

# Impactul factorului de inflamare a erorilor (VIF) asupra valorilor calculate ale testului Student în cazul regresiiilor lineare multiple (Impact of VIF on the Student test statistics in case of multiple linear regressions)

---

---

Florin Marius Pavelescu<sup>1</sup>

## Abstract

*This paper demonstrates that if we deal with multiple regressions the Student Test statistics are sensibly influenced by VIF (Variance Inflation Factor). The author brings arguments in favour of the idea that the study of (multi) collinearity, has to consider not only critical values of VIF, but also the interactions of VIF with the other modeling factors of coefficients of collinear refraction ( $T_{nk}$ ), especially the coefficients of correlation mediated by dependent variable of the explanatory variables ( $r_{jk}$ ), which emphasize the differentiation of the absolute values of Pearson coefficients of correlation of explanatory variables and the dependent variable. This way, it is possible to reveal the role of different factors in the occurrence of harmful collinearity and the advantages of use of a Transformed Form of Student Test Statistics (TFST).*

*Also, a re-grouping of the modeling factors TFST is proposed, in line with the assumptions related to premises of harmful (multi) collinearity occurrence, presented in Glauber and Farrar (1967), and with the comments of Belsey (1991) related to consequences of (multi) collinearity on Student Test statistics. A numerical example, related to the estimation of the index of the fixed capital formation in Romania, during the period 1991-2009, as been modeled by four explanatory, confirms the assumptions made upon the major impact of VIF on Student Test statistics in case of a multiple linear regression. The respective example shows also that harmful collinearity may occurs even in conditions of relatively low values of VIF, if the differentiation of the coefficients  $r_{jk}$  is very high.*

**Keywords:** coefficients of collinear refraction ( $T_{nk}$ ), coefficients of correlation mediated by dependent variable of the explanatory variables ( $r_{jk}$ ), rank of explanatory variables, order of generators of collinearity, Transformed Form of Student Test Statistics (TFST), Reference Value of TFST

**JEL Codes:** C13, C20, C51, C52

---

<sup>1</sup> Institute for National Economy, Center for Macroeconomic Modelling, Bucharest, Romania. E-mail: pavelescu.florin@yahoo.com

## Rezumat

În această lucrare se demonstrează faptul că, în cazul unei regresii lineare multiple, valorile calculate ale Testului Student sunt puternic influențate de Factorul de Inflamare a Dispersiei Erorilor (VIF). Autorul aduce argumente în favoarea ideii că în studierea intensității (multi)colinearității este necesar să se ia în considerare nu doar anumite valori, definite ca fiind critice ale VIF, ci și interacțiunea dintre VIF și alți factori modelatori ai coeficienților de refracție colineară ( $T_{nk}$ ), în mod special coeficienții de corelație mediată de variabila dependentă dintre variabilele explicative ( $r_{jk}$ ), care relevă diferențierea valorilor absolute ale coeficienților de corelație Pearson dintre variabilele explicative și variabila dependentă. Astfel, este posibil să se evidențieze rolul diferiților factori în apariția colinearității toxice, precum și avantajele utilizării Formei Transformate a valorilor calculate ale Testului Student (TFST).

De asemenea, este propusă o regrupare a factorilor modelatori ai TFST în concordanță cu ipotezele adoptate în Glauber și Farrar (1967) referitoare la premisele apariției (multi)colinearității toxice și cu comentariile făcute în D.Belsey (1991) în legătură cu consecințele (multi)colinearității asupra valorilor calculate ale Testului Student. Un exemplu numeric referitor la estimarea indicelui formării brute a capitalului fix în România în perioada 1991-2009, ca fiind modelată de patru variabile explicative, confirmă ipoteza impactului major al VIF asupra valorilor calculate ale Testului Student în cazul unei regresii lineare multiple. Respectivul exemplu relevă faptul că este posibilă apariția colinearității toxice chiar și în condițiile unor valori relativ scăzute ale VIF, dacă diferențierea coeficienților ( $r_{jk}$ ) este foarte accentuată.

**Cuvinte-cheie** coeficienți de refracție colineară ( $T_{nk}$ ), coeficienți de corelație mediată de variabila dependentă între variabilele explicative ( $r_{jk}$ ), rangul variabilei explicative, ordinul generatorului colinearității, Forma Transformată a Valorilor Calculate ale Testului Student (TFST), Valoarea de Referință a TFST

**Clasificare JEL** : C13, C20, C51, C52

## Introducere

Între indicatorii avuți în vedere în literatura de specialitate pentru relevarea intensității colinearității în cazul unei regresii lineare multiple, Factorul de Inflamare a Dispersiei Erorilor (VIF)<sup>2</sup> este des utilizat. De regulă, se consideră că nivelul colinearității este

---

<sup>2</sup> Factorul de Inflamare a Dispersiei Erorilor aferent variabilei explicative  $x_k$  într-o regresie liniară cu  $n$  variabile explicative ( $VIF_{nxk}$ ) denumit în limba engleză "Variance Inflation Factor" se calculează după formula:

$$VIF_{nxk} = \frac{1}{1 - R_{(n-1),xk}^2}, \quad \text{unde: } R_{(n-1),xk}^2, \quad \text{coeficientul de determinare al regresiei lineare}$$

$$x_k = a_{n-1} + \sum_{j=1}^n c_j \cdot x_{j \neq k}$$

deosebit de ridicat, dacă VIF aferent uneia dintre variabilele explicative este mai mare decât 5<sup>3</sup>.

Luarea în considerare doar a nivelului VIF în sine pentru a caracteriza perturbațiile determinate de colinearitate asupra valorilor estimate ale paramerilor și valorilor calculate ale unor teste statistice, precum coeficientul de determinare sau testul Student, nu este însă suficientă, deoarece există destule situații în care apar valori stranii ale indicatorilor menționați anterior, chiar și în condițiile unor valori relativ scăzute ale VIF. Pe această bază, în unele lucrări (A. Alin, 2010) se concluzionează că în fapt colinearitatea este fenomen care nu se limitează doar la o intensă corelație de tip Pearson dintre variabilele explicative.

În aceste condiții, devine deosebit de importantă studierea unor proprietăți algebrice ale valorilor calculate ale testului Student, prin intermediul cărora se pot oferi explicații ale rezultatelor obținute în cazul estimării parametrilor regresii lineare multiple.

## 1. Proprietăți ale valorilor calculate ale testului Student

Dacă pentru estimarea parametrilor unei regresii lineare de tipul  $y = a_n + \sum_{k=1}^n b_k \cdot x_k$ , se utilizează metoda celor mai mici pătrate și se ia în considerație formula de calcul a testului Student, precum și proprietățile valorilor estimate ale parametrilor  $x_k$ , relevate în F. M. Pavelescu (1986, 2005, 2009)<sup>4</sup>, valorile calculate ale testului Student aferente variabilelor explicative  $x_k$  ( $t_{bnk}$ ) se pot exprima conform formulei:

$$t_{bnk} = [m - (n + 1)]^{1/2} * \frac{R(x_k; y)}{(1 - R_n^2)^{1/2}} * \left( \frac{(R_{jl})_n}{(R_{jl})_{n-1, l \neq k}} \right)^{1/2} * T_{nk} \quad (1),$$

unde:

m= numărul de observații,

<sup>3</sup> Considerarea nivelului de 5 al VIF ca fiind un nivel care indică o colinearitate intensă are la bază observațiile dintr-o serie de lucrări și articole dedicate estimărilor econometrice, de la începutul începutul anilor 1960, unde se remarcă faptul că, în condițiile unui coeficient de determinare în regresiiile dintre variabilele explicative de peste 0,8, există o probabilitate foarte mare de apariție a distorsiunilor generate de fenomenul (multi)colinearității.

Regula empirică a stabilirii unei valori critice a VIF pentru relevarea intensității colinearității este destul de laxă. Astfel, există propuneri pentru a se considera că există o (multi)colinearitate intensă dacă  $VIF > 10$ , ceea ce corespunde unui coeficient de determinare a regresiei dintre variabilele explicative de 0,9.

Uneori, se apreciază că nivelul critic de 5 sau 10 al VIF trebuie să fie depășit de valoarea cea mai mare a respectivului indicator în cazul regresiei avute în vedere pentru a semnaliza apariția unei (multi)colinearități intense, care distorsionează rezultatele estimării, în timp ce există opinii care susțin că respectivele niveluri critice este necesar să fie depășite de media aritmetică a VIF, pentru a considera că se manifestă o (multi)colinearitate intensă.

<sup>4</sup> Astfel, se poate demonstra că în cazul unei regresii lineare simple de forma  $y = a_1 + b_{1k} \cdot x_k$ , valoarea estimată a

parametrului  $b_{1k}$  este dată de formula  $b_{1k} = \frac{\text{cov}(y; x_k)}{\text{var}(x_k)}$ , unde:

$\text{cov}(y; x_k)$  = covarianța dintre variabila dependentă (y) și variabila explicativă  $x_k$ .  
 $\text{var}(x_k)$  = dispersia variabilei explicative  $x_k$

În cazul unei regresii lineare multiple, respectiv  $y = a_n + \sum_{k=1}^n b_k \cdot x_k$ , valoarea estimată a parametrului  $b_{nk}$  se determină conform formulei:  $b_{nk} = b_{1k} \cdot T_{nk}$

$n$  = numărul de variabile explicative luate în considerare în cadrul regresiei lineare  
 $R(x_k, y)$  = coeficientul de corelație Pearson dintre variabila explicativă  $x_k$  și variabila dependentă  $y$ .

$R^2_{ny}$  = coeficientul de determinare al regresiei lineare  $y = a_n + \sum_{k=1}^n b_k \cdot x_k$

$(R_{jl})_n$  = determinantul matricii coeficienților de corelație Pearson dintre toate variabilele explicative

$(R_{jl})_{n-1, j \neq k}$  = determinantul matricii coeficienților de corelație Pearson dintre variabilele explicative, exclusiv variabila explicativă  $x_k$ .

$T_{nk}$  = coeficientul de refracție colineară<sup>5</sup>.

Se poate demonstra (F. M. Pavelescu, 2012, și 2013) că în cazul regresiiilor lineare cu două și respectiv trei variabile explicative există relația<sup>6</sup>:

$$\left( \frac{(R_{jl})_n}{(R_{jl})_{n-1, l \neq k}} \right)^{1/2} = (1 - R^2_{(n-1), xk})^{(1/2)} \quad (2)$$

Dacă se are în vedere (F. M. Pavelescu 1986, 2012 și 2013) se poate ajunge, în cazul regresiiilor lineare cu două și respectiv trei variabile explicative, la următoarea formulă de calcul pentru  $T_{nk}$ :

<sup>5</sup> În lucrările publicate de autor în perioada 1986-2005, factorul  $T_{nk}$  a fost denumit "coeficient de aliniere", iar în perioada martie 2009-iunie 2014 "coeficient de aliere la hazardul colinearității". Respectiv cele denumiri puneau în evidență valorile indicatorului, care în multe cazuri apăreau ca fiind stranii, mai cu seamă în contextul regresiiilor lineare cu cel puțin trei variabile explicative. De asemenea, în F. M. Pavelescu (2010), pornind de la considerațiile expuse în D. Douglas, D. Clader, J. Cristy, P. Michaelis, D. Belsey (2003) se arăta că factorul  $T_{nk}$  are unele similitudini cu raportul dintre zgomot și semnal, care este utilizat în științele tehnice.

Denumirile date factorului  $T_{nk}$  menționate anterior au o serie de dezavantaje legate în special de lungimea acestora. Pe de altă parte, asimilarea indicatorului menționat anterior cu raportul dintre zgomot și semnal nu oferă o explicație convingătoare a modificării valorilor estimate ale parametrilor în cazul regresiiilor lineare multiple în raport cu cele obținute în cazul regresiiilor lineare simple.

Aceste aspecte au fost relevate în cursul dezbaterii draftului prezentei lucrări prezentate la 24 iunie 2014, în cadrul Seminarului de Modelare Macroeconomică ale INCE. Autorul a luat în considerare criticile exprimate cu acest prilej. Analizând cu atenție efectele produse de factorul  $T_{nk}$  asupra valorilor estimate ale parametrilor unei regresii multiple, se poate găsi o anumită similitudine între efectele colinearității și cele ale refracției atmosferice.

Astfel, refracția atmosferică reprezintă "curbarea treptată a razelor de lumină care vin de la aștri prin refracția lor în straturile atmosferice, a căror densitate crește pe măsura apropierii de sol, și are drept rezultat înălțarea poziției aparente a aștrilor față de orizont" (cf. Mic Dicționar Enciclopedic, Editura Enciclopedică Română, București, 1972).

Pornind de la efectele refracției în cazul fenomenelor atmosferice, considerăm că se poate vorbi despre "efectul de refracție" determinat de (multicolinearitate) în cazul unei regresii lineare multiple sau mai pe scurt despre "refracția colineară". În aceste condiții, factorul  $T_{nk}$  poate fi denumit "coeficient de refracție colineară".

Pe de altă parte, este important să se aibă în vedere că factorul  $T_{nk}$  nu influențează în mod direct numai valorile estimate ale parametrilor în cazul regresiiilor lineare multiple sau valorile calculate ale Testului Student, ci și coeficientul de determinare al regresiiilor lineare multiple, respectiv cu  $n$  variabile explicative ( $R^2_{ny}$ ). Astfel,

în F. M. Pavelescu (2003) este demonstrată relația:  $R^2_{ny} = \sum_{k=1}^n R^2(x_k; y) \cdot T_{nk}$

De asemenea, factorul  $T_{nk}$ , influențează în mod indirect Coeficientul de determinare ajustat, precum și valoarea calculată a testului Fisher în cazul regresiiilor lineare multiple.

<sup>6</sup> Existența unei influențe a VIF asupra valorilor calculate ale Testului Student a fost sugerată de Acad. E Dobrescu în cursul dezbaterii organizate la 15 mai 2013 în cadrul Seminarului de Macromodelare Economică referitoare la lucrarea F. M. Pavelescu - Factorii modelatori ai valorilor calculate a testului Student în cazul unei regresii lineare trifactoriale".

$$T_{nk} = \frac{1 - p_{jkmed} \cdot R_{(n-1),xk}^2}{1 - R_{(n-1),xk}^2} \quad (3), \text{ unde:}$$

$p_{jkmed}$  = media aritmetică ponderată a rapoartelor  $p_{jk}$ ., definite astfel:

$$p_{jk} = \frac{r_{jk}}{R(x_j; x_k)} \quad (4)$$

$$r_{jk} = \frac{|R(x_j; y)|}{|R(x_k; y)|} \quad (5)$$

N.B. În F. M. Pavelescu (2005) rapoartele  $r_{jk}$  au fost definite drept “coeficienți de corelație dintre variabilele explicative  $x_j$  și  $x_k$ , mediate de variabila dependentă, relativ la variabila explicativă  $x_k$ .”<sup>7</sup>

Faptul că formulele (2) și (3) se pot demonstra în mod riguros în cazul unor regresii lineare cu două și respectiv trei variabile explicative, conduce la ideea că formulele respective sunt valabile și în cazul unor regresii lineare cu  $n$  variabile explicative. Demonstrația riguroasă a proprietăților menționate anterior este dificilă pentru  $n > 3$ , în principal din cauza multiplicării explozive a numărului de factori care concură la obținerea coeficienților de determinare ai regresiiilor lineare. În vederea surmontării acestui impediment, pentru relevarea valabilității respectivelor formule de calcul și în cazul unor regresii lineare cu mai mult de trei variabile explicative, se va prezenta un exemplu numeric prin intermediul căruia să se poată confirma ipoteza că formulele (2) și (3) sunt valabile în cazul general (cu  $n$  variabile explicative).

În aceste condiții, formula pentru determinarea valorilor calculate ale Testului Student aplicat unei regresii lineare devine:

$$t_{bnk} = \sqrt{(m-n-1)} \cdot R(x_k, y) \cdot \sqrt{\frac{1 - R_{(n-1),xk}^2}{1 - R_{ny}^2}} \cdot \frac{1 - p_{jkmed} \cdot R_{(n-1),xk}^2}{1 - R_{(n-1),xk}^2} \quad (6)$$

Formula (6) arată cauzele apariției unor valori calculate ale Testului Student în varianta standard a acestuia, care la prima vedere apar ca fiind de-a dreptul stranii. Se observă că factorul care determină apariția colinearității toxice, respectiv valori ale parametrilor estimați ale căror semne sunt contrare celor ale coeficienților de corelație Pearson dintre variabila explicativă analizată și variabila dependentă ( $R(x_k, y)$ ) este coeficientul de refracție colineară ( $T_{nk}$ ).

În consecință, pentru a se evidenția rapid apariția colinearității toxice în cadrul unei regresii lineare multiple se poate face apel la forma transformată a valorilor calculate ale Testului Student (TFST) definită conform formulei (F. M. Pavelescu, 2013):

<sup>7</sup> În F. M. Pavelescu (2010) au fost redefiniți respectivii coeficienți pentru ca valoarea lor să fie cuprinsă între 0 și 1, așa cum este cazul coeficienților de corelație Pearson. În acest caz, a fost avută în vedere ierarhizarea coeficienților de corelație Pearson  $R(x_k, y)$ , din punct de vedere al valorii absolute a acestora. Drept urmare, au fost definiți “coeficienți de corelație mediată de variabila rezultativă între variabila explicativă de rangul  $k$  și variabila explicativă  $k$ ”. ( $R(x_{k+1}, x_k)_{medy}$ ).

În aceste condiții, rapoartele  $r_{jk}$  pot fi scrise și sub forma:  $r_{jk} = \frac{R(x_j; x_p)_{medy}}{R(x_k; x_p)_{medy}}$ , unde  $x_p$  = variabila explicativă

principală a regresiei lineare multiple, respectiv variabila explicativă care în mod absolut este cel mai intens corelată în sens Pearson cu variabila dependentă.

$$TFST_{bnk} = \sqrt{(m-n-1)} \cdot |R(x_k; y)| \cdot \sqrt{\frac{1-R_{(n-1)Xk}^2}{1-R_{ny}^2}} \cdot T_{nk} \quad (7)$$

$$\text{echivalent cu: } TFST_{bnk} = t_{bnk} \cdot \frac{|R(x_k; Y)|}{R(x_k; Y)} \quad (8)$$

Se observă că  $TFST_{bnk}$  se obține prin multiplicarea valorilor calculate ale Testului Student în varianta standard cu scalarul  $\frac{|R(x_k; Y)|}{R(x_k; Y)}$ . Este important de subliniat faptul că folosirea

$TFST$  poate avea loc numai în cazul regresiei lineare, în condițiile în care se mențin toate celelalte ipoteze referitoare la Testul Student standard<sup>8</sup>.

Un alt avantaj al definirii și utilizării  $TFST_{bnk}$  este posibilitatea identificării rangurilor variabilelor explicative în calitate de descriptori ai comportamentului variabilei rezultative. Astfel, se pot defini rangurile variabilelor explicative în calitate de descriptori ai comportamentului variabilei rezultative după cum urmează: **variabila explicativă principală a modelului de regresie cu n variabile explicative** este cea care, în mod absolut, se corelează cel mai puternic din punct de vedere Pearson cu variabila dependentă, în timp ce **variabila explicativă de rangul n** este cea care în mod absolut este cel mai slab corelată cu variabila dependentă.

În mod similar, se pot defini ordinele generatorilor de colinearitate în funcție de mărimea VIF aferent fiecărei variabile explicative. Astfel, generatorul principal al colinearității este variabila explicativă care în cadrul regresiei are VIF maxim, iar generatorul colinearității de ordinul n este cea care are VIF minim<sup>9</sup>.

## 2. Valențe cognitive ale regrupării factorilor modelatori ai TFST

În condițiile în care Forma Transformată a valorilor calculate ale Testului Student este un produs dintre mai mulți factori, se poate pune problema regrupării acestora în cadrul unor indicatori cu o putere explicativă mai mare.

Astfel, pornind de la considerațiile expuse în D. Belsey (1991), potrivit cărora în mod normal valorile calculate ale Testului Student ar trebui să reflecte valorile coeficientului de corelație Pearson dintre variabila explicativă analizată și variabila dependentă, se poate defini indicatorul “**Valoarea de referință a Formei Transformate a valorilor calculate ale Testului Student pentru variabila  $x_k$** ” ( $RV_{studnk}$ ), respectiv:

$$RV_{studnk} = (m-n-1)^{(1/2)} \cdot |R(x_k, y)| \quad (9)$$

$RV_{studnk}$  este singura componentă a Formei Transformate a valorilor calculate ale Testului Student care nu este influențată în mod direct de fenomenul colinearității.

Pe de altă parte, avem în vedere faptul că în D. Ferrar și R. Glauber (1967) se considera că se poate defini colinearitatea toxică nu numai atunci când semnul parametrilor estimați este contrar celui anticipat, în fapt cel al coeficientului de corelație Pearson dintre variabila

<sup>8</sup> Propunerea privind modul de utilizare a TFST ca o etapă a estimării parametrilor unei regresii lineare multiple în vederea detectării rapide a colinearității toxice este descrisă în F. M. Pavelescu (2013).

<sup>9</sup> Ideia definirii unor ranguri ale variabilelor explicative în calitate de descriptori ai comportamentului variabilelor rezultative și a unor ordine ale generatorilor de colinearitate a fost expusă în F M Pavelescu (2010) pentru cazul regresiei lineare cu trei variabile explicative.

explicativă analizată și variabila dependentă, ci și atunci când pătratul coeficientului de corelație Pearson dintre variabilele explicative este mai mare decât coeficientul de determinare al regresiei lineare multiple. Pornind de la ipotezele respective, se poate defini și indicatorul “Multiplicatorul ajustat al Valorii de Referință a Formei Transformate a valorilor calculate ale Testului Student” ( $AdMRV_{nk}$ ), definit prin relația :

$$AdMRV_{nk} = \sqrt{\frac{1 - R_{(n-1),xk}^2}{1 - R_{ny}^2}} \quad (10)$$

În definirea indicatorului  $AdMRV_{nk}$  se are în vedere faptul că factorul  $\sqrt{\frac{1}{1 - R_{ny}^2}}$  acționează ca un multiplicator al  $RV_{studnk}$ , dacă nu s-ar manifesta fenomenul de (multi)colinearitate. În practică, însă datorită existenței multicolarității, impactul respectivului multiplicator este atenuat de rădăcina pătrată a toleranței, definită ca inversul VIF.

Drept urmare, pentru determinarea valorilor calculate ale Testului Student Corectat aplicat unei regresii lineare cu n variabile explicative ( $TFST_{bnk}$ ) se poate apela și la formula:

$$TFST_{bnk} = RV_{studnk} \cdot AdMRV_{nk} \cdot T_{nk} \quad (11)$$

### 3. Un exemplu numeric. Estimarea funcției formării brute a capitalului fix în România pentru perioada 1991-2009

Pentru a testa existența proprietăților algebrice demonstrate anterior ale  $TFST_{bnk}$  și în condițiile unei regresii lineare cu mai mult de trei variabile explicative, precum și impactul pe care VIF îl are în modelarea respectivului indicator, s-a estimat o funcție a indicelui formării brute a capitalului fix în România pentru perioada 1991-2009 luându-se în considerare patru variabile explicative.

Ca variabile explicative au fost avute în vedere: a) Indicele formării brute a capitalului fix în anul precedent ( $IGFCFc(-1)$ ), b) Indicele exportului din anul precedent ( $IXc(-1)$ ) c) Indicele importului din anul curent ( $IMc$ ), d) Indicele importului din anul precedent ( $IMc(-1)$ )

Estimarea regresiei lineare a condus la următorul rezultat:

$$IGFCFc = 0.0752 + 0.4993 * IGFCFc(-1) - 0.1309 * IXc(-1) + 0.4845 * IMc + 0.0839 * IMc(-1)$$

(0.2142) (2.6563) (-0.5157) (2.5340) (0.3577)

$$R^2_{4y} = 0.56145.$$

N.B. În paranteze sunt prezentate valorile calculate ale testului Student standard.

$R^2_{4y}$  = coeficientul de determinare al regresiei cu patru variabile explicative

Pentru determinarea coeficienților de corelație Pearson dintre fiecare variabilă explicativă și variabila dependentă ( $R(x_k, y)$ ), precum și coeficientul de refracție colineară ( $T_{4k}$ )<sup>10</sup> au fost calculate următoarele regresii lineare simple:

<sup>10</sup> Coeficientul  $T_{4k}$  a fost obținut în prezenta lucrare prin împărțirea valorilor estimate ale parametrului corespunzător variabilei explicative analizate în cazul regresiei lineare multiple la parametrul obținut în cazul regresiei lineare simple. Cu alte cuvinte  $T_{4k}$  a fost determinat pe baza formulei  $T_{4k} = \frac{b_{4k}}{b_{1k}}$

$$\text{IGFCFc} = 0.5899 + 0.4469 * \text{IGFCFc}(-1)$$

$$(2.9625) \quad (2.3919)$$

$$R^2 (\text{IGFCFc}(-1), \text{IGFCFc}) = 0.2518,$$

$$R (\text{IGFCFc}(-1), \text{IGFCFc}) = 0.5018$$

$$T_{4\text{IGFCFc}(-1)} = 1,1746$$

$$\text{IGFCFc} = 0.8467 + 0.2070 * \text{IXc}(-1)$$

$$(3.7285) \quad (0.9535)$$

$$R^2 (\text{IXc}(-1), \text{IGFCFc}) = 0.0508,$$

$$R (\text{IXc}(-1), \text{IGFCFc}) = 0.2253$$

$$T_{4\text{IXc}(-1)} = -0,6326$$

$$\text{IGFCFc} = 0.6093 + 0.4331 * \text{IMc}$$

$$(3.5760) \quad (2.6897)$$

$$R^2 (\text{IMc}, \text{IGFCFc}) = 0.2985,$$

$$R (\text{IMc}, \text{IGFCFc}) = 0.5464$$

$$T_{4\text{IMc}} = 1,1188$$

$$\text{IGFCFc} = 1.2459 - 0.1732 * \text{IMc}(-1)$$

$$(5.9442) \quad (-0.8946)$$

$$R^2 (\text{IMc}(-1), \text{IGFCFc}) = 0.0450,$$

$$R (\text{IMc}(-1), \text{IGFCFc}) = - 0.2120$$

$$T_{4\text{IMc}(-1)} = -0,4845$$

Se observă că pentru calcularea FTST este necesară utilizarea scalarului  $\frac{R(x_k; Y)}{R(x_k)}$  doar în

cazul variabilei explicative IMc(-1).

Pentru determinarea VIF au fost estimate următoarele regresii lineare cu trei variabile explicative:

$$\text{IGFCFc}(-1) = 0.8227 + 0.7039 * \text{IXc}(-1) + -0.1832 * \text{IMc} + -0.2866 * \text{IMc}(-1)$$

$$(1.9002) \quad (2.3654) \quad (-0.7089) \quad (-0.9140)$$

$$R^2_{3\text{IGFCFc}(-1)} = 0.2970$$

$$\text{IXc}(-1) = -0.4123 + 0.3652 * \text{IMc} + 0.6172 * \text{IMc}(-1) + 0.3859 * \text{IGFCFc}(-1)$$

$$(-1.2097) \quad (2.1471) \quad (3.4766) \quad (2.3654)$$

$$R^2_{3\text{IXc}(-1)} = 0.5900$$

$$\text{IMc} = 1.4520 - 0.8318 * \text{IMc}(-1) - 0.1770 * \text{IGFCFc}(-1) + 0.6437 * \text{IXc}(-1)$$

$$(5.0042) \quad (-3.5735) \quad (-0.7089) \quad (2.1471)$$

$$R^2_{3\text{IMc}} = 0.4617$$

$$\text{IMc}(-1) = 1.0909 - 0.1841 * \text{IGFCFc}(-1) + 0.7230 * \text{IXc}(-1) - 0.5528 * \text{IMc}$$

$$(4.1224) \quad (-0.9140) \quad (3.4766) \quad (-3.5735)$$

$$R^2_{3\text{IMc}(-1)} = 0.6203$$



Pe baza regresiiilor lineare prezentate anterior, se determină rangurile variabilelor explicative în calitate de descriptori ai comportamentului variabilei rezultative, precum și ordinele generatorilor de colinearitate, pe de altă parte (tabelul nr. 1).

**Tabelul nr. 1**

**Rangurile variabilelor explicative în calitate de descriptori ai comportamentului variabilei explicative și ordinele generatorilor de colinearitate în cazul estimării indicelui funcției formării brute a capitalului fix în România în perioada 1991-2009**

Variabila Explicativă	/R (x <sub>k</sub> , y)/	Rang Explicativă	Variabilă	VIF <sub>xk</sub>	Ordin Colinearitate
IGFCFc(-1)	0.5018	2		1.4226	4
IXc(-1)	0.2253	3		2.4388	2
IMc	0.5437	1		1.8577	3
IMc(-1)	0.2120	4		2.6333	1

Calculat pe baza datelor din Anuarul Statistic

Se observă o marcată diferențiere a intensității corelației dintre variabilele explicative și variabila dependentă, valoarea absolută a coeficientului de corelație Pearson fiind cuprinsă între 0.2120 și 0.5437. În aceste condiții, raportul dintre valoarea maximă și valoarea minimă a respectivilor coeficienți este de 2,5766, ceea ce crează premisele unei presiuni destul de ridicate a coeficienților de corelație mediată de variabila dependentă<sup>11</sup> și se favorizează în mod implicit apariția colinearității toxice. Un alt factor care determină apariția colinearității toxice este asimetria dintre rangurile variabilelor explicative și ordinele generatorilor de colinearitate.

Dacă se are în vedere că numărul gradelor de libertate este de 14, se pot identifica factorii modelatori ai TFST<sub>b4k</sub>. În tabelul nr. 2 este prezentat calculul TFST<sub>b4k</sub>, așa cum reiese din formula (12).

**Tabelul nr. 2**

**Determinarea TFST<sub>b4k</sub> prin evidențierea rolului valorii de referință (RV<sub>stud4xk</sub>), multiplicatorului ajustat al valorii de referință (AdMRV<sub>stud4xk</sub>) și a coeficientului de refracție colineară (T<sub>4k</sub>)**

Variabila explicativă	RV <sub>stud4xk</sub>	AdMRV <sub>stud4xk</sub>	T <sub>4k</sub>	TFST <sub>b4k</sub>
IGFCFc(-1)	1.8775	1.2661	1.1175	2.6563
IXc(-1)	0.8431	0.9669	-0.6326	-0.5157
IMc	2.0443	1.1079	1.1188	2.5340
IMc(-1)	0.7934	0.9305	-0.4845	-0.3577

Calculat pe baza datelor din Anuarul Statistic

Produsul celor trei factori modelatori luați în considerare în cadrul formulei (12) este practic egal cu valorile absolute calculate pentru Testul Student obținute în cazul primei estimări a indicilor formării brute a capitalului fix prin intermediul ecuației de regresie liniară cu patru variabile explicative.

<sup>11</sup> Ideia că rapoartele  $p_{jk}$  reprezintă o "presiune a coeficienților de corelație mediată de variabila rezultativă asupra coeficienților de corelație Pearson dintre variabilele explicative  $x_j$  și  $x_k$ " a pornit de la analogia cu indicatorul "presiunea cererii totale (asupra capacităților de producție)" definit în E. Dobrescu (2006) ca raportul dintre cererea totală și stocul de capital fix.

Calculul  $TFST_{b4k}$ , prin evidențierea impactului exercitat de coeficientul de refracție colineară ( $T_{4k}$ ), relevă faptul că în polarizarea respectivului indicator, care se concretizează în cele din urmă prin apariția colinearității toxice, diferențierea coeficienților de corelație Pearson ( $R(x_k, y)$ ) joacă un rol însemnat. Astfel, se observă că variabilele explicative cu  $RV_{stud4xk}$  și  $AdMRV_{stud4xk}$  supraunitare dețin și coeficienți  $T_{nk}$  pozitivi și supraunitari. În schimb, variabilele explicative mai slab corelate în sens Pearson cu variabila dependentă induc un grad de colinearitate mai ridicat în cadrul ecuației de regresie avute în vedere și care este cuantificat prin intermediul VIF.

**Exemplul numeric referitor la estimarea ecuației de regresie a formării brute a capitalului fix în România, pe parcursul perioadei 1991-2009, validează formulele (2) și (3) și pentru cazul în care numărul de variabile explicative este mai mare decât trei. Pe această bază, se evidențiază impactul major al VIF asupra valorilor calculate ale testului Student în formă standard sau în formă transformată aplicat unei regresii lineare multiple în cazul general, respectiv cu  $n$  variabile explicative.**

#### **4. Consecințe ale utilizării TFST și ale relevării impactului VIF asupra formei și intensității multicolarității în regresii lineare multiple**

Definirea și utilizarea TFST, în contextul luării în considerație a rolului VIF asupra mărimii respectivului indicator, permite relevarea unor aspecte care contribuie la explicarea obținerii unor rezultate ale estimărilor, care, la prima vedere, apar ca fiind stranii. Respectivul aspecte vor fi trecute în continuare în revistă.

A) Definirea TFST permite să se evidențieze **dublul caracter al variabilelor explicative folosite în cadrul unei regresii lineare multiple** și anume:

**descriptori ai comportamentului variabilei dependente**, calitate relevată de valoarea absolută a coeficienților de corelație Pearson  $R(x_k; y)$  și

**generatori de colinearitate**, calitate relevată de mărimea VIF

B. **Dacă se iau în considerare cele două criterii de clasificare a variabilelor explicative menționate anterior, se pot identifica un număr de  $n!$  (permutări de  $n$ ) de cazuri posibile de distribuire a variabilelor explicative în cadrul modelelor de regresie liniară<sup>12</sup>.**

Drept urmare, cea mai mică probabilitate de apariție a colinearității toxice este atunci când rangurile variabilelor explicative coincid cu ordinele generatorilor de colinearitate. În mod similar, premisele pentru apariția colinearității toxice sunt cu atât mai mari, cu cât divergențele dintre rangurile variabilelor explicative și ordinele generatorilor de colinearitate sunt mai accentuate.

C. **Prezența VIF în cadrul formulei de calcul a TFST relevă faptul că pe măsură ce numărul de variabile se majorează, crește probabilitatea apariției colinearității toxice, sau /și a polarizării valorilor calculate ale respectivului test statistic.**

D. **Valorile ridicate ale VIF reprezintă condiții favorizante ale apariției colinearității toxice, dar apariția respectivului fenomen este determinată în cele din urmă de relația dintre factorii  $p_{jkmed}$  și coeficienții de determinare  $R^2_{(n-1)xk}$ . Cu alte cuvinte,**

<sup>12</sup> *Excepția de la regulă în ceea ce privește numărul de cazuri ale distribuției rangurilor variabilelor explicative și ordinilor generatorilor de colinearitate apare în cazul regresiei lineare cu două variabile explicative. Deoarece VIF este același pentru ambele variabile explicative, există un singur caz al ierarhizării variabilelor explicative, respectiv cel după rangurile determinate de intensitatea corelației Pearson dintre variabila dependentă și variabilele explicative.*

colinearitatea toxică în cazul variabilei explicative  $x_k$  apare atunci când presiunea medie exercitată de coeficienții de corelație mediată de variabila dependentă este mai mare decât inversul coeficientului de determinare  $R^2_{(n-1)x_k}$ . În fapt, diferențierea foarte mare a valorilor absolute ale coeficienților de corelație Pearson dintre variabilele explicative și variabila dependentă poate conduce la apariția colinearității toxice, chiar și în condițiile unor valori relativ scăzute ale VIF.

Drept urmare, **stabilirea unor praguri ale VIF pentru identificarea prezenței multi(colinearității)** și cu deosebire a celei toxice are un anumit grad de convenționalitate.

**E. În cazul în care se dorește stabilirea unei valori reper a VIF pentru identificarea unei colinearități accentuate, din punctul de vedere al logicii Testului Student, se poate lua în considerare cazul când VIF = 4.** Dacă  $VIF=4$ , se determină înjumătățirea valorii de referință a formei transformate a testului Student ( $RV_{studxk}$ ).

**F. Utilizarea TFST pune în evidență faptul că mărirea numărului de observații se dovedește utilă numai când este vorba de manifestarea în cadrul regresiei lineare multiple a colinearității degradante**<sup>13</sup>. În acest caz, prin creșterea numărului de observații, se poate asigura majorarea TFST și, implicit, depășirea valorilor tabelate ale Testului Student. În cazul colinearității toxice, mărirea numărului de observații contribuie la mărirea valorii absolute a TFST, dar nu are cum să determine schimbarea semnului negativ, în contextul menținerii corelațiilor inițiale dintre variabilele explicative și variabila dependentă.

**G. În cazul apariției unor valori pozitive, dar foarte scăzute ale  $TFST_{bnk}$  este recomandabil să se releve factorul principal al respectivei situații și anume: a) o corelație slabă între variabila explicativă analizată și variabila dependentă, b) un nivel ridicat al VIF sau c) un nivel foarte scăzut, apropiat de zero, al coeficientului de refracție colineară ( $T_{nk}$ ).**

## Referințe bibliografice

A. Alin (2010) – Multicollinearity. Focus Article, *Willey Interdisciplinary Reviews: Computational Statistics, Vol. 2, Issue 3*

D Belsey (1976)– Multicollinearity: Diagnosing its presence and assessing the potential damage it causes least square estimation, *NBER Working Paper no. 154, Cambridge, Massachusetts, October*.

D. Belsey (1991)- Conditional diagnostics: collinearity and weak data in regression, Wiley Series in Probability John Wiley, New York.

C. Conrad (2006)-Applied Regression Analysis. Lectures Notes, Department of Economics, Pomona College, Claremont, California, Spring.

E. Dobrescu- (2006) - Macromodels of the Romanian Market Economy, Editura Economică, București

---

<sup>13</sup> Conform accepțiunii date de D. Belsey (1991) colinearitatea poate fi clasificată ca fiind: a) slabă (acceptabilă) b) degradantă și c) toxică. Colinearitatea degradantă este caracterizată prin valori calculate scăzute ale testului Student, dar ale căror semne sunt aceleași cu cele ale coeficienților de corelație Pearson dintre variabilele explicative și variabila rezultativă.

- C.Dougherty (2002)- Introduction to econometrics-Second Edition, Oxford University Press.
- D. Douglas, D.Clader, J.Cristy, P. Michaelis, D.Belsey (2003)- Test about harmful collinearity among predictor variables used in modeling global temperature, *Climate research vol. 24, June*.
- D. Farrar, R. Glauber (1967)- Multicollinearity in regression analysis: The problem revisited, *Review of Economics and Statistics, February*.
- D.Jula – Introducere în econometrie, Editura Professional Consulting, București, 2003
- G. Maddala - Introduction to econometrics, Mc Millan Publishing Company, 1988
- W. Greene (1993)-Econometric Analysis, Mc Millan Publishing Company, New York.
- A. Isaic - Maniu, C. Mitruț, V. Voineagu (1996)- Statistica pentru managementul afacerilor, Editura Economică, București, 1996
- F. M. Pavelescu (1986) - Some considerations regarding the Cobb-Douglas production function estimated parameters. A new approach, *Revue Roumaine des Sciences Sociales Tome 30 No 1-2*.
- F. M. Pavelescu (2003)- Proprietăți ale coeficientului de determinație, *Revista Română de Statistică nr. 4*
- F. M. Pavelescu (2005) - Impact of collinearity on the estimated parameters and values of classical Statistical tests in conditions of multifactorial linear OLS regressions *Romanian Journal of Economic Forecast no.2*.
- F .M. Pavelescu (2009)- A review of Student Test properties in condition of multifactorial linear regression, *Romanian Journal of Economic Forecast no.1*
- F .M. Pavelescu (2010)- An Extensive Study on the Disturbances Generated by Collinearity in a linear Regression Model with Three Explanatory Variables, *Romanian Journal of Economics no.2 (40)*.
- F. M. Pavelescu (2012)- Interdependența dintre tipul de colinearitate și valorile calculate ale testului Student în cazul unei regresii lineare cu două variabile explicative *Caietul de Studiu nr. 25 al Seminarului de Modelare Macroeconomică*,
- F. M. Pavelescu (2013) - Factorii modelatori ai valorii calculate a testului Student în cazul unei regresii lineare trifactoriale, *Caietul de Studiu nr. 26 al Seminarului de Modelare Macroeconomică al INCE*
- R.Wiliams (2008) - Sociology Graduate Statistics, University of Notre Dame, Indiana.