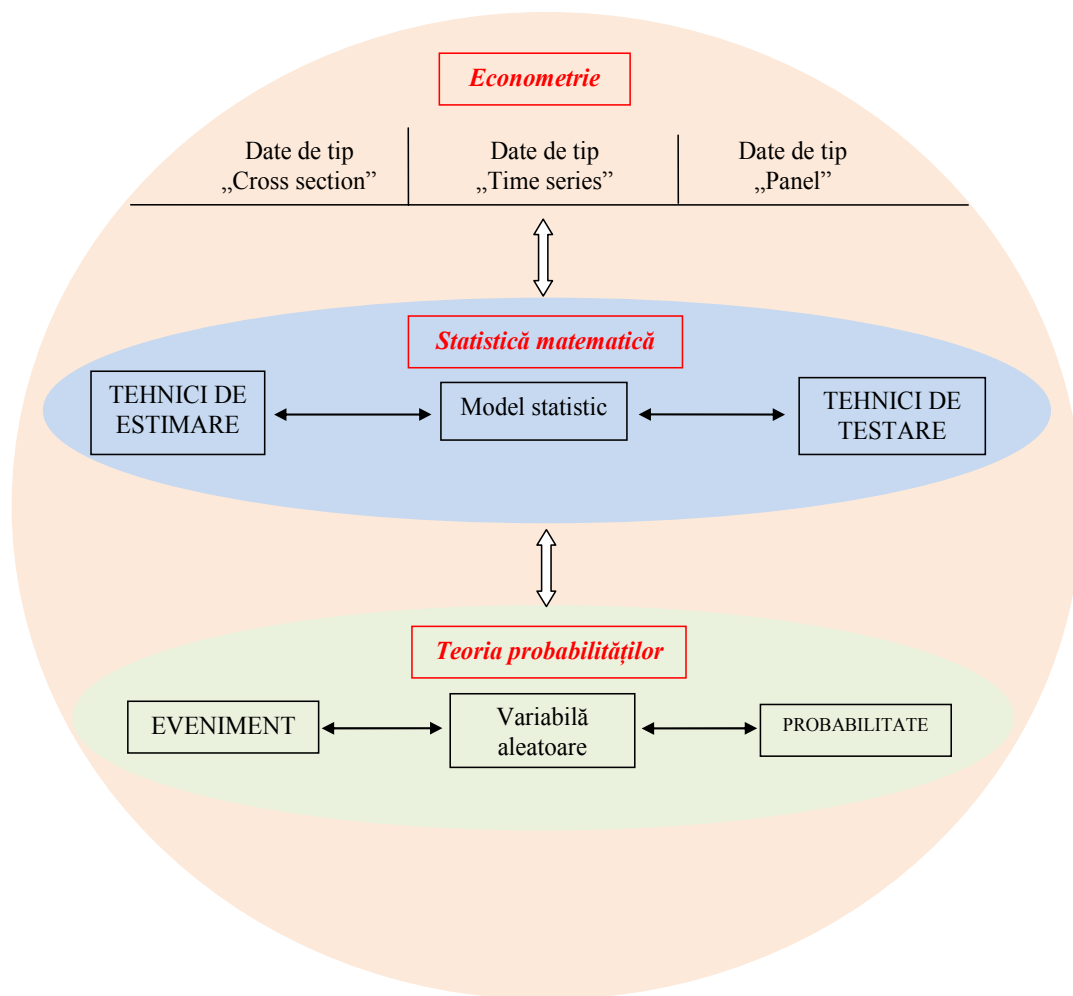


CUPRINS

Introducere	4
1 Noțiuni privind teoria probabilităților.....	13
2 Noțiuni privind statistica matematică	26
3 Modelul clasic de regresie liniară	35
4 Abateri de la ipotezele modelului clasic de regresie I	46
5 Abateri de la ipotezele modelului clasic de regresie II	51
6 Tehnici de modelare și previzionare a seriilor de timp unidimensionale	57
7 Tehnici de modelare și previzionare a seriilor de timp multidimensionale	81
8 Tehnici de analiză a dinamicii pe termen lung a variabilelor economice.....	91
9 Tehnici de analiză a modelelor de tip panel data.....	98
10 Tehnici de estimare a modelelor de tip DSGE.....	103
Bibliografie selectivă	108

Introducere

Acest suport de nivel pentru nivelul de complexitate 3 este conceput pentru a oferi cursanților o fundație riguroasă și accesibilă privind principiile teoriei probabilităților, utilizarea metodelor de inferență statistică în domeniul macroeconomic, precum și tehnicile oferite de către programul econometric **EViews** în ceea ce privește construirea, rezolvarea, estimarea, verificarea și alegerea de modele econometrice adecvate. Toate capitolele suportului cuprind exemple privind utilizarea diverselor tehnici econometrice cu ajutorul programului **EViews**.



Econometria presupune totalitatea metodelor și tehnicilor de analiză a dinamicii variabilelor economice, precum și a interconexiunilor dintre acestea. Econometria utilizează o mare parte din tehnicile de inferență statistică puse la dispoziție de către **statistica matematică**. De asemenea, **teoria probabilităților** oferă noțiunile fundamentale necesare pentru înțelegerea statisticii și a econometriei. Econometria reprezintă o îmbinare armonioasă între teoria economică, modelarea economică, statistica economică și statistica matematică. Teoria economică propune o serie de ipoteze care, în general, sunt de natură calitativă. Modelarea economică transpune aceste ipoteze în limbaj cantitativ prin intermediul modelelor economice care pot fi utilizate pentru previzionarea variabilelor de interes. Statistica economică are drept scop principal colectarea, prelucrarea și prezentarea datelor economice sub forma de grafice și tabele. Statistica matematică oferă multe instrumente de analiză a acestor datelor. Totuși, un practician are deseori nevoie de metode speciale având în vedere natura unică a datelor economice. Rolul econometriei constă în punerea la dispoziție a unor astfel de tehnici care permit testarea unui model economic și transpunerea acestuia într-un model econometric, care poate fi utilizat efectiv pentru a previziona evoluția variabilelor economice.

Modelele econometrice presupun utilizare a trei tipuri de date, respectiv:

- **date de tip „cross-section”** – acestea presupun observații în ceea ce privește o caracteristică, obținute *la un anumit moment dat, pentru mai mulți agenți economici*;
- **date de tip „time series”** – acestea presupun observații în ceea ce privește o caracteristică, obținute *la mai multe momente de timp, pentru un agent economic dat*;
- **date de tip „panel”** – acestea combină ambele dimensiuni, presupunând observații în ceea ce privește o caracteristică, obținute *la mai multe momente de timp, pentru mai mulți agenți economici*.

Metodologia utilizată, în general, de către econometrie pentru analiza unui fenomen economic se poate încadra de-a lungul următoarelor linii:

- identificarea *teoriei economice* care explică fenomenul respectiv;
- specificarea *modelului teoretic* în format matematic;
- specificarea *modelului econometric*;
- obținerea *datelor* corespunzătoare;
- *estimarea parametrilor* modelului econometric;
- *testarea statistică a ipotezelor* propuse de teoria economică;
- *previzionarea variabilelor* din cadrul modelului econometric;
- utilizarea modelului econometric pentru *fundamentarea deciziilor de politică economică*.

Un exemplu clasic în ceea ce privește utilizarea metodologiei econometrice îl reprezintă analiza consumului privat. Binecunoscuta *teorie economică* propusă de către Keynes, în ceea ce privește legătura dintre consum și venitul disponibil, presupune că, în medie, agenții economici își majorează consumul pe măsură ce venitul lor disponibil crește, însă cu o „viteză” mai mică decât a acestuia. În limbaj cantitativ, această teorie se poate transpune sub forma unei relații funcționale între consum și venit disponibil, cu condiția suplimentară că derivata acestei funcții în raport cu venitul disponibil are valori cuprinse între 0 și 1.

Deși Keynes a postulat o relație pozitivă între consum și venit, acesta nu a specificat forma exactă a acestei relații funcționale dintre cele două variabile. Se poate construi astfel un *model teoretic* în cadrul căruia se presupune că această relație este liniară, respectiv $C = \alpha + \beta Y$, unde C reprezintă consumul, Y venitul disponibil, iar β o măsură a înclinației marginale spre consum. Acest model teoretic al funcției de consum prezintă, însă, un interes limitat pentru un practician, pentru că modelul presupune că există o relație exactă sau deterministă între consum și venit.

În realitate, legătura dintre cele două variabilele economice este, în general, inexactă deoarece consumul este influențat și de alți factori. Pentru a permite existența unei relații inexacte între cele două variabile economice, un *model econometric* presupune că funcția de consum se poate reprezenta sub forma $C = \alpha + \beta Y + \varepsilon$, unde ε este termenul de eroare, reprezentat printr-o variabilă aleatoare, concept fundamental din teoria probabilităților, cu caracteristici statistice bine-definite. Mai exact, modelul econometric prezentat mai sus este un caz particular de model de regresie liniară, tip de model care va fi analizat în cadrul cursului de față.

Determinarea parametrilor modelului econometric, pe baza *datelor* avute la dispoziție, presupune utilizarea unor metode și tehnici de *estimare statistică*. Însă, trebuie avut grijă să se aleagă acel instrument de estimare care se potrivește datelor pentru care se realizează analiza. Astfel, dacă dispunem de observații privind consumul la un moment dat al mai multor familii, adică avem la dispoziție date de tip „cross-section”, este potrivită estimarea prin metoda OLS. Dacă, însă, analiza se realizează la nivel agregat și dispunem de dinamica consumului și a venitului pentru o anumită perioadă, adică avem la dispoziție date de tip „time series”, este mai potrivit să se utilizeze tehnici de estimare bazate pe identificarea relațiilor de cointegrare dintre cele două serii. Alegerea corectă a metodei de estimare, lucru asupra căruia se va insista pe parcursul acestui curs, este esențială pentru obținerea unor estimatori „buni”, adică a celor estimatori care surprind în mod corect realitatea în ceea ce privește legătura dintre cele două variabile economice.

În continuare, trebuie verificat dacă modelul econometric estimat pentru seriile de date analizate este în concordanță cu teoria economică de la care s-a plecat. Această verificare se poate realiza prin metode riguroase privind *testarea ipotezelor statistice*, tehnici care vor fi prezentate în cadrul acestui curs. Astfel, în cazul exemplului analizat, se testează ipoteza „nulă” conform căreia $0 < \beta < 1$, sau, altfel spus, se testează dacă înclinația marginală spre consum este pozitivă și subunitară.

În cazul în care modelul estimat este în concordanță cu teoria economică, se poate utiliza acest model pentru a *previziona* o valoare viitoare a variabilei dependente, în acest caz consumul, în funcție de valoarea viitoare anticipată a variabilei independente, în acest caz venitul disponibil. De asemenea, modelul estimat poate fi utilizat pentru *fundamentarea deciziilor de politică economică*. Astfel, în exemplul analizat, pornind de la înclinația marginală spre consum estimată pe baza datelor, se poate calcula „multiplicatorul” lui Keynes, care poate fi utilizat pentru a realiza diverse scenarii și simulări legate de impactul modificării sistemului de taxare asupra sectorului privat.

În continuare, este prezentat pe scurt conținutul acestui suport. ***Primele două capitole*** prezintă partea nevăzută a econometriei, fundamentele acesteia. Fără a avea o serie de noțiuni elementare privind teoria probabilităților și statistica matematică, practicienii percep econometria ca un amalgam de formule și tehnici fără nici un sens. Detaliile privind utilizarea tehnicilor implementate în programe econometrice se estompează rapid pentru cei care nu înțeleg motivele utilizării procedurilor pe care încearcă să le aplice. Multe instituții recunosc, în prezent, nevoia unui studiu mai riguros al teoriei probabilităților, al principiilor statisticii matematice, precum și al metodelor econometrice avansate, în scopul de a-și pregăti specialiștii pentru o înțelegere pe termen lung a tehnicilor statistice utilizate în domeniul economic.

În ***capitolul 1*** sunt schițate o serie de noțiuni privind teoria probabilităților, cum ar fi eveniment, probabilitate, probabilitate condiționată, variabilă aleatoare. Pentru a evidenția importanța conceptului de variabilă aleatoare este prezentat modul în care se poate analiza în **EViews** funcția de densitate de repartiție a unei variabile aleatoare, respectiv a unui vector de variabile aleatoare. Înțelegerea conceptului de distribuție a unei variabile aleatoare este esențială pentru înțelegerea conceptelor legate de inferența statistică.

În ***capitolul 2*** sunt prezentate pe scurt concepte de bază în ceea ce privește statistica matematică. Astfel, pentru început este analizat conceptul de distribuție

asimptotică. În continuare, sunt evidențiate tehnici privind procedura de estimare a unui model statistic. Sunt prezentate o serie de definiții necesare pentru o înțelegere de durată a econometriei, cum ar fi conceptul de estimator „bun”. În a treia parte a acestui capitol sunt discutate tehnici privind procedura de testare a ipotezelor statistice. Pentru a evidenția importanța tehnicilor de inferență statistică, este prezentat modul în care se poate realiza în **EViews** un test statistic privind media unei populații statistice. Acest exemplu, oferă cursanților instrumentele necesare înțelegerii unor teste mai sofisticate legate de modelele econometrice discutate în capitolele următoare.

În *următoarele trei capitole* atenția se îndreaptă asupra tehnicilor econometrice utilizate, în special, pentru analiza interconexiunilor dintre **date de tip „cross-section”**. Totuși, aceste tehnici stau la baza unor metode avansate utilizate pentru investigarea dinamicii celorlalte tipuri de date economice. În *capitolul 3* este prezentat modelul clasic de regresie liniară, ipotezele acestuia, modul de estimare al unei ecuații de regresie în **EViews**, precum și modalitatea de a interpreta estimatorii și testele statistice rezultate în urma acestei estimări. Modelul clasic de regresie are la bază o serie de ipoteze, întâmplându-se foarte rar ca în practică toate acestea să fie îndeplinite. În *capitolul 4* sunt prezentate tehnicile puse la dispoziție de **EViews** pentru a rezolva problemele induse de nerespectarea acestor ipoteze în ceea ce privește inovațiile ecuației de regresie. Astfel, sunt analizate metodele care pot contabiliza existența heteroskedasticității și a autocorelării erorilor, în special tehnicile robuste de determinare a erorilor standard pentru estimatori. În *capitolul 5* este adusă în discuție „problema de endogenitate”, care este destul de frecventă în practică, și care invalidează multe din rezultatele modelului clasic de regresie. Este prezentat modul în care se poate rezolva această problemă în cadrul programului econometric **EViews** prin intermediul variabilelor instrumentale.

Fenomenele economice se pot analiza din perspectivă statică sau dinamică. Analiza statică urmărește descrierea relațiilor care există între variabile la un

anumit moment. Analiza în dinamică a fenomenelor economice încearcă să surprindă modul în care variabilele evoluează în timp, modul în care se schimbă relațiile care există între variabile, precum și eventualele relații de cauzalitate care funcționează între acestea. În *următoarele trei capitole* atenția este concentrată asupra tehnicilor econometrice utilizate, în special, pentru analiza dinamicii și interconexiunilor dintre **date de tip „time series”**. O serie de timp reprezintă o succesiune de înregistrări ale unei variabile economice (de exemplu produsul intern brut – PIB, rata inflației, rata de dobândă, cursul de schimb, indicele BET etc.) corespunzătoare unei frecvențe: zilnică, lunară, trimestrială sau anuală. Analiza seriilor de timp are ca obiective: stabilirea proprietăților statistice, cum ar fi medie, varianță, autocorelație, autocorelație parțială; identificarea relațiilor care există pe termen lung între variabile; modelarea unei serii de date reale, prin identificarea modelului prin care se poate reprezenta cel mai bine seria respectivă; prognoza pe termen scurt a unei serii de timp pe baza modelului selectat pentru aceasta.

În *capitolul 6* sunt prezentate tehnicile de analiză puse la dispoziție de către **EViews** pentru analiza dinamicii seriilor de timp unidimensionale. Sunt evidențiate modalitățile de modelare a serilor de timp prin intermediul proceselor de tip ARMA (AutoRegressive Moving Average), precum și modul în care natura autocorelației dintre valorile seriei respective oferă informații în ceea ce privește selectarea unei specificații pentru modelul ARMA. De asemenea, este discutat conceptul de staționaritate, concept fundamental pentru modelarea adecvată a dinamicii unei serii de timp, și sunt prezentate instrumentele puse la dispoziție de către **EViews** pentru testarea staționarității unei serii,

În *capitolul 7* sunt prezentate tehnicile de analiză puse la dispoziție de către **EViews** pentru analiza dinamicii și a interacțiunii dintre serii de timp multidimensionale. Astfel, în cadrul capitolului sunt descrise estimarea modelelor de tip VAR (Vector AutoRegressive), precum și metodele de analiză ale acestor modele, cum ar fi funcția de răspuns la impuls și descompunerea varianței. În *capitolul 8* sunt prezentate tehnicile de analiză puse la dispoziție de către **EViews**

pentru analiza dinamicii pe termen lung dintre variabilele economice. Acest capitol descrie modelele și instrumentele pentru testarea prezenței relațiilor de cointegrare dintre mai multe variabile nestaționare, precum și procesul de estimare și analiză a modelelor de tip VEC (Vector Error Correction).

În aceste note de curs se prezintă elementele esențiale, care formează baza analizei seriilor de timp. Prezentarea este gândită pentru un nivel de complexitate mediu, punând accent pe introducerea intuitivă a conceptelor și concretizarea acestora prin exemple elocvente. De asemenea, se argumentează la un nivel corespunzător unele din cele mai importante rezultate.

În **capitolul 9** atenția este concentrată asupra tehnicilor econometrice utilizate pentru analiza dinamicii și interconexiunilor dintre **date de tip „panel”**, care reprezintă o îmbinare dintre metodele descrise în capitolele anterioare. Sunt descrise etapele necesare în cadrul **EViews** pentru specificarea și pentru estimarea unui model de tip panel, precum și metodele puse la dispoziție de către programul econometric pentru a discrimina între o specificație cu efecte fixe și o specificație cu efecte aleatoare.

În **capitolul 10** este prezentată structura de bază a unui script pentru **Dynare**, program care poate fi utilizat pentru estimarea și analiza dinamicii modelelor de tip DSGE. De asemenea, este fundamentată, pe scurt, procedura de log-liniarizare a unui model DSGE, operațiune importantă în cadrul metodologiei de estimare. Într-un rând se amintește de tendința recentă din literatura de specialitate de a utiliza pentru estimarea modelelor DSGE tehnici de bayesiene de inferență statistică, în încheierea capitolului sunt prezentate noțiuni fundamentale privind econometria bayesiană.

În final, trebuie subliniat ce reprezintă și ce nu reprezintă acest suport. Acest suport nu este un manual de econometrie. Acest suport nu poate suplini, în nici un fel, un manual de econometrie de nivel introductiv, cum ar fi Brooks (2008) sau Wooldridge (1999), sau un manual de nivel intermediar, cum ar fi Greene (2008) sau Enders (2004), sau un manual de nivel avansat, cum ar fi Hamilton

(1994). De asemenea, acest suport nu este un manual de utilizare al programului econometric **EViews**. Acest suport reprezintă o schiță, la nivel intermediar, a tehnicilor econometrice care apar în cadrul analizei fenomenelor economice, cu accent pe utilizarea acestora cu ajutorul programului econometric **EViews**. De asemenea, acesta reprezintă o modalitate de a prezenta ceea ce este esențial, mai precis tot ceea ce nu trebuie uitat după ce se vor fi fost uitate toate celelalte detalii. Este demn de menționat faptul că utilizarea acestui suport este complementară cu prezența activă în timpul orelor de curs.

Parcurgerea suportului, precum și audierea cursului nu presupun cunoștințe anterioare privind teoria probabilităților sau statistica matematică și construiește în mod eficient subiectul "de la zero". Cursanții vor dobândi astfel pregătirea necesară pentru o înțelegere matură și durabilă a metodelor statistice și econometrice de inferență și vor fi pregătiți pentru a citi și înțelege texte de econometrie de nivel intermediar și avansat.

1 Noțiuni privind teoria probabilităților

Un agent economic rațional, cel mai probabil, va prefera să diminueze incertitudinea privind rezultatul unei situații în orice context de luare a unei decizii în care profitul, utilitatea, precum și bunăstarea sunt afectate. Teoria probabilităților pune la dispoziția unui agent de decizie o serie de instrumente utilizate pentru a distinge probabilul de improbabil, în cazul unor decizii, și oferă managerilor, economiștilor, organismelor de reglementare și consumatorilor informații care pot fi folosite pentru a clasifica rezultatele potențiale ale deciziilor lor în ceea ce privește probabilitatea de apariție. Astfel, este posibilă luarea unor decizii care să maximizeze probabilitatea de apariție a unui rezultat dorit, sau care să reducă probabilitatea de apariție a unor rezultate dezastruoase.

1.1 Evenimente și probabilitatea de apariție a acestora

Noțiunea de experiment este foarte des utilizată în domeniul statisticii pentru a face referire la orice activitate pentru care rezultatul sau starea finală a unei activități nu pot fi specificate în avans, dar pentru care poate fi identificată o mulțime conținând toate rezultatele posibile ale acelei activități. Înainte de a analiza probabilitatea de apariție a unui anumit rezultat al unui experiment, este necesar să se identifice ce rezultate sunt posibile. Aceasta conduce la definirea spațiului stărilor unui experiment.

Spațiul stărilor reprezintă o mulțime, notată în acest suport cu Ω , care conține toate rezultatele posibile ale unui experiment. Altfel, la aruncarea unui zar, rezultatele posibile se referă la apariția uneia din cele șase fețe ale acestuia. Ca urmare, în cazul experimentului care constă în aruncarea unui zar, $\Omega = \{1,2,3,4,5,6\}$.

Entitățile fundamentale cărora li se atribuie probabilități în cadrul teoriei probabilităților sunt submulțimi ale spațiului stărilor. Un **eveniment** reprezintă o submulțime a spațiului stărilor, sau, altfel spus, reprezintă o colecție de elemente

posibile ale unui experiment. De exemplu, în cazul experimentului care constă în aruncarea unui zar, submulțimea $A = \{1\}$ reprezintă evenimentul că a apărut fața 1 a zarului, iar submulțimea $B = \{1,2\}$ reprezintă evenimentul că a apărut fața 1 sau fața 2 a zarului. Trebuie menționat faptul că nu orice submulțime a lui Ω este un eveniment în sensul descris de către teoria probabilităților. Totuși, în cadrul acestui suport, de nivel introductiv, nu se va analiza în amănunt această distincție. Mulțimea tuturor evenimentelor care pot fi observate în urma derulării unui experiment poartă numele de **spațiul evenimentelor** și este notat cu \mathcal{F} .

Dacă A este un eveniment, se notează cu $\bar{A} := \Omega - A$ *evenimentul contrar* al evenimentului A , acesta conținând toate elementele din spațiul stărilor care nu se regăsesc în submulțimea A . Există două evenimente deosebite, respectiv *evenimentul imposibil* reprezentat de mulțimea vidă (\emptyset) și *evenimentul sigur*, reprezentat de mulțimea care conține toate stările posibile, adică de către spațiul stărilor Ω . De exemplu, în cazul experimentului care constă în aruncarea unui zar, se consideră că este imposibil ca zarul să rămână pe muchie și, astfel, să nu apară nici o față. Ca urmare, este sigur că va apărea cel puțin o față. Evenimentul imposibil este evenimentul contrar al evenimentului sigur.

Unul din obiectivele teoriei probabilităților este acela de a elabora o măsură cu care să se poată cuantifica posibilitatea de apariție a diverselor evenimente înglobate în spațiul stărilor experimentului analizat. O **funcție de probabilitate** este o funcție $P: \mathcal{F} \rightarrow [0,1]$ care îndeplinește următoarele proprietăți:

- $P(\Omega) = 1$;
- Pentru orice n evenimente, A_1, A_2, \dots, A_n disjuncte două câte două avem că $P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$.

Un **câmp de probabilitate**, cunoscut și sub numele de spațiu de probabilitate, este un triplet (Ω, \mathcal{F}, P) care este format din mulțimea tuturor stărilor posibile, mulțimea tuturor evenimentelor observabile și o funcție de probabilitate cu care se cuantifică posibilitatea de apariție a acestor evenimente.

Înainte de realizarea unui experiment, evenimentul care este absolut sigur să apară este evenimentul Ω , deoarece rezultatul unui experiment trebuie să fie un element al spațiului stărilor. În continuare este studiat efectul pe care informațiile suplimentare privind apariția unui eveniment îl au asupra probabilității altor evenimente asociate experimentului analizat. În special, în cazul în care este cunoscut faptul că rezultatul experimentului este un element dintr-o submulțime B , a spațiului stărilor, se pune problema identificării cantitative a efectului acestei informații suplimentare asupra probabilităților celorlalte evenimente. Cum ar trebui să fie definită probabilitatea unui eveniment A , având în vedere informațiile suplimentare care arată că a avut loc evenimentul B ?

Astfel, este convenabil să se introducă noțiunea de **probabilitate condiționată**. Vom nota cu $P(A|B)$ probabilitatea evenimentului A condiționată de evenimentul B . Avem $P(\Omega|B)=1$, dar, în plus, și $P(B|B)=1$, fiind evident faptul că evenimentul B devine “sigur” condiționat de informația suplimentară că acest eveniment a avut loc. Astfel, în mod intuitiv, rezultă faptul că

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Probabilitatea condiționată poate fi utilizată și pentru definirea, într-o manieră intuitivă, a conceptului de independență a două evenimente. Astfel, dacă evenimentul A este independent de evenimentul B , atunci este evident că informația suplimentară generată de apariția evenimentului B nu poate aduce nici o îmbunătățire în ceea ce privește cuantificarea posibilității de apariție a evenimentului A . Astfel, în acest caz rezultă că $P(A|B) = P(A)$. Mai exact, două evenimente A și B se numesc **independente** dacă $P(A \cap B) = P(A)P(B)$. Trebuie evidențiat faptul că, dacă evenimentul A sau evenimentul B au probabilitate de apariție zero, atunci ele sunt independente. De asemenea, dacă două evenimente sunt independente atunci și evenimentele contrare vor fi independente între ele.

1.2 Tipuri de variabile aleatoare

1.2.1 Variabile aleatoare unidimensionale

Rezultatele mai multor experimente sunt exprimate în mod inerent sub formă de numere reale. De exemplu, măsurarea înălțimii sau a greutateii unei persoane, sau observarea prețului și a cantității de echilibru de pe o piață. Spațiul stărilor asociat cu aceste tipuri de experimente sunt submulțimi ale mulțimii numerelor reale.

Există, însă, și experimente ale căror rezultate nu sunt numere reale și al căror spațiu al stărilor nu este în mod inerent o submulțime a unui spațiu real. De exemplu, observarea rezultatului aruncării unei monede, care este ban sau stemă, observarea rezultatului aruncării unui zar, care este una din cele șase fețe ale zarului, sau observarea dacă un element selectat dintr-un ansamblu este defect sau nu. Este atât convenabil cât și util convertirea acestor spații abstracte într-un subspațiu format din numere reale, conversie realizată prin asocierea unui număr real pentru fiecare rezultat din spațiul stărilor original. O astfel de procedură ar putea fi privită ca o „codificare” a rezultatelor unui experiment prin diverse numere reale. În plus, este posibil ca rezultatele unui experiment să nu poată fi de interes direct într-un cadru dat; în schimb, s-ar putea să prezinte interes o anumită combinație a acestora exprimată printr-un număr real. Conceptul de variabilă aleatoare poate fi utilizat pentru a caracteriza rezultatele unui experiment ca o submulțime de numere reale.

Fie (Ω, \mathcal{F}, P) un câmp de probabilitate. O **variabilă aleatoare unidimensională** în raport cu (Ω, \mathcal{F}, P) este o funcție definită pe spațiul stărilor (Ω) și are valori numere reale, $X : \Omega \rightarrow R$. Astfel, prin utilizarea conceptului de variabilă aleatoare, toate experimentele cu rezultate univariate pot fi interpretate ca având spațiul stărilor compus din elemente reale. Mai exact, spațiul stărilor indus, $R[X]$, reprezintă o submulțime a unui spațiu real. O variabilă aleatoare se numește **variabilă aleatoare discretă** dacă spațiul stărilor indus, $R[X]$, este o mulțime finită

sau numărabilă. Dacă spațiul stărilor indus este un interval, atunci vorbim despre o **variabilă aleatoare continuă**.

Un avantaj major al utilizării doar a spațiilor de stări cu valori reale constă în faptul că toate instrumentele matematice elaborate pentru sistemul numerelor reale sunt disponibile atunci când se analizează proprietățile unui astfel de spațiu. În practică, o dată ce câmpul de probabilitate indus a fost identificat, câmpul de probabilități inițial (Ω, \mathcal{F}, P) poate fi, în general, ignorat în scopul definirii evenimentelor variabilelor aleatoare și a probabilității de apariție a acestora. De fapt, cel mai adesea se alege câmpul indus de probabilitate direct de la începutul unui experiment, acordând mai puțină atenție câmpului de probabilitate original.

O simplificare importantă, în ceea ce privește analiza variabilelor aleatoare, este realizată prin introducerea conceptului de funcție de repartiție. Astfel, **funcția de repartiție a unei variabile aleatoare** X , cunoscută în literatura de specialitate sub prescurtarea CDF (*eng. Cumulative Distribution Function*), este o funcție definită pe mulțimea numerelor reale cu valori în mulțimea numerelor reale, $F_X : \mathbb{R} \rightarrow \mathbb{R}$ astfel încât pentru orice număr real x avem că $F_X(x) = P(X \leq x)$. Altfel spus, $F_X(x)$ reprezintă un număr care cuantifică probabilitatea evenimentului ca variabila aleatoare X să ia valori mai mici sau egale cu numărul real x .

Dacă două variabile aleatoare X și Y au proprietatea că $F_X = F_Y$ se spune că cele două variabile aleatoare sunt **identic distribuite**.

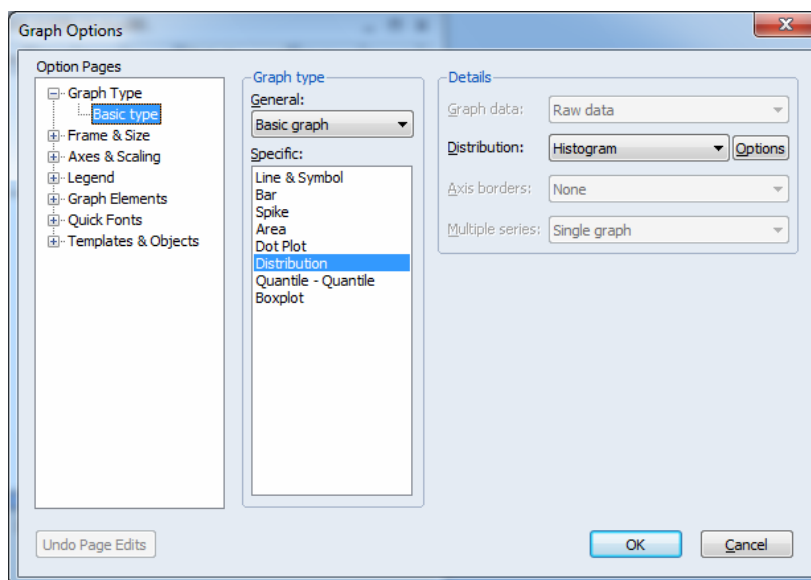
Trebuie menționat faptul că, dacă X este variabilă aleatoare discretă, atunci F_X nu este funcție continuă, graficul funcției prezentând „salturi”. Pe de altă parte însă, dacă X este o variabilă aleatoare continuă, atunci se poate arăta că F_X este o funcție continuă.

În cazul unei variabile aleatoare continue, analiza acesteia se poate simplifica și mai mult prin apelarea la conceptul de funcție de densitate de repartiție. Astfel, **funcția de densitate de repartiție a unei variabile aleatoare**

continue X , cunoscută în literatura de specialitate sub prescurtarea PDF (*eng. Probability Distribution Function*), este o funcție definită pe mulțimea numerelor reale cu valori în mulțimea numerelor reale, $f_X : \mathbb{R} \rightarrow \mathbb{R}$ astfel încât pentru orice

$$\text{număr real } a \text{ avem că } F_X(a) = \int_{-\infty}^a f_X(x) dx.$$

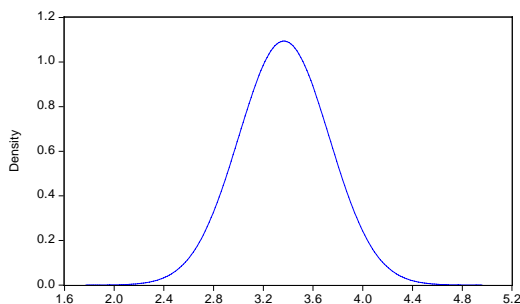
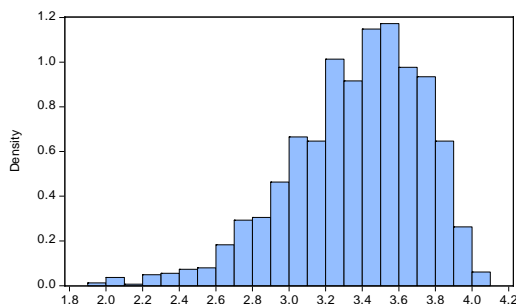
Views oferă posibilitatea de a afișa o varietate de așa-numite grafice analitice (*i.e.* grafice rezultate prin prelucrarea statistică a datelor brute) Caracteristica centrală a acestor grafice constă în faptul că afișează un rezumat al datelor originale. O clasă importantă de astfel de grafice sunt cele care cuantifică vizual funcția de densitate de repartiție a unei variabile aleatoare, calculată utilizând o serie de date privind realizările respectivei variabile aleatoare. Pentru a afișa un astfel de grafic se utilizează opțiunea **Graph**.



În cadrul căsuței de dialog care este afișată se selectează tipul graficului **Distribution**, iar în zona **Details** se precizează tipul de grafic de distribuție dorit, cum ar fi histogramă (**Histogram**), grafic de distribuție determinat prin funcții de

tip „kernel” (**Kernel Density**) sau grafic de distribuție determinat prin utilizare unui model parametric (**Theoretical Distribution**). Astfel, în figura de mai jos sunt prezentate două grafice de tip distribuție determinate pe baza acelorași date, respectiv o histogramă și o funcție de densitate de repartiție asociată unei variabile distribuită normal. O variabilă aleatoare X are o **distribuție normală** cu medie μ și varianță σ^2 , notată cu $X \sim \Phi(\mu, \sigma^2)$, dacă funcția de densitate de repartiție este

dată de formula:
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$



Un grafic de densitate de repartiție oferă informații în ceea ce privește probabilitatea de apariție a anumitor evenimente. Astfel, se observă că este „mai probabil” ca variabila aleatoare considerată să ia valori în jurul valorii de 3,5 decât să ia valori în jurul valorii de 1,5 sau în jurul valorii de 4,5.

1.2.2 Variabile aleatoare multidimensionale

În secțiunea precedentă a acestui capitol, a fost analizat conceptul de variabilă aleatoare unidimensională, pe spațiul stărilor fiind definită numai o funcție reală. Conceptul unei variabilă aleatoare multidimensională este o extensie a conceptului variabilei aleatoare unidimensională, pe spațiul stărilor experimentului analizat fiind definite două sau mai multe funcții reale. Conceptul

de variabilă aleatoare multidimensională se aplică în mod firesc și, în general, oricărui experiment în care, pentru fiecare rezultat al experimentului, sunt observate două sau mai multe caracteristici.

Fie (Ω, \mathcal{F}, P) un câmp de probabilitate. O **variabilă aleatoare multidimensională** în raport cu (Ω, \mathcal{F}, P) , cunoscut și sub denumirea de vector aleatoriu, este definită ca o funcție pe spațiul stărilor (Ω) și cu valori vectori având componente numere reale, $X : \Omega \rightarrow R^n$.

În cadrul acestui suport ne vom limita atenția asupra cazului bidimensional, $X = (X_1, X_2)$. Ca și în cazul unidimensional, se poate construi un câmp de probabilitate indus de vectorul aleator, unde spațiul stărilor, $R[X]$, este o submulțime formată din perechi de numere reale. Ne vom concentra atenția asupra variabilelor aleatoare multidimensionale continue caracterizate prin faptul că $R[X]$ este formată din produse carteziene dintre două intervale (*i.e.* dreptunghiuri). În mod analog cu cazul unidimensional, se poate defini **funcția de repartiție a unui vector** aleator, $F_X : R^2 \rightarrow R$ astfel încât avem că $F_X(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2)$ pentru orice vector real (x_1, x_2) . Altfel spus, $F_X(x_1, x_2)$ reprezintă un număr care cuantifică probabilitatea evenimentului ca, în același timp, variabila aleatoare X_1 să ia valori mai mici sau egale cu numărul real x_1 și variabila aleatoare X_2 să ia valori mai mici sau egale cu numărul real x_2 .

De asemenea, se poate defini **funcția de densitate de repartiție a unui vector** aleator X , $f_X : R^2 \rightarrow R$ astfel încât, pentru orice vector real (a_1, a_2) avem că

$$F_X(a_1, a_2) = \int_{-\infty}^{a_1} \int_{-\infty}^{a_2} f_X(x_1, x_2) dx_1 dx_2 .$$

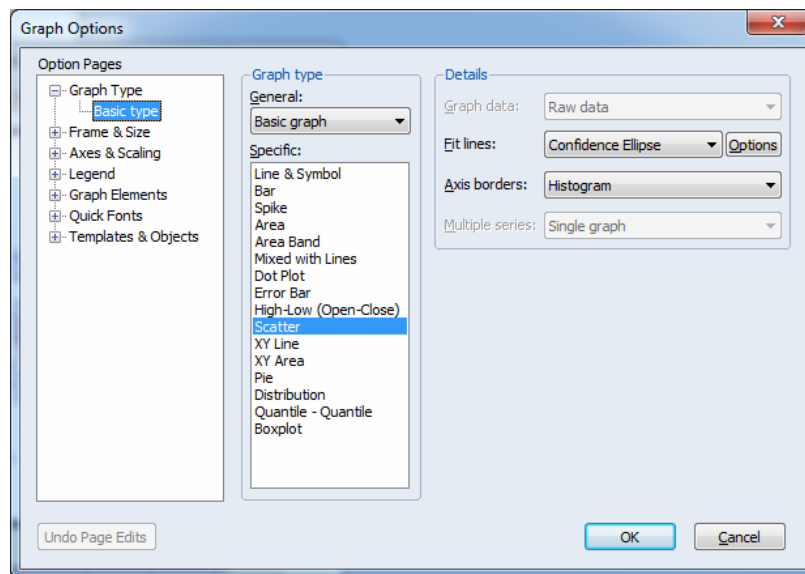
În cazul analizei vectorilor bidimensionali este de interes să se extragă informații cu privire la una din componentele acestui vector. Acest lucru se poate realiza prin intermediul repartiției marginale a unui vector aleator. Astfel, se poate defini **funcția de repartiție marginală** a unei componente a unui vector aleator (de exemplu componenta 1) ca fiind o funcție cu caracteristici asemănătoare

funcției de repartiție a unei variabile aleatoare, $F_{X_1} : R \rightarrow R$, astfel încât pentru orice număr real x_1 avem că $F_{X_1}(x_1) = \lim_{x_2 \rightarrow \infty} F_X(x_1, x_2)$. De asemenea, există o

funcție de densitate de repartiție marginală a unei componente a unui vector

aleator, $f_{X_1} : R \rightarrow R$, unde $f_{X_1}(a_1) = F'_{X_1}(a_1) = \int_{-\infty}^{a_1} \int_{-\infty}^{\infty} f_X(x_1, x_2) dx_1 dx_2$.

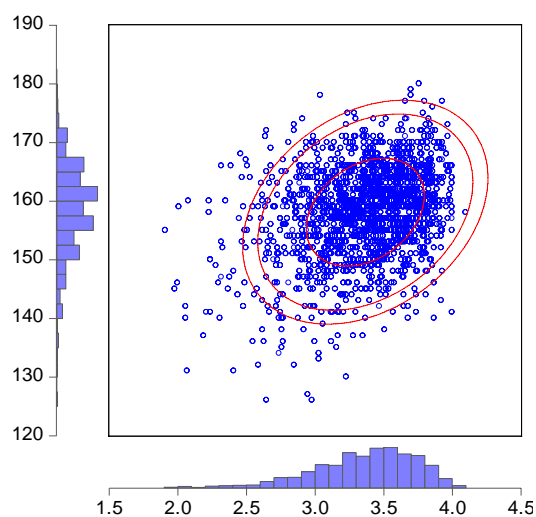
Eviews permite vizualizarea funcției de densitate bidimensională, precum și a funcțiilor de densitate marginală pentru un vector format din două componente. Programul econometric cuantifică aceste funcții utilizând seriile de date privind realizările respectivelor variabile aleatoare. Pentru realizarea unui astfel de grafic se selectează cele două serii și se utilizează opțiunea **Open as Group**, după care se utilizează opțiunea **Graph** cu setările din figura de mai jos.



Mai exact, în cadrul căsuței de dialog care este afișată se selectează tipul graficului **Scatter**, iar în zona **Details** se precizează că se dorește afișarea

distribuțiilor marginale prin selectarea în căsuța **Axis borders** a opțiunii **Histogram**. Astfel, în figura de mai jos sunt prezentate, pe același grafic

- *densitatea de repartiție bidimensională* prin intermediul unui nor de puncte și prin elipse care demarchează anumite zone de probabilitate;
- *densitățile de repartiție marginală*, pentru cele două variabile aleatoare din componență vectorului aleatoriu, prin intermediul a două histograme.



Acest grafic de densitate de repartiție oferă informații în ceea ce privește probabilitatea de apariție, în același timp, al unor valori pentru cele două variabile aleatoare. Astfel, este „puțin probabil” ca variabila 1 (reprezentată pe orizontală) să ia valori în jur de 2,5 și, în același timp, variabila 2 (reprezentată pe verticală) să ia valori în jur de 140. În schimb, este „mai probabil” ca variabila 1 să ia valori în jur de 3,5 și, în același timp, variabila 2 să ia valori în jur de 155.

Un alt aspect important, în ceea ce privește vectorii aleatori, se referă la conceptul de repartiție condiționată a unei componente a vectorului în raport cu o altă componentă a acestuia.

Să presupunem că se cunoaște câmpul de probabilitate corespunzător unui experiment care implică rezultatele unei variabile aleatoare bidimensională $X = (X_1, X_2)$ și se urmărește analiza probabilităților evenimentului $X_1 \in A$ dat fiind faptul că $X_2 \in B$. Acest aspect poate fi analizat prin intermediul **funcției de densitate de repartiție condiționată** a componentei 1 a vectorului în raport cu componenta 2 a acestuia. Aceasta este o funcție cu valori reale, $f_{X_1|X_2} : R \rightarrow R$ și se poate determina utilizând funcția de densitate a întregului vector și funcția de densitate marginală a componentei 2, respectiv $f_{X_1|X_2}(x_1 | x_2) = \frac{f_X(x_1, x_2)}{f_{X_2}(x_2)}$. De asemenea, se poate defini și **funcția de repartiție condiționată** a componentei 1 a vectorului în raport cu componenta 2, ca fiind o funcție $F_{X_1|X_2} : R \rightarrow R$, unde

$$F_{X_1|X_2}(x_1 | x_2) = \int_{-\infty}^{x_1} f_{X_1|X_2}(u | x_2) du .$$

Noțiunea de repartiție condiționată dintre două variabile aleatoare, sau echivalent între două componente ale unui vector aleator, poate fi utilizată și pentru definirea, într-o manieră intuitivă, a conceptului de independență a două variabile aleatoare. Astfel, dacă variabila aleatoare X_1 este independentă de variabila aleatoare X_2 , atunci este evident că informația suplimentară generată de apariția unei realizări a variabilei X_2 , nu poate aduce nici o îmbunătățire în ceea ce privește cuantificarea posibilității de apariție a unei realizări a variabilei aleatoare X_1 . Astfel, în acest caz, rezultă că $f_{X_1|X_2}(x_1 | x_2) = f_{X_1}(x_1)$. Mai exact, două variabile aleatoare X_1 și X_2 se numesc **independente** dacă $f_X(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$.

Variabile aleatoare X_1, X_2, \dots, X_n se numesc **i.i.d. (independente și identic distribuite)** dacă sunt independente și $F_{X_1} = F_{X_2} = \dots = F_{X_n}$.

1.3 Statistici descriptive

În **Eviews**, opțiunea **Histogram and Stats** afișează distribuția de frecvență a seriei analizate sub forma unei histogramme. Histograma împarte diametrul seriei (*i.e.* distanța între valorile maxime și minime), în mai multe intervale de lungime egală și afișează numărul de observații care se încadrează în fiecare interval.

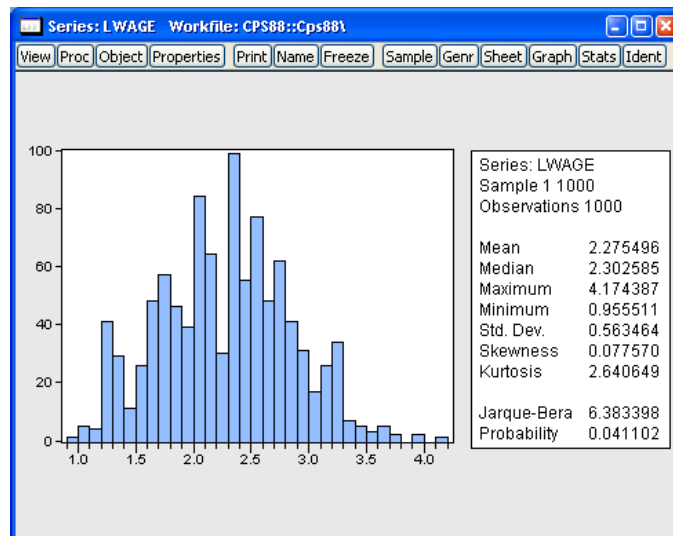
Lângă histogramă sunt afișate o serie de **statistici descriptive**, care numere care oferă o imagine de ansamblu asupra distribuției seriei analizate.

- **Mean** reprezintă media seriei de date, obținută prin însumarea tuturor valorilor și împărțirea la numărul de observații.

- **Median** reprezintă mediana seriei de date, definită ca fiind valoarea din mijlocul seriei atunci când valorile acesteia sunt ordonate în ordine crescătoare. Mediana este o măsură robustă a „tendenței centrale”, fiind mai puțin sensibilă față de valori aberante decât media.

- **Max** și **Min** reprezintă valoarea maximă și, respectiv, minimă înregistrată de serie pentru eșantionul curent.

- **Std. Dev.** reprezintă abaterea standard, care, așa cum s-a mai discutat, este o măsură a dispersiei sau a „împrăștierii” valorilor seriei față de medie.



• **Skewness** este o măsură a asimetriei funcției de densitate de repartiție a seriei în jurul valorii sale medii. Acest indicator se după formula:

$$S = \frac{1}{N} \sum_{i=1}^N \left(\frac{y_i - \bar{y}}{\sigma} \right)$$

unde σ este un estimator pentru abaterea standard. Statistica skewness pentru o distribuție simetrică, cum ar fi distribuția normală, este întotdeauna zero. O valoare pozitivă semnifică faptul că distribuția are „coada” din partea dreaptă mai lungă, iar o valoare negativă implică faptul că distribuția are „coada” din partea stângă mai lungă.

• **Kurtosis** este o măsură a amplitudinii funcției de densitate, a aplatizării acesteia în raport cu funcția de densitate a distribuției normale. Această statistică se calculează după formula:

$$K = \frac{1}{N} \sum_{i=1}^N \left(\frac{y_i - \bar{y}}{\sigma} \right)^4$$

Valoarea acestei statistici pentru o serie distribuită normal este 3. Dacă statistica înregistrează o valoare mai mare decât 3, se spune că distribuția este **leptokurtică**, iar dacă are o valoare mai mică decât 3, distribuția este **platikurtică**.

2 Noțiuni privind statistica matematică

În acest capitol, ne vom îndrepta atenția asupra unor concepte și metode care sunt legate în mod explicit de problema de inferență statistică. Instrumentele teoriei probabilităților, prezentate în capitolul precedent, se utilizează, în principal, la analiza unor întrebări precum: "Având în vedere un câmp de probabilitate, ce se poate deduce despre caracteristicile rezultate ale unui experiment?" Pe de altă parte însă, inferența statistică va analiza această întrebare dintr-o altă perspectivă: "Având în vedere caracteristicile rezultatelor unui experiment, ce putem deduce despre câmpul de probabilitate care a generat aceste rezultate?".

Termenul de **inferență statistică** se referă la procesul inductiv de generare de informații cu privire la caracteristicile unei populații din lumea reală sau unui proces, prin analizarea unui eșantion de obiecte, de rezultate din cadrul populației respective sau al procesului respectiv. Problemele de inferență statistică implică analiza unor observații legate de eșantioane privind o populație sau un proces, dar care, în cele din urmă, conduce la concluzii în ceea ce privește caracteristicile întregii populații.

2.1 Comportamentul asimptotic al variabilelor aleatoare

În cadrul acestei secțiuni se analizează o serie de rezultate referitoare la caracteristicilor funcțiilor care depind de n variabile aleatoare atunci când n este mare. În anumite cazuri, anumite tipuri de funcții care depind de n variabile aleatoare pot converge în diverse moduri către o constantă, sau către o distribuție „limită”. Există o serie de motive care induc importanța studiului "comportamentului asimptotic" al unei serii de variabile aleatoare. În practică, astfel de funcții care depind de n variabile aleatoare sunt utilizate pentru estimarea parametrilor unui model macro-econometric sau pentru testarea ipotezelor. În acest

caz, n reprezintă numărul de observații disponibile referitoare la experimentul care este analizat.

Pentru a putea evalua și compara proprietățile acestor proceduri statistice, și chiar pentru a putea defini metode de estimare sau de testare a ipotezelor statistice este necesar să se stabilească caracteristicile unei funcții care depinde de un număr mare de variabile aleatoare. Din păcate, deseori în practica statistică și econometrică se întâmplă ca densitatea de probabilitate reală a unei astfel de combinații să fie prea dificil de determinat în forma analitică. Teoria asimptotică identifică metode care furnizează aproximări potrivite ale distribuției de probabilitate atunci când n este suficient de mare, și, prin urmare oferă, de asemenea, un mijloc de evaluare, comparare, precum și de definire a diverse proceduri de inferență statistică.

O diferență fundamentală între o serie de numere reale și o serie de variabile aleatoare se referă la faptul că elementele din cea de a doua serie cuprind variabile aleatoare care pot lua diverse valori în mulțimea numerelor reale. Având în vedere că seria este aleatoare, întrebările referitoare la convergență și mărginire nu pot fi verificate ca fiind fără echivoc adevărate sau false, așa cum se întâmplă în cazul unui șir de numere reale, dar lor li se poate atribui o probabilitate de apariție, în contextul câmpului de probabilitate asociat evenimentului analizat.

Conceptul de **convergență în distribuție** analizează dacă șirul de funcții de repartiție asociat șirului de variabile aleatoare converge sau nu către o funcție de repartiție limită. În multe cazuri de interes practic, convergența în distribuție poate fi, de asemenea, caracterizată în termeni de convergență a unei serii de funcții de densitate către o funcție de densitate limită. Utilitatea conceptului rezidă în capacitatea de a obține o aproximare a funcției de repartiție sau de densitate, atunci când n este suficient de mare (unde "suficient de mare" înseamnă că funcțiile de repartiție asociate variabilelor din șir CDF sunt „aproape” funcția de repartiție limită). O astfel de aproximare este extrem de utilă atunci când formula analitică a funcției de repartiție/densitate este foarte dificil, sau imposibil, de determinat, însă

aproximarea oferită de funcția de repartiție limită este mai ușor de definit sau analizat.

În sensul cel mai general al termenului, o **distribuție asimptotică** pentru o funcție care depinde de n variabile aleatoare este o distribuție, care oferă o aproximare a distribuției reale a acestei funcții atunci când n este mare (*i.e.* tinde la infinit). În acest caz, dacă șirul are o distribuție limită, deoarece această distribuție limită poate fi interpretată ca o aproximare a distribuției reale a funcției pentru n mare, distribuția limită poate fi considerată ca fiind o distribuție asimptotică a funcției respective, ce depinde de un număr mare de variabile aleatoare.

Conceptul de **convergență în probabilitate** invocă întrebarea dacă rezultatele variabilelor aleatoare din cadrul șirului sunt aproape de rezultatele unei variabile aleatoare „limită”, cu grad ridicat de probabilitate, atunci când n este suficient de mare. În caz afirmativ, rezultatele pentru această variabilă aleatoare „limită” pot servi ca o aproximare a rezultatelor variabilei aleatoare din cadrul șirului, pentru n suficient de mare.

Celelalte două tipuri de distribuții prezintă o importanță mai mică în analiza econometrică și, ca urmare, nu vor fi analizate în profunzime.

Fie (Ω, \mathcal{F}, P) un câmp de probabilitate și $(X_n)_n$ variabile aleatoare i.i.d. cu $E[X_n] = \mu, \sigma[X_n] = \sigma$. Teorema numerelor mari se referă la convergența șirului format din media aritmetică a n variabile aleatoare i.i.d. Mai precis, **teorema numerelor mari** este reflectată în formula $\frac{X_1 + X_2 + \dots + X_n}{n} \xrightarrow{p} \mu$. De

asemenea, se poate obține o caracterizare în ceea ce privește modul de convergență al seriei formate din media aritmetică a n variabile aleatoare i.i.d. Acest lucru este

descriș de **teorema limită centrală**, respectiv $\frac{X_1 + X_2 + \dots + X_n - n\mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{d} \Phi(0,1)$. În

mod echivalent, teorema limită centrală arată faptul că șirul studiat are o

distribuție asimptotică descrisă de către o distribuție normală,

$$\frac{X_1 + X_2 + \dots + X_n}{n} \overset{a}{\sim} \Phi\left(\mu, \frac{\sigma^2}{n}\right).$$

Distribuția normală este des întâlnită în inferența statistică. O variabilă aleatoare distribuită normal cu medie 0 și varianță 1 se spune că are o **distribuție normală standard**. Alte distribuții întâlnite în analiza asimptotică a variabilelor aleatoare sunt **distribuția χ^2** (citit hi pătrat) și **distribuția Student** sau t . Valorile funcțiilor de repartiție pentru toate aceste distribuții se pot obține din tabele specializate care se găsesc în cărțile de statistică și econometrie, sau pot fi calculate cu ajutorul programului **Eviews**.

În acest moment, cititorul se poate întreba de ce problema particulară privind convergența în distribuție sau în probabilitate, definite mai sus, merită o astfel de atenție în mod explicit. Răspunsul este dat de faptul că un număr mare de proceduri de estimare a parametrilor și de testare a ipotezelor în econometrie și statistică sunt definite ca funcții de variabile aleatoare.

2.2 Estimatori statistici

Problema de estimare examinată în această secțiune se concentrează asupra estimării, determinării valorii unor parametri necunoscuți sau a unor funcții care depind de acești parametri, și care reprezintă caracteristici de interes legate de câmpul de probabilitate asociat unui set de experimente din domeniul economic, sociologic sau al științelor naturale. Rezultatele generate de această colecție de experimente se presupun a fi rezultatele unui eșantion aleator, cu o funcție de densitate de repartiție care depinde de un parametru, a cărei valoare nu este cunoscută, dar care se dorește a fi determinată. Trebuie menționat, că eșantionul aleator nu trebuie să fie o colecție de variabile aleatoare i.i.d. Obiectivul procedurilor de **estimare statistică** constă în determinarea unei combinații între variabilele aleatoare din eșantion (*i.e.* o funcție care depinde de acestea) care să

posede o serie de proprietăți și care să îl facă util pentru a deduce caracteristicile de interes necunoscute.

Un **model statistic** pentru un eșantion aleatoriu constă într-o funcție de densitate de repartiție a eșantionului care depinde de un parametru care poate lua valori într-un spațiu bine determinat, acesta definind setul de potențiali candidați pentru densitatea adevărată a eșantionului. Modelul statistic prezintă contextul probabilistic în care are loc procedura de estimare parametrică. Odată ce modelul statistic a fost specificat, centrul de interes cade pe estimarea valorilor unor (sau tuturor) parametri, sau pe estimarea valorilor unor funcții care depind de parametrii problemei.

O **statistică** reprezintă o funcție care depinde de variabilele aleatoare din cadrul modelului statistic analizat. Trebuie subliniat faptul că, fiind o combinație de variabile aleatoare, o statistică este la rândul ei o variabilă aleatoare. Ca urmare, o statistică este caracterizată printr-o funcție de repartiție, printr-o funcția de densitate de repartiție și i se poate calcula media și varianța.

Un **estimator** este o statistică utilizată cu scopul de a estima parametrul sau vectorul de parametri asociat modelului statistic de interes. Deci, **estimatorul este o variabilă aleatoare**. O realizare a acestei variabile aleatoare, obținută pentru un set specific de date, poartă numele de estimare a parametrului respectiv. Fiind o variabilă aleatoare, un estimator este caracterizat de o funcție de densitate de repartiție și de o funcție de repartiție, iar unui estimator i se poate calcula media și varianța. Abaterea medie pătratică a unui estimator poartă numele de **eroarea standard (en. standard error)** asociată estimatorului.

Determinarea de **estimatori „buni”** constituie unul din obiectivele statisticii matematice și, în special, al econometriei. Estimatorii sunt comparați pe baza unei serii de caracteristici. Proprietățile de tip „small sample” sunt acele proprietăți valabile indiferent de mărimea eșantionului disponibil. Proprietățile de tip „large sample” sunt valabile asimptotic, adică în cazul în care mărimea eșantionului este foarte mare, tinzând la infinit. În acest caz, estimatorii sunt

comparați pe baza distribuției lor asimptotice. Un **estimator nedepășat** are proprietatea că media sa este egală cu valoarea reală a parametrului care se dorește estimat. Proprietatea de nedepășare este o proprietate de tip „small sample”. Un **estimator consistent** are proprietatea că acesta converge în probabilitate către valoarea reală a parametrului. Proprietatea de consistență este o proprietate de tip „large sample”. În general, se consideră că un **estimator este „bun” dacă este consistent**. Estimatorii, așa cum s-a mai precizat, sunt variabile aleatoare și, ca urmare, acestora li se poate calcula varianța. Dintre doi estimatori care sunt „buni”, adică sunt consistenți, este „mai bun” estimatorul cu varianță mai mică. Estimatorul care este caracterizat prin faptul că are cea mai mică varianță, în cadrul unei clase de estimatori ai aceluiași parametru, poate fi considerat „cel mai bun” estimator și poartă denumirea de **estimator eficient**.

Există o serie de metode de estimare statistică care generează estimatori „buni” (i.e. nedepășați și/sau consistenți), dintre care menționăm **MLE** (Maximum Likelihood Estimation), **GMM** (Generalized Method of Moments) și **LS** (Least Squares). De exemplu, parametrii unei ecuații de regresie pot fi estimați, în mod consistent, prin una din aceste metode, în funcțiile de setul de ipoteze de la care se pornește. Trebuie menționat faptul că MLE este singura metodă de estimare care asigură obținerea „celor mai buni” estimatori (i.e. consistenți și eficienți).

2.3 Teste statistice

Al doilea grup important de proceduri de inferență statistică se referă la testarea ipotezelor statistice. Procedurile de testare se bazează pe construirea unei statistici pe baza unui eșantion, care va permite analistului să decidă, cu o probabilitate rezonabilă, dacă datele din eșantion ar fi fost generate de un proces care este caracterizat de o anumită proprietate testată. Procedura implică o specificare a acestei ipoteze, denumită, de obicei, **ipoteza nulă**, precum și a unei ipoteze alternative, notate convențional cu H_0 și, respectiv H_1 . Procedura de testare în sine constă în determinarea unei reguli, bazată pe calculul unei statistici,

care dictează dacă ipoteza nulă ar trebui să fie respinsă sau nu. De exemplu, ipoteza nulă ar putea fi că un parametru este egal cu o valoare specificată. O posibilă regulă de decizie poate afirma că ipoteza trebuie să fie respinsă în cazul în care o estimare a acestui parametru este „prea departe” de această valoare specificată. Însă regula, procedura de testare nu este infailibilă, putând genera erori. Există două tipuri de erori care pot apărea atunci când se efectuează un test statistic:

- **eroarea de tip I** - procedura de testare respinge ipoteza nulă a testului, atunci când aceasta, în realitate, este adevărată;
- **eroarea de tip II** - procedura de testare acceptă (nu respinge) ipoteza nulă a testului, atunci când aceasta, în realitate, este falsă;

Nivelul de semnificație al unui test statistic este probabilitatea ca procedura de testare să respingă ipoteza nulă a testului, atunci când aceasta este adevărată sau, altfel spus, reprezintă probabilitatea de apariție a unei erori de tip I. Nivelul de semnificație marginală a unui test se numește **p-value**. Având la dispoziție p-value asociat unui test, se poate spune rapid dacă se respinge sau se acceptă ipoteza nulă. În cazul mării majorități a testelor puse la dispoziție de către EViews, acesta calculează p-value și astfel interpretarea rezultatelor testului respectiv este relativ simplă.

Eviews pot fi utilizat pentru a realiza teste statistice privind media, mediana și varianța unei serii de date. Se selectează **View/Descriptive Statistics & Tests/Simple Hypothesis Tests** și se va afișa căsuța de dialog **Series Distribution Tests**.

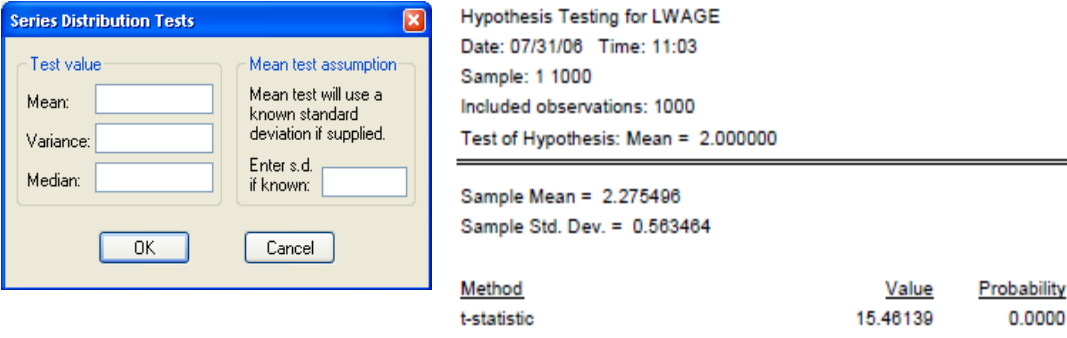
Opțiunea **Mean Test** efectuează un test care are ca ipoteză nulă că media seriei (μ) este egală cu o valoare specificată m , iar ca ipoteză alternativă că media nu este egală cu m :

$$H_0 : \mu = m , H_1 : \mu \neq m$$

Dacă nu se specifică deviația standard a seriei analizate, care de obicei nu este cunoscută, **EViews** raportează o statistică t calculat astfel:

$$t = \frac{\bar{X} - m}{s / \sqrt{N}}$$

unde \bar{X} reprezintă media de selecție a eșantionului, s este un estimator nedeplasat pentru deviația standard a eșantionului și N este numărul de observații în cadrul eșantionului.



The image shows two parts of the EViews software interface. On the left is the 'Series Distribution Tests' dialog box, which has fields for 'Test value' (Mean, Variance, Median) and 'Mean test assumption' (Mean test will use a known standard deviation if supplied, Enter s.d. if known). On the right is the output window titled 'Hypothesis Testing for LWAGE', showing the date, time, sample size (1000), and included observations (1000). It reports a 'Test of Hypothesis: Mean = 2.000000'. Below this, it shows 'Sample Mean = 2.275496' and 'Sample Std. Dev. = 0.563464'. At the bottom, a table displays the test results:

Method	Value	Probability
t-statistic	15.46139	0.0000

Dacă X este normal distribuit, sub ipoteza nulă, statistica t urmează o distribuție Student-t cu $N-1$ grade de libertate. Chiar și în cazul în care X nu are o distribuție normală, se poate realiza testarea ipotezei respective, deoarece statistica t are o distribuție asimptotică normală standard, putându-se vorbi în acest caz de un test z .

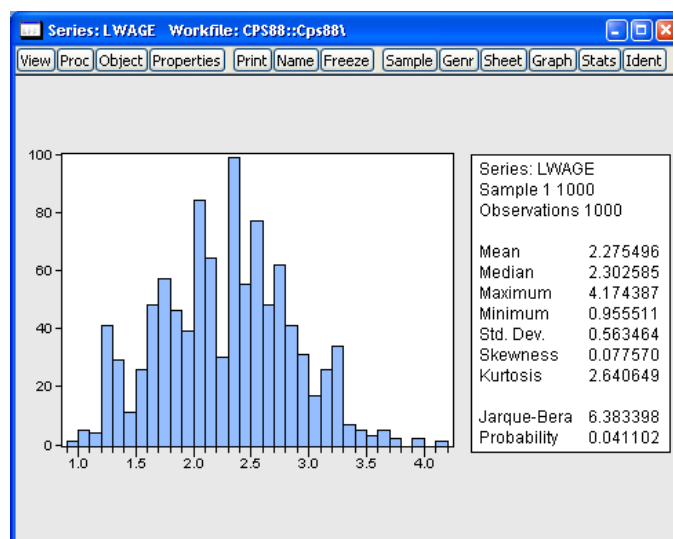
Pentru a efectua un test de medie, se testează valoarea mediei sub ipoteza nulă în câmpul editabil **Mean**. Dacă se dorește calcularea statisticii condiționate pentru o deviere standard cunoscută, de asemenea, se testează o valoare pentru abaterea standard în câmpul de editare corespunzător. Valoarea probabilității raportate este de fapt un p-value, sau nivelul de semnificație marginal față de ipoteza alternativă. Dacă această valoare a probabilității este mai mică decât nivelul de semnificație fixat testului, să spunem 5%, vom respinge ipoteza nulă.

Un alt test des întâlnit în practică este **testul Jarque-Bera** utilizat pentru a verifica dacă o serie de date este distribuită normal. Ipoteza nulă a acestui test constă în faptul că seria este normal distribuită. Statistica este o măsură a distanței dintre indicatorii Skewness și Kurtosis ai seriei analizate față de cele ale distribuției normale. Statistica se calculează astfel:

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

unde S este indicatorul Skewness, iar K este indicatorul Kurtosis. Sub ipoteza nulă a unei distribuții normale, statistica Jarque-Bera este distribuită χ^2 cu 2 grade de libertate.

În **Eviews**, pentru a efectua un test Jarque-Bera, se alege opțiunea **Histogram and Stats**.



Probabilitatea raportată reprezintă p-value asociat testului, adică este probabilitatea ca statistica Jarque-Bera să depășească (în valoare absolută), valoarea observată sub ipoteza nulă. O valoare „mică” pentru p-value indică respingerea ipotezei nule că seria are o distribuție normală.

3 Modelul clasic de regresie liniară

Modelul clasic de regresie liniară reprezintă una dintre tehnicile statistice cele mai versatile și mai des utilizate în analiza economică, și nu numai. În continuare, sunt descrise și analizate tehnicile de regresie puse la dispoziție de către programul econometric **Eviews**: specificarea și estimarea unui model de regresie, efectuarea de analize simple de diagnostic și utilizarea rezultatelor estimării în analize suplimentare.

3.1 Ipotezele modelului clasic de regresie

Modelul clasic de regresie liniară presupune o ca sunt îndeplinite o serie de ipoteze. Validitatea rezultatelor unei regresii este strâns legată de validitatea ipotezelor presupuse.

Ipoteza 1 (I1). Între variabila dependentă (y) și variabila independentă (x) există o **relație liniară**. Însă, datorită faptului că pot exista și alți factori de influență, această relație nu este deterministă ci are loc „în medie”. Factorii care nu au fost luați în considerare, în mod explicit, își manifestă influența prin intermediul unei variabile reziduale, ε . Astfel, modelul presupune că există următoarea relație:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Ipoteza 1 este o ipoteză fundamentală, fără îndeplinirea căreia nu se poate discuta despre un model *liniar* de regresie.

În cazul în care se utilizează mai mulți factori de influență, folosind o notație matricială, o ecuație de regresie liniară (multidimensională) poate fi scrisă sub forma:

$$y = X\beta + \varepsilon$$

unde y este un vector N -dimensional care conține observațiile privind variabila dependentă, X este o matrice de dimensiune $N \times K$ cu variabile

independente sau explicative, β este un vector K dimensional al coeficienților de regresie și ε este un vector N dimensional reprezentând inovațiile asociate ecuației, adică aceea componentă din dinamica variabilei dependente care nu este captată de către cele K variabile independente. N reprezintă numărul de observații, iar K este numărul de regresori (variabile explicative) din partea dreaptă a ecuației.

Ipoteza 2 (I2). Variabila explicativă, x , este **exogenă**, în sensul că nu este corelată cu variabila reziduală:

$$COV(x_i, \varepsilon_i) = 0.$$

Ipoteza 2 este o ipoteză esențială, ne-îndeplinirea acesteia conducând la invalidarea a multe rezultate privind modelul liniar de regresie și generând așa numita **„problemă de endogenitate”** care va fi analizată în cadrul Capitolului 5.

Ipoteza 3 (I3). Variabilele reziduale, cunoscute și sub denumirea de reziduuri, inovații sau erori, sunt **„sferice”**, adică sunt homoskedastice și nu sunt autocorelate. Homoskedasticitatea se referă la faptul că toate variabilele reziduale care apar au aceeași varianță:

$$VAR(\varepsilon_i) = \sigma^2.$$

Lipsa autocorelării se referă la faptul că oricare două variabile reziduale nu sunt corelate:

$$COV(\varepsilon_i, \varepsilon_j) = 0.$$

Ipoteza 3 este o ipoteză importantă, ne-îndeplinirea acesteia conducând la invalidarea a unor rezultate privind modelul liniar de regresie și generând așa numitele **„problemă de heteroskedasticitate”** și **„problemă de autocorelare”** care vor fi analizate în cadrul Capitolului 4.

Ipoteza 4 (I4). Variabilele reziduale au o **distribuție normală**.

Ipoteza 4 este o ipoteză mai puțin importantă decât celelalte trei ipoteze analizate. Ne-îndeplinirea acesteia nu conduce la invalidarea rezultatelor importante privind estimare și inferența modelului liniar de regresie.

3.2 Estimatorul OLS

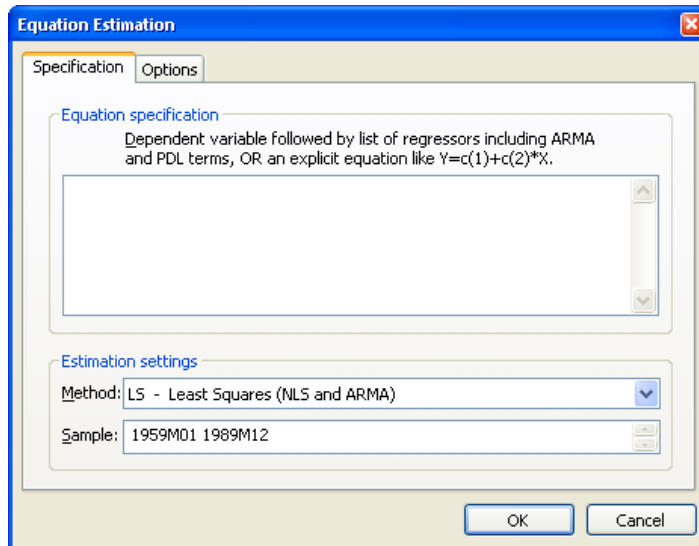
Estimatorul OLS presupune minimizarea distantei dintre valoarea efectivă a variabilei dependente și cea presupusă de către relația liniară care există între aceasta și variabila explicativă. Altfel spus, metoda „celor mai mici pătrate” presupune faptul că parametrii modelului se determină astfel încât să se minimizeze distanța dintre norul de puncte reprezentat de observațiile efective și o dreaptă care aproximează cel mai bine relația dintre variabila dependentă și cea independentă.

Ca orice estimator, estimatorul OLS este o variabilă aleatoare, o procedură de determinare a valorilor parametrilor modelului liniar de regresie. Proprietățile estimatorului OLS depind de gradul de îndeplinire a ipotezelor modelului clasic de regresie.

Dacă sunt **îndeplinite ipotezele I1, I2, I3, I4** estimatorul OLS este „**cel mai bun**” estimator posibil, în sensul că este **consistent** și **eficient**. De asemenea, acesta are o **distribuție normală**.

Dacă sunt **îndeplinite doar ipotezele I1, I2, I3** estimatorul OLS este, în continuare, un estimator „**bun**”, dar nu mai este „cel mai bun” estimator posibil, în sensul că este **consistent**, dar nu este **eficient**. De asemenea, în aceste condiții, estimatorul OLS are o **distribuție asimptotic normală**.

Estimarea unei ecuații de regresie se realizează în **Eviews** cu ajutorul unui obiect de tip *equation*. Pentru a crea un astfel de obiect trebuie să selectați **Object/New Object.../Equation** sau **Quick/Estimate Equation...** din meniul principal. În continuare, va fi afișată caseta de dialog **Equation Estimation**. În cadrul acestei casete de dialog este necesară precizarea a trei aspecte: specificațiile ecuației, metoda de estimare și eșantionul care urmează să fie utilizat în estimare.



Cea mai simplă modalitate de a specifica o ecuație liniară de regresie, este de a preciza lista variabilelor care se doresc a fi utilizate în ecuație. În primul rând, trebuie să includă numele variabilei dependente, urmată de o listă a variabilelor explicative. Având specificată ecuația, în continuare trebuie aleasă o metodă de estimare. Prin apăsarea pe lista de opțiuni **Method** se va observa un meniu listă cu metode de estimare. În cadrul modelului clasic de regresie estimarea se realizează cu ajutorul metodei celor mai mici pătrate (OLS). Alte metode de estimare vor fi analizate în capitolele ulterioare. **Eviews** oferă o serie de opțiuni de estimare. Aceste opțiuni permit estimarea ecuației, mai exact de calculul de erori standard care sunt robuste față de abaterile ipotezelor modelului clasic, abateri legate de existența heteroskedasticității și/sau a autocorelării inovațiilor. Aceste opțiuni de estimare a erorilor standard sunt discutate în capitolul următor. Prin apăsarea butonului din căsuța de dialog **Equation Estimation**, **Eviews** afișează o fereastră care conține un ecran cu rezultatele estimării:

Dependent Variable: LOG(M1)
 Method: Least Squares
 Date: 08/18/97 Time: 14:02
 Sample: 1959:01 1989:12
 Included observations: 372

	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.699912	0.164954	-10.30539	0.0000
LOG(IP)	1.765866	0.043546	40.55199	0.0000
TB3	-0.011895	0.004628	-2.570016	0.0106
R-squared	0.886416	Mean dependent var		5.663717
Adjusted R-squared	0.885800	S.D. dependent var		0.553903
S.E. of regression	0.187183	Akaike info criterion		-0.505429
Sum squared resid	12.92882	Schwarz criterion		-0.473825
Log likelihood	97.00980	Hannan-Quinn criter.		0.008687
F-statistic	0.000000			

Coloana denumită **Coefficient** descrie coeficienții estimați. Estimatorul OLS pentru coeficienții ecuației de regresie este calculat prin formula:

$$b = (X'X)^{-1} X'y$$

Coloana denumită **Variable** precizează numele variabilei căreia îi corespunde paramerul din coloana alăturată.

Fiecare parametru astfel estimat măsoară contribuția marginală a variabilei independente respective asupra variabilei dependentă, în condițiile în care toate celelalte variabile explicative nu își modifică valorile. Dacă este prezent, coeficientul asociat variabilei denumită *C* este, de fapt, constanta din ecuația de regresie, reprezentând nivelul mediu al variabilei dependente atunci când toate variabile independente sunt zero. Ceilalți parametri pot fi interpretați ca fiind indicatori de senzitivitate ai relației dintre variabila independentă corespunzătoare și variabila dependentă, presupunând că valorile tuturor celorlalte variabile nu se modifică.

3.3 Inferența statistică a modelului clasic de regresie

Coloana denumită **Std. Error** prezintă erorile standard estimate ale estimărilor coeficienților ecuației de regresie. **Erorile standard** măsoară fiabilitatea statistică a estimării unui parametru – cu cât sunt mai mari erorile standard, cu atât este mai mult „zgomot statistic” în estimări. Matricea de varianță-covarianța a coeficienților estimați, în cazul estimatorului OLS, se calculează astfel:

$$COV(b) = s^2 (X'X)^{-1}, s^2 = \hat{\varepsilon}'\hat{\varepsilon} / (N - K), \hat{\varepsilon} = y - Xb$$

unde $\hat{\varepsilon}$ reprezintă reziduul ecuației. Erorile standard ale coeficienților estimați sunt rădăcinile pătrate ale elementelor de pe diagonala matricei de varianță-covarianță a coeficienților. Întreaga matrice de varianță-covarianță se poate vizualiza prin utilizarea opțiunii **View/Covariance Matrix**.

Coloana denumită **t-statistic** prezintă **statistica testului t** care se calculează ca raportul dintre mărimea coeficientul estimat și eroarea sa standard și este utilizat pentru a testa **ipoteza nulă că respectivul coeficient este egal cu zero**. Pentru a interpreta valoarea statisticii t , trebuie să se examineze probabilitatea de a observa această valoare a statisticii având în vedere că respectivul coeficient este, în realitate, egal cu zero. Această probabilitate este descrisă în continuare.

Ultima coloană a ecranului cu rezultate, numită **Prob.**, prezintă probabilitatea de a obține o valoare teoretică a statisticii t la fel de mare în modul ca cea observată în urma calculării acesteia pentru eșantionul considerat. Calculul acestei probabilități se realizează plecându-se de la ipoteza că erorile sunt normal distribuite sau că estimatorul considerat are o distribuție asimptotică normală. De exemplu, în cazul în care inovațiile sunt distribuite normal, statistica t are o distribuție de tip Student- t cu $N-K$ grade de libertate.

Această probabilitate este, de asemenea, cunoscută sub numele de **p-value** și reprezintă nivelul de semnificație marginală a testului. Având la dispoziție p-value asociat acestui test, se poate spune rapid dacă se respinge sau se acceptă

ipoteza nulă conform căreia coeficientul este zero, față de ipoteza alternativă care presupune că parametrul este diferit de la zero. De exemplu, dacă se dorește efectuarea testului la un nivel de semnificație de 5%, un p-valoare mai mic de 5% poate fi considerat o dovadă pentru a respinge ipoteza nulă conform căreia paramerul respectiv este zero.

Statistica **R-squared** măsoară „succesul” cu care ecuația de regresie estimată reușește să explice valoarea variabilei dependente în cadrul eșantionului. În mod normal, această statistică poate fi interpretată ca fracțiunea din varianța variabilei dependente explicată de variabilele independente. Statistica este egală cu 1 în cazul în care ecuația de regresie se potrivește perfect și zero în cazul în care nu se potrivește mai bine decât media variabilei dependente.

O problemă majoră în ceea ce privește utilizarea statisticii **R-squared**, ca o măsură a “potrivirii” modelului la datele disponibile, se referă la faptul că valoarea acestei statistici nu scade niciodată pe măsură ce se adaugă mai mulți regresori. Astfel, în cazuri extreme, se poate obține o statistică egală cu 1 dacă se includ atât de mulți regresori independenți, câte observații sunt în eșantion. Statistica **Adjusted R-squared** reprezintă o alternativă, aceasta având avantajul că “penalizează” adăugarea de regresori care nu contribuie la puterea explicativă a modelului. Astfel, această statistică poate scădea pe măsură ce sunt adăugați regresori, iar pentru modelele pentru care “potrivirea” la date nu este foarte bună, poate fi chiar negativ.

Statistica **Durbin-Watson** reprezintă o măsură a corelației seriale a reziduurilor. Ca o regulă desprinsă din experiență, în cazul în care DW are o valoare mai mică de 2, există dovezi de corelație serială pozitivă. Există teste mai puternice pentru analiza existenței corelației seriale în reziduurile ecuației de regresie, cum ar fi testul Q, precum și testul Breusch-Godfrey, ambele oferind un cadru de testare mai general decât testul Durbin-Watson.

Criteriul informațional Akaike (**Akaike Information Criterion - AIC**) este adesea folosit în selecția între două modele estimate pentru același set de date.

Astfel, modelele caracterizate prin valori mai mici ale AIC sunt de preferat celor caracterizate de valori mai mari. **Schwarz Criterion** (SC) reprezintă o alternativă a criteriului AIC folosit însă în același scop.

F-statistic reprezintă statistica asociată testului care are drept ipoteză nulă că toți coeficienții din regresie (mai puțin constanta) sunt zero. Valoarea p-value asociată, notată **Prob(F-statistic)**, este nivelul de semnificație marginal al acestui test. În cazul în care valoarea p-value este mai mică decât nivelul de semnificație testat, să zicem 1%, se respinge ipoteza nulă conform căreia toți coeficienții sunt egali cu zero.

În cazul în care regresorii sunt caracterizați printr-un grad înalt de coliniaritate, **Eviews** poate întâmpina dificultăți în calculul estimatorului OLS. În astfel de cazuri, programul va emite un mesaj de eroare, "**Near singular matrix**". Dacă apare acest mesaj de eroare, trebuie verificat dacă regresorii sunt coliniari. Regresorii sunt exact coliniari dacă un regresor poate fi scris ca o combinație liniară a altor regresori. În caz de coliniaritate, matricea XX nu este inversabilă și, astfel, estimatorul OLS nu poate fi calculat.

3.4 Efectuarea de prognoze bazate pe modelul liniar de regresie

Modelul liniar de regresie poate fi utilizat pentru realizarea de prognoze cu privire la variabila dependentă, dacă se cunosc informații privitoare la variabila explicativă. Astfel dacă se consideră o entitate pentru care valoarea variabilei independente este x^0 , atunci **previziunea** cu privire la valoarea variabilei dependente se poate calcula utilizând formula:

$$\hat{y}^0 = \hat{\beta}_0 + \hat{\beta}_1 x^0$$

unde $\hat{\beta}_0$ și $\hat{\beta}_1$ reprezintă estimatorii OLS ai parametrilor modelului de regresie.

Pentru a realiza o prognoză în **Eviews** se utilizează butonul **Forecast** din cadrul unei obiect de tip *equation*.

Equation: EQ_1 Workfile: WAGE1::Wage1

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: LOG(WAGE)
 Method: Least Squares
 Date: 10/16/11 Time: 01:50
 Sample (adjusted): 1 526
 Included observations: 526 after adjustments

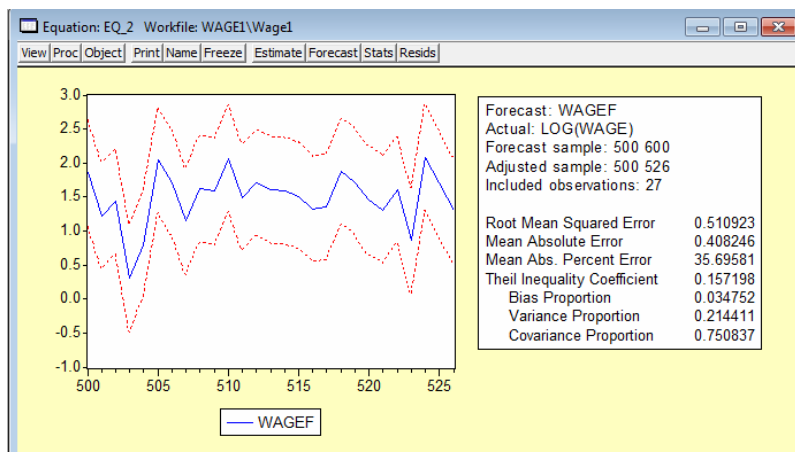
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.583773	0.097336	5.997510	0.0000
EDUC	0.082744	0.007567	10.93534	0.0000

R-squared	0.185806	Mean dependent var	1.623268
Adjusted R-squared	0.184253	S.D. dependent var	0.531538
S.E. of regression	0.480079	Akaike info criterion	1.374061
Sum squared resid	120.7691	Schwarz criterion	1.390279
Log likelihood	-359.3781	Hannan-Quinn criter.	1.380411
F-statistic	119.5816	Durbin-Watson stat	1.801328
Prob(F-statistic)	0.000000		

Eroare de prognoză asociată acestei previziuni se poate calcula utilizând formula:

$$\hat{y}^0 - y^0 = (\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1)x^0 + \varepsilon$$

Varianța erorii de prognoză depinde de incertitudinea în ceea ce privește valoarea parametrilor, precum și de incertitudinea în ceea ce privește variabila reziduală. Magnitudinea varianței erorii de prognoză determină mărimea **intervalului de încredere** asociat unei previziuni.



Intervalul de încredere cu probabilitate de 95% asociat unei previziuni punctuale \hat{y}^0 poate fi aproximat prin $(\hat{y}^0 - 2\sigma_{\hat{y}^0}, \hat{y}^0 + 2\sigma_{\hat{y}^0})$. Astfel, intervalul de încredere este simetric, în jurul previziunii punctuale. Abaterea medie pătratică a erorii de prognoză este dată de formula $\sigma_{\hat{y}^0} = \sqrt{se(\hat{\beta}_0)^2 + [x^0 se(\hat{\beta}_1)]^2 + \sigma_\varepsilon^2}$. În general, datorită faptului că eroarea standard asociată unui parametru este mică, abaterea medie pătratică a erorii de prognoză este influențată în special de către abaterea medie pătratică a variabilei reziduale, adică $\sigma_{\hat{y}^0} \approx \sigma_\varepsilon$. Abaterea medie pătratică a variabilei reziduale se regăsește în cadrul unui output **Eviews** al unei ecuații de regresie sub denumirea de **S.E. of regression**.

Se întâmplă deseori ca variabila dependentă unei ecuații de regresii să fie **logaritm natural dintr-o variabilă economică** de interes:

$$\ln(y_i) = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Astfel dacă se consideră o entitate pentru care valoarea variabilei independente este x^0 , atunci **previziunea** cu privire la logaritm din valoarea variabilei de interes este:

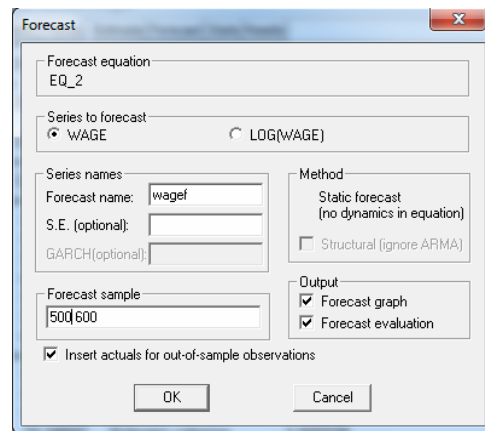
$$\hat{\ln}(y^0) = \hat{\beta}_0 + \hat{\beta}_1 x^0$$

unde $\hat{\beta}_0$ și $\hat{\beta}_1$ reprezintă estimatorii OLS ai parametrilor modelului de regresie.

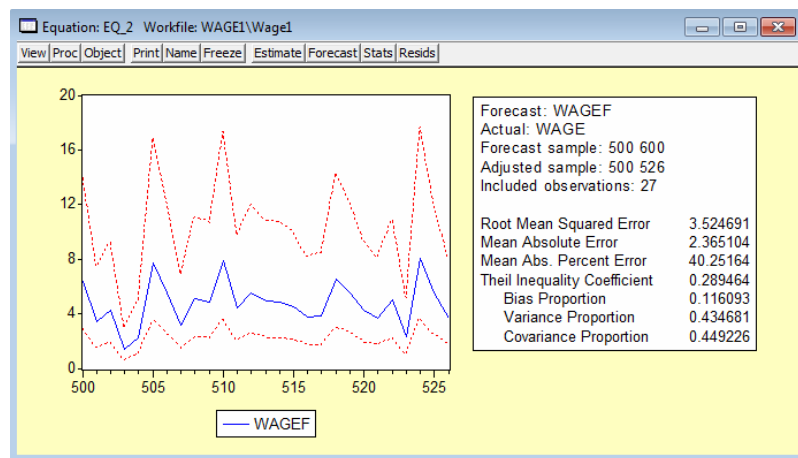
Trebuie subliniat faptul că obținerea previziunii pentru variabila economică de interes, \hat{y}^0 , nu este o simplă transformare exponențială a previziunii pentru logaritm din variabila de interes, $\hat{\ln}(y^0)$. Altfel spus $\hat{y}^0 \neq e^{\hat{\ln}(y^0)}$. **Formula de transformare** între cele două previziuni depinde de gradul de îndeplinire al ipotezelor modelului de regresie. Se poate arăta faptul că dacă toate cele patru ipoteze (I1, I2, I3, I4) ale modelului liniar de regresie se respectă atunci previziunea pentru variabila de interes este $\hat{y}^0 = e^{\frac{1}{2}\sigma_\varepsilon^2} e^{\hat{\ln}(y^0)}$. Formula prezentată

anterior nu mai este valabilă dacă reziduurile modelului nu sunt distribuite normal, sau altfel spus, dacă ipoteza I4 nu se respectă.

Eviews permite efectuarea automată de previziuni pentru o variabilă de interes chiar și în cazul în care variabila dependentă a ecuației de regresie estimată este logaritmul natural din aceea variabilă.



Trebuie evidențiat faptul că intervalul de încredere asociat prognozei unei variabile de interes, în cazul în care variabila dependentă a regresiei este logaritmul din variabila respectivă, nu mai este simetric, așa cum se poate observa din figura următoare.



4 Abateri de la ipotezele modelului clasic de regresie I

O ipoteză importantă a modelului clasic de regresie este aceea potrivit căreia inovațiile sunt homoskedastice (i.e. au aceeași varianță) și nu sunt autocorelate. În lipsa acestei ipoteze estimatorii pentru erorile standard ale coeficienților din modelul de regresie, calculați cu formula clasică, nu sunt consistenți.

4.1 Problema de heteroskedasticitate

Pentru a testa dacă inovațiile modelului de regresie sunt homoskedastice, **EViews** pune la dispoziție **testul White**. Ipoteza nulă a acestui test specifică faptul că inovațiile sunt homoskedastice. Statistica testului White presupune construirea a unei ecuații de regresie “auxiliară”: reziduurile la pătrat ale ecuației principale sunt regresate în funcție de variabilele explicative și pătratul acestora. **Eviews** calculează p-value asociat testului, ca urmare interpretarea rezultatelor acestui test este relativ simplă.

În condițiile existenței heteroskedasticității, **estimatorul OLS** este, în continuare, **consistent, distribuit asimptotic normal**, însă **erorile standard** calculate pe baza formulei clasice **nu sunt corecte** și, astfel, **testul t** bazat pe aceste erori standard **este eronat**.

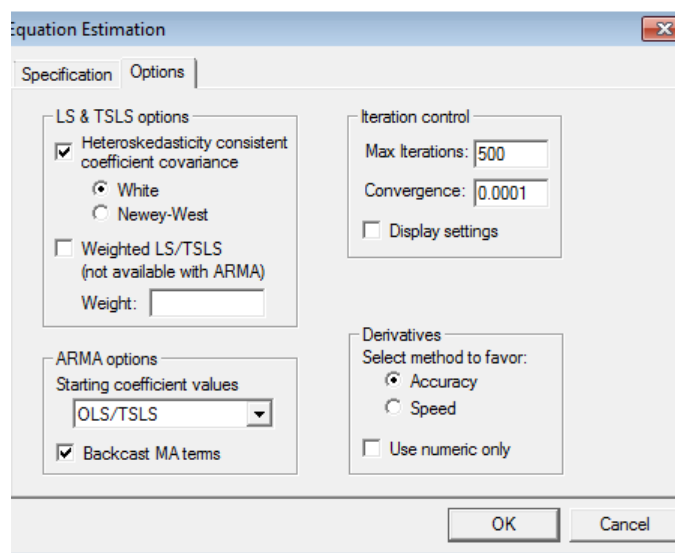
Există **două soluții** în ceea ce privește **“problema heteroskedasticității”**

- păstrăm estimatorul OLS, dar “corectăm” erorile standard;
- înlocuim estimatorul OLS cu un nou estimator.

Prima soluție se utilizează în cazul în care *forma heteroskedasticității nu este cunoscută*, care reprezintă cazul cel mai frecvent întâlnit în practică. Estimatorul OLS oferă estimări consistente ale parametrilor în prezența

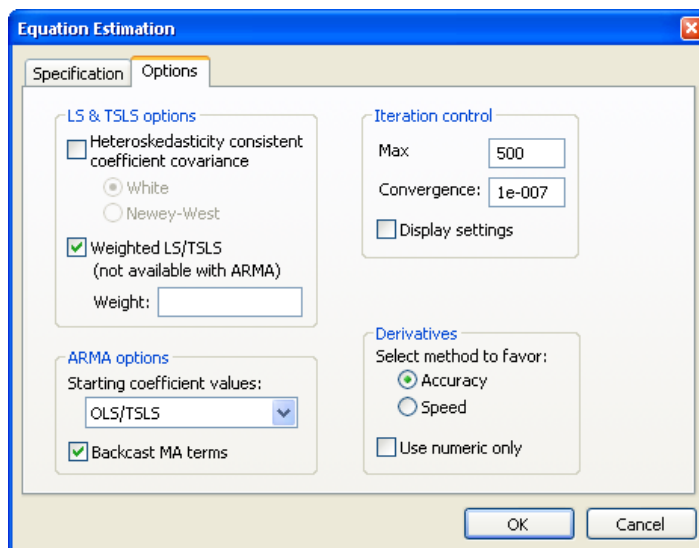
heteroskedasticității, dar erorile standard calculate utilizând metoda OLS ar fi incorecte și nu ar trebui să fie utilizate pentru inferență.

White (1980) a obținut un estimator consistent față de heteroskedasticitate a matricei de varianță-covarianță care oferă estimări corecte ale erorilor standard ale parametrilor modelului liniar de regresie în prezența heteroskedasticității sub formă necunoscută. **Eviews** oferă opțiunea de a utiliza **estimatorul White pentru erorile standard („robust standard errors” de tip White)** în locul formulei standard OLS pentru calculul acestora. Astfel, se deschide caseta de dialog **Equation Estimation** și se specifică ecuația ca și mai înainte, apoi se apasă butonul **Options** și se face clic pe caseta de selectare a metodei de calcul a covarianței **Heteroskedasticity Consistent Covariance** și apoi se face clic pe opțiunea **White**.



Cea de a doua soluție se utilizează în cazul în care se presupune că în ceea ce privește inovațiile modelului liniar de regresie există *heteroskedasticitate sub o formă cunoscută*. Mai exact, se face ipoteza că există o serie ale cărei valori sunt proporționale cu inversele erorilor standard. În această situație, foarte rar întâlnită în practică, există posibilitatea să se utilizeze metoda ponderată a celor mai mici

pătrate (WLS), cu ponderile date de această serie, pentru a corecta heteroskedasticitatea. **Eviews** calculează **estimatorul WLS** prin împărțirea ponderilor din serie cu media acestora, înmulțind apoi toate datele pentru fiecare observație cu aceste ponderi scalate.



Pentru a estima o ecuație cu ajutorul metodei WLS, se deschide meniul principal și se selectează **Quick/Estimate Equation...**, apoi se alege **LS—Least Squares (NLS and ARMA)** din lista mobilă corespunzătoare. Se introduce specificația ecuației și eşantionul în fila **Specification**, apoi se selectează fila **Options** și se face clic pe opțiunea **Weighted LS/TSLS**. Se completează căsuța **Weight** cu numele seriei care conține ponderile.

4.2 Problema de autocorelare

Pentru a testa dacă inovațiile modelului de regresie sunt autocorelate, **Eviews** pune la dispoziție **corelograma**, **testul Ljung–Box** și **testul Breusch – Godfrey**.

Corelograma reprezintă o modalitate „vizuală” de a verifica dacă există corelație între valoarea de la momentul t și valoarea de la un moment anterior al variabilei reziduale.

Ipoteza nulă a **testului Ljung–Box** specifică faptul că inovațiile nu sunt autocorelate. **Eviews** calculează p-value asociat testului, ca urmare interpretarea rezultatelor acestui test este relativ simplă.

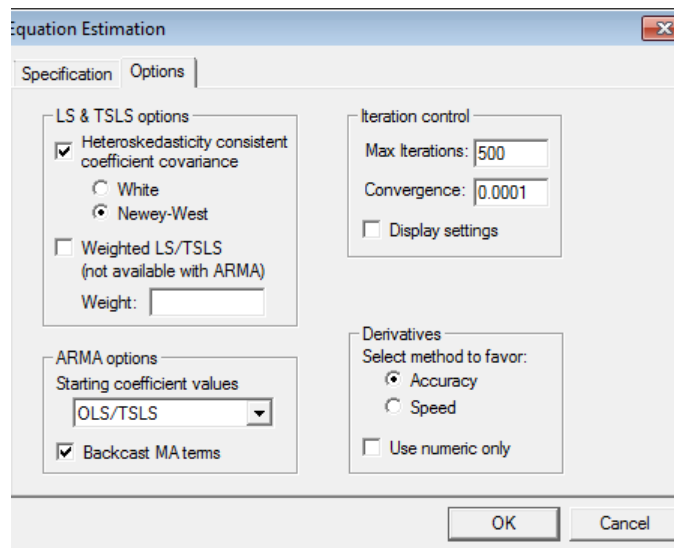
Ipoteza nulă a **testului Breusch – Godfrey** specifică faptul că inovațiile nu sunt autocorelate. Statistica acestui test presupune construirea a unei ecuație de regresie “auxiliară”: reziduurile ecuației principale sunt regresate în funcție de variabilele explicative și de “lag-uri” ale reziduurilor. De asemenea, **Eviews** calculează p-value asociat acestui test.

În condițiile existenței autocorelării, **estimatorul OLS** este, în continuare, **consistent, distribuit asimptotic normal**, însă **erorile standard** calculate pe baza formulei clasice **nu sunt corecte** și, astfel, **testul t** bazat pe aceste erori standard **este eronat**.

Există **două soluții** în ceea ce privește **“problema autocorelării”**

- păstrăm estimatorul OLS, dar “corectăm” erorile standard;
- înlocuim estimatorul OLS cu un nou estimator.

Prima soluție este mult mai des utilizată în practică, deoarece nu presupune că forma autocorelării este cunoscută. Estimatorul White pentru matricea de varianță-covarianță analizat mai sus presupune că reziduurile ecuației estimate nu sunt autocorelate. Newey și West (1987) au propus un estimator robust al erorilor standard mai general care este coerent atât în prezența heteroskedasticității, cât și a autocorelării de formă necunoscută a inovațiilor. Pentru a utiliza în **Eviews** **estimatorul Newey-West pentru erorile standard („robust standard errors” de tip Newey-West)**, se selectează fila **Options** din **Equation Estimation**. Se apasă caseta **Heteroskedasticity Consistent Covariance** și se selectează opțiunea **Newey-West**.



Cea de a doua soluție se utilizează în cazul în care se presupune că în ceea ce privește inovațiile modelului liniar de regresie există *autocorelare sub o formă cunoscută*. În această situație, foarte rar întâlnită în practică, există posibilitatea să se utilizeze metoda generalizată a celor mai mici pătrate (FGLS). De exemplu, **Eviews** poate utiliza **estimatorul FGLS** pentru cazul în care reziduurile ecuației de regresie au o dinamică descrisă de un proces AR(1), un tip de proces analizat pe larg în Capitolul 6.

În ceea ce privește tehnicile de **estimare robustă a erorilor standard**, mai trebuie reținut faptul că:

- utilizarea acestor estimări robuste pentru matricea de varianță-covarianță nu modifică estimările parametrilor.
- nu există nici un impediment în combinarea diferitelor metode de contabilizare a heteroskedasticității sau a corelației seriale din cadrul inovațiilor. De exemplu, estimatorul WLS poate fi însoțit de estimări robuste pentru erorile standard.

5 Abateri de la ipotezele modelului clasic de regresie II

O ipoteză fundamentală a modelului clasic de regresie este aceea potrivit căreia variabilele din partea dreaptă (i.e. variabilele explicative) sunt necorelate cu inovațiile modelului. Pentru simplitate, se va face referire la variabilele care sunt corelate cu reziduurile, ca variabile **endogene** și variabile care nu sunt corelate cu reziduurile, ca **exogene**. Dacă această ipoteză este încălcată, apare așa numita **„problemă de endogenitate”**, care este o problemă majoră întâlnită în practică. Astfel, în condițiile existenței endogenității, **estimatorul OLS nu este „bun”**, în sensul că **nu este consistent**.

Există o serie de situații binecunoscute în care apare problema de endogenitate:

- din cadrul ecuației de regresie au fost omise variabile relevante;
- variabile explicative sunt măsurate cu erori;
- există variabile determinate endogen în partea dreaptă a ecuației ca în cazul modelelor cu ecuații simultane.

Există **două soluții** în ceea ce privește **“problema de endogenitate”**

- „corectam” endogenitatea astfel încât estimatorul OLS să redevină „bun”;
- înlocuim estimatorul OLS cu un nou estimator .

Prima soluție poate fi utilizată în cazul în care unul din factorii explicativi este neobservabil, însă se poate utiliza o **variabilă „proxy”** pentru a aproxima valoarea acestuia. În practică, însă, este mult mai des întâlnită a doua variantă.

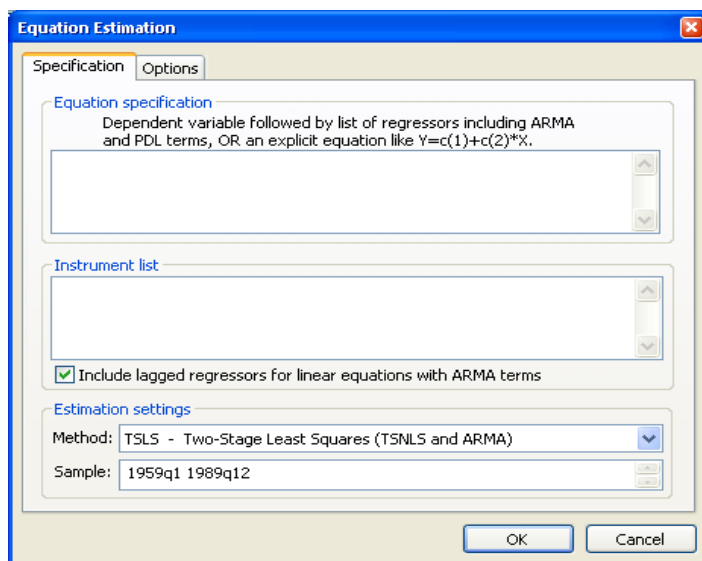
5.1 Estimatorul TSLS

Abordarea standard în cazurile în care variabilele din partea dreaptă sunt corelate cu reziduurile este să se estimeze ecuația de regresie utilizând **metoda variabilelor instrumentale**.

Ideea din spatele metodei variabilelor instrumentale este de a găsi un set de variabile, numite instrumente sau **variabile instrumentale**, care îndeplinesc următoarele condiții:

- sunt **relevante**, adică sunt corelate cu variabilele explicative din ecuație
- sunt **exogene**, adică sunt necorelate cu erorile.

Aceste variabile instrumentale sunt utilizate pentru a elimina corelația dintre variabilele din partea dreaptă și inovațiile ecuației de regresie.



Metoda **Two-stage least squares (TSLS)** reprezintă un caz special al metodei variabilelor instrumentale. După cum sugerează și numele, există două etape distincte în cadrul TSLS. Prima etapă implică estimarea unei regresii OLS pentru fiecare variabilă din model în funcție de setul de variabile instrumente. A

doua etapă reprezintă o regresie a ecuației originale, cu toate variabilele înlocuite cu valorile rezultate din regresiiile din prima etapă. Coeficienții acestei regresii sunt estimatorii TSLS ai parametrilor modelului de regresie.

Pentru a determina estimatorul TSLS, în programul **Eviews**, se deschide caseta de specificare a ecuației prin alegerea opțiunii **Object/New Object.../Equation...** sau **Quick/Estimate Equation...** Se alege TSLS din lista mobilă **Method**, iar caseta de dialog se va modifica pentru a include o casetă **Instrument list** în care trebuie introdusă lista instrumentelor.

Este demn de menționat că există o serie de **reguli în ceea ce privește instrumentele**:

- pentru a calcula estimatorul TSLS, trebuie să existe cel puțin la fel de multe instrumente ca și coeficienții din ecuație;
- orice variabilă explicativă care este exogenă, adică nu este corelată cu variabila reziduală, ar trebui să fie inclusă ca variabilă instrumentală;
- constanta este întotdeauna un instrument adecvat, astfel încât **Eviews** o va adăuga la lista de instrumente.

5.2 Modele cu ecuații simultane

Există o serie de situații în care există simultaneitate între două variabile economice. De exemplu, în cazul cantității de echilibru și al prețului de echilibru de pe o piață oarecare. Teoria economică presupune faptul că echilibrul pe piață se stabilește la intersecția dintre funcția de cerere și cea de ofertă. Funcția de cerere presupune că există o relație inversă între cantitatea cerută și preț, iar funcția de ofertă presupune că există o relație directă între cantitatea oferită și preț. Însă, nu se pot observa cantitatea cerută și cantitatea oferită, ci numai cantitatea de echilibru de pe piață. Astfel, cantitatea și prețul de echilibru „verifică” atât ecuația cererii cât și ecuația ofertei, formându-se, astfel, un **model cu ecuații simultane**.

Un model cu ecuații simultane prezintă o **formă structurală**, cea presupusă de teoria economică, în care există simultaneitate între cele două variabile de

interes, y_1 și y_2 , fiecare din cele două variabile fiind la rândul lor influențate de factori exogeni, z_1 și, respectiv z_2 :

$$\begin{aligned} y_1 &= \alpha_1 y_2 + \beta_1 z_1 + u_1 \\ y_2 &= \alpha_2 y_1 + \beta_2 z_2 + u_2 \end{aligned}$$

unde u_1 și u_2 reprezintă reziduurile structurale ale modelului.

Prin rezolvarea efectivă a sistemului de ecuații simultane, se determină **forma redusă** a modelului, fiecare variabilă de interes, y_1 și y_2 , fiind exprimată numai în funcție factorii exogeni, z_1 și z_2 . De exemplu, prin rezolvarea sistemului de mai sus rezultă forma redusă pentru variabila y_2 :

$$y_2 = \pi_{21} z_1 + \pi_{22} z_2 + v_2$$

unde parametrii π_{21} și π_{22} sunt o combinație (funcție) de parametrii structurali (i.e. α_1 , α_2 , β_1 și β_2), iar variabila reziduală în formă redusă, v_2 este o combinație (funcție) de reziduurile structurale ale modelului.

Simultaneitatea dintre variabilele y_1 și y_2 induce endogenitate datorită faptului că există corelație între aceste variabile și reziduurile structurale ale modelului. Mai exact, deoarece se poate arăta că $COV(y_2, u_1) \neq 0$, prima ecuație din cadrul sistemului de ecuații simultane nu poate fi estimat, în mod consistent, prin estimatorul OLS. Datorită endogenității indusă de către simultaneitate, un model cu ecuații simultane trebuie să fie estimat prin intermediul estimatorului TSLS.

Un exemplu clasic în acest sens în constituie sistemul de ecuații simultane care apare în cazul analizei preț-cantitate pe o piață oarecare. Teoria economică presupune existența a două ecuații structurale, respectiv ecuația asociată funcției de cerere și cea asociată funcției de ofertă.

Având în vedere că se pot observa doar cantitățile și prețurile de echilibru, forma structurală a sistemului este dată de

$$\begin{cases} q_t = \alpha_1 p_t + \beta_1 v_t + \varepsilon_t^d \\ q_t = \alpha_2 p_t + \varepsilon_t^s \end{cases}$$

Prin rezolvare efectivă a sistemului se obține forma redusă a acestuia:

$$\begin{cases} p_t = \frac{\beta_1}{\alpha_2 - \alpha_1} v_t + \frac{\varepsilon_t^d - \varepsilon_t^s}{\alpha_2 - \alpha_1} \\ q_t = \frac{\beta_1 \alpha_2}{\alpha_2 - \alpha_1} v_t + \frac{\alpha_2 \varepsilon_t^d - \alpha_1 \varepsilon_t^s}{\alpha_2 - \alpha_1} \end{cases}$$

Se poate arăta faptul că $E[p_t \varepsilon_t^s] = COV(p_t, \varepsilon_t^s) = -\frac{\sigma_s^2}{\alpha_2 - \alpha_1} \neq 0$ sau altfel

spus variabila p este „endogenă” în cadrul ecuației „de ofertă”, $q_t = \alpha_2 p_t + \varepsilon_t^s$.

Astfel, dacă se încercă estimarea parametrului structural α_2 prin estimatorul OLS rezultatele obținute nu sunt corecte deoarece, în această situație, estimatorul OLS

nu este consistent. Mai exact, se poate arăta că $\alpha_2^{OLS} \xrightarrow{p} \alpha_2 - \frac{1}{\alpha_2 - \alpha_1} \frac{\sigma_s^2}{\sigma_p^2}$.

În mod similar, variabila p este, de asemenea, „endogenă” în cadrul ecuației

„de cerere”, $q_t = \alpha_1 p_t + \beta_1 v_t + \varepsilon_t^d$, deoarece $E[p_t \varepsilon_t^d] = COV(p_t, \varepsilon_t^d) = \frac{\sigma_d^2}{\alpha_2 - \alpha_1} \neq 0$.

Ca urmare, estimarea parametrului structural α_1 prin OLS nu este consistentă.

Se pune problema identificării unei variabile instrumentale care să ne permită estimarea consistentă a cel puțin unuia dintre parametrii structurali α_1 și α_2 . Deoarece, variabila v apare în ecuația structurală a cererii, dar nu și în ecuația structurală a ofertei vom investiga dacă aceasta este o variabilă instrumentală „bună” pentru variabila p în cadrul ecuației ofertei. Într-adevăr, această variabilă este

- *relevantă*, deoarece $E[v_t p_t] = COV(v_t, p_t) = \frac{\beta_1}{\alpha_2 \alpha_1} E[v^2] \neq 0$
- *exogenă*, deoarece $E[v_t \varepsilon_t^s] = COV[v_t, \varepsilon_t^s] = 0$

Ca urmare, parametrul structural α_2 poate fi estimat în mod consistent utilizând estimatorul TSLS în care drept variabilă instrumentală pentru variabila p utilizăm variabila v .

6 Tehnici de modelare și previzionare a seriilor de timp unidimensionale

Capitolul de față se axează pe specificarea și estimarea modelelor cu care se pot analiza dinamica seriilor de timp unidimensionale. **Eviews** calculează diverse statistici pentru o serie de timp și afișează aceste statistici, în diverse forme, cum ar fi sub formă de foaie de calcul, tabel sau grafic. Acestea variază de la afișarea graficului seriei respective, până la estimatori ne-parametrici pentru densitatea seriei respective, bazați pe funcții de tip „kernel”. De asemenea, **Eviews** dispune de o serie de proceduri care pot fi utilizate pentru a modifica și a analiza seria respectivă, cum ar fi diferite metode de ajustare sezonieră, metode de nivelare exponențială, precum și tehnici de filtrare statistică, dintre care cel mai utilizat în practică este filtrul Hodrick- Prescott.

6.1 Noțiuni introductive

Unul din principalele concepte cu care operează econometria seriilor de timp este cel de **proces stocastic** (aleator). Un proces stocastic este un șir de variabile aleatoare. Astfel, un proces stocastic este similar unui șir de numere reale, cu mențiunea că elementele unui proces stocastic nu sunt numere, ci variabile aleatoare. Intuitiv, un proces aleator poate fi reprezentat prin imaginea unui tren, care trece prin fața unei gări în care se află un observator. Fiecare vagon este numerotat și conține o persoană. Cel care stă în gară observă la un anumit moment un singur vagon, luând astfel la cunoștință de cine este la geam. Observatorul nu știe cine este la geamul următorului vagon, dar știe cine poate fi și cu ce probabilitate (variabilă aleatoare). Știe, de asemenea, cine a fost în vagoanele care au trecut prin fața sa până în prezent. Acestea formează o observație din procesul aleator. O **serie de timp** este, de fapt, o realizare a unui proces stocastic. O serie de

timp este formată din câte o observație din fiecare variabilă aleatoare care compune procesul stocastic ce a generat datele.

Spre deosebire de alte concepte care implică eșantionarea de variabile aleatoare, dintr-un proces stocastic putem avea, cel mult, o observație (înregistrare, realizare). De ce? Pentru că, de exemplu, pentru variabila aleatoare prețul de închidere de astăzi putem avea o singură observație, pentru variabila aleatoare prețul de închidere de mâine la fel etc. Nu ne putem întoarce în timp pentru a putea observa din nou prețul de închidere de astăzi. Această caracteristică a proceselor stocastice complică analiza acestora, deoarece, fără a impune ipoteze suplimentare, o serie de timp se reduce la o colecție care presupune câte o observație din mai multe variabile aleatoare.

Putem surmonta această dificultate presupunând că toate variabilele aleatoare care formează procesul stocastic sunt identice. În această ipoteză, a avea câte o observație din fiecare variabilă din procesul stocastic ce a generat datele este echivalent cu a avea mai multe observații din aceeași variabilă aleatoare. În concluzie, pentru a obține rezultate relevante în legătură cu o serie de timp, este necesar ca elementele sale să fie observații ale unor variabile aleatoare identice. Cu alte cuvinte, analiza unei serii de timp este validă dacă și numai dacă procesul stocastic care a generat datele își păstrează proprietățile statistice în timp.

Pe mulțimea tuturor proceselor aleatoare de poate defini următorul operator, notat cu L , după recula $Lx_t = x_{t-1}$. Operatorul L , numit **operator de lag**, transformă șirul x_t într-un alt șir, cu aceleași elemente, dar decalate înapoi cu o perioadă (lag). Aplicând succesiv operatorul de lag obținem șiruri în care valorile lui x_t sunt decalate cu 2 (L^2) sau cu k (L^k) perioade: $L^2x_t = L(Lx_t) = Lx_{t-1} = x_{t-2}$, respectiv $L^kx_t = x_{t-k}$. Un caz special, apare pentru $k = 0$: $L^0x_t = x_t$. Combinând compunerea operatorului de lag cu operațiile obișnuite cu polinoame, obținem **polinoame de lag**. De exemplu, $B(L) = 1 - L$ poate fi „aplicat” unui proces x_t și rezultă $B(L)x_t = (1 - L)x_t = x_t - x_{t-1}$. Polinoamele de lag se pot utiliza pentru

deducerea formulei generale a unui proces stocastic, pornind de la definirea acestora prin relația de dinamică.

6.2 Conceptul de staționaritate

După cum am arătat în introducerea referitoare la procesele stocastice, analiza seriilor de timp se poate face numai cu impunerea anumitor ipoteze asupra procesului care a generat datele. În mare, aceste ipoteze ne asigură că datele înregistrate (observațiile statistice) provin din același proces și că acesta nu își modifică proprietățile statistice în timp. Această proprietate se numește staționaritate.

Seriile staționare sunt, în general, seriile definite ca rate de creștere: rata de creștere a PIB, rata inflației, rata de creștere a salariului mediu net, rentabilitățile activelor financiare, dar și serii de tipul ratelor de dobândă pe termen scurt sau al variației stocurilor. **Serii nestaționare** sunt, în general, seriile care reflectă indici cu bază fixă: indicele prețurilor de consum, indicele prețurilor producției industriale, PIB real și nominale, indicele producției industriale, indici bursieri.

Vizual, seriile staționare sunt cele care evoluează în jurul unei medii, în timp ce seriile nestaționare nu prezintă acest comportament de revenire la o valoare medie. Într-o definiție mai riguroasă, **conceptul de staționaritate** comportă două înțelesuri. În sens strict, un proces stocastic este staționar dacă momentele sale (medie, varianță, momentul de ordinul trei, patru etc.) nu depinde de timp. În sens slab, **conceptul de staționaritate în medie-covarianță** se referă la un proces stocastic pentru care primele sale două momente (media și autocovarianța) nu depind de timp.

Identificarea caracterului staționar sau nestaționar al seriilor de timp cu care lucrăm este un pas esențial în analiză. Nestaționaritatea seriilor de timp are ca rezultat:

- invalidarea unor rezultate referitoare la testele statistice sau la proprietățile datelor obținute în ipoteza că datele sunt staționare sau

- apariția de regresii eronate (en. *spurious regression*).

Regresiile eronate se referă la identificarea unor relații inexistente între serii de date fără nicio legătură, aceste relații fiind puse în evidență prin regresii simple și fiind consecința nestaționarității datelor. Printre cele mai celebre exemple de regresii eronate se numără:

- relația dintre numărul de nou-născuți și numărul de cuiburi de berze în Olanda sau cea dintre
- populația Africii de Sud și cheltuielile de cercetare-dezvoltare în SUA, utilizând date anuale pentru perioada 1971-1990.

Stabilirea caracterului staționar al unei serii de timp se realizează prin aplicarea **testelor de rădăcină unitate** (en. *unit-root*). Cele mai utilizate astfel de teste sunt **testul ADF** (Augmented Dickey-Fuller) și **testul PP** (Phillips-Perron).

Înainte de a trece la descrierea testelor ADF și PP, vom prezenta legătura dintre prezența rădăcinii unitate în dinamica seriei de timp și nestaționaritate. Rădăcina unitate se referă la prezența printre rădăcinile polinomului de lag $B(L)$ a unei rădăcini egale cu 1. După cum vom arăta în continuare, în acest caz seria este nestaționară. Să presupunem un proces de forma $x_t = \beta x_{t-1} + \varepsilon_t$, cu ε_t *white-noise*. Utilizând polinomul de lag, acesta se poate scrie ca $B(L)x_t = \varepsilon_t$, unde $B(L) = 1 - \beta L$. Dacă polinomul $B(L)$ are o rădăcină unitară, $B(1) = 0$, ceea ce implică $\beta = 1$. În acest caz, procesul este de forma $x_t = x_{t-1} + \varepsilon_t$. Media procesului respectă relația $E[x_t] = E[x_{t-1}]$, iar varianța $Var[x_t] = Var[x_{t-1}] + Var[\varepsilon_t]$. Cele două relații au fost obținute utilizând faptul că, prin definiție, procesul ε_t are media zero: $E[\varepsilon_t] = 0$ și nu este corelat cu x_{t-1} : $Corr[x_{t-1}, \varepsilon_t] = 0$. Iterând înapoi cele două relații pentru medie și varianță, obținem $E[x_t] = E[x_{t-1}] = \dots = E[x_0]$ și $Var[x_t] = Var[x_{t-1}] + \sigma^2 = Var[x_{t-2}] + \sigma^2 + \sigma^2 = \dots = Var[x_0] + t\sigma^2$. Rezultă că deși media este invariantă în timp, $E[x_t] = E[x_0]$, pentru orice t , varianța procesului depinde de timp. Prin urmare procesul este nestaționar.

Testele ADF și PP au ca ipoteză nulă faptul că seria analizată este nestaționară.

Testul ADF consideră un proces autoregresiv $AR(1)$, de forma $x_t = \mu + \varphi x_{t-1} + \varepsilon_t$, unde ε_t este de tip zgomot alb. Dacă $|\varphi| < 1$ seria x_t este staționară; dacă $\mu = 1$ seria x_t este nestaționară, iar dacă $|\varphi| > 1$ seria x_t este explozivă. În varianta fără constantă și fără trend, testul ADF se bazează pe regresia:

$$\Delta x_t = \gamma \cdot x_{t-1} + \gamma_1 \cdot \Delta x_{t-1} + \dots + \gamma_p \cdot \Delta x_{t-p} + \varepsilon_t.$$

Ipoteza nulă este $H_0 : \gamma = 0$, cu alternativa $H_1 : \gamma < 0$; $\gamma = \varphi - 1$. La testarea ipotezei nule se utilizează testul t , cu simularea valorilor critice.

Distribuția asimptotică a testului ADF este non-standard, fiind independentă de numărul de laguri, p . Caracterul non-standard al acestei distribuții impune simularea valorilor critice, care în acest caz sunt -2,57 pentru 1%; -1,94 pentru 5% și -1,62 pentru 10%. Interpretarea testului ADF se bazează pe compararea valorii calculate a testului t pentru coeficientul ρ , obținut din regresie, cu valorile critice menționate anterior. Dacă t-Statistic calculat este mai mic decât valoarea critică, se respinge ipoteza nulă, seria este staționară, iar dacă t-Statistic calculat este mai mare decât valorile critice, nu se poate respinge ipoteza nulă, seria este nestaționară.

Testul PP testează regresia $\Delta x_t = \alpha + \beta x_{t-1} + \varepsilon_t$. Distribuția t a coeficientului γ este corectată pentru a include corelațiile seriale ale ε_t .

Există și teste de staționaritate care iau staționaritatea ca ipoteză nulă. Cel mai utilizat dintre acestea este **testul KPSS** (Kwiatkowski-Phillips-Schmidt-Shin). Testul KPSS pornește de la specificația $y_t = x_t + z_t$, unde x_t este de tip *random walk*: $x_t = x_{t-1} + u_t, u_t \sim iidN(0, \sigma_u^2)$, iar z_t este staționar.

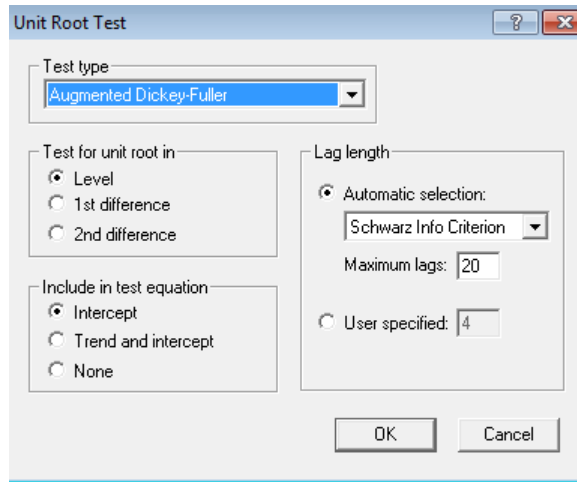
Testul KPSS este complementar testelor ADF și PP în sensul că ipoteza sa nulă este că seria este staționară, cu alternativa nestaționarității: $H_0 : \sigma_u^2 = 0 \Rightarrow y_t$ este staționară, $H_1 : \sigma_u^2 > 0 \Rightarrow y_t$ este nestaționară.

Testul KPSS are, la rândul său, o distribuție asimptotică non-standard, valorile critice pentru varianta cu constantă și fără trend fiind 0,74 pentru 1%, 0,46 pentru 5% și 0,35 pentru 10%. O valoarea a testului calculate mai mare decât aceste valori critice indică respingerea ipotezei nule.

Complementaritatea celor două tipuri de teste face posibilă utilizarea combinată, pentru a stabili staționaritatea unei serii de timp. Rezultatul combinat al testelor ADF și PP, respectiv KPSS este următorul:

- respingerea ipotezei nule ADF și imposibilitatea respingerii ipotezei nule KPSS indică un proces staționar;
- imposibilitatea respingerii ipotezei nule ADF și respingerea ipotezei nule KPSS indică o serie nestaționară;
- imposibilitatea respingerii atât a ipotezei nule ADF, cât și a celei KPSS indică insuficiența informațională a datelor analizate;
- respingerea atât a ipotezei nule ADF, cât și a celei KPSS indică necesitatea unei reprezentări alternative, printr-un proces mai complex, care poartă denumirea de integrat fracționar.

Pentru a se efectua un test de rădăcină unitate în **Eviews**, se selectează opțiunea **View – Unit Root Test...** In cadrul casetei de dialog de dialog **Unit Root Test** se selectează tipul testului care se dorește efectuat. Trebuie specificat dacă se dorește testarea pentru o rădăcină unitate a seriei, sau a prima diferență, respectiv a celei de a doua diferență a seriei. Setări avansate puse la dispoziție în cazul testului ADF permit specificarea modului în care lag-urile termenilor diferențiați vor fi incluși în ecuația de testare ADF. Se poate opta pentru ca **Eviews** să facă selecția în mod automat sau se poate specifica un număr fix, întreg și pozitiv.



Eviews calculează p-value pentru testul ADF, ceea ce face relativ simplă interpretarea rezultatelor acestui test.

În funcție de tipul de nestăționaritate al seriei de timp analizate, problema nestăționarității datelor se poate rezolva prin transformarea datelor astfel încât să se obțină serii staționare:

- diferențiere;
- eliminarea trendului: liniar, pătratic, HP.

Fiind dat procesul stocastic x_t , diferențierea acestuia produce procesul $\Delta x_t = x_t - x_{t-1} = (1 - L)x_t$.

Eliminarea trendului liniar constă în rularea regresiei $x_t = c + \alpha \cdot t + \varepsilon_t$ și extragerea reziduurilor estimate, $\hat{\varepsilon}_t$.

Trendul pătratic se elimină din regresia $x_t = c + \alpha \cdot t + \beta \cdot t^2 + \varepsilon_t$.

Filtrul Hodrick – Prescott (HP) este o metodă parametrică de eliminare a trend-ului, fiind cel mai popular instrument pentru a descompune o serie de timp într-un trend și un ciclu. Filtrul HP elimină trendul prin minimizarea unei funcții de pierdere care are următoarea formă:

$$\min_{\{\bar{y}_t\}} \left[\sum_{t=1}^T c^2_t + \lambda \sum_{t=1}^T ((\bar{y}_{t+1} - \bar{y}_t) - (\bar{y}_t - \bar{y}_{t-1})) \right]^2$$

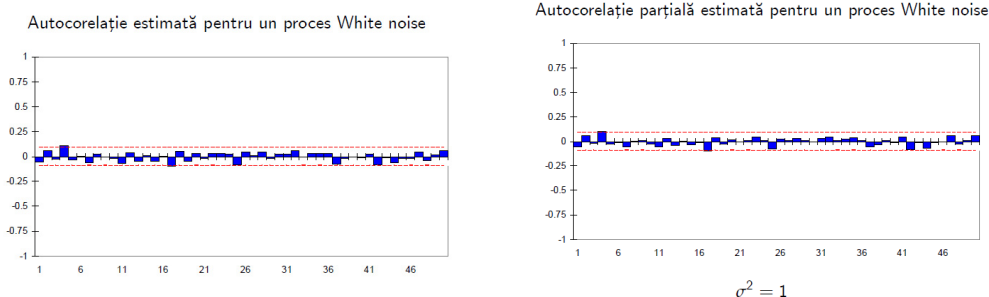
unde T reprezintă numărul total de observații. Termenul $\sum_{t=1}^T c^2_t$ este echivalent cu o sumă de pătrate de erori, deoarece $\sum_{t=1}^T c^2_t = \sum_{t=1}^T (y_t - \bar{y}_t)^2$, iar termenul $\sum_{t=1}^T ((\bar{y}_{t+1} - \bar{y}_t) - (\bar{y}_t - \bar{y}_{t-1}))^2$ asigură faptul că seria trend-ului este netedă, iar parametrul λ din problema de optimizare penalizează fluctuațiile seriei de trend. Cu cât λ este mai mare, cu atât trend-ul astfel estimat devine mai neted, iar atunci când λ tinde la infinit, trend-ul devine o linie dreaptă.

Analiza seriilor de date după ce au fost staționarizate prin diferențiere sau prin eliminarea trendului are ca posibil dezavantaj pierderi din conținutul informațional al datelor, mai ales în ceea ce privește dinamica pe termen lung. Acest dezavantaj se poate surmonta prin analiza cu metode speciale, care țin cont de nestaționaritatea datelor, cum ar fi cele bazate pe conceptul de cointegrare.

6.3 Modele ARMA

Modelele **ARMA** (AutoRegressive Moving Average) sunt folosite pe scară largă pentru a analiza dinamica seriilor de timp unidimensionale. Un model ARMA este compus din două părți: termenul autoregresiv și termenul de medie mobilă. Procesele autoregresive au la bază ideea, naturală, conform căreia valoarea curentă a unei variabile depinde de valorile sale anterioare. Aceste procese se obțin adăugând la ecuații de recurență deterministe o componentă staționară, care este, de obicei, un proces staționar cu o structură de corelație extrem de simplă. Un astfel de proces este **procesul de tip zgomot alb** (*en.* white-noise), care la momentul t ia valoarea ε_t , unde ε_t este o variabilă aleatoare distribuită normal, de medie zero și varianță constantă, $\sigma^2 = E[\varepsilon_t^2]$. Simplitatea structurii de corelație a

acestui proces rezultă din faptul că valoarea la un moment t , ε_t este necorelată cu valoarea la oricare alt moment, $s \neq t$, ε_s . Autocovarianța pentru orice rang $k > 0$ pentru un proces *white-noise* este zero. Autocorelația pentru orice rang $k > 0$ pentru un proces *white-noise* este zero. Autocorelația parțială pentru orice rang $k > 0$ pentru un proces *white-noise* este zero. Autocorelația parțială este o măsură a autocorelației directe dintre două valori ale unei serii de timp, excluzând corelația care rezultă prin efectul lagurilor intermediare.



Importanța procesului *white-noise* rezultă nu numai din folosirea sa ca bloc pentru construcția de procese autoregresive, ci și din faptul că reprezintă prototipul reziduului din regresii care implică serii de timp.

6.3.1 Procese de tip *AR(1)*

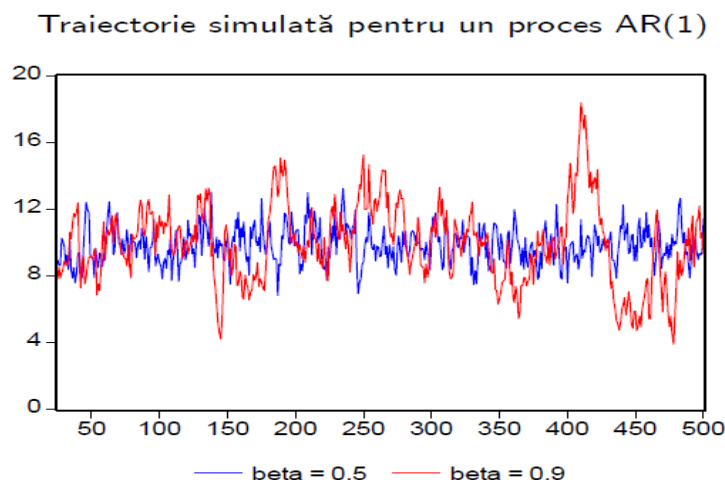
Un **proces de tip AR(1)** este un proces stocastic în care valoarea curentă depinde numai de valoarea anterioară:

$$x_t = \alpha + \beta \cdot x_{t-1} + \varepsilon_t,$$

unde ε_t este un proces de tip *white-noise*. Utilizând operatorul de lag introdus la începutul cursului, relația anterioară se poate scrie $(1-L)x_t = \alpha + \varepsilon_t$ sau $B(L)x_t = \varepsilon_t$, unde $B(L) = 1 - \beta L$ este un polinom de lag-uri.

În figura de mai jos este reprezentat un proces AR(1), pentru $\beta = 0,5$ și $\beta = 0,9$. Procesul simulat cu $\beta = 0,9$ manifestă un grad de persistență mai mare, abaterile sale de la medie fiind mai ample. Intuitiv, acest lucru se întâmplă

deoarece $\beta = 1$ corespunde unei serii nestaționare, iar cu cât β se apropie mai mult de 1, cu atât comportamentul său se apropie mai mult de cel al unei serii nestaționare.



Un interes deosebit îl prezintă determinarea teoretică a momentelor unui proces AR(1), deoarece acestea vor sta la baza identificării unei serii reale de date ca fiind de acest tip. Vom determina momentele unui proces AR(1) în ipoteza că acesta este staționar. Stabilirea condițiilor de staționaritate este un subiect mai complex, care nu va fi prezentat în acest suport de nivel mediu.

În ipoteza staționarității $E[x_t] = E[x_{t-1}]$ și ținând cont de faptul că ε_t este *white-noise*, prin urmare $E[\varepsilon_t] = 0$, aplicând operatorul de medie se obține

$$E[x_t] = \alpha + \beta E[x_{t-1}], \text{ cu notația } \mu = E[x_t]: \mu = \alpha + \beta \cdot \mu, \text{ de unde } \mu = \frac{\alpha}{1 - \beta},$$

relație care are sens numai pentru $\beta \neq 1$. În ipoteza staționarității varianța unui proces AR(1) este constantă, $\gamma_0 = Var[x_t] = E[(x_t - \mu)^2] = Var[x_{t-1}] = E[(x_{t-1} - \mu)^2]$, pentru orice t . Aplicând operatorul de varianță relației de definiție, obținem $\gamma_0 = \beta^2 \gamma_0 + \sigma^2$, deoarece varianța constantei α este zero, covarianțele dintre

aceasta și x_{t-1} și ε_t sunt zero, iar covarianța dintre x_{t-1} și ε_t este zero. Obținem

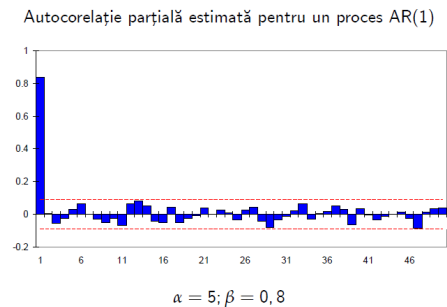
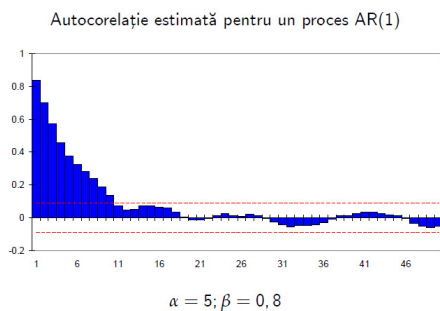
$\gamma_0 = \frac{\sigma^2}{1-\beta^2}$, relația care are sens numai pentru $|\beta| < 1$. În ceea ce privește

autocovarianța de rang k , $\gamma_k = Cov[x_t, x_{t-k}] = E[(x_t - \mu)(x_{t-k} - \mu)]$, obținem

$\gamma_k = \beta\gamma_{k-1}$. Autocorelația la rangul k se definește prin $\rho_k = \frac{\gamma_k}{\gamma_0}$. Ca urmare,

$$\rho_k = \beta^k.$$

Conform acestui rezultat, dacă $|\beta| < 1$ corelația dintre două valori aflate la distanța k scade și se apropie de zero, pe măsură ce k crește. Un proces staționar se caracterizează tocmai prin lipsa corelației între două valori aflate la distanță mare în timp. Autocorelația parțială pentru un proces AR(1) este $a_1 = \beta$ și $a_k = 0$ pentru $k > 1$, ceea ce se explică prin faptul că, prin definiție, nu există nicio legătură directă între x_{t-2} , x_{t-3} , ... etc și x_t .



Remarcăm că autocorelația scade treptat către zero, în timp ce autocorelația parțială este diferită de zero numai la primul lag, după care scade brusc la zero.

Vom demonstra în continuare că $|\beta| < 1$ implică staționaritatea procesului AR(1). Pentru aceasta, vom face apel la conceptul de polinom de lag. Să

presupunem că există polinomul de lag $\Phi(L)$, astfel încât $B(L)\Phi(L)=1$. În acest caz, $\Phi(L)$ ar fi inversul lui $B(L)$ și am putea scrie $\Phi(L)=B(L)^{-1}$. Dacă există, polinomul $B(L)^{-1}$ ne oferă posibilitatea de a rezolva ecuația prin care se definește procesul AR(1), $B(L)x_t = \alpha + \varepsilon_t$, înmulțind la stânga cu $B(L)^{-1}$. Obținem $x_t = B(L)^{-1}\alpha + B(L)^{-1}\varepsilon_t$. Deoarece pentru o serie constantă egală cu α , $L\alpha = \alpha$, soluția este $x_t = B(1)^{-1}\alpha + B(L)^{-1}\varepsilon_t$.

Pentru a stabili în ce condiții există $\Phi(L)=B(L)^{-1}$, luăm $\Phi(L)=\phi_0 + \phi_1L + \phi_2L^2 + \dots + \phi_kL^k + \dots$ și determinăm coeficienții $\phi_0, \phi_1, \dots, \phi_k, \dots$ astfel încât $B(L)\Phi(L)=1$. Observăm, mai întâi că $\Phi(L)$ este un polinom cu o infinitate de termeni. Bineînțeles că și $B(L)$ se poate scrie ca un polinom cu o infinitate de termeni: $B(L)=1 - \beta L + 0 \cdot L^2 + \dots + 0 \cdot L^k + \dots$. Atât $B(L)$, cât și $\Phi(L)$ fac parte din mulțimea polinoamelor de lag cu o infinitate de termeni.

$B(L)\Phi(L)=1$ este echivalent cu $(1 - \beta L)(\phi_0 + \phi_1L + \phi_2L^2 + \dots + \phi_kL^k + \dots)=1$, sau $\phi_0 + (-\beta\phi_0 + \phi_1)L + (-\beta\phi_1 + \phi_2)L^2 + \dots + (-\beta\phi_{k-1} + \phi_k)L^k + \dots=1$, de unde rezultă $\phi_0=1$, $\phi_1=\beta\phi_0=\beta$, $\phi_2=\beta\phi_1=\beta^2$, în general $\phi_k=\beta^k$. Polinomul de lag invers al lui $B(L)=1 - \beta L$ este $B(L)^{-1}=1 + \beta L + \beta^2L^2 + \dots + \beta^kL^k + \dots$. Scris compact $B(L)^{-1} = \sum_{k=0}^{\infty} \beta^k L^k$.

Soluția ecuației AR(1) este

$$x_t = B(1)^{-1}\alpha + \left(\sum_{k=0}^{\infty} \beta^k L^k \right) \varepsilon_t = B(1)^{-1}\alpha + \sum_{k=0}^{\infty} \beta^k \varepsilon_{t-k}.$$

Sub această formă, numită scriere MA(∞), procesul AR(1) apare ca o sumă infinită de termeni de tip *white-noise*. În ce condiții această sumă este bine definită? Aplicând operatorul de medie obținem $E[x_t] = \frac{1}{1-\beta}\alpha + \sum_{k=0}^{\infty} \beta^k E[\varepsilon_{t-k}]$.

Deoarece ε_t este staționar, prin definiție $E[\varepsilon_{t-k}] = 0$. Rezultă

$$E[x_t] = \frac{1}{1-\beta} \alpha + \left(\sum_{k=0}^{\infty} \beta^k \right) \cdot 0. \text{ Al doilea termen al mediei lui } x_t \text{ este determinat}$$

dacă și numai dacă suma $\sum_{k=0}^{\infty} \beta^k$ este finită. Altfel, se obține o nedeterminare de

forma $\infty \cdot 0$. Prin urmare, $E[x_t]$ este bine-definită dacă și numai dacă $\sum_{k=0}^{\infty} \beta^k$ este

finită, ceea ce este echivalent cu $|\beta| < 1$. În acest caz, $\sum_{k=0}^{\infty} \beta^k = \frac{1}{1-\beta}$ și $x_t = \frac{1}{1-\beta} \alpha$

și nu depinde de timp. Am obținut același rezultat ca mai sus, dar fără a presupune că procesul este staționar.

Varianța lui x_t este dată de relația

$$Var[x_t] = Var\left(\sum_{k=0}^{\infty} \beta^k \varepsilon_{t-k} \right) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \beta^k \beta^l Cov[\varepsilon_{t-k}, \varepsilon_{t-l}].$$

Deoarece, prin definiție $Cov[\varepsilon_t, \varepsilon_s] = 0$ pentru $t \neq s$ și $Cov[\varepsilon_t, \varepsilon_s] = \sigma^2$ pentru $\beta^k \beta^l Cov[\varepsilon_{t-k}, \varepsilon_{t-l}] = 0$ pentru $k \neq l$ și $\beta^k \beta^l Cov[\varepsilon_{t-k}, \varepsilon_{t-l}] = \beta^{2k} \sigma^2$ pentru

$k = l$. Rezultă $Var[x_t] = \left(\sum_{k=0}^{\infty} (\beta^2)^k \right) \sigma^2$, bine-definită numai în cazul $\beta^2 < 1$,

echivalent cu $\beta < 1$. În acest caz $Var[x_t] = \frac{1}{1-\beta^2} \sigma^2$, același rezultat pe care l-am

obținut anterior în privința varianței lui x_t , numai că de data aceasta nu am presupus că procesul este staționar.

Am obținut așadar, că $|\beta| < 1$ implică faptul că media (μ) și varianța (γ_0) procesului x_t nu depind de timp. Putem adăuga că pentru un astfel de proces covarianța este $\gamma_k = \beta^k \gamma_0$, fiind la rândul său invariantă în timp. Un proces pentru care $|\beta| < 1$ se numește stabil. În concluzie, un proces AR(1) care este stabil este și staționar. Deși reciproca nu este întotdeauna adevărată, în sensul că există procese

staționare care nu sunt stabile, vom înțelege prin condiție de staționaritate a unui proces AR(1), condiția $|\beta| < 1$.

În cele ce urmează vom oferi detaliile de calcul pentru funcția de autocorelație parțială. Prin definiție, autocorelația parțială la lagul k este $a_k = Corr[x_t, x_{t-k} | x_{t-1}, x_{t-2}, \dots, x_{t-k+1}]$. a_k este egal cu coeficientul \hat{a}_{kk} estimat din regresia $x_t = a_{k1}x_{t-1} + a_{k2}x_{t-2} + \dots + a_{kk}x_{t-k} + u_{kt}$. De exemplu, pentru un proces AR(1), autocorelația parțială la lagul 2 se obține din regresia $x_t = a_{21}x_{t-1} + a_{22}x_{t-2} + u_{2t}$. Analitic, coeficientul $a_2 = a_{22}$ se poate obține din sistemul

$$\begin{cases} \rho_1 = a_{21} + a_{22}\rho_1 \\ \rho_2 = a_{21}\rho_1 + a_{22} \end{cases},$$

rezultat aplicând succesiv coeficientul de corelație cu x_{t-1} , respectiv x_{t-2} . Pentru un proces AR(1), $\rho_1 = \beta$ și $\rho_2 = \beta^2$. Sistemul anterior are forma

$$\begin{cases} \beta = a_{21} + a_{22}\beta \\ \beta^2 = a_{21}\beta + a_{22} \end{cases}, \text{ ceea ce este echivalent cu } \begin{pmatrix} 1 & \beta \\ \beta & 1 \end{pmatrix} \begin{pmatrix} a_{21} \\ a_{22} \end{pmatrix} = \begin{pmatrix} \beta \\ \beta^2 \end{pmatrix}. \text{ Soluția este}$$

$$\begin{pmatrix} a_{21} \\ a_{22} \end{pmatrix} = \begin{pmatrix} 1 & \beta \\ \beta & 1 \end{pmatrix}^{-1} \begin{pmatrix} \beta \\ \beta^2 \end{pmatrix}. \text{ Ținând cont că } \begin{pmatrix} 1 & \beta \\ \beta & 1 \end{pmatrix}^{-1} = \frac{1}{1-\beta^2} \begin{pmatrix} 1 & -\beta \\ -\beta & 1 \end{pmatrix}, \text{ soluția}$$

este $a_{21} = \beta$ și $a_{22} = 0$.

În general, coeficientul $a_k = a_{kk}$ se obține analitic rezolvând sistemul

$$\begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} a_{k1} \\ a_{k2} \\ \vdots \\ a_{kk} \end{pmatrix}.$$

6.3.2 Procese de tip AR(2)

Un **proces de tip AR(2)** este un proces stocastic în care valoarea curentă depinde de cele mai recente două valori anterioare:

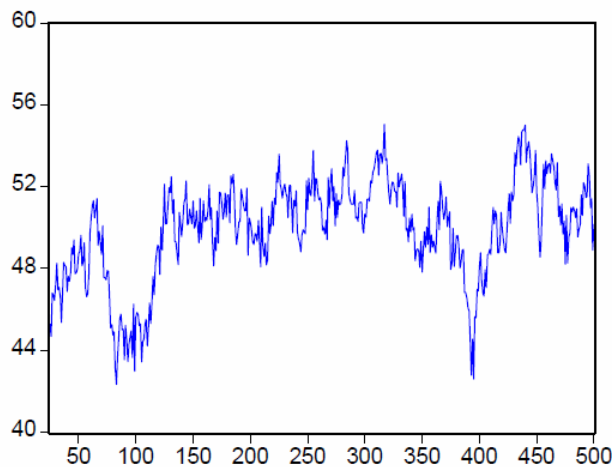
$$x_t = \alpha + \beta_1 \cdot x_{t-1} + \beta_2 \cdot x_{t-2} + \varepsilon_t,$$

unde ε_t este un proces de tip *white-noise*, cu media $E[\varepsilon_t]=0$ și varianța $Var[\varepsilon_t]=\sigma^2$.

Utilizând operatorul de lag, un proces AR(2) se scrie ca $B(L)x_t = \alpha + \varepsilon_t$, unde $B(L)=1-\beta_1L-\beta_2L^2$.

În figura de mai jos este reprezentat un proces AR(2), pentru $\beta_1 = 0,6$ și $\beta_2 = 0,3$.

Traietorie simulată pentru un proces AR(2)



$$\alpha = 5; \beta_1 = 0,6; \beta_2 = 0,3$$

Orice proces AR(2) se poate scrie matricial, ca un proces AR(1):

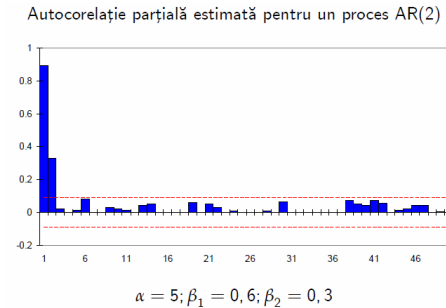
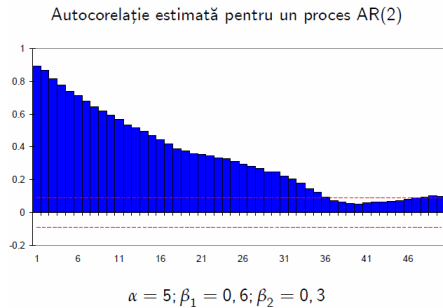
$$\begin{pmatrix} x_t \\ x_{t-1} \end{pmatrix} = \begin{pmatrix} \alpha \\ 0 \end{pmatrix} + \begin{pmatrix} \beta_1 & \beta_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ x_{t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}. \quad \text{Cu notațiile } y_t = \begin{pmatrix} x_t \\ x_{t-1} \end{pmatrix}, \quad a = \begin{pmatrix} \alpha \\ 0 \end{pmatrix},$$

$$b = \begin{pmatrix} \beta_1 & \beta_2 \\ 1 & 0 \end{pmatrix} \quad \text{și} \quad u_t = \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}, \quad \text{scrierea devine } y_t = a + by_{t-1} + u_t. \quad \text{Extrapolând}$$

rezultatul obținut în cazul procesului AR(1) la procesul AR(2), condiția de stabilitate este echivalentă cu $\lim_{k \rightarrow \infty} b^k = 0$. Convergența matricei b^k către zero este controlată de valorile proprii ale lui b . Astfel, $\lim_{k \rightarrow \infty} b^k = 0$ dacă și numai dacă $|\lambda_1|, |\lambda_2| < 1$, cu λ_1 și λ_2 valorile proprii ale matricei b . Acestea se obțin din sistemul $\lambda^2 - tr(b) \cdot \lambda + det(b) = 0$, unde $tr(b) = \beta_1 + 0 = \beta_1$, iar $det(b) = \beta_1 \cdot 0 - \beta_2 \cdot 1 = -\beta_2$. Condiția de staționaritate este ca soluțiile ecuației $\lambda^2 - \beta_1 \lambda - \beta_2 = 0$ să fie subunitare în modul.

Media unui proces AR(2) se obține aplicând operatorul de medie relației de definiție: $E[x_t] = \alpha + \beta_1 E[x_{t-1}] + \beta_2 E[x_{t-2}]$. Dacă procesul este staționar,

$$E[x_t] = E[x_{t-1}] = E[x_{t-2}] = \mu = \frac{\alpha}{1 - \beta_1 - \beta_2}.$$



Pentru a obține varianța procesului x_t se aplică operatorul de varianță relației de definiție: $Var[x_t] = \beta_1^2 Var[x_{t-1}] + \beta_2^2 Var[x_{t-2}] + 2\beta_1\beta_2 Cov[x_{t-1}, x_{t-2}] + Var[\varepsilon_t]$. Dacă procesul este staționar, $Var[x_t] = Var[x_{t-1}] = Var[x_{t-2}] = \gamma_0$, iar relația devine $\gamma_0 = \beta_1^2 \gamma_0 + \beta_2^2 \gamma_0 + 2\beta_1\beta_2 \gamma_1 + \sigma^2$. În această relație apar două necunoscute: varianța γ_0 și covarianța γ_1 . Pentru a le determina, este necesară încă o relație, pe care o obținem calculând covarianța dintre x_t și x_{t-1} : $\gamma_1 = \beta_1 \gamma_0 + \beta_2 \gamma_1$.

Soluția sistemului format cu cele două relații este

$$\gamma_0 = \frac{1 - \beta_2}{1 + \beta_2} \frac{\sigma^2}{(1 - \beta_1 - \beta_2)(1 + \beta_1 - \beta_2)} \text{ și } \gamma_1 = \frac{\beta_1}{1 - \beta_2} \gamma_0.$$

Coeficientul de corelație la lagul 1 este $\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\beta_1}{1 - \beta_2}$. Calculând corelația dintre $x_t = \alpha + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \varepsilon_t$ și x_{t-k} obținem pentru coeficientul de corelație următoarea relație de recurență: $\rho_k = \beta_1 \rho_{k-1} + \beta_2 \rho_{k-2}$. Această relație se poate transcrie matricial, în $\begin{pmatrix} \rho_k \\ \rho_{k-1} \end{pmatrix} = b \begin{pmatrix} \rho_{k-1} \\ \rho_{k-2} \end{pmatrix}$, cu soluția $\begin{pmatrix} \rho_k \\ \rho_{k-1} \end{pmatrix} = b^k \begin{pmatrix} \rho_1 \\ \rho_0 \end{pmatrix}$. Cum b^k tinde la zero dacă procesul AR(2) este staționar, corelația ρ_k va tinde la zero, pe măsură ce $k \rightarrow \infty$. Acest fapt este ilustrat în figura anterioară.

Coeficientul de corelație parțială a_k este $a_1 = \rho_1$, $a_2 = \beta_2$ și $a_k = 0$ pentru $k > 2$, deoarece x_t depinde direct numai de x_{t-1} și x_{t-2} .

Pentru un proces AR(2) funcția de autocorelație descrește treptat către zero, în timp ce funcția de autocorelație parțială descrește brusc la zero, după primele două laguri.

6.3.3 Procese de tip MA(1)

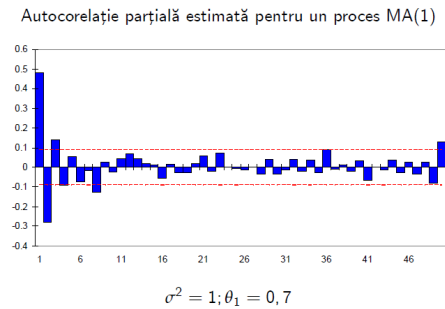
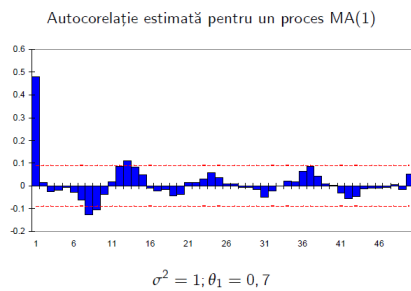
Un **proces MA(1)** este un proces rezultat prin combinarea valorilor curentă și anterioară ale procesului *white-noise*:

$$x_t = \alpha + \theta_1 \varepsilon_{t-1} + \varepsilon_t,$$

cu ε_t *white-noise*. Cu operatorul de lag, relația de definire a unui proces MA(1) este $x_t = \alpha + (1 + \theta_1 L)\varepsilon_t = \alpha + \Theta(L)\varepsilon_t$, unde $\Theta(L) = 1 + \theta_1 L$ este un polinom de lag-uri.

Un proces MA(1) este întotdeauna staționar, indiferent de valoarea lui θ_1 , deoarece este o combinație de procese staționare. Media unui proces MA(1) este $\mu = E[x_t] = \alpha$. Ținând cont de faptul că ε_t și ε_{t-1} sunt necorelate, se obține că

$\gamma_0 = Var[x_t] = (1 + \theta_1^2)\sigma^2$. Autocovarianța pentru primul lag, se poate calcula din este $\gamma_1 = Cov[x_t, x_{t-1}] = \theta_1\sigma^2$, de unde rezultă $\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\theta_1}{1 + \theta_1^2}$. În general, $\gamma_k = 0$, pentru $k > 1$. Prin urmare, $\rho_k = 0$, pentru $k > 1$. Se poate arăta că funcția de autocorelație parțială pentru un proces MA(1) descrește treptat către zero.



Din figura de mai sus rezultă că funcțiile de autocorelație pentru un proces MA(1) au următorul comportament: autocorelația scade brusc la zero, după primul lag, în timp ce autocorelația parțială descrește treptat către zero.

Pe baza unui raționament similar celui de la secțiunea dedicată proceselor de tip AR(1), se arată că dacă $|\theta_1| < 1$, polinomul de lag $\Theta(L)$ este inversabil,

$$\Theta(L)^{-1} = 1 - \theta_1 L + \theta_1^2 L^2 - \theta_1^3 L^3 + \dots = \sum_{k=0}^{\infty} (-\theta_1)^k L^k.$$

În acest caz, procesul MA(1) se numește inversabil. Înmulțind relația $x_t = \alpha + \Theta(L)\varepsilon_t$ cu $\Theta(L)^{-1}$ obținem

$$\Theta(L)^{-1} x_t = \Theta(1)^{-1} \alpha + \varepsilon_t, \text{ de unde } x_t = \Theta(1)^{-1} \alpha + \sum_{k=1}^{\infty} (-\theta_1)^k x_{t-k} + \varepsilon_t, \text{ scriere de forma}$$

AR(∞). În această formă, valoarea curentă a procesului apare ca media tuturor valorilor sale anterioare. Pe baza acestui rezultat se poate argumenta traiectoria treptat descendentă către zero a funcției de autocorelație parțială pentru un proces MA(1).

6.3.4 Procese de tip MA(2)

Un **proces MA(2)** este un proces rezultat prin combinarea valorilor curentă și anterioară ale procesului *white-noise*:

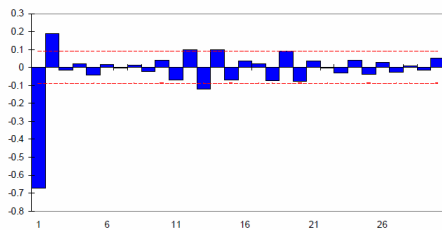
$$x_t = \alpha + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \varepsilon_t,$$

cu ε_t *white-noise*. Cu operatorul de lag, relația de definire a unui proces MA(1) este $x_t = \alpha + (1 + \theta_1 L + \theta_2 L^2) \varepsilon_t = \alpha + \Theta(L) \varepsilon_t$, unde $\Theta(L) = 1 + \theta_1 L + \theta_2 L^2$ este un polinom de lag-uri.

Un proces MA(2) este întotdeauna staționar, indiferent de valorile parametrilor θ_1 și θ_2 , deoarece este o combinație de procese staționare.

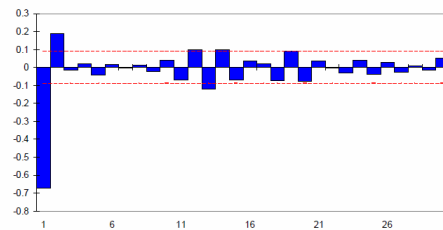
Condițiile de inversabilitate se obțin printr-un raționament analog celui utilizat pentru deducerea condițiilor de staționaritate a unui proces AR(2). Procesul este inversabil dacă și numai dacă soluțiile ecuației $\lambda^2 + \theta_1 \lambda + \theta_2 = 0$ sunt sub-unitare în modul.

Autocorelație estimată pentru un proces MA(2)



$$\sigma^2 = 1; \theta_1 = -1, 7; \theta_2 = 0, 72$$

Autocorelație estimată pentru un proces MA(2)



$$\sigma^2 = 1; \theta_1 = -1, 7; \theta_2 = 0, 72$$

Media unui proces MA(2) este $\mu = E[x_t] = \alpha$.

Varianța unui proces MA(2) este $\gamma_0 = Var[x_t] = (1 + \theta_1 + \theta_2) \sigma^2$, iar covarianța: $\gamma_1 = \theta_1 (1 + \theta_2) \sigma^2$, $\gamma_2 = \theta_2 \sigma^2$ și $\gamma_k = 0$ pentru $k > 2$. Rezultă că valoarea coeficientului de corelație este zero pentru $k > 2$.

Funcția de autocorelație parțială descrește treptat către zero. Argumentul este similar celui de la procesele de tip MA(1).

6.3.5 Procese de tip $ARMA(1,1)$

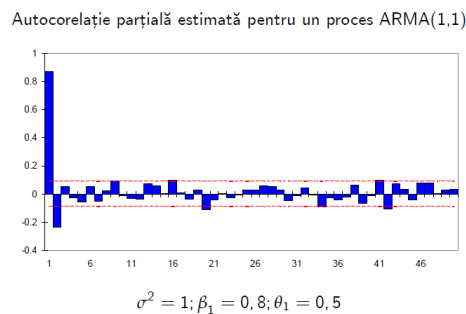
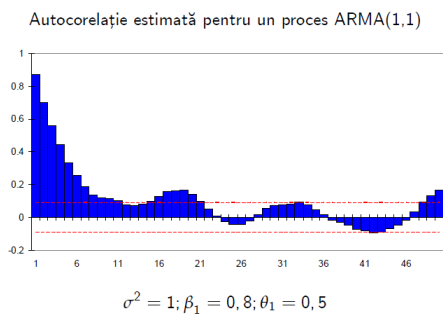
Un **proces $ARMA(1,1)$** este construit prin combinarea unui proces $AR(1)$ cu unul $MA(1)$:

$$x_t = \alpha + \beta_1 x_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t.$$

Utilizând operatorul de lag, relația de mai sus se poate scrie sub forma

$$B(L)x_t = \alpha + \Theta(L)\varepsilon_t, \text{ cu } B(L) = 1 - \beta_1 L \text{ și } \Theta(L) = 1 + \theta_1 L.$$

Fiind o combinație între procesul $AR(1)$ și $MA(1)$, procesul $ARMA(1,1)$ moștenește caracteristicile acestora în ceea ce privește funcțiile de autocorelație și autocorelație parțială. De asemenea, deoarece procesul $MA(1)$ este staționar indiferent de parametrul θ_1 , staționaritatea unui proces $ARMA(1,1)$ este guvernată de staționaritatea componentei $AR(1)$. Astfel, procesul $ARMA(1,1)$ este staționar dacă și numai dacă partea autoregresivă a acestuia este staționară.

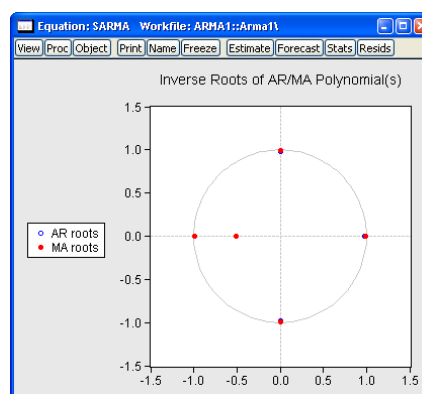
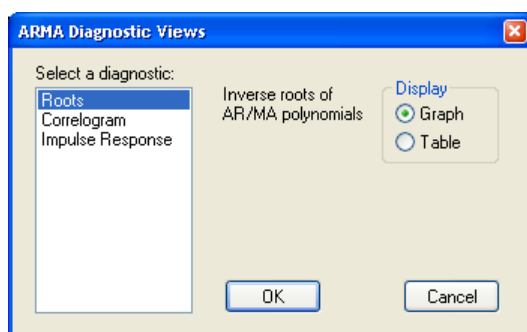


Se poate arăta faptul că funcția de autocorelație a unui proces $ARMA(1,1)$ urmează același tipar cu cea a unui proces $AR(1)$, numai că începând de la lagul 2 și nu de la lagul 1. Funcția de autocorelație parțială pentru un proces $ARMA(1,1)$ descrește, la rândul său, treptat către zero.

Estimare modelelor $ARMA$ în **Eviews** se realizează utilizând un obiect de tip *equation*. În cadrul specificației ecuației în caseta de dialog **Equation Estimation**, părțile AR și MA ale modelului sunt specificate folosind cuvintele

cheie *ar* și *ma*, ca parte a ecuației. Rezultatele estimării unui model de regresie cu specificații AR sau MA sunt similare cu cele de la o ecuație de regresie simplă, cu adăugarea unui nou bloc, care prezintă rădăcinile polinoamelor AR și MA.

Eviews oferă acces la mai multe instrumente de diagnostic, care ajută la evaluarea structurii ARMA a ecuației estimate. Pentru a afișa structura ARMA, selectați **View/ARMA Structure...** din meniul unei ecuații estimate. Se deschide căsuța de dialog **ARMA Diagnostic Views**, aceasta fiind disponibilă doar pentru modelele care includ cel puțin un termen AR sau MA și sunt estimate cu metoda celor mai mici pătrate. Există trei moduri de vizualizare disponibile: prezentarea rădăcinilor polinoamelor AR și MA, cea mai importantă, prezentarea corelogramei și analiza de tip răspuns la impuls.



Vizualizarea referitoare la rădăcinile asociate modelului afișează inversul rădăcinilor polinomului caracteristic AR și/sau MA. Rădăcinile pot fi afișate ca un grafic sau sub forma unui tabel prin selectarea butonului corespunzător. Vizualizarea sub formă grafică prezintă rădăcinile în planul complex în care pe axa orizontală este partea reală și pe axa verticală este partea imaginară a fiecărei rădăcini. În cazul în care procesul ARMA estimat este staționar, atunci toate rădăcinile AR ar trebui să fie plasate în interiorul cercului unitate. În cazul în care procesul ARMA estimat este inversabil, atunci toate rădăcinile MA ar trebui să se

plaseze în interiorul cercului unitate. Vizualizarea sub formă de tabel afișează toate rădăcinile în ordine descrescătoare a modulului (radical din suma pătratelor părților reale și imaginare).

6.4 Metodologia Box-Jenkins

Pe baza rezultatelor analitice prezentate în secțiunea anterioară, exemplificate pe exemple simulate, putem trage concluzia conform căreia fiecare din cele trei procese analizate până în prezent: AR(1), MA(1) și ARMA(1,1) prezintă un tipar caracteristic după care se comportă funcțiile de autocorelație și de autocorelație parțială. Această semnătură care identifică unic unul din cele trei procese, poate fi utilizată pentru a identifica modelul care este potrivit pentru a modela o serie de date reale.

Determinarea procesului pe care îl urmează o serie de timp reală, în funcție de structura de autocorelație.

Autocorelație	Autocorelația parțială	Concluzie
Descrește treptat la zero.	Scade brusc la zero, după primul lag.	Proces AR(1)
Scade brusc la zero, după primul lag.	Descrește treptat la zero.	Proces MA(1)
Descrește treptat la zero.	Descrește treptat la zero.	Proces ARMA(1,1)

Astfel, dacă pentru o serie de date funcțiile de autocorelație și autocorelație parțială estimate se comportă ca cele ale unui model AR(1) (autocorelația scade treptat către zero, în timp ce autocorelația parțială scade brusc, fiind semnificativă

numai pentru primul lag), putem trage concluzia că această serie se poate modela corespunzător printr-un astfel de proces. Similar pentru procesele MA(1) și ARMA(1,1). Această abordare poartă numele de **metodologia Box-Jenkins** și este sintetizată în tabelul de pe pagina anterioară.

Primul pas în formarea unui model ARIMA pentru o serie de timp este analizarea proprietăților de autocorelație a acesteia. În acest scop, se poate utiliza opțiunea **View/Correlogram...** pentru a determina **corelograma**. Natura corelației dintre valorile curente ale reziduurilor și a valorilor lor din trecut oferă informații în ceea ce privește selectarea unei specificații pentru modelul ARMA.

Deși este relativ simplă și ușor de interpretat, metodologia Box-Jenkins nu oferă un rezultat lipsit de echivoc în privința procesului ARMA prin care se poate modela o serie de date. Prezența erorilor în estimarea coeficienților de autocorelație și de autocorelație parțială face ca identificarea vizuală a traiectoriei funcțiilor de autocorelație să fie dificilă. Utilizarea metodologiei Box-Jenkins presupune o experiență vastă de lucru cu datele. Pentru selectarea mecanică și fără echivoc a unui model de tip ARMA se pot utiliza criteriile informaționale.

6.5 Modele tip *ARIMA*

Dacă o serie este nestaționară în nivel, dar staționară în prima diferență se numește **serie integrată de ordinul 1**, notată cu **I(1)**. Dacă o serie este nestaționară și trebuie diferențiată de cel puțin k ori pentru a deveni staționară, se numește integrată de ordinul k , notată cu $I(k)$. Cele mai multe serii de date economice sunt integrate de ordinul 1, cel mult 2. O serie este reprezentată corespunzător printr-un **model ARIMA(1,k,1)** dacă după ce este diferențiată de k ori, rezultatul se poate reprezenta fidel printr-un model ARMA(1,1).

Modele ARIMA (AutoRegressive Integrated Moving Average) sunt folosite pe scară largă pentru a analiza dinamica seriilor de timp unidimensionale. Un model ARIMA este compus din trei părți: termenul autoregresiv, termenul privind ordinul de integrare și termenul de medie mobilă. În analiza de tip ARIMA,

se assemblează un model de prognoză complet prin utilizarea de combinații ale celor trei blocuri descrise mai sus.

Pentru a specifica modelul ARIMA, este necesar să:

- se diferențieze variabila dependentă, dacă este necesar, pentru a ține cont de ordinul de integrare.
- se descrie modelul de regresie structural (variabilele dependente și regresori) și să se adauge termeni AR sau MA, așa cum este descris mai sus.

7 Tehnici de modelare și previzionare a seriilor de timp multidimensionale

Abordarea structurală pentru modelarea seriilor de timp utilizează teoria economică pentru a modela relația dintre variabilele de interes. Din păcate, teoria economică nu este suficient de bogată pentru a oferi o specificație dinamică, care identifică toate aceste relații. În plus, estimarea și inferența sunt complicate de faptul că variabilele endogene pot să apară în ambele părți ale unei ecuații. Aceste probleme conduc la abordări alternative, nestructurale, pentru modelarea relației dintre mai multe variabile.

Extensia la cazul multivariat a modelelor autoregresive scalare, univariate, poartă denumirea de **modele VAR** (Vector AutoRegression), modele autoregresive vectoriale. Modelele de tip VAR sunt frecvent utilizate pentru prognozarea sistemelor de serii de timp interconectate și pentru analizarea impactului dinamic al inovațiilor asupra sistemului de variabile. Abordarea VAR eludează nevoia de modelare structurală prin tratarea fiecărei variabile endogene din sistem ca pe o funcție a lag-urilor, a valorilor din trecut, a tuturor variabilelor endogene din sistem.

Printre avantajele modelelor VAR:

- nu necesită separarea clară a variabilelor în endogene și exogene;
- se pot utiliza pentru a deduce modul în care variabilele economice răspund la șocuri;
- sunt utilizate pe scară largă în modelarea macroeconomică, fiind incluse în majoritatea programelor econometrice

Cu toate acestea, modelele VAR nu sunt lipsite de dezavantaje:

- au fost criticate deseori din pricina lipsei fundamentelor teoretice;

- în situația în care presupun și existența unor interacțiuni contemporane între variabile (modele structurale) necesită impunerea de restricții suplimentare, pentru a putea fi identificate;
- interpretarea rezultatelor depinde decisiv de modul în care au fost impuse aceste restricții.

Acest capitol descrie estimarea și analiza modelelor de tip VAR. Următorul capitol descrie estimarea și analiza modelelor care permit integrarea prezenței unei relații pe termen lung între mai multe variabile din model.

7.1 Modele tip VAR(1)

Modelul VAR pornește de la ideea că toate variabilele pot fi endogene, valoarea curentă a fiecărei variabile depinzând de propriile laguri și de lagurile celorlalte variabile. Un model **VAR bidimensional cu un lag** este reprezentat prin ecuațiile

$$\begin{cases} x_t = \beta_{11}^0 + \beta_{11}^1 x_{t-1} + \beta_{12}^1 z_{t-1} + u_t^1 \\ z_t = \beta_{12}^0 + \beta_{21}^1 x_{t-1} + \beta_{22}^1 z_{t-1} + u_t^2 \end{cases}$$

Matricial, relația de mai sus se scrie ca

$$y_t = \beta_0 + B y_{t-1} + \varepsilon_t,$$

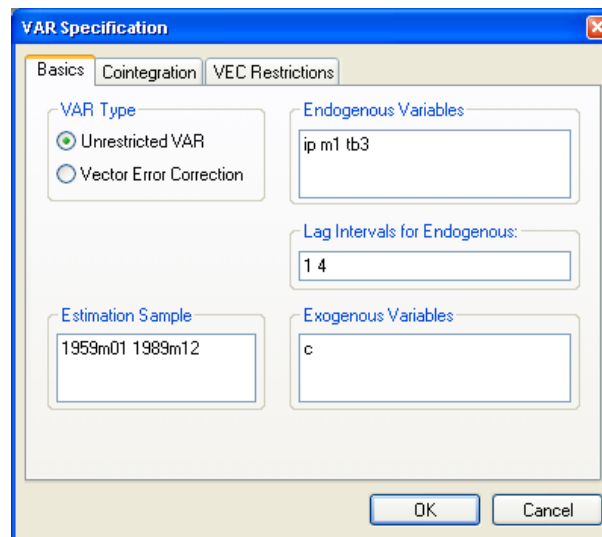
unde $y_t = (x_t, z_t)^T$, $\beta_0 = (\beta_{11}^0, \beta_{12}^0)^T$, $B = \begin{pmatrix} \beta_{11}^1 & \beta_{12}^1 \\ \beta_{21}^1 & \beta_{22}^1 \end{pmatrix}$ și $\varepsilon_t = (u_t^1, u_t^2)^T$ este

un vector de inovații, care poate fi contemporan corelat, dar este necorelat cu valorile din perioadele anterioare și cu variabilele din partea dreaptă a ecuației.

Pentru a specifica un model VAR în **Eviews**, trebuie să se creeze mai întâi un obiect *var*. Se selectează **Quick/Estimate VAR...** Fila **Basics** a căsuței de dialog va solicita definirea structurii VAR:

- Se selectează tipul de model VAR: **Unrestricted VAR**. Ceea ce s-a numit până acum un VAR este de fapt un VAR nerestricționat.

- Se introduc specificațiile lag-urilor în căsuța de editare corespunzătoare. Această informație se introduce în perechi: fiecare pereche de numere definește o serie de lag-uri.
- Se introduce numele seriilor endogene și exogene în casetele de editare corespunzătoare.



Eviews va afișa rezultatele estimării în fereastra VAR. Fiecare coloană din tabel corespunde unei ecuații din modelul VAR. Pentru fiecare variabilă din partea dreaptă, **Eviews** raportează coeficientul estimat, eroarea sa standard și statistica testului t. Deoarece nu este afișat p-value pentru testul t, interpretarea rezultatelor acestui test este mai dificilă. Ca o regulă generală, dacă valoarea în modul a a statisticii t este „mare” (i.e. mai mare decât 2,5) se respinge ipoteza nulă și astfel parametrul respectiv este semnificativ din punct de vedere statistic.

Vector Autoregression Estimates

Date: 08/02/97 Time: 09:06
Sample(adjusted): 1959M05 1989M12
Included observations: 368 after adjusting endpoints
Standard errors in () & t-statistics in []

	IP	M1	TB3
IP(-1)	1.253934 (0.05401) [23.2147]	0.253215 (0.17769) [1.42501]	0.095984 (0.05021) [1.91170]
IP(-2)	-0.187774 (0.08557) [-2.19448]	-0.230187 (0.28149) [-0.81774]	0.015590 (0.07954) [0.19601]
IP(-3)	-0.003780 (0.08558) [-0.04418]	-0.153515 (0.28146) [-0.54543]	-0.173824 (0.07953) [-2.19570]

Vector Autoregression Estimates

R-squared	0.999221	0.999915	0.968018
Adj. R-squared	0.999195	0.999912	0.966937
Sum sq. resids	113.8813	1232.453	98.39849
S.E. equation	0.566385	1.863249	0.526478
F-statistic	37950.20	347533.2	895.4048
Log likelihood	-306.3509	-744.5662	-279.4628
Akaike AIC	1.735603	4.117208	1.589472
Schwarz SC	1.873660	4.255265	1.727529
Mean dependent	70.97919	339.7451	6.333891
S.D. dependent	19.95932	198.6301	2.895381
Determinant resid covariance (dof adj.)		0.259637	
Determinant resid covariance		0.233082	
Log likelihood		-1298.537	
Akaike information criterion		7.269222	
Schwarz criterion		7.683394	

Eviews afișează informații suplimentare în partea de jos a ferestrei cu rezumatul estimării. Prima parte a informației suplimentare prezintă statistici standard ale regresiei OLS pentru fiecare ecuație. Rezultatele sunt calculate separat pentru fiecare ecuație folosind reziduurile acestora și sunt afișate în coloana corespunzătoare. Numerele prezentate în partea de jos a tabelului sunt rezumate ale statisticilor sistemului VAR ca un întreg.

După ce a fost estimat un model VAR, **Eviews** oferă posibilități diferite pentru a lucra cu acest model. Un set de instrumente de diagnosticare sunt prevăzute în cadrul structurii de meniuri **View/Lag Structure** și **View/Residual Tests** în fereastra VAR. Aceste instrumente pot ajuta la verificarea gradului de adecvare a VAR estimat.

Modelul VAR(1) are o importanță deosebită în metodologia VAR, având în vedere faptul că orice model VAR, cu un număr p de laguri se poate transforma într-un VAR(1) și invers. Vom ilustra acest fapt pentru un model VAR(2), cu două

variabile:

$$\begin{cases} x_t = \beta_{11}^0 + \beta_{11}^1 x_{t-1} + \beta_{12}^1 z_{t-1} + \beta_{11}^2 x_{t-2} + \beta_{12}^2 z_{t-2} + u_t^1 \\ z_t = \beta_{12}^0 + \beta_{21}^1 x_{t-1} + \beta_{22}^1 z_{t-1} + \beta_{21}^2 x_{t-2} + \beta_{22}^2 z_{t-2} + u_t^2 \end{cases}$$

Matricial, relația de mai sus se scrie ca

$$y_t = B_0 + B_1 y_{t-1} + B_2 y_{t-2} + \varepsilon_t,$$

unde $y_t = (x_t, z_t)^T$, $B_0 = (\beta_{11}^0, \beta_{12}^0)^T$, $B_1 = \begin{pmatrix} \beta_{11}^1 & \beta_{12}^1 \\ \beta_{21}^1 & \beta_{22}^1 \end{pmatrix}$, $B_2 = \begin{pmatrix} \beta_{11}^2 & \beta_{12}^2 \\ \beta_{21}^2 & \beta_{22}^2 \end{pmatrix}$ și

$\varepsilon_t = (u_t^1, u_t^2)^T$. Matricele din relația anterioară se pot grupa, astfel încât modelul să aibă forma unui VAR(1), astfel

$$Y_t = \begin{pmatrix} B_0 \\ 0 \end{pmatrix} + BY_{t-1} + U_t,$$

Unde $Y_t = (y_t^T, y_{t-1}^T)^T$, $B = \begin{pmatrix} B_1 & B_2 \\ I_{(2 \times 2)} & 0 \end{pmatrix}$ și $U_t = (\varepsilon_t^T, \varepsilon_{t-1}^T)^T$.

Scris cu variabilele inițiale, x_t și z_t , $Y_t = (x_t, z_t, x_{t-1}, z_{t-1})^T$. x_t și z_t se pot

obține din Y_t prin înmulțirea cu matricea $J = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$:

$$\begin{pmatrix} x_t \\ z_t \end{pmatrix} = J \cdot Y_t.$$

7.2 Funcții de răspuns la impuls

Numărul ridicat de coeficienți dintr-un model VAR (într-un model cu un lag și două variabile, fără constantă, sunt 4 coeficienți ce trebuie estimați) face relativ dificilă interpretarea relațiilor care există între variabile pornind de la acești coeficienți. În loc să se analizeze fiecare coeficient din fiecare ecuație, se poate analiza o imagine sintetică a comportamentului dinamic a modelului VAR. Această imagine este dată de funcțiile de răspuns la impuls, care descriu (sub formă de tabel sau sub formă grafică) modul în care fiecare variabilă reacționează la un șoc propriu sau la un șoc în celelalte variabile.

Un șoc apărut în cadrul ecuației variabilei i nu numai că afectează în mod direct variabila i , dar este transmis, de asemenea, tuturor celorlalte variabile endogene prin structura dinamică a modelului VAR. O **funcție de răspuns la**

impuls urmărește efectul unui șoc apărut la un moment dat într-una din inovațiile modelului asupra valorilor prezente și viitoare ale variabilelor endogene.

Cea mai simplă cale de construcție a funcțiilor de răspuns la impuls este prin iterarea succesivă a modelului, utilizând matricea estimată de coeficienți. Să presupunem, de exemplu că vrem să analizăm efectele unui șoc în variabila x_t .

Pornim cu $y_0 = (x_0, z_0)^T$, unde $x_0 = 1$ și $z_0 = 0$. După o perioadă

$$\begin{pmatrix} x_1 \\ z_1 \end{pmatrix} = \begin{pmatrix} \beta_{11}^1 & \beta_{12}^1 \\ \beta_{21}^1 & \beta_{22}^1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \beta_{11}^1 \\ \beta_{21}^1 \end{pmatrix},$$

unde pentru a surprinde numai efectele unui șoc inițial în x_t , presupunem că erorile în toate celelalte momente sunt zero.

În relația precedentă β_{11}^1 reprezintă efectul unui șoc în x_t asupra lui x_t , după o perioadă, iar β_{21}^1 reprezintă efectul unui șoc în x_t asupra lui z_t , după o perioadă.

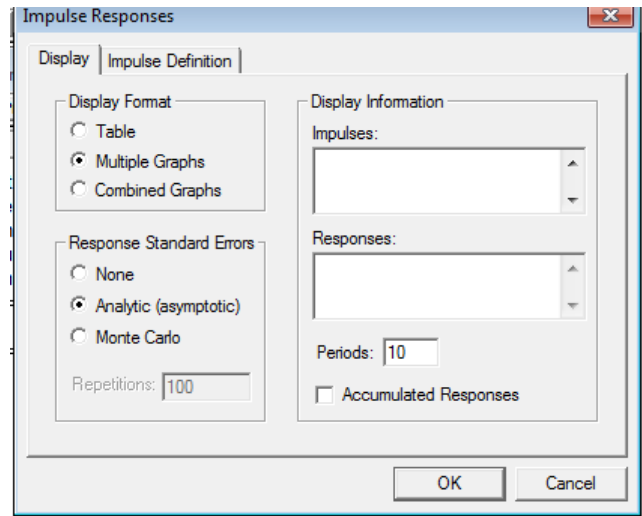
Iterând încă o perioadă:

$$\begin{pmatrix} x_2 \\ z_2 \end{pmatrix} = \begin{pmatrix} \beta_{11}^1 & \beta_{12}^1 \\ \beta_{21}^1 & \beta_{22}^1 \end{pmatrix} \begin{pmatrix} \beta_{11}^1 \\ \beta_{21}^1 \end{pmatrix} = B^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

rezultă că efectele unui șoc în x_t după două perioade se extrag din prima coloană a matricei B^2 .

Iterând în continuare, după 3, ..., n perioade, obținem că efectul (răspunsul la impuls) în variabila i la un șoc în variabila j este egal cu elementul de pe poziția (i, j) din matricea B^n , $r_i^j(n) = (B^n)_{i,j}$.

În cazul în care inovațiile sunt simultan necorelate, interpretarea răspunsului la impuls este relativ simplă. Inovația i este pur și simplu un șoc pentru variabila i . Inovațiile, cu toate acestea, sunt de obicei corelate, și pot fi privite ca având un element comun, care nu poate fi asociat cu o variabilă specifică. Pentru a interpreta, în această situație generală, răspunsurile la impuls, se obișnuiește să se aplice o transformare a inovațiilor, astfel încât acestea să devină necorelate.



În **Eviews**, pentru a obține funcțiile de răspuns la impuls, se estimează mai întâi modelul VAR, apoi se selectează **View/Impulse Response...** din bara de instrumente VAR. Se afișează o casetă de dialog cu două file: **Display** și **Impulse Definition**. În fila **Display Information** ar trebui introduse variabilele pentru care se dorește să se genereze inovații (**Impulsuri**) și variabilele pentru care se dorește să se urmărească răspunsurile (**Responses**).

7.3 Descompunerea varianței

În timp ce funcțiile de răspuns la impuls urmează efectele unui șoc apărut în dinamica unei variabile asupra unei alte variabile din VAR, **descompunerea varianței** oferă informații despre importanța relativă a fiecărei inovații privind efectul asupra dinamicii variabilelor din VAR. Pentru a obține descompunerea varianței, se selectează **View/Variance Decomposition...** din bara de instrumente a obiectului var. Practic, trebuie să se furnizeze aproximativ aceleași informații ca și mai sus în cazul specificării unei funcții de răspuns la impuls. Ca și în cazul

răspunsurilor la impuls, descompunerea varianței bazate pe metoda Cholesky, are o mare sensibilitate față de modul de ordonare al variabilelor în modelul VAR.

Vom prezenta intuiția teoretică pe care se bazează descompunerea varianței erorii de prognoză pentru un model VAR(1) cu două variabile. Pentru mai multe laguri și mai multe variabile intuiția se păstrează, dar apar unele complicații inerente dimensiunii ridicate a sistemului și numărului ridicat de laguri. Vom considera modelul VAR(1) cu două variabile: $y_t = B_0 + B_1 y_{t-1} + \varepsilon_t$, unde

$y_t = (x_t, z_t)$, $B_0 = \begin{pmatrix} \beta_{11}^0 \\ \beta_{21}^0 \end{pmatrix}$, $B_1 = \begin{pmatrix} \beta_{11}^1 & \beta_{12}^1 \\ \beta_{21}^1 & \beta_{22}^1 \end{pmatrix}$ și $\varepsilon_t = \begin{pmatrix} u_t^1 \\ u_t^2 \end{pmatrix}$. Vectorul de erori este

presupus a fi distribuit normal bi-variant, cu medie $E[\varepsilon_t] = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ și matrice de

$$\text{varianță covarianță } \Sigma = E \left[\begin{pmatrix} u_t^1 & u_t^2 \end{pmatrix} \begin{pmatrix} u_t^1 \\ u_t^2 \end{pmatrix} \right] = \begin{pmatrix} E[u_t^1 u_t^1] & E[u_t^1 u_t^2] \\ E[u_t^2 u_t^1] & E[u_t^2 u_t^2] \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}.$$

Dacă se cunosc valorile celor două variabile la momentul 0: x_0 și z_0 , prognoza pentru momentul următor va fi: $E[y_1 | y_0] = B_0 + B_1 y_0$, iar eroarea de prognoză

$$y_1 - E[y_1 | y_0] = \varepsilon_1.$$

Prognoza pentru momentul 2, pe baza informațiilor de la momentul 0 este $E[y_2 | y_0] = B_0 + B_1 E[y_1 | y_0] = B_0 + B_1 B_0 + B_1^2 y_0$, cu eroarea de prognoză

$$y_2 - E[y_2 | y_0] = \varepsilon_2 + B_1 \varepsilon_1.$$

În general, eroarea corespunzătoare prognozei de peste k perioade, pe baza informațiilor din prezent este

$$y_k - E[y_k | y_0] = \varepsilon_k + B_1 \varepsilon_{k-1} + B_1^2 \varepsilon_{k-2} + \dots + B_1^{k-1} \varepsilon_1,$$

ceea ce se traduce printr-o medie ponderată a erorilor ce vor apărea pe parcurs. Ponderile erorilor sunt cu atât mai mari, cu cât acestea sunt mai recente.

Eroarea de prognoză a variabilei x_t după două perioade este $x_2 - E[x_2 | y_0] = u_2^1 + \beta_{11}^1 u_1^1 + \beta_{12}^1 u_2^1$, iar după trei perioade:

$$x_3 - E[x_3 | y_0] = u_3^1 + \beta_{11}^1 u_2^1 + \beta_{12}^1 u_2^2 + \left((\beta_{11}^1)^2 + \beta_{12}^1 \beta_{21}^1 \right) u_1^1 + \left(\beta_{11}^1 \beta_{12}^1 + \beta_{12}^1 \beta_{22}^1 \right) u_2^1.$$

Coefficienții $\left((\beta_{11}^1)^2 + \beta_{12}^1 \beta_{21}^1 \right)$ și $\left(\beta_{11}^1 \beta_{12}^1 + \beta_{12}^1 \beta_{22}^1 \right)$ sunt extrași de pe prima linie a matricei B_1^2 . Pentru eroarea de prognoză a variabilei x_t după trei perioade am avea nevoie de prima linie a matricei B_1^3 etc. Scrierea erorii de prognoză a lui x_t se simplifică dacă utilizăm funcțiile de răspuns la impuls $r_1^j(k)$, răspunsul primei variabile (x_t) la variabila j ($j=1$ pentru x_t și $j=2$ pentru z_t) după k perioade. Utilizând răspunsurile la impuls, eroarea de prognoză a variabilei x_t după două perioade se scrie: $x_2 - E[x_2 | y_0] = r_1^1(0)u_2^1 + r_1^2(0)u_2^2 + r_1^1(1)u_1^1 + r_1^2(1)u_1^2$, iar după două perioade:

$$x_3 - E[x_3 | y_0] = r_1^1(0)u_3^1 + r_1^2(0)u_3^2 + r_1^1(1)u_2^1 + r_1^2(1)u_2^2 + r_1^1(2)u_1^1 + r_1^2(2)u_1^2. \text{ În general,}$$

$$x_k - E[x_k | y_0] = \sum_{j=0}^{k-1} r_1^1(j)u_{k-j}^1 + \sum_{j=0}^{k-1} r_1^2(j)u_{k-j}^2.$$

Varianța erorii de prognoză este

$$Var[x_k - E[x_k | y_0]] = Var \left[\sum_{j=0}^{k-1} r_1^1(j)u_{k-j}^1 + \sum_{j=0}^{k-1} r_1^2(j)u_{k-j}^2 \right].$$

Deoarece erorile sunt necorelate, atât între ecuații, la același moment, dar și în interiorul aceleiași ecuații, varianța erorii de prognoză devine

$$\begin{aligned} Var[x_k - E[x_k | y_0]] &= \sum_{j=0}^{k-1} Var[r_1^1(j)u_{k-j}^1] + \sum_{j=0}^{k-1} Var[r_1^2(j)u_{k-j}^2] \\ &= \sum_{j=0}^{k-1} (r_1^1(j))^2 Var[u_{k-j}^1] + \sum_{j=0}^{k-1} (r_1^2(j))^2 Var[u_{k-j}^2] \\ &= \left(\sum_{j=0}^{k-1} (r_1^1(j))^2 \right) \sigma_1^2 + \left(\sum_{j=0}^{k-1} (r_1^2(j))^2 \right) \sigma_2^2. \end{aligned}$$

Varianța de prognoză a variabilei x_t apare astfel ca fiind rezultatul celor două variante ale erorilor care intervin în ecuațiile VAR. Putem calcula procentul de varianță totală a erorii de prognoză care se datorează primei variabile:

$$\left(\sum_{j=0}^{k-1} (r_1^1(j))^2 \right) \sigma_1^2 / \text{Var}[x_k - E[x_k | y_0]], \text{ restul } \left(\sum_{j=0}^{k-1} (r_1^2(j))^2 \right) \sigma_2^2 / \text{Var}[x_k - E[x_k | y_0]]$$

fiind datorat varianței celei de a doua variabilă.

În general, într-un sistem cu n variabile, procentul din varianța erorii de prognoză a variabilei i care se datorează varianței erorii din ecuația variabilei j ,

$$\text{după } k \text{ perioade, este } \left(\sum_{j=0}^{k-1} (r_i^j(l))^2 \right) \sigma_i^2 / \text{Var}[x_k - E[x_k | y_0]].$$

7.4 Cauzalitate Granger

În cadrul modelelor VAR se pot stabili relațiile de **cauzalitate în sens Granger** care există între variabile. Pornind de la ideea că efectul nu poate să preceadă cauza, Granger (1969) a introdus o noțiune de cauzalitate definită astfel: dacă variabila x_t influențează variabila z_t , atunci cunoașterea valorilor lui x trebuie să amelioreze performanța prognozelor făcute asupra lui z .

În modelul VAR(1) prezentat mai sus, cauzalitatea Granger între cele două variabile se testează astfel:

- de la z la x :

$$H_0 : z \text{ nu cauzează Granger pe } x \Leftrightarrow \beta_{12}^1 = 0 ;$$

- de la x la z :

$$H_0 : x \text{ nu cauzează Granger pe } z \Leftrightarrow \beta_{21}^1 = 0 .$$

Respingerea ipotezei nule este un indiciu în favoarea cauzalității.

8 Tehnici de analiză a dinamicii pe termen lung a variabilelor economice

Acest capitol descrie procesul de estimare și analiză a modelelor de tip VEC (Vector Error Correction), precum și modelele și instrumentele pentru testarea prezenței relațiilor de cointegrare dintre mai multe variabile nestaționare.

Teoria din spatele estimării modelelor ARMA, discutate în capitolul 6, se bazează pe serii de timp staționare. O serie este staționară (în covarianță) în cazul în care media, varianța și auto-covarianțele seriei nu depind de timp. Orice serie care nu este staționară se numește nestaționară. Un exemplu clasic de serie nestaționară este procesul de tip *random walk*:

$$y_t = y_{t-1} + \varepsilon_t$$

unde ε este un proces staționar. Procesul *Random walk* este o serie staționară în prima diferență, sau proces cu ordin de integrare 1, deoarece prima diferență a procesului ($y_t - y_{t-1} = \varepsilon_t$) este un proces staționar. Ordinul de integrare, sau numărul de rădăcini unitare (unit roots), reprezintă numărul de operațiuni de diferență necesare pentru a face o serie staționară. Pentru procesul de tip random walk există o singură rădăcină unitate, deci este o serie I(1). Similar, o serie staționară este I(0).

Procedurile standard de inferență nu se aplică pentru ecuații de regresie care conțin o variabilă dependentă integrată sau regresori integrați. Prin urmare, este important să se verifice dacă o serie este staționară înainte de a o utiliza într-o regresie. Metoda formală pentru testarea staționarității unei serii este testul de unit root (rădăcină unitate), despre care s-a discutat în Capitolul 6.

Unul din dezavantajele analizei seriilor nestaționare în două etape: staționarizare prin diferențiere sau eliminarea unui trend, urmată de analiza propriu-zisă este posibilitatea pierderii în acest mod a informațiilor asupra

dinamicii pe termen lung a seriilor de timp. Acest dezavantaj poate fi adresat prin utilizarea unor metode concepute special pentru analiza seriilor nestaționare. Constatarea empirică potrivit căreia multe serii de timp macroeconomice conțin o rădăcină unitate a impulsionat dezvoltarea teoriei analizei seriilor de timp de nestaționare.

Engle și Granger (1987) au subliniat faptul că o combinație liniară a două sau mai multe serii nestaționare poate conduce la o serie care este staționară. În cazul în care există o astfel de combinație liniară staționară, se spune că seriile nestaționare care intră în aceea combinație sunt **cointegrate**. Combinația liniară staționară se numește **ecuația de cointegrare** și poate fi interpretată ca o *relație de echilibru pe termen lung între variabile*.

Avantaje ale utilizării metodologiei cointegrării:

- permite analiza seriilor de timp nestaționare, fără a pierde informațiile legate de comportamentul acestora pe termen lung;
- cointegrarea stă la baza modelelor VEC, care includ atât relația pe termen lung, cât și mecanismul de corecție a erorilor.

Dezavantaje:

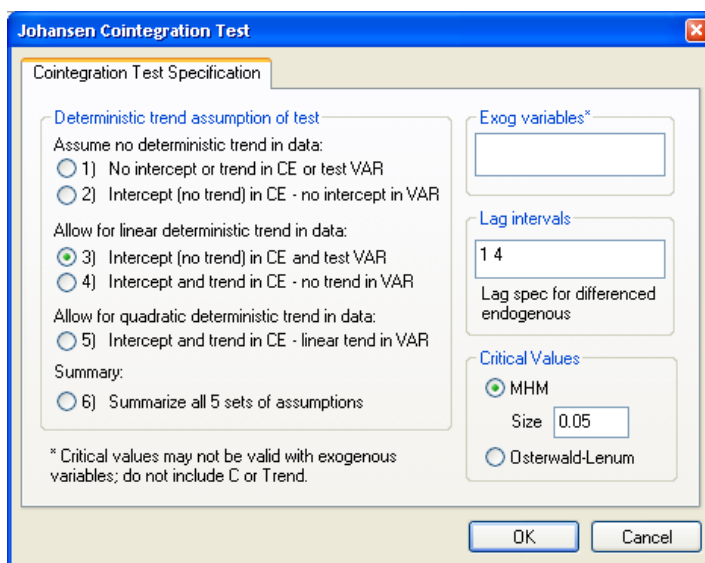
- necesită impunerea de restricții suplimentare pentru identificarea relației de cointegrare;
- metodologia dezvoltată pentru determinarea numărului de relații de cointegrare este relativ complexă;
- situațiile în care între variabile există mai multe relații de cointegrare sunt dificil de interpretat din punct de vedere economic.

8.1 Teste de cointegrare

Scopul unui **test de cointegrare** constă în a determina dacă un grup de serii nestaționare sunt cointegrate sau nu. După cum se explică în continuare, prezența unei relații de cointegrare constituie baza pentru modelele de tip VEC.

Metodologia Engle-Granger se utilizează, în special, pentru a testa existența cointegrării între două serii de date și constă în estimarea unei regresii între cele două variabile și testarea staționarității reziduurilor rezultate. Dacă seria de reziduuri este staționară, cele două serii sunt cointegrate.

Eviews implementează teste de cointegrare pe bază de modele VAR, folosind metodologia elaborată de către Johansen (1991, 1995). Pentru a efectua **testul de cointegrare Johansen**, se selectează **View/Cointegration Test...** din grup sau bara de instrumente a ferestrei VAR.



De reținut faptul că, deoarece discutăm de un test de cointegrare, acesta este valabil numai atunci când se lucrează cu serii despre care se știe că sunt nestaționare. Se pot aplica mai întâi teste ale rădăcinii unitate pentru fiecare serie din VAR. Pagina **Cointegration Test Specification** solicită informații despre test. În practică, cazurile 1 și 5 sunt rar folosite. Ar trebui utilizat cazul 1 numai dacă știți că toate seriile au medie zero. În practică, se utilizează cazul 2 dacă nici una din serii nu pare să aibă un trend. Pentru serii cu trend se utilizează cazul 3, dacă se consideră că toate trendurile sunt stocastice; dacă se consideră că unele serii au un

trend determinist, se utilizează cazul 4. Dacă nu există siguranță privind trendul care se poate folosi în ipoteză se poate alege opțiunea **Summary of all 5 trend assumptions** (cazul 6). Această opțiune indică numărul de relații de cointegrare sub fiecare din cele 5 ipoteze de trend, și apoi se va putea evalua gradul de sensibilitate al rezultatelor față de aceste ipoteze.

Date: 08/16/06 Time: 11:29
 Sample (adjusted): 1974Q3 1987Q3
 Included observations: 53 after adjustments
 Trend assumption: No deterministic trend (restricted constant)
 Series: LRM LRY IBO IDE
 Lags interval (in first differences): 1 to 1

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None	0.469677	52.71087	54.07904	0.0859
At most 1	0.174241	19.09464	35.19275	0.7814
At most 2	0.118083	8.947661	20.26184	0.7411
At most 3	0.042249	2.287849	9.164546	0.7200

Trace test indicates no cointegration at the 0.05 level
 * denotes rejection of the hypothesis at the 0.05 level
 **MacKinnon-Haug-Michelis (1999) p-values

Prima parte a tabelului cu rezultate prezintă numărul de relații de cointegrare. Două tipuri de teste statistice sunt raportate. Primul bloc prezintă așa-numitele statistici *trace* și al doilea bloc (care nu apare în figura de mai sus) prezintă statisticile *maximum eigenvalue*. Pentru a determina numărul de relații de cointegrare, condiționat de ipotezele făcute cu privire la trend, se analizează succesiv rezultatele de la $r = 0$ până la $r = K - 1$ până când nu se mai respinge ipoteza nulă asociată testului. Rezultatul acestei proceduri de testare secvențială este prezentat în partea de jos a fiecărui tabel cu rezultate.

A doua parte a ferestrei cu rezultate oferă estimări ale parametrilor din relațiile de cointegrare β și ale parametrilor de ajustare α . După cum se cunoaște, vectorul de cointegrare nu este identificat dacă nu se impun o anumită normalizare a acestuia în mod arbitrar. De obicei, se utilizează o normalizare definite în Johansen (1995). De reținut faptul că transpusa lui β este raportată prin selectarea opțiunii **Unrestricted Cointegrating Coefficients**, astfel încât pe primul rând este primul vectorul de cointegrare, pe al doilea rând este al doilea vector de cointegrare, și așa mai departe.

8.2 Modele tip VEC

Un model de tip **VEC** (Vector Error Correction) este un model VAR restricționat conceput pentru a fi utilizat în cazul seriilor nestaționare despre care se cunoaște că sunt cointegrate. Un model VEC are încorporat în carul structurii sale aceste relații de cointegrare, astfel încât urmărește să limiteze dinamica pe termen lung a variabilelor endogene astfel încât să convergă către relațiile lor de cointegrare, permițând în același timp ajustări dinamice pe termen scurt. Termenul care cuantifică relația de cointegrare este cunoscut ca termenul de corectare a erorilor (error correction), deoarece o eventuală abatere de la echilibrul pe termen lung este corectată treptat printr-o serie de ajustări parțiale pe termen scurt.

Coefficienții regresiei estimate dau relația de cointegrare, care se poate interpreta ca o relație pe termen lung între cele două variabile. Combinând relația pe termen lung cu un mecanism de ajustare pe termen scurt, se obține un model de tip VEC.

Dacă două variabile, y_{1t} și y_{2t} , respectă relația pe termen lung (sunt cointegrate) atunci $y_{1t} = \beta_1 y_{2t}$, $\beta_1 > 0$. Mecanismul de ajustare pe termen scurt conține răspunsul fiecărei variabile la abaterile de la relația pe termen lung:

$$\begin{cases} \Delta y_{1t} = \alpha_1 (y_{1t-1} - \beta_1 y_{2t-1}) + \gamma_{11}^1 \Delta y_{1t-1} + \gamma_{12}^1 \Delta y_{2t-1} + u_{1t} \\ \Delta y_{2t} = \alpha_2 (y_{1t-1} - \beta_1 y_{2t-1}) + \gamma_{21}^1 \Delta y_{1t-1} + \gamma_{22}^1 \Delta y_{2t-1} + u_{2t} \end{cases}$$

În relația de mai sus, coeficienții α_1 și α_2 reflectă viteza cu care cele două variabile se ajustează la relația pe termen lung. De exemplu, dacă $y_{1t-1} > \beta_1 y_{2t-1}$, ceea ce înseamnă că variabila y_{1t-1} este mai mare decât trebuie conform relației de echilibru, $\Delta y_{1t} < 0$, dacă $\alpha_1 < 0$, ceea ce readuce variabila la valori compatibile cu relația pe termen lung.

Matricial, relația de definiție a unui model VEC se scrie

$$\Delta y_t = \alpha \cdot \beta^T \cdot y_{t-1} + \Gamma_1 \Delta y_{t-1} + u_t,$$

unde $y_t = (y_{1t}, y_{2t})^T$, $\alpha = (\alpha_1, \alpha_2)^T$, $\beta = (\beta_1, \beta_2)^T$, $\Gamma_1 = \begin{pmatrix} \gamma_{11}^1 & \gamma_{12}^1 \\ \gamma_{21}^1 & \gamma_{22}^1 \end{pmatrix}$ și

$$u_t = (u_{1t}, u_{2t})^T.$$

Este de remarcat faptul că un model VEC se poate pune sub forma unui model VAR și invers:

$$y_t = (I + \Gamma_1 + \alpha \cdot \beta^T) y_{t-1} - \Gamma_1 y_{t-2} + u_t.$$

Deoarece modelul de tip VEC se aplică numai pentru serii cointegrate, este necesar să se ruleze mai întâi un test de cointegrare, așa cum este descris mai sus, și să se determine numărul de relații de cointegrare. Aceste informații trebuie furnizate ca parte a specificațiilor VEC.

În **Eviews**, pentru a configura un model VEC, se face clic pe butonul **Estimate** din bara de instrumente și se alege opțiunea **Vector Error Correction** din fila **VAR/VEC Specification**. În fila **VAR/VEC Specification** trebuie furnizate aceleași informații ca pentru un VAR nerestricționat, cu unele excepții:

- Termenul de trend constant sau liniar nu ar trebui să fie inclus în caseta de editare **Exogenous Series**. Specificațiile privind trendul pentru un model VEC trebuie să fie menționate în fila **Cointegration**.
- Specificațiile referitoare la lag-uri se referă la termenii constând în prima diferență a variabilelor și care cuantifică dinamica pe termen scurt din

VEC. Pentru a estima un VEC fără lag-uri în termenii constând în prima diferență, se specifică intervalul de lag "0 0".

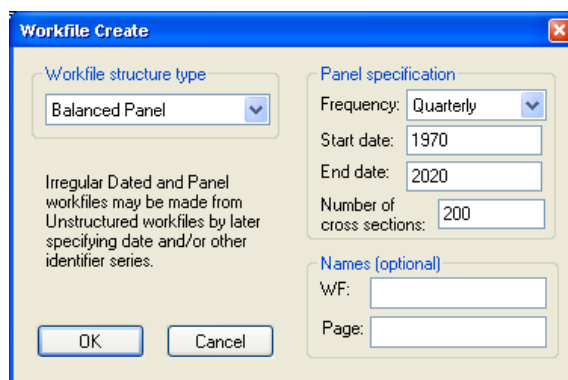
- Dacă se dorește impunerea de restricții cu privire la relațiile de cointegrare și/sau a coeficienților de ajustare, se utilizează fila **Restrictions**.

Fereastra care conține rezultatele estimării modelului VEC se compun din două părți. Prima parte raportează rezultatele procedurii Johansen, efectuată în prima etapă. Dacă nu există restricții impuse, **Eviews** va utiliza o normalizare implicită. Această normalizare implicită exprimă primele variabilele r din VEC, ca funcții de restul de $K - r$ variabile rămase, unde r este numărul relațiilor de cointegrare și K este numărul de variabile endogene. Pentru parametrii care sunt identificați în cadrul restricțiilor sunt raportate erorile standard asimptotice.

A doua parte a tabelului cu rezultate prezintă estimările, determinate în a doua etapă, privind modelul VAR, exprimat ca prima diferență a variabilelor analizate și incluzând termenii de corecție a erorilor estimate în prima etapă. Termeni de corecție a erorilor sunt notați CointEq1, CointEq2, și așa mai departe în cadrul rezultatelor. Această parte a rezultatului are același format ca și rezultatul unui VAR nerestricționat, după cum este descris în capitolul anterior.

9 Tehnici de analiză a modelelor de tip panel data

Datele de tip panel sau pool implica observații care posedă atât identificatori pentru secțiuni transversale, cât și privind evoluția în timp a acestora. Analiza acestui tip de date se realizează în **Eviews** în fișiere de lucru de tip panel.



Există două moduri de bază pentru a crea un fișier de lucru structurat cu date de tip panel. În primul rând, se poate crea un *nou* fișier de lucru care are o structură de panel. Pur și simplu se selectează opțiunea **File/New/ Workfile...** din meniul principal pentru a deschide căsuța de dialog **Workfile Create**. În continuare, se selectează **Balanced Panel** din lista mobilă **Workfile structure type** și se completează căsuța de dialog după cum se dorește.

O a doua metodă de structurare a fișierului de lucru cu date de tip panel este să se introducă prima dată date stivuite într-un fișier de lucru nestructurat și apoi să se aplice o structură pentru a crea un fișier de lucru de tip panel. Pentru a structura un fișier existent, se selectează **Proc/“Structure/Resize Current Page...”** din fereastra principală de lucru. **Eviews** deschide o căsuță de dialog **Workfile**

structure. Structura de bază a căsuței de dialog este destul de similară cu cea a căsuței de dialog **Workfile create**. În partea stângă se găsește o listă mobilă, din care se va selecta un tip de structură. Cea mai simplă metodă pentru definirea unei structuri de tip panel cu frecvență regulată este de a selecta **Dated - regular frequency**. Partea dreaptă din căsuța de dialog se modifică pentru a reflecta alegerea făcută la pasul precedent, solicitându-se descrierea structurii de date.

Clasa de bază a modelelor ce pot fi estimate folosind instrumente pentru date de tip panel poate fi scrisă astfel:

$$Y_{it} = \alpha + X'_{it}\beta + \delta_i + \gamma_t + \varepsilon_{it}$$

unde Y_{it} este variabila dependentă, X_{it} este un vector k dimensional de regresori și ε_{it} sunt inovațiile pentru M unitățile transversale și observate pentru T perioade. Termenii δ_i și γ_t reprezintă efectele specifice (aleatoare sau fixe) pentru unități ale secțiunii transversale sau pentru anumite perioade de timp.

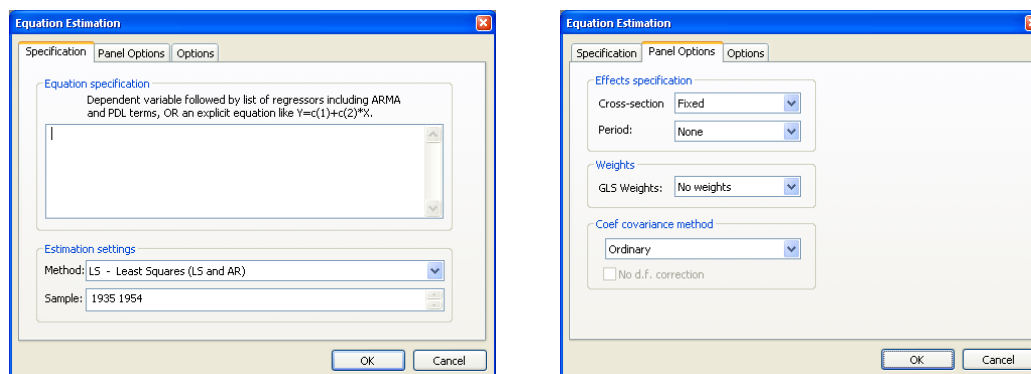
Prezența efectelor specifice transversale sau temporale poate fi surprinsă și analizată utilizând tehnici pentru efecte fixe și pentru efecte aleatoare. Se pot specifica modelele ce conțin efecte într-una sau ambele dimensiuni, de exemplu, un efect fix în dimensiunea secțiunii transversale, un efect aleator în dimensiunea perioadei sau un efect fix în secțiunea transversală și un efect aleatoriu în dimensiunea perioadei. Trebuie evidențiat faptul că, totuși, cele cu efecte aleatorii în ambele dimensiuni pot fi estimate numai în cazul în care panelul este echilibrat, astfel încât fiecare secțiune transversală are același set de observații temporale.

Specificațiile cu **efecte fixe** sunt tratate folosind o abordare simplă care constă în eliminarea mediei variabilei dependente la nivel transversal sau temporal și apoi utilizarea unei ecuații de regresie utilizând datele rezultate. Specificațiile cu **efecte aleatoare** presupun că efectele corespunzătoare δ_i și γ_t sunt realizări ale unor variabile aleatoare independente cu medie zero și varianță finită. Cel mai important, specificația bazată pe efecte aleatoare presupune faptul că efectul

specific este necorelat cu inovațiile ecuației. **Eviews** prelucrează modele cu efecte aleatoare folosind tehnici de tip FGLS.

Eviews permite estimarea ecuațiilor de tip panel utilizând metoda OLS sau metoda variabilelor instrumentale, cu corecții pentru efectele fixe sau aleatoare, atât în dimensiunea secțiunii transversale, cât și în cea temporală, erorile de tip AR și erorile standard robuste.

Primul pas în estimarea unei ecuații de tip panel este construirea unui obiect *equation* prin utilizarea opțiunii **Object/New Object.../Equation** sau **Quick/Estimate Equation...** din meniul principal. **Eviews** va detecta prezența structurii de tip panel și în loc de căsuța de dialog a ecuației de regresie standard se va deschide căsuța de dialog specifică structurii de tip panel. Trebuie deschisă lista mobilă **Method** pentru a alege metoda de estimare, fiind disponibile **LS - Least Squares (and AR)**, adică metoda OLS, **TSLS - Two-Stage Least Squares (and AR)**, adică metoda variabilelor instrumentale, precum și **GMM / DPD - Generalized Method of Moments / Dynamic Panel Data** utilizată în cazul în care în partea dreaptă a ecuației se găsesc lag-uri ale variabilei dependente.



Căsuța de dialog pentru estimarea OLS conține mai multe pagini în care trebuie introduse specificația modelului, opțiunile de estimare de tip panel și opțiunilor generale de estimare. Specificația ecuației se introduce în caseta

Equation specification și mărimea eșantionului în caseta **Sample**. În general, marea majoritate a specificațiilor permise în cadrul ecuațiilor de regresie standard poate fi utilizate și în cazul datelor panel. Se pot, de exemplu, include termeni AR. Totuși, nu se pot include termeni MA într-un model cu date panel.

Apoi, în cadrul filei **Panel Options** se specifică setări specifice de estimare. În primul rând, trebuie luate în considerare efectele specifice transversale sau temporale folosind lista mobilă **Effects specification**. În mod implicit, **Eviews** presupune că nu există efecte specifice, astfel încât ambele casete combo sunt setate la **None**. Se pot schimba setările implicite pentru a permite, fie efecte fixe (**Fixed**) sau aleatoare (**Random**), fie în secțiunea transversală, fie în dimensiunea perioadei, fie în ambele.

De asemenea, trebuie să se specifice o metodă de calcul pentru matricea de varianță-covarianță a estimatorilor. Se poate utiliza lista mobilă **Coef covariance method**, pentru a selecta una din diversele metode robuste disponibile pentru calcularea erorilor standard. Modul de calcul al covarianței poate fi ales să fie robust sub diferite ipoteze, de exemplu, corelarea generală a observațiilor într-o secțiune transversală sau heteroskedasticitate la nivel transversal.

Eviews oferă instrumente pentru testarea semnificației statistice a estimărilor privitoare la efectele specifice fixe. Pentru a testa semnificația efectelor trebuie estimat în primul **View/Fixed/Random Effects Testing/Redundant Fixed Effects – Likelihood Ratio**. **Eviews** va estima specificațiile corespunzătoare restricționate și va afișa rezultatele testului statistic, precum și rezultatele specificațiilor restricționate.

O ipoteza centrală în estimarea specificațiilor care conțin efecte aleatoare constă în faptul că aceste efecte sunt necorelate cu variabilele explicative. O metodă des utilizată pentru testarea acestei ipoteze este utilizarea **testului Hausman** pentru a compara estimările coeficienților obținute prin estimare cu efecte fixe și prin estimarea cu efecte aleatoare. Pentru a efectua testul Hausman, trebuie estimat în primul rând un model cu efecte aleatoare. În continuare, se

selectează **View/Fixed/Random Effects Testing/Correlated Random Effects-Hausman Test**. **Eviews** estimează în mod automat specificațiile corespunzătoare bazate pe efecte fixe, calculează statisticile testelor și afișează rezultatele și ecuațiile auxiliare.

Interesul din ce în ce mai mare în ceea ce privește datele panel, precum și disponibilitatea crescândă a acestora a condus la extinderea a diferite teste statistice pentru date panel. Literatura de specialitate recentă s-a concentrat asupra testelor de cointegrare în cadru panel. **Eviews** poate calcula mai multe tipuri de teste de cointegrare în panel. Pentru a realiza un test de cointegrare cu date panel în **Eviews**, se deschide un grup care conține seriile de interes și se selectează **Views/Cointegration Test...** pentru a afișa caseta de dialog în care se specifică parametrii utilizați în relația de cointegrare.

10 Tehnici de estimare a modelelor de tip DSGE

10.1 Structura de bază a unui program Dynare

Rezolvarea și estimarea modelelor dinamice de echilibru general stocastic (DSGE) nu se poate realiza în **Eviews** presupunând utilizarea unor pachete software specializate cum ar fi **Dynare** care necesită însă cunoștințe avansate de utilizare a mediului de programare **Matlab**. Prezentăm în continuare **structura de bază a unui program de tip Dynare**, care poate fi utilizat pentru estimarea unui model de tip DSGE. Un astfel de program este constituit din mai multe secțiuni, denumite blocuri.

În cadrul primului bloc trebuie specificate care sunt variabilele exogene, care sunt variabilele endogene și care sunt variabilele predeterminate în cadrul modelului respectiv. De asemenea, trebuie menționați care sunt parametrii modelului.

În cadrul celui de-al doilea bloc trebuie să se specifice valorile parametrilor sau modul de estimare al acestora. Pentru estimarea parametrilor unui model DSGE se pot utiliza, de obicei, trei metode: **GMM** (Generalized Method of Moments), **MLE** (Maximum Likelihood Estimation), precum și o **metoda bazată pe tehnici bayesiene**.

Al treilea bloc conține descrierea ecuațiilor modelului. Acest bloc trebuie să înceapă cu cuvântul cheie „model” și să se finalizeze prin cuvântul cheie „end”. Ecuațiile modelului pot fi specificate fie în forma brută rezultată din rezolvarea problemelor de optimizare a agenților economici, fie în formă liniară. Cea de-a doua specificație presupune însă din partea cercetătorului un efort suplimentar, datorită faptului că ecuațiile brute trebuie **log-liniarizate** în jurul punctului de echilibru pe termen lung. Operațiunea de log-liniarizare presupune dezvoltarea în

serie Taylor a logaritmului variabilelor de interes în jurul valorii de echilibru pe termen lung și concentrarea atenției asupra primului termen din această serie.

Al patrulea bloc definește valorile inițiale ale variabilelor din cadrul modelului. Acest bloc trebuie să înceapă cu cuvântul cheie „initval” și să se finalizeze prin cuvântul cheie „end”. Aceste valori inițiale trebuie alese cu grijă, astfel încât să aibă sens economic, deoarece au o influență foarte mare asupra valorilor de echilibru pe termen lung, care sunt calculate de către program prin metode numerice de rezolvare a sistemelor de ecuații neliniare. Alternativ, pot fi specificate în mod direct valorile de echilibru pe termen lung al unor variabile. În acest caz trebuie utilizat cuvântul cheie „endval” în locul cuvântului cheie „initval”.

În cadrul următorului bloc sunt specificate care din variabile joacă rolul șocurilor exogene care influențează dinamica sistemului. Deoarece avem de a face cu variabile aleatoare, trebuie să se specifice distribuția acestora, în special varianța fiecărui șoc, precum și corelația dintre acestea. Acest bloc trebuie să înceapă cu cuvântul cheie „shocks” și să se finalizeze prin cuvântul cheie „end”.

În cadrul următoarelor două blocuri trebuie să se specifice metoda prin care vor fi determinate valorile de echilibru pe termen lung, respectiv metoda cu ajutorul căreia va fi rezolvat și simulat modelul DSGE supus analizei.

10.2 Operațiunea de log-liniarizare

Aproximarea log-liniară a relațiilor de optimalitate asociate unui model DSGE reprezintă un caz particular al unei metode mai generale, respectiv metoda perturbațiilor. Log-liniarizarea unui model DSGE este preferabilă în cazul în care relațiile acestuia sunt „ușor” neliniare, nefiind necesare corecții de tip convexitate. Aproximarea log-liniară conduce, de asemenea, o interpretare intuitivă a variabilelor din cadrul aproximării, respectiv deviații procentuale față de o valoare pe termen lung.

Pentru a exemplifica tehnica de log-liniarizare vom pleca de la o relație de optimalitate, des întâlnită în modelele DSGE

$$\frac{u_c}{u_N} = \frac{1}{F_N}$$

Pentru a nu crește inutil dificultatea, folosim o funcție de utilitate logaritmică $u(c, N) = \ln(c) + \ln(1 - N)$ și o funcție de producție de tip Cobb-Douglas $F(K, N) = AN^\alpha K^{1-\alpha}$. Ca urmare, relația de optim de mai sus se scrie sub forma:

$$\frac{1 - N}{c} = \frac{1}{A\alpha N^{\alpha-1} K^{1-\alpha}}$$

prin logaritizarea căreia avem că:

$$\ln(1 - N) - \ln(c) + \ln(\alpha) - (1 - \alpha)[\ln(N) - \ln(K)] + \ln(A) = 0$$

În continuare, dezvoltăm în **serie Taylor**, în jurul punctului de echilibru pe termen lung sau **steady-state** (notat cu o bară deasupra variabilei de interes), fiecare termen al relației de mai sus și păstrăm doar primul termen al acestei dezvoltări.

Astfel, avem că:

$$\ln(1 - N) \approx \ln(1 - \bar{N}) - \frac{1}{1 - \bar{N}}(N - \bar{N}) = \ln(1 - \bar{N}) - \frac{\bar{N}}{1 - \bar{N}} \left(\frac{N - \bar{N}}{\bar{N}} \right) = \ln(1 - \bar{N}) - \frac{\bar{N}}{1 - \bar{N}} \hat{N}$$

$$\ln(c) \approx \ln(\bar{c}) + \frac{1}{\bar{c}}(c - \bar{c}) = \ln(\bar{c}) + \hat{c}$$

$$\ln(N) \approx \ln(\bar{N}) + \frac{1}{\bar{N}}(N - \bar{N}) = \ln(\bar{N}) + \hat{N}$$

$$\ln(K) \approx \ln(\bar{K}) + \frac{1}{\bar{K}}(K - \bar{K}) = \ln(\bar{K}) + \hat{K}$$

$$\ln(A) \approx \ln(\bar{A}) + \frac{1}{\bar{A}}(A - \bar{A}) = \ln(\bar{A}) + \hat{A}$$

Ca urmare, relația optim de mai sus se poate rescrie utilizând deviații procentuale față de steady-state (notate cu căciuliță deasupra variabilei de interes):

$$-\frac{\bar{N}}{1-\bar{N}}\hat{N}-\hat{c}-(1-\alpha)[\hat{N}-\hat{K}]+\hat{A}=0$$

Bineînțeles că procedura prezentată mai sus este o aproximare „bună” a realității doar în cazul în care abaterea procentuală față de steady-state nu este „prea mare” sau altfel spus poate fi utilizată în situații „normale”. Totuși, în situații excepționale, de „criză”, abaterile față de valorile de echilibru pe termen lung pot să înregistreze valori „mai mari”, caz în care trebuie să se apeleze la metoda mai generală a perturbațiilor.

10.3 Noțiuni elementare privind econometria bayesiană

Regula lui Bayes, descoperită în secolul 17 de către preotul și matematicianul englez Thomas Bayes, oferă o reprezentare alternativă a probabilităților condiționate și reprezintă o noțiune fundamentală care stă la baza econometriei bayesiene. În cazul simplu, care se referă la două evenimente A și B ,

regula lui Bayes este dată prin relațiile $P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B)+P(A|\bar{B})P(\bar{B})}$ și

$$P(\bar{B}|A) = \frac{P(A|\bar{B})P(\bar{B})}{P(A|B)P(B)+P(A|\bar{B})P(\bar{B})}.$$

Pentru a surprinde intuiția regulii lui Bayes să observăm pentru început că spațiul stărilor analizat este partiționat într-o colecție de evenimente disjuncte care ale căror probabilități diferite de zero sunt cunoscute, în cazul de față B și \bar{B} . De asemenea, se consideră un eveniment A a cărui probabilități condiționate de fiecare dintre evenimente din partiție sunt cunoscute. Având în vedere această informație de fond, regula lui Bayes oferă mijloacele pentru reevaluarea probabilităților evenimente din partiția spațiului stărilor având în vedere informația suplimentară că evenimentul A are loc.

Astfel, probabilitățile evenimentelor din partiție sunt, de fapt, "actualizate" în lumina noilor informații furnizate de apariția evenimentului A . Această interpretare a regulii Bayes a dus la utilizarea termenilor de **probabilitate apriori**

care se referă la probabilitatea calculată înainte de a avea informații cu privire la apariția evenimentului A , respectiv **probabilitate aposteriori** care se referă la probabilitatea „actualizată”, calculată ca o probabilitate condiționată, după ce există informații cu privire la apariția evenimentului A

Bibliografie selectivă

- Baltagi, B.H., (2005), *Econometric Analysis of Panel Data*, 3rd Edition, John Wiley & Sons
- Bobeica G., (2010), *Econometria seriilor de timp*, note de curs, Master DOFIN
- Brooks, C., (2008), *Introductory Econometrics for Finance*, 2nd Edition, Cambridge University Press
- Canova, F., (2007), *Methods for Applied Macroeconomic Research*, Princeton University Press
- Enders, W., (2004), *Applied Econometric Time Series*, 2nd Edition, Wiley
- Greene, W.H., (2008), *Econometric Analysis*, 6th Edition, Prentice Hall
- Hamilton, J., (1994), *Time Series Analysis*, Princeton University Press
- Lutkepohl, H. și M. Kratzig, (2004), *Applied Time Series Econometrics*, Cambridge University Press
- Maddala, G.S., (2001) *Introduction to Econometrics*, John Wiley & Sons.
- Mittelhammer, R.C., (1996), *Mathematical Statistics for Economics and Business*, Springer
- Necula C., (2010a), *Bazele econometriei*, note de curs, Master DOFIN
- Necula C., (2010b), *Econometria piețelor de capital*, note de curs, Master DOFIN
- Necula C., (2010c), *Econometrie avansată*, note de curs, Master BPM
- Wooldridge, J., (1999), *Introductory Econometrics: A Modern Approach*, South-Western College Pub
- ***, *Eviews 6 User Guide*
- ***, *Dynare User Guide*