



**UNIVERSITATEA ECOLOGICĂ DIN
BUCUREȘTI
FACULTATEA DE MANAGEMENT
FINANCIAR**



Program de studii universitare de licență: Finanțe și Bănci

Forma de învățământ: învățământ cu frecvență redusă

Curs de autoinstruire pentru disciplina:

ECONOMETRIE

Coordonator de disciplină:

Prof.univ.dr. Dorin JULA

București

2022

ECONOMETRIE

CUPRINS

INTRODUCERE	8
Date privind coordonatorul de disciplină.....	8
Date despre disciplină	8
Obiectivele disciplinei	8
Competențe acumulate după parcurgerea cursului	9
Resurse și mijloace de lucru	10
Structura cursului	11
Cerințe preliminare	13
Durata medie de autoinstruire.....	13
Evaluarea.....	13
Unitatea de învățare 1. Modelul econometric.....	14
1.1. Obiective	14
1.2. Conținut.....	14
1.2.1. Modelul econometric – prezentare generală	14
1.2.2. Schema generală a procesului de modelare econometrică.....	16
1.2.3. Elementele unui model econometric.....	18
1.2.3.1. Variabilele modelului.....	19
1.2.3.2. Parametrii modelului	20
1.2.3.3. Ecuațiile modelului	21
1.2.3.4. Datele utilizate.....	22
1.3. Rezumat	23
1.4. Test de autoevaluare a cunoștințelor.....	24
1.5. Bibliografie recomandată.....	25
Unitatea de învățare 2. Modelul linear unifactorial	27

2.1. Obiective	27
2.2. Conținut.....	27
2.2.1. Ecuația de regresie lineară simplă.....	27
2.2.2. Metoda celor mai mici pătrate.....	28
2.3. Problemă rezolvată.....	30
2.4. Rezumat	32
2.5. Test de autoevaluare a cunoștințelor.....	33
2.6. Bibliografie recomandată.....	33
2.7. Teme de control	34
Unitatea de învățare 3. Ipotezele de aplicare ale modelului linear	
unifactorial. Proprietăți ale estimatorilor	35
3.1. Obiective	35
3.2. Conținut.....	35
3.2.1. Ipotezele de aplicare ale modelului linear unifactorial	35
3.2.2. Proprietăți ale estimatorilor.....	38
3.2.3. Teorema Gauss -Markov	40
3.3. Rezumat	41
3.4. Test de autoevaluare a cunoștințelor.....	41
3.5. Bibliografie recomandată.....	42
Unitatea de învățare 4. Modelul linear multifactorial.....	
44	
4.1. Obiective	44
4.2. Conținut.....	44
4.2.1. Ecuația de regresie lineară multifactorială.....	44
4.2.2. Forma matriceală a modelului linear multifactorial de regresie	45
4.2.3. Estimarea parametrilor din modelul linear multifactorial.....	46
4.3. Problemă rezolvată.....	47
4.4. Rezumat	50
4.5. Test de autoevaluare a cunoștințelor.....	51
4.6. Bibliografie recomandată.....	51

Unitatea de învățare 5. Ipotezele de aplicare ale modelului linear multifactorial. Proprietăți ale estimatorilor	53
5.1. Obiective	53
5.2. Conținut.....	53
5.2.1. Ipotezele de aplicare ale modelului linear multifactorial.....	53
5.2.2. Proprietăți ale estimatorilor.....	54
5.3. Rezumat	58
5.4. Test de autoevaluare a cunoștințelor.....	59
5.5. Bibliografie recomandată.....	60
Unitatea de învățare 6. Testarea semnificației parametrilor din modelul linear unifactorial de regresie	61
6.1. Obiective	61
6.2. Conținut.....	61
6.2.1. Dispersia estimatorilor în modelul linear unifactorial	61
6.2.2. Teste de semnificație în cazul modelului unifactorial	62
6.3. Problemă rezolvată.....	64
6.4. Rezumat	65
6.5. Test de autoevaluare a cunoștințelor.....	66
6.6. Bibliografie recomandată.....	66
Unitatea de învățare 7. Testarea semnificației parametrilor din modelul linear multifactorial de regresie.....	67
7.1. Obiective	67
7.2. Conținut.....	67
7.2.1. Dispersia estimatorilor în modelul linear multifactorial	67
7.2.2. Teste de semnificație în cazul modelului unifactorial	68
7.3. Problemă rezolvată.....	70
7.4. Rezumat	72
7.5. Test de autoevaluare a cunoștințelor.....	72
7.6. Bibliografie recomandată.....	73

Unitatea de învățare 8. Acuratețea ajustării	74
8.1. Obiective	74
8.2. Conținut.....	74
8.2.1. Coeficientul de determinare	74
8.2.1. Coeficientul de determinare corectat	76
8.3. Problemă rezolvată.....	77
8.4. Rezumat	79
8.5. Test de autoevaluare a cunoștințelor.....	79
8.6. Bibliografie recomandată.....	80
8.7. Teme de control	80
Unitatea de învățare 9. Specificarea modelului.....	81
9.1. Obiective	81
9.2. Conținut.....	81
9.2.1. Criterii pentru specificarea modelului.....	81
9.2.1.1. Criteriul informațional Akaike (AIC)	82
9.2.1.2. Criteriul informațional Schwartz (SIC).....	82
9.2.1.3. Criteriul Hannan-Quinn	82
9.2.2. Erori de specificare a modelului linear de regresie.....	83
9.3. Problemă rezolvată.....	85
9.4. Rezumat	86
9.5. Test de autoevaluare a cunoștințelor.....	87
9.6. Bibliografie recomandată.....	87
Unitatea de învățare 10. Multicolinearitatea.....	89
10.1. Obiective	89
10.2. Conținut.....	89
10.2.1. Consecințe ale multicolinearității.....	89
10.2.2. Identificarea multicolinearității	90
10.2.2.1. Criteriul lui Klein	91
10.2.2.2. Factorul de inflație al dispersiei	92

10.2.3. Atenuarea multicolarității.....	92
10.3. Problemă rezolvată.....	95
10.4. Rezumat	96
10.5. Test de autoevaluare a cunoștințelor.....	96
10.6. Bibliografie recomandată.....	97
Unitatea de învățare 11. Heteroscedasticitatea erorilor	98
11.1. Obiective	98
11.2. Conținut.....	98
11.2.1. Consecințe ale heteroscedasticității.....	98
11.2.2. Testarea heteroscedasticității.....	99
11.2.2.1. Testul Goldfeld–Quandt.....	99
11.2.2.2. Testul White	100
11.3. Problemă rezolvată.....	102
11.4. Rezumat	104
11.5. Test de autoevaluare a cunoștințelor.....	104
11.6. Bibliografie recomandată.....	105
Unitatea de învățare 12. Autocorelarea erorilor	106
12.1. Obiective	106
12.2. Conținut.....	106
12.2.1. Consecințe ale autocorelării erorilor	106
12.2.2. Testarea autocorelării erorilor	107
12.2.2.1. Testul Durbin – Watson	107
12.2.2.2. Testul Lagrange.....	109
12.3. Problemă rezolvată.....	111
12.4. Rezumat	113
12.5. Test de autoevaluare a cunoștințelor.....	113
12.6. Bibliografie recomandată.....	114
Unitatea de învățare 13. Normalitatea distribuției erorilor	115
13.1. Obiective	115

13.2. Conținut.....	115
13.2.1. Consecințe ale nerespectării ipotezei de normalitate a distribuției erorilor din modelul linear de regresie	115
13.2.2. Testarea ipotezei de normalitate a distribuției erorilor din modelul linear de regresie. Testul Jarque-Bera.....	116
13.3. Problemă rezolvată.....	118
13.4. Rezumat	120
13.5. Test de autoevaluare a cunoștințelor.....	120
13.6. Bibliografie recomandată.....	121
Unitatea de învățare 14. Prognoza econometrică.....	122
14.1. Obiective	122
14.2. Conținut.....	122
14.2.1. Prognoza în cazul modelului linear unifactorial de regresie.....	122
14.2.2. Prognoza în cazul modelului linear multifactorial de regresie	123
14.3. Problemă rezolvată.....	125
14.4. Rezumat	126
14.5. Test de autoevaluare a cunoștințelor.....	127
14.6. Bibliografie recomandată.....	127
14.7. Teme de control	128
Răspunsuri la testele de autoevaluare a cunoștințelor	129

Lista figurilor din text

Figura 1-1. Schema generală a procesului de modelare econometrică.....	17
Figura 1-2. Etapele unui studiu econometric	18
Figura 2-1: Dreapta de regresie.....	28
Figura 2-2. Modelul unifactorial	31
Figura 4-1. Modelul bi-factorial de regresie	49
Figura 12-1: Testul bilateral Durbin – Watson pentru autocorelarea erorilor ..	108

Figura 13-1. Testul Jarque-Bera.....	119
Figura 14-1. Prognoza econometrică	126

INTRODUCERE

Cursul este destinat, în principal, studenților *Facultății de Management Financiar* din cadrul Universității Ecologice din București – învățământ cu frecvență redusă (IFR), dar poate constitui un util material de studiu și pentru studenții la învățământ cu frecvență (IF).

Date privind coordonatorul de disciplină

Nume și prenume:	Prof.univ.dr. Dorin JULA
E-mail:	dorin.jula@ueb.education

Date despre disciplină

Anul de studiu:	III
Semestrul:	1

Obiectivele disciplinei

Obiectivul general al disciplinei	Însușirea principalelor concepte teoretice, a categoriilor economice, a principiilor și a modului de funcționare a mecanismului economic
Obiectivele specifice	▪ prezentarea noțiunilor și conceptelor teoretice fundamentale referitoare la comportamentul agenților economici, ▪ formarea cunoștințelor de baza privind problemele teoretico-metodologice ale economiei de piață, ▪ prezentarea principalelor concepte teoretice, a categoriilor economice, a principiilor și a modului de funcționare a mecanismului economic;

	▪ prezentarea modului de utilizare a instrumentelor de analiză și calcul economic din corpul metodologic general al științei economice; ▪ aplicarea principiului abordării ecologice în activitatea economică
--	--

Competențe acumulate după parcurgerea cursului

Competențe profesionale	▪ Identificarea, definirea, explicarea și interpretarea conceptelor, teoriilor, metodelor și instrumentelor de natură economică ▪ Aplicarea conceptelor, teoriilor, metodelor și instrumentelor de natură economică în entitățile private și publice pentru rezolvarea de probleme specifice ▪ Evaluarea critică a conceptelor, metodelor și instrumentelor de natură economică folosite pentru rezolvarea de probleme ▪ Valorificarea conceptelor, teoriilor, metodelor și instrumentelor de natură financiară în elaborarea de proiecte/lucrări ▪ Identificarea și definirea metodelor, tehnicilor și instrumentelor de culegere, analiză și interpretare a datelor referitoare la o problemă economico-financiară ▪ Explicarea metodelor, tehnicilor și instrumentelor de culegere, analiză și interpretare a datelor referitoare la o problemă economico-financiară ▪ Aplicarea metodelor, tehnicilor și instrumentelor de culegere, analiză și interpretare a datelor referitoare la o problemă economico-financiară
-------------------------	---

	▪ Efectuarea de analize economico-financiare curente pe baza datelor și informațiilor culese ▪ Aplicarea cunoștințelor, metodelor, tehnicilor și instrumentelor pentru realizarea lucrărilor economico-financiare
Competențe transversale	

Resurse și mijloace de lucru

Pentru o pregătire temeinică, vă sugerăm să efectuați toate testele de evaluare, astfel încât rezultatul pregătirii dumneavoastră să fie cât mai obiectiv. De asemenea, vă sugerăm să vă alcătuiți propriile dumneavoastră planuri și scheme, acestea ajutându-vă la o mai bună sistematizare a cunoștințelor dobândite.

Fiecare unitate de învățare debutează cu prezentarea obiectivelor pe care trebuie să le atingeți – din punctul de vedere al nivelului de cunoștințe – prin studierea temei respective. Vă recomandăm să le citiți cu atenție și apoi, la sfârșitul unității de învățare, să le revedeți pentru a verifica dacă le-ați atins în întregime.

Testele de evaluare prezente la sfârșitul fiecărei unități de învățare vă vor ajuta să verificați modalitatea specifică de învățare și să vă îmbunătățiți cunoștințele. Timpul recomandat rezolvării testelor este de 30 de minute. Rezolvați testele numai după studierea în întregime a unei unități de învățare și nu vă uitați la răspunsuri decât după rezolvarea testului și numai dacă nu reușiți să găsiți răspunsul prin recitirea cursului.

La fiecare unitate de învățare există indicată o bibliografie selectivă pe care vă sfătuim să o parcurgeți.

Structura cursului

Cursul de *Econometrie* este alcătuit din 14 **unități de învățare**:

- Unitatea de învățare 1. Modelul econometric
- Unitatea de învățare 2. Modelul linear unifactorial
- Unitatea de învățare 3. Ipotezele de aplicare ale modelului linear unifactorial.
Proprietăți ale estimatorilor
- Unitatea de învățare 4. Modelul linear multifactorial
- Unitatea de învățare 5. Ipotezele de aplicare ale modelului linear multifactorial.
Proprietăți ale estimatorilor
- Unitatea de învățare 6. Testarea semnificației parametrilor din modelul linear unifactorial de regresie
- Unitatea de învățare 7. Testarea semnificației parametrilor din modelul linear multifactorial de regresie
- Unitatea de învățare 8. Acuratețea ajustării
- Unitatea de învățare 9. Specificarea modelului
- Unitatea de învățare 10. Multicolinearitatea
- Unitatea de învățare 11. Heteroscedasticitatea erorilor
- Unitatea de învățare 12. Autocorelarea erorilor
- Unitatea de învățare 13. Normalitatea distribuției erorilor
- Unitatea de învățare 14. Prognoza econometrică

La sfârșitul fiecărei unități de învățare sunt prezentate modele de teste de autoevaluare care vă vor ajuta să stabiliți singuri ritmul de învățare și necesitățile proprii de repetare a unor teme.

La sfârșitul suportului de curs sunt prezentate răspunsurile corecte la testele de autoevaluare a cunoștințelor.

Totodată, sunt formulate **3 teme de control**, în vederea evaluării pe parcurs, astfel:

Tema 1 – Capitolul 2: "Realizați un referat de maximum 5 pagini în care să prezentați contribuția la dezvoltarea econometriei a unui laureat al premiului Nobel pentru economie";

Tema 2 – Capitolul 8: "Din site-ul Institutului Național de Statistică (www.insse.ro), de la secțiunea Tempo On-line selectați trei serii de date anuale, începând cu 1990 și până în prezent. Analizați econometric legătura dintre variabilele respective (ecuația de regresie, estimarea coeficienților, testele de semnificație, acuratețea ajustării)";

Tema 3 – Capitolul 14: "Din site-ul Băncii Naționale a României (www.bnro.ro), de la secțiunea *Statistică*, subsecțiunea *Bază de date interactivă* selectați serii lunare referitoare la baza monetară și inflație. Construiți, testați și aplicați un model de prognoză a inflației".

Temele de control se prezintă de către fiecare student, iar rezultatele acestei evaluări se comunică de către cadrul didactic în mod direct și nemijlocit studenților.

Cursul de *Econometrie* poate fi studiat atât în întregime, potrivit ordinii prestabilite a capitolelor, dar se poate și fragmenta în funcție de interesul propriu mai accentuat pentru anumite teme. Însă, în vederea susținerii examenului este obligatorie parcurgerea tuturor celor 14 capitole și efectuarea testelor prezentate.

Cerințe preliminare

Pentru o mai bună înțelegere a materiei, este necesară coroborarea cunoștințelor dobândite în cadrul acestei discipline cu cele acumulate la *Economie* (Microeconomie și Macroeconomie) și la *Statistică*.

Durata medie de autoinstruire

Durata medie de învățare estimăm a fi de aproximativ 28 de ore, iar pentru examen ar fi necesar un studiu de o săptămână.

Studierea fiecărei unități de învățare necesită un efort minim estimat la 2-4 ore (precizat la fiecare temă), astfel încât să se fixeze temeinic cunoștințele specifice acestui domeniu.

Evaluarea

La încheierea studierii cursului, este obligatorie realizarea celor trei teme de control. Înainte de examen este indicat să parcurgeți din nou toată materia, cu atenție, durata estimată pentru această activitate fiind de aproximativ o săptămână.

Forma de evaluare (E-examen, C-colocviu/test final, LP-lucrări de control)		E
Stabilirea notei finale (procentaje)	– evaluarea finală	60%
	– activități aplicative /laborator/ lucrări practice/proiect etc.	20%
	– teste pe parcursul semestrului	10%
	– teme de control	10%

Unitatea de învățare 1. Modelul econometric



1.1. Obiective

În cadrul acestei unități de învățare sunt prezentate cunoștințele referitoare la structura și elementele unui model econometric, precum și schema generală a procesului de modelare.

Durata medie de parcurgere a acestei unități de învățare este de aproximativ 2 ore.



1.2. Conținut

1.2.1. Modelul econometric – prezentare generală

Denumirea de *econometrie* provine din combinarea cuvintelor grecești *oikonomia* – economia (de la *oikos* – casă, gospodărie și *nomos* – lege) și *metron* – măsură. Deci, etimologic, econometria presupune aplicarea unor tehnici de măsurare în economie.

În sens restrâns, econometria este definită ca o aplicație a statisticii matematice în economie, astfel încât prin analiza și prelucrarea datelor economice să se ofere un suport empiric modelelor construite de economia matematică¹. Sau, într-o altă formulare, mai succintă, sarcina de bază a econometriei este să umple empiric structurile teoretice².

¹ Samuelson P.A., Koopmans T.C., Stone J.R.N., 1954, *Report of the evaluative committee for Econometrica*, în *Econometrica*, 22, pag.141-146.

² "The basic task of econometrics ... is to put empirical flesh and blood on theoretical structures", Johnston J., 1984, *Econometric Methods*, 3rd edition, McGraw-Hill pag. 5.

În sens larg, econometria este înțeleasă ca o știință de graniță între economie, matematică și statistică (Thomas R.L., 1997³ și Ramanathan R., 1992⁴).

Pornind de la relațiile definite de teoria economică, econometria clasică se concentrează asupra următoarelor tipuri de analiză⁵:

- (a) testarea și validarea teoriei economice (în sensul confruntării teoriei cu realitatea economică, inclusiv testarea ipotezelor privind comportamentul economic), concret, testarea ideilor, noțiunilor, conceptelor teoretice reprezentate prin asemenea relații;
- (b) estimarea relațiilor dintre variabilele economice, adică măsurarea fiecărei relații și estimarea parametrilor pe care relațiile respective îi presupun;
- (c) prognoza evoluțiilor și a comportamentelor economice.

Pentru testarea corespondenței dintre teoria economică și realitate (fapte, obiecte, procese), econometria realizează o combinare a tehnicilor de analiză specifice economiei matematice cu cele din statistica economică și inferența statistică. Economia matematică exprimă teoriile și ideile economice în formă matematică. Aceste exprimări sunt de cele mai multe ori calitative, în sensul că nu presupun întotdeauna utilizarea datelor numerice. Statistica economică se ocupă cu identificarea, colectarea și analiza datelor economice și exprimarea lor într-o formă inteligibilă pentru utilizatori. Ca o sinteză a acestor două abordări, econometria preia ecuațiile din economia matematică și le confruntă cu datele economice. În acest proces (econometric), prin utilizarea unor tehnici de inferență se testează adecvarea teoriei la realitatea economică, prin validarea ecuațiilor

³ Thomas R.-L., 1997, *Modern Econometrics: An Introduction*, 2nd ed., Harlow, Longman, pp. 1-3.

⁴ Ramanathan R., 1992, *Introductory Econometrics*, Second Edition, The Dryden Press, Harcourt Brace College Publishers, Orlando, USA, pp. 3-11.

⁵ Thomas R.-L., 1997, *Modern Econometrics: An Introduction*, 2nd edition, Harlow, Longman, p. 1.

respective. De obicei, tehnicile utilizate în scopul validării empirice a ecuațiilor care modelează un anumit aspect al realității economice sunt cele de inferență statistică.

1.2.2. Schema generală a procesului de modelare econometrică

Pornind de la schema generală privind procesul de modelare și de la elementele prezentate, poate fi construită o schiță a procesului de modelare econometrică. Astfel Spanos (1986)⁶, consideră că, de cele mai multe ori, variabilele din modelul teoretic nu corespund în mod direct cu seriile de date observate. În această situație, în construirea modelului sunt generate date printr-un proces specific. Un astfel de *proces generator de date* poate porni de la izolearea fenomenelor care ne interesează de orice alte influențe. De exemplu, într-un model teoretic care analizează cererea de monedă sunt incluse ca variabile explicative venitul, nivelul prețurilor, rata dobânzii ș.a. În statistica oficială⁷ există mai multe serii de date care pot reprezenta candidați verosimili pentru măsurarea variabilelor respective. Este posibil însă ca nici una dintre seriile statistice să nu corespundă exact descrierii teoretice. În aceste condiții, concluziile modelului pot fi false. De aceea, mecanismul prin care sunt identificate datele utilizate în model, numit *procesul de generare a datelor (PGD)*⁸ are un rol important în modelarea econometrică. Se afirmă, de altfel, că modelul econometric trebuie interpretat ca o aproximare a PGD (în engleză: *Data Generating Process – DGP*).

⁶ Spanos A., 1986, *Statistical foundations of econometric modelling*, Cambridge University Press, pp. 19-22.

⁷ *Anuarele Statistice și Buletinele statistice lunare* ale Institutului Național de Statistică, *Buletinele lunare și Rapoartele anuale* ale Băncii Naționale a României, *Buletinele de informare* ale Ministerului Finanțelor Publice, *Eurostat* ș.a.

⁸ În engleză: *Data Generating Process (DGP)*.

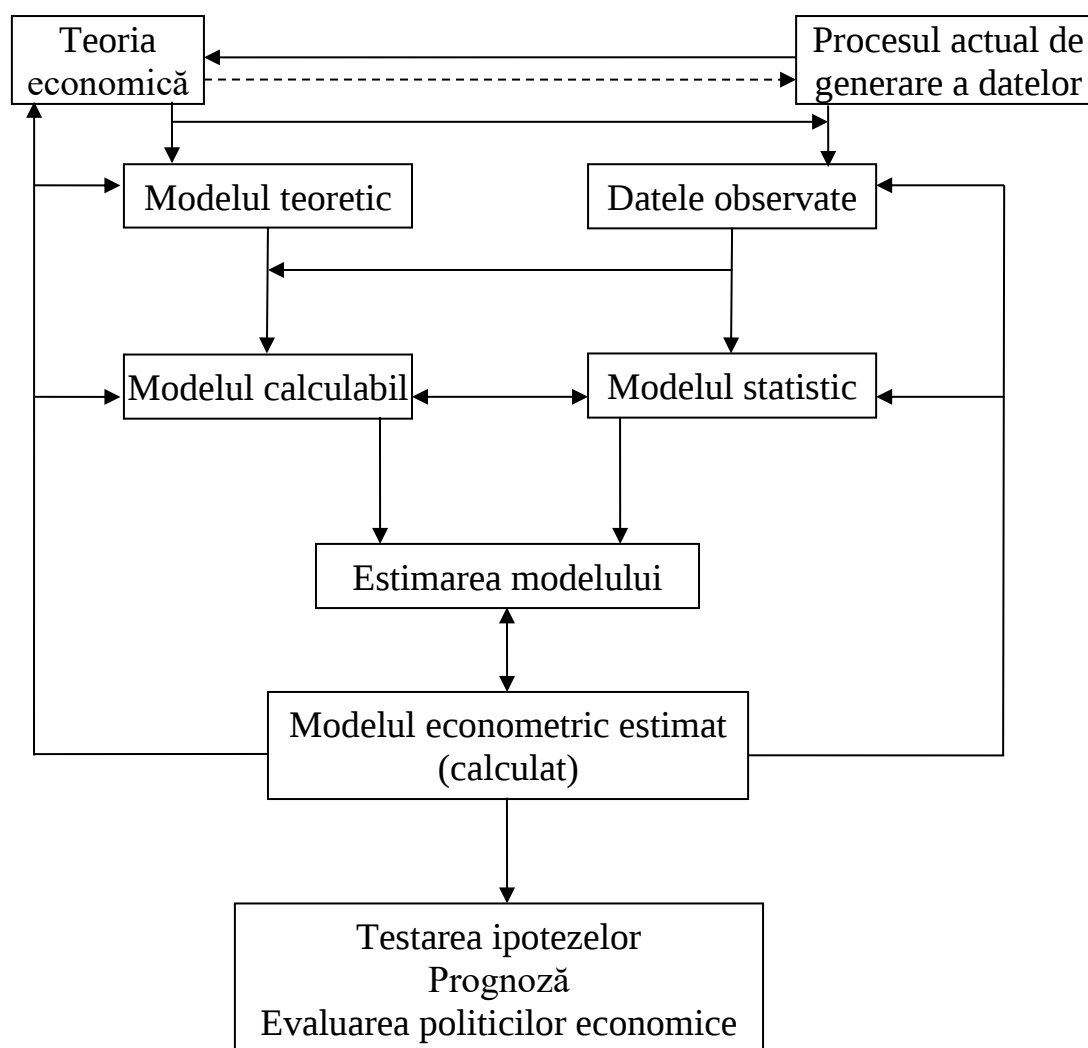


Figura 1-1. Schema generală a procesului de modelare econometrică

Sursa: adaptat după Spanos A., 1986, *Statistical foundations of econometric modelling*, Cambridge University Press, Cambridge (fig.1.1: *The textbook approach to econometric modeling*, p. 16 și fig.1.2: *An approach to econometric modelling*, p. 21).

Un studiu econometric presupune, cel puțin, parcurgerea următoarelor etape (figura 1-2):

- (1) formularea modelului pornind de la teoria considerată adecvată pentru explicarea evoluției unui proces economic,
- (2) realizarea unei cercetări selective pentru generarea unor serii de date referitoare la procesul respectiv,
- (3) estimarea modelului,
- (4) testarea ipotezelor teoretice folosite în construirea modelului

(5) interpretarea rezultatelor.

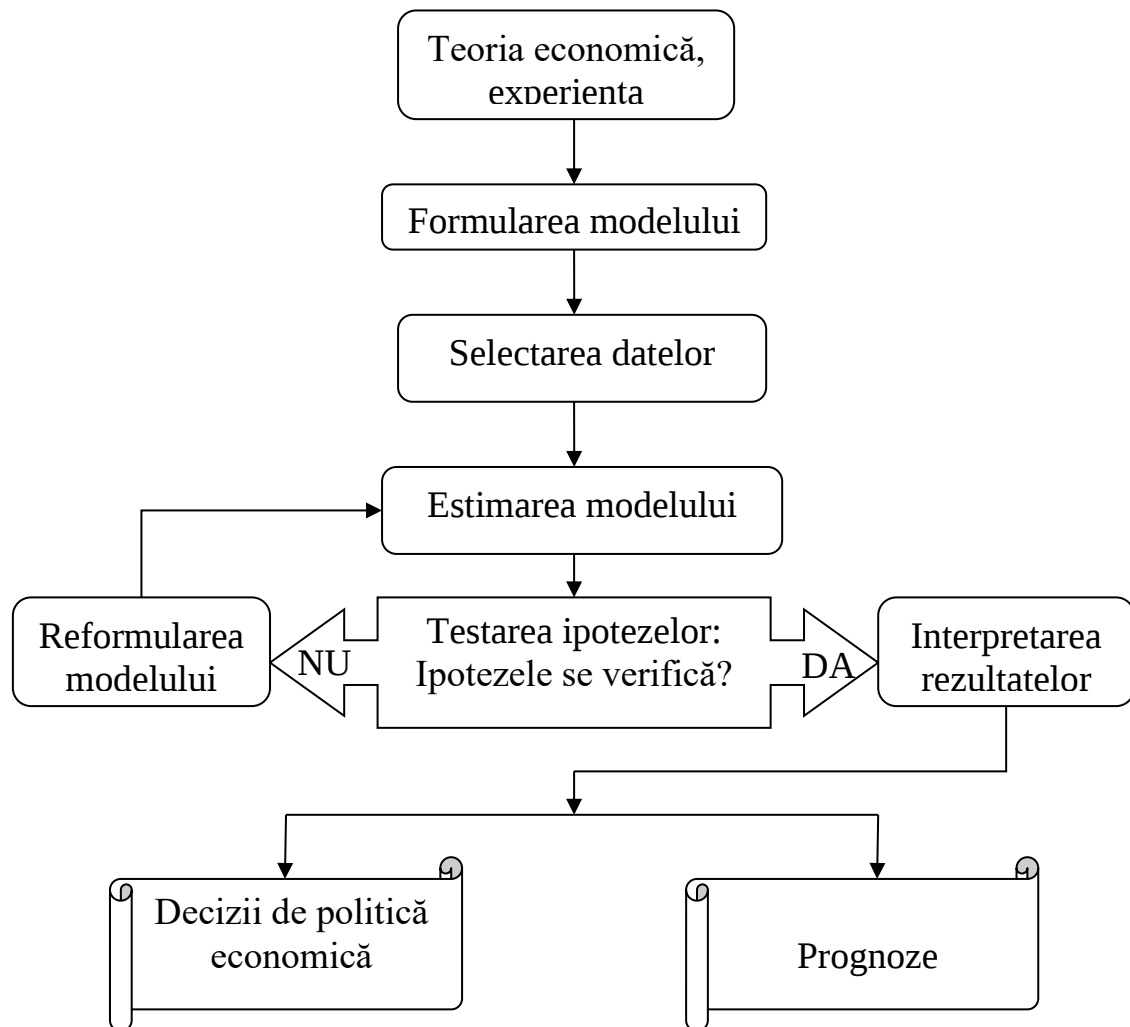


Figura 1-2. Etapele unui studiu econometric

1.2.3. Elementele unui model econometric

Într-o abordare generală, un model econometric reprezintă un ansamblu de relații interdependente care descriu legăturile dintre valorile unui anumit număr de variabile economice, într-un context dat. Elementele unui model economico-matematic sunt variabilele, ecuațiile și parametrii modelului. De asemenea, rezolvarea modelului presupune existența unor serii de date, care să prezinte starea și/sau evoluția (distribuția) variabilelor din model.

1.2.3.1. Variabilele modelului

Într-un model econometric pot fi distinse trei tipuri de variabile: variabile endogene, variabile exogene și variabile de abatere (eroare).

Sunt *endogene* sau *explicate* acele variabile ale căror valori sunt obținute prin rezolvarea modelului. De obicei, numărul variabilelor endogene este egal cu numărul ecuațiilor din model.

Sunt *exogene* acele variabile pentru care starea și evoluția sunt determinate de factori exteriori sistemului a cărui funcționare este studiată cu ajutorul modelului. În cadrul variabilelor exogene, o categorie specială este constituită din *variabilele de decizie*. Variabilele de decizie sunt instrumente a căror evoluție poate fi controlată sau simulată de un agent economic. Sunt utilizate atunci când se urmărește măsurarea impactului unei decizii sau a unei situații posibile asupra evoluției uneia sau mai multor variabile endogene, sau asupra echilibrului economic descris de model. Printr-o astfel de variabilă se poate simula răspunsul sistemului la un șoc extern (de exemplu, creșterea prețului petrolului) sau la o măsură de politică economică (de exemplu, modificarea ratei reescontului sau a ratei rezervelor obligatorii).

Exogenele sunt întotdeauna variabile explicative, însă nu orice variabilă explicativă din model este și exogenă. De aceea, sunt numite *predeterminate* sau *explicative* acele variabile ale căror valori sunt cunoscute a priori și sunt utilizate pentru explicarea stării și evoluției variabilelor endogene. Pe lângă exogene, predeterminate pot fi și *variabilele retardate* (variabilele *cu efect întârziat*). Adică, este posibil ca printre factorii explicativi ai evoluției unei variabile endogene să se găsească valoarea variabilei respective sau valorile altor endogene calculate pentru un moment anterior. De exemplu, nivelul șomajului la momentul t depinde de șomajul la momentul $t-1$, deoarece absorbția excesului de ofertă de pe piața muncii nu se produce instantaneu, ci este un proces care se realizează în timp. La fel, consumul curent depinde de nivelul consumului înregistrat la

momentul anterior, deoarece familiile au tendința de a-și conserva un anumit nivel de trai, eventual prin diminuarea economisirii, sau chiar printr-un proces de dezechilibru. De asemenea, în funcția ofertei, decizia de producție la momentul t depinde de prețul pe piață la momentul $t-1$.

Variabilele de abatere (sau *erorile*) reprezintă discrepanțele între evoluția anticipată a unei variabile și evoluția reală a variabilei respective. Aceste discrepanțe (erori) pot proveni din perturbații aleatoare cărora nu li se poate atribui o semnificație economică, din neincluderea în model a unor variabile semnificative, sau includerea unor variabile explicative mai puțin semnificative, din erori de măsurare sau din imposibilitatea măsurării unor factori de influență (elemente psiho-sociale, variabile care țin de mediul cultural sau politic ș.a.).

1.2.3.2. Parametrii modelului

Parametrii modelului pot fi definiți ca fiind caracteristicile cantitative ale sistemului studiat. Dacă vectorul parametrilor este fixat, comportamentul sistemului depinde numai de starea și evoluția variabilelor explicative și de forma legăturilor dintre variabile. În cazul general, fiecărei colecții de valori ale parametrilor îi corespunde o mulțime de traiectorii ale sistemului, mulțime numită câmp de comportament⁹. La limită, fiecărei colecții de valori ale parametrilor îi corespunde o singură traiectorie a sistemului.

În modelele econometrice, se acceptă ipoteza că forma legăturii dintre variabile este cunoscută, iar parametrii modelului nu sunt cunoscuți. În rezolvarea modelului se calculează estimatori ai parametrilor respectivi pornind de la o cercetare selectivă privind starea și evoluția variabilelor din model.

⁹ Fedorenko N.P., Kantorovici L.V., Danilov-Danilian V.I., Konüs A.A., Maiminas E.Z., Ceremniș I.N., Cerneak I.I., 1979, *Dicționar de matematică și cibernetică în economie*, Editura Științifică și Enciclopedică, București, pag. 439.

1.2.3.3. Ecuațiile modelului

Prin ecuațiile unui model se încearcă surprinderea legăturilor dintre variabilele endogene și cele explicative. Într-un model econometric pot apărea ecuații de comportament, ecuații de definiție și ecuații contabile (de echilibru).

Ecuațiile de comportament sunt construite pe baza unei teorii economice și descriu, în formă funcțională, comportamentul unui agent economic. Comportamentul, într-un anumit context, al agentului respectiv este apreciat prin starea și evoluția unei variabile economice. La fel, se atașează o variabilă fiecărui factor care explică un anumit comportament. Intensitatea influenței unui factor explicativ asupra comportamentului economic al agentului considerat este dată de mărimea parametrului care stabilește legătura dintre valorile celor două variabile.

Ecuațiile sau relațiile de definiție sunt utilizate pentru precizarea unor noțiuni sau determinarea unor variabile. În ecuațiile de definiție nu intervin parametri necunoscuți. De exemplu, într-un model în care sunt estimate nivelul ocupării (PO) și numărul de șomeri (S), populația activă (PA) se determină prin însumarea valorilor celor două variabile, iar rata șomajului se definește prin raportul dintre numărul șomerilor și populația activă. Indicele de creștere al unei variabile se calculează ca raport între starea variabilei respective la momentul t și starea la momentul $t-1$. Ritmul de creștere se definește scăzând valoarea 1 din indicele de creștere ș.a.m.d. De asemenea, pot fi definite anumite variabile intermediare care să ajute la simplificarea ecuațiilor, sau care să asigure respectarea anumitor condiții necesare pentru rezolvarea modelului. De exemplu, pentru atenuarea corelației dintre erorile din model se procedează la diferențierea seriilor de date (se calculează diferența dintre valoarea la momentul t și valoarea la momentul $t-1$), sau în anumite situații se operează în model cu logaritmul valorilor înregistrate prin cercetarea selectivă.

Ecuațiile de echilibru sunt utilizate pentru asigurarea coerenței modelului. De exemplu, într-un model macroeconomic suma dintre consumul și economiile

famiiliilor trebuie să fie egal cu venitul minus impozitele plătite de famiiliile respective. De asemenea, dacă în model se calculează atât populația ocupată (PO) și nivelul șomajului (S), cât și populația activă (PA), atunci $PA = PO + S$ este o relație de echilibru prin care se poate testa corectitudinea estimării și coerența modelului. Ecuatiile de echilibru nu exclud interpretările și analizele teoretice. De exemplu, o ecuație de echilibru poate fi scrisă în ipoteza adaptării perfecte și instantanee a ofertei la cererea pe piața unui produs. Or, un asemenea comportament chiar dacă poate fi acceptat în anumite modele, nu are un caracter universal valabil.

1.2.3.4. Datele utilizate

Pentru estimarea parametrilor din ecuațiile modelului sunt utilizate anumite date economice referitoare la evoluția fenomenelor analizate. Datele economice reprezintă reflectări cantitative sau calitative ale dimensiunii, stării și evoluției proceselor și fenomenelor economice. Aceste date pot fi descrise pornind de la mai multe criterii. Astfel, ca prezentare, datele pot fi disponibile ca serii de timp (cronologice), de distribuție sau de tip panel.

Seriile de timp corespund unor observații efectuate asupra stării unui fenomen economic la intervale regulate de timp (evoluția cursului de schimb, dinamica prețurilor, a consumului, a veniturilor etc.).

Seriile de distribuție (repartiție), numite și serii în tăietură transversală sau serii de date instantanee reflectă starea, structura și relațiile care există între diferitele componente ale unui sistem economic, la un moment dat. De exemplu, aceste serii reflectă distribuția în spațiu a diferitelor caracteristici ale unui fenomen – șomajul regional, localizarea activităților etc., structura unor agregate (structura sectorială a unei economii, structura ocupării, structura pe elemente a costului de producție ș.a.), sau starea unei variabile la un moment dat într-un anumit eșantion (de exemplu, venitul și consumul famiiliilor).

Datele de tip panel combină seriile de timp și datele în structură transversală. Principalul avantaj al analizei de tip *panel* constă în aceea că permite o mai mare flexibilitate în modelarea diferențelor înregistrate în comportamentele individuale.

Recapitulare

Elementele unui model economico-matematic sunt variabilele, ecuațiile și parametrii modelului. De asemenea, rezolvarea modelului presupune existența unor serii de date, care să prezinte starea și/sau evoluția (distribuția) variabilelor din model.



1.3. Rezumat

În sens restrâns, econometria este definită ca o aplicație a statisticii matematice în economie, astfel încât prin analiza și prelucrarea datelor economice să se ofere un suport empiric modelelor construite de economia matematică. Sau, într-o altă formulare, mai succintă, sarcina de bază a econometriei este să umple empiric structurile teoretice. În sens larg, econometria este înțeleasă ca o știință de graniță între economie, matematică și statistică.

Econometria clasică se concentrează asupra următoarelor tipuri de analiză: (a) testarea și validarea teoriei economice (în sensul confruntării teoriei cu realitatea economică, inclusiv testarea ipotezelor privind comportamentul economic), concret, testarea ideilor, noțiunilor, conceptelor teoretice reprezentate prin asemenea relații; (b) estimarea relațiilor dintre variabilele economice, adică măsurarea fiecărei relații și estimarea parametrilor pe care relațiile respective îi presupun; (c) prognoza evoluțiilor și a comportamentelor economice.

Elementele unui model economico-matematic sunt variabilele, ecuațiile și parametrii modelului. Rezolvarea modelului presupune existența unor serii de date, care să prezinte starea și/sau evoluția (distribuția) variabilelor din model.



1.4. Test de autoevaluare a cunoștințelor

1) În modelul funcției de consum de tip keynesian linear

$$\text{Consum} = C_0 + c \cdot \text{Venit} + e,$$

unde C_0 este consumul autonom și c este înclinația marginală spre consum, semnificația econometrică a variabilelor este următoarea:

- a. *Consumul* este variabila endogenă, *venitul* este variabila exogenă și e este variabila de eroare;
- b. *Consumul* este variabila exogenă, *venitul* este variabila endogenă și e este variabila de eroare;
- c. *Consumul* și *venitul* sunt variabile endogene modelului econometric și e este variabila de eroare;
- d. *Consumul* și *venitul* sunt variabile exogene modelului econometric și e este variabila de eroare
- e. Singura variabilă exogenă din model este variabila de eroare.

2) În modelul funcției de consum de tip

$$(\text{Consum})_t = C_0 + c \cdot (\text{Venit})_t + a \cdot (\text{Consum})_{t-1} + e_t,$$

unde C_0 este consumul autonom, c este înclinația marginală spre consum, iar t este indicele timpului,

Semnificația econometrică a variabilelor este următoarea:

- a. *Consumul* la momentul t și *consumul* la momentul $t-1$ sunt variabile endogene, *venitul* este variabila exogenă și e este variabila de eroare;
- b. *Consumul* la momentul t și *consumul* la momentul $t-1$ sunt variabile exogene, *venitul* este variabila endogenă și e este variabila de eroare;
- c. *Consumul* la momentul t este variabilă endogenă, *venitul* și *consumul* la momentul $t-1$ sunt variabile exogene, iar e este variabila de eroare;

- d. *Consumul* la momentul t , *consumul* la momentul $t-1$ și *venitul* și sunt variabile exogene, iar e este variabila de eroare;
- e. *Consumul* la momentul t , *consumul* la momentul $t-1$ și *venitul* și sunt variabile endogene, iar e este variabila de eroare.

3) În sistemul

Ecuția 1: $(\text{Consum})_t = a_0 + a_1 \cdot (\text{Venit})_t + a_2 \cdot (\text{Consum})_{t-1} + e_t$,

Ecuția 2: $(\text{Venit})_t = (\text{Consum})_t + (\text{Investiții})_t$

unde a_0 , a_1 și a_2 sunt parametrii modelului, iar t este indicele timpului, natura ecuațiilor din model este următoarea:

- a. Ambele ecuații sunt de comportament;
- b. Prima ecuație este de comportament, cea de-a doua este o *relație de echilibru*;
- c. Prima ecuație este de comportament, cea de-a doua este o *relație de definiție*.
- d. Prima ecuație este o *relație de definiție*, cea de-a doua este o *relație de echilibru*;
- e. Ambele ecuații sunt relații de echilibru.



1.5. Bibliografie recomandată

- a. Jula, Dorin; Jula, Nicolae-Marius, 2022. *Econometrie*, ediția a 5^a. Editura Mustang, București
- b. Jula, Nicoleta; Jula, Dorin, 2017. *Modelare economică. Modele econometrice și de optimizare*, Editura Mustang, București

Alternativa web:

- c. Hansen, Bruce E. 2017. *Econometrics*. Download for free at:
<http://www.ssc.wisc.edu/~bhansen/econometrics/Econometrics.pdf>
- d. Diebold, Francis X., 2013-2016. *Econometrics*. Department of Economics, University of Pennsylvania. Download for free at:
<http://www.ssc.upenn.edu/~fdiebold/Teaching104/Econometrics.pdf>

Unitatea de învățare 2.

Modelul linear unifactorial



2.1. Obiective

În cadrul acestui capitol, studenților le sunt prezentate noțiunile de bază privind construcția ecuației de regresie pentru modelul linear factorial și metoda celor mai mici pătrate de rezolvare a modelului.

Durata medie de parcurgere a acestei unități de învățare este de aproximativ 3 ore.



2.2. Conținut

2.2.1. Ecuația de regresie lineară simplă

Acceptăm ipoteza că *variabila endogenă Y depinde linear de variabila exogenă X* și să notăm $M(Y/X)$ valoarea anticipată a lui Y, condiționată de valoarea X. Adică, în ipoteza de linearitate menționată, acceptăm că

$$M(Y / X) = a_0 + a_1X$$

unde a_0 și a_1 sunt parametrii modelului.

Valorile înregistrate ale variabile endogen (Y) nu sunt întotdeauna egale cu valorile anticipate $[M(Y/X)]$ la un nivel dat al variabilei exogene X. Valoarea reală (înregistrată) Y diferă de valoarea anticipată din cauza unor factori de natură diversă. În aceste condiții, pentru ca modelul care evaluează evoluția lui Y să se apropie cât mai mult de realitate, variabila endogenă ar trebui să fie scrisă:

$$Y = M(Y / X) + e,$$

unde e este o variabilă de eroare.

Relația precedentă se numește *ecuația de regresie* a lui Y în funcție de X. Dacă se acceptă ipoteza de linearitate, atunci:

$$Y = a_0 + a_1X + e$$

unde Y este variabila endogenă (variabila explicată prin model), X este variabila exogenă (variabila explicativă), a_0 și a_1 sunt parametrii modelului, iar e reprezintă eroarea sau abaterea dintre valoarea anticipată a endogenei și valoarea efectiv înregistrată. Pentru un eșantion de n înregistrări, ecuația de regresie este:

$$Y_t = a_0 + a_1X_t + e_t, t = 1, 2, \dots, n.$$

Recapitulare

Relația $Y = M(Y / X) + e$ se numește *ecuația de regresie* a lui Y în funcție de X. Dacă se acceptă ipoteza de linearitate, atunci în ecuația de regresie poate fi scrisă într-o formă echivalentă astfel:

$$Y = a_0 + a_1X + e.$$

2.2.2. Metoda celor mai mici pătrate

Prin reprezentarea într-un sistem de axe (XOY) a punctelor de coordonate (X_t, Y_t) se obține un nor de puncte (așa ca în figura 2-1). În același sistem de axe, orice cuplu de parametri (a_0, a_1) determină o dreaptă de ecuație $Y = a_0 + a_1X$.

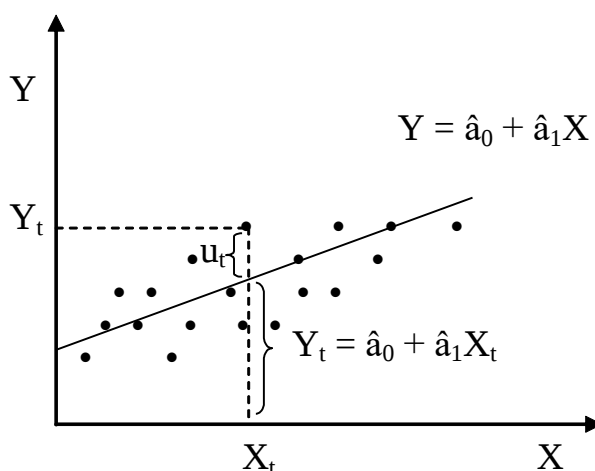


Figura 2-1: Dreapta de regresie

O dreaptă este complet identificată atunci când sunt fixați parametrii a_0 și a_1 . Notăm $\hat{a}_0$, respectiv $\hat{a}_1$ acele valori ale parametrilor pentru care distanța dintre valorile calculate în model și valorile înregistrate efectiv este minimă. Grafic, în figura 2-1, alegem, pentru eșantion, acea dreaptă $\hat{Y} = \hat{a}_0 + \hat{a}_1 X$ pentru care ansamblul abaterilor u_t să fie cât mai mici.

Pentru a evita fenomenul de compensare a erorilor, soluția adoptată, de obicei, constă în însumarea pătratelor abaterilor și determinarea acelor valori ale parametrilor care să ducă la minimizarea sumei respective. De aceea, procedura este cunoscută sub denumirea de *metoda celor mai mici pătrate*.

Grafic, criteriul aplicat în cazul metodei celor mai mici pătrate este următorul:

dreapta care asigură cea mai bună ajustare a punctelor empirice (dreapta de regresie) este aceea pentru care se minimizează suma pătratelor abaterilor dintre punctele de pe grafic și punctele care au aceiași abscisă pe dreapta de regresie, abaterile fiind măsurate vertical.

Analitic, notăm $F(\hat{a}_0, \hat{a}_1)$ suma pătratelor abaterilor u dintre valorile înregistrate ale variabilei Y și valorile calculate $\hat{Y}$. Cu alte cuvinte, $F(\hat{a}_0, \hat{a}_1)$ măsoară suma pătratelor valorilor variabilei reziduale:

$$F(\hat{a}_0, \hat{a}_1) = \sum_{t=1}^n u_t^2 = \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 = \sum_{t=1}^n (Y_t - \hat{a}_0 - \hat{a}_1 X_t)^2$$

În funcția F sunt necunoscuți doar estimatorii $\hat{a}_0$ și $\hat{a}_1$, deoarece valorile X_t și Y_t reprezintă observațiile din eșantionul selectat, deci sunt numere reale cunoscute. Metoda celor mai mici pătrate constă în determinarea, pentru un set dat de observații, a acelor valori $\hat{a}_0$ și $\hat{a}_1$ care minimizează funcția $F(\hat{a}_0, \hat{a}_1)$.

Se poate demonstra¹⁰ că punctul $(\hat{a}_0, \hat{a}_1)$ obținut prin metoda celor mai mici pătrate este soluția *sistemului de ecuații normale*:

¹⁰ Pentru demonstrație vezi: Jula, Dorin, 2003. *Introducere în econometrie*, Editura Professional Consulting, București, 2003, capitolul II, anexa 2.A.1

$$\begin{cases} \sum_{t=1}^n Y_t = n\hat{a}_0 + \hat{a}_1 \sum_{t=1}^n X_t \\ \sum_{t=1}^n X_t Y_t = \hat{a}_0 \sum_{t=1}^n X_t + \hat{a}_1 \sum_{t=1}^n X_t^2 \end{cases}$$

Soluțiile sistemului sunt:

$$\begin{cases} \hat{a}_0 = \bar{Y} - \hat{a}_1 \bar{X} \\ \hat{a}_1 = \frac{\sum_{t=1}^n (X_t - \bar{X})(Y_t - \bar{Y})}{\sum_{t=1}^n (X_t - \bar{X})^2} \end{cases}$$

Modelul estimat este:

$$\hat{Y}_t = \hat{a}_0 + \hat{a}_1 X_t,$$

iar reziduurile din ecuația de regresie sunt:

$$u_t = Y_t - \hat{Y}_t = Y_t - (\hat{a}_0 + \hat{a}_1 X_t).$$

Recapitulare

Dacă ecuația de regresie este $Y_t = a_0 + a_1 X_t + e_t$, atunci parametrii modelului estimați prin metoda celor mai mici pătrate sunt soluții ale sistemului:

$$\begin{cases} \sum_{t=1}^n Y_t = n\hat{a}_0 + \hat{a}_1 \sum_{t=1}^n X_t \\ \sum_{t=1}^n X_t Y_t = \hat{a}_0 \sum_{t=1}^n X_t + \hat{a}_1 \sum_{t=1}^n X_t^2 \end{cases}.$$

2.3. Problemă rezolvată

Fie următoarele serii de date, înregistrate în 20 momente diferite de timp:

t	Y_t	X_t	t	Y_t	X_t
1	4.84	-2.12	11	9.61	-0.38
2	26.27	9.03	12	18.62	4.44
3	18.25	3.54	13	28.90	11.22
4	13.57	3.10	14	20.63	5.45

t	Y _t	X _t	t	Y _t	X _t
5	16.83	4.23	15	14.31	2.27
6	22.53	6.34	16	37.15	14.61
7	17.70	4.28	17	18.30	5.27
8	35.75	11.14	18	23.89	8.31
9	5.46	-2.54	19	20.57	2.89
10	20.69	4.44	20	14.29	2.67

Reprezentați grafic relația dintre cele două variabile și calculați estimatorii din ecuația de regresie, prin metoda celor mai mici pătrate.

Rezolvare:

Scriem modelul econometric în forma:

$$Y_t = a_0 + a_1 X_t + e_t.$$

Legătura dintre cele două variabile este prezentată în figura următoare.

Pentru calculul estimatorilor $\hat{a}_0$ și $\hat{a}_1$ folosim sistemul:

$$\begin{cases} \sum_{t=1}^n Y_t = n\hat{a}_0 + \hat{a}_1 \sum_{t=1}^n X_t \\ \sum_{t=1}^n X_t Y_t = \hat{a}_0 \sum_{t=1}^n X_t + \hat{a}_1 \sum_{t=1}^n X_t^2 \end{cases}$$

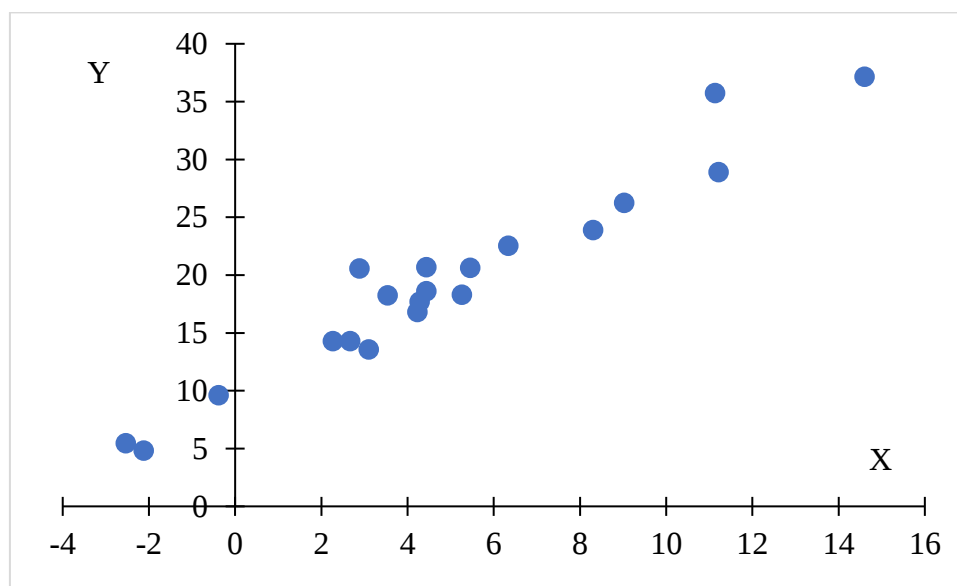


Figura 2-2. Modelul unifactorial

Pornind de la datele din tabel calculăm:

	n	$\sum X_t$	$\sum Y_t$	$\sum X_t Y_t$	$\sum (X_t)^2$
	20	98.17	388.15	2582.65	840.79
Rezultă	$\hat{a}_0 = 10.14316$				
	$\hat{a}_1 = 1.88735$				

Ecuația de regresie este:

$$\hat{Y}_t = 10.14316 + 1.88735 \cdot X_t, \text{ sau } Y_t = 10.14316 + 1.88735 \cdot X_t + u_t.$$

Observație:

Soluția EViews-10 pentru problema precedentă este prezentată în tabelul următor:

Dependent Variable: Y

Method: Least Squares

Sample: 1 20

Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	10.14316	0.707997	14.32656	0.0000
X	1.887350	0.109195	17.28425	0.0000

Soluțiile sunt marcate cu bold. Celelalte informații oferite în tabel vor fi descrise în capitolele următoare.



2.4. Rezumat

Se notează cu Y variabila explicată și X – variabila explicativă. Ecuația de regresie este: $Y = f(X, e)$, unde Y este variabila endogenă, X – variabila exogenă, iar e reprezintă un factor perturbator, de natură aleatoare.

Modelul linear de regresie este $Y = a_0 + a_1 X + e$, iar parametrii modelului, calculați prin metoda celor mai mici pătrate sunt soluții ale sistemului:

$$\begin{cases} \sum_{t=1}^n Y_t = n\hat{a}_0 + \hat{a}_1 \sum_{t=1}^n X_t \\ \sum_{t=1}^n X_t Y_t = \hat{a}_0 \sum_{t=1}^n X_t + \hat{a}_1 \sum_{t=1}^n X_t^2 \end{cases},$$

respectiv

$$\begin{cases} \hat{a}_0 = \bar{Y} - \hat{a}_1 \bar{X} \\ \hat{a}_1 = \frac{\sum_{t=1}^n (X_t - \bar{X})(Y_t - \bar{Y})}{\sum_{t=1}^n (X_t - \bar{X})^2} \end{cases}.$$



2.5. Test de autoevaluare a cunoștințelor

Fie următoarele serii de date, înregistrate în 20 momente diferite de timp:

t	X _t	Y _t	t	X _t	Y _t
1	5.23	24.74	11	3.45	15.71
2	6.23	27.15	12	8.27	34.77
3	2.34	14.09	13	3.59	20.52
4	5.44	25.52	14	6.35	27.92
5	6.49	27.68	15	4.22	20.15
6	6.52	29.77	16	7.51	32.78
7	4.40	21.08	17	6.37	27.52
8	4.70	24.04	18	5.59	24.86
9	6.59	28.34	19	5.40	23.45
10	5.57	20.15	20	3.17	17.53

Reprezentați grafic relația dintre cele două variabile și calculați estimatorii din ecuația de regresie, prin metoda celor mai mici pătrate.



2.6. Bibliografie recomandată

- a. Jula, Dorin; Jula, Nicolae-Marius, 2022. *Econometrie*, ediția a 5^a. Editura Mustang, București

- b.** Jula, Nicoleta; Jula, Dorin, 2017. *Modelare economică. Modele econometrice și de optimizare*, Editura Mustang, București

Alternativa web:

- a.** Hansen, Bruce E. 2017. *Econometrics*. Download for free at:
<http://www.ssc.wisc.edu/~bhansen/econometrics/Econometrics.pdf>
- b.** Diebold, Francis X., 2013-2016. *Econometrics*. Department of Economics, University of Pennsylvania. Download for free at:
<http://www.ssc.upenn.edu/~fdiebold/Teaching104/Econometrics.pdf>

2.7. Teme de control

Realizați un referat de maximum 5 pagini în care să prezentați contribuția la dezvoltarea econometriei a unui laureat al premiului Nobel pentru economie.

Unitatea de învățare 3.

Ipotezele de aplicare ale modelului linear unifactorial. Proprietăți ale estimatorilor



3.1. Obiective

Metoda celor mai mici pătrate duce la calculul celor mai buni estimatori ai modelului unifactorial de regresie lineară, dacă sunt respectate anumite ipoteze de aplicare. În cadrul acestui capitol sunt prezentate ipotezele respective și proprietățile estimatorilor.

Durata medie de parcurgere a acestei unități de învățare este de aproximativ 3 ore.



3.2. Conținut

3.2.1. Ipotezele de aplicare ale modelului linear unifactorial

Estimarea standard, prin metoda celor mai mici pătrate a parametrilor din ecuația de regresie se bazează pe o serie de ipoteze referitoare la forma dependenței dintre variabile, la variabila explicativă și la variabila de abatere.

Ipoteza I-1: *linearitatea modelului.* Oricare ar fi cuplul (X_t, Y_t) , legătura dintre cele două variabile este lineară.

Acceptăm ipoteza că între variabila dependentă (endogenă) Y și variabila independentă, explicativă (exogenă) X există o relație lineară, perturbată însă de variabila de abatere (eroarea) e . Linearitatea se referă la forma legăturii dintre variabile, nu la modul de exprimare a variabilelor respective.

Ipoteza I-2: *variabila X are dispersia nenulă și finită.* Nu toate valorile variabilei exogene sunt egale între ele. Cu alte cuvinte, dispersia de selecție a variabilei X nu este zero, dar este finită.

Dacă nu toate valorile observate pentru X sunt egale între ele, atunci X va avea o dispersie diferită de zero. Dacă la eșantionul inițial se adaugă și alte observații, în mod normal suma pătratelor abaterilor dintre valorile înregistrate și media acestor valori va crește. Ipoteza I-2 presupune că această creștere a dispersiei, atunci când are loc, nu este fără limită. Adică, suma $\sum (X_t - \bar{X})^2$ crește mai puțin rapid decât crește dimensiunea eșantionului.

Ipoteza I-3: (ipoteza de exogeneitate) *variabila X nu este corelată cu erorile din ecuația de regresie.* Variabila X nu este aleatoare. O variantă a acestei ipoteze presupune că X nu este o variabilă aleatoare.

Atunci când X este o variabilă aleatoare, valorile variabilei respective nu trebuie să fie corelate cu erorile. În versiunea strictă a ipotezei, se presupune că sunt înregistrate valorile X_t ale variabilei independente fără erori de observație, iar aceste mărimi sunt păstrate fixe dacă se repetă experimentul (selecția).

Următoarele patru ipoteze (I-4, I-5, I-6 și I-7) se referă la distribuția erorilor.

Ipoteza I-4: *erorile sunt aleatorii, cu media zero.* Erorile et din modelul linear unifactorial sunt variabile aleatoare cu media zero: $M(e_t) = 0$

Între valorile Y_t ale variabilei endogene Y și anticipările realizate pe baza modelului de regresie există abaterile e_t . În cadrul acestor abateri (erori) nu se înregistrează vreo componentă sistematică și/sau o tendință de evoluție.

Ipoteza I-5: *dispersia erorii este constantă.* Erorile e_t sunt identic distribuite, cu o dispersie constantă și finită

Lipsa acestei proprietăți poartă denumirea de *heteroscedasticitate*. Potrivit ipotezei I-5, distribuția variabilei de abatere este relativ constantă, indiferent care ar fi sub-secvența de valori e_t din populația analizată.

Ipoteza I-6: *erorile nu sunt autocorelate*. Covarianța dintre oricare două valori ale variabilei de abatere este zero: e_t și e_p sunt independente, oricare ar fi $t \neq p$

Ipotezele I-6 și I-7 pot fi scrise concentrat astfel: matricea de varianță-covarianță a erorilor este $M(ee') = \sigma_e^2 I$, unde σ_e^2 este valoarea comună a dispersiei erorilor, iar I este matricea unitate. Un caz special important este cel al lipsei autocorelării între termenii succesivi ai valorilor variabilei de abatere, pentru seriile de timp.

Ipoteza I-7: *erorile sunt normal distribuite*. Fiecare variabilă aleatoare e_t este distribuită normal în jurul mediei sale egală cu zero.

Ipoteza I-7 implică faptul că, pentru modelul linear unifactorial, densitatea lui Y condiționată de X este normal distribuită, de medie $a_0 + a_1 X$ și de dispersie constantă σ_e^2 .

Ipotezele de la I-4 la I-7 pot fi însumate în următoarea afirmație: abaterile (erorile) e_t sunt presupuse independente, normal distribuite, cu media zero și dispersia constantă, oricare ar fi t .

Recapitulare

Ipotezele de fundamentare a modelului linear unifactorial de regresie sunt următoarele:

- I-1: Modelul de regresie este linear în raport cu parametrii
- I-2: Variabila exogenă X are dispersia nenulă, dar finită
- I-3: Variabila exogenă X nu este corelată cu erorile din ecuația de regresie (ipoteza de exogeneitate)

- I-4: Erorile et sunt variabile aleatoare cu media zero
- I-5: Erorile nu sunt heteroscedastice
- I-6: Erorile sunt independente (lipsa autocorelării erorilor)
- I-7: Erorile sunt distribuite normal

3.2.2. Proprietăți ale estimatorilor

Pornind de la ipotezele prezentate, pot fi demonstrate o serie de proprietăți ale estimatorilor calculați prin metoda celor mai mici pătrate pentru parametrii modelului linear unifactorial¹¹.

Proprietatea P-1: *estimatorii sunt lineari.* Estimatorii $\hat{\alpha}_0$ și $\hat{\alpha}_1$ din modelul linear unifactorial, calculați prin metoda celor mai mici pătrate, sunt lineari.

Deoarece valorile X sunt constante (I-3) se demonstrează că estimatorii sunt combinații lineare ale valorilor de selecție Y_t .

Proprietatea P-2: *estimatorii sunt nedeplasați.* Dacă variabila exogenă X nu este aleatoare, sau chiar dacă X este variabilă aleatoare, dar este independentă de variabila de abatere e , estimatorii obținuți prin metoda celor mai mici pătrate sunt nedeplasați (nedistorsionați).

Un estimator este nedeplasat (nedistorsionat) dacă media estimatorului este egală cu valoarea parametrului pe care îl estimează.

Proprietatea P-2': *estimatorii sunt consistenți.* Estimatorii $\hat{\alpha}_0$ și $\hat{\alpha}_1$ din modelul linear unifactorial, calculați prin metoda celor mai mici pătrate, sunt consistenți.

Proprietatea P-2' este specifică selecțiilor de volum mare. Un estimator este consistent dacă pe măsură ce numărul observațiilor crește, valoarea estimatorului se apropie de valoarea parametrului estimat. La limită, un estimator este consistent

¹¹ Pentru demonstrația proprietăților vezi Jula, Dorin, 2003. *Introducere în econometrie*, Editura Professional Consulting, București, cap.2, anexele 2.A.6 – 2.A.10, pag.59-70.

dacă, atunci când dimensiunea eșantionului tinde la infinit, valoarea estimatorului tinde spre valoarea parametrului estimat.

Proprietatea P-3: *estimatorii sunt eficienți.* Estimatorii nedeplasați $\hat{a}_0$ și $\hat{a}_1$ din modelul linear unifactorial, calculați prin metoda celor mai mici pătrate, sunt eficienți.

Un estimator al unui parametru este eficient dacă este nedeplasat și nu există un alt estimator nedeplasat care are o dispersie mai mică. Pentru selecțiile de volum mare se analizează proprietatea de eficiență asimptotică. Un estimator este asimptotic eficient dacă este consistent și nu există un alt estimator consistent care are o dispersie mai mică.

Proprietatea P-4: *estimatorii sunt normal distribuiți.* Estimatorii $\hat{a}_0$ și $\hat{a}_1$ din modelul linear unifactorial, calculați prin metoda celor mai mici pătrate, sunt variabile aleatoare cu o distribuție normală.

Dacă potrivit ipotezei I-7, abaterile (erorile) sunt variabile aleatoare distribuite normal atunci distribuția pentru selecții repetate a estimatorilor $\hat{a}_0$ și $\hat{a}_1$ este, de asemenea, normală. În modelul de regresie lineară $Y_t = a_0 + a_1 X_t + e_t$, variabila X_t nu este aleatoare, iar parametrii a_0 și a_1 sunt constante necunoscute.

Se demonstrează că orice combinație lineară a unor variabile distribuite normal este, la rândul său, o variabilă aleatoare distribuită normal. În consecință, estimatorii $\hat{a}_0$ și $\hat{a}_1$ obținuți prin metoda celor mai mici pătrate sunt normal distribuiți.

Proprietatea P-5: *estimatorii sunt de maximă verosimilitate.* Dacă ipotezele I-1 – I-7 ale modelului unifactorial de regresie lineară sunt respectate, atunci estimatorii $\hat{a}_0$ și $\hat{a}_1$ calculați prin metoda celor mai mici pătrate, sunt de maximă verosimilitate.

Dacă, în modelul de regresie, valorile variabilei endogene Y_t depind de valorile X_t prin intermediul unor parametri (necunoscuți), atunci, fixând valorile

X_t , orice set de estimatori ai parametrilor modelului generează o serie de valori $\hat{Y}_t$.

Un estimator este verosimil dacă generează valori plauzibile pentru variabila endogenă. Sunt de maximă verosimilitate acei estimatori care, fiind fixate valorile X_t , generează cu o probabilitate maximă valorile observate ale variabilei endogene Y_t . Se demonstrează că, dacă ipotezele modelului de regresie lineară (de la I-1 la I-7) sunt respectate, atunci estimatorii obținuți prin metoda celor mai mici pătrate sunt de maximă verosimilitate.

Recapitulare

Proprietățile estimatorilor obținuți prin metoda celor mai mici pătrate pentru parametrii din modelul linear unifactorial sunt următoarele:

- P-1: estimatorii sunt lineari
- P-2: estimatorii sunt nedeplasați
- P-2': estimatorii sunt consistenți
- P-3: estimatorii sunt eficienți
- P-4: estimatorii sunt normal distribuiți
- P-5: estimatorii sunt de maximă verosimilitate.

3.2.3. Teorema Gauss -Markov

Proprietatea P-3 scrisă împreună cu precizările privind linearitatea modelului (I-1) și cu proprietățile variabilei exogene și ale variabilei de abatere este cunoscută sub denumirea de teorema Gauss-Markov:

Dacă sunt verificate ipotezele de la I-1 la I-6, atunci estimatorii $\hat{\alpha}_0$ și $\hat{\alpha}_1$ ai parametrilor modelului linear unifactorial sunt cei mai buni estimatori lineari nedeplasați, în sensul că, dintre toți estimatorii lineari nedeplasați, au cea mai mică dispersie.

În literatura de specialitate se folosește expresia BLUE (**B**est **L**inear **U**nbiased **E**stimators) pentru estimatorii parametrilor din modelul de regresie lineară calculați prin metoda celor mai mici pătrate, atunci când estimatorii respectivi îndeplinesc condițiile din teorema Gauss-Markov.

Recapitulare

Dacă modelul de regresie lineară respectă ipotezele I-1 la I-6, atunci estimatorii parametrilor din modelul de regresie lineară calculați prin metoda celor mai mici pătrate sunt cei mai buni estimatori lineari nedeplasați (BLUE).



3.3. Rezumat

Dacă în modelul de regresie lineară, variabila exogenă are dispersia nenulă, dar finită și este independentă față de erori, iar erorile sunt variabile aleatoare independente între ele, normal distribuite, cu medie zero și dispersia constantă, *atunci* estimatorii obținuți prin metoda celor mai mici pătrate sunt lineari, nedeplasați (consistenți), eficienți, normal distribuiți și de maximă verosimilitate.



3.4. Test de autoevaluare a cunoștințelor

1) Modelul $(Y_t)^2 = a_0 + a_1(X_t)^2 + e_t$:

- este nelinear, deoarece variabila exogenă este $(X_t)^2$;
- este nelinear, deoarece variabila endogenă este $(Y_t)^2$;
- este nelinear, deoarece atât variabila exogenă, cât și variabila endogenă sunt la pătrat;
- este linear deoarece legătura dintre variabile este lineară;

- e. este linear deoarece modelul econometric nu ține seama de legătura dintre variabile.

2) Modelul $Y_t = a_0 + a_1 \ln(X_t) + e_t$:

- a. este nelinear, deoarece variabila exogenă este în formă logaritmică;
- b. este nelinear, deoarece variabila endogenă nu este în formă logaritmică;
- c. este linear deoarece legătura dintre variabile este lineară;
- d. este linear deoarece modelul econometric nu ține seama de legătura dintre variabile;
- e. forma prezentată nu permite evaluarea naturii modelului.

3) Dacă, în modelul de regresie lineară $Y_t = a_0 + a_1 X_t + e_t$, variabila X are dispersia nulă, atunci:

- a. variabila X nu are nicio semnificație în model;
- b. variația înregistrată de Y nu este explicată de X, impactul variabilei X se asociază termenului liber din model;
- c. variația înregistrată de Y nu este explicată de X, impactul variabilei X se asociază erorii din model;
- d. modelul econometric nu depinde de variația lui X;
- e. variabila endogenă are, de asemenea, dispersia nulă;



3.5. Bibliografie recomandată

- a. Jula, Dorin; Jula, Nicolae-Marius, 2022. *Econometrie*, ediția a 5^a. Editura Mustang, București
- b. Jula, Nicoleta; Jula, Dorin, 2017. *Modelare economică. Modele econometrice și de optimizare*, Editura Mustang, București

Alternativa web:

- a. Hansen, Bruce E. 2017. *Econometrics*. Download for free at:
<http://www.ssc.wisc.edu/~bhansen/econometrics/Econometrics.pdf>
- b. Diebold, Francis X., 2013-2016. *Econometrics*. Department of Economics, University of Pennsylvania. Download for free at:
<http://www.ssc.upenn.edu/~fdiebold/Teaching104/Econometrics.pdf>

Unitatea de învățare 4. Modelul linear multifactorial



4.1. Obiective

În cadrul acestui capitol sunt extinse noțiunile prezentate în capitolele precedente pentru cazul în care interdependențelor lineare între variabilele economice sunt multiple.

Durata medie de parcurgere a acestei unități de învățare este de aproximativ 3 ore.



4.2. Conținut

4.2.1. Ecuația de regresie lineară multifactorială

Modelul linear unifactorial este extins la un model, tot de regresie lineară, însă cu dependențe multiple, de tipul:

$$Y = a_0 + a_1X_1 + a_2X_2 + \dots + a_kX_k + e,$$

unde Y este variabila endogenă, $X_1, X_2, \dots, X_k$ sunt k variabile explicative, $a_0, a_1, a_2, \dots, a_k$ sunt $k+1$ parametri necunoscuți, iar e este variabila de abatere (eroarea) din ecuația de regresie.

Eroarea e reflectă, la fel ca în modelul linear unifactorial, influența elementelor calitative necuantificabile, a celor care depind de comportamentul uman nepredictibil, sau a altor factori cu influență minoră, alții decât $X_1, X_2, \dots, X_k$.

Parametrul a_0 modelează comportamentul autonom al variabilei endogene, iar parametrii a_i cuantifică intensitatea influenței factorului X_i asupra variabilei Y .

4.2.2. Forma matriceală a modelului linear multifactorial de regresie

Pentru modelul $Y_t = a_0 + a_1X_{1t} + a_2X_{2t} + \dots + a_kX_{kt} + e_t$, scriem X_{it} valoarea variabilei i , în înregistrarea t , unde $i = \overline{1, k}$, iar $t = \overline{1, n}$ astfel:

$$\begin{cases} Y_1 = a_0 + a_1X_{11} + a_2X_{21} + \dots + a_kX_{k1} + e_1 \\ Y_2 = a_0 + a_1X_{12} + a_2X_{22} + \dots + a_kX_{k2} + e_2 \\ Y_3 = a_0 + a_1X_{13} + a_2X_{23} + \dots + a_kX_{k3} + e_3 \\ \vdots \\ Y_t = a_0 + a_1X_{1t} + a_2X_{2t} + \dots + a_kX_{kt} + e_t \\ \vdots \\ Y_n = a_0 + a_1X_{1n} + a_2X_{2n} + \dots + a_kX_{kn} + e_n \end{cases}$$

Introducem următoarele notații:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{pmatrix}, X = \begin{pmatrix} 1 & X_{11} & X_{21} & \dots & X_{k1} \\ 1 & X_{12} & X_{22} & \dots & X_{k2} \\ 1 & X_{13} & X_{23} & \dots & X_{k3} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{1n} & X_{2n} & \dots & X_{kn} \end{pmatrix}, A = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix}, e = \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_n \end{pmatrix}$$

unde:

- Y este un vector coloană, de dimensiuni $n \times 1$, care are drept componente cele n înregistrări ale variabilei explicate (endogene),
- X este o matrice de dimensiuni $n \times (k+1)$, care conține în prima coloană (atașată termenului liber) constanta 1, iar în celelalte k coloane înregistrările pentru fiecare dintre cele k variabile explicative;
- A este un vector coloană, de dimensiuni $(k+1) \times 1$, care include cei $k+1$ parametri ai modelului;
- e este un vector coloană, de dimensiuni $n \times 1$, care include cele n valori ale variabilei de abatere (erorile din ecuație de regresie)

Cu aceste notații, sistemul poate fi scris matriceal astfel:

$$Y = XA + e$$

Notăm estimatorii parametrilor din ecuația de regresie cu $\hat{a}_0, \hat{a}_1, \dots, \hat{a}_k$, unde $\hat{a}_i$ este estimatorul parametrului a_i . Atunci, $\hat{Y}$ – valorile calculate, pe baza modelului, pentru variabila endogenă se deduc din ecuația de regresie lineară multiplă:

$$\hat{Y}_t = \hat{a}_0 + \hat{a}_1 X_{1t} + \hat{a}_2 X_{2t} + \dots + \hat{a}_k X_{kt}.$$

Variabila reziduală (u_t) este diferența dintre valorile înregistrate (Y_t) și valorile estimate ($\hat{Y}_t$):

$$Y_t - \hat{Y}_t = u_t, \text{ oricare ar fi } t = 1, 2, \dots, n$$

sau
$$Y_t = \hat{a}_0 + \hat{a}_1 X_{1t} + \hat{a}_2 X_{2t} + \dots + \hat{a}_k X_{kt} + u_t, t = \overline{1, n}$$

Ecuațiile precedentă poate fi scrise sub formă matriceală astfel:

$$Y = X\hat{A} + u$$

unde $\hat{Y} = X\hat{A}$, iar $\hat{A}$ și u sunt vectorii atașați estimatorilor, respectiv variabilei reziduale.

$$\hat{A} = \begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_k \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_n \end{pmatrix}$$

Valorile $\hat{A}$ și u depind de eșantionul selectat și de metoda de estimare aleasă.

4.2.3. Estimarea parametrilor din modelul linear multifactorial

Cea mai cunoscută procedură de calcul a estimatorilor pentru parametrii modelului linear multifactorial este metoda celor mai mici pătrate. La fel ca în cazul modelului linear unifactorial, se caută acele valori ale estimatorilor care minimizează suma pătratelor reziduurilor (diferențelor dintre valorile înregistrate statistic pentru Y și valorile estimate pe baza ecuației de regresie). Adică, sunt alese acele valori $\hat{a}_0, \hat{a}_1, \dots, \hat{a}_k$, pentru care valoarea funcției F este minimă, unde F se calculează astfel:

$$F = \sum_{t=1}^n u_t^2 = \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 = \sum_{t=1}^n (Y_t - \hat{a}_0 - \hat{a}_1 X_{1t} - \hat{a}_2 X_{2t} - \dots - \hat{a}_k X_{kt})^2$$

Expresia precedentă se scrie matriceal $F = u'u$, unde prin u' am notat transpusul vectorului u , unde $u = Y - X\hat{A}$. Scriem succesiv:

$$\begin{aligned} F = u'u &= (Y - X\hat{A})'(Y - X\hat{A}) = Y'Y - Y'X\hat{A} - \hat{A}'X'Y + \hat{A}'X'X\hat{A} = \\ &= Y'Y - 2\hat{A}'X'Y + \hat{A}'X'X\hat{A} \end{aligned}$$

Pentru determinarea punctului de minim se egalează cu zero derivata lui F în raport cu vectorul estimatorilor $\hat{A}$:

$$\frac{\partial F}{\partial \hat{A}} = -2X'Y + 2X'X\hat{A} = 0$$

Expresia matriceală precedentă este scrierea concentrată a $k+1$ ecuații în care elementele necunoscute sunt componentele vectorului $\hat{A}$:

$$X'X\hat{A} = X'Y$$

Relația de mai sus reprezintă sistemul de ecuații normale din modelul linear multifactorial. Dacă vectorii atașați fiecărei variabile explicative X_i sunt linear independenți, atunci matricea $X'X$ nu este singulară și sistemul poate fi rezolvat în raport cu vectorul estimatorilor $\hat{A}$:

$$\hat{A} = (X'X)^{-1}X'Y$$

Din condițiile de ordinul II rezultă¹² că vectorul $\hat{A}$ minimizează valoarea funcției $F(\hat{A}) = u'u$. Expresia respectivă reprezintă formula de calcul a estimatorilor pentru parametrii din modelul linear multifactorial.

4.3. Problemă rezolvată

Fie următoarele serii de date, înregistrate în 20 momente diferite de timp:

¹² Vezi Jula, Dorin, 2003. *Introducere în econometrie*, Editura Professional Consulting, București, cap.2, anexa 3.A.2, pag. 123.

t	Y_t	X_{1t}	X_{2t}	t	Y_t	X_{1t}	X_{2t}
1	17.71	5.23	3.02	11	10.12	3.45	1.74
2	15.26	6.23	4.09	12	15.66	8.27	6.27
3	11.48	2.34	1.42	13	8.47	3.59	4.45
4	37.95	5.44	-1.81	14	11.85	6.35	5.23
5	28.27	6.49	0.91	15	17.08	4.22	1.64
6	15.26	6.52	5.18	16	29.68	7.51	2.34
7	0.75	4.40	6.05	17	31.05	6.37	0.22
8	25.98	4.70	1.00	18	3.25	5.59	6.43
9	18.01	6.59	3.72	19	24.73	5.40	0.50
10	17.39	5.57	0.55	20	14.90	3.17	1.66

Reprezentați grafic variabilele și calculați estimatorii din ecuația de regresie, prin metoda celor mai mici pătrate.

Rezolvare:

Scriem modelul econometric în forma $Y_t = a_0 + a_1X_{1t} + a_2X_{2t} + e_t$, sau, matriceal: $Y = XA + e$.

Variabile Y_t , X_{1t} și X_{2t} sunt prezentate grafic în figura următoare. Pentru calculul

estimatorilor $\hat{a}_0$, $\hat{a}_1$ și $\hat{a}_2$ folosim relația $\hat{A} = (X'X)^{-1}X'Y$.

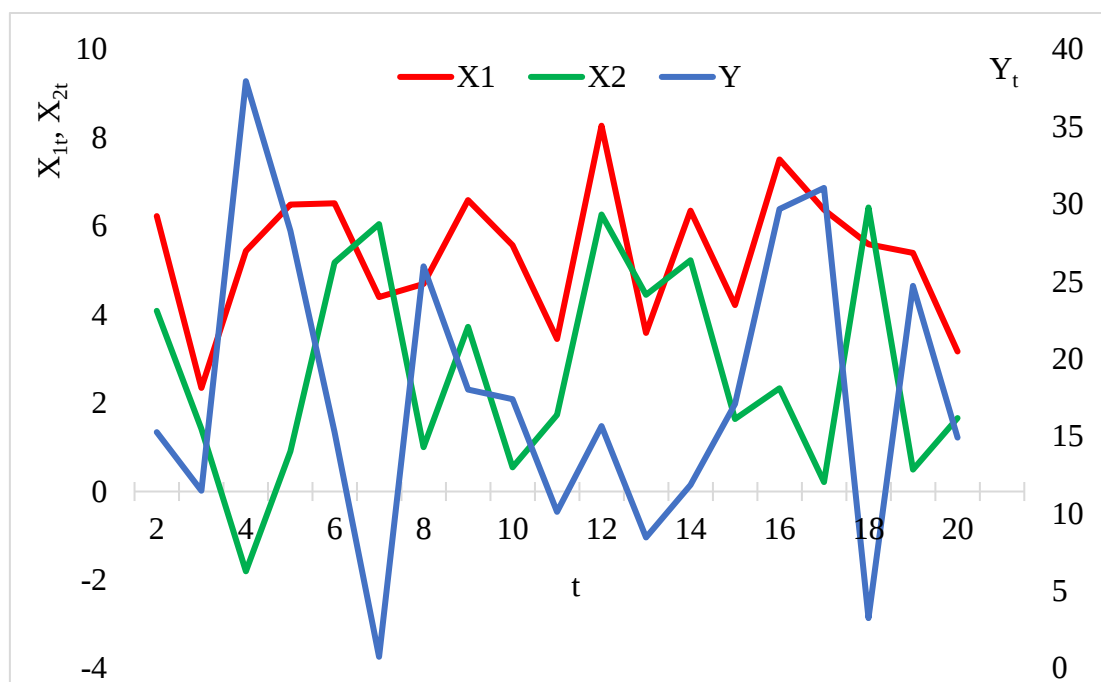


Figura 4-1. Modelul bi-factorial de regresie

Elementele folosite în calcul sunt următoarele:

(X'X)			(X'X) ⁻¹			X'Y
20.000	107.431	54.6013	0.7176	-0.1209	-0.0067	354.855
107.431	620.576	310.124	-0.1209	0.0245	-0.0040	2006.754
54.6013	310.124	252.548	-0.0067	-0.0040	0.0103	652.9018

Rezultă $\hat{A} = (X'X)^{-1}X'Y$, adică

$$\hat{A} = \begin{matrix} 7.70335 \\ 3.72858 \\ -3.65884 \end{matrix}$$

Modelul estimat este:

$$Y_t = 7.70335 + 3.72858 \cdot X_{1t} - 3.65884 \cdot X_{2t}.$$

Soluția EViews-10 pentru problema precedentă este prezentată în tabelul următor:

Dependent Variable: Y
Method: Least Squares
Sample: 1 20

Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.703353	2.654574	2.901917	0.0099
X ₁	3.728578	0.490786	7.597154	0.0000
X ₂	-3.658842	0.318214	-11.49806	0.0000

Soluția este prezentată cu bold, în tabelul precedent.

Recapitulare

Modelul multifactorial de regresie lineară reprezintă o extindere a modelului linear unifactorial de regresie pentru situația în care există un număr mai mare de factori de influență ai variabile endogene.

Modelul multifactorial de regresie lineară

$$Y_t = a_0 + a_1X_{1t} + a_2X_{2t} + \dots a_kX_{kt} + e_t,$$

poate fi scris concentrat (matriceal) astfel:

$$Y = XA + e.$$

Estimatorii parametrilor din modelul multifactorial de regresie lineară $Y = XA + e$ sunt soluții ale ecuației $\hat{A} = (X'X)^{-1}X'Y$.



4.4. Rezumat

Cea mai cunoscută procedură de calcul a estimatorilor pentru parametrii modelului linear multifactorial este metoda celor mai mici pătrate. Metoda celor mai mici pătrate caută acele valori ale estimatorilor care minimizează suma pătratelor reziduurilor (diferențelor dintre valorile înregistrate statistic pentru Y și valorile estimate pe baza ecuației de regresie).

Estimatorii parametrilor din modelul multifactorial de regresie lineară $Y = XA + e$ se calculează ca soluții ale ecuației

$$\hat{A} = (X'X)^{-1}X'Y.$$



4.5. Test de autoevaluare a cunoștințelor

Fie următoarele serii de date, înregistrate în 20 momente diferite de timp:

t	Y_t	X_{1t}	X_{2t}	t	Y_t	X_{1t}	X_{2t}
1	23.96	5.23	3.02	11	13.31	3.45	1.74
2	23.57	6.23	4.09	12	28.20	8.27	6.27
3	13.25	2.34	1.42	13	14.51	3.59	4.45
4	39.58	5.44	-1.81	14	21.44	6.35	5.23
5	33.67	6.49	0.91	15	20.94	4.22	1.64
6	24.96	6.52	5.18	16	37.53	7.51	2.34
7	9.20	4.40	6.05	17	35.64	6.37	0.22
8	29.68	4.70	1.00	18	13.26	5.59	6.43
9	26.33	6.59	3.72	19	28.62	5.40	0.50
10	21.51	5.57	0.55	20	17.74	3.17	1.66

Calculați estimatorii din ecuația de regresie, prin metoda celor mai mici pătrate.



4.6. Bibliografie recomandată

- a. Jula, Dorin; Jula, Nicolae-Marius, 2022. *Econometrie*, ediția a 5^a. Editura Mustang, București
- b. Jula, Nicoleta; Jula, Dorin, 2017. *Modelare economică. Modele econometrice și de optimizare*, Editura Mustang, București

Alternativa web:

- a. Hansen, Bruce E. 2017. *Econometrics*. Download for free at:
<http://www.ssc.wisc.edu/~bhansen/econometrics/Econometrics.pdf>
- b. Diebold, Francis X., 2013-2016. *Econometrics*. Department of Economics, University of Pennsylvania. Download for free at:
<http://www.ssc.upenn.edu/~fdiebold/Teaching104/Econometrics.pdf>

Unitatea de învățare 5.

Ipotezele de aplicare ale modelului linear multifactorial. Proprietăți ale estimatorilor



5.1. Obiective

Metoda celor mai mici pătrate duce la calculul celor mai buni estimatori ai modelului multifactorial de regresie lineară, dacă sunt respectate anumite ipoteze de aplicare. În cadrul acestui capitol sunt prezentate ipotezele respective și proprietățile estimatorilor.

Durata medie de parcurgere a acestei unități de învățare este de aproximativ 3 ore.



5.2. Conținut

5.2.1. Ipotezele de aplicare ale modelului linear multifactorial

Estimarea parametrilor din ecuația de regresie multifactorială se bazează, la fel ca în cazul unifactorial, pe o serie de ipoteze referitoare la forma dependenței dintre variabile, la variabila explicativă și la variabila de abatere.

I-1M: *Linearitatea modelului.* Modelul este linear în sensul că oricare ar fi înregistrarea ($Y_t, X_{1t}, X_{2t}, \dots, X_{kt}$) selectată, forma legăturii dintre Y_t variabilele explicative X_{kt} și variabila de abatere este lineară.

Linearitatea este în raport cu parametrii și nu neapărat în raport cu variabilele exogene.

I-2M: *Ipotezele referitoare la variabilele explicative*

- a. Exogeneitate: variabilele explicative nu sunt corelate cu erorile din ecuația de regresie. Variantă: variabilele explicative nu sunt aleatoare, au valorile fixate atunci când se repetă selecția.
- b. Fiecare variabilă exogenă are dispersia nenulă, dar finită.
- c. Numărul de observații este superior numărului de parametri.
- d. Nu există nici o relație lineară între două sau mai multe variabile explicative (absența colinearității).

Ipotezele I-2Mc și I-2Md pot fi scrise concentrat astfel: matricea X este de rang $k+1$.

Ipotezele referitoare la variabila de abatere (la erori) sunt similare celor prezentate pentru modelul unifactorial de regresie lineară:

I-3M: Ipotezele referitoare la erori

- a. Erorile e_t au media nulă
- b. Erorile e_t au dispersia constantă oricare ar fi t (erorile nu sunt heteroscedastice)
- c. Erorile e_t sunt independente (nu sunt autocorelate)
- d. Erorile e_t sunt normal distribuite

Ipotezele I-3Mb și I-3Mc pot fi scrise concentrat astfel: matricea de varianță-covarianță a erorilor este $M(ee') = \sigma_e^2 I$, unde σ_e^2 este valoarea comună a dispersiei erorilor, iar I este matricea unitate.

5.2.2. Proprietăți ale estimatorilor

Dacă ipotezele modelului sunt respectate, atunci estimatorii *calculați prin metoda celor mai mici pătrate* pentru modelul multifactorial de regresie lineară au anumite proprietăți.

Proprietatea P-1M: *estimatorii sunt lineari.* Dacă variabilele explicative X nu sunt aleatoare și au valorile fixate (atunci când se repetă selecția), estimatorii parametrilor din modelul multifactorial de regresie lineară sunt funcții lineare de observațiile din eșantionul selectat.

Potrivit relației de calcul a estimatorilor pentru parametrii modelului multifactorial, $\hat{A} = (X'X)^{-1}X'Y$. Deci $\hat{A}$ este combinație lineară a valorilor Y din eșantion, cu ponderi care depind de valorile înregistrate pentru variabilele explicative.

Proprietatea P-2M: *estimatorii sunt nedeplasați.* Dacă erorile sunt variabile aleatoare cu media zero, estimatorii lineari ai parametrilor din modelul multifactorial de regresie lineară sunt nedeplasați.

Calculăm succesiv:

$$\begin{aligned}\hat{A} &= (X'X)^{-1}X'Y = (X'X)^{-1}X'(XA + e) = (X'X)^{-1}X'XA + (X'X)^{-1}X'e \rightarrow \\ \hat{A} &= A + (X'X)^{-1}X'e\end{aligned}$$

În relația precedentă se aplică operatorul de medie:

$$\begin{aligned}M(\hat{A}) &= A + (X'X)^{-1}X'M(e) = A \\ M(\hat{A}) &= M[A + (X'X)^{-1}X'e] = M(A) + M[(X'X)^{-1}X'e] = \\ &= A + (X'X)^{-1}X'M(e)\end{aligned}$$

adică estimatorii sunt nedeplasați.

Proprietatea P-2'M: *estimatorii sunt consistenți.* Dacă variabilele explicative au dispersia finită, estimatorii lineari și nedeplasați ai parametrilor din modelul multifactorial de regresie lineară sunt consistenți.

La fel ca în cazul modelului linear unifactorial, proprietatea de consistență se aplică doar pentru selecțiile de volum mare. Consistența unui estimator $\hat{a}_i$ presupune că, atunci când dimensiune eșantionului crește, valoarea estimatorului

tinde asimptotic spre valoarea parametrului a_i . La limită, când $n \rightarrow \infty$, $\hat{a}_i$ tinde în probabilitate spre a_i . Se poate demonstra că, în ipotezele prezentate, estimatorii calculați prin metoda celor mai mici pătrate pentru parametrii modelului sunt consistenți¹³.

Proprietatea P-3M: *estimatorii sunt eficienți.* Dacă variabilele explicative au dispersia finită, estimatorii lineari și nedeplasați ai parametrilor din modelul multifactorial de regresie lineară sunt eficienți.

La fel ca în cazul modelului unifactorial de regresie lineară, un estimator al unui parametru este eficient dacă nu există un alt estimator nedeplasat care are o dispersie mai mică. Se poate demonstra că, în ipotezele obișnuite, estimatorii obținuți prin metoda celor mai mici pătrate sunt eficienți¹⁴.

Proprietatea P-4M: *estimatorii sunt normal distribuiți.* Estimatorii parametrilor din modelul multifactorial de regresie lineară ($\hat{A}$), calculați prin metoda celor mai mici pătrate, sunt variabile aleatoare cu distribuție normală.

Estimatorii $\hat{A}$ sunt combinații lineare ale variabilei endogene Y . Însă, Y este o variabilă aleatoare calculată ca sumă între constantele XA (deși necunoscuți, parametrii A sunt fixați, iar elementele matricei X nu sunt variabile aleatoare) și variabila aleatoare e repartizată normal. Rezultă că valorile Y sunt variabile aleatoare repartizate normal și, în consecință, vectorul $\hat{A}$ conține elemente care sunt variabile aleatoare repartizate normal.

¹³ Pentru demonstrație vezi: Vangrevellinghe G., 1973, *Econométrie*, Hermann, Paris, pag. 41-44 și Thomas R.-L., 1997, *Modern Econometrics: An Introduction*, 2nd edition, Harlow, Longman, pag. 202-203.

¹⁴ Pentru demonstrație vezi: Spanos A., 1986, *Statistical foundations of econometric modeling*, Cambridge University Press, pag. 386-392.

Proprietatea P-5M: *estimatorii sunt de maximă verosimilitate.* Estimatorii parametrilor din modelul multifactorial de regresie lineară ($\hat{A}$), calculați prin metoda celor mai mici pătrate, sunt estimatori de maximă verosimilitate.

Un estimator este verosimil dacă generează valori plauzibile pentru variabila endogenă. În ipotezele menționate, estimatorii obținuți prin metoda celor mai mici pătrate sunt de maximă verosimilitate.

Recapitulare

În sinteză, ipotezele de aplicare ale modelului linear multifactorial sunt următoarele:

I-1M: *Linearitatea modelului.*

I-2M: *Ipotezele referitoare la variabilele explicative*

- a. Exogeneitate: variabilele explicative nu sunt corelate cu erorile din ecuația de regresie. Variantă: variabilele explicative nu sunt aleatoare, au valorile fixate atunci când se repetă selecția.
- b. Fiecare variabilă exogenă are dispersia nenulă, dar finită.
- c. Numărul de observații este superior numărului de parametri.
- d. Nu există nici o relație lineară între două sau mai multe variabile explicative (absența colinearității).

I-3M: *Ipotezele referitoare la erori*

- a. Erorile e_t au media nulă
- b. Erorile e_t au dispersia constantă oricare ar fi t (erorile nu sunt heteroscedastice)
- c. Erorile e_t sunt independente (nu sunt autocorelate)
- d. Erorile e_t sunt normal distribuite

În sinteză, proprietățile estimatorilor din modelul linear multifactorial de regresie sunt următoarele:

P-1M: *estimatorii sunt lineari.*

P-2M: *estimatorii sunt nedeplasați.*

P-2'M: *estimatorii sunt consistenți.*

P-3M: *estimatorii sunt eficienți.*

P-4M: *estimatorii sunt normal distribuiți.*

P-5M: *estimatorii sunt de maximă verosimilitate.*



5.3. Rezumat

Estimarea parametrilor din ecuația de regresie multifactorială se bazează, la fel ca în cazul unifactorial, pe o serie de ipoteze referitoare la forma dependenței dintre variabile, la variabila explicativă și la variabila de abatere.

Ipotezele referitoare la variabilele explicative sunt similare cu I-2 și I-3 din modelul linear unifactorial, cu precizarea că, în cazul modelului multifactorial, ipotezele respective se aplică fiecărei variabile exogene.

În plus, se adaugă două prezumții suplimentare: numărul observațiilor este mai mare decât numărul variabilelor explicative și, oricare ar fi două sau mai multe variabile explicative, nu există o relație lineară între valorile de selecție ale acestora.

Dacă în modelul de regresie lineară, variabilele exogene au dispersia nenulă, dar finită și sunt independentă față de erori și mutual independente, iar erorile sunt variabile aleatoare independente între ele, normal distribuite, cu medie zero și dispersia constantă, *atunci* estimatorii obținuți prin metoda celor mai mici

pătrate sunt lineari, nedeplasați (consistenți), eficienți, normal distribuiți și de maximă verosimilitate.



5.4. Test de autoevaluare a cunoștințelor

1) Modelul de tip Cobb-Douglas $Y_t = A_0 K_t^\alpha L_t^\beta e^{\varepsilon_t}$:

- a. este linear în parametrii;
- b. este nelinear în parametrii;
- c. este nelinear datorită formei exponențiale de exprimare a dependenței dintre variabile;
- d. este nelinear deoarece variabilele exogene și variabila endogenă au forme diferite de exprimare;
- e. este linear deoarece modelul econometric nu ține seama de legătura dintre variabile.

2) Modelul de tip Cobb-Douglas $Y_t = A_0 K_t^\alpha L_t^\beta + \varepsilon_t$:

- a. este linear în parametrii;
- b. este nelinear în parametrii;
- c. este nelinear datorită formei exponențiale de exprimare a dependenței dintre variabile;
- d. este nelinear deoarece variabilele exogene și variabila endogenă au forme diferite de exprimare;
- e. este linear deoarece modelul econometric nu ține seama de legătura dintre variabile.

3) Model de regresie $Y_t = a_0 + a_1 X_{1t} + a_2 \ln(X_{2t}) + e_t$:

- a. este nelinear datorită formei de exprimare a dependenței dintre variabile;
- b. este nelinear deoarece variabilele exogene și variabila endogenă au forme diferite de exprimare;
- c. este linear deoarece modelul econometric nu ține seama de legătura dintre variabile;
- d. este linear în parametrii;
- e. este nelinear în parametrii.



5.5. Bibliografie recomandată

- a. Jula, Dorin; Jula, Nicolae-Marius, 2022. *Econometrie*, ediția a 5^a. Editura Mustang, București
- b. Jula, Nicoleta; Jula, Dorin, 2017. *Modelare economică. Modele econometrice și de optimizare*, Editura Mustang, București

Alternativa web:

- a. Hansen, Bruce E. 2017. *Econometrics*. Download for free at:
<http://www.ssc.wisc.edu/~bhansen/econometrics/Econometrics.pdf>
- b. Diebold, Francis X., 2013-2016. *Econometrics*. Department of Economics, University of Pennsylvania. Download for free at:
<http://www.ssc.upenn.edu/~fdiebold/Teaching104/Econometrics.pdf>

Unitatea de învățare 6.

Testarea semnificației parametrilor din modelul linear unifactorial de regresie



6.1. Obiective

Estimatorii parametrilor din ecuația de regresie sunt determinați pe baza unei selecții și, în consecință, sunt variabile aleatoare. Ca urmare, pot lua, cu probabilități diferite, orice valoare într-un anumit interval. În cadrul acestui capitol este prezentată, pentru modelul linear unifactorial de regresie, procedura de testare a riscului ca parametrii să fie zero, adică a riscului să nu existe o legătură între variabila endogenă (Y) și variabila exogenă (X).

Durata medie de parcurgere a acestei unități de învățare este de aproximativ 3 ore.



6.2. Conținut

6.2.1. Dispersia estimatorilor în modelul linear unifactorial

Se poate demonstra că o estimare nedeplasată a dispersiei erorilor, calculată pornind de la dispersia de selecție a variabilei reziduale, este dată de expresia¹⁵:

$$s_u^2 = \frac{\sum_{t=1}^n u_t^2}{n-2}.$$

¹⁵ Pentru demonstrație vezi Jula, Dorin, 2003. *Introducere în econometrie*, Editura Professional Consulting, București, cap. II, Anexa 2.A.13, pag.74-75.

Dacă se respectă ipotezele I-5 (erorile nu sunt *heteroscedastice*) și I-6 (erorile nu sunt *autocorelate*), atunci se pot calcula estimările nedeplasate pentru dispersiile estimatorilor din modelul unifactorial de regresie lineară:

$$s_{\hat{\alpha}_0}^2 = s_u^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_t - \bar{X})^2} \right)$$

$$s_{\hat{\alpha}_1}^2 = s_u^2 \frac{1}{\sum (X_t - \bar{X})^2}$$

Abaterile standard ale variabilelor aleatoare u , $\hat{\alpha}_0$ și $\hat{\alpha}_1$ se calculează prin extragerea rădăcinii pătrate din valorile corespunzătoare ale dispersiilor:

$$s_u = \sqrt{\frac{\sum_{t=1}^n u_t^2}{n-2}}$$

$$s_{\hat{\alpha}_0} = s_u \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_t - \bar{X})^2}}$$

$$s_{\hat{\alpha}_1} = \frac{s_u}{\sqrt{\sum (X_t - \bar{X})^2}}.$$

6.2.2. Teste de semnificație în cazul modelului unifactorial

Testele de semnificație presupun parcurgerea următorului algoritm:

Pasul 1: Se formulează nulă $H_0: \hat{\alpha}_0 = 0$ și ipoteza alternativă $H_1: \hat{\alpha}_0 > 0$, pentru estimatorul $\hat{\alpha}_0$, respectiv $H_0: \hat{\alpha}_1 = 0$ și $H_1: \hat{\alpha}_1 > 0$, pentru estimatorul $\hat{\alpha}_1$.

Pasul 2: Pornind de la datele de selecție se calculează $t_{\hat{\alpha}_0} = \frac{\hat{\alpha}_0}{s_{\hat{\alpha}_0}}$ și $t_{\hat{\alpha}_1} = \frac{\hat{\alpha}_1}{s_{\hat{\alpha}_1}}$. Sub

ipoteza nulă, statisticile respective urmează o distribuție de tip t-Student cu $n - 2$ grade de libertate.

Pasul 3: Din tabelul statisticii t-Student, pornind de la numărul gradelor de libertate ($n - 2$) și de la nivelul de semnificație ales (α), se selectează o valoare $t_{n-2,\alpha}^*$ astfel încât $P(t > t^*) = \alpha$ (testul unilateral)

Pasul 4: Dacă valoarea calculată t este mare admitem că, probabil, parametrul a_0 , respectiv a_1 este semnificativ mai mare decât zero. Concret, dacă $t_{\hat{a}_0} > t_{n-2,\alpha}^*$ atunci se respinge H_0 și admitem că parametrul a_0 este semnificativ mai mare decât zero, cu un grad de încredere mai mare decât $1 - \alpha$. La fel, procedăm pentru parametrul a_1 .

Observație:

Dacă estimatorul $\hat{a}_0$ este negativ, atunci ipoteza alternativă este $H_1: a_0 < 0$ și se respinge H_0 dacă $t_{\hat{a}_0} < -t_{n-2,\alpha}^*$. O procedură mai simplă este: se calculează estimatorul în modul (în valoare absolută) și se urmează algoritmul în 4 pași prezentat. Evident, dacă $\hat{a}_1$ este negativ, se procedează în mod similar.

Recapitulare

Abaterile standard ale variabilelor aleatoare u , $\hat{a}_0$ și $\hat{a}_1$ se calculează prin extragerea rădăcinii pătrate din valorile corespunzătoare ale dispersiilor:

$$s_u = \sqrt{\frac{\sum_{t=1}^n u_t^2}{n-2}}$$

$$s_{\hat{a}_0} = s_u \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_t - \bar{X})^2}}$$

$$s_{\hat{a}_1} = \frac{s_u}{\sqrt{\sum (X_t - \bar{X})^2}}.$$

Pentru testarea semnificație estimatorilor se calculează statisticile t_{calc} care se compară cu valorile corespunzătoare din distribuția teoretică t-Student.

6.3. Problemă rezolvată

Fie următoarele serii de date, înregistrate în 20 momente diferite de timp:

t	Y _t	X _t	t	Y _t	X _t
1	4.84	-2.12	11	9.61	-0.38
2	26.27	9.03	12	18.62	4.44
3	18.25	3.54	13	28.90	11.22
4	13.57	3.10	14	20.63	5.45
5	16.83	4.23	15	14.31	2.27
6	22.53	6.34	16	37.15	14.61
7	17.70	4.28	17	18.30	5.27
8	35.75	11.14	18	23.89	8.31
9	5.46	-2.54	19	20.57	2.89
10	20.69	4.44	20	14.29	2.67

Observație: datele sunt cele din problema rezolvată la capitolul 2.3.

Testați semnificația parametrilor din modelul linear unifactorial de regresie: $Y_t = a_0 + a_1X_t + e_t$.

Rezolvare:

Estimatorii parametrilor din ecuația de regresie sunt: $\hat{a}_0 = 10.14316$ și $\hat{a}_1 = 1.88735$ (vezi problema rezolvată de la cap. 2.3). Aplicarea algoritmului de testare a semnificației parametrilor presupune calculul, în prealabil, a dispersiei estimatorilor:

$$s_{\hat{a}_0} = s_u \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_t - \bar{X})^2}} = 0.70800$$

$$s_{\hat{a}_1} = \frac{s_u}{\sqrt{\sum (X_t - \bar{X})^2}} = 0.10920$$

Pornind de la aceste valori, determinăm statisticile t :

$$t_{\hat{a}_0} = \frac{\hat{a}_0}{s_{\hat{a}_0}} = \frac{10.14316}{0.70800} = 14.327$$

și
$$t_{\hat{a}_1} = \frac{\hat{a}_1}{s_{\hat{a}_1}} = \frac{1.88735}{0.10920} = 17.284.$$

Valorile corespunzătoare din tabelul distribuției teoretice unilaterale t-Student pentru $n-2 = 20-2 = 18$ grade de libertate sunt:

α	0.10	0.05	0.01
$t_{18;\alpha}$	1.330	1.734	2.552

În evaluarea semnificației parametrilor din model, pragul standard este de 0.05 (5%). Deoarece $t_{\hat{a}_0} = 14.327 > 1.734$ și $t_{\hat{a}_1} = 17.284 > 1.734$, riscul ca parametrii a_0 și a_1 să nu difere semnificativ de zero este mai mic de 5%. De altfel, riscul este mai mic decât 1% (pragul pentru 1% este 2.552 și statisticile calculate sunt superioare acestui prag).

Soluția EViews-10 pentru problema precedentă este prezentată în tabelul următor:

Dependent Variable: Y

Method: Least Squares

Sample: 1 20

Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	10.14316	0.707997	14.32656	0.0000
X	1.887350	0.109195	17.28425	0.0000

Riscul ca estimatorii să fie zero este prezentat (cu bold), în ultima coloană a tabelului.



6.4. Rezumat

Pentru testarea semnificației parametrilor din modelul linear unifactorial de regresie se calculează dispersia de selecție a estimatorilor, apoi raportul dintre valoarea estimatorilor și abaterea standard. Raportul respectiv se compară cu valoarea corespunzătoare din distribuția unilaterală t-Statistic.



6.5. Test de autoevaluare a cunoștințelor

Fie următoarele serii de date:

t	X _t	Y _t	t	X _t	Y _t
1	5.23	24.74	11	3.45	15.71
2	6.23	27.15	12	8.27	34.77
3	2.34	14.09	13	3.59	20.52
4	5.44	25.52	14	6.35	27.92
5	6.49	27.68	15	4.22	20.15
6	6.52	29.77	16	7.51	32.78
7	4.40	21.08	17	6.37	27.52
8	4.70	24.04	18	5.59	24.86
9	6.59	28.34	19	5.40	23.45
10	5.57	20.15	20	3.17	17.53

Testați semnificația parametrilor din modelul linear unifactorial de regresie: $Y_t = a_0 + a_1X_t + e_t$.



6.6. Bibliografie recomandată

- a. Jula, Dorin; Jula, Nicolae-Marius, 2022. *Econometrie*, ediția a 5^a. Editura Mustang, București
- b. Jula, Nicoleta; Jula, Dorin, 2017. *Modelare economică. Modele econometrice și de optimizare*, Editura Mustang, București

Alternativa web:

- a. Hansen, Bruce E. 2017. *Econometrics*. Download for free at:
<http://www.ssc.wisc.edu/~bhansen/econometrics/Econometrics.pdf>
- b. Diebold, Francis X., 2013-2016. *Econometrics*. Department of Economics, University of Pennsylvania. Download for free at:
<http://www.ssc.upenn.edu/~fdiebold/Teaching104/Econometrics.pdf>

Unitatea de învățare 7.

Testarea semnificației parametrilor din modelul linear multifactorial de regresie



7.1. Obiective

În cadrul acestui capitol este prezentată, pentru modelul linear multifactorial de regresie, procedura de testare a riscului ca parametrii să fie zero, adică a riscului să nu existe o legătură între variabila endogenă (Y) și variabilele exogene ($X_1, X_2, \dots, X_k$). Analiza extinde calculele prezentate în capitolul precedent.

Durata medie de parcurgere a acestei unități de învățare este de aproximativ 3 ore.



7.2. Conținut

7.2.1. Dispersia estimatorilor în modelul linear multifactorial

În formula de calcul a estimatorilor,

$$\hat{A} = (X'X)^{-1}X'Y,$$

se înlocuiește Y cu expresia sa,

$$Y = XA + e.$$

Rezultă

$$\hat{A} = (X'X)^{-1}X'(XA + e) = A + (X'X)^{-1}X'e.$$

Matricea de covarianță a estimatorilor este dată de relația

$$V(\hat{A}) = E[(\hat{A} - A)(\hat{A} - A)'] = E[(X'X)^{-1}X'ee'X(X'X)^{-1}]$$

unde $E(.)$ este operatorul de calcul al speranței matematice (engl. *mathematical expectation*). Dacă erorile sunt independente de variabilele explicative din model, nu sunt autocorelate și nu sunt heteroscedastice, atunci:

$$V(\hat{A}) = \sigma_e^2 (X'X)^{-1}$$

unde σ_e^2 este dispersia erorilor. Se demonstrează că dispersia de selecție a reziduurilor:

$$s_u^2 = \frac{\sum_{t=1}^n u_t^2}{n - k - 1}$$

este un estimator nedeplasat al dispersiei erorilor. În aceste condiții, un estimator nedeplasat al matricei $V(\hat{A})$, matricea de covarianță a vectorului estimatorilor $\hat{A}$, se calculează astfel: $S_{\hat{A}}^2 = s_u^2 (X'X)^{-1}$. Dispersia de selecție pentru estimatorii parametrilor din ecuația de regresie se găsește pe diagonala principală a matricei $S_{\hat{A}}^2$. Elementele ne-diagonale sunt covarianțele dintre estimatori. *Abaterea standard de selecție a estimatorului $\hat{a}_i$* se calculează prin extragerea rădăcinii pătrate din dispersia estimatorului respectiv.

7.2.2. Teste de semnificație în cazul modelului unifactorial

Algoritmul urmat în testarea semnificației estimatorilor pentru modelul multifactorial de regresie lineară este următorul:

Pasul 1: Se formulează ipoteza nulă $H_0: \hat{a}_i = 0$ și ipoteza alternativă $H_1: \hat{a}_i > 0$, pentru $i = 0, 1, \dots, k$.

Pasul 2: Se calculează statistica $t_{\hat{a}_i} = \frac{\hat{a}_i}{s_{\hat{a}_i}}$, unde $\hat{a}_i$ este estimatorul obținut prin

aplicarea metodei celor mai mici pătrate pentru parametrul a_i din modelul linear de regresie multifactorială, iar $s_{\hat{a}_i}$ este un estimator nedeplasat al abaterii standard a lui $\hat{a}_i$, estimator calculat pornind de la

datele de selecție. Sub ipoteza nulă, statistica respectivă urmează o distribuție Student cu $n - k - 1$ grade de libertate, unde n este volumul eșantionului și k reprezintă numărul variabilelor exogene din model.

Pasul 3: Din tabelul statisticii Student, pornind de la numărul gradelor de libertate ($n - k - 1$) și de la nivelul de semnificație ales (α), se selectează o valoare $t_{n-k-1,\alpha}^*$ astfel încât $P(t > t^*) = \alpha$ (testul unilateral).

Pasul 4: Dacă valoarea calculată $t_{\hat{a}_i}$ este mare admitem că, probabil, parametrul a_i este mai mare decât zero. Concret, dacă $t_{\hat{a}_i} > t_{n-k-1;\alpha}^*$ atunci se respinge H_0 și admitem că parametrul a_i este semnificativ mai mare decât zero, cu un grad de încredere mai mare decât $1 - \alpha$. Dacă $\hat{a}_i$ este negativ, atunci ipoteza alternativă este $H_1: \hat{a}_i < 0$ și se respinge H_0 dacă $t_{\hat{a}_i} < -t_{n-k-1;\alpha}^*$ $t_{a_i} < -t^*$. O altă soluție constă în calcularea statisticilor t în modul (în valoare absolută) și se urmează procedura prezentată.

Respingerea ipotezei nule $H_0: a_i = 0$ semnifică faptul că datele disponibile nu oferă argumente în favoarea ipotezei potrivit căreia estimatorii din ecuația de regresie sunt nuli, adică a faptului că, statistic, nu există o legătură între variabila endogenă Y și variabila explicativă X_i .

Recapitulare

Dispersia estimatorilor din modelul linear multifactorial de regresie se

calculează astfel: $S_{\hat{A}}^2 = s_u^2 (X'X)^{-1}$, unde $s_u^2 = \frac{\sum_{t=1}^n u_t^2}{n - k - 1}$.

Pentru testarea semnificație estimatorilor se calculează statisticile t_{calc} care se compară cu valorile corespunzătoare din distribuția teoretică t-Student.

7.3. Problemă rezolvată

Fie următoarele serii de date, înregistrate în 20 momente diferite de timp:

t	Y _t	X _{1t}	X _{2t}	t	Y _t	X _{1t}	X _{2t}
1	17.71	5.23	3.02	11	10.12	3.45	1.74
2	15.26	6.23	4.09	12	15.66	8.27	6.27
3	11.48	2.34	1.42	13	8.47	3.59	4.45
4	37.95	5.44	-1.81	14	11.85	6.35	5.23
5	28.27	6.49	0.91	15	17.08	4.22	1.64
6	15.26	6.52	5.18	16	29.68	7.51	2.34
7	0.75	4.40	6.05	17	31.05	6.37	0.22
8	25.98	4.70	1.00	18	3.25	5.59	6.43
9	18.01	6.59	3.72	19	24.73	5.40	0.50
10	17.39	5.57	0.55	20	14.90	3.17	1.66

Observație: datele sunt cele din problema rezolvată la capitolul 4.3.

Testați semnificația parametrilor din modelul linear multifactorial de regresie: $Y_t = a_0 + a_1X_{1t} + a_2X_{2t} + e_t$.

Rezolvare:

Estimatorii parametrilor din ecuația de regresie sunt:

$$\hat{A} = \begin{matrix} 7.70335 \\ 3.72858 \\ -3.65884 \end{matrix}$$

(vezi problema rezolvată de la cap. 4.3). Aplicarea algoritmului de testare a semnificației parametrilor presupune calculul, în prealabil, a dispersiei estimatorilor. Dispersia estimatorilor din modelul linear multifactorial de regresie se calculează astfel:

$$S_{\hat{A}}^2 = s_u^2 (X'X)^{-1},$$

unde
$$s_u^2 = \frac{\sum_{t=1}^n u_t^2}{n - k - 1} = \frac{166.9294}{20 - 2 - 1} \text{ și}$$

$$(X'X)^{-1} = \begin{vmatrix} 0.71764 & -0.12087 & -0.00672 \\ -0.12087 & 0.02453 & -0.00399 \\ -0.00672 & -0.00399 & 0.01031 \end{vmatrix}$$

Rezultă abaterile standard ale estimatorilor:

$$s_{\hat{a}_0} = 2.654574,$$

$$s_{\hat{a}_1} = 0.490786$$

și $s_{\hat{a}_2} = 0.318214$

(calculate prin extragerea rădăcinii pătrate din dispersia estimatorilor). Valorile calculate ale statisticilor t sunt:

$$t_{\hat{a}_0} = 2.902$$

$$t_{\hat{a}_1} = 7.597$$

$$t_{\hat{a}_2} = -11.498.$$

Valoarea identificată în tabelul distribuției teoretice unilaterale t-student pentru 20-2-1 grade de libertate sunt prezentate în tabelul următor:

α	0.10	0.05	0.01
$t_{17;\alpha}$	1.333	1.740	2.567

În evaluarea semnificației parametrilor din model, pragul standard este de 0.05 (5%).

Deoarece $t_{\hat{a}_0} = 2.902 > 1.740$, $t_{\hat{a}_1} = 7.597 > 1.740$ și $|t_{\hat{a}_2}| = 11.498 > 1.740$, riscul ca parametrii a_0 , a_1 și a_2 să nu difere semnificativ de zero este mai mic de 5%.

De altfel, riscul este mai mic decât 1% (pragul pentru 1% este 2.567 și statisticile calculate sunt superioare acestui prag).

Soluția EViews-10 pentru problema precedentă este prezentată în tabelul următor:

Dependent Variable: Y

Method: Least Squares

Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.703353	2.654574	2.901917	0.0099
X ₁	3.728578	0.490786	7.597154	0.0000
X ₂	-3.658842	0.318214	-11.49806	0.0000

Riscul ca estimatorii să fie zero este prezentat (cu bold), în ultima coloană a tabelului.



7.4. Rezumat

Pentru a calcula un estimator nedepășit al covarianței dintre estimatorii $\hat{a}_i$ și $\hat{a}_j$ se înmulțește d_{ij} – elementul aflat la intersecția dintre linia i ($i = 0, 1, 2, \dots, k$) și coloana j ($j = 0, 1, 2, \dots, k$) în matricea $(X'X)^{-1}$ cu dispersia constantă a valorilor variabilei reziduale.

Pornind de la dispersia de selecție a estimatorilor se calculează raportul dintre valoarea estimatorilor și abaterea standard. Raportul respectiv se compară cu valoarea corespunzătoare din distribuția unilaterală t-Statistic.



7.5. Test de autoevaluare a cunoștințelor

Fie următoarele serii de date, înregistrate în 20 momente diferite de timp:

t	Y _t	X _{1t}	X _{2t}	t	Y _t	X _{1t}	X _{2t}
1	23.96	5.23	3.02	11	13.31	3.45	1.74
2	23.57	6.23	4.09	12	28.20	8.27	6.27
3	13.25	2.34	1.42	13	14.51	3.59	4.45
4	39.58	5.44	-1.81	14	21.44	6.35	5.23

5	33.67	6.49	0.91	15	20.94	4.22	1.64
6	24.96	6.52	5.18	16	37.53	7.51	2.34
7	9.20	4.40	6.05	17	35.64	6.37	0.22
8	29.68	4.70	1.00	18	13.26	5.59	6.43
9	26.33	6.59	3.72	19	28.62	5.40	0.50
10	21.51	5.57	0.55	20	17.74	3.17	1.66

Testați semnificația parametrilor din modelul linear multifactorial de regresie: $Y_t = a_0 + a_1X_{1t} + a_2X_{2t} + e_t$.



7.6. Bibliografie recomandată

- a. Jula, Dorin; Jula, Nicolae-Marius, 2022. *Econometrie*, ediția a 5^a. Editura Mustang, București
- b. Jula, Nicoleta; Jula, Dorin, 2017. *Modelare economică. Modele econometrice și de optimizare*, Editura Mustang, București

Alternativa web:

- a. Hansen, Bruce E. 2017. *Econometrics*. Download for free at:
<http://www.ssc.wisc.edu/~bhansen/econometrics/Econometrics.pdf>
- b. Diebold, Francis X., 2013-2016. *Econometrics*. Department of Economics, University of Pennsylvania. Download for free at:
<http://www.ssc.upenn.edu/~fdiebold/Teaching104/Econometrics.pdf>

Unitatea de învățare 8. Acuratețea ajustării



8.1. Obiective

Prin metoda celor mai mici pătrate se determină acea ecuație de regresie pentru care suma pătratelor abaterilor dintre datele înregistrate și cele calculate este cea mai mică posibilă, pentru clasa de modele respective. În cadrul acestui capitol sunt prezentate principalele noțiuni referitoare la evaluarea acurateții ajustării.

Durata medie de parcurgere a acestei unități de învățare este de aproximativ 3 ore.



8.2. Conținut

8.2.1. Coeficientul de determinare

Se demonstrează¹⁶ că, dacă estimatorii sunt determinați pe baza metodei celor mai mici pătrate, iar ecuația de regresie conține și termenul liber, atunci variația totală (VT) se descompune în variația explicată de model (VM) și variația reziduală (VR):

$$VT = VM + VR:$$

$$\sum_{t=1}^n (Y_t - \bar{Y})^2 = \sum_{t=1}^n (\hat{Y}_t - \bar{Y})^2 + \sum_{t=1}^n u_t^2 .$$

¹⁶ Pentru demonstrație vezi: Jula, Dorin, 2003. *Introducere în econometrie*, Editura Professional Consulting, București, 2003, capitolul II, anexa 2.A.3

Pornind de la relația precedentă, se definește *coeficientul de determinare*, notat cu R^2 , ca fiind partea din variația lui Y față de medie care este explicată de model, pentru eșantionul analizat:

$$R^2 = \frac{VM}{VT} = \frac{VT - VR}{VT} = 1 - \frac{VR}{VT},$$

adică

$$R^2 = 1 - \frac{\sum_{t=1}^n u_t^2}{\sum_{t=1}^n (Y_t - \bar{Y})^2}$$

Prin modul de definire, coeficientul de determinare este o mărime pozitivă și subunitară. Cu cât R^2 este mai aproape de unu, cu atât modelul se apropie mai mult de procesul economic modelat.

Pentru modelul linear multifactorial de regresie, la fel ca în cazul modelului unifactorial de regresie lineară, *variația totală a lui Y* (VT) se calculează prin însumarea pătratelor variațiilor individuale, adică prin expresia:

$$VT = (Y - \bar{Y})(Y - \bar{Y})' = \sum_{t=1}^n (Y_t - \bar{Y})^2$$

De asemenea, *variația totală explicată de model* (VM) se calculează astfel:

$$VM = (\hat{Y} - \bar{Y})(\hat{Y} - \bar{Y})' = \sum_{t=1}^n (\hat{Y}_t - \bar{Y})^2$$

În sfârșit, prin analogie, se definește *variația reziduală* (VR) prin expresia:

$$VR = u'u = \sum_{t=1}^n u_t^2.$$

Calitatea ajustării este măsurată, la fel ca în cazul modelului unifactorial, prin *coeficientul de determinare R^2* .

Coeficientul de determinare nu poate fi interpretat atunci când ecuația de regresie nu conține termen liber și crește atunci când în ecuația de regresie sunt adăugate variabile explicative suplimentare (chiar dacă aceste variabile au mică relevanță în explicarea variației lui Y). Pentru a contracara acest efect, Theil a

propus un coeficient modificat, notat $\bar{R}^2$ și denumit în mod uzual R^2 – ajustat, coeficient care penalizează adăugarea de variabile explicative suplimentare, nerelevante.

8.2.1. Coeficientul de determinare corectat

Pentru a se evita tentația de construire a unor modele explicative de dimensiuni mari, irelevante din punct de vedere al teoriei economice (și care din punct de vedere al coeficientului de determinare sunt "mai bune") s-a propus construirea altor măsuri ale calității ajustării. Cea mai utilizată corecție este cea introdusă de Theil prin *coeficientul de determinare ajustat*, notat $\bar{R}^2$ și care folosește dispersia în locul variației.

$$\bar{R}^2 = 1 - \frac{s_u^2}{s_Y^2}$$

unde s_u^2 este dispersia de selecție a reziduurilor din modelul de regresie, iar s_Y^2 este dispersia de selecție a variabilei endogene:

$$s_u^2 = \frac{\sum_{t=1}^n u_t^2}{n - k - 1}$$

$$s_Y^2 = \frac{\sum (Y_t - \bar{Y})^2}{n - 1}$$

Dacă ținem seama de formula de calcul a coeficientului R^2 , $\bar{R}^2$, se calculează astfel:

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1} (1 - R^2)$$

unde k este numărul variabilelor exogene din model.

Recapitulare

Coeficientul de determinare, notat cu R^2 , se calculează astfel:

$$R^2 = 1 - \frac{\sum_{t=1}^n u_t^2}{\sum_{t=1}^n (Y_t - \bar{Y})^2}$$

Coeficientul de determinare corectat, notat cu $\bar{R}^2$, se calculează astfel:

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1} (1 - R^2)$$

unde R^2 este coeficientul de determinare simplu, n – reprezintă numărul de înregistrări, iar k – numărul de variabile explicative din model.

8.3. Problemă rezolvată

Fie următoarele serii de date, înregistrate în 20 momente diferite de timp:

t	Y_t	X_{1t}	X_{2t}	t	Y_t	X_{1t}	X_{2t}
1	17.71	5.23	3.02	11	10.12	3.45	1.74
2	15.26	6.23	4.09	12	15.66	8.27	6.27
3	11.48	2.34	1.42	13	8.47	3.59	4.45
4	37.95	5.44	-1.81	14	11.85	6.35	5.23
5	28.27	6.49	0.91	15	17.08	4.22	1.64
6	15.26	6.52	5.18	16	29.68	7.51	2.34
7	0.75	4.40	6.05	17	31.05	6.37	0.22
8	25.98	4.70	1.00	18	3.25	5.59	6.43
9	18.01	6.59	3.72	19	24.73	5.40	0.50
10	17.39	5.57	0.55	20	14.90	3.17	1.66

Observație: datele sunt cele din problemele rezolvate la capitolele 4.3 și 7.3.

Testați acuratețea ajustării pentru modelul linear multifactorial de regresie:

$$Y_t = a_0 + a_1 X_{1t} + a_2 X_{2t} + e_t.$$

Rezolvare:

Calculăm coeficientul de determinare (R^2) și coeficientul de determinare ajustat ($\bar{R}^2$):

$$R^2 = 1 - \frac{\sum_{t=1}^n u_t^2}{\sum_{t=1}^n (Y_t - \bar{Y})^2} = 1 - \frac{166.9294}{1697.8687} = 0.9017$$

Interpretare: aproximativ 90.17% din variația față de medie a variabilei exogene Y ar putea fi explicată de model.

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1} (1 - R^2) = 1 - \frac{20-1}{20-2-1} (1 - 0.9017) = 0.8901.$$

Soluția EViews-10 pentru problema precedentă este prezentată în tabelul următor:

Dependent Variable: Y

Method: Least Squares

Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.703353	2.654574	2.901917	0.0099
X ₁	3.728578	0.490786	7.597154	0.0000
X ₂	-3.658842	0.318214	-11.49806	0.0000
R-squared	0.901683	Mean dependent var	17.74276	
Adjusted R-squared	0.890116	S.D. dependent var	9.453122	
S.E. of regression	3.133589	Akaike info criterion	5.259716	
Sum squared resid	166.9294	Schwarz criterion	5.409076	
Log likelihood	-49.59716	Hannan-Quinn criter.	5.288873	
F-statistic	77.95500	Durbin-Watson stat	1.632575	
Prob(F-statistic)	0.000000			

Soluțiile sunt marcate cu bold. Celelalte informații oferite în partea inferioară a tabelului vor fi descrise în capitolele următoare.



8.4. Rezumat

Coeficientul de determinare (R^2) evaluează partea din variația lui Y față de medie care este explicată de model, pentru eșantionul analizat. Interpretarea obișnuită a coeficientului de determinare este următoarea: din variația totală față de medie a variabilei exogene (Y), $[(R^2) \cdot 100\%]$ ar putea fi explicată de model.

Pentru calculul *coeficientului de determinare ajustat* folosim, în calcule, dispersia, în locul variației față de medie.



8.5. Test de autoevaluare a cunoștințelor

Fie următoarele serii de date, înregistrate în 20 momente diferite de timp:

t	Y_t	X_{1t}	X_{2t}	t	Y_t	X_{1t}	X_{2t}
1	23.96	5.23	3.02	11	13.31	3.45	1.74
2	23.57	6.23	4.09	12	28.20	8.27	6.27
3	13.25	2.34	1.42	13	14.51	3.59	4.45
4	39.58	5.44	-1.81	14	21.44	6.35	5.23
5	33.67	6.49	0.91	15	20.94	4.22	1.64
6	24.96	6.52	5.18	16	37.53	7.51	2.34
7	9.20	4.40	6.05	17	35.64	6.37	0.22
8	29.68	4.70	1.00	18	13.26	5.59	6.43
9	26.33	6.59	3.72	19	28.62	5.40	0.50
10	21.51	5.57	0.55	20	17.74	3.17	1.66

Testați acuratețea ajustării pentru modelul linear multifactorial de regresie:

$$Y_t = a_0 + a_1X_{1t} + a_2X_{2t} + e_t.$$



8.6. Bibliografie recomandată

- a. Jula, Nicoleta; Jula, Dorin, 2017. *Modelare economică. Modele econometrice și de optimizare*, Editura Mustang, București
- b. Jula, Dorin; Jula, Nicolae-Marius, 2017. *Proгноza economică. Aplicații financiare*, Editura Mustang, București

Alternativa web:

- c. Hansen, Bruce E. 2017. *Econometrics*. Download for free at:
<http://www.ssc.wisc.edu/~bhansen/econometrics/Econometrics.pdf>
- d. Diebold, Francis X., 2013-2016. *Econometrics*. Department of Economics, University of Pennsylvania. Download for free at:
<http://www.ssc.upenn.edu/~fdiebold/Teaching104/Econometrics.pdf>

8.7. Teme de control

Din site-ul Institutului Național de Statistică (www.insse.ro), de la secțiunea Tempo On-line selectați trei serii de date anuale, începând cu 1990 și până în prezent. Analizați econometric legătura dintre variabilele respective (ecuația de regresie, estimarea coeficienților, testele de semnificație, acuratețea ajustării).

Unitatea de învățare 9. Specificarea modelului



9.1. Obiective

În cadrul acestui capitol sunt prezentate principalele criterii folosite pentru evaluarea măsurii în care modelul econometric linear este bine specificat, precum și noțiuni referitoare la erorile de specificare.

Durata medie de parcurgere a acestei unități de învățare este de aproximativ 3 ore.



9.2. Conținut

9.2.1. Criterii pentru specificarea modelului

Dacă prin includerea unei (unor) variabile suplimentare suma pătratelor reziduurilor scade mai repede decât numărul gradelor de libertate, din punct de vedere econometric se justifică reținerea în model a variabilei (variabilelor) respective.

De asemenea, se poate demonstra următoarea proprietate: *dacă valoarea absolută a testului t pentru un parametru din ecuația de regresie lineară multiplă este mai mică decât 1, atunci, eliminând din model variabila explicativă asociată, valoarea coeficientului de determinare corectat $\bar{R}^2$ va crește; dacă se elimină o variabilă pentru care t statistic este mai mare decât 1, valoarea coeficientului de determinare corectat $\bar{R}^2$ se va reduce¹⁷.*

¹⁷ Ramanathan Ramu, 1992, *Introductory Econometrics*, Second Edition, The Dryden Press, Harcourt Brace College Publishers, Orlando, USA, pag. 170.

9.2.1.1. Criteriul informațional Akaike (AIC)

Unul dintre cele mai cunoscute modalități de specificare a modelelor econometrice este *criteriul informațional Akaike* (Akaike information criterion – AIC). Acest criteriu este definit astfel:

$$AIC = \left(\frac{VTR}{n} \right) \cdot e^{\frac{2(k+1)}{n}} = \frac{1}{n} \left(\sum_{t=1}^n u_t^2 \right) \cdot e^{\frac{2(k+1)}{n}}$$

sau, în expresie logaritmică

$$\ln(AIC) = \ln \left(\frac{\sum u_t^2}{n} \right) + \frac{2(k+1)}{n}$$

O condiție pentru includerea unei noi variabile explicative este ca *prin această re-specificare a modelului să se obțină o valoare mai mică pentru AIC* sau, echivalent, pentru $\ln(AIC)$.

9.2.1.2. Criteriul informațional Schwartz (SIC)

O altă procedură cunoscută este *criteriul Schwartz*. Acest criteriu presupune calculul expresiei:

$$SIC = \left(\frac{VTR}{n} \right) \cdot n^{\frac{k+1}{n}} = \frac{1}{n} \left(\sum_{t=1}^n u_t^2 \right) \cdot n^{\frac{k+1}{n}}$$

sau, în expresie logaritmică:

$$\ln(SC) = \ln \left(\frac{\sum u_t^2}{n} \right) + \frac{(k+1)}{n} \ln(n)$$

La fel ca în cazul testului AIC, o variabilă suplimentară este admisă dacă noua valoare obținută pentru criteriul SCHWARTZ este inferioară celei calculate pentru modelul inițial.

9.3.1.3. Criteriul Hannan-Quinn

Criteriul Hannan-Quinn presupune calculul următoarei valori:

$$HQ = \left(\frac{VTR}{n} \right) \cdot \ln(n)^{\frac{2(k+1)}{n}} = \frac{1}{n} \left(\sum_{t=1}^n u_t^2 \right) \cdot \ln(n)^{\frac{2(k+1)}{n}},$$

sau, în expresie logaritmică

$$\ln(HQ) = \ln\left(\frac{\sum u_t^2}{n}\right) + \frac{2(k+1)}{n} \ln(\ln(n))$$

Interpretarea este la fel ca pentru AIC și SIC.

9.2.2. Erori de specificare a modelului linear de regresie

Este posibil ca în construcția modelului să fie omise variabile explicative importante, sau, din contră, să fie incluse variabile irelevante. Ambele situații reprezintă erori de specificare.

Consecințele omiterii unor variabile importante în specificarea modelului pot fi sintetizate astfel¹⁸:

Consecințe ale omiterii unor variabile importante

- a. Dacă o variabilă importantă omisă este corelată cel puțin cu o variabilă inclusă în model, atunci estimatorii parametrilor reținuți în model sunt deplasați și nu sunt consistenți;
 - b. Chiar dacă variabilele omise nu sunt corelate cu variabilele reținute în model, estimatorul termenului liber ($\hat{\alpha}_0$) este, în general, deplasat;
 - c. Dispersiile estimate pentru parametrii variabilelor reținute în model sunt estimatori deplasați ai dispersiilor reale și, în consecință, testul t privind semnificația estimatorilor nu este valid.
-

¹⁸ Kmenta, Jan, 1986, *Elements of Econometrics*, New York: Macmillan, pag. 394

Consecințele includerii unor variabile nerelevante în specificarea modelului pot fi sintetizate astfel¹⁹:

Consecințe ale includerii unor variabile nerelevante

- a. Dacă o variabilă explicativă nerelevantă este inclusă în model, atunci estimatorii parametrilor pentru toate celelalte variabile din model sunt nedeplasați și consistenți;
 - b. Dispersiile estimate pentru parametrii variabilelor din model sunt mai mari decât în cazul neinclunderii variabilelor nerelevante și deci estimatori nu sunt eficienți;
 - c. Deoarece dispersiile estimate pentru parametrii modelului sunt nedeplasate, testul t privind semnificația estimatorilor este valid.
-

Recapitulare

Criteriile informaționale utilizate pentru specificarea modelului se calculează astfel:

$$\ln(\text{AIC}) = \ln\left(\frac{\sum u_t^2}{n}\right) + \frac{2(k+1)}{n}$$

$$\ln(\text{SC}) = \ln\left(\frac{\sum u_t^2}{n}\right) + \frac{(k+1)}{n} \ln(n)$$

$$\ln(\text{HQ}) = \ln\left(\frac{\sum u_t^2}{n}\right) + \frac{2(k+1)}{n} \ln(\ln(n))$$

Un model este mai bun dacă valoarea criteriilor informaționale este mai mică.

¹⁹ Ramanathan, Ramu, 1992. *Introductory Econometrics*, Second Edition, The Dryden Press, Harcourt Brace College Publishers, Orlando, USA, pag. 188.

Este posibil ca în construcția modelului să fie omise variabile explicative importante, sau, din contră, să fie incluse variabile irelevante. Ambele situații reprezintă erori de specificare.

9.3. Problemă rezolvată

Fie următoarele serii de date, înregistrate în 20 momente diferite de timp:

t	Y_t	X_{1t}	X_{2t}	t	Y_t	X_{1t}	X_{2t}
1	17.71	5.23	3.02	11	10.12	3.45	1.74
2	15.26	6.23	4.09	12	15.66	8.27	6.27
3	11.48	2.34	1.42	13	8.47	3.59	4.45
4	37.95	5.44	-1.81	14	11.85	6.35	5.23
5	28.27	6.49	0.91	15	17.08	4.22	1.64
6	15.26	6.52	5.18	16	29.68	7.51	2.34
7	0.75	4.40	6.05	17	31.05	6.37	0.22
8	25.98	4.70	1.00	18	3.25	5.59	6.43
9	18.01	6.59	3.72	19	24.73	5.40	0.50
10	17.39	5.57	0.55	20	14.90	3.17	1.66

Observație: datele sunt cele din problemele rezolvate la capitolele 4.3, 7.3 și 8.3.

Testați specificarea modelului linear multifactorial de regresie:

$$Y_t = a_0 + a_1X_{1t} + a_2X_{2t} + e_t.$$

Rezolvare:

Presupunem că specificarea modelului este:

$$(M1) \quad Y_t = a_0 + a_1X_{1t} + e_t.$$

Valorile calculate pentru criteriile informaționale sunt:

Akaike info criterion 7.331826

Schwarz criterion 7.431399

Hannan-Quinn criter. 7.351264

Presupunem că specificarea modelului este:

$$(M2) \quad Y_t = a_0 + a_2X_{2t} + e_t.$$

Valorile calculate pentru criteriile informaționale sunt:

Akaike info criterion	6.640207
Schwarz criterion	6.739780
Hannan-Quinn criter.	6.659645

Presupunem că specificarea modelului este:

$$(M3) \quad Y_t = a_0 + a_1 X_{1t} + a_2 X_{2t} + e_t.$$

Valorile calculate pentru criteriile informaționale sunt:

Akaike info criterion	5.259716
Schwarz criterion	5.409076
Hannan-Quinn criter.	5.288873

Se observă că pentru modelul complet (M3) valorile calculate pentru criteriile informaționale sunt cele mai mici. În concluzie, M3 este preferat modelelor M1 și M2.



9.4. Rezumat

Dacă valoarea absolută a testului t pentru un parametru din ecuația de regresie lineară multiplă este mai mică decât 1, atunci, eliminând din model variabila explicativă asociată, valoarea coeficientului de determinare corectat $\bar{R}^2$ va crește; dacă se elimină o variabilă pentru care t statistic este mai mare decât 1, valoarea coeficientului de determinare corectat $\bar{R}^2$ se va reduce.

Un model este mai bun dacă valoarea criteriilor informaționale (Akaike Information Criterion – AIC, Schwartz Information Criterion – SIC, Hannan-Quinn Information Criterion) este mai mică.



9.5. Test de autoevaluare a cunoștințelor

Fie următoarele serii de date, înregistrate în 20 momente diferite de timp:

t	Y_t	X_{1t}	X_{2t}	t	Y_t	X_{1t}	X_{2t}
1	23.96	5.23	3.02	11	13.31	3.45	1.74
2	23.57	6.23	4.09	12	28.20	8.27	6.27
3	13.25	2.34	1.42	13	14.51	3.59	4.45
4	39.58	5.44	-1.81	14	21.44	6.35	5.23
5	33.67	6.49	0.91	15	20.94	4.22	1.64
6	24.96	6.52	5.18	16	37.53	7.51	2.34
7	9.20	4.40	6.05	17	35.64	6.37	0.22
8	29.68	4.70	1.00	18	13.26	5.59	6.43
9	26.33	6.59	3.72	19	28.62	5.40	0.50
10	21.51	5.57	0.55	20	17.74	3.17	1.66

Testați specificarea modelului linear multifactorial de regresie:

$$Y_t = a_0 + a_1X_{1t} + a_2X_{2t} + e_t.$$



9.6. Bibliografie recomandată

- a. Jula, Dorin; Jula, Nicolae-Marius, 2022. *Econometrie*, ediția a 5^a. Editura Mustang, București
- b. Jula, Nicoleta; Jula, Dorin, 2017. *Modelare economică. Modele econometrice și de optimizare*, Editura Mustang, București

Alternativa web:

- a. Hansen, Bruce E. 2017. *Econometrics*. Download for free at:
<http://www.ssc.wisc.edu/~bhansen/econometrics/Econometrics.pdf>

- b.** Diebold, Francis X., 2013-2016. *Econometrics*. Department of Economics, University of Pennsylvania. Download for free at:
<http://www.ssc.upenn.edu/~fdiebold/Teaching104/Econometrics.pdf>

Unitatea de învățare 10. Multicolinearitatea



10.1. Obiective

În cadrul acestui capitol sunt prezentate noțiunile fundamentale referitoare la consecințele, identificarea și atenuarea fenomenului de multicolinearitate.

Durata medie de parcurgere a acestei unități de învățare este de aproximativ 3 ore.



10.2. Conținut

10.2.1. Consecințe ale multicolinearității

Renunțarea la ipoteza de independență a variabilelor explicative din modelele de regresie creează probleme în ceea ce privește estimarea parametrilor și calitatea estimatorilor.

Sintetic, consecințele prezenței fenomenului de multicolinearitate sunt următoarele²⁰:

Consecințe ale multicolinearității

-
- Dacă două sau mai multe variabile explicative din modelul de regresie multiplă sunt perfect corelate, estimatorii parametrilor nu pot fi calculați prin metoda celor mai mici pătrate.
-

²⁰ Ramanathan, Ramu, 1992. *Introductory Econometrics*, Second Edition, The Dryden Press, Harcourt Brace College Publishers, Orlando, USA, pag.233-245, pentru consecințele a, b, c și e; Thomas, Richard Leighton, 1997. *Modern Econometrics: An Introduction*, 2nd edition, Harlow, Longman, pag.237-244, pentru consecința d.

-
- b. Dacă anumite variabile explicative sunt relativ puternic corelate, estimatorii obținuți prin metoda celor mai mici pătrate sunt lineari, normal distribuiți, nedeplasați, consistenți și de maximă verosimilitate.
 - c. Efectul multicolarității se manifestă în creșterea abaterii standard a estimatorilor calculați pentru parametrii modelului, ceea ce reduce valoarea testului t statistic (Student). Aceasta face estimatorii mai puțin semnificativi (posibil chiar nesemnificativi). Totuși, testul t rămâne valid.
 - d. Se reduce precizia estimatorilor calculați pentru parametrii modelului, în sensul că abaterea standard mare duce la creșterea intervalului de încredere în care sunt garantați parametrii.
 - e. Deoarece covarianța între variabilele explicative corelate relativ puternic poate fi mare (în valoare absolută), interpretarea parametrilor individuali este dificilă.
-

10.2.2. Identificarea multicolarității

Este posibil ca fenomenul de multicolaritate să fie prezent atunci când:

-
- a. *Coeficienții de corelație lineară, calculați pentru perechile de variabile explicative din model, sunt mari în valoare absolută* (sunt, în modul, apropiați de +1). Deoarece efectele multicolarității depind și de alți factori (de exemplu, dispersia erorilor și dispersia variabilelor explicative), nu există valori prag ale coeficienților de corelație, praguri a căror depășire să semnifice prezența multicolarității. Existența unor coeficienți de corelație mari (în valoare absolută) între variabila endogenă Y și variabilele explicative X_i nu indică neapărat prezența fenomenului de multicolaritate, ci, din contră, pot semnală faptul că sunt șanse mari ca modelul să fie bine specificat (factorii reținuți ca variabile explicative să fie semnificativi, în sensul definit prin testele statistice de semnificație).
-

-
- b. *Determinantul matricei $(X'X)$ are valori în apropierea lui zero.* La fel, nu există un prag fundamentat statistic astfel încât situarea valorii $\det(X'X)$ sub acest prag să semnifice prezența multicolinearității.
- c. *Coeficientul de determinare R^2 este mare, iar valorile testelor t (Student), calculate pentru parametrii modelului sunt mici.* Nici acest criteriu nu are o valoare absolută, deoarece valorile mici ale testelor de semnificație ar putea proveni de la dispersiile mari ale estimatorilor, generate de prezența fenomenului de multicolinearitate.
- d. *Estimatorii parametrilor sunt sensibili la specificarea modelului.* Dacă existența unor valori mari ale coeficienților de corelație lineară poate indica prezența multicolinearității, reciproca acestei afirmații nu este întotdeauna adevărată: fenomenul de multicolinearitate poate fi prezent chiar dacă valorile coeficienților de corelație lineară între variabilele explicative nu sunt mari. Aceasta deoarece este posibil ca o variabilă exogenă să fie corelată cu două sau mai multe alte variabile. În asemenea situații, valorile estimatorilor pot să se modifice semnificativ (chiar cu schimbare de semn) atunci când în respecificarea modelului sunt eliminate anumite variabile explicative, sau sunt introduse variabile suplimentare.
-

10.2.2.1. Criteriul lui Klein

Criteriul lui Klein²¹ se bazează pe compararea coeficienților de corelație lineară (Pearson) între variabilele explicative și coeficientul de determinare R^2 . Procedura este următoarea:

- se calculează coeficienții de corelație lineară r_{ij} , între variabilele explicative x_i și x_j , pentru $i, j \in \{1, 2, \dots, k\}$
- se estimează, prin metoda celor mai mici pătrate, ecuația de regresie:

²¹ Klein, Lawrence R., 1962, *An introduction to econometrics*, Greenwood Press, p. 101.

$$Y_t = a_0 + a_1X_{1t} + a_2X_{2t} + \dots + a_kX_{kt} + e_t$$

și se calculează coeficientul de determinare R^2 .

- se compară r_{ij} , pentru $i \neq j$ cu R^2 . Dacă $|r_{ij}| > R^2$ atunci acceptăm presupunerea că există colinearitate între variabilele explicative.

10.2.2.2. Factorul de inflație al dispersiei

Pentru modelul, $Y_t = a_0 + a_1X_{1t} + a_2X_{2t} + \dots + a_kX_{kt} + e_t$, dispersia estimatorului $\hat{a}_i$ se calculează astfel²²:
$$s_{\hat{a}_i}^2 = \frac{s_u^2}{\sum_{t=1}^n (X_{it} - \bar{X}_i)^2} \cdot \frac{1}{1 - R_i^2}.$$

Factorul de inflație al dispersiei pentru estimatorul $\hat{a}_i$ este: $VIF_i = \frac{1}{1 - R_i^2},$

unde prin R_i^2 am notat coeficientul de determinare din ecuația de regresie în care factorul de influență X_i este calculat în funcție de toate celelalte variabile din model.

Se consideră că o valoare $VIF > 4$ (echivalent cu $R_i^2 > 0.75$) arată prezența fenomenului de multicolaritate, iar $VIF > 10$ (echivalent cu $R_i^2 > 0.90$) indică un fenomen sever de multicolaritate.

10.2.3. Atenuarea multicolarității

Metode de atenuare a multicolarității

- Eliminarea unor variabile explicative.* Deoarece multicolaritatea este generată de legătura puternică existentă între două sau mai multe variabile explicative, o măsură imediată de atenuare a fenomenului respectiv este eliminarea uneia dintre variabilele implicate. Selectarea variabilei care va fi

²² Greene William H., 2002, *Econometric Analysis*, Fifth Edition, Prentice Hall, p.56-58.

eliminată se face fie prin analiză teoretică (economică), fie pornind de la testele statistice de semnificație. În ultima situație, se elimină variabila pentru care valoarea testului t este mică, sau variabila pentru care corelația cu Y este mai slabă.

- b. *Realizarea unor observații suplimentare asupra variabilelor din model (se mărește volumul eșantionului).* Dacă este posibilă mărirea eșantionului, atunci crește precizia estimatorilor și se reduce efectul negativ al multicolinearității. De asemenea, prin adăugarea unor valori suplimentare, cresc variațiile înregistrate de exogenele X_i (cu excepția situației în care valorile adăugate suplimentar sunt egale cu media de selecție a variabilelor respective). Adică, numitorul din relațiile

$$s_{\hat{a}_i}^2 = \frac{s_u^2}{\sum_{t=1}^n (X_{it} - \bar{X}_i)^2} \cdot \frac{1}{1 - R_i^2} = \frac{s_u^2}{\sum_{t=1}^n (X_{it} - \bar{X}_i)^2} \cdot VIF_i$$

crește prin mărirea eșantionului (VIF_i este factorul de inflație al dispersiei pentru estimatorul $\hat{a}_i$). Dacă, prin aplicarea acestei proceduri, valoarea R_i^2 nu se modifică, sau scade, atunci dispersia estimatorilor se reduce (R_i^2 este coeficientul de determinare din ecuația de regresie în care factorul de influență X_i este calculat în funcție de toate celelalte variabile din model). În acest mod, efectele negative ale fenomenului de multicolinearitate sunt atenuate.

- c. *Prelucrarea primară a datelor (calculul ritmurilor de modificare, a sporurilor, indicilor, logaritmare valorilor observate).* Este posibil, de exemplu în cazul seriilor de timp, ca anumite variabile economice să evolueze pe traiectorii asemănătoare, fără ca între acestea să existe o legătură directă. Astfel, în medii economice puternic inflaționiste, variabilele au în timp valori crescătoare dacă sunt exprimate nominal, chiar dacă, în prețuri constante, tendința reală este de scădere. În aceste condiții, evitarea

multicolinearității indusă artificial se poate realiza prin includerea în model a ritmurilor (sau a indicilor) de modificare a variabilelor respective.

- d. *Regresia ridge*. Este o tehnică mecanică, arbitrară, care constă în transformarea matricei $(X'X)$ astfel încât determinantul matricei obținute să fie diferit de zero. De exemplu, se adaugă un număr fiecărui element de pe diagonala principală a matricei $(X'X)$, astfel încât în aplicarea procedurii de estimare a parametrilor modelului se folosește matricea $(X'X + \alpha I_{k+1,k+1})$, unde α este o constantă, iar I este matricea unitate. Chiar dacă în acest mod calitatea statistică a estimării poate fi îmbunătățită (crește eficiența estimatorilor), principalul inconvenient constă în dificultatea interpretării economice a rezultatelor obținute (estimatorii sunt ușor deplasați).

Recapitulare

Renunțarea la ipoteza de independență a variabilelor explicative din modelele de regresie creează probleme în ceea ce privește estimarea parametrilor și calitatea estimatorilor.

Fenomenul de multicolinearitate este prezent atunci când coeficienții de corelație lineară, calculați pentru perechile de variabile explicative din model, sunt mari în valoare absolută. De asemenea, o valoare a factorului de inflația al dispersiei (VIF) mai mare de 10 indică un fenomen sever de multicolinearitate.

Sintetic, cele mai utilizate proceduri folosite pentru atenuarea fenomenului de multicolinearitate sunt următoarele:

- Eliminarea unor variabile explicative
- Realizarea unor observații suplimentare asupra variabilelor din model (se mărește volumul eșantionului)
- Prelucrarea primară a datelor (calculul ritmurilor de modificare, a sporurilor, indicilor, logaritmare valorilor observate etc.)
- Regresia ridge

10.3. Problemă rezolvată

Fie următoarele serii de date, înregistrate în 20 momente diferite de timp:

t	Y_t	X_{1t}	X_{2t}	t	Y_t	X_{1t}	X_{2t}
1	17.71	5.23	3.02	11	10.12	3.45	1.74
2	15.26	6.23	4.09	12	15.66	8.27	6.27
3	11.48	2.34	1.42	13	8.47	3.59	4.45
4	37.95	5.44	-1.81	14	11.85	6.35	5.23
5	28.27	6.49	0.91	15	17.08	4.22	1.64
6	15.26	6.52	5.18	16	29.68	7.51	2.34
7	0.75	4.40	6.05	17	31.05	6.37	0.22
8	25.98	4.70	1.00	18	3.25	5.59	6.43
9	18.01	6.59	3.72	19	24.73	5.40	0.50
10	17.39	5.57	0.55	20	14.90	3.17	1.66

Observație: datele sunt cele din problemele rezolvate la capitolele 4.3, 7.3 - 9.3.

Testați prezența fenomenului de colinearitate în modelul linear multifactorial de regresie:

$$Y_t = a_0 + a_1 X_{1t} + a_2 X_{2t} + e_t.$$

Rezolvare:

Criteriul lui Klein:

– coeficientul de corelație lineară între X_1 și X_2 este;

$$\text{correl}(X_1, X_2) = 0.2508$$

– coeficientul de determinare din ecuația de regresie este:

$$R^2 = 0.9017.$$

Deoarece $|r_{ij}| < R^2$, criteriul lui Klein nu indică prezența fenomenului de colinearitate între variabilele explicative.

Factorul de inflație al dispersiei:

– în ecuația de regresie $X_2 = b_0 + b_1X_1 + \varepsilon_t$, $R_2^2 = 0.0629$.

– factorul $VIF_i = \frac{1}{1 - R_2^2} = 1.067$.

Deoarece $VIF < 4$, criteriul nu indică prezența fenomenului de colinearitate între variabilele explicative.



10.4. Rezumat

În prezența multicolarității se reduce precizia estimatorilor calculați pentru parametrii modelului, în sensul că abaterea standard mare duce la creșterea intervalului de încredere în care sunt garantați parametrii. De asemenea, interpretarea parametrilor individuali este dificilă. Dacă estimatorii calculați pentru parametrii modelului sunt semnificativi din punct de vedere statistic, sensul și intensitatea influenței sunt în concordanță cu teoria economică, atunci eventuala prezență a fenomenului de multicolaritate poate fi neglijată. Dacă însă, influențele supra calității estimatorilor sunt negative, atunci există mai multe soluții pentru eliminarea multicolarității.



10.5. Test de autoevaluare a cunoștințelor

Fie următoarele serii de date, înregistrate în 20 momente diferite de timp:

t	Y_t	X_{1t}	X_{2t}	t	Y_t	X_{1t}	X_{2t}
1	23.96	5.23	3.02	11	13.31	3.45	1.74

t	Y _t	X _{1t}	X _{2t}	t	Y _t	X _{1t}	X _{2t}
2	23.57	6.23	4.09	12	28.20	8.27	6.27
3	13.25	2.34	1.42	13	14.51	3.59	4.45
4	39.58	5.44	-1.81	14	21.44	6.35	5.23
5	33.67	6.49	0.91	15	20.94	4.22	1.64
6	24.96	6.52	5.18	16	37.53	7.51	2.34
7	9.20	4.40	6.05	17	35.64	6.37	0.22
8	29.68	4.70	1.00	18	13.26	5.59	6.43
9	26.33	6.59	3.72	19	28.62	5.40	0.50
10	21.51	5.57	0.55	20	17.74	3.17	1.66

Testați prezența fenomenului de colinearitate în modelul linear multifactorial de regresie:

$$Y_t = a_0 + a_1X_{1t} + a_2X_{2t} + e_t.$$



10.6. Bibliografie recomandată

- a. Jula, Dorin; Jula, Nicolae-Marius, 2022. *Econometrie*, ediția a 5^a. Editura Mustang, București
- b. Jula, Nicoleta; Jula, Dorin, 2017. *Modelare economică. Modele econometrice și de optimizare*, Editura Mustang, București

Alternativa web:

- a. Hansen, Bruce E. 2017. *Econometrics*. Download for free at:
<http://www.ssc.wisc.edu/~bhansen/econometrics/Econometrics.pdf>
- b. Diebold, Francis X., 2013-2016. *Econometrics*. Department of Economics, University of Pennsylvania. Download for free at:
<http://www.ssc.upenn.edu/~fdiebold/Teaching104/Econometrics.pdf>

Unitatea de învățare 11. Heteroscedasticitatea erorilor



11.1. Obiective

În prezentul capitol sunt analizate problemele legate de *testarea* modului în care se respectă ipoteza privind distribuția erorilor cu o dispersie constantă, *consecințele* nerespectării acestei ipoteze și procedurile de *atenuare* a fenomenului respectiv.

Durata medie de parcurgere a acestei unități de învățare este de aproximativ 3 ore.



11.2. Conținut

Proprietatea erorilor de a nu avea o dispersie constantă se numește *heteroscedasticitate*.

11.2.1. Consecințe ale heteroscedasticității

Dacă fenomenul de heteroscedasticitate a erorilor din modelul de regresie lineară este ignorat, iar pentru estimarea parametrilor se folosește metoda celor mai mici pătrate, atunci, sintetic, cele mai importante consecințe ale ignorării fenomenului de heteroscedasticitate a erorilor sunt următoarele consecințe ale ignorării fenomenului de heteroscedasticitate a erorilor:

- a. Estimatorii parametrilor din model sunt nedeplasați și consistenți.
- b. Estimatorii parametrilor din model nu sunt eficienți (există estimatori care au o dispersie mai mică).

- c. Estimatorii calculați pentru dispersia și covarianța parametrilor sunt deplasați, nu sunt consistenți și nu sunt eficienți.
- d. Testul t Student aplicat pentru analiza semnificației estimatorilor nu este valid. Dacă dispersia erorilor și variația factorului explicativ sunt pozitiv corelate, atunci dispersia corectă a parametrului a_1 este subestimată, astfel încât calculele sugerează o precizie a estimării mai bună decât este în realitate.
- e. Estimatorii parametrilor nu au proprietatea de maximă verosimilitate.

11.2.2. Testarea heteroscedasticității

11.2.2.1. Testul Goldfeld–Quandt

Procedura Goldfeld–Quandt este folosită pentru testarea ipotezei nule H_0 , care presupune lipsa heteroscedasticității erorilor, contra ipotezei alternative H_1 , care admite faptul că dispersia erorilor este corelată cu valorile uneia dintre variabilele explicative (de exemplu, dispersia erorilor crește pe măsură ce cresc valorile acelei variabile explicative care este considerată ca fiind relevantă). Pentru aplicarea testului se admite faptul că o astfel de variabilă explicativă relevantă poate fi identificată. Procedura este următoarea:

1. Se identifică, dintre variabile explicative prezente în model, o variabilă notată Z , de care este legată dispersia erorilor. Să presupunem că σ_t^2 este corelată pozitiv cu Z_t . Se aranjează toate observațiile din eșantionul reținut pentru analiză în ordinea crescătoare a valorilor Z_t .
2. Eșantionul de dimensiune n se divide în două părți de dimensiuni n_1 și n_2 , după eliminarea a m observații situate la mijlocul eșantionului. Numărul observațiilor eliminate este arbitrar dar, în mod obișnuit, se elimină între $1/6$ și $1/5$ dintre observații. Valorile n_1 și n_2 trebuie să fie mai mari decât numărul parametrilor din model.

3. Se calculează estimatorii pentru parametrii modelului separat pentru primele n_1 observații și pentru ultimele n_2 observații (se aplică metoda celor mai mici pătrate separat pentru primele n_1 observații și pentru ultimele n_2 observații).
4. Se calculează VTR_1 – suma totală a reziduurilor din modelul estimat pentru primele n_1 observații și VTR_2 – suma totală a reziduurilor din modelul estimat pentru ultimele n_2 observații:

$$VTR_1 = \sum_{t=1}^{n_1} u_t^2 \quad \text{respectiv} \quad VTR_2 = \sum_{t=n-n_2+1}^n u_t^2$$

5. Dacă numărul variabilelor explicative din model este k , iar modelul conține și termen liber (a_0) atunci

$$F_c = \frac{\frac{VTR_2}{n_2 - (k+1)}}{\frac{VTR_1}{n_1 - (k+1)}} = \frac{\frac{\sum_{t=n-n_2+1}^n u_t^2}{n_2 - (k+1)}}{\frac{\sum_{t=1}^{n_1} u_t^2}{n_1 - (k+1)}}$$

urmează o distribuție F cu $n_1 - (k+1)$ și $n_2 - (k+1)$ grade de libertate. Ipoteza nulă (lipsa heteroscedasticității) este respinsă dacă F_c calculat prin relația **Eroare! Fără sursă de referință.** este mai mare decât F identificat în tabelele distribuției teoretice. Dacă F_c este subunitar, atunci se calculează $1/F_c$ și se compară cu valorile din tabelele distribuției teoretice, deoarece ipoteza alternativă H_1 este, de obicei, $\sigma_2^2 > \sigma_1^2$, unde σ_2^2 este dispersia erorilor pentru ultima parte a seriei, iar σ_1^2 este dispersia erorilor pentru prima parte a seriei.

11.2.2.2. Testul White

Procedura White este următoarea:

1. Se estimează ecuația de regresie inițială prin metoda celor mai mici pătrate și se calculează

$$u_t = Y_t - (\hat{a}_0 + \hat{a}_1 X_{1t} + \hat{a}_2 X_{2t} + \dots + \hat{a}_k X_{kt}), t = 1, 2, \dots, n$$

2. Se construiește ecuația auxiliară:

$$u_t^2 = \alpha_0 + \alpha_1 Z_{1t} + \alpha_2 Z_{2t} + \dots + \alpha_p Z_{pt} + \varepsilon_t$$

pentru care se calculează coeficientul de determinare multiplă R^2 . Dacă oricare dintre parametrii din ecuația auxiliară sunt semnificativi, valoarea coeficientului de determinare R^2 va fi semnificativă.

3. Se testează semnificația coeficientului R^2 calculat pentru ecuația auxiliară. Dacă $nR^2 < \chi_p^2(0.05)$, atunci **nu** respingem ipoteza nulă (erorile **nu** sunt heteroscedastice), la un prag de peste 5%. În expresia nR^2 , n reprezintă volumul eșantionului.

White sugerează și modul de selectare a variabilelor relevante Z . Astfel, dacă modelul de regresie este

$$Y_t = a_0 + a_1 X_{1t} + a_2 X_{2t} + e_t, t = 1, 2, \dots, n$$

atunci ecuația auxiliară se scrie:

$$u_t^2 = \alpha_0 + \alpha_1 X_{1t} + \alpha_2 X_{2t} + \alpha_3 X_{1t}^2 + \alpha_4 X_{2t}^2 + \alpha_5 X_{1t} X_{2t} + \varepsilon_t$$

unde

$$u_t = Y_t - (\hat{a}_0 + \hat{a}_1 X_{1t} + \hat{a}_2 X_{2t}), t = 1, 2, \dots, n$$

Ipoteza nulă $H_0: \alpha_1 = \alpha_2 = \dots = \alpha_5 = 0$ este respinsă dacă

$$nR^2 > \chi_5^2(0.05).$$

Având în vedere faptul că, din tabelul distribuției χ^2 pentru pragul $\alpha = 0.05$ și 5 grade de libertate se identifică $\chi_5^2(0.05) = 11.0705$, relația precedentă este echivalentă cu: $nR^2 > 11.0705$. Adică, dacă nR^2 este mai mare decât pragul de 11.0705, atunci erorile e_t din ecuația inițială sunt heteroscedastice.

Recapitulare

Dacă în estimarea dispersiei este folosită relația $\text{Var}(\hat{A}) = \sigma_e^2 (X'X)^{-1}$ și erorile sunt heteroscedastice, atunci estimatorii obținuți vor fi deplasați față de dispersia reală a parametrilor estimați și nu vor fi eficienți.

11.3. Problemă rezolvată

Fie următoarele serii de date, înregistrate în 20 momente diferite de timp:

t	Y _t	X _{1t}	X _{2t}	t	Y _t	X _{1t}	X _{2t}
1	17.71	5.23	3.02	11	10.12	3.45	1.74
2	15.26	6.23	4.09	12	15.66	8.27	6.27
3	11.48	2.34	1.42	13	8.47	3.59	4.45
4	37.95	5.44	-1.81	14	11.85	6.35	5.23
5	28.27	6.49	0.91	15	17.08	4.22	1.64
6	15.26	6.52	5.18	16	29.68	7.51	2.34
7	0.75	4.40	6.05	17	31.05	6.37	0.22
8	25.98	4.70	1.00	18	3.25	5.59	6.43
9	18.01	6.59	3.72	19	24.73	5.40	0.50
10	17.39	5.57	0.55	20	14.90	3.17	1.66

Observație: datele sunt cele din problemele rezolvate la capitolele 4.3, 7.3 - 10.3.

Testați prezența fenomenului de heteroscedasticitate a erorilor în modelul linear de regresie:

$$Y_t = a_0 + a_1 X_{1t} + a_2 X_{2t} + e_t.$$

Rezolvare:

Aplicăm testul White. Ecuația auxiliară este

$$u_t^2 = \alpha_0 + \alpha_1 X_{1t} + \alpha_2 X_{2t} + \alpha_3 X_{1t}^2 + \alpha_4 X_{2t}^2 + \alpha_5 X_{1t} X_{2t} + \varepsilon_t,$$

pentru care $R^2 = 0.11727$. Rezultă $nR^2 = 20 \cdot 0.11727 = 2.3454$. Cum $nR^2 < \chi_5^2(0.05) = 11.0705$, datele nu indică prezența fenomenului de heteroscedasticitate a erorilor.

Soluția EViews-10 pentru problema precedentă este prezentată (cu bold) în tabelul următor:

Heteroskedasticity Test: White

F-statistic	0.371978	Prob. F(5,14)	0.8595
Obs*R-squared	2.345399	Prob. Chi-Square(5)	0.7996
Scaled explained SS	3.915248	Prob. Chi-Square(5)	0.5617

Test Equation:

Dependent Variable: RESID²

Method: Least Squares

Sample: 1 20

Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-17.46217	49.45915	-0.353062	0.7293
(X ₁) ²	-1.597045	2.366290	-0.674915	0.5107
X ₁ *X ₂	0.545218	2.193015	0.248616	0.8073
X ₁	14.99270	21.19049	0.707520	0.4909
(X ₂) ²	-0.053966	0.903378	-0.059739	0.9532
X ₂	-4.741261	12.58023	-0.376882	0.7119
R-squared	0.117270	Mean dependent var	8.346471	
Adjusted R-squared	-0.197991	S.D. dependent var	18.40808	
S.E. of regression	20.14815	Akaike info criterion	9.087427	
Sum squared resid	5683.272	Schwarz criterion	9.386147	
Log likelihood	-84.87427	Hannan-Quinn criter.	9.145740	
F-statistic	0.371978	Durbin-Watson stat	1.859719	
Prob(F-statistic)	0.859504			



11.4. Rezumat

Dacă în calcule sunt folosiți estimatorii deplasati ai dispersiilor, atunci testele statistice nu sunt valide, iar intervalele de încredere pentru parametrii modelului vor fi determinate într-un mod incorect. În prezența heteroscedasticității erorilor, calculele sugerează o precizie a estimării mai bună decât este în realitate.



11.5. Test de autoevaluare a cunoștințelor

Fie următoarele serii de date, înregistrate în 20 momente diferite de timp:

t	Y_t	X_{1t}	X_{2t}	t	Y_t	X_{1t}	X_{2t}
1	23.96	5.23	3.02	11	13.31	3.45	1.74
2	23.57	6.23	4.09	12	28.20	8.27	6.27
3	13.25	2.34	1.42	13	14.51	3.59	4.45
4	39.58	5.44	-1.81	14	21.44	6.35	5.23
5	33.67	6.49	0.91	15	20.94	4.22	1.64
6	24.96	6.52	5.18	16	37.53	7.51	2.34
7	9.20	4.40	6.05	17	35.64	6.37	0.22
8	29.68	4.70	1.00	18	13.26	5.59	6.43
9	26.33	6.59	3.72	19	28.62	5.40	0.50
10	21.51	5.57	0.55	20	17.74	3.17	1.66

Testați prezența fenomenului de heteroscedasticitate a erorilor în modelul linear multifactorial de regresie $Y_t = a_0 + a_1X_{1t} + a_2X_{2t} + e_t$ și ecuația auxiliară $u_t^2 = \alpha_0 + \alpha_1X_{1t}^2 + \alpha_2X_{2t}^2 + \varepsilon_t$.



11.6. Bibliografie recomandată

- a. Jula, Dorin; Jula, Nicolae-Marius, 2022. *Econometrie*, ediția a 5^a. Editura Mustang, București
- b. Jula, Nicoleta; Jula, Dorin, 2017. *Modelare economică. Modele econometrice și de optimizare*, Editura Mustang, București

Alternativa web:

- a. Hansen, Bruce E. 2017. *Econometrics*. Download for free at:
<http://www.ssc.wisc.edu/~bhansen/econometrics/Econometrics.pdf>
- b. Diebold, Francis X., 2013-2016. *Econometrics*. Department of Economics, University of Pennsylvania. Download for free at:
<http://www.ssc.upenn.edu/~fdiebold/Teaching104/Econometrics.pdf>

Unitatea de învățare 12. Autocorelarea erorilor



12.1. Obiective

În prezentul capitol sunt analizate problemele legate de *testarea* modului în care se respectă ipoteza privind autocorelarea erorilor, *consecințele* nerespectării acestei ipoteze și procedurile de *atenuare* a fenomenului respectiv.

Durata medie de parcurgere a acestei unități de învățare este de aproximativ 3 ore.



12.2. Conținut

12.2.1. Consecințe ale autocorelării erorilor

Dacă fenomenul de autocorelare a erorilor din modelul de regresie lineară este ignorat, iar pentru estimarea parametrilor se folosește metoda celor mai mici pătrate, atunci, sintetic, consecințele ignorării fenomenului de autocorelare a erorilor sunt următoarele:

Consecințe ale ignorării fenomenului de autocorelare a erorilor

- a. Estimatorii parametrilor din model sunt nedepasați și consistenți.
- b. Estimatorii parametrilor din model nu sunt eficienți și nu au proprietatea de maximă verosimilitate.
- c. Estimatorii calculați pentru dispersia și covarianța parametrilor sunt depasați, nu sunt consistenți și nu sunt eficienți.

- d. Testul t statistic (Student) aplicat pentru analiza semnificației estimatorilor nu este valid.
- e. Valorile t -Student calculate pentru estimarea semnificației parametrilor sunt supradimensionate, ceea ce sugerează o semnificație a parametrilor mai mare decât este în realitate.
- f. Abaterea standard a erorilor este subdimensionată față de valoarea reală și, în consecință, coeficientul de determinare R^2 este supradimensionat, ceea ce indică o ajustare mai bună decât este în realitate.

12.2.2. Testarea autocorelării erorilor

12.2.2.1. Testul Durbin – Watson

Testul Durbin – Watson este cea mai cunoscută procedură utilizată pentru identificarea autocorelării de ordinul întâi a erorilor din modelele de regresie lineară.

Statistica Durbin – Watson se calculează astfel:

$$dw = \frac{\sum_{t=2}^n (u_t - u_{t-1})^2}{\sum_{t=1}^n u_t^2} \in [0, 4]$$

unde u_t sunt valorile variabilei reziduale din ecuația de regresie lineară, ecuație estimată pornind de la datele din eșantionul selectat.

Procedura descrisă de Durbin și Watson pentru testarea autocorelării erorilor în modelul AR(1) de tipul

$$Y_t = a_0 + a_1 X_{1t} + a_2 X_{2t} + \dots + a_k X_{kt} + e_t, t = \overline{1, n}$$

$$e_t = \rho e_{t-1} + \varepsilon_t$$

este următoarea:

1. Se estimează modelul prin metoda celor mai mici pătrate și se calculează reziduurile u_t .
2. Se calculează statistica

$$dw = \frac{\sum_{t=2}^n (u_t - u_{t-1})^2}{\sum_{t=1}^n u_t^2}$$

- 3a. Pentru a testa ipoteza $H_0: \rho = 0$ (lipsa autocorelării erorilor), contra ipotezei alternative $H_1: \rho \neq 0$ (prezența fenomenului de autocorelare a erorilor) se aplică testul Durbin – Watson bilateral. Din tabelele testului bilateral Durbin – Watson se selectează (pornind de la nivelul de semnificație ales – de obicei, 0.05), valorile critice d_L și d_U , pentru k – numărul de variabile explicative din model și n – dimensiunea eșantionului.
 - Se acceptă ipoteza H_0 dacă $d_U \leq dw \leq 4 - d_U$
 - Se respinge H_0 dacă $dw \leq d_L$ sau $dw \geq 4 - d_U$.
 - Dacă $d_L \leq dw \leq d_U$ sau $4 - d_U \leq dw \leq 4 - d_L$, testul este neconcludent.

Grafic, situațiile care pot să apară atunci când valoarea calculată a statisticii dw se găsește în diferite puncte de pe dreapta $[0, 4]$ sunt descrise în diagrama din figura 12.1.

0	d_L	d_U	2	$4 - d_U$	$4 - d_L$	4
Autocorelare pozitivă	Testul nu este concludent	Erorile nu sunt autocorelate	Testul nu este concludent	Autocorelare negativă		

Figura 12-1: Testul bilateral Durbin – Watson pentru autocorelarea erorilor

Dezavantaje ale testului Durbin-Watson

- Un prim dezavantaj al testului Durbin-Watson constă în faptul că *se aplică doar pentru identificarea autocorelării de ordinul I*. Adică, testul DW nu

depistează, de exemplu, fenomenul de sezonabilitate, de tipul $e_t = \rho e_{t-4} + \varepsilon_t$, fenomen des întâlnit în economie.

- Testul se aplică doar *dacă modelul de regresie are termen liber; dacă modelul nu are termen liber, valorile critice din tabelele DW nu sunt valabile* - se folosesc pragurile calculate de Farebrother (1980)²³.
- Un alt dezavantaj al testului Durbin – Watson rezidă în faptul că *rezultatele ar putea fi eronate dacă se aplică pentru modelele care conțin variabile decalate în timp*.

12.2.2.2. Testul Lagrange

Testul construit pe baza multiplicatorilor Lagrange pentru identificarea fenomenului de autocorelare a erorilor (testul LM) a fost propus de Breusch²⁴ și Godfrey²⁵. Acest test are o aplicabilitate mai mare decât testul Durbin – Watson, deoarece nu este restricționat la autocorelarea de ordinul I a erorilor. Aplicarea testului LM presupune parcurgerea următoarei proceduri:

1. Se rezolvă ecuația

$$Y_t = a_0 + a_1X_{1t} + a_2X_{2t} + \dots + a_kX_{kt} + e_t$$

prin metoda celor mai mici pătrate și se calculează reziduurile u_t .

2. Se construiește ecuația auxiliară

$$u_t = b_0 + b_1X_{1t} + \dots + b_kX_{kt} + \rho_1u_{t-1} + \rho_2u_{t-2} + \dots + \rho_pu_{t-p} + v_t.$$

pentru $t = p + 1, p + 2, \dots, n$.

²³ Farebrother, Richard William., 1980. "The Durbin-Watson Test for Serial Correlation when there is no Intercept in the Regression", *Econometrica*, 48, pp. 1553-1563.

²⁴ Breusch Trevor S., 1978. *Testing for autocorrelation in dynamic linear models*, în *Australian Economic Papers*, 17, pag. 334-355

²⁵ Godfrey Leslie G., 1988. *Specification Tests in Econometrics*, Cambridge University Press

3. Se rezolvă ecuația auxiliară de regresie și se calculează $(n - p)R^2$, unde R^2 este coeficientul de determinare multiplă. Dacă $(n - p)R^2 > \chi_p^2(\alpha)$, atunci se respinge ipoteza nulă $H_0: \rho_1 = \rho_2 = \dots = \rho_p$ și se admite ipoteza alternativă: H_1 : *cel puțin un parametru ρ_i este semnificativ diferit de zero*, ceea ce înseamnă că erorile sunt autocorelate.

În alegerea ordinului p al procesului autoregresiv $AR(p)$ se pornește de la analiza economică și statistică a datelor disponibile. În detaliu, această problemă este studiată în modelele de tip Box-Jenkins privind seriile de timp.

Observație:

Programul EViews consideră ecuația auxiliară pentru $t = 1, \dots, n$ și setează la zero valorile u_{t-p} care lipsesc (datorită aplicării operatorului de întârziere).

Recapitulare

În prezența autocorelării erorilor, estimatorii calculați prin metoda celor mai mici pătrate pentru parametrii modelului de regresie lineară rămân nedeplasați și consistenți, deoarece în demonstrarea proprietăților respective nu s-a folosit ipoteza de lipsă a autocorelării erorilor (ca, de altfel, nici ipoteza de lipsă a heteroscedasticității). În schimb, estimatorii nu sunt eficienți, în sensul că există estimatori ai parametrilor modelului care au o dispersie mai mică decât dispersia estimatorilor calculați prin metoda celor mai mici pătrate.

Testul Durbin-Watson se aplică pentru identificarea autocorelării erorilor de gradul I.

Testul construit pe baza multiplicatorilor Lagrange pentru identificarea fenomenului de autocorelare a erorilor nu este restricționat la autocorelarea de ordinul I a erorilor și rămâne valid în prezența regresorilor construiți pornind de

la starea variabilei dependente în perioadele trecute. În consecință, are o aplicabilitate mai mare decât testul Durbin – Watson.

12.3. Problemă rezolvată

Fie următoarele serii de date, înregistrate în 20 momente diferite de timp:

t	Y _t	X _{1t}	X _{2t}	t	Y _t	X _{1t}	X _{2t}
1	17.71	5.23	3.02	11	10.12	3.45	1.74
2	15.26	6.23	4.09	12	15.66	8.27	6.27
3	11.48	2.34	1.42	13	8.47	3.59	4.45
4	37.95	5.44	-1.81	14	11.85	6.35	5.23
5	28.27	6.49	0.91	15	17.08	4.22	1.64
6	15.26	6.52	5.18	16	29.68	7.51	2.34
7	0.75	4.40	6.05	17	31.05	6.37	0.22
8	25.98	4.70	1.00	18	3.25	5.59	6.43
9	18.01	6.59	3.72	19	24.73	5.40	0.50
10	17.39	5.57	0.55	20	14.90	3.17	1.66

Observație: datele sunt cele din problemele rezolvate la capitolele 4.3, 7.3 - 10.3.

Testați prezența fenomenului de autocorelare a erorilor în modelul linear de regresie:

$$Y_t = a_0 + a_1X_{1t} + a_2X_{2t} + e_t.$$

Rezolvare:

$$\text{Testul Durbin-Watson: } dw = \frac{\sum_{t=2}^n (u_t - u_{t-1})^2}{\sum_{t=1}^n u_t^2} = 1.633.$$

Din tabelul Durbin-Watson, pentru $n = 20$ și $k = 2$ se identifică: $d_L = 1.10$ și $d_U = 1.54$. Deoarece $d_U < dw < 4 - d_U$ nu se respinge ipoteza nulă (lipsa autocorelării erorilor, de ordinul I).

Testul Lagrange: considerăm $p = 4$. Ecuația auxiliară este:

$$u_t = 1.9648 - 0.2769X_{1t} - 0.1665X_{2t} + \\ + 0.2815u_{t-1} - 0.4022u_{t-2} - 0.0728u_{t-3} - 0.0125u_{t-4},$$

pentru care

$$R^2 = 0.17367.$$

Calculăm $nR^2 = 3.473$.

Din tabelul distribuției teoretice χ^2 , pentru $\alpha = 0.05$ și numărul gradelor de libertate $df = 4$ se identifică pragul

$$\chi^2(0.05; 4) = 9.488.$$

Deoarece

$$nR^2 < \chi^2(0.05; 4),$$

respingem ipoteza de autocorelare a erorilor, cel puțin până la pragul 4.

Soluția EViews-10 pentru problema precedentă este prezentată (cu bold) în tabelul următor:

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.683053	Prob. F(4,13)	0.6161
Obs*R-squared	3.473398	Prob. Chi-Square(4)	0.4819

Test Equation:

Dependent Variable: RESID

Method: Least Squares

Sample: 1 20

Included observations: 20

Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.964813	3.168773	0.620055	0.5459
X ₁	-0.276869	0.583901	-0.474171	0.6432
X ₂	-0.166452	0.368316	-0.451927	0.6588
RESID ₍₋₁₎	0.281520	0.302480	0.930704	0.3690
RESID ₍₋₂₎	-0.402227	0.310446	-1.295645	0.2176
RESID ₍₋₃₎	-0.072801	0.294371	-0.247309	0.8085
RESID ₍₋₄₎	-0.012549	0.291992	-0.042976	0.9664
R-squared	0.173670	Mean dependent var	3.66E-15	
Adjusted R-squared	-0.207713	S.D. dependent var	2.964078	

Variable	Coefficient	Std. Error	t-Statistic	Prob.
S.E. of regression	3.257404	Akaike info criterion	5.468955	
Sum squared resid	137.9388	Schwarz criterion	5.817461	
Log likelihood	-47.68955	Hannan-Quinn criter.	5.536987	
F-statistic	0.455369	Durbin-Watson stat	2.027747	
Prob(F-statistic)	0.828851			



12.4. Rezumat

În prezența autocorelării erorilor este afectată calitatea estimatorilor calculați prin metoda celor mai mici pătrate pentru parametrii modelului de regresie. Autocorelarea erorilor presupune existența unei covarianțe nenule între erorile din ecuația de regresie. Pentru identificarea autocorelării de ordinul I a erorilor se folosește testul Durbin-Watson. Testul construit pe baza multiplicatorilor Lagrange pentru identificarea fenomenului de autocorelare a erorilor nu este restricționat la autocorelarea de ordinul I a erorilor.



12.5. Test de autoevaluare a cunoștințelor

Fie următoarele serii de date, înregistrate în 20 momente diferite de timp:

t	Y_t	X_{1t}	X_{2t}	t	Y_t	X_{1t}	X_{2t}
1	23.96	5.23	3.02	11	13.31	3.45	1.74
2	23.57	6.23	4.09	12	28.20	8.27	6.27
3	13.25	2.34	1.42	13	14.51	3.59	4.45
4	39.58	5.44	-1.81	14	21.44	6.35	5.23
5	33.67	6.49	0.91	15	20.94	4.22	1.64
6	24.96	6.52	5.18	16	37.53	7.51	2.34
7	9.20	4.40	6.05	17	35.64	6.37	0.22

t	Y_t	X_{1t}	X_{2t}	t	Y_t	X_{1t}	X_{2t}
8	29.68	4.70	1.00	18	13.26	5.59	6.43
9	26.33	6.59	3.72	19	28.62	5.40	0.50
10	21.51	5.57	0.55	20	17.74	3.17	1.66

Testați prezența fenomenului de autocorelare a erorilor în modelul linear multifactorial de regresie $Y_t = a_0 + a_1X_{1t} + a_2X_{2t} + e_t$, pentru $p = 2$.



12.6. Bibliografie recomandată

- a. Jula, Dorin; Jula, Nicolae-Marius, 2022. *Econometrie*, ediția a 5^a. Editura Mustang, București
- b. Jula, Nicoleta; Jula, Dorin, 2017. *Modelare economică. Modele econometrice și de optimizare*, Editura Mustang, București

Alternativa web:

- a. Hansen, Bruce E. 2017. *Econometrics*. Download for free at:
<http://www.ssc.wisc.edu/~bhansen/econometrics/Econometrics.pdf>
- b. Diebold, Francis X., 2013-2016. *Econometrics*. Department of Economics, University of Pennsylvania. Download for free at:
<http://www.ssc.upenn.edu/~fdiebold/Teaching104/Econometrics.pdf>

Unitatea de învățare 13.

Normalitatea distribuției erorilor



13.1. Obiective

Normalitatea distribuției erorilor reprezintă o condiție esențială pentru evaluarea calității estimatorilor calculați prin metoda celor mai mici pătrate, în cazul modelelor lineare de regresie. Aceasta deoarece majoritatea rezultatelor referitoare la regresia lineară au fost dezvoltate pornind de la ipoteza normalității distribuției erorilor. În prezentul capitol sunt analizate problemele legate de *testarea* ipotezei de normalitate a distribuției erorilor, *consecințele* nerespectării acestei ipoteze și procedurile de *atenuare* a fenomenului respectiv.

Durata medie de parcurgere a acestei unități de învățare este de aproximativ 3 ore.



13.2. Conținut

13.2.1. Consecințe ale nerespectării ipotezei de normalitate a distribuției erorilor din modelul linear de regresie

În sinteză, consecințele nerespectării ipotezei de normalitate sunt următoarele:

- a. Estimatorii parametrilor din model sunt nedeplasați și consistenți.
- b. Estimatorii parametrilor din model nu sunt eficienți.
- c. Estimatorii parametrilor din model nu au proprietatea de maximă verosimilitate.
- d. Testul t statistic (Student) aplicat pentru analiza semnificației estimatorilor nu este valid

Recapitulare

Normalitatea distribuției erorilor reprezintă o condiție esențială pentru evaluarea calității estimatorilor calculați prin metoda celor mai mici pătrate, în cazul modelelor lineare de regresie. Aceasta deoarece majoritatea rezultatelor referitoare la regresia lineară au fost dezvoltate pornind de la ipoteza normalității distribuției erorilor.

13.2.2. Testarea ipotezei de normalitate a distribuției erorilor din modelul linear de regresie. Testul Jarque-Bera

În literatura de specialitate sunt discutate două clase de teste pentru evaluarea distribuției normale a unei serii aleatorii:

- teste bazate pe statistici descriptive ale eșantionului (teste privind asimetria, teste asupra coeficientului de aplatizare, teste combinate (omnibus) – testul Jarque-Bera, testul D'Agostino-Pearson ș.a.;
- teste de tip "cea mai bună ajustare" (engl. – best fit), bazate pe compararea distribuției empirice cu o distribuție dată – în acest caz, distribuția normală, de cele mai multe ori pornind de la funcția de distribuție cumulativă (ex. testele Kolmogorov-Smirnov, Lilliefors, Anderson-Darling, Cramér-von Mises, Shapiro-Wilk, sau Shapiro-Francia.

Unul dintre cele mai cunoscute teste privind normalitatea erorilor este **testul Jarque-Bera** (Bera & Jarque C., 1981 și 1982)²⁶, test care folosește momentele centrate ale reziduurilor pentru a estima distribuția erorilor. Un estimator nedeplasat pentru momentul centrat de ordinul p al erorilor, construit pornind de la reziduurile din ecuația de regresie este μ_p , calculat astfel:

$$\mu_p = \frac{\sum_{t=1}^n u_t^p}{n}$$

În forma consacrată a testului Jarque-Bera, se utilizează *coeficientul de asimetrie* (S) calculat astfel:

$$S = \frac{\mu_3}{\sqrt{\mu_2^3}} = \frac{\left(\frac{1}{n} \sum_{t=1}^n u_t^3 \right)}{\left(\frac{1}{n} \sum_{t=1}^n u_t^2 \right)^{\frac{3}{2}}}, \text{ echivalent cu } S^2 = \frac{\mu_3^2}{\mu_2^3}$$

și *coeficientul de aplatizare* K (kurtosis) determinat pe baza relației:

$$K = \frac{\mu_4}{\mu_2^2} = \frac{\left(\frac{1}{n} \sum_{t=1}^n u_t^4 \right)}{\left(\frac{1}{n} \sum_{t=1}^n u_t^2 \right)^2}, \text{ sau } \text{excesul } (K - 3).$$

Dacă *modelul de regresie lineară conține și termenul liber*, atunci suma reziduurilor este zero și statistica Jarque-Bera este

$$JB = n \left[\frac{S^2}{6} + \frac{(K - 3)^2}{24} \right].$$

²⁶ Bera, Anil K., Jarque, Carlos M., 1981. "Efficient Tests for Normality, Heteroscedasticity, and Serial Independence of Regression Residuals: Monte Carlo Evidence", în *Economics Letters*, 7, pp. 313-318 și Bera, Anil K., Jarque, Carlos M., 1982. "Model Specification Tests: A Simultaneous Approach", în *Journal of Econometrics*, 20, pp. 59-82

Pentru distribuția normală, asimetria S (skewness) și excesul sunt zero (coeficientul de aplatizare este 3). Cu alte cuvinte, ipoteza nulă din testul Jarque-Bera se formulează astfel:

$$H_0: S = 0 \text{ și } K = 3,$$

(parametrul S este comparat cu zero, iar coeficientul de aplatizare este comparat cu 3). Dacă $JB > \chi^2_2(\alpha)$, atunci, la pragul de semnificație α , se respinge H_0 , conform căreia erorile sunt normal distribuite.

În aplicarea statisticii JB trebuie să se țină seama de faptul că testele de acest tip sunt asimptotice, astfel încât interpretarea rezultatelor are sens doar pentru dimensiuni mari ale eșantionului.

Recapitulare

Testul Jarque-Bera în forma standard se aplică doar dacă modelul conține și termen liber.

13.3. Problemă rezolvată

Fie următoarele serii de date, înregistrate în 20 momente diferite de timp:

t	Y_t	X_t	t	Y_t	X_t
1	4.84	-2.12	11	9.61	-0.38
2	26.27	9.03	12	18.62	4.44
3	18.25	3.54	13	28.90	11.22
4	13.57	3.10	14	20.63	5.45
5	16.83	4.23	15	14.31	2.27
6	22.53	6.34	16	37.15	14.61
7	17.70	4.28	17	18.30	5.27
8	35.75	11.14	18	23.89	8.31
9	5.46	-2.54	19	20.57	2.89
10	20.69	4.44	20	14.29	2.67

Observație: datele sunt cele din problema rezolvată la capitolul 2.3.

Testați normalitatea distribuției erorilor din modelul linear unifactorial de regresie: $Y_t = a_0 + a_1X_t + e_t$.

Rezolvare:

Testul Jarque-Bera:

$$S = 1.225495$$

$$K = 3.977533$$

$$JB = n \left[\frac{S^2}{6} + \frac{(K-3)^2}{24} \right] = 5.802.$$

Din tabelul distribuției teoretice χ^2 , pentru $\alpha = 0.05$ și numărul gradelor de libertate $df = 2$ se identifică pragul $\chi^2(0.05; 2) = 5.991$. Deoarece $JB < \chi^2(0.05; 2)$, nu se respinge H_0 , conform căreia erorile sunt normal distribuite.

Soluția EViews-10 pentru problema precedentă este prezentată în figura următoare:

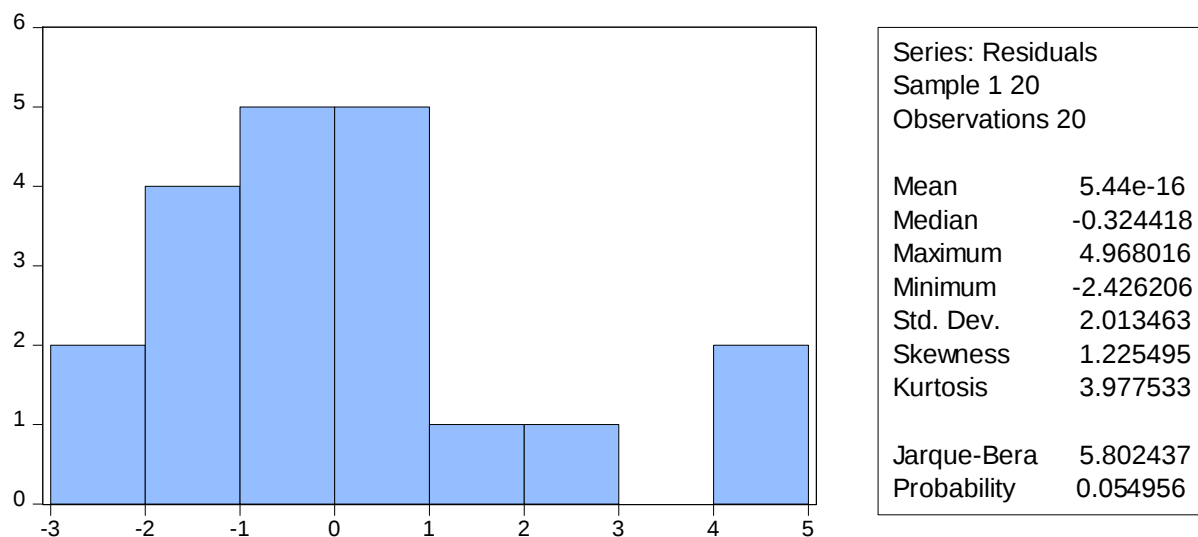


Figura 13-1. Testul Jarque-Bera

13.4. Rezumat



Dacă erorile din modelul linear de regresie nu sunt repartizate normal, estimatorii calculați prin metoda celor mai mici pătrate nu sunt estimatori de maximă verosimilitate. De asemenea, dacă estimatorii nu vor urma o lege de distribuție normală, atunci testele de semnificație prezentate, bazate pe statistica t-Student nu sunt valide. Aceasta deoarece, prin definiție, statistica t-Student se construiește prin raportarea estimatorului – ca variabilă aleatoare repartizată normal la radical dintr-o variabilă aleatoare repartizată χ^2 , divizată prin numărul gradelor de libertate.

Mai mult, testele uzuale privind heteroscedasticitatea și autocorelarea erorilor (de exemplu, testul Durbin-Watson) nu sunt valide, deoarece pornesc de la aceeași ipoteză, a normalității distribuției erorilor.



13.5. Test de autoevaluare a cunoștințelor

Fie următoarele serii de date, înregistrate în 20 momente diferite de timp:

t	Y_t	X_{1t}	X_{2t}	t	Y_t	X_{1t}	X_{2t}
1	23.96	5.23	3.02	11	13.31	3.45	1.74
2	23.57	6.23	4.09	12	28.20	8.27	6.27
3	13.25	2.34	1.42	13	14.51	3.59	4.45
4	39.58	5.44	-1.81	14	21.44	6.35	5.23
5	33.67	6.49	0.91	15	20.94	4.22	1.64
6	24.96	6.52	5.18	16	37.53	7.51	2.34
7	9.20	4.40	6.05	17	35.64	6.37	0.22
8	29.68	4.70	1.00	18	13.26	5.59	6.43
9	26.33	6.59	3.72	19	28.62	5.40	0.50
10	21.51	5.57	0.55	20	17.74	3.17	1.66

Testați normalitatea distribuției erorilor din modelul linear multifactorial de regresie:

$$Y_t = a_0 + a_1X_{1t} + a_2X_{2t} + e_t.$$



13.6. Bibliografie recomandată

- a. Jula, Dorin; Jula, Nicolae-Marius, 2022. *Econometrie*, ediția a 5^a. Editura Mustang, București
- b. Jula, Nicoleta; Jula, Dorin, 2017. *Modelare economică. Modele econometrice și de optimizare*, Editura Mustang, București

Alternativa web:

- a. Hansen, Bruce E. 2017. *Econometrics*. Download for free at:
<http://www.ssc.wisc.edu/~bhansen/econometrics/Econometrics.pdf>
- b. Diebold, Francis X., 2013-2016. *Econometrics*. Department of Economics, University of Pennsylvania. Download for free at:
<http://www.ssc.upenn.edu/~fdiebold/Teaching104/Econometrics.pdf>

Unitatea de învățare 14. Prognoza econometrică



14.1. Obiective

În sens econometric, *prognoza* reprezintă o anticipare cantitativă a unor evenimente sau condiții viitoare, pornind de la un set de informații disponibile. În prezentul capitol sunt prezentate principalele tehnici de prognoză utilizate pentru modelele lineare (unifactoriale și multifactoriale).

Durata medie de parcurgere a acestei unități de învățare este de aproximativ 3 ore.



14.2. Conținut

Realizarea unei prognoze presupune anticiparea a două elemente. Pe de o parte, prognoza implică determinarea valorii medii a variabilei Y la un moment viitor $n+1$, sau, mai general, $f = n+p$ (unde $p \geq 1$), atunci când parametrii ecuației de regresie sunt estimați pe baza unei selecții realizate la momentele $1, 2, \dots, n$. Pe de altă parte, este necesar calculul împrăștierii probabile a valorilor prognozate în jurul mediei estimate (calculul dispersiei erorilor de prognoză).

14.2.1. Prognoza în cazul modelului linear unifactorial de regresie

Valoarea anticipată a variabilei endogene, la orizontul de prognoză f (notată $\hat{Y}_f$) se calculează astfel:

$$\hat{Y}_f = \hat{a}_0 + \hat{a}_1 X_f,$$

unde $\hat{a}_0$ și $\hat{a}_1$ sunt estimatorii din ecuația de regresie, iar X_f este valoarea anticipată a variabilei exogene X la momentul de prognoză f .

Un estimator nedeplasat al dispersiei erorilor calculat pornind de la dispersia de selecție a variabilei reziduale este dat de expresia:

$$s_u^2 = \frac{\sum_{t=1}^n u_t^2}{n-2}.$$

Un estimator nedeplasat pentru dispersia erorilor de prognoză este:

$$s_f^2 = s_u^2 \left(1 + \frac{1}{n} + \frac{(X_{n+1} - \bar{X})^2}{\sum (X_t - \bar{X})^2} \right)$$

Intervalul de prognoză pentru Y_f se calculează astfel:

$$\hat{Y}_f - t_{n-2;\alpha} s_f \leq Y_f \leq \hat{Y}_f + t_{n-2;\alpha} s_f$$

unde $t_{n-2;\alpha}$ este valoarea din distribuția teoretică t-Student, *testul bilateral*, pentru $n-2$ grade de libertate (n fiind dimensiunea eșantionului), valoare corespunzătoare unui grad de încredere de $(1-\alpha)$.

14.2.2. Prognoza în cazul modelului linear multifactorial de regresie

Să presupunem că modelul de regresie lineară,

$$Y = XA + e,$$

conține k regresori (variabile explicative) și este estimat pornind de la o selecție de volum n , astfel încât vectorul estimatorilor calculați pentru parametrii modelului prin metoda celor mai mici pătrate este dat de relația:

$$\hat{A} = (X'X)^{-1}X'Y.$$

Atunci, prognoza pentru momentul f poate fi construită astfel:

$$\hat{Y}_f = X_f \hat{A}$$

Un estimator nedeplasat al dispersiei erorilor de prognoză, estimator calculat pe baza datelor din eșantion este dat de expresia:

$$s_f^2 = s_u^2 \left[1 + X_f (X'X)^{-1} X_f' \right]$$

unde
$$s_u^2 = \frac{\sum_{t=1}^n u_t^2}{n - k - 1}$$

Un interval de încredere pentru valorile de prognoză, calculat cu un grad de încredere de $(1-\alpha)$, este dat de relația următoare:

$$\hat{Y}_f - t_{n-k-1;\alpha} s_f \leq Y_f \leq \hat{Y}_f + t_{n-k-1;\alpha} s_f,$$

unde $t_{n-k-1;\alpha}$ este valoarea din distribuția teoretică t-Student, *testul bilateral*, pentru $n-k-1$ grade de libertate (n fiind dimensiunea eșantionului, iar k – numărul variabilelor exogene din model), valoare corespunzătoare unui grad de încredere de $(1-\alpha)$.

Recapitulare

Pentru modelul linear unifactorial de regresie, dispersia erorilor de prognoză se calculează astfel:

$$s_f^2 = s_u^2 \left(1 + \frac{1}{n} + \frac{(X_{n+1} - \bar{X})^2}{\sum (X_t - \bar{X})^2} \right),$$

unde
$$s_u^2 = \frac{\sum_{t=1}^n u_t^2}{n - 2}$$

Pentru modelul linear multifactorial de regresie, dispersia erorilor de prognoză se calculează astfel:

$$s_f^2 = s_u^2 \left[1 + X_f (X'X)^{-1} X_f' \right]$$

unde
$$s_u^2 = \frac{\sum_{t=1}^n u_t^2}{n - k - 1}.$$

14.3. Problemă rezolvată

Fie următoarele serii de date, înregistrate în 20 momente diferite de timp:

t	Y _t	X _{1t}	X _{2t}	t	Y _t	X _{1t}	X _{2t}
1	17.71	5.23	3.02	11	10.12	3.45	1.74
2	15.26	6.23	4.09	12	15.66	8.27	6.27
3	11.48	2.34	1.42	13	8.47	3.59	4.45
4	37.95	5.44	-1.81	14	11.85	6.35	5.23
5	28.27	6.49	0.91	15	17.08	4.22	1.64
6	15.26	6.52	5.18	16	29.68	7.51	2.34
7	0.75	4.40	6.05	17	31.05	6.37	0.22
8	25.98	4.70	1.00	18	3.25	5.59	6.43
9	18.01	6.59	3.72	19	24.73	5.40	0.50
10	17.39	5.57	0.55	20	14.90	3.17	1.66

Observație: datele sunt cele din problemele rezolvate la capitolele 4.3, 7.3 - 10.3.

Prognozați dinamica variabilei endogene Y pentru momentul f = 21, dacă X_{1,21} = 4 și X_{2,21} = 2.

Rezolvare:

Elementele folosite în calcul sunt următoarele (vezi rezolvarea problemei din capitolul 4-3):

$$(X'X)^{-1} = \begin{vmatrix} 0.7176 & -0.1209 & -0.0067 \\ -0.1209 & 0.0245 & -0.0040 \\ -0.0067 & -0.0040 & 0.0103 \end{vmatrix} \quad \hat{A} = \begin{vmatrix} 7.70335 \\ 3.72858 \\ -3.65884 \end{vmatrix}$$

Prognoza la momentul f = 21 este:

$$\hat{Y}_{21} = X_{21}\hat{A},$$

unde $X_{21} = [1 \ 4 \ 2],$

adică $\hat{Y}_{21} = 15.3$

Dispersia erorilor de prognoză este:

$$s_{21}^2 = s_u^2 \left[1 + X_{21}(X'X)^{-1}X'_{21} \right]$$

iar abaterea standard se calculează ca radical din valoarea respectivă. Rezultă

$$s_f = 3.378.$$

Pentru

$$df = 20 - 2 - 1 = 17 \text{ grade de libertate}$$

și $\alpha = 0.1$, valoarea identificată în tabelul distribuției teoretice bilaterale t-Student este $t_{17;0.1} = 1.74$. Atunci, $t_{17;0.1} \cdot s_f = 5.9$. Aceasta înseamnă că valoarea prognozată a variabilei endogene la momentul $f = 21$ se găsește, cu o probabilitate de 90% în intervalul $[9.4; 21.2]$.

Soluția EViews-10 pentru problema precedentă este prezentată în figura următoare:

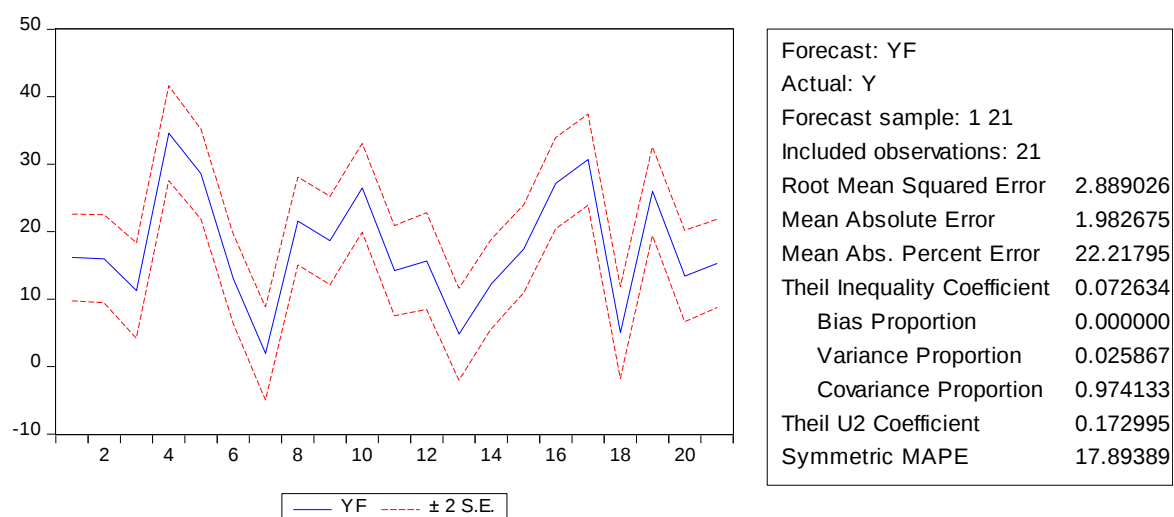


Figura 14-1. Prognoza econometrică



14.4. Rezumat

Dispersia erorilor de prognoză depinde de dispersia erorilor din model, de dimensiunea eșantionului, de variația în eșantion a valorilor exogene și de distanța dintre valoarea anticipată a variabilei explicative și media variabilei respective.

Intervalul de prognoză pentru Y_f se calculează astfel:

– pentru modelul unifactorial $\hat{Y}_f - t_{n-2;\alpha} s_f \leq Y_f \leq \hat{Y}_f + t_{n-2;\alpha} s_f$

– pentru modelul multifactorial $\hat{Y}_f - t_{n-k-1;\alpha} s_f \leq Y_f \leq \hat{Y}_f + t_{n-k-1;\alpha} s_f$

unde $t_{n-k-1;\alpha}$ este valoarea din distribuția teoretică t-Student, *testul bilateral*, pentru $n-k-1$ grade de libertate (n fiind dimensiunea eșantionului, iar k – numărul variabilelor exogene din model), valoare corespunzătoare unui grad de încredere de $(1-\alpha)$.



14.5. Test de autoevaluare a cunoștințelor

Fie următoarele serii de date, înregistrate în 20 momente diferite de timp:

t	Y_t	X_{1t}	X_{2t}	t	Y_t	X_{1t}	X_{2t}
1	23.96	5.23	3.02	11	13.31	3.45	1.74
2	23.57	6.23	4.09	12	28.20	8.27	6.27
3	13.25	2.34	1.42	13	14.51	3.59	4.45
4	39.58	5.44	-1.81	14	21.44	6.35	5.23
5	33.67	6.49	0.91	15	20.94	4.22	1.64
6	24.96	6.52	5.18	16	37.53	7.51	2.34
7	9.20	4.40	6.05	17	35.64	6.37	0.22
8	29.68	4.70	1.00	18	13.26	5.59	6.43
9	26.33	6.59	3.72	19	28.62	5.40	0.50
10	21.51	5.57	0.55	20	17.74	3.17	1.66

Prognozați dinamica variabilei endogene Y pentru momentul $f = 21$, dacă $X_{1,21} = 5.4$ și $X_{2,21} = 2.7$.



14.6. Bibliografie recomandată

- a. Jula, Dorin; Jula, Nicolae-Marius, 2022. *Econometrie*, ediția a 5^a. Editura Mustang, București

- b. Jula, Nicoleta; Jula, Dorin, 2017. *Modelare economică. Modele econometrice și de optimizare*, Editura Mustang, București

Alternativa web:

- a. Hansen, Bruce E. 2017. *Econometrics*. Download for free at:
<http://www.ssc.wisc.edu/~bhansen/econometrics/Econometrics.pdf>
- b. Diebold, Francis X., 2013-2016. *Econometrics*. Department of Economics, University of Pennsylvania. Download for free at:
<http://www.ssc.upenn.edu/~fdiebold/Teaching104/Econometrics.pdf>

14.7. Teme de control

Din site-ul Băncii Naționale a României (www.bnro.ro), de la secțiunea *Statistică*, subsecțiunea *Bază de date interactivă* selectați serii lunare referitoare la baza monetară și inflație. Construiți, testați și aplicați un model de prognoză a inflației.

Răspunsuri la testele de autoevaluare a cunoștințelor

Unitatea de învățare 1. Modelul econometric

- 1) a
- 2) c
- 3) b

Unitatea de învățare 2. Modelul linear unifactorial

Soluțiile problemei sunt prezentate în tabelul EViews următor:

Dependent Variable: Y

Method: Least Squares

Sample: 1 20

Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.962899	1.328707	4.487747	0.0003
X	3.430279	0.238532	14.38079	0.0000

Unitatea de învățare 3. Ipotezele de aplicare ale modelului linear unifactorial.

Proprietăți ale estimatorilor

- 1) d
- 2) c
- 3) b

Unitatea de învățare 4. Modelul linear multifactorial

Soluțiile problemei sunt prezentate în tabelul EViews următor:

Dependent Variable: Y

Method: Least Squares

Sample: 1 20

Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.703353	2.654574	2.148500	0.0464
X ₁	4.728578	0.490786	9.634702	0.0000
X ₂	-2.658842	0.318214	-8.355517	0.0000

Unitatea de învățare 5. Ipotezele de aplicare ale modelului linear multifactorial. Proprietăți ale estimatorilor

- 1) a
- 2) b
- 3) d

Unitatea de învățare 6. Testarea semnificației parametrilor din modelul linear unifactorial de regresie

Soluțiile problemei sunt prezentate în tabelul EViews următor:

Dependent Variable: Y

Method: Least Squares

Sample: 1 20

Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.962899	1.328707	4.487747	0.0003
X	3.430279	0.238532	14.38079	0.0000

Rezultatele sunt prezentate cu bold în tabelul precedent. Riscul ca parametrii a_0 și a_1 să nu fie semnificativ diferiți de zero este mai mic de 1%.

Unitatea de învățare 7. Testarea semnificației parametrilor din modelul linear multifactorial de regresie

Soluțiile problemei sunt prezentate în tabelul EViews următor:

Dependent Variable: Y

Method: Least Squares

Sample: 1 20

Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.703353	2.654574	2.148500	0.0464
X ₁	4.728578	0.490786	9.634702	0.0000
X ₂	-2.658842	0.318214	-8.355517	0.0000

Rezultatele sunt prezentate cu bold în tabelul precedent. Riscul ca parametrul a_0 (termenul liber) să nu fie semnificativ diferit de zero este 2.34% ($0.0464:2 = 0.0234$, deoarece EViews prezintă rezultatele pentru testul t-Student bilateral), iar riscul ca parametrii a_1 și a_2 să fie de zero este mai mic de 1%.

Unitatea de învățare 8. Acuratețea ajustării

Soluțiile problemei sunt prezentate, cu bold, în tabelul EViews-10 următor:

Dependent Variable: Y

Method: Least Squares

Sample: 1 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.703353	2.654574	2.148500	0.0464
X1	4.728578	0.490786	9.634702	0.0000
X2	-2.658842	0.318214	-8.355517	0.0000
R-squared	0.884719	Mean dependent var	23.84438	
Adjusted R-squared	0.871156	S.D. dependent var	8.729919	
S.E. of regression	3.133589	Akaike info criterion	5.259716	
Sum squared resid	166.9294	Schwarz criterion	5.409076	
Log likelihood	-49.59716	Hannan-Quinn criter.	5.288873	
F-statistic	65.23269	Durbin-Watson stat	1.632575	
Prob(F-statistic)	0.000000			

Unitatea de învățare 9. Specificarea modelului

Modelul	AIC	BIC	HQ
M1: $Y_t = a_0 + a_1X_{1t} + e_t$	6.7903	6.8899	6.8097
M2: $Y_t = a_0 + a_2X_{2t} + e_t$	7.0254	7.1250	7.0449

M3: $Y_t = a_0 + a_1X_{1t} + a_2X_{2t} + e_t$	5.2597	5.4091	5.2889
---	--------	--------	--------

Valorile criteriilor informaționale (AIC, SIC și HQ) sunt cele mai mici pentru modelul complet (M3).

Unitatea de învățare 10. Multicolinearitatea

Criteriul lui Klein:

$$\text{correl}(X_1, X_2) = 0.2508$$

$$R^2 = 0.2508.$$

Deoarece $|r_{ij}| < R^2$, criteriul lui Klein nu indică prezența fenomenului de colinearitate între variabilele explicative.

Factorul de inflație al dispersiei:

$$R_2^2 = 0.0629.$$

$$\text{VIF}_i = \frac{1}{1 - R_2^2} = 1.067.$$

Deoarece $\text{VIF} < 4$, criteriul nu indică prezența fenomenului de colinearitate între variabilele explicative.

Unitatea de învățare 11. Heteroscedasticitatea erorilor

Ecuția auxiliară: $u_t^2 = \alpha_0 + \alpha_1 X_{1t}^2 + \alpha_2 X_{2t}^2 + \varepsilon_t$. Soluțiile problemei sunt prezentate, cu bold, în tabelul EViews-10 următor.

Heteroskedasticity Test: White

F-statistic	0.630797	Prob. F(5,14)	0.5442
Obs*R-squared	1.381690	Prob. Chi-Square(5)	0.5012
Scaled explained SS	2.306499	Prob. Chi-Square(5)	0.3156

Test Equation:

Dependent Variable: RESID²

Method: Least Squares

Sample: 1 20

Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	13.29590	9.389591	1.416025	0.1748
$(X_1)^2$	-0.026314	0.284688	-0.092432	0.9274
$(X_2)^2$	-0.327299	0.319477	-1.024484	0.3200
R-squared	0.069085	Mean dependent var	8.346471	
Adjusted R-squared	-0.040435	S.D. dependent var	18.40808	
S.E. of regression	18.77656	Akaike info criterion	8.840576	
Sum squared resid	5993.504	Schwarz criterion	8.989936	
Log likelihood	-85.40576	Hannan-Quinn criter.	8.869733	
F-statistic	0.630797	Durbin-Watson stat	1.966739	
Prob(F-statistic)	0.544173			

Unitatea de învățare 12. Autocorelarea erorilor

Soluțiile problemei (pentru $p = 2$) sunt prezentate, cu bold, în tabelul EViews-10 următor:

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	1.530002	Prob. F(4,13)	0.2485
Obs*R-squared	3.388708	Prob. Chi-Square(4)	0.1837

Test Equation:

Dependent Variable: RESID

Method: Least Squares

Sample: 1 20

Included observations: 20

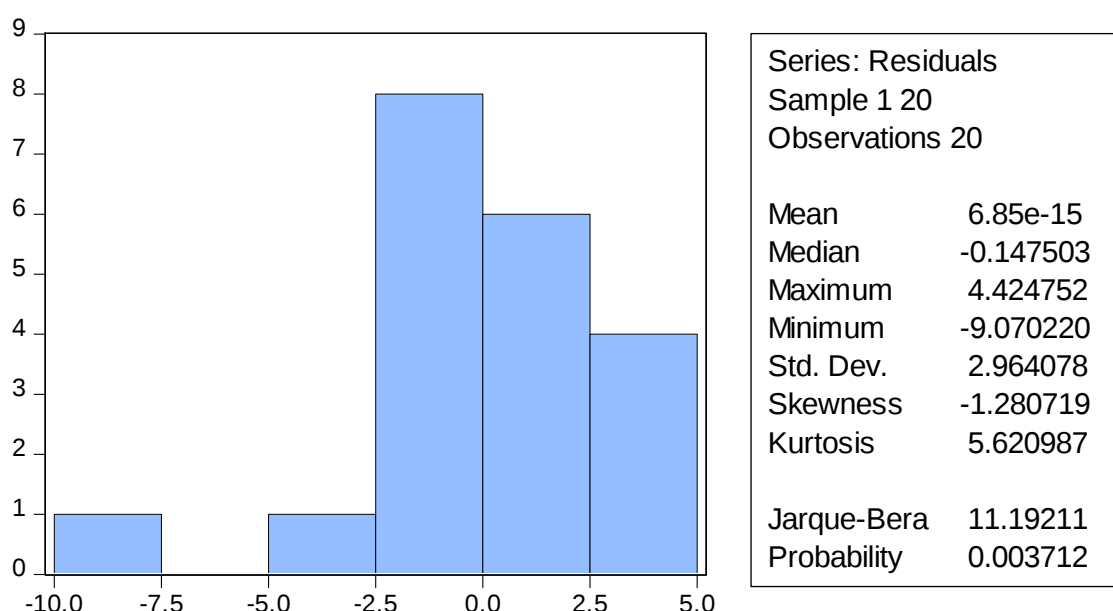
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.160742	2.870358	0.752778	0.4632
X_1	-0.325161	0.514974	-0.631412	0.5373
X_2	-0.144846	0.320623	-0.451764	0.6579
RESID ₍₋₁₎	0.311679	0.260137	1.198130	0.2495
RESID ₍₋₂₎	-0.418198	0.265601	-1.574539	0.1362
R-squared	0.169435	Mean dependent var	6.85E-15	

Adjusted R-squared	-0.052048	S.D. dependent var	2.964078
S.E. of regression	3.040238	Akaike info criterion	5.274066
Sum squared resid	138.6457	Schwarz criterion	5.522999
Log likelihood	-47.74066	Hannan-Quinn criter.	5.322661
F-statistic	0.765001	Durbin-Watson stat	2.084693
Prob(F-statistic)	0.564252		

Unitatea de învățare 13. Normalitatea distribuției erorilor

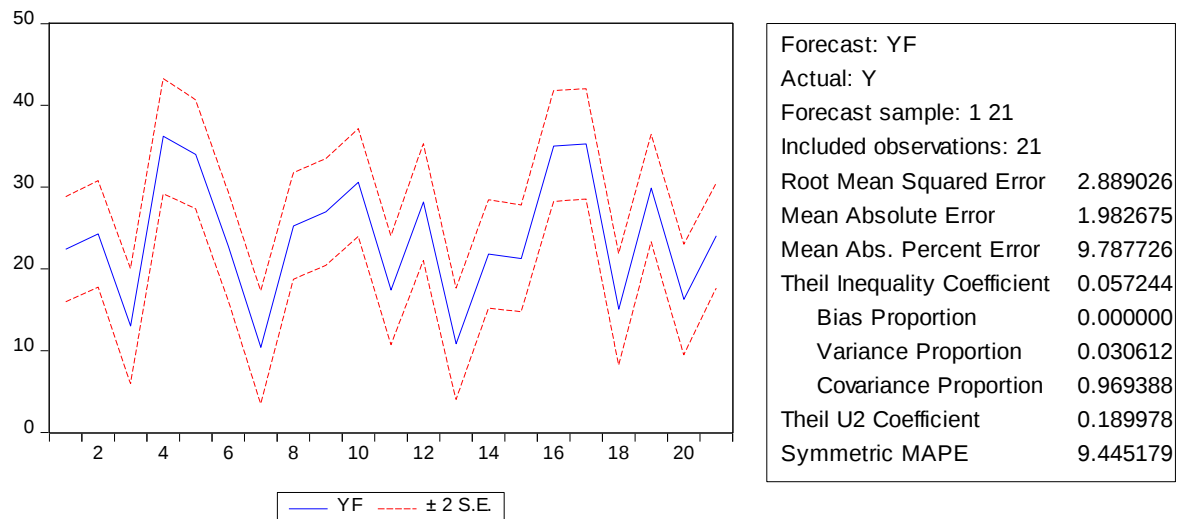
Soluțiile problemei sunt prezentate în figura următoare (generată cu programul EViews-10):



Erorile din model nu sunt normal distribuite.

Unitatea de învățare 14. Utilizarea modelelor econometrice în prognoză

Soluțiile problemei sunt prezentate în figura următoare (generată cu programul EViews-10):



Abaterea standard a erorilor de prognoză este $s_f = 3.4$, iar $\hat{Y}_{21} = 24.1$. Pentru $t_{17;0.1} = 1.74$, rezultă $t_{17;0.1} \cdot s_f = 5.9$. Aceasta înseamnă că valoarea prognozată a variabilei endogene la momentul $f = 21$ se găsește, cu o probabilitate de 90% în intervalul $[18.2; 30.0]$.