Ekonometrinis modeliavimas su *EViews:* praktinis gidas

Vita Karpuškienė Povilas Lastauskas

Versija 10-03

Turinys

Apie Praktinį gidą

Šio „koncentruoto“ gido tikslas yra padėti išmokti modeliuoti ekonominius reiškinius kompiuterio pagalba, pasitelkiant šiuolaikinius ekonometrijos mokslo metodus.

Jis pravers ne tik VU EF studentams, klausantiems kursą „Ekonometrija II“ bei atliekantiems praktines užduotis, bet ir vėliau darbuojantis ar rašant kursinius darbus – visur kur susidursite su empirine analize.

Dėl ekonometrijos mokslo spartaus progreso, negalime pateikti net mažos dalies pasiekimų. Kita vertus, mūsų tikslas yra kitas: praktinių žinių suteikimas, norintiems atlikti ekonomiškai ir statistiškai patikimą analizę. Dėl šios priežasties, gidas nėra vien procedūrų aprašymas, bet drauge ir esminės teorijos šaltinis. Pastaroji yra būtina, jog modelis būtų patikimas, o išvados pagrįstos. Nemažiau svarbu yra suvokti modelio ribotumą, rezultatų pagrįstumą bei galiojimo ribas ir jautrumą prielaidų galiojimui.

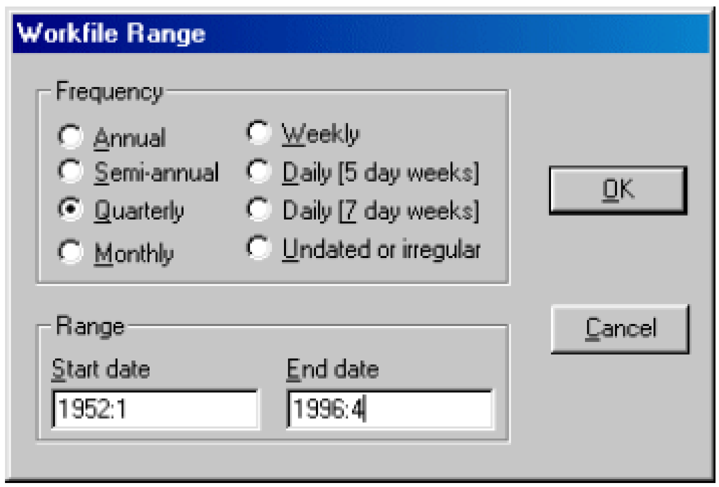
EViews ir ekonominių reiškinių modeliavimas

EViews programa skirta sudėtingai duomenų analizei, regresijai įvertinti, taipogi prognozavimui *Windows* aplinkoje. Ši programa taipogi gali būti skirta moksliniams tyrimams, finansinei analizei, makroekonominiam prognozavimui, bendrajai ekonominei politikai, simuliacijai, kaštų analizei, t.t.

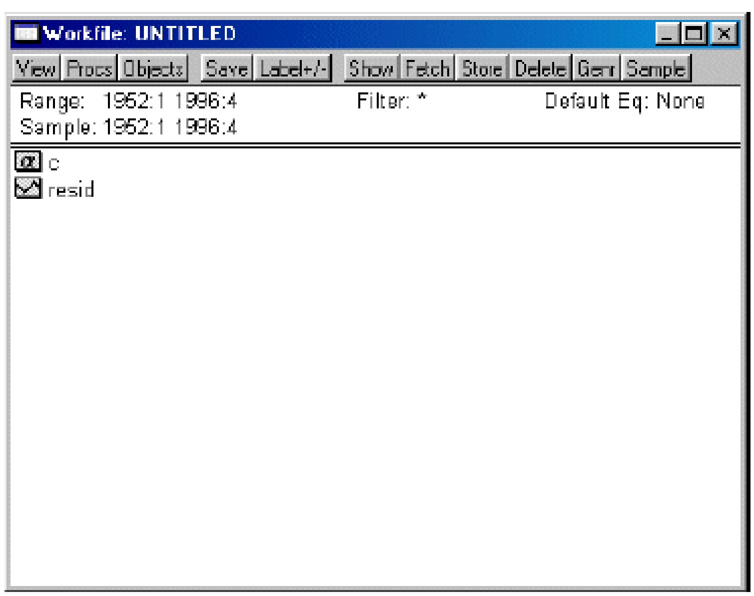
Nors ir būdamas galingas ir efektyvus ekonometrinės analizės įrankis, *EViews* yra draugiškas vartotojo atžvilgiu. Vartotojo sąsaja patogi ir paprasta, o puiki bei plati **Help** funkcija leidžia nesudėtingai išmokti naudotis visomis programos galimybėmis, maža to, ji taipogi paaiškina metodologiją bei įrankius, kuriais remiantis yra suformuojamas vienoks ar kitoks rezultatas. Tad šis trumpas vadovas suteiks tik pagrindinius gebėjimus regresinei analizei bei prielaidų tikrinimui, nes visa kita gali būti nesudėtingai „atrasta“ naudojantis programa.

Pirmieji žingsniai – duomenų įvedimas

Pasirinkę **File/New/Workfile**, pamatysite dialogo langą, kuriame turite pasirinkti, kokiu dažniu bus pateikti laiko eilučių duomenys (metiniai, pusmetiniai, ketvirtiniai, t.t.) bei pradžios ir pabaigos datas. Esant kryžminiams arba be nurodytos datos duomenims pasirinkite opciją „Undated or irregular“.



Atlikę šiuos veiksmus ir nuspaudę **OK**, išvysime kitą langą (darbalaukį):



Pastebėkime, jog kiekvienas *EViews* darbalaukis visuomet turės du elementus – *c* (koeficientų vektorius) ir *resid* (paklaidos), beje, atkreiptinas dėmesys, jog *resid* išlaikys paskutiniosios atliktos regresijos paklaidas, tad, norėdami jas išsaugoti prieš „prasukant“ kitą regresiją, išsaugokite suteikdami kitą pavadinimą (pvz., *resid_bvp*). Šiuo atveju galime importuoti duomenis iš kitų šaltinių, žinoma, dažniausiai naudojamas yra *MS Excel*. Paprasčiausias metodas – copy-paste, naudojamas duomenis perkelti iš atverto *.xls failo (darbalaukyje spustelkite dešiniuoju pelės mygtuku, rinkitės **New Object/Series**, **Edit+/-** ir **Paste**). Kitas būdas, naudojamas duomenims importuoti tiesiai iš *Excel* failo (tuomet nereikia atidaryti pačio failo, žymėti duomenų, t.t. – *EViews* pats tai atliks, jeigu duomenys bus tinkamai suvesti), pasirinkus **File/Import/Read Text-Lotus-Excel**, galop norėdami rankiniu būdu suvesti duomenis, naudokite jau aprašytą metodą – spragtelkite dešiniuoju pelės klavišu, **New Object/Series**, **Edit+/-** ir surašykite duomenis. Juos suvedę, spustelkite **Edit+/-** ir, uždarant langą, pavadinkite duomenis. Kitas rankinio suvedimo būdas – **Procs/Generate series**, įveskite lange x=0. Gausite duomenų masyvą su visais nuliais – jį atvėrę, vėlgi naudokitės **Edit+/-** opcija duomenims suvesti ir patvirtinti.

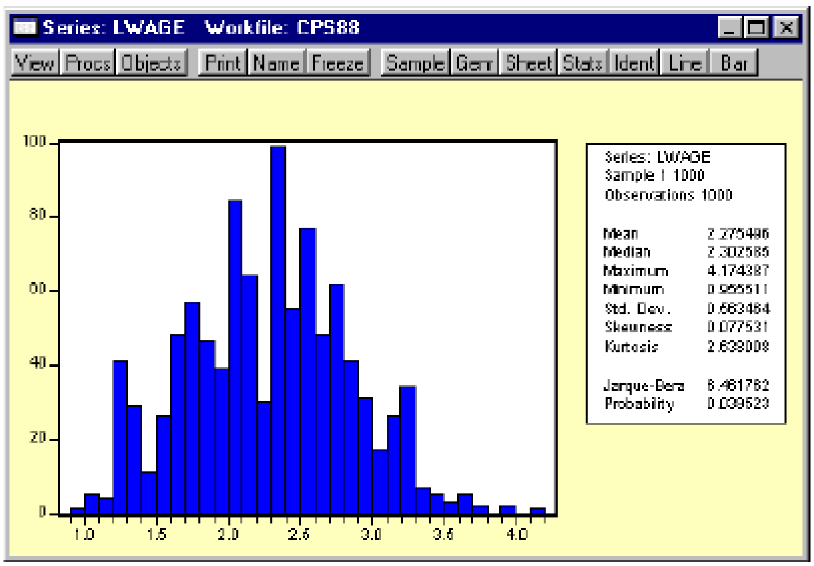
Suteikdami pavadinimus duomenų eilutėms, venkite bet kurio iš šių, nes jie jau „rezervuoti“, t.y. naudojami kaip funkcijos, operatoriai – ABS, ACOS, AR, ASIN, C, CON, CNORM, COEF, COS, D, DLOG, DNORM, ELSE, NDIF, EXP, LOG, LOGIT, LPT1, LPT2, MA, NA, NRND, PDL, RESID, RND, SAR, SIN, SMA, SQR ir THEN.

Aprašomoji statistika

- **Quick – Series Statistics – Histogram and Stats**

Histograma

Histograma yra bene svarbiausia aprašomosios statistikos grafinė priemonė (atvaizdavimo būdas), leidžianti pamatyti duomenų pasiskirstymą, suskaidžius visą duomenų seką nuo mažiausio iki didžiausio į lygias dalis. Prie histogramos yra pateikiama svarbiausių rodiklių lentelė, kurioje reikšmės gautos pagal duomenis jūsų stebimoje atrankoje. Trumpai aptarkime juos bei gautų reikšmių interpretaciją.



Vidurkis – vidutinė stebėjimų reikšmė ($\bar{y} = \frac{1}{n}\sum_{i=1}^{n} X_i$, kur n – stebėjimų skaičius, X_i – i-ojo stebėjimo skaitinė vertė).

Mediana – vidurinis stebėjimų sekos, išdėstytos nuo mažiausios iki didžiausios reikšmės, narys (esant lyginiam stebėjimų skaičiui – vidurinių narių vidurkis).

Minimumas ir maksimumas – didžiausia ir mažiausia stebėjimų, esančių atrankoje, reikšmė.

Standartinis nuokrypis

$$s = \sqrt{\sum_i (y_i - \bar{y})^2 / (N-1)}$$

čia: N – stebėjimų skaičius, $\bar{y}$ - duomenų eilutės vidurkis

Asimetrija

$$S = 1/N \sum_i ((X_i - \bar{X})/\hat{\sigma})^3$$

$\hat{\sigma}$ – standartinio nuokrypio įvertis, paremtas paslinktu dispersijos įverčiu[1]: $\hat{\sigma} = s\sqrt{(N-1)/N}$.

Normaliojo skirstinio sklaida yra lygi „0“, beje, kaip ir bet kurio simetriško skirstinio. Gavus teigiamą reikšmę turime skirstinį su ilga dešiniąja uodega ir, atvirkščiai, neigiama reikšmė indikuoja apie ilgą kairiąją uodegą.

Ekscesas

$$K = 1/N \sum_i ((y_i - \bar{y})/\hat{\sigma})^4$$

čia: $\hat{\sigma}$ – standartinio nuokrypio įvertis, paremtas paslinktu dispersijos įverčiu: $\hat{\sigma} = s\sqrt{(N-1)/N}$. Normaliojo skirstinio ekscesas lygus „3”.

Jarque-Bera statistika

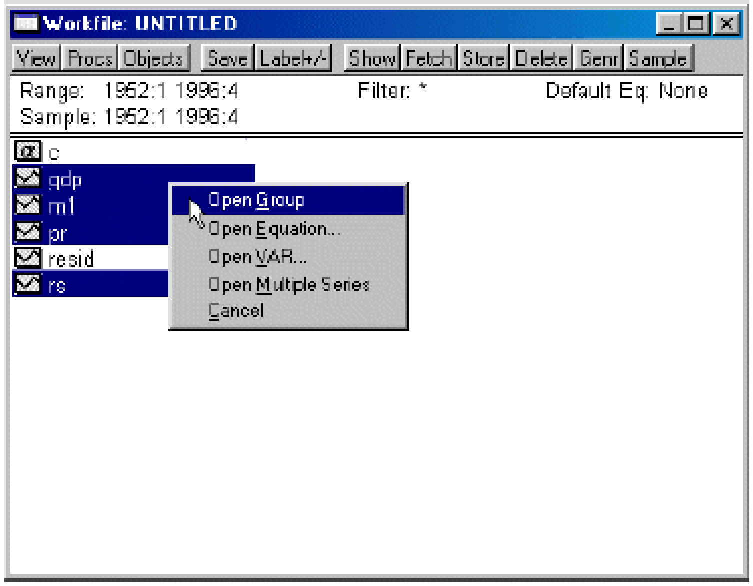
$$Jarque - Berra = \frac{N-k}{6}\left(S^2 + \frac{(K-3)^2}{4}\right)$$

čia: S – sklaida, K – ekscesas, k – įvertintų koeficientų skaičius. Jarque-Bera statistika pasiskirsčiusi pagal χ_2^2 skirstinį, postuluojama nulinė hipotezė yra $H_0 : X \sim (\mu, \sigma^2)$, tad jeigu duomenys X yra pasiskirstę ne pagal Gauso skirstinį, gausime labai mažą p reikšmę (ir galėsime atmesti H_0 būdami statistiškai tikri dėl nenormalaus pasiskirstymo).

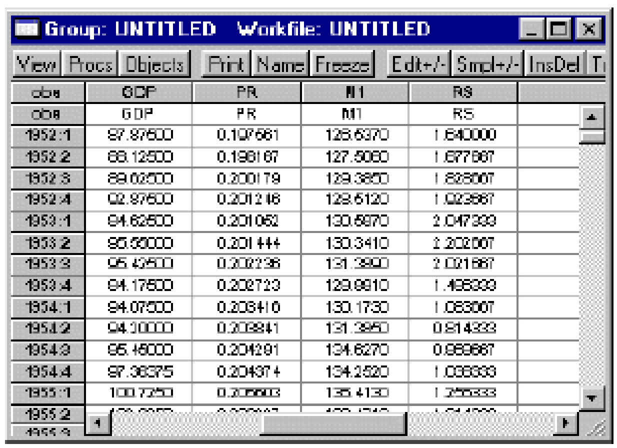
Keleto kintamųjų statistika

Pažymėję grupę kintamųjų, paspaudžiame dešinuoju pelės mygtuku ir pasirenkame grupės opciją „Open Group“.

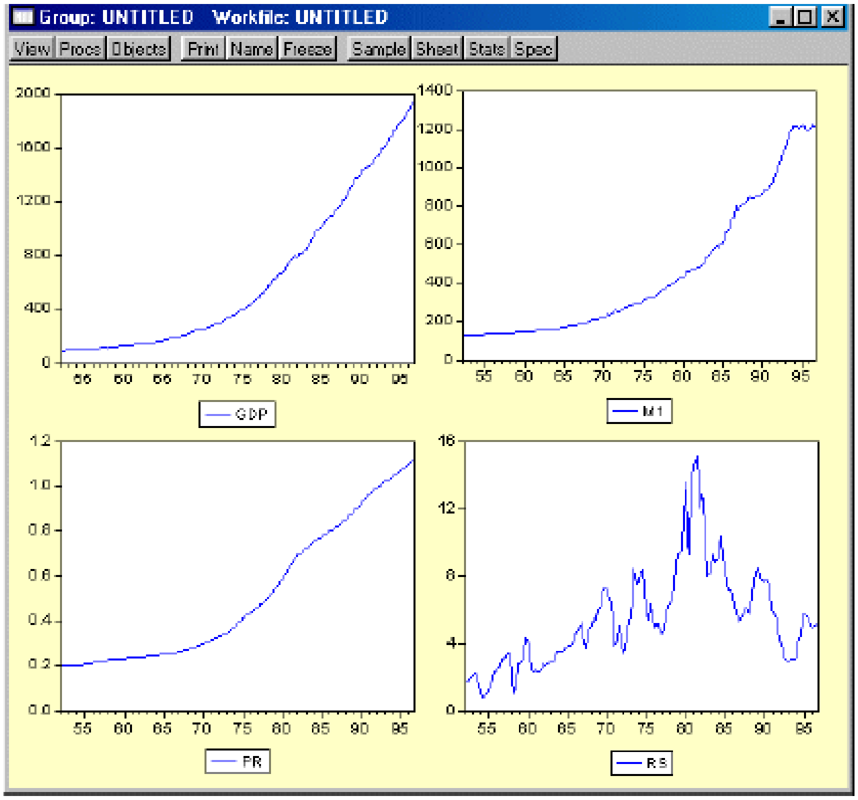
[1] Paslinktumas kyla mažose imtyse, nors didelėms imtims tai nedaro įtakos, nes $\frac{N-1}{N} = 1 - \frac{1}{N}$, $\lim_{N\to\infty}[1 - \frac{1}{N}] = 1$. Nepaslinktas yra *s*, žiūrėti skyrelį *Apibendrinimas* bei *Priedą*.



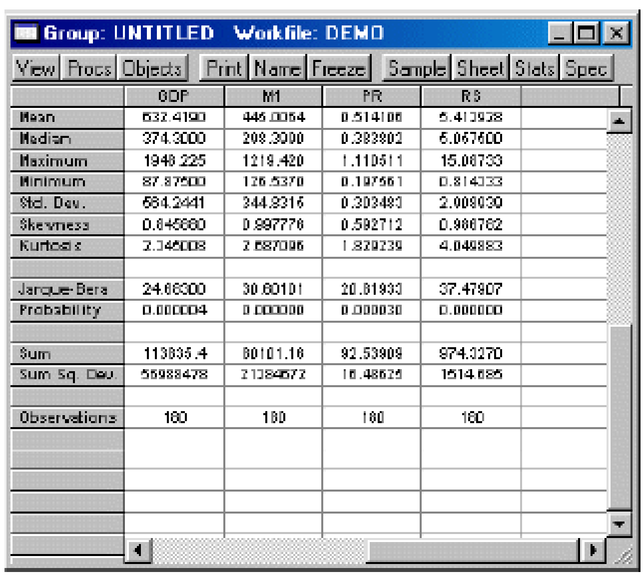
Pavyzdyje buvo pažymėti 4 duomenų masyvai, tad bus atvertas langas su pasirinktomis duomenų eilutėmis. Jos reikalingos ne tik tam, kad būtų galima sugretinti reikalingus duomenis, bet ir grafinei analizei.



Pavyzdžiui, norėdami pamatyti visų keturių duomenų grafikus viename lange, pasirinkime **View/Multiple Graphs/Line**. Išvysime langą, leidžiantį matyti kelių duomenų dinamines pozicijas.



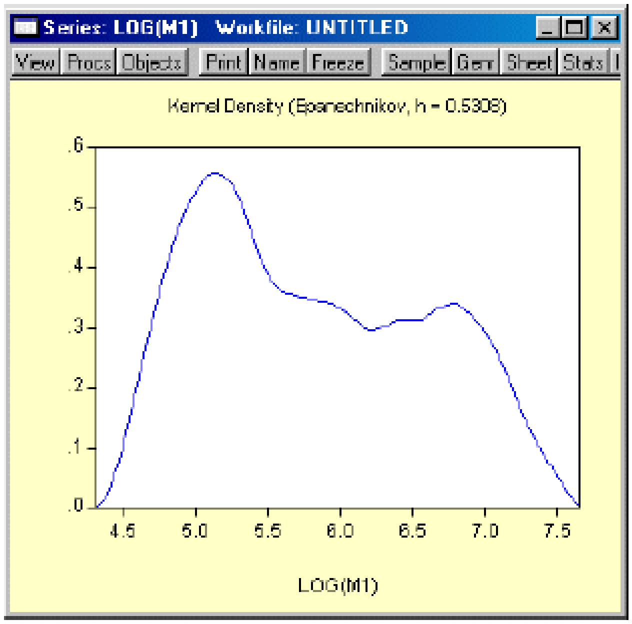
● **Quick/Group statistics/Descriptive statistics** duos tokį rezultatą:


Group: UNTITLED Workfile: DEMO

	GDP	M1	PR	RS
Mean	632.4190	445.0064	0.514106	5.413938
Median	374.3000	208.3000	0.383802	5.067600
Maximum	1948.225	1219.420	1.110511	15.08733
Minimum	87.87500	126.5370	0.197561	0.814333
Std. Dev.	584.2441	344.8316	0.303483	2.908939
Skewness	0.845660	0.997776	0.592712	0.986782
Kurtosis	2.345008	2.687096	1.829239	4.049883
Jarque-Bera	24.66300	30.60101	20.81933	37.47907
Probability	0.000004	0.000000	0.000030	0.000000
Sum	113835.4	80101.16	92.53909	974.3270
Sum Sq. Dev.	56988478	21384672	16.48625	1514.685
Observations	180	180	180	180

Vieno kintamojo statistika

Suprantama, jeigu galime rasti grupės eilučių analizę, tuomet dar paprasčiau bus nagrinėti individualias duomenų rūšis. Pažymėję kurį nors duomenų, kuriuos norime tirti, pavadinimą, spaudžiame **Quick/Show** ir patvirtinę gauname aprašomosios statistikos rodiklių langą. Tuomet einame į **View/Descriptive Statistics/Histogram and Stats** ir taipogi gausime histogramą. Įdomėsnis variantas – išglodinta histogramos versija, kurią išvysime pasirinkę **View/Distribution Graphs/Kernel Density**.



Pasinaudoję **Quick** galimybėmis, galėsime atvaizduoti kintamąjį pačiomis įvairiausiomis formomis – stulpeliais, linijomis, sklaidos diagrama, linija ir stulpeliais (apjungtas variantas) – tam tereikia pasirinkti **Quick/Graph/Line Graph** (arba **Scatter Diagram**, **Bar chart**, **Mixed Bar & Line**, t.t.).

Beje, atkreiptinas dėmesys, jog nebūtina naudotis **Quick** funkcija aprašomajai ir grafinei statistikai – pakanka atverti duomenų langą (*spreadsheet*) ir spustelėjus **View** gausime galimybę duomenis atvaizduoti tiesiniu (linijiniu) grafiku (**View/line graph**), stulpeline diagrama (**View/bar graph**), taipogi galėsime tikrinti aprašomosios statistikos rodiklius (**View/Tests for Descriptive Stats**), t.t. Šis metodas taikytinas ir turint keletą kintamųjų – tiesiog juos reikia pažymėti (**Ctrl**) ir atverti vieną langą su pasirinktų kintamųjų duomenimis (**Open as a Group**).

Kintamųjų generavimas. Funkcijos bei operatoriai *EViews*

Šis skyrelis trumpai apžvelgs, kokios funkcijos yra leidžiamos *EViews* bei kokie operatoriai yra atpažįstami programos. Pirma, trumpai galime nusakyti, kam reikalingos šios žinios – jos pravers generuojant naujus kintamuosius, vykdant simuliacijas, suvedinėjant modelio funkcinę išraišką, t.t.

Tad pradėkime nuo kintamųjų generavimo. Paprasčiausias būdas – įvedame **Genr** į formulių įvesties lauką ir spaudžiame **OK**. Gausime langą, kuriame turėsime nurodyti lygtį, pagal kurią bus sugeneruota duomenų eilutė. Ekvivalentūs metodai – **Procs/Generate Series** arba **Quick/Generate Series**. Įvesdami formulę pradėkite nuo naujo duomenų pavadinimo ir lygybės ženklo (pvz., „lx=log(x)“, gausime logaritmuotą kintamojo x eilutę, pavadintą „lx“).

EViews atpažįsta šiuos ženklus ir juos interpretuoja taip:

^ kėlimas laipsniu, pvz., x^2 (x^2)

* daugyba, pvz., x*y

/ dalyba, pvz, x/y

+ sudėtis, pvz, $x + y$

– atimtis, pvz., $x – y$

@LOG(X) – natūrinis (ln) kintamojo x logaritmas

@EXP(X) – eksponentinė x funkcija, e^x

@ABS(X) – x modulis (absoliutinė vertė)

@SQRT(X) – kvadratinė šaknis iš x

@SIN(X), @COS(X) – atitinkamos trigonometrinės funkcijos

@INV(X) – atvirkštinė x funkcija, t.y. $1/x$

@CEILING(X) – suapvalina x iki mažiausio sveikojo skaičiaus, <u>didesnio</u> arba lygaus x, pvz., @CEILING(3,45) = 4, @CEILING(6) = 6

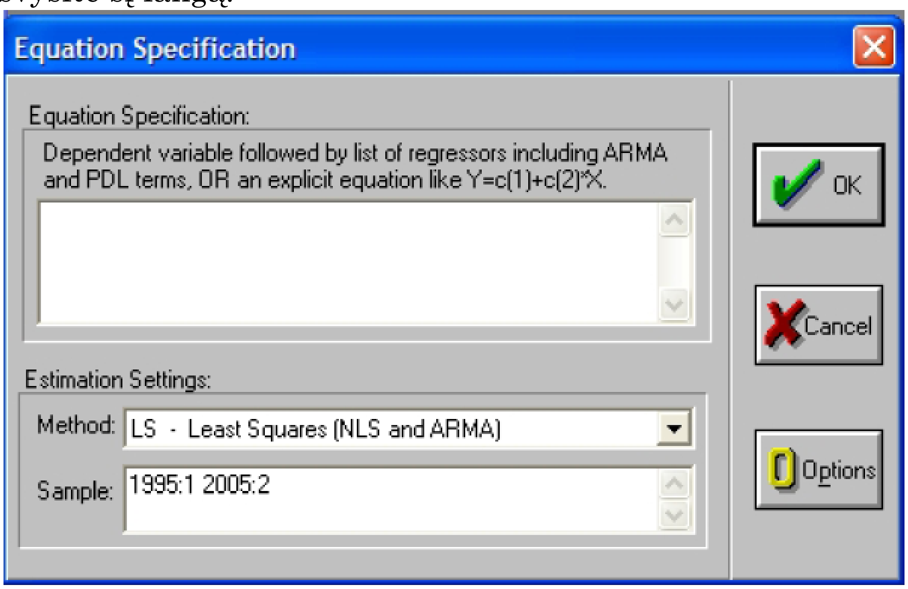
@FLOOR(X) – suapvalina *x* iki didžiausio sveikojo skaičiaus, <u>mažesnio</u> arba lygaus *x*, pvz, @FLOOR(2,23) = 2, @FLOOR(6) = 3

Mažiausių kvadratų metodas (MKM)

Norėdami įvertinti regresiją mažiausių kvadratų metodu (MKM), pasirinkite tokią žingsnių seką:

● **Quick/Estimate Equation**

Tuomet išvysite šį langą:



Įveskite lygties kintamuosius, pradėdami nuo priklausomojo kintamojo. Po kiekvieno įvesto kintamojo pavadinimo darykite tarpelį, be to, jeigu norite, kad modelis turėtų laisvąjį narį, nepamirškite nurodyti „c“ tarp priklausomojo ir pirmojo nepriklausomojo kintamojo. Pavyzdžiui, norime nustatyti, ar teorinė priklausomybė tarp investicijų ir BVP, realių palūkanų normų ir vyriausybės išlaidų (išstūmimo efektas) yra statistiškai reikšminga. Nesigilindami į gautų įverčių reikšmes (dydžius) bei ženklus, pažvelkime, kaip modelis pateikiamas *EViews* ir koks rezultatas gaunamas.

Įvedame: I C G BVP R, čia I – investicijos (regresantas), C – laisvojo nario koeficientas, kurį *EViews* apskaičiuos (konstanta), G, BVP, R yra atitinkamai vyriausybės išlaidų, BVP ir realių palūkanų normų duomenų eilučių pavadinimai. Rezultatą gausime toka forma:

Dependent Variable: I
Method: Least Squares
Sample: 1995:1 2005:2
Included observations: 22

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-8429.883	2081.156	-4.050578	0.0008
G	0.848156	0.371456	2.283327	0.0348
BVP	0.340755	0.046345	7.352537	0.0000
R	171.4737	52.01009	3.296931	0.0040
R-squared	0.913789	Mean dependent var	5942.153	
Adjusted R-squared	0.899420	S.D. dependent var	1780.438	
S.E. of regression	564.6534	Akaike info criterion	15.67327	
Sum squared resid	5739002.	Schwarz criterion	15.87164	
Log likelihood	-168.4059	F-statistic	63.59656	
Durbin-Watson stat	2.085201	Prob(F-statistic)	0.000000	

Kitas būdas įvesti tokiai lygčiai yra vietoje lygties kintamųjų išvardymo įrašyti pilną priklausomybę:

I=C(1)+C(2)*G+C(3)*BVP+C(4)*R

Tiesinėms regresijoms pirmumas teikiamas anksčiau pateiktam metodui, nes antrajame yra didesnė įvedimo klaidos tikimybė, be to, rezultatai pateikiami ne tokia patogia forma bei nepilna (vietoje kintamųjų vardų, matysime kintamųjų numerį, negausime F statistikos bei jos p reikšmės; šis lygties įvedimo metodas yra labiau skirtas netiesiniams modeliams, nes *EViews* iškart suvoks, jog netiesine forma apibrėžtą modelį reikia skaičiuoti naudojant netiesinius mažiausius kvadratus):

Dependent Variable: I
Method: Least Squares
Sample: 1995:1 2005:2
Included observations: 22
I=C(1)+C(2)*G+C(3)*BVP+C(4)*R

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-8429.883	2081.156	-4.050578	0.0008
C(2)	0.848156	0.371456	2.283327	0.0348
C(3)	0.340755	0.046345	7.352537	0.0000
C(4)	171.4737	52.01009	3.296931	0.0040
R-squared	0.913789	Mean dependent var	5942.153	
Adjusted R-squared	0.899420	S.D. dependent var	1780.438	
S.E. of regression	564.6534	Akaike info criterion	15.67327	
Sum squared resid	5739002.	Schwarz criterion	15.87164	
Log likelihood	-168.4059	Durbin-Watson stat	2.085201	

Įvertina regresija gali būti išsaugota spustelėjus **Name**. Beje, paspaudus **Estimate** galėsime modifikuoti jau įvertintą regresiją – taip nereikės įvedinėti naujos regresijos.

Hipotezių tikrinimas

Hipotezėmis galime tikrinti, esant dauginei regresijai:

- Vieną apribojimą, vieną koeficientą
- Vieną apribojimą, daugiau negu vieną koeficientą
- Daugiau negu vieną apribojimą

Dažniausiai taikomų testų ribotumas:

- *t*-testas yra ribotas – galime taikyti tik vienam apribojimui
- *F*-testas yra ribotas – galime tirti tik dvipuses hipotezes

Turime bet kokią konstantą (*c*) ir modelį $Y_i = \beta_1 + \beta_2 X_{2i} + ... + \beta_k X_{ki} + u_i$, tuomet hipotezės ir tinkami testai:

$H_0 : \beta_2 = c$ prieš $H_1 : \beta_2 > c$	*t-testas*
$H_0 : \beta_2 = c$ prieš $H_1 : \beta_2 \neq c$	*t-testas* arba *F-testas*
$H_0 : \beta_2 + \beta_3 = c$ prieš $H_1 : \beta_2 + \beta_3 \neq c$	*t-testas* arba *F-testas*
$H_0 : \beta_2 = \beta_3 = \beta_4 = 0$ prieš $H_1 :$ bent vienas $\beta_2, \beta_3, \beta_4 \neq 0$	*F-testas*
$H_0 : \beta_2 + \beta_3 = c, \beta_4 - \beta_5 = 0$ prieš $H_1 : \beta_2 + \beta_3 \neq c$ ir/arba $\beta_4 - \beta_5 \neq 0$	*F-testas*

Aptarsime kiek sudėtingesnius hipotezių tikrinimo variantus su *EViews*, tirdami parametrų apribojimus.

Parametrų apribojimai

Pradėkime nuo parametrų apribojimų testavimo. Naudosime *Wald* testą, tinkantį tikrinti tiek vieną, tiek keletą apribojimų regresijos koeficientams. Iš esmės galime tikrinti tiek tiesinius, tiek ir netiesinius apribojimus, o skirtumas bus naudojamas skirstinys – netiesiniams apribojimams *Wald* taps asimptotiniu χ^2 testu, o esant tiesiniams gali būti traktuojamas kaip *F*-testas. Naudojimas:

- **View/Coefficient Tests/Wald Coefficient Restrictions**

Įvedame apribojimus atskirdami juos kableliu.

Parametrų stabilumas gali būti testuojamas naudojant Chow lūžio arba prognozės testą. Renkamės:

● **View/Stability Tests/Chow Breakpoint Test** arba **View/Stability Tests/ Chow Forecast Test**. Abu yra *F*-testai, bet *EViews* pateikia ir *log likelihood ratio* rezultatus.

○ Pavyzdys

Apribojimų tikrinimo testai labai naudingi esant teoriniam pagrindui numanyti vienokią ar kitokią koeficientų priklausomybę arba esant poreikiui patikrinti modeliuotojo keliamas prielaidas. Nagrinėkime klasikinį ekonominį pavyzdį – gamybinę Cobb-Douglas funkciją. Ją apibrėžiame taip: $\log y_i = \beta_1 + \beta_2 \log l_i + \beta_3 \log k_i + u_i$[2]. Tuomet yra įprasta tikrinti, ar galioja pastovi masto grąža (matematiškai – ar funkcija tiesiškai homogeninė). Paprasčiausia tai atlikti *F* testu (*Wald* testu) – tikriname apribojimą $\beta_2 + \beta_3 = 1$. Turime tekstilės įmonių Maroke 1995 metais duomenis, kur *y*, *l*, *k* yra atitinkamai gamyba, darbas ir kapitalas. Gauname tokius rezultatus (suvedame **ls log(y) c log(L) log(K)**):

Dependent Variable: LOG(Y)
Method: Least Squares
Sample: 1 520
Included observations: 520

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.574703	0.129100	19.94347	0.0000
LOG(L)	0.907820	0.036282	25.02136	0.0000
LOG(K)	0.207026	0.024618	8.409408	0.0000

R-squared	0.798382	Mean dependent var	7.422180
Adjusted R-squared	0.797602	S.D. dependent var	1.734664
S.E. of regression	0.780402	Akaike info criterion	2.347737
Sum squared resid	314.8671	Schwarz criterion	2.372279
Log likelihood	-607.4117	F-statistic	1023.628
Durbin-Watson stat	2.031563	Prob(F-statistic)	0.000000

[2] Nesunku įsitikinti, jog įprastai apibrėžiama Cobb-Douglas funkcija yra ekvivalenti aprašytajai: $y_i = A l_i^{\beta_2} k_i^{\beta_3} e^{u_i}$, ją logaritmavę gauname $\log y_i = \log A + \beta_2 \log l_i + \beta_3 \log k_i + u_i = \beta_1 + \beta_2 \log l_i + \beta_3 \log k_i + u_i$, kur $\log A = \beta_1$.

Tuomet gauname tokį *EViews* rezultatą (**View/Coefficient Tests/Wald Coefficient Restrictions**, įvedame C(2)+C(3)=1):

Wald Test:
Equation: Untitled

Null Hypothesis:	C(2)+C(3)=1		
F-statistic	20.49441	Probability	0.000007
Chi-square	20.49441	Probability	0.000006

Statistinė interpretacija – *F* testo statistika yra lygi 20,4944 (*p* reikšmė lygi 0,0000), esant H_0, ji pasiskirsčiusi pagal *F*-skirstinį su 1 ir 517 laisvės laipsnių. Kritinė *F* reikšmė esant 5% reikšmingumo lygmeniui yra 3,84, taigi atmetame H_0, nes 20,4944>3,84 (arba palyginame *p* reikšmės metodu, 0,0000<0,05). Reziumuojame, jog gamybinė funkcija nepasižymi pastovia masto grąža (statistinė išvada).

Verta pastebėti, jog *Wald* testas tikrina tik dvipusę hipotezę. Jeigu norite atlikti vienpusę hipotezę, teks taikyti *t*–testą. Pavyzdžiui, jeigu postuluojame $H_0: \beta_2 + \beta_3 = 1$ su alternatyvia hipoteze dėl didėjančios masto grąžos $H_A: \beta_2 + \beta_3 > 1$; tuomet į modelį įtraukiame naują parametrą $\phi = \beta_2 + \beta_3 - 1$ (hipotezės taps $H_0: \phi = 0$, $H_1: \phi > 0$). Reparametrizuotas modelis įgaus tokią išraišką: $\log y_i - \log l_i = \beta_1 + \phi \log l_i + \beta_3 (\log k_i - \log l_i) + u_i$, o įvertintas modelis bus:

Dependent Variable: LOG(Y)-LOG(L)
Method: Least Squares
Sample: 1 520
Included observations: 520

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.574703	0.129100	19.94347	0.0000
LOG(L)	0.114847	0.025369	4.527075	0.0000
LOG(K)-LOG(L)	0.207026	0.024618	8.409408	0.0000

R-squared	0.155991	Mean dependent var	3.731788
Adjusted R-squared	0.152726	S.D. dependent var	0.847825
S.E. of regression	0.780402	Akaike info criterion	2.347737
Sum squared resid	314.8671	Schwarz criterion	2.372279
Log likelihood	-607.4117	F-statistic	47.77636
Durbin-Watson stat	2.031563	Prob(F-statistic)	0.000000

Taigi galime apibendrinti: vienpusis *t*-testas, kur $H_0: \phi = 0$ ir $H_1: \phi > 0$, *t*-statistika lygi 4,5271 (*p*=0,0000) ir yra, esant nulinei hipotezei, pasiskirsčiusi pagal t skirstinį su 517 laisvės laipsnių. Kritinė vienpusio testo reikšmė lygi 1,645, tad išvada atmesti H_0 (4,5271>1,645 arba p reikšmė 0,00<0,05). Užfiksuota (statistiškai reikšminga) didėjanti masto grąža.

Klasikinės tiesinės regresijos modelio prielaidų tikrinimas

Kai įvertinote regresiją, pats laikas diagnozuoti jos „sveikatą“, t.y. paklaidų normalumą, Gauss-Markov prielaidas[3], funkcinę formą, t.t.

Multikolinearumas

Multikolinearumas (interkoreliacija) yra situacija, kai regresoriai tarpusavyje koreliuoja, taip sukeldami problemų atskirti kiekvieno atskiro veiksnio efektą priklausomam kintamajam. Savaime suprantama, tobulas multikolinearumas aspkritai neleistų rasti parametrų. Šio teiginio teisingumu lengva įsitikinti žvelgiant į matricinę MKM įverčio išraišką: $\hat{\beta} = (X'X)^{-1}X'Y$, tada $(X'X)^{-1} = \frac{1}{|X'X|}[\mathrm{adj}(X'X)]$ būtų neapskaičiuojamas, nes $|X'X| = 0$, esant X matricos stulpeliams tiesiškai priklausomiems. Kitaip tariant, regresorių stebėjimų matrica X privalo būti pilno stulpelių rango. Tobulas multikolinearumas išvengiamas atidžiai parenkant modelio kintamuosius.

Paprastai susiduriame su netobulu multikolinearumu, nuo kurio laipsnio priklauso ir problemos dydis: įverčio dispersija yra apibrėžiama $\mathrm{var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$, tad, esant multikolinearumui, dispersijos ir standartinės paklaidos išaugs (nes jas „didins“ $(X'X)^{-1}$).

EViews mums teikia galimybę įvertinti multikolinearumo problemą per porines koreliacijas. Koreliacijų matrica gali būti gauta **Quick/Group**

---

[3] Priminimas – Gauss-Markov sąlygos dėl paklaidų reikalauja: 1) $E(u_i) = 0 \forall i$ arba paklaidų vidurkis lygus nuliui; 2) $\sigma^2_{u_i} = \sigma^2_u \forall i$ arba homoskedastiškumas; 3) $\sigma_{u_i u_j} = 0$ arba neautokoreliacija; 4) $\sigma_{X_i u_i} = 0$ arba regresorių (silpnasis) egzogeniškumas, t.y. regresorių ir paklaidų nekoreliacija.

Statistics/Correlations arba formulių lauke įvedame **cor *x y z*** (kur *x*, *y*, *z* yra kintamieji), tada, kai $r_{ij} \to 1$, multikolinearumas taps problema. Dauginės koreliacijos koeficientą tarp x_i ir kitų aiškinančiųjų veiksnių (R_i^2) išnaudoja VIF statistika, kuri randama taip: $VIF(\hat{\beta}_i) = \frac{1}{1 - R_i^2}$. Ją randame skaičiuodami regresijas. Galiausiai, suklusti turime, jei modelio R^2 aukštas, o *t*-statistikos mažos.

Multikolinearumas iš esmės gali būti sprendžiamas keičiant (koreguojant) modelį bei įtraukiant naujų veiksnių, nes nėra kokių nors mechaninių veiksmų ar taisyklių sumažinti problemą. Tiesa, netobulas multikolinearumas nekelia bėdų įverčiams – jie išlieka BLUE, *F*- ir *t*- testai galioja (jei G-M prielaidos ir normalumas tenkinami), *s.e.*↑, *t*↓, regresijos koeficientai jautrūs specifikacijai.

Heteroskedastiškumo testai

Heteroskedastiškumas – prielaidos $\sigma_{u_i}^2 = \sigma_u^2 \forall i$ pažeidimas, kuris nepaslenka MKM įverčių, bet jų dispersijos tampa neteisingos, *t*- ir *F*- testai nebegalioja, MKM tampa nebeefektyvūs.

Paprasčiausiai atliekamas testas *EViews*, skirtas heteroskedastiškumo aptikimui, yra **White testas** (asimptotinis χ^2 testas, preziumuojamas paklaidų normalumas).

- Tegu turime tokį modelį: $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$.
- Įvertinę modelį, gauname $\hat{u}_t$.
- Įvertiname naują regresiją, kur regresantas yra paklaidos, pakeltos kvadratu, o regresoriai – pirmame modelyje naudoti veiksniai, kvadratu pakelti veiksniai ir jų tarpusavio sandaugos. Mūsų atveju būtų: $\hat{u}_t = \delta_0 + \delta_1 X_{1t} + \delta_2 X_{2t} + \delta_3 X_{1t}^2 + \delta_4 X_{2t}^2 + \delta_5 X_{1t} X_{2t} + v_t$.
- Formuluojame hipotezes: $H_0 : \delta_1 = \delta_2 = ... = \delta_k = 0$, H_1 : bent vienas $\delta \neq 0$.

- Randame LM statistiką ($LM = nR^2$), čia n – stebėjimų skaičius naujoje regresijoje, R^2 naujosios regresijos determinacijos koeficientas.
- Atmetame H_0, jei $LM > \chi^2_{k,\alpha}$.

Spaudžiame:

● **View/ Residual Tests/ White Heteroscedasticity (cross terms** arba **no cross terms)**. Bendru atveju, White testas yra efektyvesnis su kryžminiais nariais („cross terms“), bet jeigu tai stipriai sumažins laisvės laipsnių skaičių, verta testą atlikti be kryžminių narių („no cross terms“).

Dar vienas testas – **Goldfeld-Quandt**, taip pat gali būti atliktas su *EViews*, tik teks atlikti daugiau rankinio darbo.

- Testas ribotesnis negu LM testai, nes tikrina tik vieną įtarimą dėl heteroskedastiškumo. Tegu galioja $\sigma_i^2 = \sigma^2 X_i^2$. Tada surūšiuojame (nuo didžiausios iki mažiausios reikšmės) duomenis pagal *X* kintamąjį, kuris ir yra atsakingas už heteroskedastiškumą.
- Įvertiname dvi naujas regresijas įprastai su pirmais ir paskutiniais 37,5% stebėjimų, bet pasirinkimas paprastai varijuoja tarp 1/6 ir 1/3 duomenų. Kitaip tariant, praleidžiame *c* vidurinių duomenų.
- Gavę abiejų regresijų RSS, randame *F*-statistiką: $F = \frac{RSS_1}{RSS_2}$, kur RSS_1 yra *RSS* iš regresijos su mažesnėmis X_i reikšmėmis. Jeigu turime *n* duomenų ir *k* parametrų (su laisvuoju nariu), tai *F*-statistika, esant homoskedastiškumo hipotezei ($H_0 : RSS_1 = RSS_2$), pasisikirsčiusi pagal $F_{0,5(n-c-2k),0,5(n-c-2k)}$ skirstinį.
- Atmetame H_0, jei $F > F_{0,5(n-c-2k),0,5(n-c-2k)}$.

Pradėkime nuo duomenų surūšiavimo (spaudžiame **Procs/Sort Series**, tačiau išrūšiavę neišsaugojame duomenų – antraip prarasime pradžioje buvusią tvarką). Atliekame regresijas su dviem sub-atrankomis. Patogiausia naudoti šį algoritmą (čia *y* – priklausomas kintamasis, *x* – regresorius, vardus pakeičiame pagal savo duomenis; *start* ir *end* nurodo pirmos ir antros atrankos pradžios ir pabaigos taškus):

smpl start end

ls y c x

scalar rss1=@ssr

--

smp1 start end

ls y c x

scalar rss2=@ssr

Tada sugeneruojame *F*-statistiką: **genr F_GQ=RSS1/RSS2** ir palyginame su kritine reikšme (ją galime gauti **genr f=@qfdist(.95,n1-k,n2-k)**, kur *n1* ir *n2* yra mūsų subatrankų dydžiai, o *k* – parametrų skaičius.

Dar viena galimybė – **Glejser testas** (labai panašus į toliau einantį Breusch-Pagan LM testą). Jo žingsniai tokie:

- Tegu turime tokį modelį: $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$.
- Įvertinę modelį, gauname $\hat{u}_t$.
- Įvertiname naują regresiją, kur regresantas yra absoliutinės paklaidos, o regresoriai – veiksniai, atsakingi už paklaidų heteroskedastiškumą. Pavyzdžiui: $|\hat{u}_t| = \delta_0 + \delta_i X_{it} + \delta_k X_{kt} + ... + \delta_m X_{mt} + v_t$.
- Formuluojame hipotezes: $H_0 : \delta_i = \delta_k = ... = \delta_m = 0$, $H_1 :$ bent vienas $\delta \neq 0$.
- Randame LM statistiką ($LM = nR^2$), čia n – stebėjimų skaičius naujoje regresijoje, R^2 naujosios regresijos determinacijos koeficientas.
- Atmetame H_0, jei $LM > \chi^2_{k,\alpha}$, kur k – parametrų skaičius naujoje regresijoje.

EViews Glejser testas gali būti atliktas įvertinus regresiją, išsaugant likutinius dydžius. Sugeneruojame juos: **genr ut=resid**, kadangi reikalingos absoliutinės paklaidų reikšmės, jas gauname taip: **genr absut=abs(ut)** ir, galop, apskaičiuojame naująją regresiją su heteroskedastiškumą įtakojančiais veiksniais: **ls absut c x1 x2...** Kaip paprastai, randame LM statistiką ($LM = nR^2$) ir ją palyginame su kritine reikšme.

Breusch-Pagan LM testas. Įvertiname regresiją, gauname paklaidas, tuomet nusprendžiame, kurie kintamieji yra atsakingi dėl paklaidų heteroskedastiškumo. Įvertiname papildomą regresiją (kvadratu pakelti likutiniai dydžiai), kur X yra mūsų spėjami paklaidų dispersiją lemiantys veiksniai: $\hat{u}_i^2 = a_0 + a_1 X_{ji} + a_2 X_{zi} + \ldots + a_m X_{mi} + e_i$. Nulinė hipotezė postuluoja homoskedastiškumą ($H_0 : a_1 = a_2 = \ldots = a_m = 0$), tuo tarpu alternatyvi, kaip įprasta, teigia, jog bent vienas koeficientas a_i yra nenulinis. Randame $LM = nR^2$ statistiką ir ją lyginame su kritine reikšme iš χ^2 skirstinio su m laisvės laipsnių.

○ Pavyzdys

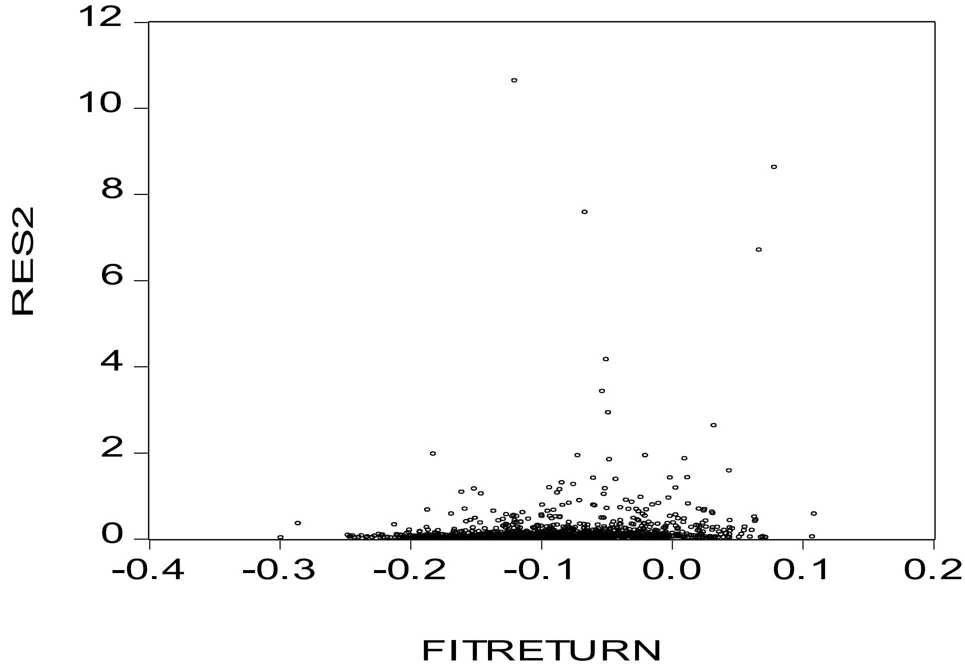
Turime kryžminius 1159 JAV kompanijų duomenis, tiksliau metinės grąžos (*return*), skolos ir aktyvų santykį (*debta*), turto (aktyvų) buhalterinę vertę (*netcap*). Sugeneruoame kintamąjį *size* suvesdami į formulių lauką taip: **genr size=log(netcap)**. Tuomet apskaičiuojame modelį. Rezultatai:

Dependent Variable: RETURN
Method: Least Squares
Sample: 1 1159
Included observations: 1159

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.082958	0.033210	2.497956	0.0126
DEBTA	-0.184649	0.063209	-2.921233	0.0036
SIZE	-0.028359	0.005890	-4.814982	0.0000

R-squared	0.025109	Mean dependent var	-0.086460
Adjusted R-squared	0.023422	S.D. dependent var	0.405974
S.E. of regression	0.401191	Akaike info criterion	1.013827
Sum squared resid	186.0631	Schwarz criterion	1.026912
Log likelihood	-584.5127	F-statistic	14.88649
Durbin-Watson stat	2.018539	Prob(F-statistic)	0.000000

Tikriname dėl heteroskedastiškumo. Pirma, pradėsime nuo grafinės analizės. Matome, jog įvertintas modelis sugretintas su kvadratine paklaida nėra „švarus“ (galime įžvelgti sisteminę priklausomybę, įtarimų kelia anomalinių reikšmių paplitimas augant įvertintam *return*):



Atlikime formalizuotų testų,pradėdami nuo Breusch-Pagan varianto. Šis testas reikalauja pačiam tyrėjui nustatyti, kurie kintamieji daro įtaką paklaidų dispersijai. Pavyzdžiui, numanome, jog galioja tokia heteroskedastiškumo forma: $\sigma^2_{u_i} = \delta_0 + \delta_1 size_i + \delta_2 size_i^2$. Tuomet $\sigma^2_{u_i}$ apskaičiuojame iš gautojo modelio, pateikto aukščiau, pakėlę likutinius dydžius kvadratu (pvz., **genr res2=resid*resid**). Įvertiname numanomą modelį:

Dependent Variable: RES2
Method: Least Squares
Sample: 1 1159
Included observations: 1159

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.596403	0.071926	8.291895	0.0000
SIZE	-0.153317	0.031281	-4.901265	0.0000
SIZE^2	0.010358	0.003133	3.306250	0.0010
R-squared	0.044160	Mean dependent var	0.160538	
Adjusted R-squared	0.042507	S.D. dependent var	0.573282	
S.E. of regression	0.560966	Akaike info criterion	1.684271	
Sum squared resid	363.7730	Schwarz criterion	1.697357	
Log likelihood	-973.0351	F-statistic	26.70401	
Durbin-Watson stat	2.074375	Prob(F-statistic)	0.000000	

Mūsų atveju $H_0: \delta_1 = \delta_2 = 0$, kai $H_1: \delta_1$ ir/arba $\delta_2 = 0$, LM testo statistika lygi $nR^2 = 1159 * 0{,}044 = 50{,}99$ ir, esant H_0, yra pasiskirsčiusi pagal skirstinį su 2 laisvės laipsniais (kritinė reikšmė, esant 5% reikšmingumo lygmeniui, yra 5,99; jei neturime statistinių lentelių, ją galime sugeneruoti *EViews*: **genr chi=@qchisq(.95,2)**), todėl atmetame homoskedastiškumo hipotezę (50,99>5,99).

○ **Perspėjimas!** Priėmus šią išvadą, akivaizdu, jog standartinės paklaidos yra paslinktos, o *t*- ir *F*-testai negalioja.

Nenuostabu, jog prie tokios pačios išvados prieitume ir pasinaudoję kitų testų pagalba. Pavyzdžiui, White testas (su kryžminiais nariais) duotų rezulatą, kur LM statistikos *p*=0,000, tad homoskedastiškumas atmestas bet kuriame reikšmingumo lygmenyje. Beje, *EViews* suteikia galimybę spręsti heteroskedastiškumo problemas su **White kovariacijų matrica** MKM įverčiams. Šis metodas pagrįstas tokia logika – tegu turime modelį $Y_i = \beta_1 + \beta_2 X_{2i} + u_i$, tuomet homoskedastiškų paklaidų dispersija:

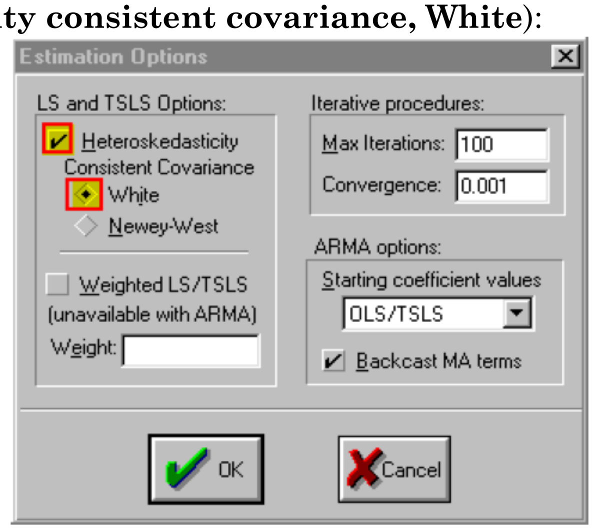
$E(u_i^2) = \sigma_u^2 \Rightarrow Var(b_2) = \frac{\sigma_u^2}{nVar(X)}$, heteroskedastiškumas: $E(u_i^2) = \sigma_{u_i}^2 \Rightarrow Var(b_2) = \frac{\sum w_i \sigma_{u_i}^2}{nVar(X)}$

su $w_i = \frac{(X_i - \bar{X})^2}{nVar(X)}$. Esant homoskedastiškumui randame $\sigma_{b_2}^2 = \frac{s_u^2}{nVar(X)}$, kur

$s_u^2 = \frac{\sum e_i^2}{n-2}$, tuo tarpu esant heteroskedastiškumui naudojame $\sigma_{b_2}^2 = \frac{\sum w_i s_{u_i}^2}{nVar(X)}, s_{u_i}^2 = e_i^2$.

White'as (1980) pademonstravo, jog didelėse atrankose šis įvertis yra suderintas. Be to, šio metodo privalumas – nereikia daryti jokių prielaidų dėl heteroskedastiškumo formos.

Atlikime šį testą *EViews*. Renkamės **Quick/Estimate equation/Options** (**Heteroskedasticity consistent covariance, White**):



Tuomet palyginame rezultatus :

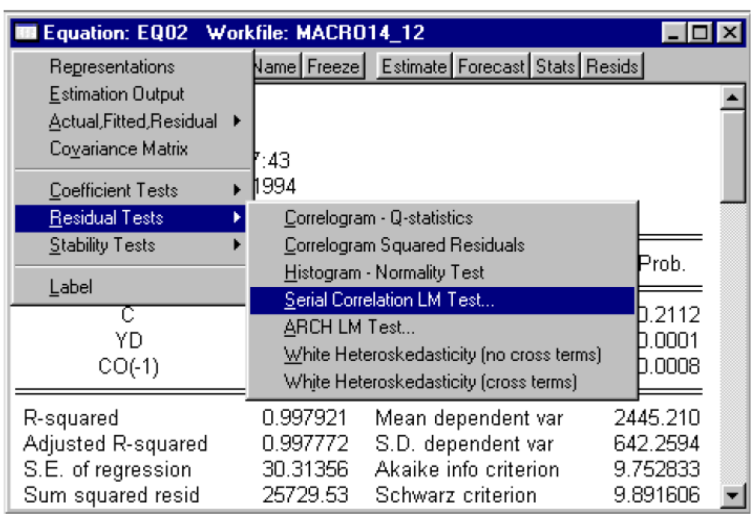
	MKM (įverčiai ir st. paklaidos)		White Heteroskedasticity (įverčiai ir st. paklaidos)	
C	0.082958	0.033210	0.082958	0.043040
DEBTA	-0.184649	0.063209	-0.184649	0.074720
SIZE	-0.028359	0.005890	-0.028359	0.006525

Kaip ir tikėtasi, parametrų įverčiai išliko identiški, tik pasikeitė jų paklaidos. Vis dėlto DEBTA ir SIZE statistinis reikšmingumas nepakito.

Autokoreliacija

Autokoreliacija – prielaidos $\sigma_{u_i u_j} = 0$ pažeidimas, kuris nepaslenka MKM įverčių, bet MKM koeficientų dispersijos klaidingos, *t*- ir *F*- testai negalioja, MKM nebeefektyvūs. Žinoma, šios išvados galioja, jei nėra kitų modelio prielaidų pažeidimų. Pavyzdžiui, esant priklausomam kintamajam su pavėlavimu tarp regresorių + autokoreliacija paklaidose = MKM įverčiai tampa paslinkti.

Tiriant autokoreliaciją, pristatome alternatyvą *Durbin-Watson* testui – **Breusch Godfrey LM** testą. Spaudžiame **View/Residual Test/Serial Correlation LM Test** ir įvedame lagų skaičių.



Breusch Godfrey yra asimptotinis χ^2 testas, kuris gerokai lankstesnis už *Durbin Watson*, nes leidžia įvertinti aukštesnės negu pirma eilė autokoreliaciją, be to, gali būti naudojamas esant endogeninių kintamųjų su pavėlavimais tarp regresorių (kaip žinia, *Durbin Watson*[4] tokiu atveju nepriimtinas).

Breusch Godrey LM testas preziumuoja, jog turime tokį daugines regresijos modelį: $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \ldots + \beta_m X_{mt} + u_t$, o paklaida yra tokios formos: $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \ldots + \rho_p u_{t-p} + \varepsilon_t$ (*p*-os eilės autokoreliacija). Tada, apjungę šias dvi lygtis, ir gausime modelį, kurį tikrina Breusch Godfrey LM testas:

- $e_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_k X_{kt} + \rho_1 e_{t-1} + \ldots + \rho_p e_{t-p} + \varepsilon_t$
- $H_0: \rho_1 = \ldots = \rho_p = 0$
- Suskaičiuojame LM (nR^2)
- Atmetame H_0, kai $LM > \chi^2_{p,\alpha}$ (α – reikšmingumo lygmuo)

○ Pavyzdys

Tiriame laiko eilutę – metinius JAV duomenis apie pinigų atsargą šalyje (*lms*), pajamas (*ly*) ir palūkanų normą (*ir*). Pinigų atsarga ir pajamos yra logaritmuoti dydžiai, o palūkanų norma išreikšta %. Randame modelio įverčius (**ls lms c ly ir**).

Dependent Variable: LMS
Method: Least Squares
Sample: 1900 1989
Included observations: 90

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.771121	0.042931	-17.96183	0.0000
LY	0.940341	0.019729	47.66228	0.0000
IR	-0.082912	0.005413	-15.31597	0.0000

R-squared	0.963653	Mean dependent var	0.977332
Adjusted R-squared	0.962817	S.D. dependent var	0.692195
S.E. of regression	0.133474	Akaike info criterion	-1.157049
Sum squared resid	1.549942	Schwarz criterion	-1.073722
Log likelihood	55.06718	F-statistic	1153.298
Durbin-Watson stat	0.561128	Prob(F-statistic)	0.000000

Pirma, tiriame pirmos eilės autokoreliaciją su Durbin-Watson statistika. Manome, jog paklaida išreiškiama taip: $u_t = \rho u_{t-1} + \varepsilon_t$, o nulinė hipotezė postuluoja pirmos eilės autokoreliacijos nebuvimą: $H_0: \rho = 0$, o alternatyvi – teigiamą autokoreliaciją $(H_1: \rho > 0)$. Turime n = 90, k = 2 (be laisvojo nario),

[4] Dėl šios kritikos, Durbin'as pasiūlė vadinamąją *Durbin's h* statistiką modeliams su regresantais, turinčiais vieną ar keletą vėlavimų, ir įtrauktiems tarp aiškinančiųjų veiksnių (regresorių). Dar vienas variantas *Durbin's m* testas.

tuomet gauname $d_L = 1{,}61$ bei $d_U = 1{,}70$. Apskaičiuota DW statistika lygi DW = 0,56 $(< d_L)$, taigi atmetame nulinę hipotezę ir reziumuojame, jog modelis pasižymi reikšminga teigiama pirmos eilės autokoreliacija.

Kaip jau minėta yra lankstesnių testų – taigi atlikime vieną iš jų – *Breusch Godfrey* testą. Atlikę nurodytus veiksmus, bei pasirinkę du vėlavimus (lagus) paklaidos išraiškoje, gauname:

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	52.76162	Probability	0.000000
Obs*R-squared	49.84742	Probability	0.000000

Test Equation:
Dependent Variable: RESID
Method: Least Squares

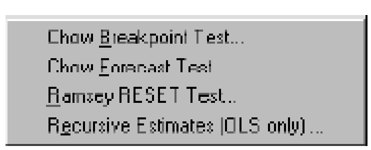
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.002870	0.029013	-0.098922	0.9214
LY	-0.002836	0.013516	-0.209794	0.8343
IR	0.001973	0.003826	0.515699	0.6074
RESID(-1)	0.915324	0.104472	8.761422	0.0000
RESID(-2)	-0.270260	0.108149	-2.498950	0.0144

R-squared	0.553860	Mean dependent var	-3.05E-16
Adjusted R-squared	0.532865	S.D. dependent var	0.131966
S.E. of regression	0.090195	Akaike info criterion	-1.919727
Sum squared resid	0.691491	Schwarz criterion	-1.780849
Log likelihood	91.38772	F-statistic	26.38081
Durbin-Watson stat	1.816933	Prob(F-statistic)	0.000000

Kaip įprasta, aprašome testo rezultatus. Buvo matuota tokia modelio paklaida: $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$; su tokiomis hipotezėmis: $H_0 : \rho_1 = \rho_2 = 0$ ir $H_1 : \rho_1$ ir/arba ρ_2 nelygūs nuliui. LM statistika lygi nR^2, taigi 90*0,55 = 49,84, esant nustatytai nulinei hipotezei, LM pasiskirsčiusi pagal χ^2 skirstinį su 2 laisvės laipsniais (kritinė 5% reikšmingumo lygmens reikšmė yra 5,99). Atmetame H_0, nes 49,84>5,99. Reziumuojame dar kartą patvirtinę autokoreliacijos problemą modelyje.

Funkcinė forma

Tikriname modelio funkcinę formą naudodamiesi RESET testu (**View/Stability Tests/Ramsey RESET Test**):



Pavyzdžiui, norime patikrinti, ar gamybai modeliuoti tinka Cobb-Douglas funkcinė išraiška. Panaudojame du narius (angl. *fitted terms*), tuomet *EViews*, skaičiuodama RESET testą, naudos tokį modelį: $\log y_i = \delta_1 + \delta_2 \log l_i + \delta_3 \log k_i + \delta_4 (\log \hat{y}_i)^2 + \delta_5 (\log \hat{y}_i)^3 + \xi$, kur $\log \hat{y}_i$ yra įvertina lygtis $\log \hat{y}_i = b_1 + b_2 \log l_i + b_3 \log k_i$, o hipotezės yra $H_0 : \delta_4 = 0, \delta_5 = 0$ prieš alternatyvią $H_A : \delta_4 = 0$ ir/arba $\delta_5 = 0$. Gauname tokį *EViews* rezultatą:

Ramsey RESET Test:

F-statistic	2.838941	Probability	0.059403
Log likelihood ratio	5.701634	Probability	0.057797

Test Equation:
Dependent Variable: LOG(Y)
Method: Least Squares
Sample: 1 520
Included observations: 520

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.643559	0.287349	9.199827	0.0000
LOG(L)	-0.519417	0.802086	-0.647583	0.5175
LOG(K)	-0.118979	0.185320	-0.642022	0.5211
FITTED^2	0.191104	0.118388	1.614216	0.1071
FITTED^3	-0.007410	0.005150	-1.438806	0.1508

R-squared	0.800581	Mean dependent var	7.422180
Adjusted R-squared	0.799032	S.D. dependent var	1.734664
S.E. of regression	0.777641	Akaike info criterion	2.344465
Sum squared resid	311.4335	Schwarz criterion	2.385367
Log likelihood	-604.5609	F-statistic	516.8744
Durbin-Watson stat	2.013868	Prob(F-statistic)	0.000000

Kaip matome, F testo statistika yra lygi 2,8389 (p reikšmė lygi 0,0594) bei, esant H_0, yra pasiskirsčiusi pagal F-skirstinį su 2 ir 515 laisvės laipsniais. Kritinė F reikšmė (5% lygmuo) lygi 3,00, tad negalime atmesti H_0, nes 2,8389<3,00 arba 0,059>0,05 (p reikšmės metodas).

○ **Perspėjimas!** Kadangi esame ties atmetimo riba, verta svarstyti reikšmingumo lygmens pakeitimą, mat II tipo klaida čia svarbesnė, t.y. nenorime klaidingai priimti išvadą, jog nėra specifikacijos klaidų. Taigi būtų prasminga, priėmus 10% reikšmingumo lygmenį, reziumuoti, jog turime modelio funkcinės formos (specifikacijos) klaidų.

Norėdami išbandyti įvairias funkcines formas *EViews*, tiesiog tinkamai apibrėžkite funkciją formulių lauke. Pateikiame keletą pavyzdžių:

Funkcinė forma	**Lygtis**	***EViews***
Tiesinė	$Y=\beta_0+\beta_1 X_1+\beta_2 X_2$	Y C X_1 X_2
Log-log	$\ln Y=\beta_0+\beta_1 \ln X_1+\beta_2 \ln X_2$	log(Y) C log(X_1) log(X_2)
Lin-log	$Y=\beta_0+\beta_1 \ln X_1+\beta_2 X_2$	Y C log(X_1) X_2
Log-lin	$\ln Y=\beta_0+\beta_1 X_1+\beta_2 X_2$	log(Y) C X_1 X_2
Polinomas	$Y=\beta_0+\beta_1 X_1+\beta_2 X_1^2+\beta_3 X_2$	Y C X_1 X_1^2 X_2
Atvirkštinė	$Y=\beta_0+\beta_1 (1/X_1)+\beta_2 X_2$	Y C 1/ X_1 X_2
Su fiktyviu poslinkio kintamuoju	$Y=\beta_0+\beta_1 X_1+\beta_2 D_1$	Y C X_1 D_1
Su fiktyviu poslinkio ir posūkio kintamaisiais	$Y=\beta_0+\beta_1 X_1+\beta_2 D_1+\beta_3 D_1 X_1$	Y C X_1 D_1 D_1* X_1
Skirtuminė lygtis	$Y=\beta_0+\beta_1 (X_t-X_{t-1})$	Y C d(X)
Logaritminė skirtuminė lygtis	$Y=\beta_0+\beta_1 (\ln X_t-\ln X_{t-1})$	Y C dlog(X)
Vieno periodo procentinis pokytis	$Y=\beta_0+\beta_1 [(X_t-X_{t-1})/X_t]$	Y C pch(X)

Apibendrinimas: KTRM prielaidų pažeidimo pasekmės įverčiams ir dispersijoms

Turime regresiją[5] $Y=X\beta+u$, kurios MKM parametro įvertis $\hat{\beta}=(X'X)^{-1}X'Y$, jo dispersijos įvertis $Var(\hat{\beta})=\sigma^2(X'X)^{-1}$ bei imties dispersija $Var(X)=\frac{1}{n}\sum_i (X_i-\bar{X})^2$. Pastaroji yra paslinktas populiacijos dispersijos įvertinys, tuo tarpu nepaslinktas σ^2 įvertis yra $s^2=\frac{1}{n-1}\sum_i (X_i-\bar{X})^2$. Intuityvus paslinktumo paaiškinimas – imties dispersijos įvertis yra skaičiuojamas naudojant imties vidurkį $\bar{X}$, o ne tikrąjį populiacijos vidurkį μ. Įrodymą rasite *Prieduose*.

[5] Čia stebėjimų matricos *X* pirmasis stulpelis yra sudarytas iš vienetų, taigi modelis įtraukia ir laisvąjį narį.

Aptarėme klasikinės tiesinės regresijos modelio (KTRM) prielaidų pažeidimo atvejus. Reziumuojame lentele, kuri pagelbės atliekant praktinius ekonometrinius darbus, primins prielaidų svarbą bei poveikį įverčiams bei dispersijoms. Tiek įverčių nepaslinktumas, tiek ir jų matavimo tikslumas yra reikšmingi modelio kokybės kriterijai. Žinoma, svarbus ir testavimas, kuris remiasi įvertintomis standartinėmis paklaidomis.

Pažeidimas	Poveikis β_i (ir dispersijų įverčiams)
Multikolinearumas (t.y. stipri koreliacija tarp stebėjimų X_i)	Nepaslinktas (įvertintos dispersijos s^2 yra paslinktos bei padidintos)
Heteroskedastiškumas (t.y. $\sigma_i^2 \neq \sigma^2$ arba nepastovi dispersija)	Nepaslinktas (įvertintos dispersijos s^2 yra paslinktos bei padidintos)
Autokoreliacija (serijinė koreliacija, $\sigma_{u_i,u_j} \neq 0$ visiems $i \neq j$)	Nepaslinktas (įvertintos dispersijos s^2 yra paslinktos)
Praleisti reikšmingi regresoriai	Paslinktas: poslinkio „kryptis“ priklauso nuo $Cov(X,u)$ ženklo
Matavimo klaidos X_i	Paslinktas (link nulio, t.y. $\beta_i \to 0$)
Matavimo klaidos Y	Nepaslinktas (įvertintos dispersijos paslinktos)
X_i endogeniškumas (ypač akivaizdi problema simultaninių lygčių atveju)	Paslinktas: gali apskritai pakeisti ženklą

Kita ekonometrijos sritis domisi dinamine ekonometrija bei laiko eilutėmis – modeliais, kurie turi reikšmingų skirtumų, lyginant su kryžminių duomenų analize.

Dinaminiai ekonometrijos modeliai

Šio skyriaus tikslas išmokti modeliuoti ekonominius kintamuosius, išdėstytus laike, pastarojo veiksnio įtaką žymint sutartiniu simboliu *t*. Taigi kintamojo X_t seką, išdėstytą pagal laiko indeksą *t*, mes ir vadinsime *laiko eilute*. Pastaroji gali būti begalinė (teorinėje analizėje), visgi modeliuojant ekonominius reiškinius, duomenys ribos tiek sekos pradžią, tiek ir jos pabaigą.

Vėluojančio poveikio modeliai (paskirstyto lago modeliai)

Analizei pasirinkome autoregresinį paskirstyto vėlavimo modelį (angl. Autoregressive Distributed Lag (ADL) model), aprašomą šia lygtimi: $Y_t = \beta_0 + \sum_i \beta_i X_{t-i} + \sum_j \theta_j Y_{t-j} + u_t$. Nagrinėkime specialųjį atvejį: $Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \theta_1 Y_{t-1} + u_t$

Modelio išraiška atskleidžia pavadinimo kilmę:

- $\beta_1 X_t + \beta_2 X_{t-1}$ yra paskirstytas vėlavimas (lagas) egzogeninio kintamojo X_t laiko eilutėje;
- $\theta_1 Y_{t-1}$ yra autoregresinis Y_t narys, t.y. endogeninis kintamasis su lagu.

Tiek X_t, tiek ir Y_t didžiausias yra pirmasis vėlavimas (lagas), taigi modelis trumpinamas kaip ADL(1, 1).

Tolesnei analizei pravers lago operatorius *L*. Pastarasis yra tiesinė funkcija, pasižyminti šiomis savybėmis:

- $LX_t = X_{t-1}$
- $L\alpha X_t = \alpha X_{t-1} = \alpha L X_t$
- $L(X_t + Y_t) = X_{t-1} + Y_{t-1} = LX_t + LY_t$
- $L^2 X_t = LLX_t = L X_{t-1} = X_{t-2}$

Taigi pradžioje užrašytas modelis, gali būti perrašytas, naudojant lago operatorių: $Y_t = \beta_0 + \beta_1 X_t + \beta_2 L X_t + \theta_1 L Y_t + u_t = \beta_0 + (\beta_1 + \beta_2 L) X_t + \theta_1 L Y_t + u_t$ arba dar trumpiau: $(1-\theta_1 L) Y_t = \beta_0 + (\beta_1 + \beta_2 L) X_t + u_t$. Ši išraiška labai patogi nustatyti, kokį ilgojo laikotarpio poveikį X_t turi Y_t. Preziumuojama, kad ilgojo laikotarpio pusiausvyra yra tokia, kad $X_t = X_{t-1} = LX_t$ bei $Y_t = Y_{t-1} = LY_t$. Tam, jog šios lygybės būtų patenkintos, reikia lago operatorių prilyginti vienetui: $L=1$. Tada lygtis transformuojama į $Y_t = \beta_0 (1-\theta_1)^{-1} + (\beta_1 + \beta_2)(1-\theta_1)^{-1} X_t + w_t$ Taigi ilgojo laikotarpio efektas yra

$$\frac{\partial Y}{\partial X} = \frac{\beta_1 + \beta_2}{1-\theta_1}.$$

Ilgojo laikotarpio arba pusiausviros X_t ir Y_t reikšmės vaidina svarbų vaidmenį adaptyvių (supaprastintų) lūkesčių modeliuose. Lūkesčių modeliavimas vaidina reikšmingą vaidmenį tokių kintamųjų kaip investicijos, taupymas ar paklausa turtui analizėje. Adaptyvių lūkesčių procesas gelbsti sprendžiant lūkesčių modeliavimo problemą. Tiesa, šis modelis gerokai supaprastintas ir nevisada atspindi realią modeliuojamo kintamojo dinamiką.

Idėja: elementarus „mokymosi“ procesas, kurio metu faktinė kintamojo reikšmė yra lyginama su lauktina (tikėtina) reikšme. Jei faktinė reikšmė didesnė už lauktiną, kito periodo tikėtina reikšmė padidinama, ir atvirkščiai. Preziumuojama, jog šio pokyčio dydis yra proporcingas nesutapimui tarp faktinės ir lauktinos reikšmės. Tegu X_t yra modeliuojamas kintamasis, X_t^e yra lauktina reikšmė, esant t–1 laikotarpio informacijai; kitaip tariant, $X_t^e = X_t | \Omega_{t-1}$, kur Ω_{t-1} yra informacinė aibė laiko momentu t–1. Tada, anot adaptyvių lūkesčių: $X_{t+1}^e - X_t^e = \lambda(X_t - X_t^e)$ arba $X_{t+1}^e = \lambda X_t + (1-\lambda) X_t^e$, kur $0 \leq \lambda \leq 1$.

○ **Pavyzdys**

Modeliuokime tokį kintamąjį Y_t, kuris priklauso nuo tikėtinos kitų metų regresoriaus reikšmės X_{t+1}^e: $Y_t = \beta_1 + \beta_2 X_{t+1}^e + u_t$. Suprantama, X_{t+1}^e yra neišmatuojamas ir turi būti pakeistas tokiomis reikšmėmis, kurios prieinamos laiko momentu t bei atspindi X_{t+1}^e. Čia pravers adpatyvūs lūkesčiai. Laikome, jog $X_{t+1}^e = \lambda X_t + (1-\lambda) X_t^e$, tada įstatę gauname $Y_t = \beta_1 + \beta_2 \lambda X_t + \beta_2 (1-\lambda) X_t^e + u_t$. Vėlgi problema – X_t^e neišmatuojamas, bet galime taikyti adaptyvių lūkesčių taisyklę ir šiam kintamajam (kitaip tariant, jeigu specifikacija yra teisinga, tai ji galioja ir laiko momentui t, ir t–1). Tada gauname $Y_t = \beta_1 + \beta_2 \lambda X_t + \beta_2 \lambda (1-\lambda) X_{t-1} + \beta_2 (1-\lambda)^2 X_{t-1}^e + u_t$. Šitaip tęsdami k kartų, gausime:

$$Y_t = \beta_1 + \beta_2 \lambda X_t + \beta_2 \lambda (1-\lambda) X_{t-1} + \beta_2 \lambda (1-\lambda)^2 X_{t-2} + \ldots + \beta_2 \lambda (1-\lambda)^{k-1} X_{t-k+1} + \beta_2 (1-\lambda)^k X_{t-k+1}^e + u_t .$$

Kadangi svoriai $0 \leq \lambda \leq 1$, tai augant k, paskutinis narys artėja prie nulio, nes $\lim_{k \to \infty} (1-\lambda)^k = 0$. Tokia vėlavimo struktūra, kuomet svoriai geometriškai mažėja, yra apibūdinama *Koyck'o skirstiniu*.

○ **Perspėjimas!** Norėdami įvertinti aukščiau išvestos regresijos parametrus, turėsite naudoti netiesinius įvertinimo metodus. MKM netiks dėl kelių priežasčių:

ypač akivaizdi multikolinearumo problema ir iš to kylantis įverčių nepastovumas; taipogi įverčių nesuderinamumas su teoriniu modeliu, nes, įvertinant parametrus MKM būdu, apribojimai nėra integruoti į skaičiavimo procedūrą. Alternatyva – Koyck'o transformacija, aprašyta apačioje, kuri, esant patenkintoms tam tikroms sąlygoms, suderinama su MKM.

Išdėstę reikalingą teorijos dalį, galime parodyti, kodėl pusiausviros X_t ir Y_t reikšmės yra svarbios adaptyvių lūkesčių modeliuose. Iš paskutinės lygties, t.y.

$$Y_t=\beta_1+\beta_2\lambda X_t+\beta_2\lambda(1-\lambda)X_{t-1}+\beta_2\lambda(1-\lambda)^2X_{t-2}+\ldots+\beta_2\lambda(1-\lambda)^{k-1}X_{t-k+1}+\beta_2(1-\lambda)^kX^e_{t-k+1}+u_t$$

matome, jog trumpalaikis efektas yra $\frac{\partial Y_t}{\partial X_t}=\beta_2\lambda$. Ilgojo laikotarpio arba pusiausviro X_t poveikis pusiausviram Y_t yra surandamas iš[6]

$$\bar{Y}_t=\beta_1+\beta_2\lambda\bar{X}+\beta_2\lambda(1-\lambda)\bar{X}+\beta_2\lambda(1-\lambda)^2\bar{X}+\ldots=\beta_1+\beta_2\bar{X}[\lambda+\lambda(1-\lambda)+\lambda(1-\lambda)^2\ldots]=\beta_1+\beta_2\bar{X}$$

ir yra lygus $\frac{\partial\bar{Y}}{\partial\bar{X}}=\beta_2$.

Adaptyvių lūkesčių modeliai gali būti nagrinėjami ir vadinamosios *Koyck'o transformacijos* pagalba: priklausomas kintamasis išreiškiamas baigtiniu skaičiumi nepriklausomų kintamųjų, kurie yra išmatuojami. Prisiminkime nagrinėjamą modelį: $Y_t=\beta_1+\beta_2X^e_{t+1}+u_t$, kur $X^e_{t+1}=\lambda X_t+(1-\lambda)X^e_t$. Tada teisinga išraiška būtų ir ši: $Y_{t-1}=\beta_1+\beta_2X^e_t+u_{t-1}$, iš kurios išreiškiame: $\beta_2X^e_t=Y_{t-1}-\beta_1-u_{t-1}$. Įstatome į anksčiau gautą išraišką: $Y_t=\beta_1+\beta_2\lambda X_t+\beta_2(1-\lambda)X^e_t+u_t=\beta_1\lambda+(1-\lambda)Y_{t-1}+\beta_2\lambda X_t-(1-\lambda)u_{t-1}+u_t$. Taigi matome, kad modelis gali būti transformuotas į $Y_t=\tilde{\beta}_1+\tilde{\beta}_2X_t+\tilde{\beta}_3Y_{t-1}+\xi_t$, kur $\tilde{\beta}_1=\beta_1\lambda$, $\tilde{\beta}_2=\beta_2\lambda$, $\tilde{\beta}_3=1-\lambda$ ir $\xi_t=u_t-(1-\lambda)u_{t-1}$. Trumpalaikis ir ilgalaikis efektai nepakito palyginant Koyck'o transformaciją su pirmine analize, kitaip tariant, $\frac{\partial Y_t}{\partial X_t}=\beta_2\lambda$ ir $\frac{\partial\bar{Y}}{\partial\bar{X}}=\beta_2$.

○ **Pastebėjimas!** MKM metodas dabar pritaikomas, nes tereikia įvertinti du koeficientus. Tiesa, problema yra endogeninis kintamasis su vėlavimu, kuris

[6] Nesunku įsitikinti: $\lambda+\lambda(1-\lambda)+\lambda(1-\lambda)^2\ldots=C$. Tada $(1-\lambda)\lambda+\lambda(1-\lambda)^2+\lambda(1-\lambda)^3\ldots=(1-\lambda)C$, o atėmę vieną iš kito gausime $\lambda=C-(1-\lambda)C\Rightarrow\lambda=\lambda C\Rightarrow C=1$.

įtrauktas kaip regresorius. Rezultatas: paslinkti įverčiai. Jeigu paklaida u_t buvo baltasis triukšmas, tai ξ_t pasižymi pirmos eilės autokoreliacija. Drauge su endogeniniu pavėluotu kintamuoju, tai sukelia dar didesnę problemą – įverčiai būtų ne tik paslinkti, bet ir nesuderinti, t.y. išliktų paslinkti imties dydžiui augant iki begalybės. MKM naudojimas leistinas, jeigu ξ_t yra baltasis triukšmas (tokiu atveju, u_t nėra baltasis triukšmas). Šią prielaidą reikia tikrinti (EViews tai galima padaryti *Ljung–Box* ir *Box–Pierce* testų pagalba).

Norintieji išmatuoti polinominį paskirstyto lago modelį $Y_t = \beta_0 + \sum_i \beta_i X_{t-i} + u_t$ be autoregresinio nario, gali pasinaudoti *EViews* integruota funkcija PDL (angl. *Polynomial Distributed Lag*). Tokiam modeliui bus panaudota Almon procedūra, transformuojanti modelį taip, kad multikolinearumo problema būtų sumažinta, reikalingų matuoti parametrų skaičius būtų nedidelis, o MKM būtų geriausi ir nepaslinkti. Formulės lauke įrašykite **Y C PDL(*X*, *q*, *r*)**, kur ***q*** yra vėlavimų skaičius modelyje, o ***r*** yra parametrų polinomo laipsnis.

Laiko eilučių modeliai[7]

Prieš pradėdami laiko eilučių modeliavimą, prisiminkime ir sugriežtinkime apibrėžimus. Sakėme, jog laiko eilutė – tai seka, išdėstytą pagal laiko indeksą *t*. Taikomosios ekonometrijos kontekste dažnai pasitaiko labiau „praktinis“ apibrėžimas: laiko eilutė – tai periodiškų stebėjimų, fiksuotų tam tikrais laiko momentais, visuma. Išties, modeliuodami susiduriame su viena realizacija, bet teoriškai X_t yra atsitiktinis kintamasis, galintis generuoti daugelį realizacijų bei pasiskirstęs pagal tam tikrą skirstinį. Šio skirstinio momentai yra apibrėžiami taip:

Matematinė viltis (vidurkis)	$\mu_t = E(X_t), t=1,\dots,T$
Dispersija	$\gamma_t(0) = Var(X_t) = E[(X_t - \mu_t)^2], t=1,\dots,T$
Kovariacija	$\gamma_t(\tau) = Var(X_t, X_{t-\tau}) = E[(X_t - \mu_t)(X_{t-\tau} - \mu_{t-\tau})], t=1,\dots,T$

Bet, kaip minėjome, turime tik vieną duomenų realizaciją. Norėdami užtikrinti, jog negausime klaidingų regresijų, reikalaujame, jog aprašyti

[7] Gide neaptariamas klasikinis laiko eilučių išskaidymasb bei sezoniškumas – šias temas galima rasti paskaitų medžiagoje bei statistikos knygose.

momentai nekistų laike. Kitaip tariant, norime, jog laiko eilutė pasižymėtų *silpnuoju* (arba *kovariacijos*, arba *antros eilės*) *stacionarumu*. Taigi reikalavimai momentams apibrėžiami taip:

Matematinė viltis (vidurkis)	$\mu = E(X_t)$
Dispersija	$\gamma(0) = Var(X_t) = \sigma^2 = E[(X_t - \mu)^2] < \infty$
Kovariacija	$\gamma(\tau) = Var(X_t, X_{t-\tau}) = E[(X_t - \mu)(X_{t-\tau} - \mu)] < \infty$

Žinoma, egzistuoja ir *stipriojo stacionarumo* sąvoka. Pastaroji ypač ribojanti, mat reikalauja *visų* atsitiktinio kintamojo X_t skirstinio momentų nekintamumo.

○ **Pastaba!** Standartinė prielaida apie X_t normalųjį pasiskirstymą (tuomet X_t yra vadinamas Gauso procesu) garantuoja stiprųjį stacionarumą, jeigu užtikrintas silpnasis stacionarumas. Taip yra todėl, kad pirmųjų dviejų momentų, vidurkio μ ir dispersijos σ^2, pakanka tiek ir silpnajam stacionarumui užtikrinti, tiek ir normaliojo skirstinio specifikacijai. Prisiminkite, jog normaliai pasiskirsčiusio X_t tikimybinė tankio funkcija aprašoma taip: $f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}$.

Stacionarumo vaidmuo

Jau užsiminėme, kuomet apibrėžėme stacionarumą, jog norime išvengti klaidingų regresijų. Jų prigimtis ir slypi nestacionarių laiko eilučių regresijose, kuomet nustatomas statistiškai reikšmingas ryšys tarp visiškai nesusijusių nestacionarių kintamųjų. Kitaip tariant, regresija tarp nepriklausomų nestacionarių kintamųjų, indikuojanti stiprų statistinį ryšį naudojant *t*, *F* statistikas bei pasižyminti aukšta determinacijos koeficiento R^2 reikšme, ir yra vadinama *klaidinga regresija*.

○ **Pastaba!** Klaidingos regresijos išvada išlieka galioti ir asimptotiškai, t.y. – statistinis ryšys išlieka reikšmingas tarp nesusijusių kintamųjų, taigi net ir labai didelės imtys šios problemos nepadeda išspręsti (Yule, 1926).

Ekonomistai dažniausiai taiko skirtumines transformacijas tam, kad paverstų nestacionarų kintamąjį stacionariu, kitaip tariant, integruoja. Stacionarus kintamasis yra integruotas nuline eile (žymime, $X_t \sim$ I(0)). Dažniausiai ekonominiai kintamieji yra integruoti pirma eile, $X_t \sim$ I(1).

○ Pavyzdys

Tegu turime pirmos eilės autoregresinį modelį (AR(1), plačiau rasite kitame skyrelyje):

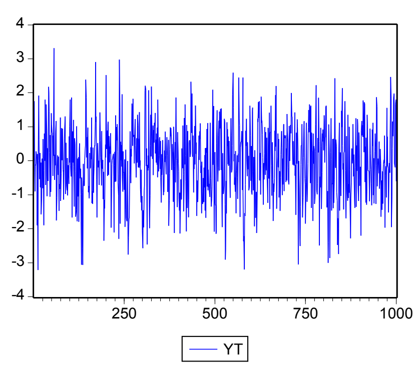
$x_t = \rho x_{t-1} + \varepsilon_t$, kur ε_t yra normaliai pasiskirstęs atsitiktinis kintamasis su nuliniu vidurkiu ir dispersija σ^2 bei kovariacija $Cov(\varepsilon_s, \varepsilon_t) = 0 \; \forall s \neq t$. Kitais žodžiais tariant, ε_t yra baltasis triukšmas.

Nagrinėkime du atvejus:

- Tegu $-1 < \rho < 1$, tada AR(1) procesas yra stacionarus. Taip yra todėl, kad, išreiškę x_t naudodami vien ε_t, gauname: $x_t = \varepsilon_t + \rho\varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} \ldots$ Sutraukime: $x_t = \sum_j \rho_j \varepsilon_{t-j}$. Iš pastarosios išraiškos nesunku įsitikinti, jog $E[x_t] = E[\sum_j \rho_j \varepsilon_{t-j}] = 0$, nes $E[\varepsilon_{t-j}] = 0 \; \forall j$. Dispersija: $Var[x_t] = Var[\sum_j \rho_j \varepsilon_{t-j}] = \sigma^2 + \rho^2\sigma^2 \ldots$ Taip yra, nes $Var[\rho^j \varepsilon_{t-j}] = \rho^{2j}\sigma^2$. Sutraukime: $Var[\rho^j \varepsilon_{t-j}] = \dfrac{\sigma^2}{1-\rho^2}$, nes $1 + \rho^2 + \rho^4 \ldots = \dfrac{1}{(1-\rho^2)}$. Remiantis panašiais argumentais, gauname $Cov[x_t, x_s] = \rho^{t-s}\dfrac{\sigma^2}{1-\rho^2}$. Reziumė: x_t yra stacionarus procesas, nes pasižymi pastoviu vidurkiu (=0), pastovia dispersija $(\dfrac{\sigma^2}{1-\rho^2})$ bei kovariacija $(\rho^{t-s}\dfrac{\sigma^2}{1-\rho^2})$. Pastarieji nepriklauso nei nuo *t*, nei nuo *s*; kovariacija priklauso nuo laiko skirtumo *t–s*.
- *EViews* sugeneruokime stacionarų procesą (pradėkite nuo **New/Workfile/Unstructured/Undated ir 1000 observations/OK**):

$y_t = 0{,}3 * y_{t-1} + e_t \; ; \; e_t \sim (0, \sigma_e^2)$

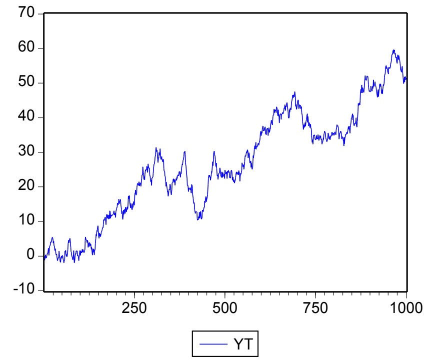
```
smpl 1 1
genr yt=0
smpl 2 1000
genr yt=0.3*yt(-1)+nrnd
smpl 1 1000
plot yt
```



- Tegu $\rho = 1$, tada $Var[x_t] \to \infty$, taigi procesas nestacionarus. Beje, šis procesas $x_t = x_{t-1} + \varepsilon_t$ yra vadinamas *atsitiktiniu klaidžiojimu* ir yra *vienetinės šaknies proceso* pavyzdys.
- *EViews* sugeneruokime vienetinės šaknies procesą:

  $y_t = y_{t-1} + e_t$; $e_t \sim (0, \sigma_e^2)$

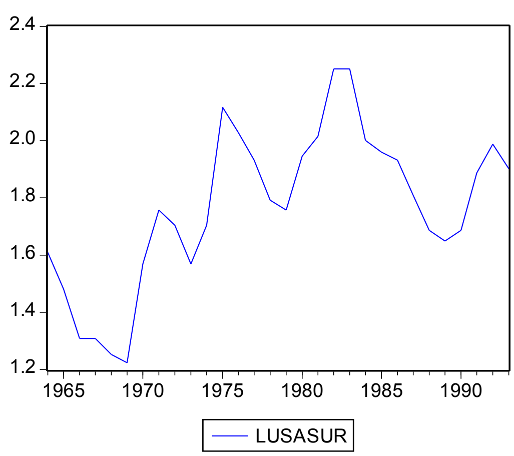
```
smpl 1 1
genr yt=0
smpl 2 1000
genr yt=yt(-1)+nrnd
smpl 1 1000
plot yt
```

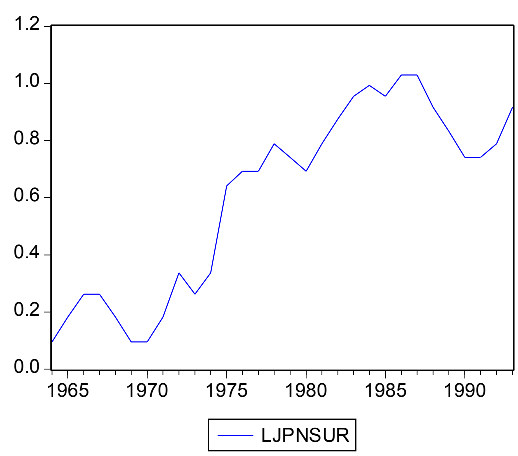


○ **Pastaba!** Vienetinės šaknies šaknies procesai ir yra nestacionarūs integruoti procesai, kurie stacionariais paverčiami taikant skirtumines transformacijas.

○ Pavyzdys

Nagrinėkime ekonominį pavyzdį. Turime OECD šalių duomenis apie standartizuotą nedarbo lygį, žymimą šalies trumpiniu bei pabaiga SUR (angl. standardised unemployment rate). Logaritmuoti atitikmenys žymimi L prieš pavadinimą. Duomenys yra istoriniai (1964 – 1993) bei metiniai. Pradėkime nuo grafinės analizės[8] – paveiksle matote JAV bei Japonijos nedarbo lygio kitimo tendenciją minėtu laikotarpiu.





[8] Plačiau apie grafinę analizę *EViews* aplinkoje rasite pirmuosiuose skyriuose.

Stacionarumu pavaizduotos eilutės nepasižymi. Teiginiui patvirtinti, nubrėžkime kolerogramą.

○ **Apibrėžimas:** Korelograma – tai autokoreliacijos funkcijos, t.y. autokoreliacijos koeficiento ρ_k kaip funkcijos nuo *k*, grafikas.

○ **Apibrėžimas:** *Autokoreliacijos koeficientu* vadinamas koreliacijos koeficientas, kuriuo matuojama vienos laiko eilutės skirtingų stebėjimų tiesinė priklausomybė.

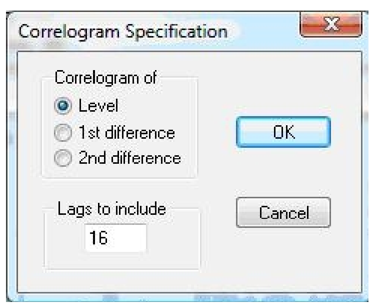
Matematiškai: $\rho(Y_t, Y_{t-k}) = \rho_k = E[(\frac{X_t - E(X_t)}{\sigma_{X_t}})(\frac{X_{t-k} - E(X_{t-k})}{\sigma_{X_{t-k}}})]$,

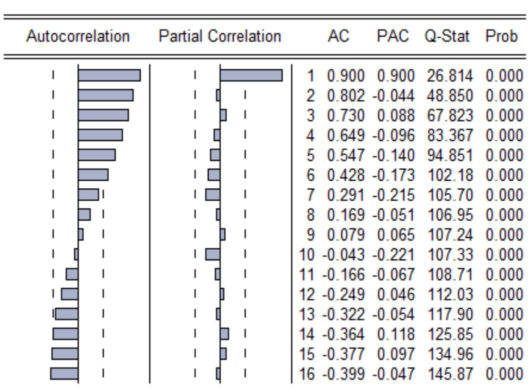
na, o esant stacionariam procesui, dispersija nepriklauso nuo laiko, todėl

$\sigma_{X_t} = \sigma_{X_{t-k}} = \gamma_0$. Gauname $\rho(Y_t, Y_{t-k}) = \rho_k = \frac{\gamma_k}{\gamma_0}$, kur γ_k yra *k*-ojo vėlavimo

kovariacija. Pavyzdys: $\rho(Y_t, Y_t) = \frac{\gamma_0}{\gamma_0} = 1$.

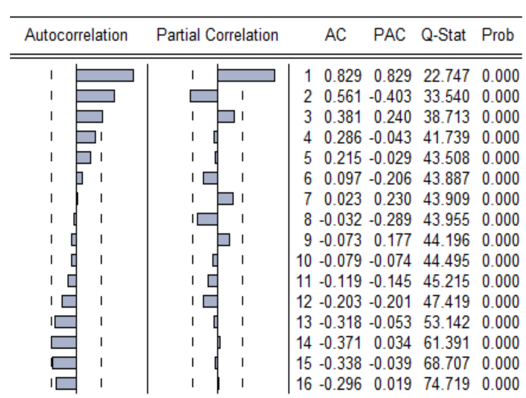
Išsiaiškinę reikalingus terminus, galime EViews aplinkoje nubrėžti, o po to interpretuoti korelogramą (**View/Correlogram/Level bei parinktas vėlavimų skaičius**):





Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.900	0.900	26.814	0.000
		2	0.802	-0.044	48.850	0.000
		3	0.730	0.088	67.823	0.000
		4	0.649	-0.096	83.367	0.000
		5	0.547	-0.140	94.851	0.000
		6	0.428	-0.173	102.18	0.000
		7	0.291	-0.215	105.70	0.000
		8	0.169	-0.051	106.95	0.000
		9	0.079	0.065	107.24	0.000
		10	-0.043	-0.221	107.33	0.000
		11	-0.166	-0.067	108.71	0.000
		12	-0.249	0.046	112.03	0.000
		13	-0.322	-0.054	117.90	0.000
		14	-0.364	0.118	125.85	0.000
		15	-0.377	0.097	134.96	0.000
		16	-0.399	-0.047	145.87	0.000

Japonija



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.829	0.829	22.747	0.000
		2	0.561	-0.403	33.540	0.000
		3	0.381	0.240	38.713	0.000
		4	0.286	-0.043	41.739	0.000
		5	0.215	-0.029	43.508	0.000
		6	0.097	-0.206	43.887	0.000
		7	0.023	0.230	43.909	0.000
		8	-0.032	-0.289	43.955	0.000
		9	-0.073	0.177	44.196	0.000
		10	-0.079	-0.074	44.495	0.000
		11	-0.119	-0.145	45.215	0.000
		12	-0.203	-0.201	47.419	0.000
		13	-0.318	-0.053	53.142	0.000
		14	-0.371	0.034	61.391	0.000
		15	-0.338	-0.039	68.707	0.000
		16	-0.296	0.019	74.719	0.000

JAV

Paveiksluose pateikta informacija: autokoreliacijos funkcija, dalinės koreliacijos funkcija, Q-statistika ir *p* lygmuo.

○ **Apibrėžimas:** *Dalinės koreliacijos koeficientas* – tai koreliacijos koeficientas, kuriuo matuojama vienos laiko eilutės skirtingų stebėjimų tiesinė priklausomybė eliminavus tarpinių narių poveikį.

○ **Apibrėžimas:** *Dalinės koreliacijos funkcija* – tai funkcija, parodanti atsitiktinio proceso dalinės autokoreliacijos koeficiento priklausomybę nuo vėlavimo *k* tarp Y_t ir Y_{t-k}.

Apibrėžę reikiamas sąvokas, grįžkime prie paveikslo. Punktyrinės vertikalios linijos žymi dviejų standartinių paklaidų atstumą nuo koreliacijos koeficientų. Kadangi visų *Q*-statistikų *p* reikšmės yra lygios nuliui, procesas akivaizdžiai nėra baltasis triukšmas. Dar daugiau, autokoreliacijos išlieka reikšmingos net ir vėlesniuose laguose (nėra statistiškai lygūs nuliui), be to, pirmos eilės, o JAV atveju, pirmųjų dviejų eilių dalinės autokoreliacijos yra statistiškai reikšmingos. Lėtas autokoreliacijų reikšmingumo gesimas indikuoja apie proceso nestacionarumą. Tad prieš imantis analizės, reikės stacionarizuoti eilutę, o tuomet bandyti išsiaiškinti, su kokiu procesu susiduriame. Tai gali būti AR, MA arba mišrieji ARMA procesai.

○ **Interpretacija:** jeigu AC gęsta (daugmaž) geometriškai, panašu, jog susiduriame su AR procesu; jeigu AC pasieka nulį vos po kelių vėlavimų – tai indikacija, jog nagrinėjamas MA procesas. PAC, visiškai dingstantis po *p* vėlavimo, indikuoja apie grynąjį AR(*p*) procesą, o MA proceso PAC asimptotiškai, be staigių nutrūkimų, artėja prie nulio. Taigi AC yra tinkamas MA, o PAC naudotinas AR eilės nustatymui. Vėliau bus aptarti ARMA procesai, kurie turi tiek AR, tiek MA būdingų savybių derinį.

AR, MA ir ARIMA modeliai

Apibendrinkime ir aprašykime *p*-os eilės AR procesą AR(*p*): $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t$. Taigi einamojo laikotarpio Y_t yra modeliuojamas kaip svertinis Y_t su vėlavimais iki *p*-ojo periodo vidurkis plius atsitiktinė einamojo periodo paklaida ε_t. Stacionarumas: su AR procesu susijusio polinomo $1-\phi_1 x-\phi_2 x^2-...-\phi_p x^p=0$ šaknys už vienetinio apskritimo ribų. Ekvivalentus reikalavimas: polinomo $x^p-\phi_1 x^{p-1}-\phi_2 x^{p-2}-...-\phi_p=0$ absoliučios šaknų reikšmės turi būti didesnės už vienetą.

MA apibendrintas procesas gali būti užrašytas kaip $Y_t=\varepsilon_t+\theta_1\varepsilon_{t-1}+\theta_2\varepsilon_{t-2}+...+\theta_q\varepsilon_{t-q}$. Suprantame, jog MA procesas yra svertinis stacionarių baltųjų triukšmų vidurkis. Taigi MA yra visada stacionarus. Tuomet, apjungus AR ir MA, gauname ARMA(*p*,*q*) procesą, apibūdinamą $Y_t=\phi_1 Y_{t-1}+\phi_2 Y_{t-2}+...+\phi_p Y_{t-p}=\varepsilon_t+\theta_1\varepsilon_{t-1}+\theta_2\varepsilon_{t-2}+...+\theta_q\varepsilon_{t-q}$. Pasinaudokime vėlavimų operatoriumi *L* ir perrašykime ARMA(*p*,*q*) taip: $\phi(L)Y_t=\theta(L)\varepsilon_t$, čia $\phi(L)=1-\phi_1 L-\phi_2 L^2-...-\phi_p L^p$ ir $\theta(L)=1-\theta_1 L-\theta_2 L^2-...-\theta_p L^p$. Kaip jau minėjome, ARMA proceso stacionarumas priklauso tik nuo AR dalies. Nepamirškite, kad bet kuris MA procesas $Y_t=\varepsilon_t+\theta_1\varepsilon_{t-1}+\theta_2\varepsilon_{t-2}+...+\theta_q\varepsilon_{t-q}$ gali būti išreikštas kaip begalinis AR procesas, jeigu MA yra „apverčiamas“, t.y., jeigu polinomo $1+\theta_1 x+\theta_2 x^2+...+\theta_q x^q$ šaknys yra už vienetinio apskritimo ribų.

Tiesa, retai ekonomikos kintamieji yra stacionarūs. Dėl to susiduriame netiesiogiai su Y_t, o su X_t, kuris modeliuojamas kaip ARIMA(*p*,*d*,*q*) procesas, t.y. integruotas procesas. Kitaip tariant, $Y_t=\Delta^d X_t$. Perrašytas modelis vėlavimų pagalba yra $\phi(L)\Delta^d X_t=\phi(L)(1-L)^d X_t=\theta(L)\varepsilon_t$. Kaip nuo teorinių samprotavimų pereiti prie ARIMA modeliavimo su realiais duomenimis bei jų panaudojimo prognozavimui? Atsakymą turi paruošę statistikai George Box ir Gwilym Jenkins.

Box-Jenkins procedūra

Išbaigta Box-Jenkins procedūra apima šiuos žingsnius:

1. ARMA proceso stacionarumo nustatymas;
2. Užtikrinamas stacionarumas integruojant laiko eilutę;
3. ARMA proceso *p* ir *q* eilės nustatymas;
4. ARMA modelio ir jo alternatyvų vertinimas;

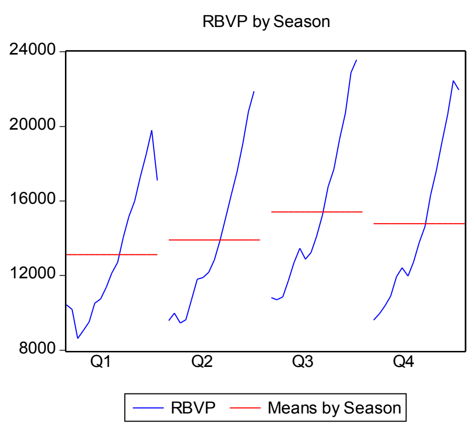
5. Modelio diagnostika.

○ Pavyzdys

Nagrinėkime Lietuvos BVP dinamiką nuo 1993 K1 iki 2009 K1, arba iš viso 65 ketvirčius. Norime sukonstruoti modelį, specifikuojamą ARIMA pagalba, kuriuo remiantis būtų galima atlikti trumpojo laikotarpio prognozes. Duomenis gavome iš Tarptautinio valiutos fondo Tarptautinės finansų statistikos duomenų bazės. Išskaičiavome realųjį BVP turėdami nominaliojo bei BVP defliatoriaus reikšmes. Prisiminkite, jog $RBVP = \dfrac{NBVP}{BVP\ defliatorius} \times 100$, kur RBVP yra realusis, o NBVP nominalusis BVP. Tolesniam tyrimui naudosime Box-Jenkins procedūrą.

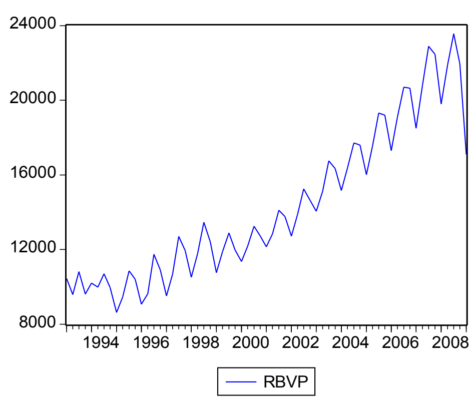
1. *Pradėkime nuo grafinės analizės*

● View/graph/seasonal stacked line



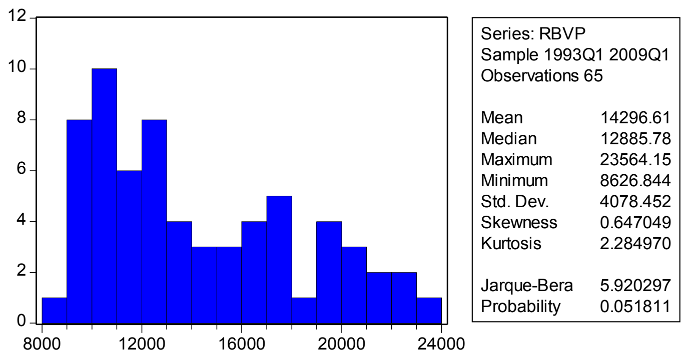
Aptinkame sezoniškumo įrodymų – apie jų įtakos eliminavimą kalbėsime vėliau.

● View/graph/line



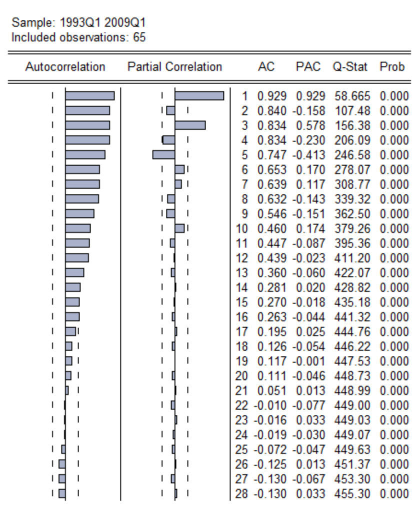
Duomenys turi akivaizdų teigiamą trendą ir nėra stacionarūs. Visgi stacionarumas turėtų būti nagrinėjimas ir remiantis statistiniais testais bei kolerogramos pagalba.

● **View/descriptive statistics/histogram and stats**



Ši lentelė buvo aptarta pirmuose skyriuose. Pastebime, jog duomenys normaliai pasiskirstę su 5% reikšmingumo lygmeniu, o nenormalus pasiskirstymas diagnozuojamas taikant 10% reikšmingumo lygmenį. Antro tipo klaida čia svarbesnė, taigi laikome, jog duomenys nėra gausinio pasiskirstymo.

Kaip minėjome, BVP pasižymi sezoniškumu. Stacionaramui taipogi svarbu eliminuoti sezoniškumo įtaką. *EViews* pastaroji apibrėžiama taip: **genr srbvp=d(rbvp,n,s)**, kur *n* – kintamųjų skirtumų eilė, *s* – sezoniškumo eilė. Kadangi tiriame ketvirtinius duomenis, turėsime **srbvp=d(rbvp,0,4)**. Patikriname gautos eilutės stacionarumą remdamiesi korelograma, **View/correlogram/level**.


Sample: 1993Q1 2009Q1
Included observations: 65

	AC	PAC	Q-Stat	Prob
1	0.929	0.929	58.665	0.000
2	0.840	-0.158	107.48	0.000
3	0.834	0.578	156.38	0.000
4	0.834	-0.230	206.09	0.000
5	0.747	-0.413	246.58	0.000
6	0.653	0.170	278.07	0.000
7	0.639	0.117	308.77	0.000
8	0.632	-0.143	339.32	0.000
9	0.546	-0.151	362.50	0.000
10	0.460	0.174	379.26	0.000
11	0.447	-0.087	395.36	0.000
12	0.439	-0.023	411.20	0.000
13	0.360	-0.060	422.07	0.000
14	0.281	0.020	428.82	0.000
15	0.270	-0.018	435.18	0.000
16	0.263	-0.044	441.32	0.000
17	0.195	0.025	444.76	0.000
18	0.126	-0.054	446.22	0.000
19	0.117	-0.001	447.53	0.000
20	0.111	-0.046	448.73	0.000
21	0.051	0.013	448.99	0.000
22	-0.010	-0.077	449.00	0.000
23	-0.016	0.033	449.03	0.000
24	-0.019	-0.030	449.07	0.000
25	-0.072	-0.047	449.63	0.000
26	-0.125	0.013	451.37	0.000
27	-0.130	-0.067	453.30	0.000
28	-0.130	0.033	455.30	0.000

2. *Stacionarizavimas*[9].

Korelograma indikuoja, kad eilutė, panaikinus sezoniškumo įtaką, tebėra nestacionari. Stacionarizuojame: ieškome tokios transformacijos, kurios rezultatas būtų stacionari laiko eilutė. Pasitelksime vienetinės šaknies testus: **View/unit root test-augmented Dickey Fuller/level, 1st ir 2nd difference**. Tada pirmos ar antros eilės skirtuminė transformacija gaunama **genr srbvp1=d(srbvp)** bei **genr srbvp2=d(srbvp,2)**.

Null Hypothesis: SRBVP has a unit root

Exogenous: Constant

Lag Length: 0 (Automatic based on SIC, MAXLAG=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.335830	0.6074
Test critical values:	1% level	-3.544063	
	5% level	-2.910860	
	10% level	-2.593090	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: D(SRBVP,2) has a unit root

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-8.768488	0.0000

Radome, jog antrasis skirtumas yra pakankamas eilutei stacionarizuoti. Nubrėžkime stacionarios eilutės korelogramą. Kaip matome, Q-statistika reikšminga visuose pasirinktuose laguose, taigi modelio paklaidos autokoreliuotos. Iš AC ir PAC taipogi įtariame, jog procesas yra mišrus.

Sample: 1993Q1 2009Q1
Included observations: 59

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.542	-0.542	18.257	0.000
		2	0.099	-0.277	18.870	0.000
		3	0.003	-0.121	18.871	0.000
		4	-0.002	-0.046	18.871	0.001
		5	0.100	0.141	19.543	0.002
		6	-0.217	-0.117	22.737	0.001
		7	0.193	0.008	25.316	0.001
		8	-0.031	0.091	25.386	0.001
		9	-0.106	-0.077	26.198	0.002
		10	0.129	0.032	27.419	0.002
		11	-0.095	-0.003	28.089	0.003
		12	0.079	0.003	28.572	0.005
		13	-0.084	-0.023	29.120	0.006
		14	0.062	0.025	29.425	0.009
		15	-0.025	-0.042	29.475	0.014

[9] Skaitytojui palikta užduotis išnagrinėti atvejį, kai eilutė yra nestacionari dėl dispersijos. Tokiu atveju reikėtų duomenis transformuoti, paprastai logaritmuojant.

3. *ARMA eilės nustatymas.*

Iš korelogramos parenkame alternatyvas. Pastaroji sufleruoja, jog tai galėtų būti ARMA(2,1). Dėl dalinės koreliacijos antrojo vėlavimo reikšmingumo ties pačia atmetimo riba, vertėtų patikrinti ir ARMA(1,1). Kadangi remiamės imties informacija, o taikome teorinių AC ir PAC įžvalgas, pravartu patikrinti daugiau alternatyvų, pavyzdžiui, ARMA(2,2), ARMA(3,1), MA(1), AR(2).

Alternatyvas įvertiname, pvz., ARMA(2,1) suvesime (**ls srbvp2 c ar(1) ar(2) ma(1)**) bei įsitikiname, jog visos proceso šaknys yra vienetinio apskritimo viduje (**View/ARMA structure/roots-graph**).

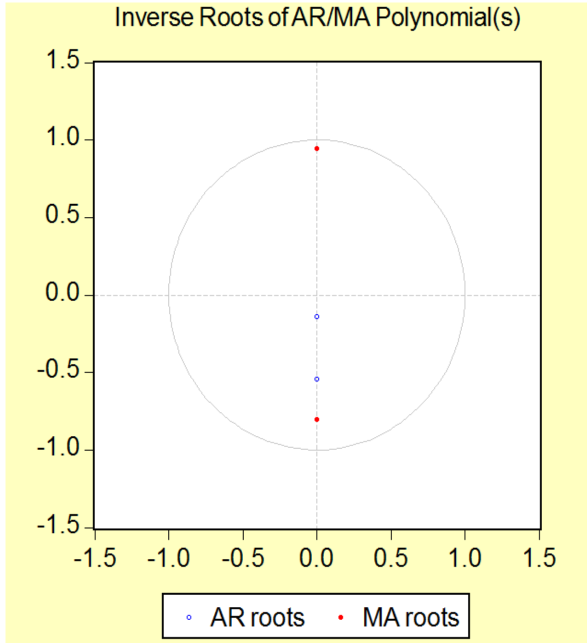
Mūsų atveju, kiek netikėtai, tvariausias modelis buvo ARMA(2,2). Jo koreguotas determinacijos koeficientas buvo didžiausias, o Akaike info ir Schwarz kriterijai mažiausi tarp alternatyvų.

Dependent Variable: SRBVP2
Method: Least Squares
Sample (adjusted): 1995Q1 2009Q1
Included observations: 57 after adjustments
Convergence achieved after 19 iterations
Backcast: 1994Q3 1994Q4

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-9.140883	9.095411	-1.004999	0.3196
AR(1)	-0.682334	0.199344	-3.422888	0.0012
AR(2)	-0.075229	0.147468	-0.510138	0.6121
MA(1)	-0.142152	0.162264	-0.876049	0.3850
MA(2)	-0.760561	0.145167	-5.239227	0.0000

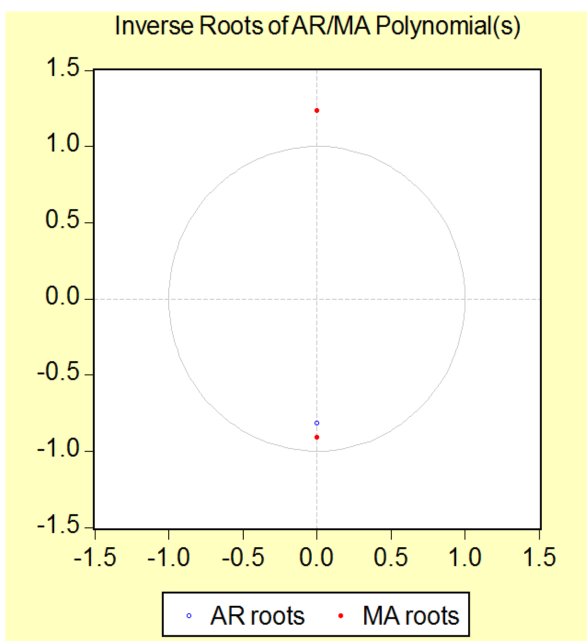
R-squared	0.531128	Mean dependent var	-46.66017
Adjusted R-squared	**0.495061**	S.D. dependent var	751.0742
S.E. of regression	533.7063	Akaike info criterion	**15.48120**
Sum squared resid	14811805	Schwarz criterion	**15.66041**
Log likelihood	-436.2142	F-statistic	14.72612
Durbin-Watson stat	1.663030	Prob(F-statistic)	0.000000

Inverted AR Roots	-.14	-.54
Inverted MA Roots	.95	-.80



Kaip minėta, korelograma tėra imties atspindys, todėl dažnai pasitaiko, jog *a priori* geriausi modeliai, parinkti remiantis teoriniais dydžiais, neatitiks statistiškai tvariausių. Apsistojame ties ARMA(2,2) modeliu.

○ **Pastebėjimas!** Įdomus ARMA (1,2) modelio rezultatas. Pažvelkite į šio modelio šaknis vienetiniame apskritime.



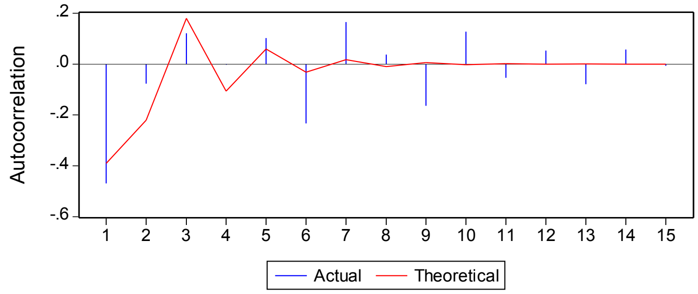
Taigi susiduriame su problema – atvirkštinės MA šaknys yra didesnės už vienetą (arba MA šaknys mažesnės už vienetą) ir visas procesas yra neapverčiamas. Kitaip tariant, MA proceso negalima išreikšti kaip begalinio AR proceso. Taigi modelis yra *stacionarus* (mat atvirkštinės AR šaknys yra vienetinio apskritimo viduje, kitaip, AR šaknys yra už vienetinio apskritimo ribų), bet *neapverčiamas*. Kokios pasekmės? Jeigu mūsų tikslas yra ARMA modelį naudoti prognozėms, tai neapverčiamumas sukels sunkumų, nes negalėsime MA perrašyti naudojantis priklausomojo kintamojo praeities reikšmėmis, o prognozė priklausys nuo ateities reikšmių. Visgi Hamilton (1994) pademonstravo kaip gauti MA modelio reprezentaciją, kad atvirkštinės šaknys būtų vienetinio apskritimo viduje. *EViews* ši problema išsprendžiama išjungiant *MA backcasting* procedūrą (tai *EViews* padaro automatiškai, kai modelis yra neapverčiamas).

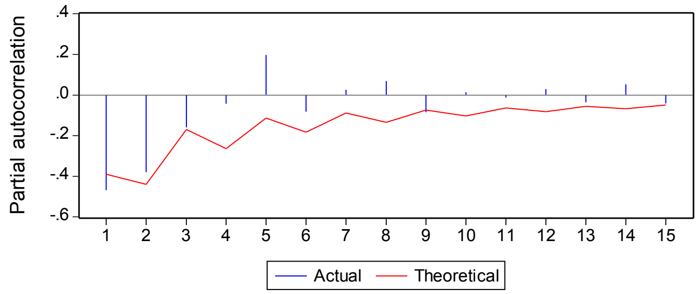
4. *Parinkto modelio tvarumo ir adekvatumo įvertinimas bei likučių analizė*. Remdamiesi Akaike informaciniu, Schwarz bei koreguoto determinacijos koeficiento kriterijais, parenkame geriausią modelį. Tikriname pasirinkto modelio ARMA struktūrą (**View/ARMA structure/correlogram/graph arba table**) bei ištiriame paklaidų elgseną (**View-Residual Tests-Correlogram Q-statistics**).

Korelograma parodys ryšį tarp struktūrinių likučių ir modelio likučių parinktam vėlavimų skaičiui. Gerai specifikuotam modeliui, teorinės ir modelio autokoreliacijos bei dalinės autokoreliacijos turi būti „arti“ viena kitos.

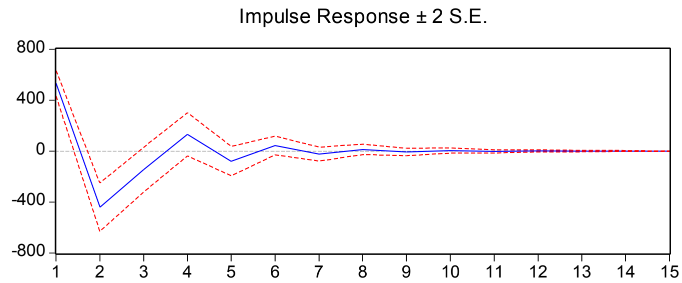
○ **Apibrėžimas:** *Struktūriniai likučiai* – tai likučiai gaunami pašalinus egzogeninių kintamųjų įtaką, bet ne ARMA narius. Mūsų modelyje egzogeninis kintamasis yra laisvasis narys.

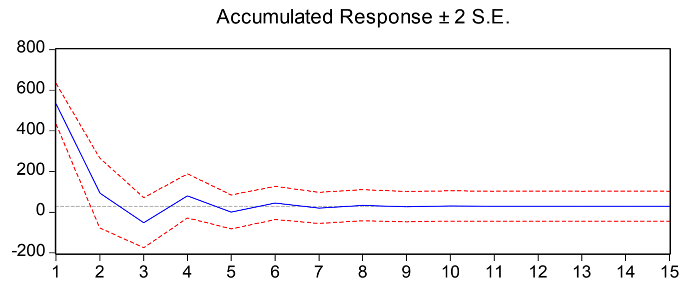
Mūsų atveju ryšys tarp autokoreliacijų išlaikytas, pagrindinės tendencijos tarp išmatuotų bei teorinių autokoreliacijų akivaizdžios.





EViews suteikia galimybę ištirti atsakus į impulsus, pasirinkus **View/ARMA structure/impulse response/vėlavimų skaičius**. Ši opcija tiria ARMA dalies reakciją į vieno laikotarpio inovacinį šoką. Tipinis yra vieno standartinio nuokrypio šokas (kitokį šoką turėtumėte nurodyti patys **Impulse–User specified**). *EViews* atvaizduoja ir kumuliatyvų (suminį) šoką, kuris atspindį atsakų į impulsus sumą. Pastarasis kartojamas kas žingsnį ir visuomet yra tokio paties dydžio (paprastai vieno standartinio nuokrypio). Stacionariam ARMA procesui atsakai į impulsus artės prie nulio, o akumuliuotiems šokams – prie ilgojo laikotarpio (pusiausvirinės) reikšmės. Mūsų ARMA procesas yra stacionarus, nes atsakai artėja prie nulio, o pusiausvirinė reikšmė taipogi nulis.





Galiausiai, ištirkime ar mūsų modelio likučiai nėra autokoreliuoti. Kaip žinia, gerai specifikuoti ARMA modelio likučiai turėtų priminti baltąjį triukšmą. Pasinaudokie Q-statistika: **View/Residual Tests/Correlogram-Q-statistic**. Mūsų modelio ARMA(2,2) atveju gausime:

Date: 09/30/09 Time: 16:36
Sample: 1995Q1 2009Q1
Included observations: 57
Q-statistic probabilities adjusted for 4 ARMA term(s)

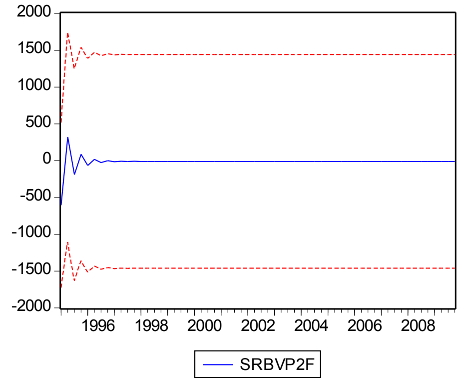
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.027	-0.027	0.0454	
		2	0.050	0.050	0.1999	
		3	0.031	0.034	0.2589	
		4	0.127	0.127	1.2828	
		5	-0.054	-0.051	1.4702	0.225
		6	-0.169	-0.189	3.3509	0.187
		7	0.015	0.000	3.3655	0.339
		8	0.067	0.080	3.6710	0.452
		9	-0.110	-0.084	4.5261	0.476
		10	0.081	0.115	4.9967	0.544
		11	-0.072	-0.088	5.3804	0.614
		12	0.048	-0.012	5.5559	0.697
		13	-0.128	-0.094	6.8007	0.658
		14	-0.005	-0.017	6.8025	0.744
		15	-0.065	-0.065	7.1451	0.787

Kaip matome iš paveikslo nė vieno vėlavimo AC ar PAC yra statistiškai reikšminga. Prisiminkime, jog Q-statistika, kitaip žinoma kaip Ljung-Box postuluoja, jog nėra paklaidų autokoreliacijos. Žiūrime į p reikšmės ir matome, jog negalime atmesti nulinės hipotezės. Taigi likučiai yra visiškai atsitiktiniai, taigi nėra būtinybės ieškoti kito modelio.

5. *Prognozavimas.*

Galop, prognozavimas, kuris gali būti statinis arba dinaminis. Renkamės įvertinto modelio lange **Proc/Forecast**. Nurodykite **Forecast Name** bei **S.E.**

(optional), taip *EViews* sukurs naujas eilutes su prognoze bei standartine jos paklaida. Dinaminio prognozavimo atveju gauname tokį vaizdą:



Forecast: SRBVP2F
Actual: SRBVP2
Forecast sample: 1993Q1 2009Q4
Adjusted sample: 1995Q1 2009Q4
Included observations: 57

Root Mean Squared Error	692.9246
Mean Absolute Error	524.1919
Mean Abs. Percent Error	98.43128
Theil Inequality Coefficient	0.823400
Bias Proportion	0.001980
Variance Proportion	0.880355
Covariance Proportion	0.117665

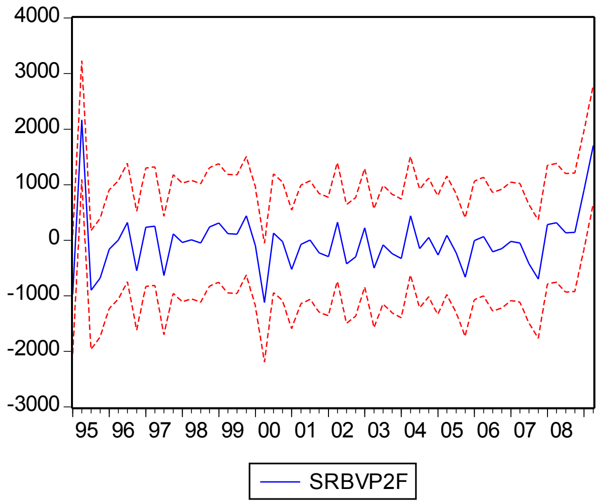
Pastarasis nėra labai įdomus iš ekonominės perspektyvos (dėl to ARIMA metodologija ir yra skirta trumpalaikėms, o ne ilgalaikėms prognozėms). Kaip matome, dinaminio prognozavimo atveju prognozė labai greitai konverguoja į konstantą (mūsų atveju į nulį).

○ **Dėmesio!** Nulinė reikšmė nereiškia, jog prognozuojamas Lietuvos RBVP yra nulis. Nepamirškite, kad čia prognozuojama dukart diferencijuoti ir su pašalintu sezoniškumu duomenys. Tada jie aprašomi kaip $X_t = (1-L^4)(1-L)^2 Y_t$. Taigi prognozė sako, jog vėlavimo operatorius turi būti L=1 ir absoliutinės reikšmės bus lygios[10].

Dinaminis prognozavimas rėmėsi prognozuojamomis priklausomo kintamojo su vėlavimais reikšmėmis, jog numatytume einamojo periodo priklausomojo kintamojo reikšmę. Statinis prognozavimas remiasi imties duomenimis, t.y. naudojamos priklausomo kintamojo su vėlavimais reikšmės, jeigu tokios yra prieinamos. Norėdami išplėsti prognozuojamą laikotarpį už imties laikotarpio, nueikite į darbalaukį, kuriame matote nagrinėjamus kintamuosius, dukart

[10] Jeigu norime būti preciziški, tai turėdami šeštos eilės lygtį, gausime šiuos atsakymus: -1, -*i*, *i*, 1, 1, 1, kur $i=\sqrt{-1}$.

spustelėkite **Range** ir nustatykite pakeitimus. Tuomet grįžkite į **Forecast** ir pasirinkite kokio tipo prognozės pageidaujate.



Forecast: SRBVP2F
Actual: SRBVP2
Forecast sample: 1993Q1 2009Q2
Adjusted sample: 1995Q1 2009Q2
Included observations: 57

Root Mean Squared Error	509.7609
Mean Absolute Error	382.9999
Mean Abs. Percent Error	116.0130
Theil Inequality Coefficient	0.414526
Bias Proportion	0.001187
Variance Proportion	0.270061
Covariance Proportion	0.728752

Raudonos punktyrinės linijos nurodo pasikliautinio intervalo ribas. Šalia pateikta informacija nurodo prognozės tikslumą (RMSE, MAE ir MAPE yra paklaidos, kuo jos mažesnės, tuo prognozė tikslesnė). Tai, jog MAPE yra daugiau negu 100% rodo, kad RBVP prognozavimas yra sudėtingas ir ARMA modelis nepaaiškina labai didelės duomenų, išeinančių už imties ribų, variacijos dalies remdamasis statine prognoze. Vertinga informacija yra prognozės paklaidos išskaidymas į tris komponentes: poslinkio (matuoja, kiek prognozės vidurkis skiriasi nuo imtyje esančių duomenų vidurkio), dispersijos (panašiai matuojami dispersijų skirtumai) ir kovariacijos (įeina visa likusi nesisteminė prognozės paklaidos dalis). Geriausios prognozės yra tos, kurios yra nepaslinktos (poslinkio proporcija yra arti nuliui) bei su kuo mažesne dispersijos proporcija. Kaip matome, mūsų prognozė turi neblogas statistines savybes (nepaslinkta, o didžiausia paklaidos dalis tenka kovariacijos proporcijai), kita vertus prognozuojamas vos vienas ketvirtis, išeinantis už imties ribų.

VAR

Toliau trumpai aptarsime apibendrintą autoregresinį modelį, kitaip žinomą vektorinės autoregresijos (VAR) vardu. Su šia sąvoka glaudžiai susijęs Granger priežastingumas.

Granger priežastingumas

1969 metais ekonometras ir Nobelio premijos laureatas Clive Granger pristatė techninį priežastingumo apibrėžimą: jis tarė, jog jeigu *X* yra *Y* priežastis, tuomet, žinodami *X* praeities reikšmes, galėsime geriau prognozuoti *Y* negu žinodami tik *Y* praeities reikšmes. Kitais žodžiais tariant, jei dabarties *Y* matematinė viltis, esant duotoms *Y* praeities reikšmėms, priklauso nuo *X* praeities reikšmių, yra įrodymas, jog *X* veikia *Y*. Galiojant atvirkštiniam ryšiui, kai *Y* paveikia *X*, turėsime „grįžtamojo ryšio“ atvejį. Atlikus įprastą koeficientų prie *X* testą, ir radus, jog bent vienas iš jų yra nenulinis, sakysime, jog *X* yra Granger *Y* priežastis – svarbu pabrėžti, jog tai nesako, jog *Y* yra *X* rezultatas. Šiuo atveju susiduriame su įprasta statistine įžvalga – koreliacija tarp kintamųjų nesufleruoja priežastingumo.

Testas realizuojamas tiriant vektorinės autoregresijos modelį (VAR):

$$\begin{aligned} y_t &= \alpha_1 + \sum_i \beta_i x_{t-i} + \sum_j \gamma_j y_{t-j} + \varepsilon_{1t} \\ x_t &= \alpha_2 + \sum_i \theta_i x_{t-i} + \sum_j \delta_j y_{t-j} + \varepsilon_{2t} \end{aligned},$$

kur laikome

$$\sigma_{\varepsilon_{1t},\varepsilon_{2t}} = 0 \,\forall t,\ [\varepsilon_{1t} \quad \varepsilon_{2t}]' \sim N(\mathbf{0}, \Sigma_\varepsilon)$$

Kitaip sakant, Granger priežastingumas nepasireikš, jei galios $f(x_t | x_{t-i}, y_{t-i}) = f(x_t | x_{t-i})$, kur $i \in [1, n]$. Tad galime tikėtis keturių rezultatų: x su vėlavimais grupė statistiškai reikšmingai skiriasi nuo nulio, o y – ne, y su vėlavimais grupė statistiškai reikšmingai skiriasi nuo nulio, o x – ne, y ir x su vėlavimais grupės statistiškai reikšmingai skiriasi nuo nulio arba abi statistiškai nesiskiria nuo nulio.

○ Pavyzdys

Nagrinėjame Lietuvos ekonomikos augimui darančių įtaką veiksnių Granger priežastingumą. Turime RBVP metinius augimo duomenis ir išlaidas moksliniams tyrimams ir technologinei plėtrai (MTTP). Norime sužinoti, ar spartesnis MTTP finansavimas gali padėti paaiškinti spartesnį augimą, o galbūt ūkio plėtra teikia daugiau galimybių finansuoti MTTP veiklą. Galų gale, efektas gali būti abipusis ar jokio nebūti.

Pažymime eilutes, kurias naudosime tyrime. Spusteliame dešiniuoju pelės mygtuku **Open/As Group**. Atsidariusiame lange spaudžiame **View/Granger Causality...** ir renkamės, kiek vėlavimų įtraukti į modelį. Mūsų atveju turime duomenis nuo 1995 iki 2008, taigi vos 14 stebėjimų kiekvienoje eilutėje. Išbandykime įvairias vėlavimų skaičiaus kombinacijas.

Lags: 1

Null Hypothesis:	Obs	F-Statistic	Probability
MTTP does not Granger Cause GROWTH	12	0.00382	0.95209
GROWTH does not Granger Cause MTTP		0.08694	0.77479

Lags: 2

Null Hypothesis:	Obs	F-Statistic	Probability
MTTP does not Granger Cause GROWTH	11	0.20768	0.81807
GROWTH does not Granger Cause MTTP		0.24417	0.79077

Lags: 3

Null Hypothesis:	Obs	F-Statistic	Probability
MTTP does not Granger Cause GROWTH	10	0.36186	0.78700
GROWTH does not Granger Cause MTTP		3.90833	0.14620

Taigi negalime atmesti nulinės hipotezės, jog nė vienas kintamasis nėra kito Granger priežastis, išskyrus trijų lagų atvejį, kai parinkus 15% reikšmingumo lygmenį, galėtume atmesti nulinę hipotezę, jog augimas nėra MTTP Granger priežastis. Patikimos išvados negali būti daromos turint tiek nedaug stebėjimų, visgi matome, jog statistinio pagrindo teigti, jog praeities mokslinių tyrimų duomenys paaiškina einamąjį ekonomikos augimą, nėra. Augimo duomenys taipogi mažai tepadeda aiškinti investicijas į mokslą.

Panelinių duomenų modeliai

Duomenys, sudaryti iš kryžminių bei laiko eilučių, pvz., keleto valstybių pasirinkto ekonominio kintamojo stebėjimas tam tikru periodu arba kelių vartotojų pirmenybių dinamika tam tikrame laiko intervale, bus vadinami *paneliniais* (arba tiesiog *panèle*[11]).

Kodėl paneliniai duomenys?

- Daugiau duomenų negu tiriant vien kryžminius ar laiko eilučių modelius. Tai pagerina analizę tuo, jog galime tirti variaciją tiek grupėje (t.y. kaip kinta ekonominis vienetas laike), tiek ir tarp grupių (koks kitimas, lyginant su kitais tiriamais ekonominiais vienetais).
- Galimybė identifikuoti bei atskirti hipotezes, nes galime stebėti tuos pačius ekonominius vienetus kintant laikui. Pavyzdys: vidutinis moterų dalyvavimo darbo rinkoje dydis yra 50%. Tada paaiškinimas gali būti dvejopas: 1) kiekviena moteris turi ½ tikimybės dalyvauti darbo rinkoje; 2) 50% moterų dirba visą darbo laiką, o 50% – apskritai nedirba. Tada hipotezės bus labai skirtingos: (1) atveju darbuotojų kaita labai didelė, (2) kaitos apskritai nėra. Tik paneliniai duomenys leidžia atskirti šiuos atvejus.
- Galimybė kontroliuoti nepastebėtą heterogeniškumą (angl. *unobserved heterogeneity*) individo lygmenyje. Paneliniai duomenys leidžia analizuoti latentinių (nepastebėtų) likučių prigimtį; ši tema labai plačiai nagrinėjama mikroekonometrijos rėmuose.

Bendrasis tiesinis panelinių duomenų modelis

Nagrinėsime bendrąjį panelinį modelį: $y_{it}=\alpha_i+\boldsymbol{\beta}'_{1i}\boldsymbol{x}_{1it}+\boldsymbol{\beta}'_{2i}\boldsymbol{x}_{2i}+\boldsymbol{\beta}'_{3i}\boldsymbol{x}_{3t}+\varepsilon_{it}$, kuriame $i=1,\ldots,N$ bei $t=1,\ldots,T$ Kintamieji turi tokią reikšmę:

[11] Pastaroji skiriasi nuo $pan\tilde{e}l\grave{e}s$.

- $\boldsymbol{x}_{1it}$ yra vektorius, kurio elementai kinta tiek laike, tiek ir tarp individų ar kitų tiriamų subjektų. Pavyzdžiui, firmoms tai būtų pelnai ar pardavimų apimtys, asmenims – pajamos ar vartojimas.
- $\boldsymbol{x}_{2i}$ yra vektorius, kurio elementai skiriasi tarp tiriamų subjektų. Pavyzdžiui, individų charakteristikos, kurios nekinta laike.
- $\boldsymbol{x}_{3t}$ yra vektorius, kurio elementai kinta tik laike. Pavyzdžiui, palūkanų normos, klimato sąlygos mažoje ekonomikoje.

Akivaizdu, jog kuomet $N=1$ ir T didelis, turime laiko eilutę; kai $N>1$, o $T=1$, turime kryžminių duomenų modelį. Panelinis modelis turės $N>1$ ir $T>1$.

Homogeninių nuolydžio parametrų atvejis

Jeigu parametrai bus homogeninai tiek tarp *i*, tiek ir tarp *t*, tai gausime *homogeninių nuolydžio parametrų panelinių duomenų modelį*. Formaliai, $y_{it}=\alpha_i+\boldsymbol{\beta}'\boldsymbol{x}_{it}+\varepsilon_{it}$, kur $Var(\varepsilon_{it})=\sigma_\varepsilon^2$, $\boldsymbol{x}_{it}=[\boldsymbol{x}_{1it}\ \boldsymbol{x}_{2i}\ \boldsymbol{x}_{3t}]'$, $\boldsymbol{\beta}=[\beta_1\ \beta_2\ \beta_3]'$. Dar labiau apribotas modelis bus toks, kuriame $\alpha_i=\alpha$. Tokiu atveju panelinis modelis bus vadinamas *jungtiniu modeliu*, nes jis „sujungtas“ iš atskirų laiko eilučių. MKM metodas, taikomas tokiems modeliams, vadinamas Jungtiniu MKM (JMKM, angl. Pooled Ordinary Least Squares, POLS). Kiekvienas laiko momentas laikomas nepriklausomu informacijos vienetu. Tam, kad įverčiai būtų suderinti, reikia *iid* ε_{it} pasiskirstymo (tam, kad kiekvienas laiko momentas išties suteiktų nepriklausomą informaciją, nepersiklojančią su kitų laiko momentų duomenimis); $Cov(\varepsilon_{it},\boldsymbol{x}_{it})=0$, šis reikalavimas jau matyta kryžminių duomenų modeliuose nepaslinktumo sąlyga. Pakankama MKM nepaslinktumo sąlyga, esant patenkintiems aukčiau aprašytiems reikalavimams, yra $N\to\infty$ arba $T\to\infty$.

○ **Pastaba!** Taikant jungtinį MKM, reikia nepamiršti, jog standartinės paklaidos turi būti koreguotos paneliniams duomenims.

Labai svarbus yra α_i ryšys su likučiais. Kuomet jis buvo pastovus, užteko aptartų sąlygų, jog MKM įverčiai artėtų prie tikrųjų reikšmų, augant imčiai.

Jeigu visgi α_i kinta, modelis gali būti perrašytas: $y_{it}=\boldsymbol{\beta}'\boldsymbol{x}_{it}+v_{it}$, kur $v_{it}=\alpha_i+\varepsilon_{it}$. Dabar jau galime užrašyti JMKM suderintumo reikalavimus: $E(\boldsymbol{x}_{it}, v_{it})=0$, $E(\boldsymbol{x}_{it}, \alpha_i)=0$ ir $E(\boldsymbol{x}_{it}, \varepsilon_{it})=0$.

○ **Pastaba!** Net esant patenkintoms aukščiau aprašytoms sąlygoms, v_{it} pasižymės serijine koreliacija. Kadangi v_{it} priklauso nuo α_i visiems *t*, tai koreliacija tarp v_{it} ir v_{is} nemažės augant $|t-s|$. Jeigu į modelį įtrauktas priklausomas kintamasis su vėlavimu, t.y. y_{it-1}, tai $E(\boldsymbol{x}_{it}, \alpha_i)\neq 0$.

Išvada – svarbu žinoti α_i pasiskirstymą. Du variantai:

- jei α_i yra nestochastiniai, specifiniai individui bei laike nekintantys parametrai, kurie koreliuoja su $\boldsymbol{x}_i$, tai turime *fiksuotų efektų modelį.*
- jei α_i yra stochastiniai ir pasiskirstę $\alpha_i \overset{iid}{\sim} N(0, \sigma_\alpha^2)$, be to, nepriklausomi nuo $\varepsilon_{it} \sim iid\,(0, \sigma_\varepsilon^2)$ bei nekoreliuoja su $\boldsymbol{x}_{it}$, tada turime *atsitiktinių efektų modelį.*

○ **Pastaba!** Pagrindinė prielaida – ar atsitiktiniai efektai nekoreliuoja su $\boldsymbol{x}_i$.

Taigi galime apibendrinti, pasinaudoję $y_{it}=\boldsymbol{\beta}'\boldsymbol{x}_{it}+v_{it}$: jeigu tam tikro tipo endogeniškumas yra modelyje, tai kalbame apie fiksuotus efektus, kurie leidžia, atlikus duomenų transformaciją, pašalinti nepastebėtus α_i iš suminės paklaidos v_{it}. Kaip tai daroma? Suieškomi vidurkiai, apskaičiuoti pagal laiko veiksnį, t.y. $\bar{y}_i=\frac{1}{T}\sum_t y_{it}$, atitinkamai randami $\bar{\boldsymbol{x}}_i$, $\bar{\varepsilon}_i$, atkreiptinas dėmesys, jog ši transformacija skaičiuoja laiko vidurkį, nes α_i nepriklauso nuo laiko, dėl to ir vidurkis pagal laiko veiksnį lygus pačiam α_i.

Atimkime iš bendro modelio vidurkinį ir gausime: $y_{it}-\bar{y}_i=\alpha_i-\alpha_i+\boldsymbol{\beta}'\boldsymbol{x}_{it}-\boldsymbol{\beta}'\bar{\boldsymbol{x}}_i+\varepsilon_{it}-\bar{\varepsilon}_i$ arba $y_{it}-\bar{y}_i=(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)'\boldsymbol{\beta}+\varepsilon_{it}-\bar{\varepsilon}_i,\ i=1,\dots,N.$

Taigi koreliuojantis su regresoriais α_i yra pašalintas iš regresijos, todėl nesukelia paslinktumo. Tokiu metodu gautas rezultatas vadinamas *grupiniu įvertiniu* (angl. *within group estimator*), nes aiškina priklausomojo kintamojo variaciją aplink vidurkį, remdamasis regresorių variacijomis aplink vidurkius. Panašus metodas yra vietoj vidurkio iš pagrindinės regresijos atimti tokią pat regresiją, tik su vieno periodo vėlavimu. Tuomet taipogi išvengsime poslinkio, atsirandančio dėl nepastebėto heterogeniškumo.

○ **Pastaba!** Šių metodų esminis trūkumas – regresorių, kurie nekinta arba nedaug kinta laike, įtaka prarandama, nes po duomenų transformacijos jų dydis artėja prie nulio, taigi ir parametrai prie šių kintamųjų yra neišmatuojami.

Dar vienas fiksuotų efektų įvertinimo būdas – mažiausių kvadratų fiktyviųjų kintamųjų metodas. Šiuo atveju į modelį inkorporuojami nepastebėti efektai. Tai daroma α_i pakeičiant $\sum_i \alpha_i A_i$, modelis tampa $y_{it}=\sum_i \alpha_i A_i+\boldsymbol{\beta}'\boldsymbol{x}_{it}+\varepsilon_{it}$. Taigi α_i dabar modeliuojami kaip koeficientai prie fiktyviojo *i*-ojo individo kintamojo; tada $\alpha_i A_i$ yra fiksuotas efektas priklausomam individo *i* kintamajam y_i. Toks modelis gali būti apskaičiuotas MKM būdu.

○ **Pastaba!** Turite vieną subjektą neįtraukti į regresiją, mat pagal jį lyginsite efektus likusiems subjektams. Tokiu būdu laisvasis narys[12] β_1 bus būdingas tik šiam subjektui, o likusieji bus apskaičiuoti atitinkamai prie β_1 pridedant ar atimant α_i; įtraukus visus susidursite su „fiktyvių kintamųjų spąstais“, kurių rezultatas – tobulas multikolinearumas.

Kitas būdas – nenaudoti laisvojo nario ir visi α_i bus *i*-ojo subjekto laisvieji nariai. Visgi esminis trūkumas yra laisvės laipsnių sunaudojimas, kuomet didelis subjektų skaičius *N* lems *N* papildomų parametrų skaičiavimą.

Vis dėlto anksčiau aptartas grupinis metodas yra ekvivalentus mažiausių kvadratų fiktyviųjų kintamųjų metodui. Taigi kaskart įtraukti fiktyvių kintamųjų nėra prasmės – praktikoje naudojamas grupinis metodas, visgi pravartu suvokti, jog pastarasis sutampa su fiksuotų efektų modeliavimu, įtraukiant fiktyvius kintamuosius.

Apibendrinimo nusipelno ir atsitiktiniai efektai: jeigu aukščiau išdėstytos prielaidos yra patenkintos, tada MKM bus nepaslinkti ir suderinti, visgi neefektyvūs. Kitaip tariant, egzistuoja įvertiniai su mažesnėmis dispersijomis, o standartinės paklaidos yra klaidingos. Neefektyvumas atsiranda dėl to, kad kovariacija tarp suminių paklaidų v_{it} ir v_{is} nėra lygi nuliui: $Cov(v_{it}, v_{is}) = E(\alpha_i + \varepsilon_{it})(\alpha_i + \varepsilon_{is}) = E(\alpha_i \alpha_i) + E(\alpha_i \varepsilon_{is}) + E(\varepsilon_{it} \alpha_i) + E(\varepsilon_{it} \varepsilon_{is}) = \sigma_\alpha^2$.

Darėme prielaidą, jog $E(\alpha_i) = 0$, tada $E(v_{it}) = E(\varepsilon_{it}) + E(\alpha_i) = 0$. Dabar jau matome sprendimo būdą: šią informaciją panaudoti skaičiuojant įverčius, t.y.

---

[12] Nepamirškite, kad mūsų žymėjimuose $\boldsymbol{x}_{it}$ gali būti sudarytas iš vienetų, pavyzdžiui, $\boldsymbol{x}_{1t} = [1\ 1\ 1 \ldots 1]'$, tada β_1 bus laisvasis narys.

taip transformuoti paklaidas, kad jos atitiktų klasikines prielaidas. Tokiai transformacijai naudojami apskaičiuojami[13] apibendrinti kvadratai (angl. *feasible generalised least squares*).

Fiksuoti ar atsitiktiniai efektai? Hausman testas

Iš diskusijos apie fiksuotus ir atsitiktinius efektus ima aiškėti, kokioms sąlygoms esant priimtinesnis vienas metodas už kitą. Šiomis įžvalgomis pasinaudosime konstruodami Hausman testą.

Nulinė hipotezė yra (H_0): α_i ir $\boldsymbol{x}_i$ yra tarpusavyje nepriklausomai pasiskirstę. Tokiu atveju abeji, ir fiksuoti, ir atsitiktiniai efektai, yra suderinti. Visgi esant šiai nulinei hipotezei, tikrasis modelis yra atsitiktinių efektų, nes fiksuoti efektai yra neefektyvūs dėl fiktyviųjų kintamųjų panaudojimo (laisvės laipsniai sunaudojami, kai tai nėra reikalinga). Jeigu visgi α_i ir $\boldsymbol{x}_i$ nėra neprikluosimi, taigi H_0 yra atmetama, tada atsitiktiniai efektai įgis nepastebėto heterogeniškumo poslinkį ir sistemiškai skirsis nuo fiksuotų efektų įverčių.

Tegu atsitiktinių ir fiksuotų įverčių skirtumas bus žymimas $\hat{\kappa}=b_{RE}-b_{WG}$. Tada H_0 postuluos, jog asimptotiškai įverčiai sutampa, $plim\,\hat{\kappa}=0$ bei $Cov(\hat{\kappa}, b_{RE})=0$. Taigi akivaizdu, jog $Var(\hat{\kappa})=Var(b_{RE})-Var(b_{WG})$. Tuomet Hausman statistika pasiskirsčiusi pagal χ_k^2 skirstinį, kur k yra laike kintančių regresorių skaičius: $\hat{H}=\hat{\kappa}'[Var(\hat{\kappa})]^{-1}\hat{\kappa}\sim\chi_k^2$.

[13] Ne įprasti, o apskaičiuojami apibendrinti kvadratai reikalingi dėl to, kad nežinoma tikroji dispersijos-kovariacijos matrica yra pakeičiama įvertinta matrica, naudojantis imties informacija.

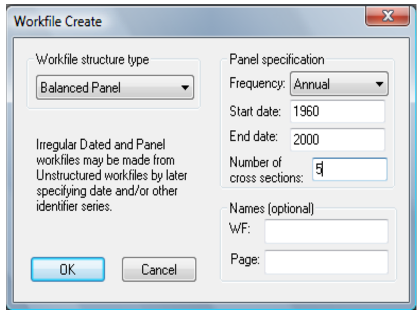
Atsitiktiniai efektai ar MKM? Breusch–Pagan LM testas

Jeigu Hausman testas indikuoja, jog priimtinesni yra atsitiktiniai efektai (negalime atmesti nulinės hipotezės), tada vertėtų išsiaiškinti, ar apskritai modelyje yra nepastebėtų efektų. Kitais žodžiais tariant, galbūt modelis yra puikiai specifikuotas ir suminė paklaida, $v_{it}=\alpha_i+\varepsilon_{it}$, išties yra lygi ε_{it}. Tokiu atveju jungtiniai MKM būtų pranašesni už atsitiktinius procesus, nes (1) MKM bus efektyvesni, mat nebandysime primesti neesančios grupių viduje autokoreliacijos, taip pat (2) MKM yra patikimesni baigtinės imties atveju, mat atsitiktiniai efektai remiasi asimptotinėmis savybėmis, kurių galiojimas mažose imtyse nebūtinai išpildomas.

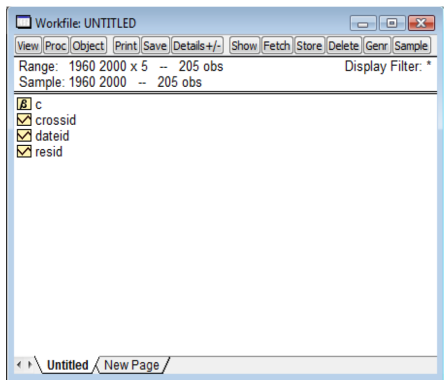
Pasinaudosime Breusch–Pagan LM testu, kurio postuluojama nulinė hipotezė yra atsitiktinių efektų nebuvimas. Testo statistika pasiskirsčiusi pagal χ_1^2. Apibendrintas algoritmas pateikiamas prieduose P3.

○ Pavyzdys

EViews turi „suprasti“, jog dirbame su paneliniais duomenimis. Aprašysime vieną būdą, kaip galima sukurti panelę, nors galimybių yra bent keletas. Pradedame **File/New/Workfile...** ir atveriame **Workfile Create** langą. Iš **Workfile structure type** išrenkame **Balanced Panel**. Tada pažymime, kokio dažnio yra duomenys, koks periodas ir kiek ekonominių vienetų stebime.



Suformuota panelė turi 205 stebėjimų (41 stebėjimas 5 šalims). *EViews* suformavo laiko ir šalies ID eilutes.

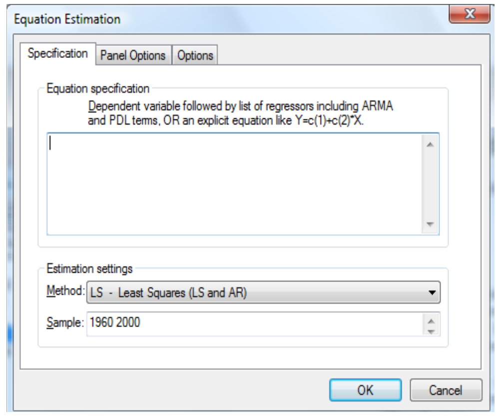


Dabar galime importuoti arba įrašyti duomenis. Atkreipkite dėmesį, jog paprastai duomenys išdėstomi vertikaliai: pagal šalis ir metus. Pavyzdžiui, turime penkias šalis ir kintamąjį *bvp*. Tada pastarasis bus 205x1 dimensijos, t.y. 205 eilutės ir 1 stulpelis. Duomenys išdėstomi taip:

$$\begin{pmatrix} x_{1960}^{1} \\ \dots \\ x_{2000}^{1} \\ x_{1960}^{2} \\ \dots \\ x_{2000}^{2} \\ \dots \\ x_{2000}^{5} \end{pmatrix}.$$

Modeliui įvertinti renkamės **Object/New Object...** ir **Equation**.

Jeigu tinkamai apibrėžėte panelę, lygties langas bus modifikuotas panelėms (bus naujaąselė **Panel Options**).



Suvedame lygtį įprasta forma (pvz., Y C X) bei spaudžiame **Panel Options**. Apskaičiuojame dvi lygtis: vieną kurioje **Effects specification–Cross-section** yra **Fixed**, o kitą **Random**. Taip įvertinsime α_i poveikį abiem, fiksuotų ir atsitiktinių efektų, atvejais. Taigi galėsime pamatyti, ar parametrų įverčiai reikšmingai keičiasi, pakeitus specifikaciją. Statistinis reikšmingumas, žinomas, tikrinamas testais.

Hausman testas *EViews* atliekamas pačios programos ir nereikia atskirai įvertinti papildomų regresijų. Pirma, įvertinkite atsitiktinius efektus. Tada rinkitės **View/Fixed/Random Effects Testing/Correlated Random Effects-Hausman Test**. Programa apskaičiuos reikalingas fiksuotų efektų specifikacijas, testo statistikas ir pateiks rezultatus su papildomomis lygtimis, reikalingomis Hausman testui. Atminkite, jog nulinė hipotezė postuluoja, jog priimtinesnis atsitiktinių efektų modelis.

Jeigu atmetate nulinę hipotezę, likite prie fiksuotų efekto modelio. Tuo tarpu neatmetus nulinės, dar raskite Breusch-Pagan LM testo rezultatus (šiuo atveju nulinė hipotezė postuluos, jog modelis nepasižymi atsitiktiniais efektais). Jis aprašytas skyrelyje apie heteroskedastiškumo testus. Deja, *EViews* neskaičiuoja šio testo automatiškai. Pritaikykite vieną iš P4 priede pateikiamų trumpų programėlių, suvesdami jas *EViews* viršutiniame lange (jos veiks priklausomai nuo jūsų *EViews* naudojamos versijos).

Įvertintą Breusch-Pagan LM lyginame su χ_1^2 kritine reikšme. Jeigu atmetate nulinę, likite prie atsitiktinių efektų specifikacijos, jeigu ne – naudokite JMKM. Tada tiesiog nespecifikuokite efektų (palikite prie **Effects Specification: None**).

Literatūra

Asteriou, D. *Applied econometrics: a modern approach using EViews and Microfit*, Palgrave Macmillan: New York, 2006.

Casteren, P.H.F.M., *Econometrics*, Department of Quantitative Economics, University of Amsterdam: The Netherlands, 2004.

Dougherty, C., *Introduction to Econometrics*, Oxford: Oxford University Press, 2006.

EViews 4 User's Guide, Quantitative Micro Software: USA, 2002.

Greene, W., *Econometric Analysis*, Prentice Hall: USA, 2008.

Johnson, R. R., *A Guide to Using EViews with Using Econometrics: A Practical Guide*, University of San Diego: USA, 2000.

Lecture notes on Econometrics, exercises, solutions, etc., Department of Quantitative Economics, University of Amsterdam: The Netherlands, 2006.

Verbeek, M., *A guide to modern econometrics*, Wiley: Chichester , 2000.

Priedai

P1. Įrodymas, jog yra s^2 nepaslinktas populiacijos dispersijos įvertis

Tekste teigėme, jog σ^2 nepaslinktas įvertis yra s^2, t.y. $E(s^2)=\sigma^2$, apibrėžtas $s^2=\frac{1}{n-1}\sum_i (X_i-\bar{X})^2$. Perrašykime nuokrypius, įvesdami populiacijos vidurkį, kuris nepakeičia išraiškos $(X_i-\bar{X})^2=[(X_i-\mu)-(\bar{X}-\mu)]^2$. Pakelkime kvadratu: $(X_i-\bar{X})^2=(X_i-\mu)^2-2(X_i-\mu)(\bar{X}-\mu)+(\bar{X}-\mu)^2$. Sumuokime pagal $i\in$ [1; n]:

$\sum_i (X_i-\bar{X})^2=\sum_i (X_i-\mu)^2-2(\bar{X}-\mu)\sum_i (X_i-\mu)+n(\bar{X}-\mu)^2$. Pastebėkime, jog $\sum_i (X_i-\mu)=\sum_i X_i-n\mu=n\bar{X}-n\mu=n(X-\mu)$. Tuomet

$\sum_i (X_i-\bar{X})^2=\sum_i (X_i-\mu)^2-2n(\bar{X}-\mu)^2+n(\bar{X}-\mu)^2=\sum_i (X_i-\mu)^2-n(\bar{X}-\mu)^2$.

Galiausiai, pritaikę matematinės vilties operatorių, gauname $E[\sum_i (X_i-\bar{X})^2]=E[\sum_i (X_i-\mu)^2]-nE[(\bar{X}-\mu)^2]$.

Prisiminkime dispersijų apibrėžimus:

$$E[(X_i-\mu)^2]=\sigma^2$$

bei

$$E[(\bar{X}-\mu)^2]=\sigma^2_{\bar{X}}$$

gauname $E[\sum_i (X_i-\bar{X})^2]=n\sigma^2-n\sigma^2_{\bar{X}}$,

nes $E[\sum_i (X_i-\mu)^2]=\sum_i E[(X_i-\mu)^2]=n\sigma^2$.

Jeigu stebėjimai yra atrinkti nepriklausomai vienas nuo kito, tuomet galioja $\sigma_{X_1+X_2}=\sigma_{X_1}+\sigma_{X_2}+2\sigma_{X_1X_2}=\sigma_{X_1}+\sigma_{X_2}$, t.y. stebėjimų kovariacija lygi nuliui.

Tiesa, nulinė kovariacija nėra pakankama nepriklausomumo sąlyga, tiesiog dydžiai yra nekoreliuoti. Taigi ryšys yra vienpusis: NEpriklausoma imtis reiškia ir tarpusavio NEkoreliaciją, bet NEkoreliuoti dydžiai gali būti ir priklausomi, ir nepriklausomi!

Ši įžvalga leis perrašyti $\sigma^2_{\bar{X}} = \frac{1}{n}\sigma^2_{\frac{1}{n}(X_1+X_2+\ldots+X_n)} = \frac{1}{n^2}\sigma^2_{X_1+X_2+\ldots+X_n}$, taigi

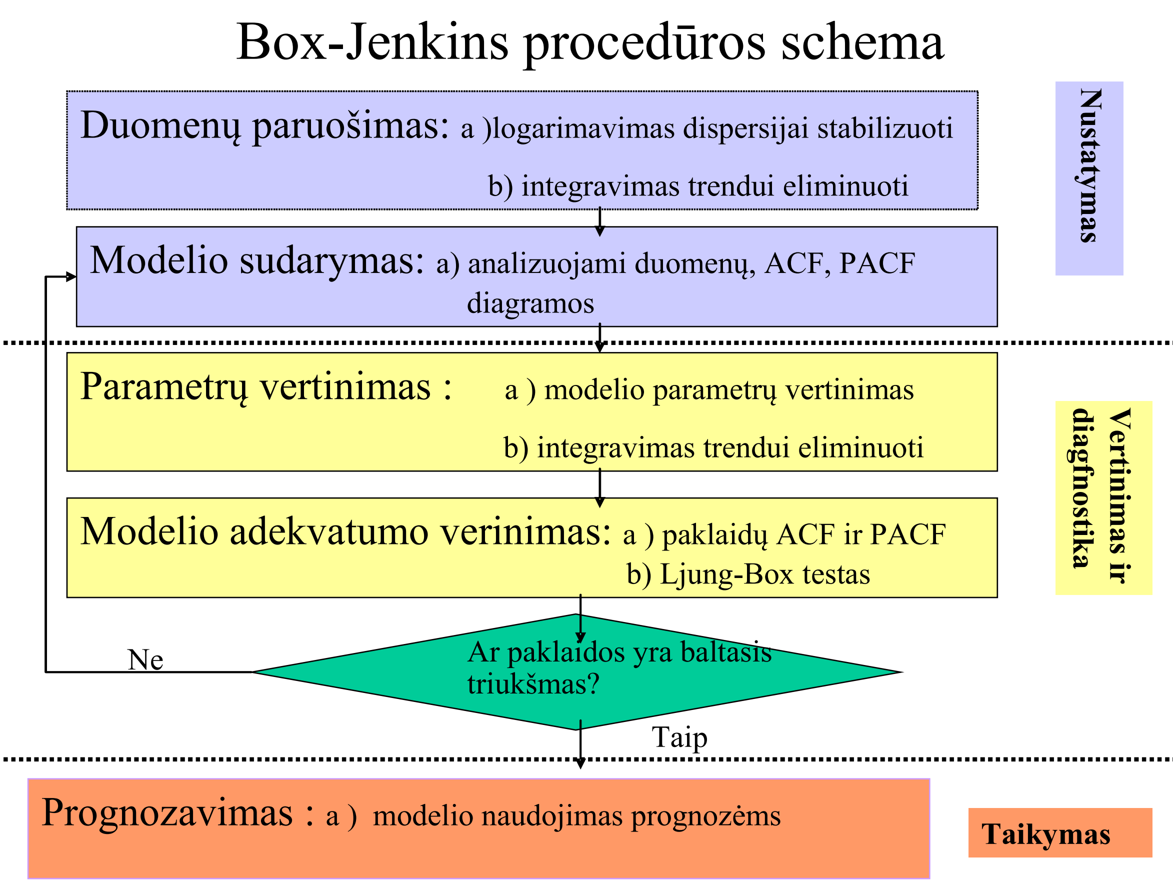
$$\sigma^2_{\bar{X}} = \frac{1}{n^2}(\sigma^2_{X_1}+\sigma^2_{X_2}+\ldots+\sigma^2_{X_n}) = \frac{1}{n^2}(\sigma^2+\sigma^2+\ldots+\sigma^2) = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}.$$

Tuomet $E[\sum_i (X_i - \bar{X})^2] = n\sigma^2 - n\sigma^2_{\bar{X}} = n\sigma^2 - n\frac{\sigma^2}{n} = (n-1)\sigma^2$.

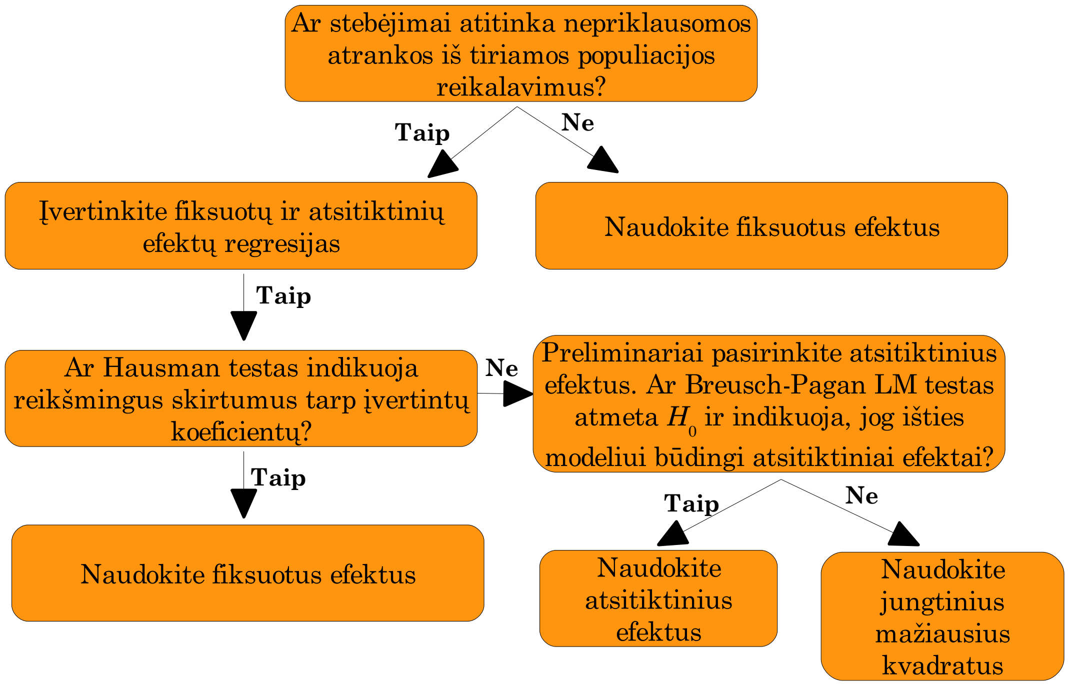
Galime užbaigti įrodymą:

$$E(s^2) = \frac{1}{n-1}E[\sum_i (X_i - \bar{X})^2] = \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2.$$

P2. Box-Jenkins procedūros etapų iliustracija



P3. Panelinių duomenų regresinio modelio parinkimas



Iš Dougherty (2006)

P4. Breusch-Pagan LM testo realizavimas EViews

1 programėlė

Įvertinkite lygtį ir išsaugokite N ir T

```
equation e1.ls() Y c X
!n=e1.@ncross
!t=e1.@npers
```

Gaukite likučius

```
e1.makeresids e1resids
```

Raskite likučius, pakeltus kvadratu

```
series e1residssq=e1resids*e1resids

!sum_resid_sq = @sum(e1residssq)
```

Raskite kryžminių duomenų sumų kvadratų sumą

```
!sum_sq_sums=0
for !i=1 to !n
  smpl if @crossid=!i
  !sum=@sum(e1resids)
  !sum_sq_sums=!sum_sq_sums+!sum^2
  smpl @all
next

delete e1resids
delete e1residssq
```

Rezultatai

```
Table BP
bp.setwidth(1) 25
BP(2,1)="Breusch-Pagan LM Test:"
BP(1,2)="Value"
BP(1,3)="Prob."
!LM=!n*!t/(2*(!t-1)) * (!sum_sq_sums/!sum_resid_sq - 1)^2
BP(2,2)=!lm
!prob=1-@cchisq(!lm,1)
bp(2,3)=!prob
Show bp
```

2 programėlė

```
' apskaičiuokite lygtį remdamiesi fiksuotais efektais
equation eq01.ls(cx=f) Y C X

' sukurkite likučių vektorių
eq01.makeresid presid
```

```
' perrašykite į naują puslapį
pageunstack(page=unstack) id01 id02 @ presid

'  apskaičiuokite necentruotą koreliaciją ir likučių (pagal subjektus) stebėjimų skaičių
do presid.cov(out=a) ucor obs

' raskite grupių skaičių
scalar n = @columns(acorr)

' raskite stebėjimų vektorių
vector t = @vech(aobs)

' pataisykite koreliacių matricą
acorr = acorr - @identity(n)

' raskite koreliacijų matricos vektorius
vector vcor = @vech(acorr)
vector vcor2 = @epow(vcor, 2)

' apskaičiuokite Breusch-Pagan LM statistiką
scalar cdlm = @inner(vcor2, t) - n*(n-1)/2
cdlm = cdlm / @sqrt(n * (n-1))
scalar cdlmpval = 2*@cnorm(-@abs(cdlm))
```