16. AN ANALYSIS MODEL FOR THE DISTURBANCES GENERATED BY COLLINEARITY IN THE CONTEXT OF THE OLS METHOD

Florin Marius PAVELESCU*

Abstract

Under the conditions of OLS use in order to perform multiple linear regressions, both the estimated parameters values and also computed values of some statistical tests such as coefficient of determination, Fisher test or Student test are influenced by collinearity[1]. The respective influence is revealed by the values of the coefficient of alignment to collinearity hazard. For this reason, this paper presents an analysis model which identifies the factors and their influences on the above-mentioned indicator. On the one hand, we quantify the factors contribution to the arithmetical mean of coefficients of alignment to collinearity hazard, having in view that the respective indicator reveals the collinearity impact on a linear regression model as a whole. On the other hand, we emphasize the necessary conditions for the positivity of all coefficients of alignment to collinearity hazard, in order to avoid the occurrence of "unexpected signs" in case of some estimated parameters. Also, we bring some clarifications and extensions of the other concepts previously proposed by the author such as: the main and secondary explanatory variable, coefficient of mediated by resultative variable correlation between explanatory variables.

Key-words: arithmetical mean of coefficients of allignament to collinearity hazard, main explanatory variables, secondary explanatory variables, coefficient of correlation between explanatory variables mediated by resultative variable, structural constraints of a multiple linear regression model.

JEL Classification: C13, C20, C51, C52

1. The usefulness of an analysis model of factors contributions to computed values of coefficients of allignament to collinearity hazard

Under the conditions of OLS use, the coefficient of allignament to collinearity hazard related to explanatory variable x_k in a linear regression with n explanatory variables

* nstitute of National Economy, The Romanian Academy, e-mail: pavelescu.florin@yahoo.com

[1] It is to be noticed that collinearity generates distortions in estimations results not only when OLS is used, but also in other estimation methods.

(T_{nk}) may be written as a ratio between two determinants (F. M. Pavelescu, 1986), as follows:

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

where:

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

$$r_{jk} = \frac{R(x_j; y)}{R(x_k; y)} \qquad (2)$$

where: R (x_j; y) = Pearson coefficient of correlation between explanatory variables x_j and resultative variable;

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

Ratio r_{jk} was defined in (F. M. Pavelescu, 2005) as "a coefficient of correlation between explanatory variables x_j and x_k mediated by resultative variable, related to explanatory variable x_k".

In case of a multiple linear regression, the coefficient of alignment to collinearity hazard has an impact both on estimated parameters related to explanatory variables (b_{nk}), and also on the computed values of some statistical tests such as: coefficient of determination (R_n^2), Fisher test ($F_{m,n}$), Student test (t_{bnk}).

Therefore, the parameters values estimation of a linear regression with n explanatory variables, respectively

$$y = a_n + \sum_{k=1}^{n} b_{nk} * x_k \qquad (3)$$

can be written (F. M. Pavelescu, 1986):

$$b_{nk} = b_{1k} * T_{nk} \qquad (4)$$

where: b_{1k} = estimated value of parameter b in case of unifactorial regression $y = a_1 + b_{1k} * x_k$.

For this reason, b_{1k} may be considered as the proper estimated value of the parameter. As a consequence, b_{nk} represents the derived estimated value of the parameter taken into account, that is, influenced by the intensity of (multi)collinearity between explanatory variables.

On the other hand, the computation formula for the coefficient of determination in case of a linear regresion with n explanatory variables (R^2_n)[2] may be written:

[2] The coefficients of alignment to collinearity hazard influence also the computed values of the Fisher and Student tests. The correlations between the Fisher and Student tests computed values and the implications of coefficients of alignment to collinearity values for the respective correlation are discussed in F.M. Pavelescu (2009).

$$R_n^2 = \sum_{k=1}^{n} R^2(x_k, y) * T_{nk}\ , \qquad (5)$$

equivalent with:

$$R_n^2 = (\sum_{k=1}^{n} R^2(x_k, y)) * (T_{nk})_{meds} * (1 + sc\,\text{cov}(R^2(x_k; y); T_{nk})) \qquad (6)$$

where:

$(T_{nk})_{meds}$= arithmetical mean of coefficients of alignment to collinearity hazard.

sccov($R^2(x_k$, y) = structural component of covariance between $R^2(x_k$:y) and T_{nk}.

$$sc\,\text{cov}(R^2(x_k; y); T_{nk}) = \frac{n * \sum_{k=1}^{n} R^2(x_k; y) * T_{nk}}{(\sum_{k=1}^{n} R^2(x_k; y)) * (\sum_{k=1}^{n} T_{nk})} - 1 \qquad (7)$$

It is important to mention that the occurrence of negative values for some coefficients of alignment to collinearity hazard emphasize the surpassing of a critical level of collinearity that leads to unfeasible results (F.M. Pavelescu, 2004)[3]. Also, it has to take into account both the individual values and the arithmetical mean of coefficients of alignment to collinearity hazard. This way, significant information could be furnished on the efficiency of utilization of explanatory variables as a whole in a multiple linear regression.

In these conditions, building an analysis model in order to identify the factors and their contributions to the acquirement of values for each coefficient of alignment to collinearity hazard and their arithmetical mean become a useful approach in order to improve the selection of explanatory variables in a linear regression model.

2. Factorial analysis of coefficients of alignment to collinearity hazard in case of regression with two explanatory variables

In a linear regression with two explanatory variables, if we take into account the absolute values of the Pearson coefficient of correlation between the resultative variable and the explanatory variables we can define as the main explanatory variable (x_p) the explanatory variable which is more intensively correlated with the resultative variable. The other explanatory variable is defined as a secondary one (x_s). The

[3] Usually, the existence of an important degree of collinearity is revealed by very small computed values of the Student test and high values of coefficient of determination. But, also the unexpected signs of the estimated parameters are considered a case of collinearity. (C. Conrad, 2006). In fact, the respective situation is a consequence of the negative values of some coefficients of alignment to the collinearity hazard. The idea that an unexpected sign of the estimated parameters is a wrong one and consequently the result obtained is incorrect appear also in W.H. Greene (1993, pg.210)

coefficients of alignment to collinearity hazard for the main explanatory variable (T_{2p}) and for the secondary variable (T_{2s}), respectively are:

$$T_{2p} = \frac{1 - R(x_p; x_s) * r}{1 - R^2(x_p; x_s)} \quad (8)$$

$$T_{2s} = \frac{1 - R(x_p; x_s) * (1/r)}{1 - R^2(x_p; x_s)} \quad (9)$$

$$r = \frac{R(x_s; y)}{R(x_p; y)} \quad (10)$$

where: R (x_p, y) = Pearson coefficient of correlation between the main explanatory variable and the resultative variable;

R (x_s, y) = Pearson coefficient of correlation between the secondary explanatory variable and the resultative variable;

R(x_p, x_s) = Pearson coefficient of correlation between explanatory variables;

r = coefficient of correlation between explanatory variables x_s and x_p mediated by resultative variable, related to explanatory variable $x_{p.}$

It may be noticed that /r/<1. As a rule, R(x_p;x_s)*r>0. In these conditions, both coefficients of alignment to collinearity hazard are positive if /r/>/R (x_p;x_s)/.

If /r/</R (x_p;x_s)/, the values of the respective coefficients are: T_p >1 and T_s <0.

It can be concluded that from the coefficients of alignment to collinearity hazard point of view the main explanatory variable takes the “benefits” of the absolute values differentiation of Pearson coefficients of correlation between resultative variables and the explanatory variables, while the secondary explanatory variable take the “costs” of above-mentioned differentiation. Also, the factors r, T_{2p} and T_{2s} have an important impact on the value of the coefficient of partial correlation between explanatory variables x_p and x_s (see Annex 1).

It may be noticed that:

$$T_p > \frac{1}{1 + / R(x_p; x_s) /} \text{ (11), and } T_s < \frac{1}{1 + / R(x_p; x_s) /} \quad (12)$$

As an exception to the rule, **the situation of anticollinearity can be defined if R(x_p;x_s)*r<0**. In this case, both coefficients of alignment to collinearity hazard are bigger than 1 (F. M. Pavelescu, 2009).

The arithmetical mean of the coefficients of alignment to collinearity hazard $(T_{2ps})_{meds}$ takes the following values:

$(T_{2k})_{meds}$<0.5, if /r/>/R(x_p;x_s)/ and R(x_p;x_s)*r>0

0.5 <$(T_{2k})_{meds}$<1 if /r/>/R(x_p;x_s)/ and R(x_p;x_s)*r>0

$(T_{2k})_{meds}$ >1 if R(x_p;x_s)*r<0.

The factors that determine the arithmetical mean of the coefficients of alignment to collinearity hazard are, on the one hand, the absolute value of Pearson coefficients of correlation between explanatory variables (R(x_p;x_s)) and on the other hand, the absolute value of the coefficient of correlation between explanatory variables x_p and x_s, mediated by resultative variable, related to the main explanatory variable (r).

If $R(x_p;x_s)*r>0$ the factorial influences on the obtaining of the respective values are:

a) influence of the absolute value of Pearson coefficient of correlation between the two explanatory variables (infl. $/R(x_p;x_s)/$)

$$\inf l./R(x_p;x_s)/=\frac{1}{1+/R(x_p;x_s)/} \quad (13)$$

b) influences of the value of the coefficient of correlation between the explanatory variables x_p and x_s mediated by the resultative variable, related to main explanatory variable on the coefficient of alignment to collinearity hazard of main explanatory variable (infl.r (T_p)), of secondary explanatory variable (infl.r (T_s)) and on the arithmetical mean of the two above-mentioned coefficients (infl. r (($T_{2ps})_{meds}$) may be determined using the following formulae:

$$\inf l.r(T_p)=\frac{1-R(x_p;x_s)*r}{1-/R\ (x_p;x_s)/} \quad (14)$$

$$\inf l.r(T_s)=\frac{1-R(x_p;x_s)*\frac{1}{r}}{1-/R\ (x_p;x_s)/} \quad (15)$$

$$\inf l.r(T_{2ps})_{med}=\frac{(1-\frac{R(x_p;x_s)*(r+\frac{1}{r})}{2})}{1-/R\ (x_p;x_s)/} \quad (16)$$

If the situation of anticollinearity occurs, it has to be taken into account that usually $R(x_p;x_s)$ takes absolute values near zero. As a consequence, for correctness reasons, in the computation formulae $/R(x_p;x_s)/$ has to be replaced with $-/R(x_p;x_s)/$. This way, it can emphasize the fact that very small absolute values of Pearson coefficient of correlation between explanatory variables and the divergence of signs in case of the two kinds of coefficients of correlation between explanatory variables taken into account (Pearson and mediated by resultative variable) enable anticollinearity.

3. Modelling factors of the coefficients of alignment to collinearity hazard values in case of linear regressions with n explanatory variables (n>2)

In case of a linear regression with more than two expalanatory variables the methodology of factor contribution quantification of coefficients of alignment to collinearity hazard values increases its complexity. Due to this fact, for the factor influence identification a series of changes in the computation formulae are operated, as follows:

a) the highest absolute value of Pearson coefficient of corelation between explanatory variables is emphasized. Afterwards, if it is neccesary, the determinants computation is transformed, so that the respective indicator to be positive and then it is denoted by R_{max}.

b) explanatory variables are ordered having in view the absolute values of Pearson coefficients of correlation between explanatory variables and resultative variable. Then, the explanatory variable that in absolute terms is most intensively corelated with the resultative variable is defined as the primordial explanatory variable of the multiple linear regression model and is denoted with x_1, while the explanatory variable which in absolute terms is least correlated with the resultative variable, is denoted by x_n.

On this base, we compute the coefficients of correlation mediated by resultative variable between explanatory variables, specially in case of sucessive explanatory variables ($r_{k+1,\ k}$), and also in case of the weakest and strongest correlated explanatory variables with the resultative variable (r_{nmin}), as follows:

$$r_{k+1,k} = \frac{R(x_{k+1}; y)}{R(x_k; y)} \tag{17}$$

$$r_{n,1} = r_{n\min} = \frac{R(x_n; y)}{R(x_1; y)}. \tag{18}$$

Also, the computation formula is transformed, so that the positivity of all coefficients of correlation mentioned above is obtained. As a rule, the respective transformations determine positive values both for R_{max} and for each of coefficients $r_{k+1,\ k}$.

In this context, the explanatory variables ranking related to the intensity of correlation in absolute terms with resultative variable enable **a redefinition of the notion of correlation between explanatory variables, mediated by resultative variable.**

We suggest that the respective indicator to be computed only related to the explanatory variable that is stronger correlated with resultant variable. On this basis, the **indicator $r_{k+i,\ k}$, may be defined as coefficient of correlation between explanatory variables x_{k+i} şi x_k mediated by resultative variable.** This way, the respective indicator takes positive and underunitary values. As a consequence, the **expression $r_{k,\ k+i}$ may be defined as the inverse of the coefficient of correlation between explanatory variables x_{k+i} şi x_k mediated by resultative variable.**

Therefore, we are able to conclude that the coefficients of alignment to collinearity hazard values are determinated by the size of $R_{max,}$ and r_{nmin}, on the one hand, and by the differentiation of the two kinds of coefficients of correlation between explanatory variables (Pearson and mediated by resultant variable), on the other hand.

In order to reveal the above-mentioned factors influences for obtaining the arithmetical mean of coefficients of alignment to collinearity hazard it is necessary to compute also other indicators:

a) **arithmetical mean of coefficients of alignment to collinearity hazard in conditions of non-differentiation** of two kinds of coefficients of correlation between explanatory variables (Pearson and mediated by resultant variable) **(T_{nR}).**

Because in this situation all $R(x_j, x_k)=R_{max}$, ($j\neq k$) and all $r_{k+1,k}=1$

$$T_{nR} = \frac{1}{1+(n-1)*R_{\max}} \quad (19)$$

b) arithmetical mean of coefficients of alignment to collinearity hazard in conditions of non-differentiation of Pearson coefficients of correlation between explanatory variables, on the one hand, and of particular differentiation of coefficients of correlation between explanatory variables mediated by resultant variable, on the other hand ($(T_{nRrpdk})_{med}$).

$$(T_{nRpdfk})_{med} = \frac{1+(n-1)*R_{\max} - (\frac{1}{n})*\frac{R_{\max}}{r_{n\min}}*(\sum_{k=1}^{n} r_{k1})^2}{1+(n-2)*R_{\max} - (n-1)*R_{\max}^2}. \quad (20)$$

If R_{max} and r_{nmin} are given, the maximum value of $(T_{nRefk})_{med}$ is obtained if $r_{k+1,k} = \sqrt[n-1]{r_{n\min}}$.

It may be noticed that in this case the coefficients of correlation between explanatory mediated by resultant variable are differentiated in a geometrical progression. We define this situation as **an ordered differentiation of coefficients of correlation between explanatory variables mediated by resultative variable.** Also, for this case we denote the arithmetical mean of coefficients of alignment to collinearity hazard by $(T_{nRrordk})_{med}$

$$(T_{nRrordk})_{med} = \frac{1+(n-1)*R_{\max} - (\frac{1}{n})*(\frac{1-r_{n\min}*\sqrt[n-1]{r_{n\min}}}{1-\sqrt[n-1]{r_{n\min}}})^2*\frac{R_{\max}}{r_{n\min}}}{1+(n-2)*R_{\max} - (n-1)*R_{\max}^2}. \quad (21)$$

It can be observed that all coefficients of alignment to collinearity hazard are positive if:

$$\frac{r_{n\min}*(1-\sqrt[n-1]{r_{n\min}})}{1-r_{n\min}*\sqrt[n-1]{r_{n\min}}} > \frac{R_{\max}}{1+(n-1)*R_{\max}} \quad (22)$$

The minimum value of arithmetical mean of coefficients of alignment to collinearity hazard in the context of the differentiation of **coefficients of correlation between explanatory variables mediated by resultative variable** and non-differentiation of Pearson coefficients of correlation between explanatory variables is acquired in two situations if:

1) coefficients of correlation between explanatory variables mediated by resultative variable are differentiated just between primordial explanatory variable and the explanatory variable of rank 2 ($r_{21} = r_{n\min}$ şi $r_{k+1,k} = 1$ (k=2…n)).

In this case, the condition which has to be fullfilled in order that all the coefficients of alignment to collinearity hazard to be positive is: $r_{n\min} > R_{\max}$ (23)

2) **coefficients of correlation between explanatory variables mediated by resultative variable** are not differentiated between the first (n-1) explanatory

variables, the differentiation taking place just between the explanatory variables of rank n-1 şi n ($r_{n;n-1} = r_{n\min}$, $r_{k+1,k} = 1$ (k=1…n-1)).

Therefore, the necessary condition for the positivity of all coefficients of alignment to collinearity hazard is:

$$r_{n\min} > \frac{1+(n-2)*R_{\max}}{(n-1)*R_{\max}}. \tag{24}$$

It may be noticed that the weakest constraint in order to obtain the positivity of all coefficient of allignment to collinearity hazard in the context of non-differention of Pearson coefficients of correlation between explanatory is $r_{n\min} > R_{\max}$ when $r_{21} = r_{n\min}$ and $r_{k+1,k} = 1$ (k=2…n). For these reasons we define the respective situation as a **standard differentiation of coefficients of correlation between explanatory variables mediated by resultative variable.** In this case, the arithmetical mean of coefficients of alignment to collinearity hazard $(T_{nRrstk})_{med}$ can be computed by the formula:

$$(T_{nRrstk})_{med} = \frac{1+(n-2)*R_{\max} - \frac{(n-1)}{n}*R_{\max}*\frac{(1+r_{n\min}^2)}{r_{n\min}}}{1+(n-2)*R_{\max} - (n-1)*R_{\max}^2} \tag{25}$$

The strongest constraint in order to obtain the positivity of all coefficients of allignment to collinearity hazard in the context of non-differention of Pearson coefficients of correlation between explanatory is: $r_{n\min} > \frac{1+(n-2)*R_{\max}}{(n-1)*R_{\max}}$ and $r_{n;n-1} = r_{n\min}$, $r_{k+1,k} = 1$ (k=1…n-1). This situation may be called “the most postponed differentiation of **coefficients of correlation between explanatory variables mediated by resultative variable**”. In this case, the arithmetical mean of coefficients of alignment to collinearity hazard is equal to the arithmetical mean obtained in case of standard differentiation of coefficients of correlation between explanatory variables mediated by the resultant variable.

c) arithmetical mean of coefficients of alignment to collinearity hazard in conditions of non-differentiation of Pearson coefficients of correlation between explanatory variables, on the one hand, and of actual differentiation of coefficients of correlation between explanatory variables mediated by resultative variable, on the other hand ($(T_{nRrefk})_{med}$).

$$(T_{nRrefk})_{med} = \frac{1+(n-1)*R_{\max} - n*R_{\max}*\frac{r_{n1medaritm}}{r_{n1medharm}}}{1+(n-2)*R_{\max} - (n-1)*R_{\max}^2} \tag{26}$$

Analogously, the computation formulae can be writtren for factors contribution related to each individual coefficient of allignment to collinearity hazard.

Also, in the context of non-differentiation of Pearson coefficients of correlation between explanatory variables the notion of main and respectively secondary explanatory variable can be extended. Therefore, **an explanatory variable can be considered as a main one if:**

$$T_{nRrefk} > \frac{1}{1+(n-1)*R_{\max}} \tag{27}$$

and as secondary one if:

$$T_{nRrefk} < \frac{1}{1+(n-1)*R_{\max}} \tag{28}$$

The idea issued in conditions of regression with two explanatory variables remain also in the context of regression with n explanatory variables (n>2), respectively, related to the computed values, **the main explanatory variables take the "benefits" of the differentiation of coefficients of correlation between explanatory variables mediated by resultative variable., while the secondary explanatory variables take the "costs" of the above-mentioned differentiation.**

4. Quantification of factor contribution to coefficients of alignment to collinearity hazard values in case of linear regressions with n explanatory variables (n>2)

On the basis of indicators computed values mentioned above, one can emphasize, on one hand, the influences of $R_{max,,}$and r_{nmin} size, and, on the other hand, of the differentiation of the two kinds of coefficients of correlation between explanatory variables (Pearson and mediated by resultative variable) on the arithmetical mean of coefficients of allignment to collinearity hazard.

Hence, we can identify five influences on arithmetical mean of the coefficients of allignment to collinearity hazard, respectively:

1) influence of the R_{max} value (ΔR_{max} (T_{nmed}))

$$\Delta R_{\max}(T_{nmed}) = \frac{1}{1+(n-1)*R_{\max}} \tag{29}$$

2) influence of r_{nmin} value in the context of standard differentiation of coefficients of correlation between explanatory variables mediated by resultative variable ($\Delta r_{nminst}(T_{nmed})$)

$$\Delta r_{n\min st}(T_{nmed}) = \frac{1+(n-2)*R_{\max} - \frac{(n-1)}{n}*R_{\max}*\frac{1+r_{n\min}^{2}}{r_{n\min}}}{1-R_{\max}} \tag{30}$$

3) influence of ordered differentiation of coefficients of correlation between explanatory variables mediated by resultative variable ($\Delta r_{n\,ord}(T_{nmed})$)

$$\Delta r_{nord}(T_{nmed}) = \frac{1+(n-1)*R_{max} - (\frac{1}{n})*(\frac{1-r_{n\min} * \sqrt[n-1]{r_{n\min}}}{1-\sqrt[n-1]{r_{n\min}}})*\frac{R_{max}}{r_{n\min}}}{1-(n-2)*R_{max}*(\frac{1+r_{n\min}^2}{r_{n\min}})} \quad (31)$$

4) influence of actual differentiation of coefficients of correlation between explanatory variables mediated by resultative variable ($\Delta r_{n\,ef}(T_{nmed})$).

$$\Delta r_{nef}(T_{nmed})) = \frac{1+(n-1)*R_{max} - n*R_{max}*\frac{r_{1medaritm}}{r_{1medharm}}}{1+2*R_{max} - (\frac{1}{n})*(\frac{1-r_{n\min} * \sqrt[n-1]{r_{n\min}}}{1-\sqrt[n-1]{r_{n\min}}})^2*\frac{R_{max}}{r_{n\min}}} \quad (32)$$

5) influence of the differentiation of Pearson coefficients of correlation between explanatory variables related to R_{max} (Δ diff.R(x_j;x_k) (T_{nmed}))

$$\Delta diff.R(x_j;x_k)(T_{nmed}) = \frac{1}{(T_{nk})_{med}} * \frac{1+(n-1)*R_{max} - n*R_{max}*\frac{r_{1medaritm}}{r_{1medharm}}}{1+(n-2)*R_{max}-(n-1)*R_{max}^2} \quad (33)$$

Analoguously, the computation formulae can be written in order to detect the factors contribution to each individual coefficient of allignment to collinearity hazard computed values. Consequently, we can identify the primary cause that determines the occurrence of negative values for the above-mentioned indicator related to some explanatory variables.

5. A numerical example. Quantification of factorial influences to the coefficient of allignment to collinearity hazard values in case of a Felstein-Horioka model with error-correction mechanism estimated for Romania

The analysis model of the factors influences on the coefficient of allignment to collinearity hazard values previously presented will be practically applied to the estimation of the Feldstein-Horioka model with error-correction mechanism for Romania's economy during 1990-2005. The respective econometric model is: $(I/Y)_t = a + b*(S/Y)_t + c*(I/Y)_{t-1} + d*(S/Y)_{t-1}$, where:

$(I/Y)_{t-1}$, $(I/Y)_t$ = weight of investments (gross capital formation) in gross domestic product in year t-1 and t, respectively.

$(S/Y)_{t-1}$, $(S/Y)_t$= weight of conventional savings in gross domestic product in year t-1, and t respectively.

In order to determine the (multi)collinearity impact on the feasability of estimated parameters values, several kinds of linear regressions were made, as follows:

a) estimation of parameters b, c and d proper values, with the help of regressions with a single explanatory variable:

$(I/Y)_t$=9.397 + 0.691*$(S/Y)_t$
(3.350) (5.063)
R^2_1=0.664, R^2_{1adj}=0.638, F=25.638 AIC= 4.786 SCH=4.880 D-W=0.856

$(I/Y)_t$=5.496 + 0.747*$(I/Y)_{t-1}$
(1.534) (5.030)
R^2_1=0.660 R^2_{1Fadj}=0.634, F=25.310 AIC= 4.795 SCH=4.889 D-W=1.752

$(I/Y)_t$=9.659 + 0.662*$(S/Y)_{t-}$
(2.673) (3.838)
R^2_1=0.531 R^2_{adj}=0.495, F=14.732 AIC= 5.118 SCH=5.212 D-W=0.751

N.B. In brackets are presented the computed values of Student test and:
R^2_1= coefficient of determination, R^2_{1adj}= adjusted coefficient of determination
F= computed values of Fisher test, AIC = Akaike informational criterion
SCH= Schwartz criterion, D-W= computed values of Durbin-Watson test

b) estimation of linear regression for the error-correction mechanism model:

$(I/Y)_t$=5.628+ 0.468*$(I/Y)_{t-1}$ + 0.474*$(S/Y)_t$ – 0.146*$(S/Y)_{t-1}$
(1.681) (1.944) (1.818) (-0.493)
R^2_3=0.752 R^2_{3adj}=0.684, F=11.105 AIC= 4.749 SCH=4.937 D-W=1.586

Comparison of the estimated parameters values obtained in the two kinds of linear regressions shows that an unexpected sign occurs in case of explanatory variable $(S/Y)_{t-1.i}$ when the model with correction-mechanism is estimated. This is a consequence of a negative coefficient of allignament to collinearity hazard related to the above-mentioned explanatory variable in the linear regression with three explanatory variables.

In order to determine the causes that lead to a negative value in case of one of the coefficients of allignament to collinearity hazard we have firstly to compute (Table 1) all the respective coefficients (T_{3k}) and Pearson coefficient of correlation between resultative variable and explanatory variables.

Table 1

Values of coefficients of allignament to collinearity hazard and of coefficients of mediated correlation related to primordial explanatory variable in case of linear regression

$$(I/Y)_t = a + b*(S/Y)_t + c*(I/Y)_{t-1} + d*(S/Y)_{t-1}$$

Explanatory variable	T_{3k}	$R((I/Y)_t, (X_k))$	R_{k1}
$(S/Y)_t$	0.6860	0.8146	1.0000
$(I/Y)_{t-1}$	0.6265	0.8128	0.9978
$(S/Y)_{t-1}$	-0.2205	0.7289	0.8948

Computations based on: NBR-Annual Report1997, Bucharest1998, NBR Financial National Accounts 1998-2006,Bucharest,2007

On this basis, we can determinate the primordial explanatory variable, that is $(S/Y)_t$ and compute the coefficients of correlation mediated by resultative variable related to primordial explanatory variable (r_{k1}).

It can be noticed that, on one hand, the negative coefficient of allignament to collinearity hazard occurs at the explanatory variable, relatively weakest correlated with the resultative variable $(S/Y)_{t-1}$. It is to be noticed that the value of the coefficient of correlation between the first two explanatory variables $(S/Y)_t$ şi $(S/Y)_{t-1}$ mediated by resultative variable is very near of 1.

The matrix of Pearson coefficients of correlation between the explanatory variables (table no.2) shows that the respective indicator takes a maximum value of 0.8610, for $R((S/Y)_t,(S/Y)_{t-1})$ and a minimum one of 0.7744 for $R((S/Y)_t,(I/Y)_{t-1})$. Therefore, the coefficient of alignment to collinearity hazard values in conditions of non-differentiation of coefficients of mediated corelation by resultative variable and respectively Pearson coefficients of correlation between explanatory variables is 0.3674 (Table 3).

Table 2

Values of Pearson coeficients of correlation between explanatory variables in case of linear regression

$$(I/Y)_t = a + b*(S/Y)_t + c*(I/Y)_{t-1} + d*(S/Y)_{t-1}$$

Explanatory variable	$(S/Y)_t$	$(I/Y)_{t-1}$	$(S/Y)_{t-1}$
$(S/Y)_t$	1.0000	0.7744	0.8610
$(I/Y)_{t-1}$	0.7744	1.0000	0.8023
$(S/Y)_{t-1}$	0.8610	0.8023	1.0000

Table 3

The coefficiens of alignment to collinearity hazard values depending on the differention of coefficients of correlation between explanatory variables in case of linear regression

$$(I/Y)_t = a + b*(S/Y)_t + c*(I/Y)_{t-1} + d*(S/Y)_{t-1}$$

Explanatory variable	$(S/Y)_t$	$(I/Y)_{t-1}$	$(S/Y)_{t-1}$	Aritmetical mean
T_{3R}	0.3674	0.3674	0.3674	0.3674
T_{3Rrst}	0.8465	0.0997	0.0997	0.3486
T_{3Rrord}	0.7300	0.3603	-0.0305	0.3533
T_{3Rref}	0.6119	0.5975	-0.1625	0.3489
T_{3k}	0.6860	0.6265	-0.2205	0.3640

Due to the fact that $r_{3\min} > R_{\max}$, in conditions of non-differention of Pearson coefficients of correlation and of standard differention of coefficients of correlation between explanatory variables mediated by resultative variable, all coefficiens of alignment to collinearity hazard (T_{3Rrst}) are positive. But if the ordered differentiation of coefficient of correlation between explanatory variables mediated by resultative variable is taken into account a negative coefficient of alignment to collinearity hazard occurs. The respective negativity becomes more marked in the context of actual differentiation of coefficients of correlation mediated by resultative variable and respectively Pearson coefficients of correlation between explanatory variables.

The factorial influence computation reveals that the maximum value of the Pearson coefficients of correlation between explanatory variables (R_{max}) has a major contribution to aritmetical mean of the coefficiens of alignment to collinearity hazard. Therefore, the ratio between T_{3k} and T_{3R} aritmetical means is 99.07%. (Table 4).

Table 4

Factors influences in obtaining the coefficiens of alignment to collinearity hazard values in case of linear regression

$$(I/Y)_t = a + b*(S/Y)_t + c*(I/Y)_{t-1} + d*(S/Y)_{t-1}$$

Explanatory variable	$(S/Y)_t$	$(I/Y)_{t-1}$	$(S/Y)_{t-1}$	Aritmetical mean
T_{3R}	0.3674	0.3674	0.3674	0.3674
T_{3Rrst}	2.3041	0.2713	0.2713	0.9489
T_{3Rrord}	0.8624	3.6160	-0.3060	1.0135
T_{3Rref}	0.8381	1.6582	5.3309	0.9877
T_{3k}	1.1211	1.0485	1.3568	1.0431

The standard differentiation of coefficients of correlation between expalnatory variables mediated by resultative variable determine a decrease with 5.11% of the arithmetical mean of coefficients of allignment to collinearity hazard, while the ordered differentiation determine an increase of 1.35%. Related to the ordered differentiation, the actual differentiation of coefficients of correlation between expalnatory variables mediated by resultative variable generates a decrease by 1.23% in the above-mentioned aritmetical mean and is a consequence of the fact that the transition to r_{nmin} is practically made between the penultimate and last explanatory variable. The actual differentiation of Pearson coefficients of correlation between the explanatory variables determines an increase in arithmetical mean at the same time with an increased polarization of the coefficients of alignment to collinearity hazard.

Because $r_{3\min} > R_{\max}$, all the three bifactorial linear regressions possible to be performed with the above-mentioned explanatory variables have positive coefficients of allignment to collinearity hazard. The values of estimated parameters and also the values of coefficients of allignment to collinearity hazard and their aritmetical mean are presented below.

$(I/Y)_t$=5.457+ 0.392*$(S/Y)_t$ + 0.418*$(I/Y)_{t-1}$

(1.693) (1.978) (2.013)

R^2_2=0.746 R^2_{2adj}=0.704, F=17.650 AIC= 4.637 SCH=4.779 D-W=1.426

$T_{(S/Y)t}$= 0.5673, $T_{(I/Y)t}$=0.5596, T_{2med} = 0.5635, T_{2R}= 0.5636 $\frac{T_{2med}}{T_{2R}} = 0.9998$

$(I/Y)_t$=5.256+ 0.588*$(I/Y)_{t-1}$ + 0.196*$(S/Y)_{t-1}$
(1.441) (1.944) (0.784)
R^2_2=0.677 R^2_{2adj}=0.623, F=12.587 AIC= 4.878 SCH=5.019 D-W=1.538

$T_{(I/Y)t-1}$= 0.7871, $T_{(S/Y)t-1}$=0.2961, T_{2med} = 0.5416, T_{2R}= 0.5549, $\frac{T_{2med}}{T_{2R}} = 0.9760$

$(I/Y)_t$=8.972 + 0.614*$(S/Y)_t$ + 0.096*$(S/Y)_{t-1}$
(2.814) (2.205) (0.324)
R^2_2=0.666 R^2_{2adj}=0.611, F=11.989 AIC= 4.911 SCH=5.052 D-W=0.814

$T_{(S/Y)t}$= 0.8886, $T_{(I/Y)t}$=0.1450, T_{2med} = 0.5168, T_{2R}= 0.5373, $\frac{T_{2med}}{T_{2R}} = 0.9618$

It may be noticed that, like in the case of linear regression with three explanatory variables, the main contribution to the arithmetical mean of coefficients of alignment to collinearity hazard is brought by the maximum absolute value of Pearson coefficients of correlation between explanatory variables.

The previously made parameter estimations permit to determine the efficiency of addition of a new explanatory variable in the linear regression model. If the absolute value of the coefficient of determination is considered as the main criterion, the primordial explanatory variable is $(S/Y)_t$. From the point of view of the Durbin-Watson test, at a level of significance of 2.5%[4], in this case the errors appears to be autocorrelated.

Having in view the computed values of coefficient of determination and of the Durbin-Watson test in the simple linear regressions, the next explanatory variable which is recommended to be added to the econometric model is $(I/Y)_t$. This new explanatory variable addition to linear regression model can be considered an efficient one.

For both the explanatory variables, the coefficients of alignment to collinearity hazard are positive. In comparison with the simple linear regression mentioned-above, the linear regression with two explanatory variables ($(S/Y)_t$ and $(I/Y)_{t-1}$) shows improvement of all statistical tests computed values[5]. Also, it is important to mention that from the point of view of the Durbin-Watson test, at a level of significance of 2.5%, the errors are not autocorrelated.

The addition of a third explanatory variable ($(S/Y)_{t-1}$) in the linear regression model proves to be inefficient. This appreciation is mainly supported by the occurrence of a negative coefficient of alignment to collinearity hazard. But at the same time, all the computed values of statistical tests taken into account are worse in the linear

[4] At a level of significance of 2.5% and 15 observations, the critical tabelled values are:
a) for a linear regression with one explanatory variable dL= 0,95 and dU= 1,23;
b) for a linear regression with two explanatory variables dL= 0,83 and dU= 1,40;
c) for a linear regression with three explanatory variables dL= 0.71 and dU= 1.61.

[5] The coefficient of determination increases from 0,664 to 0,746, the adjusted coefficient of determination increases from 0,664 to 0,704, the Akaike informational criterion decreases from 4,786 to 4,637, the Schwartz criterion decreases from 4,880 to 4,779 and the computed value of Durbin-Watson test increase from 0,856 to 1,426.

regression with three explanatory variables in comparison with the linear regression with two explanatory variables[6]. At a level of significance of 2.5% the computed value of the Durbin-Watson test shows that the errors cannot be considered as autocorrelated or not autocorrelated.

6. Conclusions on collinearity impact on linear multiple regressions. Proposals for rules designed to increase the feasibility of OLS estimated parameters

The considerations previously made show that in the case of multiple linear regressions, parameters estimated by the OLS are influenced by collinearity. The respective phenomenun is synthetically revealed by the arithmetical mean of the coefficients of allignment to collinearity hazard. For this reason, in evaluating the quality and feasibility of estimated parameters is necessary to consider the coefficients of alignment to collinearity hazard values among statistical tests tests which validate econometric models. In order to avoid "statistical illusions", in a multiple linear regression all the coefficients of alignment to collinearity hazard have to be positive at least.

Therefore, the identification of modelling factors and the quantification of their influence on the coeficients of alignment to collinearity hazard individual values and the arithmetical mean may be a useful approach for assessing the estimation of econometric models.

In analyzing the coefficients of adjustment to collinearity hazard it is important to consider factors shaped into two groups, as follows: a) structural constraints; and b) features of differentiation of Pearson correlations between the explanatory variables, on the one hand, and between the respective variables and the resultative variable, on the other hand.

This way, three structural constraints can be identified in estimating linear regression parameters, namely:

1) **The maximum absolute value of the Pearson coefficients of correlation between explanatory variables (R_{max}).** This indicator determines the critical value of the coefficients of alignment to collinearity hazard. If R_{max} value is increased, the effect is a low arithmetical mean for the coefficient of alignment to collinearity hazard and their polarization. However, it is important to note that large values of R_{max} do not directly determine the occurrence of negative coefficients of alignment to collinearity hazard, but favour the respective phenomenon.

2) **Number of explanatory variables.** An increase in number of explanatory variables determines a reduction in individual values and arithmetical mean for the coefficients

[6] The coefficient of determination increases from 0,746 to 0,752, the adjusted coefficient of determination decreases from 0,704 to 0,684, the Akaike informational criterion increases from 4,637 to 4,749, the Schwartz criterion increases from 4,779 to 4,937 and the computed value of the Durbin-Watson test increase from 1,426 to 1,586.

of alignment to collinearity hazard, and enhanced constraints to be met for all mentioned above factors are positive.

3) **The minimum value of coefficients of correlation between expalnatory variables mediated by resultative variable ($r_{n\ min}$).** The above mentioned indicator has a direct influence on the fulfilment of the condition that all coefficients of alignment to collinearity hazard are positive. The weakest constraint that have to be met in order that all coefficient of alignment to collinearity hazard are positive is: $r_{n\min} > R_{\max}$. Also, it is to be noticed that the value decrease in r_{nmin} determines the diminution in the arithmetical mean of the coefficients of alignment to collinearity hazard

The distribution features of two kinds of coefficients of correlation between expalnatory variables (Pearson and mediated by resultative variable, respectively) have complex effects on the coefficients of alignment to collinearity hazard.

Hence, the differentiation of the absolute values of coefficients of correlation mediated by explanatory variable has as a consequence the establishing of an "order" of the explanatory variables in the regression model.

If the Pearson coefficients of correlation between explanatory variables are not differentiated we can establish the following interdependences between the distribution of coefficients r_{k1} and the sign of coefficients T_{nk}, as follows:

a) $r_{n\min} < R_{\max}$ there is at least one negative coefficient T_{nk}, regardless of the differentiation of coefficients r_{k1}

b) $R_{\max} < r_{n\min} < \frac{(n-1) * R_{\max}}{1+(n-2) * R_{\max}}$, the occurence of negative coefficients T_{nk} is dependent on the distribution of coefficients r_{k1}.

c) $r_{n\min} > \frac{(n-1) * R_{\max}}{1+(n-2) * R_{\max}}$, all coefficients T_{nk} are positive, regardless of the distribution of coefficients r_{k1}.

The differentiation of the Pearson coeficients of corelation between the explanatory variables essentialy determinates the “hazard occurence” in coefficients T_{nk}. In other words, the respective differentiation can result in the challenging of the „order” of the coefficients T_{nk} as regards their values established by coefficients r_{k1}. The trend towards "chaos" of the coefficients r_{k1} is stressed by the increase in the explanatory variable number. A global indicator for the trend towards “order” or to “chaos” of the coefficients of alignment to collinearity hazard may be the comparison between the value of coeficient of determination computed for the linear regression model (R_n^2) and product between the sum of the squares of the Pearson coefficients of correlation between resultative variable and the explanatory variables and the simple arithmetical mean of coefficients of alignment to collinearity hazard (SRT). **If $R_n^2 > SRT$, we may speak about a trend towards “order” of the coefficients of alignment to**

collinearity hazard. If $R_n^2 < SRT$, we may speak about a trend towards "chaos" of the coefficients of alignment to collinearity hazard[7].

Therefore, in order to improve the feasibility of estimated parameters of a multiple linear regression model using the OLS method we should consider the following rules:

1) Before the estimation of a multiple linear regression it is necessary to compute the values of the two kinds of coefficients of correlation between explanatory variables (Pearson and mediated by resultative variable).

2) Avoid to use a set of explanatory variables highly correlated with each other or a set of explanatory variables that present very differentiated correlations with the resultative variable.

3) **Test the fulfilment of constraints** $r_{n\min} > R_{n\max}$ **and** $r_{n\min} > \frac{(n-1)*R_{\max}}{1+(n-2)*R_{\max}}$, **respectively.** This way, we may emphasize the possibility to ensure positivity for all the coefficients of allignament to collinearity hazard. Subsequently, after the regression has been made, it can be set to the role played by the distribution of the Pearson coefficients of correlation between explanatory variables in shaping the indicator values mentioned above.

4) **Undertake a careful selection of explanatory variables**, in order to assure that in the regression equation are included a reasonable number of parameters estimated. This way, it can be avoided the occuring of negative coefficients of alignment to collinearity hazard, due to an excessive number of explanatory variables.

5). Use of regression in main components.

6) Use of regression with composite explanatory variables.

---

[7]Having in view that $R_n^2 = (\sum_{k=1}^{n} R^2(x_k, y)) * (T_{nk})_{meds} * (1 + sc\,\text{cov}(R^2(x_k; y); T_{nk}))$, and $SRT = (\sum_{k=1}^{n} R^2(x_k; y)) * (T_{nk})_{meds}$ it may be admitted that a positive covariance between the square of the Pearson coefficients of correlation between the resultant variable and explanatory variables, on one hand, and the coefficients of alignment to collinearity hazard, on other hand, means a trend towards "order" and a negative covariance means a trend towards "chaos", the affirmations made above can be demonstrated.

Annex 1

Interdependencies between the coefficients of correlation between explanatory variables mediated by resultative variable, coefficients of alignment to collinearity hazard, coefficients of partial correlation and "standard" Student test in case of a linear regression with two explanatory variables

The coefficients of correlation between explanatory variables mediated by resultative variable and the coefficients of alignment to collinearity hazard in case of a linear regression with two explanatory variables have a considerable impact on the computed value of the coefficient of partial correlation between the respective explanatory variables.

The coefficients of partial correlation for explanatory variable x_p (Rpart (x_p;y) and for explanatory variable x_s (Rpart (x_p;y) are computed using the following formulae:

$$Rpart(x_p;y) = \frac{R(x_p;y) - R(x_s;y) * R(x_p;x_s)}{\sqrt{(1-R^2(x_s;y)) * (1-R^2(x_s;x_p))}}$$

and

$$Rpart(x_s;y) = \frac{R(x_s;y) - R(x_p;y) * R(x_p;x_{s)}}{\sqrt{(1-R^2(x_p;y)) * (1-R^2(x_s;x_p))}},$$

equivalent with:

$$Rpart(x_p;y) = R(x_p;y) * T_{2p} \frac{\sqrt{(1-R^2(x_p;x_s))}}{\sqrt{(1-R^2(x_s;y) * r^2)}}$$

$$Rpart(x_s;y) = R(x_s;y) * T_{2s} \frac{\sqrt{(1-R^2(x_p;x_s))}}{\sqrt{(1-R^2(x_s;y))}}$$

As a consequence, the ratio between the coefficients of the absolute values of partial correlation related to the two explanatory variables may be written:

$$\frac{/\, Rpart(x_{s:};y)\, /}{/\, Rpart(x_p;y)\, /} = r * \frac{T_{2s}}{T_{2p}} * \sqrt{\frac{(1-r^2 * R^2(x_p;y))}{(1-R^2(x_p;y))}}$$

In other words, the coefficient of correlation between explanatory variables mediated by the resultative variable explains the differentiation of the absolute values of coefficients of partial correlation and the unexpected sign in case of secondary variable, when the critical level of collinearity is surpassed.

The above-mentioned formulae for coefficients of partial correlation computation permit also to reveal the relationship between the respective indicators and the Student test computed values in case of a linear regression with two explanatory variables. Having in view the computation formulae for "standard" Student test (F. M.

Pavelescu, 2009)[8] in case of explanatory variable x_p (t_{bxp}) and explanatory variable x_p (t_{bxp}), respectively, we may write:

$$t_{bxp} = (m-3) * Rpart(x_p; y) * \sqrt{\frac{(1-R^2(x_p; y) * (1 + \frac{(r - R(x_p; x_s))^2}{1-R^2(x_p; x_s)})}{1-R^2(x_p; y) * r^2}}$$

and

$$t_{bxs} = (m-3) * Rpart(x_s; y) * \sqrt{\frac{(1-R^2(x_p; y) * (1 + \frac{(r - R(x_p; x_s))^2}{1-R^2(x_p; x_s)})}{1-R^2(x_p; y)}} ,$$

where: m= number of the observations.

References

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

Dougherty C. (2002), "*Introduction to Econometrics"*, Second Edition, Oxford University Press.

Greene W. (1993), "*Econometric Analysis"*, New York: McMillan Publishing Company.

Jansen W.J., G.G.Schultz, (1996), "Theory-based measurement of the saving-investment correlation with an application in Norway", *Economic Inquiry*, vol. XXXIV.

Isaic-Maniu A., C. Mitruţ, V. Voineagu, (1996), "*Statistica pentru managementul afacerilor",* Bucureşti: Editura Economică.

Jula D. (2003), "*Introducere în econometrie",* Bucureşti: Professional Consulting.

---

[8] In F.M. Pavelescu (2009), the formulae usually used to compute the values of Student test are considered as "standard" ones, because in the above-mentioned paper, a "correction" was proposed for the respective statistical test in order to identify more easily the sign of the coefficient of alignment to collinearity hazard. The computation formula for the "corrected"Student test (CST $_{bxnk}$) is:

$$CST_{bnk} = (m-n-1)^{0.5} * / R(x_k; y) / * \frac{T_{nk}}{\sqrt{(1-\sum R^2(x_k; y) * T_{nk})}} * \sqrt{\frac{(R_{jk})_n}{(R_{jk})_{n-1}}}$$

As a consequence, relationship between the "standard" and the "corrected" Student test computation formula is $CST_{bnk} = t_{bnk} * \frac{/ R(x_k; y) /}{R(x_k; y)}$. In these conditions, it was suggested that the validation of a linear regression model takes place only if the "corrected" Student test computed values are positive for all the explanatory variables. Negative values of "corrected" Student test computed values reveal the surpassing of the critical level of collinearity and the occurrence of "statistical illusions" during the estimation process.

Feldstein M., Ch. (1980), "Horioka-Domestic saving and international capital flows", *Economic Journal*, 90.

Pavelescu F. M., (1986), "Some considerations regarding the Cobb-Douglas production function estimated parameters. A new approach", *Revue Roumaine des Sciences Sociales,* 30(1-2).

Pavelescu F. M. (2004.), "Features of ordinary least square (OLS) method. Implications for estimation methodology", *Romanian Journal of Economic Forecasting,* 5(2): 85-101.

Pavelescu F. M. (2005), "Impact of collinearity on the estimated parameters and classical statistical tests values of multi-factorial linear regressions in conditions of O.L.S.", *Romanian Journal of Economic Forecasting,* 6(2): 50-71.

Pavelescu F. M. (2009), "A review of statistical properties of Student test", *Romanian Journal of Economic Forecasting,* 10(1): 63-75.

Pecican E., (1996), "*Macroeconometrie. Politici economice guvernamentale şi econometrie*," Bucureşti: Editura Economică.

Pelgrin F., S.Schich, (2004). "*National Saving-Investment Dynamics and International Capital Mobility"*, Bank of Canada Working Paper 2004-14, April 2004.

Rocha F. (2006), "Are error-correction models the best alternative to assess capital mobility in developing countries?", *Economia Aplicada*, 10(2).

Tudorel A., S. Stancu, (1994), "*Statistică. Teorie şi aplicaţi"i*, Bucureşti: Editura All.

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