

Evaluarea globală a portofoliilor de garanții bancare

- Recomandări -

Introducere

Experții evaluatori apelează, în prezent, la diverse metode statistice pentru evaluarea proprietăților imobiliare, o multitudine de soluții pe care statistica matematică le oferă pentru a fi aplicate și în acest domeniu: MRA - analiza prin regresie multiplă, serii de timp, tehnici de logică fuzzy, ANN - rețele neuronale artificiale, NFS - sisteme neuro-fuzzy și altele.

Pe plan internațional, din ambele zone de expertiză - mediul academic (universități, instituții de cercetare) și industrie (bănci, firme de evaluatori, agenții imobiliare), emerg soluții bazate pe statistica matematică și, chiar mai mult, simbioze ale acestora cu Sisteme Informatică Geografice (GIS) sau cu aplicații web.

Avantajele imediate și foarte evidente ale aplicării acestor soluții în evaluare sunt automatizarea procesului de evaluare și existența metodelor de control ale rezultatelor obținute cu acestea.

Dezavantajul principal constă în complexitatea aplicării soluțiilor propuse, fiind necesar un colectiv de specialiști din diferite domenii - programatori, statisticieni, matematicieni, evaluatori, analiști de piață - pentru a concepe și pune în funcțiune sisteme de asemenea natură și anvergură.

Întrucât în practică soluțiile pot fi multiple, prezentele recomandări privind evaluarea globală a portofoliilor de garanții bancare au caracter exemplificativ.

Modele automate de evaluare

Crearea și utilizarea unui **AVM** – model automat de evaluare (Automated Valuation Model) poate fi o soluție pentru evaluarea portofoliilor de garanții imobiliare deținute de către bănci. AVM-urile sunt soft-uri ce produc estimatori ai valorilor de piață pentru proprietatea/proprietățile imobiliare subiect, în baza analizei condițiilor de piață și a caracteristicilor proprietăților imobiliare comparabile, informații de piață colectate în prealabil. AVM sunt aplicabile în special în cazul evaluării proprietăților imobiliare prin utilizarea abordării prin piață.

Credibilitatea unui AVM și acuratețea rezultatelor obținute prin utilizarea acestuia depind de **cantitatea** și **calitatea** datelor folosite în evaluare, de experiența și pregătirea colectivului care proiectează și dezvoltă modelul. Prin cantitatea datelor se înțelege volumul eșantionului de date utilizat în evaluare. În ceea ce privește calitatea datelor, în cadrul procesului de elaborare a unui AVM este desemnată o etapă specială pentru managementul și analiza calității datelor (a se vedea etapa a 3-a).

Având în vedere imposibilitatea utilizării întregii populații de date (formată din proprietățile imobiliare identificate, care sunt oferite la vânzare la data evaluării sau tranzacționate în condiții de piață identice), este necesar să se preleveze un eșantion de date reprezentativ pentru întreaga populație. Un eșantion este reprezentativ pentru întreaga populație dacă structura lui este identică sau foarte asemănătoare cu structura populației din care a fost extras, astfel încât să fie posibilă extrapolarea concluziilor finale la toată populația din care face parte. Eșantionul va fi reprezentativ dacă se utilizează un procedeu de eșantionare obiectiv, aleator, cu probabilitate anticipată și dacă eșantionul are un volum suficient de mare de date.

Notă: Pentru proprietățile imobiliare ce formează eșantionul analizat trebuie să se cunoască atât prețurile (de tranzacționare și/ sau ofertă), cât și caracteristicile acestora.

1. Procedee de eșantionare

Cel mai utilizat mod de obținere a datelor necesare analizei statistice este în prezent **sondajul statistic**.

Sondajul statistic poate fi de tip probabilist (sau aleator) și de tip non-probabilist (sau nealeator). Diferențele dintre cele două tipuri de sondaje constau în faptul că, la procedeele probabiliste, selectarea unităților necesare pentru eșantionare este aleatoare, sub acest aspect fiecare unitate are o probabilitate cunoscută p și diferită de zero de a se afla în eșantion, în timp ce procedeele non-probabiliste selectează unitățile în mod judicios pentru ca eșantionul să prezinte cât mai bine caracteristicile esențiale ale populației din care este extras, dar nu se poate cunoaște probabilitatea unității de a se afla în acel eșantion.

Atenție! *Din motivele enumerate mai sus, în evaluarea proprietăților imobiliare prin metode statistice nu se recomandă sondajul statistic prin metode non-probabiliste.*

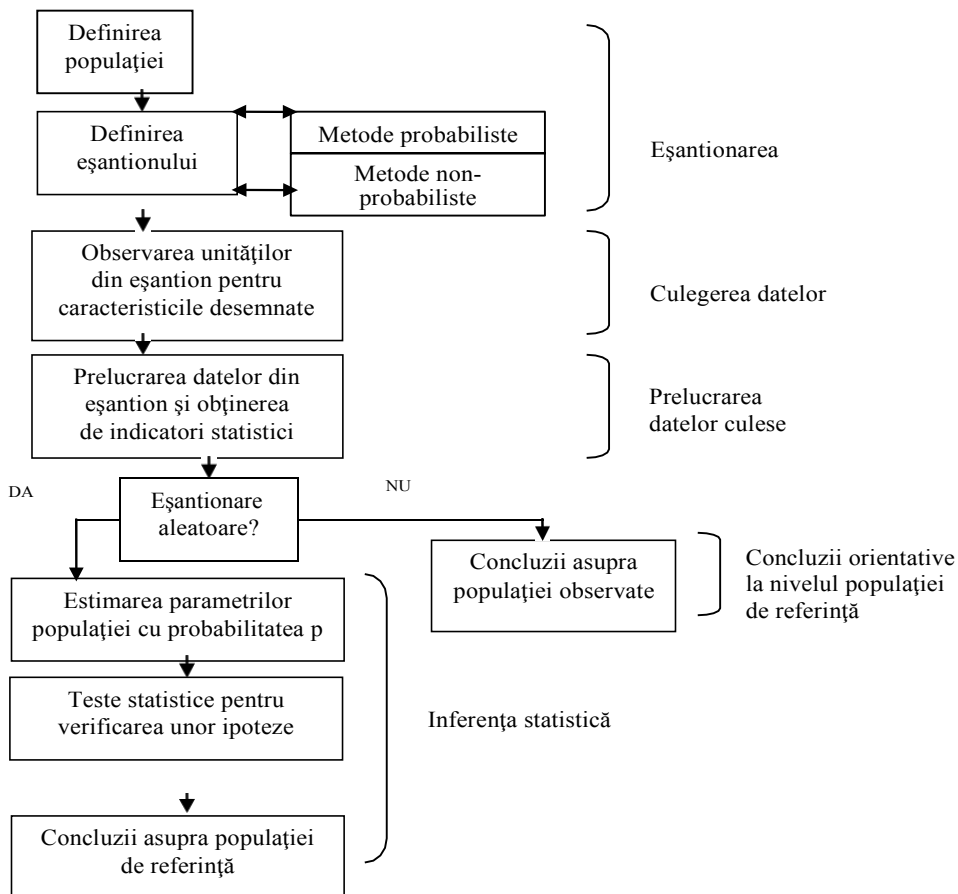


Fig. 1. Etapele cercetării prin sondaj statistic¹

Procedee probabiliste de eșantionare

a. *Eșantionarea simplă* se pretează, în special, pentru populații omogene formate din unități simple; eșantionul poate fi selectat aplicând procedeul prin revenire (procedeul bilei revenite) sau fără revenire (procedeul bilei nerevenite). Prin revenire, unitatea selectată la un moment dat se plasează înapoi în populație, astfel încât fiecare unitate a populației are șansa de a fi realesă la fiecare selecție, deci probabilitatea este constantă. Prin selectare fără revenire, unitatea aleasă nu mai este returnată în populație, probabilitatea fiind variabilă, în sensul creșterii ei pe măsura formării eșantionului.

Dintre cele două procedee, cea mai adecvată în cazul evaluării

¹ *Statistică: teorie și aplicații*, T. Andrei, S. Stancu, D. Traian Pele, Editura Economică, ediția a doua 2002.

proprietăților imobiliare ar fi eșantionarea fără revenire². Astfel avem garanția că toate unitățile eșantionului vor fi diferite, așa cum este necesar în cazul evaluării, neputând determina valoarea de piață a unei proprietăți imobiliare prin utilizarea unui eșantion care cuprinde o proprietate imobiliară de mai multe ori. Ambele procedee au însă o deficiență majoră, și anume, aceea că selecția este condiționată de abilitatea prin care sunt amestecate elementele populației. Prin urmare, în general, eșantionarea simplă nu se aplică³ în evaluare.

b. Eșantionarea cu tabele de numere aleatoare utilizează tabele Yates sau Kendell. Se numerotează toate elementele populației și apoi se extrage eșantionul pe baza tabelelor.

c. Eșantionare sistematică (mecanică) constă în ordonarea elementelor după o caracteristică, formarea unei liste, împărțirea populației în grupe de câte k unități (volume egale). k reprezintă pasul de numărare și este $\frac{N}{n}$, iar punctul de start este $j \leq k$ (N - volumul populației, n -

volumul eșantionului).

d. Eșantionarea stratificată. Populația de date este divizată în straturi cu caracteristici comune, ca de exemplu: proprietăți imobiliare rezidențiale – apartamente, vile, case ș.a.; proprietăți imobiliare industriale; proprietăți imobiliare comerciale – spații birouri, magazine ș.a.; proprietăți imobiliare agricole – terenuri arabile, livezi, ferme agro-zootehnice ș.a.; proprietăți imobiliare speciale. În vederea construirii eșantionului reprezentativ al populației la nivelul fiecărui strat se aplică eșantionarea aleatoare simplă.

Acest tip de eșantionare se pretează cel mai bine domeniului imobiliar și este deja utilizat de către evaluatori pentru eșantionarea manuală a datelor de piață necesare în evaluare, de exemplu, în funcție de vârsta clădirilor, de dimensiunile proprietății sau de locație.

Prin stratificare se asigură un grad mai ridicat de comparabilitate a proprietăților comparabile cu proprietatea subiect și se pot elimina erorile sistematice sau de selectare inadecvată a acestora. Pe de altă parte, o stratificare exagerată a datelor poate duce la imposibilitatea aplicării analizei statistice a datelor.

Se realizează în mai multe etape:

1. alegerea variabilei (variabilelor) de control (tipuri de proprietăți imobiliare și, eventual, diverse caracteristici ale acestora) după care se va face stratificarea;

² *Business Statistics in Practice*, Bruce L. Bowerman, Richard T. O'Connell, McGraw Hill, 4th edition, 2007, pg. 5.

³ *Statistică aplicată pentru modelarea evaluării imobiliare*, Iuliu Pașca, Editura Politehnica, 2008, pg. 16.

2. definirea straturilor în care se divide populația;
3. alegerea numărului de straturi în funcție de mărimea și natura populației, dar nu mai mult de 7-8 pe caracteristică/criteriu de clasificare⁴;
4. repartizarea volumului eșantionului pe straturile definite anterior:
 - a. fie proporțional cu numărul de unități din același strat al populației (numărul de unități/strat de eșantion să fie proporțional cu numărul de unități/același strat din populație):

$$p_i = \frac{N_i}{N} \text{ deci, } n_i = n \cdot p_i$$

în care :

N_i – numărul de unități din populația de referință, din stratul i

n – volumul eșantionului

N – volumul populației

n_i – numărul de unități din eșantion, în stratul i .

- b. fie optim (disproporțional), ținând seama atât de ponderea fiecărui strat în volumul total al populației, cât și de varianța fiecărui strat:

$$n_i = n \cdot \frac{N_i \cdot \sigma_i^2}{\sum_{i=1}^k N_i \cdot \sigma_i^2}, \text{ cui}=1, k$$

în care :

σ_i – abaterea standard a variabilei după care se face stratificarea, estimată pentru stratul i .

5. pentru fiecare strat se generează în mod aleatoriu unitățile care se vor include în eșantion.

Atenție! *O eroare relativă de eșantionare, situată sub 5%, ne permite să apreciem că eșantionul este reprezentativ:*

$$\frac{|\bar{x} - m|}{m} \cdot 100 \leq 5\%$$

unde:

\bar{x} - media variabilei analizate la nivelul eșantionului (valoarea de piață);

m - media variabilei analizate la nivelul populației.

⁴ *Statistica. Teorie și aplicații*, Tudorel Andrei, Stelian Stancu, Daniel Traian Pele, Editura Economică, ediția a doua, 2002, pg. 209.

2. Volumul eșantionului

Modelul automat de evaluare presupune realizarea analizei la nivelul unui eșantion de proprietăți imobiliare și generalizarea rezultatelor obținute (prin inferență statistică) pentru întreaga populație studiată. În acest sens este necesară stabilirea mărimii (volumului) adecvate a eșantionului pe baza căruia se va realiza analiza și evaluarea globală a portofoliului de garanții bancare.

Volumul n al eșantionului este influențat de:

- eroarea limită maximă admisă de reprezentativitate a eșantionului e , de 0,3;
- nivelul de semnificație α , prestabilit uzual la 5%;
- dispersia (varianța) caracteristicii în populație σ^2 ;
- volumul populației N .

Astfel, volumul eșantionului format din caracteristici măsurabile se poate determina aplicând două formule:

- pentru volume mici de eșantioane se utilizează statistica t :

$$n = \frac{t^2 \cdot \sigma^2}{e^2}$$

în care: $t_{\alpha/2}$ reprezintă statistica Student din tabele statistice t ,

corespunzător lui $\alpha/2$.

- pentru volume mari de eșantioane se folosește statistica z :

$$n = \frac{z^2 \cdot \sigma^2}{e^2}$$

$z_{\alpha/2}$ se regăsește în tabelele statistice z , corespunzător lui $\alpha/2$.

Atenție! Abaterea standard a populației, σ , de cele mai multe ori nu este cunoscută, prin urmare se poate folosi un estimator al acesteia, cum este abaterea standard s a unui eșantion preliminar, înlocuindu-se în formulele anterioare σ cu s , sau folosind σ al unei populații similară cu cea în cauză.

Atenție! Dacă n rezultat nu este o valoare rotundă, se rotunjește la următorul număr întreg.

3. Procesul de elaborare a unui model automat de evaluare

Procesul de elaborare a unui model automat de evaluare este prezentat în *Standardul pentru modele automate de evaluare*⁵ elaborat de către *International Association of Assessing Officers (IAAO)*. Acesta este un proces complex care necesită o colaborare strânsă între evaluatori, analiști ai pieței imobiliare, statisticieni și dezvoltatori de soft-uri.

Pornind de la recomandările făcute în acest standard propunem următoarea succesiune de etape necesare pentru construirea modelului automat de evaluare:

1. identificarea proprietăților subiect (portofoliul de garanții bancare de evaluat);
2. stabilirea ipotezelor și ipotezelor speciale;
3. managementul și analiza calității datelor cuprinse în eșantion;
4. stratificarea eșantionului (în cazul în care acesta nu a fost construit prin eșantionare stratificată);
5. stabilirea specificațiilor modelului;
6. calibrarea modelului;
7. testarea modelului și asigurarea calității sale;
8. validarea modelului;
9. aplicarea modelului;
10. verificarea periodică a acurateții modelului.

Primele două etape se realizează identic ca în cazul evaluării prin procesul clasic.

Etapa a 3-a. Managementul și analiza calității datelor cuprinse în eșantion

Analistul de piață trebuie să utilizeze instrumente statistice pentru a „scana” în mod sistematic și iterativ datele de piață, în vederea determinării lipsurilor, erorilor, incoerențelor, valorilor extreme (valori aberante) existente în seria de date analizată (analiza calității datelor).

În sprijinul acestei etape, softurile de analiză statistică oferă **tehnici grafice** și **tehnici cantitative** de explorare a seriilor de date. Aceste tehnici pot fi utilizate în scopul stabilirii gradului de reprezentativitate a eșantionului, pentru identificarea funcției de distribuție a frecvențelor (normală, uniformă, χ^2 , lognormală, Student, Fisher, Poisson etc.) și a caracteristicilor acesteia sau pentru detectarea valorilor aberante ale seriei analizate.

⁵ *Standard on Automated Valuation Models (AVMs)*, IAAO, 2003.

Astfel, o serie de date poate fi reprezentată sugestiv și apoi analizată vizual foarte rapid prin intermediul unei game variate de grafice statistice. În funcție de informațiile grafice pe care dorim să le obținem și de tipul variabilelor modelului (a se vedea etapa a 5-a), se pot folosi:

- pentru variabile cantitative: histograma, poligonul frecvențelor, steam-and-leaf, diagrama de împrăștiere (engl. scatter plots), box-and-wiskers, graficul de normalitate (engl. normal plot);
- pentru variabile calitative: diagrama în batoane, diagrama circulară.

Tehnicile cantitative de analiză a datelor constau în teste statistice (Student, Fisher, ANOVA, Chi-square, Shapiro-Wilk, Kolmogorov-Smirnov, Man-Withney, Durbin-Watson) și diverși indicatori statistici cum ar fi indicatorii tendinței centrale (medie, mediană, mod), respectiv indicatorii dispersiei datelor în jurul mediei (abaterea medie pătratică, varianța, coeficientul de variație ș.a.).

Notă: *Pentru luarea deciziei de stratificare a eșantionului este utilă determinarea gradului de omogenitate al acestuia. În acest sens se poate calcula coeficientul de variație (Cv), calculat ca raport între abaterea medie pătratică și media variabilei după care se va face stratificarea. Se poate considera că eșantionul este omogen dacă mărimea coeficientului de variație este mai mică sau egală cu 0,35. În cazul în care este depășită această limită se recomandă identificarea cauzelor eterogeneității eșantionului (existența unor valori extreme sau tipuri de proprietăți cu caracteristici diferite – ce formează straturile eșantionului).*

Etapa a 4-a. Stratificarea

Este procesul prin care proprietățile imobiliare care formează un eșantion eterogen, sunt grupate pe tipuri de proprietăți imobiliare în funcție de caracteristicile fizice și economice ale acestora, în vederea minimizării diferențelor din interiorul unui strat și a maximizării diferențelor dintre straturi.

Notă: *După această etapă, la nivelul fiecărui strat creat în cadrul eșantionului analizat se poate estima media valorilor de piață ale proprietăților, în scopul determinării indicelui de creștere/descreștere a valorii de piață a portofoliului de garanții bancare. Pentru estimarea acestui indice propunem următoarea formulă (derivată din modul de calcul a indicelui agregat ponderat de tip Paasche):*

$$Indice = \frac{\sum_{i=1}^m p_{i,t} \cdot \mu_{i,t}}{\sum_{i=1}^m p_{i,t} \cdot \mu_{i,t-1}} \cdot 100$$

unde:

$p_{i,t}$ = ponderea fiecărui strat în eşantionul analizat (a se vedea Eşantionarea stratificată);

μ = media valorilor de piaţă ale proprietăţilor din fiecare strat;

m = numărul de straturi din eşantionul analizat;

t = momentul actual;

$t-1$ = momentul la care s-a făcut ultima reevaluare.

Cu ajutorul acestui indice se poate identifica existenţa unei eventuale diminuări a valorii portofoliului de garanţii bancare, caz în care se dovedeşte necesară reevaluarea acestuia cu ajutorul unui AVM şi, ca urmare, parcurgerea următoarelor etape.

Etapa a 5-a. Stabilirea specificaţiilor modelului

Această etapă se referă la procesul de proiectare a structurii optime a AVM-ului şi constă în selectarea variabilelor adecvate şi definirea structurii sale. Modelul este proiectat pentru fiecare strat în parte (sub-eşantion de proprietăţi imobiliare comparabile).

Evaluatorul poate selecta, din multitudinea de variabile ce caracterizează proprietăţile imobiliare analizate (variabile explicative potenţiale), pe cele care ar putea avea impact din punct de vedere economic asupra valorii proprietăţii (variabila dependentă). Cu cât numărul de variabile explicative selectat este mai mare, cu atât creşte riscul ca informaţia conţinută de către acestea să devină redundantă. Se dovedeşte astfel utilă realizarea unei selecţii a variabilelor explicative ce vor fi reţinute în analiză. Pe de altă parte, includerea unui număr prea mic de variabile explicative ar putea determina ca modelul de regresie să nu explice îndeajuns de fidel evoluţia variabilei dependente, fapt care ar conduce la o deplasare (engl. *bias*) a estimatorilor (coeficienţii de regresie).

Alegerea variabilelor explicative trebuie să se facă de către evaluator, nu numai din perspectiva mărimii coeficienţilor de corelaţie a variabilelor explicative cu variabila dependentă sau a mărimii coeficienţilor de corelaţie dintre variabilele explicative, ci şi prin luarea în consideraţie a semnificaţiei economice ce ar putea caracteriza legătura dintre variabilele explicative cu cea dependentă.

Așa cum s-a menționat, se analizează și dacă există o corelație puternică între variabilele explicative selectate. În condițiile în care se identifică o astfel de corelație, va trebui reținută în analiză doar una dintre variabilele explicative din grupul celor corelate între ele. Dintre acestea se optează pentru cea variabilă explicativă corelată cel mai puternic cu variabila dependentă. Cu cât valoarea coeficientului de corelație se apropie de valorile 1, respectiv -1, cu atât variabilele sunt mai puternic corelate.

Variabilele pot fi *cantitative* (sau scalare), se exprimă numeric, se măsoară pe o scală și fac obiectul unor operații aritmetice sau pot fi variabile *calitative* (sau categorice), care nu se exprimă numeric (nu se măsoară pe o scală), ci prin atribute, dar pot fi codificate numeric pe categorii, și pot fi de tip:

- nominal, fără ordine internă;
- ordinal, cu o anumită ordine internă, necuantificabilă, de exemplu: starea clădirii poate fi „foarte bună” (codificată cu 0), „bună” (codificată cu 1), „satisfăcătoare” (codificată cu 2) sau „nesatisfăcătoare” (codificată cu 3).

Atenție! Este foarte important să se identifice și să se definească corect toate variabilele modelului și apoi să se codifice (transforme) corespunzător variabilele calitative.

În ceea ce privește structura modelului, aceasta poate fi aditivă (liniară), multiplicativă sau hibridă (neliniară) (a se analiza *Standard on Automated Valuation Models (AVMs)*, IAAO, 2003).

În structura aditivă, contribuția variabilelor explicative este însumată:

$$Y_i = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \dots$$

în care:

Y_i - valoarea de piață a proprietății imobiliare subiect - variabila dependentă;

β_0 - constanta modelului (sau termenul liber);

x_i - variabilele explicative;

β_i - coeficienții variabilelor explicative.

În modelul multiplicativ, contribuția variabilelor explicative este multiplicată:

$$Y_i = \beta_0 \cdot x_1^{\beta_1} \cdot x_2^{\beta_2} \cdot \dots$$

Acest tip de model este mai dificil de calibrat, deoarece variabilele trebuie transformate în format logaritmice, dar avantajul constă în faptul că permite aplicarea corecțiilor proporționale cu valoarea proprietății subiect.

Modelul hibrid (neliniar) este o combinație între modelul aditiv și cel multiplicativ:

$$Y_i = \pi_{QG} \cdot (\pi_{QC} \cdot \Sigma AC + \pi_{QT} \cdot \Sigma AT + \Sigma A)$$

în care:

π_{QG} - produsul variabilelor calitative generale (se aplică proprietății ca întreg);

π_{QC} - produsul variabilelor calitative pentru clădire;

ΣAC - suma variabilelor aditive pentru clădire;

π_{QT} - produsul variabilelor calitative pentru teren;

ΣAT - suma variabilelor aditive pentru teren;

ΣA - suma altor variabile (variabile adiționale).

Etapa a 6-a. Calibrarea modelului

Calibrarea constă în determinarea coeficienților modelului proiectat adică a parametrilor necunoscuți β . Dintre metodele de calibrare, cele mai utilizate sunt cele bazate pe metode statistice, precum regresia liniară multiplă sau regresia neliniară.

Termenul de regresie este folosit pentru descrierea oricărui proces în care una sau mai multe variabile cunoscute (caracteristici ale proprietăților analizate) sunt utilizate pentru a determina o variabilă necunoscută (în acest caz valoarea de piață a proprietăților evaluate) sau pentru previzionarea valorii acesteia. Variabilele cunoscute sunt denumite variabile explicative, iar variabilele necunoscute sunt denumite variabile dependente.

Regresia liniară simplă

Regresia liniară simplă se utilizează pentru a previziona o variabilă dependentă utilizând o singură variabilă explicativă.

Modelul regresiei liniare simple are forma:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, i = \overline{1, n}$$

în care:

β_0 - constanta, adică valoarea lui y când $x=0$;

β_1 – panta liniei de regresie;

ε_i – termenul rezidual (eroarea), cuantifică influența întâmplătoare a altor variabile asupra lui y_i și reprezintă diferența între valoarea observată (y_i) și cea estimată (\hat{y}_i) a variabilei dependente:

$$\varepsilon_i = y_i - \hat{y}_i$$

Panta liniei de regresie indică valoarea medie cu care se modifică y la o schimbare cu o unitate a valorii x . Dacă panta este pozitivă, y crește odată cu creșterea valorii lui x , iar dacă este negativă, y scade odată cu creșterea valorii lui x . Coeficienții modelului, β_0 și β_1 , precum și termenii reziduali ε_i se pot determina prin diferite metode, printre care metoda celor mai mici pătrate.

Pentru analiza regresiei liniare simple, se parcurg următoarele etape:

1. estimarea parametrilor modelului;
2. testarea și validarea modelului de regresie;
3. realizarea de predicții.

Pentru validarea modelului liniar al regresiei simple trebuie respectate o serie de ipoteze (ipoteze Gauss-Markov) și anume:

- varianța termenilor reziduali este constantă;
- termenii reziduali ε_i au media egală cu zero;
- termenii reziduali ε_i nu sunt corelați între ei;
- termenii reziduali ε_i nu sunt corelați cu variabila explicativă x_i ;
- termenii reziduali ε_i sunt normal distribuiți.

În acest sens, este necesară realizarea testării modelului în vederea asigurării calității acestuia (a se vedea etapa a 7-a).

Regresia liniară multiplă

Regresia liniară multiplă exprimă relația dintre o variabilă dependentă și cel puțin două variabile explicative, iar modelul acesteia este:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_m x_{mi} + \varepsilon_i, \quad i = \overline{1, n}$$

în care:

β_0 – termenul liber;

$\beta_1, \beta_2, \dots, \beta_m$ – coeficienții regresiei multiple;

ε_i – termenul rezidual.

Etapele și ipotezele luate în calcul, pentru realizarea modelului de regresie multiplă, sunt aceleași ca cele specificate în cazul regresiei simple.

Regresia neliniară

Dacă în urma reprezentării grafice a datelor prin intermediul diagramei de împrăștiere, se observă că punctele nu sunt grupate de-a lungul unei linii drepte, atunci se poate considera că regresia nu este liniară și soluția ar putea consta în aplicarea regresiei neliniare.

Regresia neliniară se folosește pentru a descrie legătura dintre variabila dependentă și variabilele explicative, considerând că alți factori au o acțiune constantă și neglijabilă asupra variabilei dependente.

Iată câteva din multitudinea de modele matematice care pot fi utilizate pentru aplicarea regresiei neliniare.

Modelul exponențial: $Y_i = \beta_0 \cdot \beta_1^{x_i} + \varepsilon_i$

Modelul logaritmic: $Y_i = \beta_0 + \beta_1 \cdot \lg x_i + \varepsilon_i$

Modelul parabolic de gradul doi: $Y_i = \beta_0 + \beta_1 \cdot x_i + \beta_2 \cdot x_i^2 + \varepsilon_i$

Etapa a 7-a. Testarea modelului și asigurarea calității lui

Testarea modelului are rolul de a stabili dacă acesta asigură acuratețea și integritatea necesare pentru ca estimările, realizate cu ajutorul acestuia pe baza eșantionului, să fie extrapolate prin inferență statistică la nivelul întregii populații. În acest scop se utilizează un set de date despre proprietățile imobiliare pentru care se cunosc prețurile (de tranzacționare și/sau ofertă), pe baza cărora se va face testarea. Acest set de date, folosit pentru testarea modelului, este cel pe baza căruia a fost realizată și calibrarea modelului și reprezintă 80% din volumul eșantionului (n).

Așa cum am precizat mai sus, există o multitudine de teste statistice ce pot fi aplicate pentru obținerea informațiilor necesare, unele parametrice - aplicabile seriilor de date cu distribuție a frecvențelor normală, precum testele *Student (t)*, *Fisher (F)*, și altele non-parametrice - utilizate pentru serii care nu au distribuție a frecvențelor normală, precum testele *Chi-square*, *Shapiro-Wilk*, *Kolmogorov-Smirnov*, *Man-Whitney*, *Durbin-Watson*.

Atenție! În practică, etapa de specificare a modelului și cea de calibrare se realizează în mod iterativ: specificarea modelului – testarea specificațiilor modelului prin calibrare – corectarea specificațiilor modelului – testarea specificațiilor corectate, repetând aceste operații până când se obține nivelul de încredere proiectat (α).

Se poate opta pentru includerea tuturor variabilelor explicative (selectate la etapa a 5-a) în modelul de regresie, iar apoi,

excluderea/reincluderea iterativă a acestora, analizând, după fiecare iterație, testele statistice de validare a modelului de regresie și puterea explicativă a acestuia (mărimea coeficientului de determinare). Această metodă este dificil de aplicat în condițiile în care mărimea eșantionului de date este redusă. Explicația constă în faptul că numărul de variabile explicative nu poate fi mai mare decât numărul de date analizate.

O soluție alternativă este reprezentată de includerea treptată a variabilelor explicative selectate în modelul de regresie. După includerea unei noi variabile explicative se verifică testele statistice de validare a modelului de regresie și puterea sa explicativă. Pentru a stabili care este ordinea de includere a variabilelor explicative în modelul de regresie se folosesc coeficienții de corelație a acestora cu variabila dependentă. Prima variabilă explicativă inclusă în model va fi cea caracterizată de cel mai mare coeficient de corelație cu variabila dependentă.

Interpretarea testelor și a indicatorilor statistici ce caracterizează modelul de regresie

Măsurarea semnificației unui test statistic implică estimarea gradului în care valoarea acestuia, determinată pe baza observațiilor din eșantion, se apropie de cea reală ce caracterizează întreaga populație din care s-a extras eșantionul.

Probabilitatea p (engl. p -value) asociată testului statistic, estimată la nivelul eșantionului, reprezintă probabilitatea de eroare implicată de nerespingerea rezultatului estimat ca fiind valid și reprezentativ pentru întreaga populație studiată. Ca urmare, cu cât această probabilitate este mai ridicată, cu atât este mai puțin plauzibil ca relația existentă între variabilele modelului, evidențiată la nivelul eșantionului studiat, să se mențină pentru întreaga populație statistică din care a fost extras acest eșantion. Ca rezultat, se acceptă ipoteza nulă, adică valoarea estimată pentru coeficientul asociat variabilei explicative analizate nu este semnificativ diferită de zero. În concluzie, acea variabilă explicativă nu influențează variabila dependentă studiată

Nerespingerea sau respingerea ipotezei nule depinde și de nivelul de semnificație ales (α) pentru testul statistic evaluat, respectiv a regiunii de respingere asociate acestuia. Acest nivel de semnificație reprezintă probabilitatea de a realiza o eroare de tip I (respingerea ipotezei nule atunci când aceasta este adevărată). Cu cât valoarea lui α este stabilită la un nivel mai ridicat, cu atât exigențele impuse pentru nerespingerea ipotezei nule (H_0) sunt mai mari. Majorarea nivelului de semnificație

duce la creșterea probabilității de a realiza o eroare de tip II (nerespingerea ipotezei nule atunci când aceasta este falsă) și la diminuarea puterii testului statistic realizat. Cel mai frecvent, nivelul de semnificație, ales pentru realizarea unui test statistic, este de 5%. În această situație, probabilitatea p (asociată testului statistic) mai mică decât nivelul de semnificație al testului de 5%, indică respingerea ipotezei nule.

Exemple de teste statistice utilizate pentru validarea modelelor de regresie

- ⇒ **Testul t** (Student) este aplicat pentru fiecare coeficient de regresie ca raport între valoarea estimată a acestuia și abaterea sa standard și este utilizat pentru testarea ipotezei nule, conform căreia valoarea coeficientului de regresie este egală cu zero.
- ⇒ Cu ajutorul **testului F (Fisher)** se testează ipoteza conform căreia mărimile estimate ale coeficienților de regresie (exceptând termenul liber) sunt nule, ceea ce implică testarea semnificației ecuației de regresie analizată. Dacă probabilitatea p , asociată mărimii estimate pentru acest test, este mai mică decât nivelul de semnificație ales (de exemplu, 5%), se respinge ipoteza nulă conform căreia valorile estimate pentru coeficienții regresiei sunt statistic egale cu zero.
- ⇒ **Coeficientul de determinare al ecuației de regresie R^2** arată proporția în care variația valorii variabilei dependente este explicată de către variabilele explicative selectate în ecuația de regresie. Numărul ridicat de variabile explicative incluse în model poate conduce la valori ridicate ale R^2 , fără să implice în mod direct o putere explicativă mai mare a ecuației de regresie. În acest sens, este recomandată luarea în considerare a mărimii obținute pentru R^2 -ajustat ce ține cont, în estimarea valorii sale, de numărul de variabile incluse în model.
- ⇒ **Testarea normalității legii de distribuție a termenilor reziduali** se poate face cu ajutorul testului Jarque-Bera ce măsoară diferența existentă între indicele de asimetrie, respectiv, de aplatizare al seriei de date, față de cei caracteristici distribuției normale. Un nivel al probabilității p (asociat mărimii estimate a testului) mai mic decât α presupune respingerea ipotezei nule privind distribuția normală a seriei termenilor reziduali, rezultați din ecuația de regresie.

- ⇒ **Homoscedasticitatea termenilor reziduali** (varianța constantă a valorii acestora) este analizată cu ajutorul testului dezvoltat de White. Ipoteza nulă, asociată testului, implică lipsa heteroscedasticității termenilor reziduali, independența acestora în raport de variabilele explicative ale modelului, precum și liniaritatea legăturii existente între variabila dependentă și cele explicative. Încălcarea uneia dintre aceste condiții presupune obținerea unei mărimi semnificative a testului White ($p\text{-value} < \alpha$). Ca urmare, un $p\text{-value} > \alpha$ implică faptul că niciuna dintre aceste trei condiții nu a fost încălcată. În cazul în care termenii reziduali sunt heteroscedastici se pot realiza corecțiile propuse de White și estima corect covarianța coeficienților de regresie chiar și în prezența heteroscedasticității termenilor reziduali.
- ⇒ Cel mai cunoscut test privind **autocorelarea termenilor reziduali** este reprezentat de testul Durbin-Watson. Formula aplicată este următoarea:

$$DW = \frac{\sum_{t=2}^n (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^n \varepsilon_t^2}$$

Valoarea calculată a testului se va compara cu valorile teoretice aferente testului d_L și d_U , preluate din tabelul distribuției Durbin-Watson. Regula de decizie, pe baza valorii calculate a testului, este prezentată în tabelul următor:

$0 < d < d_L$	$d_L \leq d \leq d_U$	$d_U < d < 4 - d_U$	$4 - d_U \leq d \leq 4 - d_L$	$4 - d_L < d < 4$
Autocorelare pozitivă a termenilor reziduali	Indecizie	Termenii reziduali sunt independenți - validarea modelului de regresie din această perspectivă.	Indecizie	Autocorelare negativă a termenilor reziduali

O autocorelare a termenilor reziduali cu grad mai mare de unu nu poate fi sesizată prin intermediul acestui test. Testarea existenței autocorelării termenilor reziduali poate fi analizată și prin intermediul testului Q-statistic, respectiv Breusch- Godfrey LM. În

cazul Q-statistic, în cadrul unui tabel sunt raportate valorile estimate ale testului (pentru diferitele lag-uri ale termenului rezidual) și p-value asociat. Lipsa autocorelării între termenii reziduali va implica valori ridicate ale p-value. Testul Breusch- Godfrey LM are ca ipoteză nulă lipsa autocorelării între termenii reziduali ai ecuației de regresie, fiind raportată mărimea estimată a acestuia și p-value asociat. La fel ca și în cazul testului anterior, un nivel mare al p-value reprezintă o indicație a lipsei autocorelării între termenii reziduali. În condițiile în care termenii reziduali sunt autocorelați pot fi realizate corecțiile propuse de Newey și West și astfel, poate fi estimată corect covarianța coeficienților de regresie, chiar și în prezența heteroscedasticității și autocorelării termenilor reziduali.

Etapa a 8-a. Validarea modelului

După ce modelul a fost calibrat și testat, se procedează la validarea acestuia pe baza setului de date reprezentând restul de 20% din eșantionul analizat, utilizând testele statistice amintite în etapele 3 și 7.

Etapa a 9-a. Aplicarea modelului

În cadrul modelului automat de evaluare se iau în considerare, ca și valori ale variabilelor explicative, cele corespunzătoare caracteristicilor proprietăților imobiliare subiect și se estimează astfel valoarea de piață a acestora.

Rezultatul evaluării prin aplicarea modelului poate fi o singură valoare sau un interval probabil al valorii⁶ în care se încadrează concluzia asupra valorii.

Etapa a 10-a. Verificarea periodică a acurateții modelului

Se recomandă verificarea acurateții modelului prin compararea valorilor de piață estimate cu ajutorul acestuia, cu cele estimate prin aplicarea metodelor și tehnicilor de evaluare recomandate de IVS-uri. Existența unor diferențe mari între cele două valori constituie un semnal privind necesitatea actualizării atât a eșantionului utilizat pentru analiză, cât și a modelului automat de evaluare.

⁶ Intervalul probabil al valorii este intervalul în care se regăsește valoarea, garantat cu o anumită probabilitate

Avantaje și dezavantaje ale utilizării modelelor automate de evaluare

Printre *avantajele* modelării statistice a evaluării proprietăților imobiliare se numără:

- permite utilizarea de serii nelimitate de date;
- permite folosirea unui număr oricât de mare de variabile, dacă acestea sunt necesare și relevante pentru scopul propus;
- permite determinarea acurateții de estimare a valorii de piață, acuratețe care poate fi excepțional de bună atunci când sunt folosite suficiente date de piață;
- asigură evaluări uniforme;
- modelul rezultat se poate utiliza atât pentru evaluări individuale (singulare), cât și pentru evaluare globală;
- automatizarea procesului de evaluare, deci timp redus de prelucrare și analiză a datelor și de stabilire a concluziei asupra evaluării;
- obiectivitate și eficiență.

Aceste avantaje pot fi atinse doar cu condiția esențială ca la baza modelului să stea un raționament statistico-matematic foarte serios și coerent, baza de date să fie permanent actualizată, modelul să fie testat judicios înainte de a fi utilizat efectiv în evaluare.

Dezavantajele modelării statistice sunt:

- este dificil de aplicat pentru piețele restrânse, cu puține tranzacții sau pentru piețele atipice;
- întrucât rezultatele și concluziile generate în urma aplicării modelării statistice sunt valabile numai la nivelul populației supuse studiului, evaluarea unor proprietăți atipice populației nu este credibilă;
- necesitatea unui bagaj foarte dezvoltat de cunoștințe în domeniul statisticii matematice și al soft-urilor de analiză statistică.

Bibliografie

1. *Basic Business Statistics. Concepts and Applications*, Berenson M.L., Levine D.M., Krehbiel T.C., 9th Edition, Pearson Prentice Hall, 2004;
2. *Business Statistics in Practice*, Bruce L. Bowerman, Richard T. O'Connell, McGraw Hill, fourth edition, 2007;
3. *Dicționar de statistică generală*, Alexandru Isaic-Maniu, Eugen Pecican, Viorel Gh. Vodă, Daniela Ștefănescu, Pavel Wagner, Editura Economică, 2003;
4. *Mass Appraisal Methods – An international perspective for property valuers*, Wiley-Blackwell, 2008;
5. *Practical Applications in Appraisal Valuation Modeling, Statistical Methods for Real Estate Practitioners*, M.S. Kane, M.R. Linné, J.A. Johnson, Appraisal Institute, 2004;
6. *Standardele Internaționale de Evaluare 2011*, International Valuation Standards Council, 2011;
7. *Standard on Automated Valuation Models (AVMs)*, International Association of Assessing Officers, 2003;
8. *Standard on Ratio Studies*, International Association of Assessing Officers, 2007;
9. *Statistical Analysis*, S.B. Richmond, 2nd edition, 1964;
10. *Statistică aplicată pentru modelarea evaluării imobiliare*, Iuliu Pașca, Editura Politehnică, 2008;
11. *Statistică: noțiuni fundamentale și aplicații*, Vergil Voineagu, Dana Colibabă, Giani Grădinaru, Editura ASE, 2002;
12. *Statistică teoretică și economică - note de curs*, s.l.drd. Silvia Elena Cristache, Editura ASE, 2002;
13. *Statistică teoretică și economică - teorie și aplicații*, Constantin Anghelache, Editura Economică, 2004;
14. *Statistică: teorie și aplicații*, Tudorel Andrei, Stelian Stancu, Daniel Traian Pele, Editura Economică, ediția a doua, 2002.