Modelling factors of Student test statistics dispersion in a multiple linear regression

Florin-Marius PAVELESCU[1]

Abstract. One of the main objectives of this paper is to show that some algebraical properties of the Student test statistics demonstrated in F.M. Pavelescu (2014) and the use of a proposed new indicator – Synthesis of the Transformed Form of Student Test statistics - for the case of linear regressions with three explanatory are valid for the general case of linear regressions with n explanatory variables. Also, we propose a re-grouping of the modeling factors of Transformed Form Student Test statistics, which is in line with the ideas of Glauber and Ferrar (1967) on the definition of collinearity and of Belsey (1991) on the impact of collinearity on Student Test Statistics.

In this context, the methodology proposed permits the identification of modeling factors contribution to the dispersion of Transformed Form of Student statistics.

At the end of the paper a numerical example is presented in order to show that algebraical properties of the Student test statistics demonstrated in case of a linear regression with three explanatory variables are true also when the number of explanatory variables is higher than three. The respective numerical example is also an opportunity for a practical use of the proposed factorial analysis methodology of the dispersion of the Transformed Form of the Student Test statistics.

Keywords: *Transformed Form of Student Test statistics, factorial analysis, reference value, adjusted multiplier, coefficient of collinear refraction*

JEL Classification: C13, C20, C51, C52

---

[1] Ph. D., Institute of National Economy – Romanian Academy, Bucharest, e-mail: pavelescu.florin@yahoo.com.

In linear regressions, as the number of explanatory variables increases, we may observe that the Student Test statistics related to the considered explanatory variables are more and more dispersed. Therefore, it is possible that these situations occur when some of the estimated parameters are validated. Consequently, it is very important to draw a conclusion on the quality of estimation of the linear regression as a whole. A solution to this problem may be a comprehensive research of the modelling factors of the estimated parameters and of the Student test statistics dispersion, on one hand, and the definition and practical use of an indicator for the quality of linear regression estimation as a whole, on the other hand.

1. A review of the algebraic properties of estimated parameters in case of linear regressions with n explanatory variables

Taking into consideration the demonstrations presented in F. M. Pavelescu (1986), it can be concluded that if the method of ordinary least squares (OLS) is used, in case of a multiple linear regression, i.e. $y = a_n + \sum_{k=1}^{n} b_{nk} * x_k$ (k = 1…n), both the estimated parameters and the standard Student test statistics are influenced by (multi)collinearity, quantified by the coefficient of collinear refraction (T_{nk}) because:

$$a_n = \overline{y} - \sum_{k=1}^{n} b_{nk} * \overline{x}_k \ , \tag{1}$$

where:

$\overline{y}$ = arithmetical mean of dependent variable observed values. $\overline{x}_k$

$\overline{x}$ = arithmetical mean of explanatory variable x_k observed values.

$$b_{nk} = b_{1k} * T_{nk} \ , \tag{2}$$

where: b_{1k}=estimated parameter related to explanatory variable x_k in case of unifactorial regression $y = a_1 + b_{1k} * x_k$.

$$b_{1k} = \frac{\operatorname{cov}(y; x_k)}{\operatorname{var}(x_k)} \tag{3}$$

cov $(y;x_k)$ = covariance between dependent variable and explanatory variable x_k

var (x_k) = variance of explanatory variable x_k

$$T_{nk} = \frac{\det(R_{j1}, R_{j2} \ldots R_{jk-1}, r_{jk}, R_{jk+1} \ldots R_{jn})}{\det(R_{jk})_n} \quad j = 1 \ldots n, \tag{4}$$

where:

$\det(R_{jk})_n$ = determinant of matrix of Pearson coefficient of correlation between explanatory variables x_j and x_k.

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

where:

R $(x_j; y)$ = Pearson coefficient of correlation between explanatory variables x_j and dependent variable.

R $(x_k: y)$ = Pearson coefficient of correlation between explanatory variables x_k and dependent variable.

Formula (2) shows that in a multiple regression estimated parameters b_{nk} represent the products between the estimated value of the respective parameters in case of the unifactorial linear regression (b_{1k}) and the coefficients T_{nk}. We note that coefficients T_{nk} are influenced by collinearity represented by the values of the Pearson coefficients of correlation between the explanatory variables (R_{jk}). If all the considered explanatory variables are orthogonal, i.e. all coefficients R_{jk} are equal to zero, all the coefficients T_{nk} are equal to one. As a consequence, b_{1k} may be considered as the proper value of the estimated parameter and b_{nk} as the derived value of the estimated parameter, because it is influenced by collinearity (non-orthogonality) between the explanatory variables (F.M. Pavelescu, 1997).

In F. M. Pavelescu (2012) and F. M. Pavelescu (2014) it was demonstrated that in case of linear regressions with two and three explanatory variables we may express T_{nk} as:

$$T_{nk} = \frac{1 - p_{jkwam} * R^2_{(n-1)xk}}{1 - R^2_{(n-1)xk}}, \tag{6}$$

where:

p_{jkwam}= weighted arithmetical mean of the statistics p_{jk}, where:

$$p_{jk} = \frac{R(x_j;y)}{R(x_k;y)} * \frac{1}{R(x_j;x_k)} \quad x_j \neq x_k \tag{7}$$

$R(x_j;x_k)$ = Pearson coefficients of correlation between explanatory variable x_j and explanatory variable x_k.

N.B. The weighting elements are the co-factors of the determinant $(R_{jl})_n$.

$R^2_{(n-1)xk}$= coefficient of determination of linear regression

$$x_k = a_{(n-1)} + \sum_{j=1}^{n} c_{(n-1)j} * x_j \quad (x_j \neq x_k).$$

It is important to note that statistics p_{jk} may be also written as:

$$p_{jk} = \frac{1}{b_{1k}} * \frac{b_{1j}}{c_{1j}} \ , \tag{8}$$

where:

c_{1j}= estimated proper values of parameters $c_{(n-1)j}$ from the linear regression

$$x_k = a_{(n-1)} + \sum_{j=1}^{n} c_{(n-1)j} * x_j \quad (x_j \neq x_k).$$

Formula (8) is important because it reveals that statistics p_{jk} may be viewwd as influenced by proper values of the estimated parameters of the considered linear regression and on the auxiliary linear regressions

$$x_k = a_{(n-1)} + \sum_{j=1}^{n} c_{(n-1)j} * x_j \quad (x_j \neq x_k).$$

Consequently, we may write:

$$p_{jkwam} = \frac{1}{b_{1k}} * \left(\frac{b_{1j}}{c_{1j}}\right)_{wam} \text{(9), where:}$$

$\left(\frac{b_{1j}}{c_{1j}}\right)_{wam}$ = weighted arithmetical mean of ratios $\frac{b_{1j}}{c_{1j}}$.

Formulae (6), (7), (8) and (9) can be rigorously demonstrated in case of linear regression with two and three explanatory variables. This fact sustains the idea of generalization of the respective formulae in case of linear regressions with n explanatory variables. But the demonstration of the above-mentioned formulae in case of linear regressions with more than three explanatory variables becomes difficult due to the explosive increase in number of modelling factors of coefficients T_{nk}. In order to surpass these impediments in the present paper we will build a numerical example which will confirm that the above-mentioned formulae are also valid when the number of explanatory variables are greater than three.

In fact, coefficients T_{nk} act as indicators of collinearity. If, in a linear regression, all the respective coefficients are positive, we may speak about a non-harmful collinearity. If, in a linear regression, at least one of coefficients T_{nk} are negative, we faced a harmful collinearity. In fact, negative coefficients T_{nk} are the cause of the occurrence of "wrong signs" of the estimated parameters in linear regressions.

If we carefully examine the Formula (6) we may detect that coefficients T_{nk} act for the polarization of the proper values of the considered parameters in a linear regression. The polarization of the estimated parameters b_{nk} is greater as coefficients of determination $R^2_{(n-1)xk}$ increase and statistics p_{jkwam} are different from 1. Therefore, we may establish an analogy between the coefficients T_{nk} and indices of refraction of light in different media. Consequently, we consider that coefficients T_{nk} may be named as "coefficients of linear refraction"[1].

2. Advantages of defining the Transformed Form of the Student Test statistics (TFST) and of re-grouping of its components

Coefficients of collinear refraction influence not only the size of the estimated parameters in multiple linear regressions, but also, the Student Test statistics related to the respective parameters. If we have in mind the definition of the above-mentioned statistics and formula (2), we may express the Student test statistics related to estimated parameter x_k (t_{bnk}) as follows:

---

[1] The occurrence of the coefficients T_{nk} in linear regressions, when OLS method is used, was firstly identified in F. M. Pavelescu (1986) and were named "coefficients of alignment". In F.M. Pavelescu (2009) they were re-named as "coefficient of alignment to collinearity hazard" in order to show that there are many cases, especially when the number of explanatory variable is greater than three, the values of coefficients T_{nk} are at a first sight strange and consequently may be considered a result of hazard.

$$t_{bnk} = \sqrt{m-n-1} * R(x_k; y) * \sqrt{\frac{1-R^2_{(n-1)xk}}{1-R^2_{ny}}} * T_{nk}\,, \tag{10}$$

where:

m= number of observations

n= number of explanatory variables considered in linear regression

R^2_{ny}= coefficient of determination of the linear regression

$y = a_n + \sum_{k=1}^{n} b_{nk} * x_k$ (k = 1…n)

Formula (10) shows the modelling factors of standard Student Test statistics and emphasizes the causes of the occurrence of values which at the first sight appear to be strange. It is especially the case of harmful collinearity. In the context of the respective type of collinearity, if only the absolute values of the Student Test statistics are taken into account we face “statistical illusions”. In other words, the absolute values of Student Test statistics may be high, but some of the estimated parameters have “wrong signs” (contrary to the signs of the Pearson coefficient of correlation $R(x_k;y)$). In order to detect rapidly the occurrence of harmful collinearity in a linear regression, we may use the Transformed Form of the Student Test statistics ($TFST_{bnk}$) defined by the formula presented in F. M. Pavelescu (2014), as follows:

$$TFST_{bnk} = \sqrt{m-n-1} * \left| R(x_k; y) \right| * \sqrt{\frac{1-R^2_{(n-1)xk}}{1-R^2_{ny}}} * T_{nk} \tag{11}$$

equivalent to: $$TFST_{bnk} = t_{bnk} * \frac{/\,R(x_k; Y)\,/}{R(x_k; Y)} \tag{12}$$

It is important to note that the use of $TFST_{bnk}$ is permitted only in case of linear regressions, under the condition of maintaining the other assumptions related to the use of standard Student test.

Another advantage of the use of the Transformed Form of the Student Test statistics is that we are able to highlight the double nature of the explanatory variables in multiple linear regressions. An explanatory variable in a multiple linear regression plays a role of descriptor of dependent variable behaviour, quantified by the absolute value of the Pearson coefficient of correlation $R(x_k;y)$, on the one hand, and a role of generator of collinearity, quantified by the size of variance of inflation factor ($VIF_{(n-1)xk}$), on the other hand.

It is to note that variance of the inflation factor ($VIF_{(n-1)xk}$) is determinate by the formula:

$$VIF_{(n-1)xk} = \frac{1}{1-R^2_{(n-1)xk}} \tag{13}$$

In this context, it is possible to define the explanatory variable which is the most strongly correlated in absolute terms with the dependent variable as the primordial explanatory variable of the considered linear regression (x_1). Also, we may define the notion of coefficient of correlation between explanatory variable x_k and the primordial explanatory variable mediated by the dependent variable ($r_{(k1)y}$) by the following formula:

$$r_{(k1)y} = \frac{|R(x_k;y)|}{|R(x_1;y)|} \tag{14}$$

Therefore, we may establish the hierarchy of explanatory variables in their quality of descriptors of the dependent variable by taking into account the values of coefficients $r_{(k1)y}$.

Consequently, we may write:

$$TFST_{bnk} = \sqrt{m-n-1} * |R(x_1;y)| * r_{(k1)y} * \sqrt{\frac{1-R^2_{(n-1)xk}}{1-R^2_{ny}}} * T_{nk} \tag{15}$$

Similarly, we may identify the main generator of collinearity as being the explanatory variable with the highest $VIF_{(n-1)xk}$. Also, it is possible to establish the hierarchy of generators of collinearity by the size of $VIF_{(n-1)xk}$. In other words, we may speak about the ranks of descriptors of behaviour of dependent variable and about orders of generators of collinearity[1]. It is important to note that if the rank of the descriptors of the dependent variable behaviour coincide with the order of the generators of collinearity, favourable premises occur for avoiding harmful collinearity.

---

[1] The problem of determination of the ranks of descriptors of the dependent variable behavior and of order of generators of collineary in case of a linear regression with three explanatory variables is largely examined in F.M. Pavelescu (2014).

Because the Transformed Form of Student Test statistics is a product of several factors, we may try to re-group the respective factors within other factors with a bigger explanatory power.

Therefore, if we have in mind the ideas of Belsey (1991)[1], on the one hand, and of Glauber and Ferrar (1967)[2], on the other hand, we may write:

$$TFST_{bnk} = RV_{nk} * AdMRV_{nk} * T_{nk} \tag{16}$$

where:

RV_{nk} = Reference Value of Transformed Form of Student Test for explanatory variable x_k in case of a linear regression with n explanatory variables.

$$RV_{nk} = \sqrt{m-n-1} * \left|R(x_k ; y)\right|, \tag{17}$$

equivalent with:

$$RV_{nk} = \sqrt{m-n-1} * \left|R(x_1 ; y)\right| * r_{(k1)y} \tag{18}$$

$AdMRV_{nk}$= Adjusted Multiplier of the Reference Value of the Transformed Form of the Student Test for explanatory variable x_k.

$$AdMRV_{nk} = \sqrt{\frac{1-R^2_{(n-1)xk}}{1-R^2_{ny}}} \tag{19}$$

Reference Value of the Transformed Form of the Student Test statistics (RV_{nk}) is the only one of the three above-mentioned factors which is influenced by collinearity and is in line with the position expressed in Belsey (1991), which stated that the Student test statistics should normally reveal the values of the Pearson coefficient of correlation between the analysed explanatory variable and the dependent variable.

---

[1] We refer to Belsey's idea that the Student Test statistics are normally linked with the value of the Pearson coefficient of correlation between the analyzed explanatory variable and the dependent variable.

[2] Glauber and Ferrar (1967) considered that harmful collinearity may occur not only when at least one of the estimated parameters has a wrong sign, but also when the coefficient of determination of the multiple linear regression is smaller than the Pearson coefficient of correlation between the explanatory variables.

The definition of modelling factor $AdMRV_{nk}$ is based on the assumptions of Glauber and Ferrar (1967) related to harmful collinearity. It is important to observe that in conditions of a perfect orthogonality between explanatory variables, $AdMRV_{nk}$ is equal to $\sqrt{\frac{1}{1-R_{ny}^2}}$

In other words, $AdMRV_{nk}$ acts as multiplier of RV_{nk}. If the collinearity between explanatory variables is manifest, the respective multiplier is adjusted by square root of Tolerance (inverse of $VIF_{(n-1)xk}$).

Theoretically, if $AdMRV_{nk}<1$, the probability of occurrence of harmful collinearity increases.

If $AdMRV_{nk}>1$, theoretically, there are better conditions to deal with non-harmful collinearity.

3. A factorial analysis of dispersion of the Transformed Form of the Student Test statistics

The identification of modelling factors of Student test statistics both in its standard and Transformed Form permits to find the causes which enable or hamper the validation of estimation for each of the estimated parameters. Therefore, if the number of considered explanatory variables increases, we may face situations when some of the estimated parameters are validated and the others are not. In such situations, it is difficult to draw conclusions on the quality of estimation of the considered linear regression equation as a whole. A solution to the above-mentioned problem is to build a synthetic indicator which considers both the arithmetical mean and standard deviation of the TransFormedForm of the Student test statistics. The proposed analytical tool will be named "Synthesis of the Transformed Form of the Student Test statistics"[1].

---

[1] The idea of defining a "Synthesis of the Transformed Form of the Student Test statistics" was first exposed in F.M. Pavelescu (2014) when analyzing possible ways to improve the use of the Student test in the context of linear regressions with three explanatory variables. In this context, we found that the arithmetical mean of the Transformed Form of the Student Test statistics is correlated with the square root of the Fisher Test statistics. It was demonstrated that the deviation of the respective arithmetical mean from the square root of the Fisher Test statistics is greater and greater as the intensity of collinearity increases.

We think that the quality of a linear regression implies a value for the Arithmetical Mean of the Transformed Form of the Student Test statistics ($TFSTAM_{m;n}$) higher than the tabelled one, a low level for the Coefficient of Variation in Transformed Form of Student Test statistics ($TFSTCV_{m;n}$) and there is no harmful collinearity. In this context, it is possible to obtain the validation of estimation for most of the estimated parameters.

If we have in mind the modelling factors of the TransFormedForm of Student Test statistics, we may write computation Formulae for the two components of Synthesis of TransFormedForm of Student Test statistics, which reveal the interdependence between the considered factors.

Therefore, the Arithmetical Mean of the TransFormedForm of the Student Test statistics ($TFSTAM_{m;n}$) may be expressed as:

$$TFSTAM_{m;n} = (RV_{nk})_{AM} * (AdMRV_{nk})_{AM} * I_{WAMRV_{nk}} * (T_{nk})_{AM} * I_{WAMRV_{nk}*AdMRV_{nk}} \quad (20)$$

$$I_{WAMRVnk} = \frac{(AdMRV_{nk})_{WAMRVnk}}{(AdMRV_{nk})_{AM}} \quad (21)$$

$$I_{WAMRV_{nk}*AdMRV_{nk}} = \frac{(T_{nk})_{WAM\ RVnk*AdMRVnk}}{(T_{nk})_{AM}} \quad (22)$$

where:

$(RV_{nk})_{AM}$ = simple arithmetical mean of the Reference Values of the Transformed Form of the Student Test statistics

$(AdMRV_{nk})_{AM}$ = simple arithmetical mean of the Adjusted Multipliers of the Reference Values of the Transformed Form of the Student Test statistics

$(AdMRV_{nk})_{WAMRVnk}$ = weighted arithmetical mean of the Adjusted Multipliers of the Reference Values of the Transformed Form of the Student Test statistics, the weighting factor being RV_{nk}

$(T_{nk})_{AM}$ = simple arithmetical mean of the coefficients of collinear refraction

$(T_{nk})_{WAMRVnk*AdMRVnk}$ = weighted arithmetical mean of the coefficients of collinear refraction, the weighting factor being the product RV_{nk}* $AdMRV_{nk}$

It is to note that $TFSTAM_{m;n}$ is favourably influenced by the simple arithmetical means of RV_{nk}, $AdMRV_{nk}$ and T_{nk}. The weighted arithmetical mean of $AdMRV_{nk}$ is positively influenced by the nonconcordance between the ranks of descriptors of behaviour of the dependent variable and orders of generators of collinearity, but at the price of favouring the occurrence of harmful collinearity and/or of $TFSTCV_{m;n}$. In case of weighted arithmetical mean of efficient of collinear refraction ($(T_{nk})_{WAMRVnk*AdMRVnk}$) we may expect that the respective statistics will be greater than the simple arithmetical mean if we face harmful collinarity.

$TFSTCV_{m,n}$ values may oscillate between 0 and $(n-1)^{0.5}$ if we deal with non-harmful collinearity. If negative coefficients of collinear refraction occur, $TFSTCV_{m,3}$ may take on higher values, even bigger than $(n-1)^{0.5}$.

If we determine the coefficients of variation in the reference values (RV_{nk}), of products (RV_{nk}*AdM RV_{nk}) and of the Transformed Form of the Student Test statistics ($TFST_{m;n}$) we may determine the contribution of the modelling factors to the dispersion of the Transformed Form of the Student Test statistics.

In other words, we first compute CV(RV_{nk}), CV(RV_{nk}*AdM RV_{nk}) and $TFSTCV_{m,n,}$ where:

CV(RV_{nk}) = coefficient of variation in the Reference Values of the Transformed Form of the Student Test statistics.

CV(RV_{nk}*AdM RV_{nk}) = coefficient of variation in products (RV_{nk}*AdM RV_{nk})

Then, it is possible to detect the contribution to $TFSTCV_{m,n}$ of :

a. Reference Values of the Transformed Form of the Student Test statistics (contrib. disperse. RV_{nk}), which is equal to CV(RV_{nk})
b. Adjusted Multipliers of the Transformed Form of the Student Test statistics (contrib. disperse. RV_{nk}), which is equal to (CV(RV_{nk}*AdM RV_{nk})- CV(RV_{nk}))
c. Coefficients of collinear refraction (contrib disperse T_{nk}), which is equal to ($TFSTCV_{m,n}$-CV(RV_{nk}*AdM RV_{nk})).

4. A numerical example. Estimation of an import function with four explanatory variables in the case of Romania

In order **t**o illustrate the possibility of practical use of the proposed factorial analysis methodology of the dispersion of the Transformed Form of the Student test statistics, we estimate the parameter of an import function of Romania with four explanatory variables during the period 1991-2009.

The dependent variable is the value (in euro) of imports in the current year (M)

As explanatory variables we consider:

a) Import value (in million Euro) in the previous year (M (-1))

b) Index of private consumption in the previous year (ICHc(-1))

c) Index of gross fixed capital Formation in the previous year (IGFCFc (-1)).

d) index of gross fixed capital Formation in the current year (IGFCFc)

In other words, we estimated the linear regression:

$$M = a_4 + b_{4M(-1)} * M(-1) + b_{4ICHc(-1)} * ICHc(-1) + b_{4IGFCFc(-1)} * IGFCF(-1) + b_{4IGFCFc} * IGFCFc$$

The OLS estimation has given the following results:

M=-27.8350+0.8756*M(-1)+11.6394*ICHc(-1)-8.1851*IGFCFc(-1)+27.1255*IGFCFc
(-2.0069) (8.2742) (0.8005) (-1.1365) (4.1887)

R^2_{4M} =0.9206 D-W= 2.2165 J-B= 1.3014 $F_{19;4}$= 40.5806 $(F_{19;4})^{0.5}$=6.3703

N.B. In the brackets the computed values of the standard Form of the Student Test are mentioned.

R^2_4 is the coefficient of determination for the considered import function.

D-W = Durbin-Watson Test statistics.

J-B= Jarque-Bera Test statistics.

$F_{19;4}$= Fisher Test statistics, $(F_{19;4})^{0.5}$= square root of Fisher Test statistics

Estimation of simple linear regressions in order to detect the sign of the Pearson coefficients of correlation between an explanatory variable and the dependent variable (R (x_k;y) and the coefficient of collinear refraction (T_{nk}) has given the following results:

M= 3.6357+ 0.9029*M(-1) R (M(-1); M) = 0.8993
(1.0967) (8.4764) R^2 (M(-1);M) = 0.80866

M= -58.2992+ 84.9766*ICHc(-1) R(ICHc(-1);M) = 0.6135
(-2.1048) (3.203122) R^2(ICHc(-1);M) = 0.37637

M= -9.5927+ 37.7638*IGFCFc(-1) R(IGFCFc(-1);M) = 0.5942
(-0.7259) (3.0460) R^2(IGFCFc(-1);M) = 0.35307

M= -1.9339+ 30.2936*IGFCFc R(IGFCFc(-1);M) = 0.4245
(-0.1153) (1.9329) R^2(IGFCFc(-1);M) = 0.18018

We should note that all the Pearson coefficients of correlation between each explanatory variables and the dependent variable are positive. Consequently, the correction factor $\frac{|R(x_k;y)|}{R(x_k;y)}$ used for defining the Transformed Form of the Student Test is equal to 1 for all the four explanatory variables. The Pearson coefficients of correlation between the explanatory variables and dependent variables range between 0.42488 and 0.89926. It results that the coefficients of correlation between the explanatory variables and primordial explanatory variable seen as descriptors of the dependent variable behaviour, mediated by the dependent variable ($r_{(k1)M}$) take on values between 0.4720 and 1.000 (Table 1).

In this context, it is possible to reveal the ranks of the explanatory variables, i.e. the primordial explanatory variable is M(-1), the explanatory variable of rank 2 is ICHc(-1), the explanatory variable of rank 3 is IGFCF(-1) and the explanatory variable of rank 4 is IGFCFc.

Table 1 - Coefficients of collinear refraction and coefficients of correlation mediated by dependent variable in linear regression in the case of import function of Romania, 1991-2009

X_k	b_{4X}	b_{1X}	T_{4k}	$R(X_kM)$	$r_{(k1)M}$
M(-1)	0.87563	0.90290	0.96980	0.89926	1.0000
ICHc(-1)	11.63994	84.97665	0.13698	0.61349	0.6822
IGFCF(-1)	-8.18509	37.76382	-0.21674	0.59420	0.6608
IGFCF	27.12553	30.29363	0.89542	0.42448	0.4720

The coefficient of collinear refraction in case of explanatory variable IGFCF(-1) is negative (-0.2167) and is a signal that a harmful collinearity is present in the linear regression.

In order to quantify the other modelling factors of $TFST_{19;4}$, we estimate the following four linear regressions related to each explanatory variable.

M(-1)= -53.5339+ 64.1913*ICHc(-1)+ 22.5584*IGFCFc(-1) - 7.2285*IGFCFc
(-1.7331) (2.0465) (1.3608) (-0.4607)

$R^2_{3M(-1)}$ = 0.4895

ICHc(-1)= 0.8504+ 0.1631*IGFCFc(-1) -0.0762*IGFCFc + 0.0034*M(-1)

(7.6223) (1.3506) (-0.6726) (2.0465)

$R^2_{3ICHc(-1)}$ = 0.4854

IGFCFc(-1)= -0.3058 + 0.4938*IGFCFc + 0.0049*M(-1) + 0.6648*ICHc(-1)

(-0.6228) (2.5452) (1.3608) (1.3506)

$R^2_{3IGFCF(-1)}$ = 0.5583

IGFCFc = 0.8742 - 0.0019 *M(-1) - 0.3842 *ICHc(-1) + 0.6108 *IGFCFc(-1)

(1.7315) (-0.4607) (-0.6726) (2.5452)

$R^2_{3IGFCFc}$ = 0.3112

Based on the above-mentioned linear regressions, we are able to compute the square root of the inverse of VIF (tolerance) for each of the considered explanatory variables (Table 2).

Table 2 - Square root of inverse of VIF for the considered explanatory variable in the import function of Romania, 1991-2009

X_k	R^2_{3Xk}	$(1- R^2_{3Xk})^{0.5}$
M(-1)	0.4895	0.7145
ICHc(-1)	0.4854	0.7174
IGFCF(-1)	0.5583	0.6646
IGFCF	0.3112	0.8299

We may notice that the orders of generators of collinearity are not the same with the ranks of explanatory variables seen as descriptors of the dependent variable behaviour. Therefore, the main generator of collinearity is IGFCFc(-1), the generator of collinearity of order 2 is M(-1), the generator of collinearity of order 3 is ICHc(-1) and the generator of collinearity of order 4 is IGFCFc.

The computation of the Transformed Form of the Student Test statistics based on Formula (15) produces practically the same results in absolute values when we use the Excel software (Table 3). This is in fact a confirmation that formula (6) is valid not only in case of linear regressions with two and three explanatory variables, but also in general case of a linear regression with n explanatory variables.

Table 3 - Modelling factors of the Transformed Form of the Student Test statistics in case of the import function of Romania, 1991-2009

X_k	RV_{4k}	$AdMRV_{4k}$	T_{4k}	$TFST_{b4k}$
M(-1)	3.36471	2.53570	0.96980	8.27422
ICHc(-1)	2.29549	2.54597	0.13698	0.80053
IGFCFc(-1)	2.22328	2.35854	-0.21674	-1.13654
IGFCFc	1.58824	2.94534	0.89542	4.18870

The reference values of the Transformed Form of the Student Test statistics (RV_{4k}) range between 1.58824 and 3.36471, while the adjusted multiplier ($AdMRV_{4k}$) is greater than 2.35 for all the explanatory variables. Theoretically, in this context the possibility of occurrence of harmful collinearity should be low. But this favourable premise is not sufficient for ensuring positivity for all the coefficients of collinear refraction (T_{nk}).

The simple arithmetical mean of the Transformed Form of the Student Test statistics ($TFSTAM_{19;4}$) is equal to 3.0137, and bigger than tabelled value, 2.145, respectively. The ratio $(TFSTAM_{19;4})/(F_{19;4})^{0.5}$ is equal to 0.4731. At the same time we have: $(RV_{4k})_{AM}$ = 2.3679, $(AdMRV_{4k})_{AM}$ = 2.5964, $(AdMRV_{4k})_{WAMRV4k}$ = 2.5653, $(T_{4k})_{AM}$ = 0.4464, $(T_{4k})_{WAMRV4k*AdMRV4k}$ =0.4991. At first sight, $TFSTAM_{19;4}$ takes on values which describe a moderate impact of collinearity on the estimation process. But it is very important that respective results are obtained in a context marked by the occurrence of harmful collinearity and consequently the quality of estimation is sensibly reduced.

We should note that, even in conditions of a discordance between the ranks of the descriptors of the dependent variable behaviour and the order of generators of collinearity, the weighted arithmetical mean of adjusted multipliers ($(AdMRV_{nk})_{WAMRVnk}$) is smaller than the simple one ($(AdMRV_{nk})_{AMRVnk}$). This is a consequence of the fact that the above-mentioned nonconcordance is not of large dimensions. In case of coefficients of collinear refraction the weighted arithmetical mean is bigger than the simple arithmetical mean, but in conditions of harmful collinearity.

The coefficient of variation in the Transformed Form of the Student test statistics ($TFSTAM_{19;4}$) is equal to 1.3383, while $CV(RV_{nk})$ = 0.2204 and $CV(RV_{nk}*AdM RV_{nk})$ = 0.2354.

Therefore we have: contrib. dispers. RV_{nk}= 0.2204, contrib disperse $AdMRV_{nk}$ =0.0150, contrib disperse T_{nk}= 1.1029.

The main contribution to the dispersion of the Transformed Form of the Student test statistics is given by dispersion of coefficients of collinear refraction, being a consequence of harmful collinearity.

5. Conclusions

The impact of collinearity on the estimated parameters of econometric models should not be ignored. In fact, the respective phenomenon occurs in case of all estimation methods.

The ordinary square least square (OLS) method has several advantages derived not only from the easiness of use, but also from the fact that the impact of collinearity in case of a multiple linear regression can be emphasized distinctly by the coefficient of collinear refraction (T_{nk}) or variance inflation factor (VIF_{nk}). The respective statistics have an important impact not only on the estimated parameters values, but also on the Student Test statistics.

It is important to have in mind that the values of VIF create only the premises for a certain type of collinearity and implicitly for distortions in estimated parameters and the Student Test statistics. The above-mentioned distortions are ultimately caused by the differentiation of the absolute values of Pearson coefficients of correlation between the explanatory variables and dependent variable.

In fact, distortions in estimated parameters and Student Test statistics may be manifest even in conditions of relatively moderate values of VIF, when the number of explanatory variables considered in linear regression equation becomes greater and greater.

The presence of harmful collinearity can be directly detected, if a Transformed Form of the Student Test statistics (TFST) is used. The proposed transformation of the standard Student Test statistics is mainly aimed at detecting directly the signs of the coefficients of collinear refraction and does not change the other premises of the above-mentioned statistical test.

The definition of the new statistics related to the standard OLS linear regression methodology does not exclude the use of other methods of estimation in order to obtain better results. But it is important to keep in mind that the collinearity is anyhow present, due to interdependence of the explanatory variables. Therefore, it is necessary to build and/or improve a system of indicators or statistics able to detect the impact of collinearity on estimated parameters and especially the occurrence of harmful collinearity.

An example of new indicator aimed to provide new information on the estimation quality of linear regressions as a whole could be the "Synthesis of the Transformed Form of the Student Test statistics". The respective indicator may reveal the interdependence between the Students statistics of the explanatory variables considered in a linear regression. Also, the above-mentioned indicator favours the idea that a good estimation is one in which the Student test statistics are not only higher in absolute values, but also they are less dispersed.

References

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

Belsey D. (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.

C.Dougherty (2002) – *Introduction to econometrics*, second edition, Oxford University Press.

D. Farrar, R. Glauber (1967) – "Multicollinearity in regression analysis: The problem revisited", *Review of Economics and Statistics,* February.

W. Greene (1993) – *Econometric Analysis*, Mc Millan Publishing Company, New York.

A. Isaic-Maniu, C. Mitruţ, V. Voineagu (1996) – *Statistica pentru managementul afacerilor* ("Statistics for Business Management"), Economic Publishing House, Bucharest. (in Romanian)

L. Klein (1962) – *An introduction to econometrics*, Prentice Hall.

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 (1997) – *Progresul tehnologic şi ocuparea forţei de muncă*, Editura IRLI, București.

F .M. Pavelescu (2009) – "A review of Student Test properties in condition of multifactorial linear regression", *Romanian Journal of Economic Forecasting,* No.1.

F. M. Pavelescu (2012) – "Interdependenţa dintre tipul de colinearitate şi valorile calculate ale testului Student în cazul unei regresii lineare cu două variabile resultative", *Caietul de Studiu nr. 25 al Seminarului de Modelare Macroeconomică al INCE.*

F. M. Pavelescu (2014) – "An extension of the methodology of using the Student Test in case of a linear regression with three explanatory variables", *Romanian Journal of Economics,* No.1.

E. Pecican (1996) – *Macroeconometrie. Politici economice guvernamentale şi econometrie* ("Macroeconometrics. Gouvernamental economic policies and econometrics"). Economic Publishing House, Bucharest.

R. Wiliams (2008) – *Sociology Graduate Statistics*, University of Notre Dame, Indiana.