i shundaki, $\ln(RPDI)$ $\ln(RCPE)$ dagi zarbalarga qanday javob beradi va aksincha. 2-kvadrant (shimoli-sharqiy) bir davr uchun $\ln(RCPE)$ ning $\ln(RPDI)$ ga zarbalarini ko'rsatadi, lekin juda tez o'radi. $\ln(RRDI)$ ga zarbalar $\ln(RCPE)$ (3-kvadrant)da kichikroq, ammo sezilarli javob hosil qiladi, lekin juda tez nolga tushib qolsa ham.

irf graph deb nomlangan alohida buyruq asosan irf table bilan bir xil narsani bajaradi, faqat natijalar jadval ko'rinishida emas, balki grafik sifatida ko'rinadi.

irf graph



Bu faqat irf table tomonidan grafik shaklda ishlab chiqarilgan ma'lumotlar.

Impulslar javoblar singari, FEVD lar ham jadval yoki grafik shaklda ishlab chiqarilishi mumkin.

Buyruqlar navbati bilan:

irf table fevd

irf graph fevd

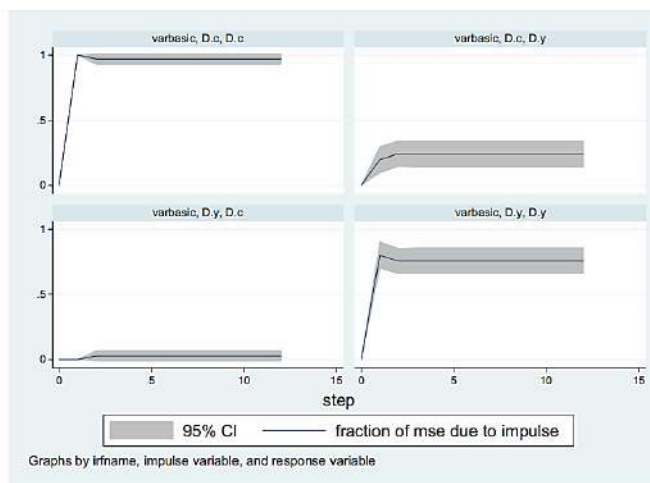
Natijalar xuddi shu tarzda, pastki qismida kalit bilan kvadrant bo'yicha joylashtirilgan.

step	(1) fevd	(1) Lower	(1) Upper	(2) fevd	(2) Lower	(2) Upper
0	0	0	0	0	0	0
1	1	1	1	.198552	.099067	.298037
2	.97297	.932729	1.01321	.2438	.146908	.340692
3	.973298	.933619	1.01298	.243198	.145308	.341087
4	.972967	.93271	1.01322	.243752	.145393	.34211
5	.972972	.932723	1.01322	.243743	.145351	.342135
6	.972967	.932709	1.01323	.24375	.145347	.342153
7	.972967	.932709	1.01323	.24375	.145346	.342154
8	.972967	.932709	1.01323	.24375	.145346	.342154
9	.972967	.932709	1.01323	.24375	.145346	.342154
10	.972967	.932709	1.01323	.24375	.145346	.342154
11	.972967	.932709	1.01323	.24375	.145346	.342154
12	.972967	.932709	1.01323	.24375	.145346	.342154

step	(3) fevd	(3) Lower	(3) Upper	(4) fevd	(4) Lower	(4) Upper
0	0	0	0	0	0	0
1	0	0	0	.801448	.701963	.900933
2	.02703	-.013212	.067271	.7562	.659308	.853092
3	.026702	-.012977	.066381	.756802	.658913	.854692
4	.027033	-.013225	.06729	.756248	.65789	.854607
5	.027028	-.01322	.067277	.756257	.657865	.854649
6	.027033	-.013225	.067291	.75625	.657847	.854653
7	.027033	-.013225	.067291	.75625	.657846	.854654
8	.027033	-.013225	.067291	.75625	.657846	.854654
9	.027033	-.013225	.067291	.75625	.657846	.854654
10	.027033	-.013225	.067291	.75625	.657846	.854654
11	.027033	-.013225	.067291	.75625	.657846	.854654
12	.027033	-.013225	.067291	.75625	.657846	.854654

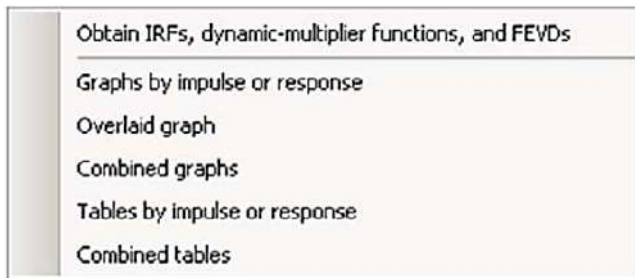
95% lower and upper bounds reported

- (1) irfname = varbasic, impulse = D.c, and response = D.c
 (2) irfname = varbasic, impulse = D.c, and response = D.y
 (3) irfname = varbasic, impulse = D.y, and response = D.c
 (4) irfname = varbasic, impulse = D.y, and response = D.y



IRF va FEVD lar uchun Statadagi jadval va grafik funksiyalari haqiqatan ham juda moslashuvchan. Til yoki dialog oynalari yordamida jadvallarni birlashtirish, grafiklarni joylashtirish va boshqa ko'p narsalarni qilishingiz mumkin. Yakuniy qutini ochish

uchun Statistics > Multivariate time series > IRF and FEVD analysis tanlang:



Bu yerda sizda bir qancha tanlovlar mavjud, shu jumladan biz muhokama qilganlarimiz. Shu bilan birga, siz ushbu dialog oynalaridan grafiklarni joylashtirish, grafiklarni birlashtirish yoki chiqishingizda foydalanish uchun jadval yoki grafiklarning muayyan qismlarini tanlash uchun ham foydalanishingiz mumkin. Bimalol tajriba qiling!

O'n uchinchi bob mavzularini mustahkamlash uchun savollar:

1. Vektor avtoregressiv (VAR) modeli nima uchun ishlatiladi?
2. Agar y va x statsionar bo'lsa, tizimni qanday baholash mumkin?
3. VEC modelini baholash qanday amalga oshiriladi?
4. tsline buyrug'i nima uchun qo'llaniladi?
5. Ixtiyoriy sxemaga ega grafiklar uchun to'liq sintaksis qanday ifodalanadi?
6. Engle-Granjer testida saqlangan qoldiqlar qay maqsadda ishlatiladi?
7. VAR modelini baholash deganda nimani tushunasiz?
8. Statada VARni hisoblashning eng oson yo'li qanday?
9. Impuls javob funksiyalari o'zgaruvchilarni sozlash yo'liga ta'sirini tushuntirib bering.
10. Impuls javob funksiyalari (IRF) va prognoz xatoliklarning parchalanishi qaysi buyruq yordamida amalga oshiriladi?

O'n uchinchi bob mavzularida qo'llanilgan buyruqlar va atamalar

ADF test	response function	varbasic
cointegration	IRF	varlmar
dfuller	irf graph	varsoc
drift	irf table	vector
Engle-Granger test	lags	autoregressive
estat bgodfrey	LM test	model
forecast error	multivariate time	vector error
variance	series	correction model
decomposition	scheme	trend
(FEVD) impulse	tsline	

O'n uchinchi bob mavzularini mustahkamlash uchun masalalarning do-file lari

Ma'lumotlar buyrug'i: `cd c:\data\poe4stata`

1-masala. **log file ni ochish.**

`log using chap13, replace text`

*-----

* Estimating a VECM

* Load the data and create a time variable

*-----

`use gdp, clear`

`gen date = q(1970q1) + _n - 1`

`format %tq date`

`tsset date`

*-----

* Plot the series to identify constants

* and trends.

*-----

`tsline aus usa, scheme(sj) name(level, replace)`

`tsline D.aus D.usa, scheme(sj) name(difference, replace)`

* Test for Unit Roots

```

* Experiment with noconst, trend, drift, and lag length
dfuller aus, regress lags(1)
dfuller usa, regress lags(3)
* Cointegrating regression
reg aus usa, noconst
predict ehat, res
tsline ehat, name(C1, replace)
* Engle-Granger Test for Cointegration
reg D.ehat L.ehat, noconst
dfuller ehat, lags(0) noconst
*-----
* VECM
*-----
regress D.aus L.ehat
regress D.usa L.ehat
drop ehat
*-----
2-masala.VAR amallarini bajarish.
*-----
use fred, clear
gen date = q(1960q1) + _n - 1
format %tq date
tsset date
*-----
* Plot the series to identify constants
* and trends.
*-----
tsline c y, legend(lab (1 "ln(RPCE)") lab(2 "ln(RPDI)")) ///
name(l1, replace) lpattern(solid dash)
tsline D.c D.y, legend(lab (1 "ln(RPCE)") lab(2 "ln(RPDI)")) ///
name(d1, replace) lpattern(solid dash)
* Stationarity Analysis
* Brute force, 1 equation at a time
qui reg L(0/1).D.c L.c

```

```

di "Lags = 1"
estat bgodfrey, lags(1 2 3)
qui reg L(0/2).D.c L.c
di "Lags = 2"
estat bgodfrey, lags(1 2 3)
qui reg L(0/3).D.c L.c
di "Lags = 3"
estat bgodfrey, lags(1 2 3)
dfuller c, lags(3)
* Use the loop to compute stats for y
forvalues p = 1/3 {
qui reg L(0/'p').D.y L.y
di "Lags =" `p'
estat bgodfrey, lags(1 2 3)
}
dfuller y, lags(0)
* Cointegration Test: Case 2
reg c y
predict ehat, res
reg D.ehat L.ehat D.L.ehat, noconst
di _b[L.ehat]/_se[L.ehat]
reg D.c D.L.c D.L.y
reg D.y D.L.c D.L.y
varbasic D.c D.y, lags(1/1) step(12) nograph
*-----
* Test residuals for autocorrelation
*-----
varlmar
* Try extending lags to 3 and repeat
quietly varbasic D.c D.y, lags(1/3) step(12)
varlmar
* There is evidence of autocorrelation so extend the lag to 3
* Selecting lags using model selection criteria
varsoc D.c D.y, maxlag(4)
* Impulse responses and variance decompositions

```

```

qui varbasic D.c D.y, lags(1/1) step(12)
irf table irf
irf table fevd
irf graph irf, name(g1, replace)
irf graph fevd, name(g2, replace)
* Combining irf and fevd in a single table
irf table irf fevd, title("Combined IRF/FEVD for C and Y")
log close

```

Mustaqil yechish uchun masalalar

1-Masala.

	Yalpi ichki mahsulot (y)	Tovar va xizmatlar eksporti (x1)
2000	3.256	0.864
2001	4.925	1.517
2002	7.450	2.353
2003	9.844	3.667
2004	12.261	4.930
2005	15.923	6.028
2006	21.125	7.820
2007	28.190	11.372
2008	38.970	15.144
2009	49.376	17.276
2010	62.388	20.668
2011	78.764	25.717
2012	97.929	25.717
2013	120.862	30.100
2014	145.846	31.414
2015	171.808	32.225
2016	199.993	36.077
2017	254.043	73.392

Yuqoridagi jadval ma'lumotidan foydalanib quyidagilarni toping.

- qisqacha xulosa statistikasi ma'lumotlarini;
- sCATTER diagrammasini yaratish;
- chiziqli regression model tuzing;
- hisoblangan yhat ni toping;
- bashoratli qiymatlarni hisoblang.

2-Masala. Yuqoridagi jadvaldan y va x_1 o'zgaruvchilari uchun quyidagilarni toping:

- kvadrati regression model tuzing;
- elastiklik qiymatini toping;
- T taqsimotning kritik qiymatlarini hisoblang;
- gipoteza testlarini amalga oshiring;
- p qiymatni toping.

3-Masala.

	Yalpi ichki mahsulot (y)	Asosiy kapitalga investitsiya (x_2)
2000	3.256	0.744
2001	4.925	1.321
2002	7.450	1.526
2003	9.844	1.978
2004	12.261	2.629
2005	15.923	3.165
2006	21.125	4.041
2007	28.190	5.903
2008	38.970	9.555
2009	49.376	12.531
2010	62.388	15.338
2011	78.764	17.953
2012	97.929	22.797
2013	120.862	28.694

2014	145.846	35.233
2015	171.808	41.670
2016	199.993	49.771
2017	254.043	68.424

Yuqoridagi jadval ma'lumotidan foydalanib, quyidagilarni toping.

- qisqacha xulosa statistikasi ma'lumotlarini;
- sCATTER diagrammasini yaratish;
- chiziqli regression model tuzing;
- hisoblangan yhat ni toping;
- bashoratli qiymatlarni hisoblang.

4-Masala. Yuqoridagi jadvaldan y va x2 o'zgaruvchilari uchun quyidagilarni toping:

- kvadrati regression model tuzing;
- elastiklik qiymatini toping;
- t taqsimotning kritik qiymatlarini hisoblang;
- gipoteza testlarini amalga oshiring;
- p qiymatni toping.

5-Masala.

	Yalpi ichki mahsulot (y)	Ishlab chiqarish (x3)
2000	3.256	0.462
2001	4.925	0.696
2002	7.450	1.079
2003	9.844	1.553
2004	12.261	2.147
2005	15.923	3.371
2006	21.125	4.597
2007	28.190	5.907
2008	38.970	9.148

2009	49.376	11.651
2010	62.388	15.115
2011	78.764	13.443
2012	97.929	16.148
2013	120.862	20.637
2014	145.846	25.363
2015	171.808	29.370
2016	199.993	33.956
2017	254.043	42.611

Yuqoridagi jadval ma'lumotidan foydalanib, quyidagilarni toping:

- qisqacha xulosa statistikasi ma'lumotlarini;
- sCATTER diagrammasini yaratish;
- chiziqli regression model tuzing;
- hisoblangan yhat ni toping;
- bashoratli qiymatlarni hisoblang.

6-Masala. Yuqoridagi jadvaldan y va x_3 o'zgaruvchilari uchun quyidagilarni toping:

- kvadrati regression model tuzing;
- elastiklik qiymatini toping;
- t taqsimotning kritik qiymatlarini hisoblang;
- gipoteza testlarini amalga oshiring;
- p qiymatni toping.

7-Masala.

	Yalpi ichki mahsulot (y)	Oziq-ovqat, ichimliklar va tamaki mahsulotlarini ishlab chiqarish(x_4)
2004	12.261	0.774
2005	15.923	0.909
2006	21.125	1.308
2007	28.190	1.715

2008	38.970	2.554
2009	49.376	3.235
2010	62.388	6.843
2011	78.764	8.981
2012	97.929	10.544
2013	120.862	13.700
2014	145.846	17.177
2015	171.808	21.889
2016	199.993	26.782
2017	254.043	28.332

Yuqoridagi jadval ma'lumotidan foydalanib quyidagilarni toping:

- qisqacha xulosa statistikasi ma'lumotlarini;
- SCATTER diagrammasini yaratish;
- chiziqli regression model tuzing;
- hisoblangan yhat ni toping;
- Bashoratli qiymatlarni hisoblang.

8-Masala. Yuqoridagi jadvaldan y va x_4 o'zgaruvchilari uchun quyidagilarni toping:

- kvadrati regression model tuzing;
- elastiklik qiymatini toping;
- T taqsimotning kritik qiymatlarini hisoblang;
- gipoteza testlarini amalga oshiring;
- p qiymatni toping.

FOYDALANILGAN ADABIYOTLAR RO'YXATI

1. Anderson, T.W. and H. Rubin (1949) "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 21, pp. 46-63
2. "Some Properties of a Modification of the Limited Information Estimator," *Econometrica*, 45, pp.939-953.
3. Principles of Econometrics, 4th Edition by R. Carter Hill, William E. Griffiths and Guay C. Lim (Wiley, 2011)
4. LEE C. ADKINS, R. CARTER HILL. Using Stata For Principles of Econometrics, Fourth Edition.2011. p-625
4. O'zbekiston Davlat statistika qo'mitasining 2010–2018-yillardagi "O'zbekiston Respublikasi yillik statistik to'plami" . <http://www.stat.uz>
2. Lee C. Adkins, R. Carter Hill. Using Stata for Principles of Econometrics, Fourth Edition. Wiley, 2011
3. Sh.I. Mustafakulov va boshqalar. Ekonometrika: O'quv qo'llanma. – Toshkent 2017, 155 b.
4. A.Ishnazarov, Sh.Nurullayeva, M.Mo'minova. "Ekonometrika asoslari" fani bo'yicha o'quv-uslubiy majmua. – Toshkent: Iqtisodiyot, 2017 yil.
5. Introduction to econometrics / James H. Stock, Mark W. Watson. 2nd ed., Boston: Pearson Addison Wesley, 2007.
6. Data analysis using regression and multilevel/hierarchical models / Andrew Gelman, Jennifer Hill. Cambridge ; New York : Cambridge University Press, 2007.
7. Gujarati D.N. "Basic Econometrics". McGraw-Hill, 4th edition, 2003 (Gu), Inc.p. 90.
8. William H. Greene. Econometric analysis, seventh edition. USA, 2012.
9. Stata basics (Stata 13 or 14), Ursina Kuhn and Oliver Lipps, FORS, Lausanne. Version August 2016
10. Sh.I. Mustafakulov. Investitsion jozibadorlikning nazariy, metodologik va amaliy talqini [Matn] /–T.: «Ma'naviyat», 2021, 352b.
11. Sh.I. Mustafakulov va boshqalar. Iqtisodiy atamalarning izohli lug'ati [Matn]: lug'at. – T.: Innovatsion rivojlantirish nashriyot-matbaa uyi, 2019,-489b.

Statistik taqsimot jadvallari

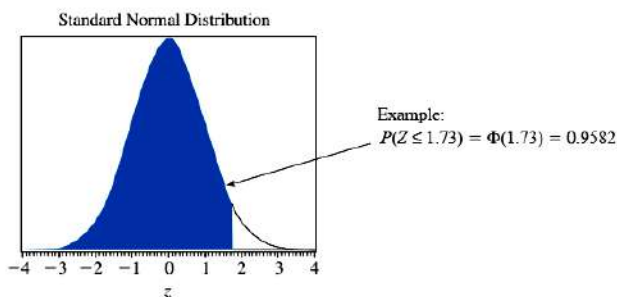


Table 1 Cumulative Probabilities for the Standard Normal Distribution
 $\Phi(z) = P(Z \leq z)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Source: This table was generated using the SAS® function PROBNORM.

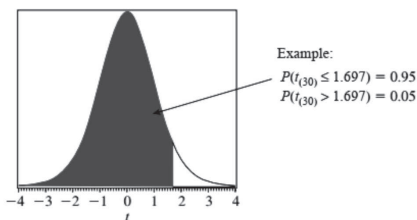


Table 2 Percentiles of the t -distribution

df	$t_{(0.90,df)}$	$t_{(0.95,df)}$	$t_{(0.975,df)}$	$t_{(0.99,df)}$	$t_{(0.995,df)}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
31	1.309	1.696	2.040	2.453	2.744
32	1.309	1.694	2.037	2.449	2.738
33	1.308	1.692	2.035	2.445	2.733
34	1.307	1.691	2.032	2.441	2.728
35	1.306	1.690	2.030	2.438	2.724
36	1.306	1.688	2.028	2.434	2.719
37	1.305	1.687	2.026	2.431	2.715
38	1.304	1.686	2.024	2.429	2.712
39	1.304	1.685	2.023	2.426	2.708
40	1.303	1.684	2.021	2.423	2.704
50	1.299	1.676	2.009	2.403	2.678
∞	1.282	1.645	1.960	2.326	2.576

Source: This table was generated using the SAS® function TINV.

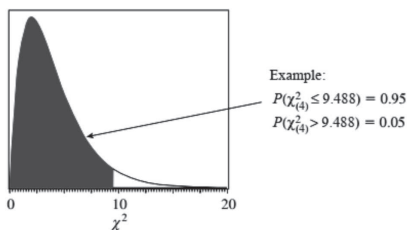


Table 3 Percentiles of the Chi-square Distribution

df	$\chi^2_{(0.90,df)}$	$\chi^2_{(0.95,df)}$	$\chi^2_{(0.975,df)}$	$\chi^2_{(0.99,df)}$	$\chi^2_{(0.995,df)}$
1	2.706	3.841	5.024	6.635	7.879
2	4.605	5.991	7.378	9.210	10.597
3	6.251	7.815	9.348	11.345	12.838
4	7.779	9.488	11.143	13.277	14.860
5	9.236	11.070	12.833	15.086	16.750
6	10.645	12.592	14.449	16.812	18.548
7	12.017	14.067	16.013	18.475	20.278
8	13.362	15.507	17.535	20.090	21.955
9	14.684	16.919	19.023	21.666	23.589
10	15.987	18.307	20.483	23.209	25.188
11	17.275	19.675	21.920	24.725	26.757
12	18.549	21.026	23.337	26.217	28.300
13	19.812	22.362	24.736	27.688	29.819
14	21.064	23.685	26.119	29.141	31.319
15	22.307	24.996	27.488	30.578	32.801
16	23.542	26.296	28.845	32.000	34.267
17	24.769	27.587	30.191	33.409	35.718
18	25.989	28.869	31.526	34.805	37.156
19	27.204	30.144	32.852	36.191	38.582
20	28.412	31.410	34.170	37.566	39.997
21	29.615	32.671	35.479	38.932	41.401
22	30.813	33.924	36.781	40.289	42.796
23	32.007	35.172	38.076	41.638	44.181
24	33.196	36.415	39.364	42.980	45.559
25	34.382	37.652	40.646	44.314	46.928
26	35.563	38.885	41.923	45.642	48.290
27	36.741	40.113	43.195	46.963	49.645
28	37.916	41.337	44.461	48.278	50.993
29	39.087	42.557	45.722	49.588	52.336
30	40.256	43.773	46.979	50.892	53.672
35	46.059	49.802	53.203	57.342	60.275
40	51.805	55.758	59.342	63.691	66.766
50	63.167	67.505	71.420	76.154	79.490
60	74.397	79.082	83.298	88.379	91.952
70	85.527	90.531	95.023	100.425	104.215
80	96.578	101.879	106.629	112.329	116.321
90	107.565	113.145	118.136	124.116	128.299
100	118.498	124.342	129.561	135.807	140.169
110	129.385	135.480	140.917	147.414	151.948
120	140.233	146.567	152.211	158.950	163.648

Source: This table was generated using the SAS® function CINV.

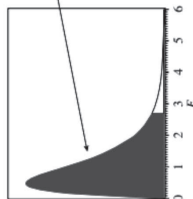


Table 4 95th Percentile for the F-distribution

v_2/v_1	1	2	3	4	5	6	7	8	9	10	12	15	20	30	60	∞
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.95	248.01	250.10	252.20	254.31
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.46	19.48	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.62	8.57	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.75	5.69	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.50	4.43	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.81	3.74	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.38	3.30	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.08	3.01	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.86	2.79	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.70	2.62	2.54
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.25	2.16	2.07
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.04	1.95	1.84
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.92	1.82	1.71
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.84	1.74	1.62
35	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11	2.04	1.96	1.88	1.79	1.68	1.56
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.74	1.64	1.51
45	4.06	3.20	2.81	2.58	2.42	2.31	2.22	2.15	2.10	2.05	1.97	1.89	1.81	1.71	1.60	1.47
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.95	1.87	1.78	1.69	1.58	1.44
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.65	1.53	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.55	1.43	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.46	1.32	1.00

Source: This table was generated using the SAS® function FINV.

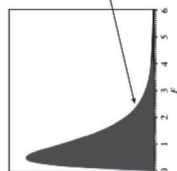


Table 5 99th Percentile for the F -distribution

v_2/v_1	1	2	3	4	5	6	7	8	9	10	12	15	20	30	60	∞
1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47	6055.85	6106.32	6157.28	6208.73	6260.65	6313.03	6365.87
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.47	99.48	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.50	26.32	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.84	13.65	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.38	9.20	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.23	7.06	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	5.99	5.82	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.20	5.03	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.65	4.48	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.25	4.08	3.91
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.21	3.05	2.87
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.78	2.61	2.42
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.54	2.36	2.17
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.39	2.21	2.01
35	7.42	5.27	4.40	3.91	3.59	3.37	3.20	3.07	2.96	2.88	2.74	2.60	2.44	2.28	2.10	1.89
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.20	2.02	1.80
45	7.23	5.11	4.25	3.77	3.45	3.23	3.07	2.94	2.83	2.74	2.61	2.46	2.31	2.14	1.96	1.74
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70	2.56	2.42	2.27	2.10	1.91	1.68
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.03	1.84	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.86	1.66	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.70	1.47	1.00

Source: This table was generated using the SAS ® function FINV.

The Rules of Summation

$$\sum_{i=1}^n x_i = x_1 + x_2 + \cdots + x_n$$

$$\sum_{i=1}^n a = na$$

$$\sum_{i=1}^n ax_i = a \sum_{i=1}^n x_i$$

$$\sum_{i=1}^n (x_i + y_i) = \sum_{i=1}^n x_i + \sum_{i=1}^n y_i$$

$$\sum_{i=1}^n (ax_i + by_i) = a \sum_{i=1}^n x_i + b \sum_{i=1}^n y_i$$

$$\sum_{i=1}^n (a + bx_i) = na + b \sum_{i=1}^n x_i$$

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^3 f(x_i, y_j) &= \sum_{i=1}^n [f(x_i, y_1) + f(x_i, y_2) + f(x_i, y_3)] \\ &= f(x_1, y_1) + f(x_1, y_2) + f(x_1, y_3) \\ &\quad + f(x_2, y_1) + f(x_2, y_2) + f(x_2, y_3) \end{aligned}$$

Expected Values & Variances

$$E(X) = x_1 f(x_1) + x_2 f(x_2) + \cdots + x_n f(x_n)$$

$$= \sum_{i=1}^n x_i f(x_i) = \sum_x x f(x)$$

$$E[g(X)] = \sum_x g(x) f(x)$$

$$\begin{aligned} E[g_1(X) + g_2(X)] &= \sum_x [g_1(x) + g_2(x)] f(x) \\ &= \sum_x g_1(x) f(x) + \sum_x g_2(x) f(x) \\ &= E[g_1(X)] + E[g_2(X)] \end{aligned}$$

$$E(c) = c$$

$$E(cX) = cE(X)$$

$$E(a + cX) = a + cE(X)$$

$$\text{var}(X) = \sigma^2 = E[X - E(X)]^2 = E(X^2) - [E(X)]^2$$

$$\text{var}(a + cX) = E[(a + cX) - E(a + cX)]^2 = c^2 \text{var}(X)$$

Marginal and Conditional Distributions

$$f(x) = \sum_y f(x, y) \quad \text{for each value } X \text{ can take}$$

$$f(y) = \sum_x f(x, y) \quad \text{for each value } Y \text{ can take}$$

$$f(x|y) = P[X = x|Y = y] = \frac{f(x, y)}{f(y)}$$

If X and Y are independent random variables, then $f(x, y) = f(x)f(y)$ for each and every pair of values x and y . The converse is also true.

If X and Y are independent random variables, then the conditional probability density function of X given that

$$Y = y \text{ is } f(x|y) = \frac{f(x, y)}{f(y)} = \frac{f(x)f(y)}{f(y)} = f(x)$$

for each and every pair of values x and y . The converse is also true.

Expectations, Variances & Covariances

$$\text{cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$$

$$= \sum_x \sum_y [x - E(X)][y - E(Y)]f(x, y)$$

$$\rho = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$$

$$E(c_1X + c_2Y) = c_1E(X) + c_2E(Y)$$

$$E(X + Y) = E(X) + E(Y)$$

$$\begin{aligned} \text{var}(aX + bY + cZ) &= a^2 \text{var}(X) + b^2 \text{var}(Y) + c^2 \text{var}(Z) \\ &\quad + 2abc \text{cov}(X, Y) + 2ac \text{cov}(X, Z) + 2bc \text{cov}(Y, Z) \end{aligned}$$

If X , Y , and Z are independent, or uncorrelated, random variables, then the covariance terms are zero and:

$$\begin{aligned} \text{var}(aX + bY + cZ) &= a^2 \text{var}(X) \\ &\quad + b^2 \text{var}(Y) + c^2 \text{var}(Z) \end{aligned}$$

Normal Probabilities

$$\text{If } X \sim N(\mu, \sigma^2), \text{ then } Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

If $X \sim N(\mu, \sigma^2)$ and a is a constant, then

$$P(X \geq a) = P\left(Z \geq \frac{a - \mu}{\sigma}\right)$$

If $X \sim N(\mu, \sigma^2)$ and a and b are constants, then

$$P(a \leq X \leq b) = P\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right)$$

Assumptions of the Simple Linear Regression Model

SR1 The value of y , for each value of x , is $y = \beta_1 + \beta_2 x + e$

SR2 The average value of the random error e is $E(e) = 0$ since we assume that $E(y) = \beta_1 + \beta_2 x$

SR3 The variance of the random error e is $\text{var}(e) = \sigma^2 = \text{var}(y)$

SR4 The covariance between any pair of random errors, e_i and e_j is $\text{cov}(e_i, e_j) = \text{cov}(y_i, y_j) = 0$

SR5 The variable x is not random and must take at least two different values.

SR6 (optional) The values of e are normally distributed about their mean $e \sim N(0, \sigma^2)$

Least Squares Estimation

If b_1 and b_2 are the least squares estimates, then

$$\hat{y}_i = b_1 + b_2 x_i$$

$$\hat{e}_i = y_i - \hat{y}_i = y_i - b_1 - b_2 x_i$$

The Normal Equations

$$nb_1 + \sum x_i b_2 = \sum y_i$$

$$\sum x_i b_1 + \sum x_i^2 b_2 = \sum x_i y_i$$

Least Squares Estimators

$$b_2 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$b_1 = \bar{y} - b_2 \bar{x}$$

Elasticity

$$\eta = \frac{\text{percentage change in } y}{\text{percentage change in } x} = \frac{\Delta y/y}{\Delta x/x} = \frac{\Delta y}{\Delta x} \cdot \frac{x}{y}$$

$$\eta = \frac{\Delta E(y)/E(y)}{\Delta x/x} = \frac{\Delta E(y)}{\Delta x} \cdot \frac{x}{E(y)} = \beta_2 \cdot \frac{x}{E(y)}$$

Least Squares Expressions Useful for Theory

$$b_2 = \beta_2 + \sum w_i e_i$$

$$w_i = \frac{x_i - \bar{x}}{\sum (x_i - \bar{x})^2}$$

$$\sum w_i = 0, \quad \sum w_i x_i = 1, \quad \sum w_i^2 = 1/\sum (x_i - \bar{x})^2$$

Properties of the Least Squares Estimators

$$\text{var}(b_1) = \sigma^2 \left[\frac{\sum x_i^2}{N \sum (x_i - \bar{x})^2} \right], \quad \text{var}(b_2) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$$

$$\text{cov}(b_1, b_2) = \sigma^2 \left[\frac{-\bar{x}}{\sum (x_i - \bar{x})^2} \right]$$

Gauss-Markov Theorem: Under the assumptions SR1–SR5 of the linear regression model the estimators b_1 and b_2 have the *smallest variance of all linear and unbiased estimators* of β_1 and β_2 . They are the **Best Linear Unbiased Estimators (BLUE)** of β_1 and β_2 .

If we make the normality assumption, assumption SR6, about the error term, then the least squares estimators are normally distributed.

$$b_1 \sim N\left(\beta_1, \frac{\sigma^2 \sum x_i^2}{N \sum (x_i - \bar{x})^2}\right), \quad b_2 \sim N\left(\beta_2, \frac{\sigma^2}{\sum (x_i - \bar{x})^2}\right)$$

Estimated Error Variance

$$\hat{\sigma}^2 = \frac{\sum \hat{e}_i^2}{N - 2}$$

Estimator Standard Errors

$$\text{se}(b_1) = \sqrt{\text{var}(b_1)}, \quad \text{se}(b_2) = \sqrt{\text{var}(b_2)}$$

t -distribution

If assumptions SR1–SR6 of the simple linear regression model hold, then

$$t = \frac{b_k - \beta_k}{\text{se}(b_k)} \sim t_{(N-2)}, \quad k = 1, 2$$

Interval Estimates

$$P[b_2 - t_{\text{se}}(b_2) \leq \beta_2 \leq b_2 + t_{\text{se}}(b_2)] = 1 - \alpha$$

Hypothesis Testing

Components of Hypothesis Tests

1. A *null hypothesis*, H_0
2. An *alternative hypothesis*, H_1
3. A *test statistic*
4. A *rejection region*
5. A *conclusion*

If the null hypothesis $H_0: \beta_2 = c$ is *true*, then

$$t = \frac{b_2 - c}{\text{se}(b_2)} \sim t_{(N-2)}$$

Rejection rule for a two-tail test: If the value of the test statistic falls in the rejection region, either tail of the t -distribution, then we reject the null hypothesis and accept the alternative.

Type I error: The null hypothesis is *true* and we decide to *reject* it.

Type II error: The null hypothesis is *false* and we decide *not* to reject it.

p -value rejection rule: When the p -value of a hypothesis test is *smaller* than the chosen value of α , then the test procedure leads to *rejection* of the null hypothesis.

Prediction

$$y_0 = \beta_1 + \beta_2 x_0 + e_0, \quad \hat{y}_0 = b_1 + b_2 x_0, \quad f = \hat{y}_0 - y_0$$

$$\widehat{\text{var}}(f) = \hat{\sigma}^2 \left[1 + \frac{1}{N} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right], \quad \text{se}(f) = \sqrt{\widehat{\text{var}}(f)}$$

A $(1 - \alpha) \times 100\%$ confidence interval, or prediction interval, for y_0

$$\hat{y}_0 \pm t_c \text{se}(f)$$

Goodness of Fit

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum \hat{e}_i^2$$

$$SST = SSR + SSE$$

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = (\text{corr}(y, \hat{y}))^2$$

Log-Linear Model

$$\ln(y) = \beta_1 + \beta_2 x + e, \quad \widehat{\ln(y)} = b_1 + b_2 x$$

$100 \times \beta_2 \approx \%$ change in y given a one-unit change in x .

$$\hat{y}_n = \exp(b_1 + b_2 x)$$

$$\hat{y}_c = \exp(b_1 + b_2 x) \exp(\hat{\sigma}^2/2)$$

Prediction interval:

$$\exp[\widehat{\ln(y)} - t_c \text{se}(f)], \quad \exp[\widehat{\ln(y)} + t_c \text{se}(f)]$$

Generalized goodness-of-fit measure $R_k^2 = (\text{corr}(y, \hat{y}_n))^2$

Assumptions of the Multiple Regression Model

$$\text{MR1} \quad y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + e_i$$

$$\text{MR2} \quad E(y_i) = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} \Leftrightarrow E(e_i) = 0.$$

$$\text{MR3} \quad \text{var}(y_i) = \text{var}(e_i) = \sigma^2$$

$$\text{MR4} \quad \text{cov}(y_i, y_j) = \text{cov}(e_i, e_j) = 0$$

MR5 The values of x_{ik} are not random and are not exact linear functions of the other explanatory variables.

$$\text{MR6} \quad y_i \sim N[(\beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK}), \sigma^2] \\ \Leftrightarrow e_i \sim N(0, \sigma^2)$$

Least Squares Estimates in MR Model

Least squares estimates $b_1, b_2, \dots, b_K$ minimize $S(\beta_1, \beta_2, \dots, \beta_K) = \sum (y_i - \beta_1 - \beta_2 x_{i2} - \cdots - \beta_K x_{iK})^2$

Estimated Error Variance and Estimator Standard Errors

$$\hat{\sigma}^2 = \frac{\sum \hat{e}_i^2}{N - K} \quad \text{se}(b_k) = \sqrt{\text{var}(b_k)}$$

Hypothesis Tests and Interval Estimates for Single Parameters

Use t -distribution $t = \frac{b_k - \beta_k}{\text{se}(b_k)} \sim t_{(N-K)}$

t -test for More than One Parameter

$$H_0: \beta_2 + c\beta_3 = a$$

When H_0 is true $t = \frac{b_2 + cb_3 - a}{\text{se}(b_2 + cb_3)} \sim t_{(N-K)}$

$$\text{se}(b_2 + cb_3) = \sqrt{\text{var}(b_2) + c^2 \text{var}(b_3) + 2c \times \text{cov}(b_2, b_3)}$$

Joint F -tests

To test J joint hypotheses,

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)}$$

To test the overall significance of the model the null and alternative hypotheses and F statistic are

$$H_0: \beta_2 = 0, \beta_3 = 0, \dots, \beta_K = 0$$

$$H_1: \text{at least one of the } \beta_k \text{ is nonzero}$$

$$F = \frac{(SST - SSE)/(K-1)}{SSE/(N-K)}$$

RESET: A Specification Test

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i \quad \hat{y}_i = b_1 + b_2 x_{i2} + b_3 x_{i3}$$

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \gamma_1 \hat{y}_i^2 + e_i, \quad H_0: \gamma_1 = 0$$

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \gamma_1 \hat{y}_i^2 + \gamma_2 \hat{y}_i^3 + e_i, \quad H_0: \gamma_1 = \gamma_2 = 0$$

Model Selection

$$\text{AIC} = \ln(SSE/N) + 2K/N$$

$$\text{SC} = \ln(SSE/N) + K \ln(N)/N$$

Collinearity and Omitted Variables

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i$$

$$\text{var}(b_2) = \frac{\sigma^2}{(1 - r_{23}^2) \sum (x_{i2} - \bar{x}_2)^2}$$

$$\text{When } x_3 \text{ is omitted, } \text{bias}(b_2^*) = E(b_2^*) - \beta_2 = \beta_3 \frac{\text{cov}(x_2, x_3)}{\text{var}(x_2)}$$

Heteroskedasticity

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

General variance function

$$\sigma_i^2 = \exp(\alpha_1 + \alpha_2 z_{i2} + \dots + \alpha_5 z_{i5})$$

Breusch-Pagan and White Tests for $H_0: \alpha_2 = \alpha_3 = \dots = \alpha_5 = 0$

$$\text{When } H_0 \text{ is true } \chi^2 = N \times R^2 \sim \chi_{(S-1)}^2$$

Goldfeld-Quandt test for $H_0: \sigma_M^2 = \sigma_R^2$ versus $H_1: \sigma_M^2 \neq \sigma_R^2$

$$\text{When } H_0 \text{ is true } F = \hat{\sigma}_M^2 / \hat{\sigma}_R^2 \sim F_{(N_M - K_M, N_R - K_R)}$$

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

$$y_i / \sqrt{x_i} = \beta_1 (1/\sqrt{x_i}) + \beta_2 (x_i/\sqrt{x_i}) + e_i / \sqrt{x_i}$$

Estimating the variance function

$$\ln(\hat{e}_i^2) = \ln(\sigma_i^2) + v_i = \alpha_1 + \alpha_2 z_{i2} + \dots + \alpha_5 z_{i5} + v_i$$

Grouped data

$$\text{var}(e_i) = \sigma_i^2 = \begin{cases} \sigma_M^2 & i = 1, 2, \dots, N_M \\ \sigma_R^2 & i = 1, 2, \dots, N_R \end{cases}$$

Transformed model for feasible generalized least squares

$$y_i / \sqrt{\hat{\sigma}_i} = \beta_1 (1/\sqrt{\hat{\sigma}_i}) + \beta_2 (x_i/\sqrt{\hat{\sigma}_i}) + e_i / \sqrt{\hat{\sigma}_i}$$

Regression with Stationary Time Series Variables

Finite distributed lag model

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \dots + \beta_q x_{t-q} + v_t$$

Correlogram

$$r_k = \sum (y_t - \bar{y})(y_{t-k} - \bar{y}) / \sum (y_t - \bar{y})^2$$

$$\text{For } H_0: \rho_k = 0, \quad z = \sqrt{T} r_k \sim N(0, 1)$$

LM test

$$y_t = \beta_1 + \beta_2 x_t + \rho \hat{e}_{t-1} + \hat{v}_t \quad \text{Test } H_0: \rho = 0 \text{ with } t\text{-test}$$

$$\hat{e}_t = \gamma_1 + \gamma_2 x_t + \rho \hat{e}_{t-1} + \hat{v}_t \quad \text{Test using } LM = T \times R^2$$

$$\text{AR}(1) \text{ error } y_t = \beta_1 + \beta_2 x_t + e_t \quad e_t = \rho e_{t-1} + v_t$$

Nonlinear least squares estimation

$$y_t = \beta_1 (1 - \rho) + \beta_2 x_t + \rho y_{t-1} - \beta_2 \rho x_{t-1} + v_t$$

ARDL(p, q) model

$$y_t = \delta + \delta_0 x_t + \delta_1 x_{t-1} + \dots + \delta_q x_{t-q} + \theta_1 y_{t-1} + \dots + \theta_p y_{t-p} + v_t$$

AR(p) forecasting model

$$y_t = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_p y_{t-p} + v_t$$

Exponential smoothing $\hat{y}_t = \alpha y_{t-1} + (1 - \alpha) \hat{y}_{t-1}$

Multiplier analysis

$$\delta_0 + \delta_1 L + \delta_2 L^2 + \dots + \delta_q L^q = (1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_p L^p) \times (\beta_0 + \beta_1 L + \beta_2 L^2 + \dots)$$

Unit Roots and Cointegration

Unit Root Test for Stationarity: Null hypothesis:

$$H_0: \gamma = 0$$

Dickey-Fuller Test 1 (no constant and no trend):

$$\Delta y_t = \gamma y_{t-1} + v_t$$

Dickey-Fuller Test 2 (with constant but no trend):

$$\Delta y_t = \alpha + \gamma y_{t-1} + v_t$$

Dickey-Fuller Test 3 (with constant and with trend):

$$\Delta y_t = \alpha + \gamma y_{t-1} + \lambda t + v_t$$

Augmented Dickey-Fuller Tests:

$$\Delta y_t = \alpha + \gamma y_{t-1} + \sum_{j=1}^m a_j \Delta y_{t-j} + v_t$$

Test for cointegration

$$\Delta \hat{e}_t = \gamma \hat{e}_{t-1} + v_t$$

Random walk: $y_t = y_{t-1} + v_t$

Random walk with drift: $y_t = \alpha + y_{t-1} + v_t$

Random walk model with drift and time trend:

$$y_t = \alpha + \delta t + y_{t-1} + v_t$$

Panel Data

Pooled least squares regression

$$y_{it} = \beta_1 + \beta_2 x_{it} + \beta_3 x_{3it} + e_{it}$$

Cluster robust standard errors $\text{cov}(e_{it}, e_{it}) = \psi_{it}$

Fixed effects model

$$y_{it} = \beta_{1i} + \beta_2 x_{it} + \beta_3 x_{3it} + e_{it} \quad \beta_{1i} \text{ not random}$$

$$y_{it} - \bar{y}_i = \beta_2 (x_{it} - \bar{x}_{2i}) + \beta_3 (x_{3it} - \bar{x}_{3i}) + (e_{it} - \bar{e}_i)$$

Random effects model

$$y_{it} = \beta_{1i} + \beta_2 x_{it} + \beta_3 x_{3it} + e_{it} \quad \beta_{1i} = \bar{\beta}_1 + u_i \text{ random}$$

$$y_{it} - \alpha \bar{y}_i = \bar{\beta}_1 (1 - \alpha) + \beta_2 (x_{2it} - \alpha \bar{x}_{2i}) + \beta_3 (x_{3it} - \alpha \bar{x}_{3i}) + v_{it}^*$$

$$\alpha = 1 - \sigma_e / \sqrt{T \sigma_u^2 + \sigma_e^2}$$

Hausman test

$$t = (b_{FEK} - b_{REK}) / \left[\frac{\text{var}(b_{FEK}) - \text{var}(b_{REK})}{2} \right]^{1/2}$$

MUNDARIJA

MUQADDIMA	4
I BOB. AMALIY EKONOMETRIKA VA STATA	
DASTURI HAQIDA QISQACHA MA'LUMOT	6
1.1. Amaliy ekonometrika fanining predmeti	6
1.2. Amaliy ekonometrika fanining ketma-ket vazifalari	8
1.3. Matematik modellarni ishlab chiqish va masalalar yechish ketma-ketligi	14
1.4. STATA dasturini ishga tushirish va o'chirish	18
1.5. STATA dasturida ma'lumotlarni kiritish va yuklash	21
1.6. Ma'lumotlarni tasvirlash va tasviriy statistikasini olish	25
1.7. Natijalarni saqlash usuli	28
II-BOB. REGRESSION MODELLAR TUZISH STATA DASTURIDA	32
2.1. Klassik chiziqli modelni yaratish uchun ma'lumotlarini yuklash	32
2.1.1. Kundalik log faylni ishga tushirish (log file) va f oydalanish (use)	33
2.1.2. Ma'lumotlarni tasvirlash (describe) va ro'yxatlash (list).....	34
2.2. Ma'lumotlarning tahliliy statistikasi (summarize)	36
2.3. Ma'lumotlar asosida SCATTER diagramsini yaratish usullari.	38
2.4. STATA dasturida regression model yaratish usuli (regress) ..	42
2.4.1. Hisoblangan qiymat ($\hat{Y}$ predict) va qoldiqlar (residuals)....	44
2.4.2. Elastiklikni hisoblash (margins)	47
2.4.3. Hisoblangan regression model chizig'ini grafikda joylashtirish (plot)	50
2.4.4. Variatsiya (var) va kovarvatsiyani (cov) hisoblash	53
2.5. Regression modelning prognoz qiymatlarini olishda stata dasturidan foydalanish (edit & set obs)	55
2.6. Chiziqsiz regression modellarni yaratish usullari	58
2.6.1. Kvadratik model (generate sqft2)	58
2.6.2. Chiziqli logarifmik model (log-lin)	64
2.7. Regression jarayonda ko'rsatkich o'zgaruvchilardan foydalanish (bysort & hist)	69
III-BOB. INTERVALLI BAHOLASH VA GIPOTEZA TESTLARI	80
3.1. T-taqsimotning kritik qiymatlarini hisoblashda stata dasturining imkoniyati (invttail)	80
3.2. Gipoteza testlarini hisoblash va tahlil qilish	83
3.2.1. Iqtisodiy gipotezaning right-tail testi (scalar tstat)	85

3.2.2. Iqtisodiy gipotezaning left-tail testi (lincom).....	86
3.2.3. Iqtisodiy gipotezani ikki tomonlama tekshirish.....	87
3.3. p-qiymatlar (p-values) ni hisoblash usullari (ttail).....	88
3.3.1. p-qiymatning o'ng tomon qiymatni tekshirish	88
3.3.2. p-qiymatning chap tomon qiymatni tekshirish.....	89
3.3.3. p-qiymatning ikki tomonli testini tekshirish	90
3.3.4. Stata tomonidan p-qiymatlarni topish (lincom).....	91
3.3.5. Chiziqli funksiyalarning parametrlarini tekshirish va baholash (estat vce)	91
IV-BOB. CHIZIQSIZ REGRESSIYALAR VA MODELLASHTIRISH MASALALARI	96
4.1. Eng kichik kvadratlarni hisoblash	96
4.1.1. Ma'lumotlarni tahrirlash	96
4.1.2. Regressiyani baholash va keyingi baholash natijalarini olish	96
4.1.3. $\hat{Y}$ hisoblangan qiymat oralig'ini yaratish	98
4.2. Moslikni o'lchash.....	99
4.2.1. Korrelyatsiya va determinatsiya R^2	100
4.3. Ma'lumotlarni kattalashtirish va o'zgartirish ta'sirlari	101
4.3.1. Liner-log funksiyasining shakli	102
4.3.2. O'rnatilgan chiziqli log modelini chizish	104
4.3.3. Grafiklarni tahrirlash	105
4.4. Qoldiqlarni tahlil olish	107
4.4.1. Jarque-Bera testi	108
4.4.2. x^2 -kvadrat taqsimotining kritik qiymatlari	110
4.4.3. x^2 -kvadrat taqsimotining p-qiymatlari	110
4.5. Polinomial modellar.....	111
4.5.1. Chiziqli munosabatni baholash va tekshirish	112
4.5.2. Kubik tenglamani baholash va tekshirish	115
4.5.3. Hosildorlikning log-chiziqli o'sish modelini baholash	117
4.6. Log-liner oylik maosh tenglamasini hisoblash	118
4.6.1. Log-liner modeli.....	119
4.6.2. Hisoblangan ish haqlarini hisoblash.....	123
4.6.3. Ish haqi grafiklarini tuzish	123
4.6.4. Determinatsiya koeffisienti R^2	124
4.6.5. Log-liner modeldagi bashorat oraliqlari	125
4.7. Log-log modeli	126
V-BOB. KO'P OMILLI REGRESSION MODELLAR.....	137
5.1. Big andy's burger barn ma'lumotlari asosida model tuzish	137

5.2. Eng kichik kvadratlar bashorati	140
5.3. Tasviriy statistika ma'lumotlarini olish aniqligi	141
5.4. Ishonch intervallari	142
5.4.1. Parametrlarning chiziqli birikmasi uchun ishonch oralig'i ...	144
5.5. Gipoteza testlari	146
5.5.1. Ikki tomonlama testlar	146
5.5.2. Bir tomonlama testlar	147
5.5.3. Chiziqli parametрни tekshirish	148
5.6. Polinomial tenglamalar	149
5.6.1. Optimal reklama: parametrlarning chiziqli bo'lmagan kombinatsiyasi	151
5.6.2. O'zaro ta'sirlar uchun omil o'zgaruvchilardan foydalanish	152
5.7. O'zgaruvchilar orasidagi o'zaro ta'sirlar	155
5.8. Modellar orasidagi moslikni baholash	158
VI-BOB. KO'P OMILLI REGRESSIYA MODELLARINI BAHOLASH	164
6.1. F-test ni hisoblash	164
6.1.1. Modelning ahamiyatini tekshirish	169
6.1.2. t- va F- testlari o'rtasidagi bog'liqlik	171
6.1.3. Qo'shimcha umumiy F-testlari	172
6.2. Namuna bo'lmagan ma'lumotlar orqali model tuzish	176
6.3. Model xususiyatlari	178
6.3.1. Yo'qolgan o'zgaruvchilar	179
6.3.2. Ahamiyatsiz o'zgaruvchilar	180
6.3.3. Model tanlash	181
6.4. Ma'lumotlar yetishmaslik, kolenearlilik va ahamiyatsizlik ...	190
VII-BOB. O'ZGARUVCHILAR INDIKATORIDAN FOYDALANISH	201
7.1. O'zgaruvchilar indikatori	201
7.1.1. O'zgaruvchilar indikatorlarini yaratish	202
7.1.2. O'zgaruvchilar indikatorining regressiyasini baholash	203
7.1.3. O'zgaruvchilar indikatorining ahamiyatlilik testi	204
7.1.4. Qo'shimcha hisob-kitoblar	205
7.1.5. O'rtacha marjinal effektni hisoblash	205
7.2. O'zgaruvchilar indikatorini qo'llash	207
7.2.1. Sifat omillarning o'zaro ta'siri	208
7.2.2. Mintaqaviy o'zgaruvchilarni qo'shish	210
7.2.3. Ikki regressiya uchun tenglik testini o'tkazish	212
7.2.4. Alohida regressiyalarni baholash	213
7.2.5. Log-liner modellariga o'zgaruvchilar indikatorining qo'llanilishi	215

7.3. Chiziqli ehtimollik model	217
VIII-BOB. GETEROSKIDASTLIK	224
8.1. Geteroskidastlik ahamiyati	224
8.2. Geteroskedastiklikni aniqlash	226
8.2.1. Qoldiq maydonlari	226
8.2.2. Lagrange multiplikator testlari	228
8.2.3. Goldfeld-Quandt testi	232
8.3. Geteroskedastik-ishlab chiqish standart xatolar	235
8.4. Umumiyashtirilgan eng kichik kvadratlar usuli	238
8.4.1. Guruhlangan ma'lumotlardan foydalangan holda GLS ...	240
8.4.2. GLS ning - umumiy holati	214
8.5. Chiziqli ehtimollik modeldagi geteroskedastiklik	243
IX-BOB. DINAMIK QATORLAR REGRESSIYASI	254
9.1. Asosiy tushunchalar	254
9.1.1. Stata-da vaqtli qatorlar grafisini ifodalash	255
9.1.2. Vaqtli qatorlar grafiklari	259
9.1.3. Stataning lagli va farqli o'zgaruvchilari	260
9.2. Cheklangan taqsimlangan kechikishlar (laglar)	262
9.3. Ketma-ket korrelyatsiya	264
9.4. Ketma-ket korrelyatsiya uchun boshqa testlar	270
9.5. Ketma-ket bog'liq xatolar bilan baholash	273
9.5.1. Eng kichik kvadratlar va HAC standart xatolar	273
9.5.2. Nochiziqli eng kichik kvadratlar	276
9.5.3. Odatda ko'proq ishlatiladigan model	278
9.6. Avtoregressiv taqsimlangan kechikish(lagli) modellari	280
9.6.1. Fillips egri chizig'i	282
9.6.2. Okun qonuni	285
9.6.3. Avtoregressiv modellar	288
9.7. Prognozlash	290
9.7.1. AR modeli bilan prognoz qilish	291
9.7.2. Eksponensial tekislash	293
9.8. Multiplikator tahlili	298
9.9. Qo'shimcha testlar.....	302
9.9.1. Durbin-Watson testi	302
9.9.2. Prais-Winsten FGLS testi.....	302
X-BOB. TASODIFIY REGRESSORLAR VA MOMENTGA	
ASOSLANGAN BAHOLASH	314
10.1. Ish haqi tenglamasining eng kichik kvadratlarini baholash	314
10.2. Ikki bosqichli eng kichik kvadratlar	316

10.3. Qo'shimcha o'zgaruvchilar bilan IV ni baholash	321
10.3.1. Qisman korrelyatsiyalarni tasvirlash	324
10.4. Endogenlik bo'yicha HAUSMAN testi	327
10.5. Qo'shimcha o'zgaruvchilarning haqiqiylikini tekshirish ...	331
10.6. Kuchsiz o'zgaruvchilar uchun sinov	332
10.7. Cragg-Donald F-statistikasini hisoblash	338
10.8. Simulyatsiya tajribasi	340
XI-BOB. SINXRON TENGLAMALI MODELLAR	352
11.1. TRUFFLE kompaniyasining taklif va talab modeli	352
11.2. Qisqartirilgan shakldagi tenglamalarni baholash	353
11.3. 2SLS truffle talabini baholash	354
11.4. 2SLS orqali truffle kompaniyasining taklif modelini baholash	360
11.5. Baliq mahsulotiga bo'lgan taklif va talab modellari	361
11.6. Baliq narxi va miqdori uchun qisqartirilgan shakllar	363
11.7. 2SLS orqali baliq mahsulotiga bo'lgan talablarni baholash	365
11.8. 2SLS ning alternativlari	366
11.9. Monte Karlo simulatsiya natijalari	372
XII-BOB. VAQT KETMA-KETLIGI MA'LUMOTLARI BILAN REGRESSIYA: STATSIONAR BO'LMAGAN O'ZGARUVCHILAR ...	382
12.1. Statsion va nostatsionar ma'lumotlar	382
12.1.1. Statada sanalarni yaratishni ko'rib chiqish	383
12.1.2. Sanalarni formatlash	384
12.1.3. Ma'lumotlarning grafigini tuzish	385
12.2. Soxta regressiyalar	390
12.3. Statsionarlik uchun unit root testlari	392
12.4. Integratsion va kointegratsion	403
12.4.1. Engle-Granger Testi	403
12.4.2. Xatolarni tuzatish modeli	405
XIII-BOB. VEKTOR XATOLARINI TUZATISH VA VEKTOR AVTOREGRESSIV MODELLARI	410
13.1. VEC va VAR modellari	410
13.2. VEC modelini baholash	411
13.3. VAR modelini baholash	416
13.4. Impuls javoblari va variatsiyalarni ayirishlari	424
Mustaqil ishlash uchun masalalar	432
Foydalanilgan adabiyotlar ro'yxati	437
Statistik taqsimot jadvallari	438
Ilovalar	443

СОДЕРЖАНИЕ

Введение.....	4
ГЛАВА 1. КРАТКАЯ ИНФОРМАЦИЯ ОБ ПРИКЛАДНОЙ ЭКОНОМЕТРИКИ И ПРОГРАММЕ STATA	6
1.1. Предмет прикладной эконометрики.....	6
1.2. Последовательные задачи науки прикладной эконометрики.....	8
1.3. Разработка математической модели и последовательность решения задач	14
1.4. Запустите и закройте программу STATA.....	18
1.5. Ввод и загрузка данных в программу STATA	21
1.6. Опишите данные и получите описательную статистику.....	25
1.7. Как сохранить результаты.....	28
ГЛАВА 2. СОЗДАНИЕ РЕГРЕССИОННЫХ МОДЕЛЕЙ В STATA	32
2.1. Загрузите данные для создания классической линейной модели.....	32
2.1.1. Зарегистрируйте файл ежедневного журнала и используйте	33
2.1.2. Описать и перечислить данные	34
2.2. Обобщение аналитических данных.....	36
2.3. Методы создания диаграмм SCATTER на основе данных ...	38
2.4. Способы как создать регрессионную модель в STATA (регресс).....	42
2.4.1. Расчетное значение ($\hat{Y}$ прогноз) и остатки	44
2.4.2. Расчет эластичности (margins)	47
2.4.3. Постройте рассчитанную линию модели регрессии ...	50
2.4.4. Расчет вариации (var) и ковариации (cov)	53
2.5. Использование программы stata для получения прогностических значений регрессионной модели (edit&setobs)	55
2.6. Методы создания нелинейных регрессионных моделей..	58
2.6.1. Квадратичная модель (generate sqft2)	58
2.6.2. Линейно-логарифмическая модель (log-lin)	64
2.7. Использование индикаторных переменных в процессе регрессии (bysort&hist)	69*
ГЛАВА 3. ИНТЕРВАЛЬНАЯ ОЦЕНКА И ПРОВЕРКА ГИПОТЕЗ	80
3.1. Возможность использования программы stata для расчета критических значений Т-распределения (invtail)	80

3.2. Расчет и анализ проверки гипотез.....	83
3.2.1. Правосторонняя проверка экономической гипотезы (scalartstat)	85
3.2.2. Левосторонний тест экономической гипотезы (lincom)	86
3.2.3. Двусторонняя проверка экономической гипотезы	87
3.3. Методы расчета р-значений (ttail).....	88
3.3.1. Проверьте значение справа от р-значения	88
3.3.2. Проверьте левое значение р-значения	89
3.3.3. Проверить двусторонний тест р-значения	90
3.3.4. Найдите р-значения по Stata (lincom).....	91
3.3.5. Проверка и оценка параметров линейных функций (estatvce)	91
ГЛАВА 4. НЕЛИНЕЙНЫЕ РЕГРЕССИИ И ПРОБЛЕМЫ МОДЕЛИРОВАНИЯ	96
4.1. Вычислить самые маленькие квадраты	96
4.1.1. Редактирование данных	96
4.1.2. Оценить регрессию и получить результаты последующей оценки.....	96
4.1.3. $\hat{Y}$ Создать расчетный диапазон значений	98
4.2. Измерение совместимости	99
4.2.1. Корреляция и определение R^2	100
4.3. Эффекты увеличения и модификации данных	101
4.3.1. Форма линейно-логарифмической функции	102
4.3.2. Нарисуйте модель бревна заданной линии	104
4.3.3. Редактировать графики	105
4.4. Остаточный анализ	107
4.4.1. Тест Jarque-Bera	108
4.4.2. Критические значения распределения χ^2 -square	110
4.4.3. р-значения χ^2 -квадратное распределение	110
4.5. Полиномиальные модели	111
4.5.1. Оцените и проверьте линейную зависимость	112
4.5.2. Оценка и проверка кубического уравнения	115
4.5.3. Оценка логарифмической модели роста производительности	117
4.6. Расчет логарифмического уравнения месячной заработной платы	118
4.6.1. Лог-линейная модель	119
4.6.2. Расчет начисленной заработной платы	123
4.6.3. Создание графиков заработной платы	124

4.6.4. Коэффициент детерминации R^2	125
4.6.5. Интервалы прогнозирования в логарифмической модели	125
4.7. Лог-лог модели	126
ГЛАВА 5. МОДЕЛИ МНОГОФАКТОРНОЙ РЕГРЕССИИ	137
5.1. Моделирование на основе данных ресторана Bigandy'sburgerbarn.....	137
5.2. Предсказание самых маленьких квадратов	140
5.3. Точность получения статистических данных	141
5.4. Доверительные интервалы	142
5.4.1. Доверительный интервал для линейной комбинации параметров.....	144
5.5. Тесты гипотез	146
5.5.1. Двусторонние тесты	147
5.5.2. Односторонние тесты	148
5.5.3. Проверить линейный параметр	149
5.6. Полиномиальные уравнения	151
5.6.1. Оптимальная реклама: нелинейная комбинация параметров	152
5.6.2. Использование факторных переменных для взаимодействий	155
5.7. Взаимодействия между переменными	158
5.8. Совместимость между моделями	158
ГЛАВА 6. ОЦЕНКА МОДЕЛЕЙ МНОГОФАКТОРНОЙ РЕГРЕССИИ	164
6.1. Расчет F-теста	164
6.1.1. Проверьте важность модели	169
6.1.2. Корреляция между t- и F-тестами	171
6.1.3. Дополнительные общие F-тесты	172
6.2. Моделирование с использованием недемонстрационных данных	176
6.3. Особенности модели	178
6.3.1. Отсутствующие переменные	179
6.3.2. Незначительные переменные	180
6.3.3. Выбор модели	181
6.4. Отсутствие информации, коллегиальность и незначительность	190
ГЛАВА 7. ИСПОЛЬЗОВАНИЕ ИНДИКАТОРА ИЗМЕНЕНИЙ	201
7.1. Переменный индикатор	201

7.1.1. Создание переменных индикаторов	202
7.1.2. Оценить регрессию переменного индикатора	203
7.1.3. Проверка значимости переменных	204
7.1.4. Дополнительные расчеты	205
7.1.5. Рассчитать средний предельный эффект	205
7.2. Используйте переменный индикатор	207
7.2.1. Взаимодействие факторов качества	208
7.2.2. Добавьте региональные переменные	210
7.2.3. Выполните тест на равенство для двух регрессий	212
7.2.4. Оценка отдельных регрессий	213
7.2.5. Применение переменных индикаторов к логарифмическим моделям	215
7.3. Линейная вероятностная модель	217
ГЛАВА 8. ГЕТЕРОСКИДАСТИК	224
8.1. Важность гетероскидаста	224
8.2. Определение гетероскедастичности	226
8.2.1. Остаточные области	226
8.2.2. Тесты множителей Лагранжа	228
8.2.3. Тест Goldfeld-Quandt	232
8.3. Стандартные ошибки гетероскедастического развития ..	235
8.4. Обобщенный метод наименьших квадратов	238
8.4.1. GLS с использованием сгруппированных данных	240
8.4.2. Общее состояние GLS	214
8.5. Гетероскедастичность в линейной вероятностной модели	243
ГЛАВА 9. ДИНАМИЧЕСКАЯ РЕГРЕССИЯ МАССИВА	254
9.1. Базовые концепты	254
9.1.1. Представление графика временных рядов в Stata	255
9.1.2. Графики временных рядов	259
9.1.3. Лаги и разные переменные стага	260
9.2. Ограниченные распределенные задержки (лаги)	262
9.3. Серийная корреляция	264
9.4. Другие тесты на серийную корреляцию	270
9.5. Оценка с последовательными ошибками	273
9.5.1. Минимальные квадраты и стандартные ошибки НАС ...	273
9.5.2. Самые маленькие квадраты без линий	276
9.5.3. Наиболее часто используемая модель	278
9.6. Авторегрессионные модели с распределенным запаздыванием	280
9.6.1. Кривая Филлипса	282

9.6.2. Окуп закон	285
9.6.3. Авторегрессионные модели	288
9.7. Прогнозирование	290
9.7.1. Прогнозирование с помощью модели AR	291
9.7.2. Экспоненциальное выравнивание	293
9.8. Мультипликационный анализ	298
9.9. Дополнительные тесты	302
9.9.1. Тест Дарбина-Ватсона	302
9.9.2. Prais-Winsten FGLS	302
ГЛАВА 10. СЛУЧАЙНЫЕ РЕГРЕССОРЫ И ОЦЕНКА	
ПО МОМЕНТАМ	314
10.1. Оцените наименьшие квадраты уравнения заработной платы	314
10.2. Самые маленькие квадраты с двумя шагами	316
10.3. Оценка IV с дополнительными переменными	321
10.3.1. Описать частичные корреляции	324
10.4. Тест HAUSMAN на эндогенность	327
10.5. Проверить допустимость дополнительных переменных	331
10.6. Тест на слабые переменные	332
10.7. Вычисление F-статистики Крэгга-Дональда	338
10.8. Опыт моделирования	340
ГЛАВА 11. СИНХРОНИЧЕСКИЕ РАВНОВЕСНЫЕ МОДЕЛИ	352
11.1. Модель спроса и предложения ТРЮФЕЛЬ	352
11.2. Оценка уравнений в сокращенной форме	353
11.3. 2SLS оценка спроса на трюфели	354
11.4. Оценка модели предложения трюфельной компании через 2SLS	360
11.5. Модели спроса и предложения на рыбную продукцию	361
11.6. Сокращенные формы для цены и количества рыбы	363
11.7. Оценка спроса на рыбную продукцию через 2SLS	365
11.8. Альтернативы 2SLS	366
11.9. Результаты моделирования методом Монте-Карло	372
ГЛАВА 12. РЕГРЕССИЯ С ДАННЫМИ ВРЕМЕНИ ДОСТАВКИ:	
НЕСТАТИСТИЧЕСКИЕ ПЕРЕМЕННЫЕ	382
12.1. Стационарные и нестационарные данные	382
12.1.1. Рассмотрите возможность создания дат в stata	383
12.1.2. Форматирование дат	384
12.1.3. Графические данные	385
12.2. Ложные регрессии	390

12.3. Единичные корневые тесты для стационарных	392
12.4. Интеграция и коинтеграция	403
12.4.1. Тесты Энгла-Грейнджера	403
12.4.2. Модель исправления ошибок	405
ГЛАВА 13. ВЕКТОРНАЯ КОРРЕКЦИЯ ОШИБОК И ВЕКТОРНЫЕ	
АВТОРЕГРЕССИВНЫЕ МОДЕЛИ	410
13.1. Модели VEC и VAR	410
13.2. Оценка модели VEC	411
13.3. Оценка модели VAR	416
13.4. Импульсные отклики и разделение вариантов	424
Вопросы для самостоятельной работы	432
Список использованной литературы	437
Таблицы статистического распределения	438
Приложения	443

N.N.MURADULLAYEV,
X.N.SABIROV

AMALIY EKONOMETRIKA

O'quv qo'llanma

Muharrir: M.Talipova
Musahhih: N.Niyazova
Kompyuterda tayyorlovchi: B.Haydarov

**Bosishga ruxsat etildi: 20.12.2024-yil
Bichimi 60x84 1/16. «Segoe UI»
garniturada raqamli bosma usulda chop etildi.
Shartli bosma tabog'i 14,75. Adadi 100. Buyurtma № 24-348**

**“Dimal” MCHJ nashriyoti, 100071,
Toshkent, Huvaydo ko'ch 2A-25.**

**“IMPRESS MEDIA” MChJ bosmaxonasida chop etildi.
Toshkent shahri, Qushbegi ko'chasi, 6-uy.**