An Extensive Study on the Disturbances Generated by Collinearity in a Linear Regression Model with Three Explanatory Variables

Author∗:
Florin-Marius PAVELESCU

Abstract. *In econometric models, linear regressions with three explanatory variables are widely used. As examples can be cited: Cobb-Douglas production function with three inputs (capital, labour and disembodied technical change), Kmenta function used for approximation of CES production function parameters, error-correction models, etc.*

In case of multiple linear regressions, estimated parameters values and some statistical tests are influenced by collinearity between explanatory variables. In fact, collinearity acts as a noise which distorts the signal (proper parameter values). This influence is emphasized by the coefficients of alignment to collinearity hazard values. The respective coefficients have some similarities with the signal to noise ratio. Consequently, it may be used when the type of collinearity is determined. For these reasons, the main purpose of this paper is to identify all the modeling factors and quantify their impact on the above-mentioned indicator values in the context of linear regression with three explanatory variables.

Key words: *types of collinearity, coefficient of mediated correlation, rank of explanatory variable, order of attractor of collinearity, mediated collinearity, anticollinearity*.

JEL classification: C13, C20, C51, C52

∗ *Florin Marius Pavelescu, Ph.D., senior researcher I, Institute of National Economy, Romanian Academy, e-mail: pavelescu.florin@yahoo.com.*

1. Characteristic features of the estimation of a linear regression with three explanatory variables by the of OLS method

If the method of ordinary least squares is used, in case of a linear regression with three explanatory variables i.e.:

$$y = a_3 + \sum_{k=1}^{3} b_{3k} \cdot x_k \tag{1}$$

the estimated values of parameters can be written (F.M. Pavelescu, 1986):

$$a_3 = y_{med} - \sum_{k=1}^{3} x_{kmed} \cdot b_{3k} \tag{2}$$

$$b_{3k} = b_{1k} * T_{3k} \tag{3}$$

where: y_{med} = arithmetical mean of resultative variable y

x_{kmed} =arithmetical mean of explanatory variable x_k

$$b_{1k} = \frac{C\,\mathrm{ov}(x_k; y)}{D^2(x_k)} \tag{4}$$

and represents the estimated value of parameter b in case of simple linear regression

$$y = a_1 + b_{1k} \cdot x_k \tag{5}$$

Cov (x_k;y) = covariance between explanatory variable x_k and resultative variable y.

D^2 (x_k) = variance of explanatory variable x_k

T_{3k}= coefficients of alignment to collinearity hazard corresponding to explanatory variable x_k.

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

$$T_{3k} = \frac{(R_{j1}, r_{jk}, R_{j3})}{(R_{jk})_3}, \tag{6}$$

where: R_{jk} = Pearson coefficient of correlation between the explanatory variables x_j and x_k.

$$r_{jk} = \frac{R(x_j; y)}{R(x_k; y)} \tag{7}$$

where: $R(x_j;y)$, $R(x_k:y)$ = Pearson correlation between the explanatory variable x_j and x_k, respectively and the resultative variable.

It is important to notice that in a multiple linear regression, the coefficients of alignment to collinearity hazard directly influence not only the values of estimated parameters but also the computed values of some indicators or statistical tests, such as the coefficient of determination or the Student test, because it can be demonstrated that:

a) Coefficient of determination (R^2_n) may be computed (F.M. Pavelescu, 2005) by the formula:

$$R_n^2 = \sum_{k=1}^{n} R^2(x_k; y) \cdot T_{nk} \tag{8}$$

equivalent to:

$$R_n^2 = \sum_{k=1}^{n} R^2(x_k; y) \cdot (T_{nk})_{med} \cdot (1 + Cv(R^2(x_k; y)) \cdot Cv(T_{nk}) \cdot R(R^2(x_k; y); T_{nk})) \tag{9}$$

where: Cv ($R^2(x_k;y)$) = coefficient of variation of squared Pearson coefficients of correlation between explanatory variables x_k and resultative variable y.

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

Cv (T_{nk}) = coefficient of variation of coefficients of alignment to collinearity hazard.

$R(R^2 (x_k;y); T_{nk})$ = Pearson coefficients of correlation between $R^2 (x_k;y)$ and T_{nk}

The values of the Student test (t_{bnk}) may be computed (F.M. Pavelescu, 2005) by the formula:

$$t_{bnk} = \left[m - (n+1)\right]^{1/2} \cdot \frac{R(x_k; y)}{(1 - R_n^2)^{1/2}} \cdot T_{nk} \cdot \left(\frac{(R_{jl})_n}{(R_{jl})_{n-1, l \neq k}}\right)^{1/2} \tag{10}$$

The computation formula presented above highlight that the coefficients of alignment to collinearity both as individual values and as arithmetical mean may sensibly influence the estimated results and their validation in case of a multiple linear regression.

2. The values of coefficients of alignment to collinearity hazard and the determination of the type of collinearity

The formula (3) shows that in a multiple linear regression the estimated values of parameters b_{3k} are the product between the estimated value of the respective parameter in case of simple linear regression (b_{1k}) and the coefficient of alignment to collinearity hazard (T_{3k}). Under these conditions, b_{1k} may be considered as the proper value of the estimated parameter and b_{3k} as the derived value of the estimated parameter, because it is influenced by collinearity between the explanatory variables (F.M. Pavelescu, 2005). But if we appeal to the concepts of “signal” and “noise”, used in (Belsey, 1991), we may define b_{1k} as the “initial signal” and b_{3k} as the “signal distorted by noise”. Consequently, T_{3k} has some similarities to the “signal to noise ratio”[1].

Because the collinearity problem occurs in all estimation methods, and consequently also in case of ordinary least squares, different procedures were proposed to deal with and classify the respective phenomenon. For example, L.R. Klein (1962) stated that “(multi) collinearity is not necessarily a problem unless it is high – relative to the overall degree of multiple correlation”. Under these conditions, the collinearity was classified as “acceptable” and “harmful”, respectively.

Acceptable collinearity means that the departures from ortogonality of the explanatory variables are small and consequently the results of the regressions are feasible. The **harmful collinearity** has been initially defined symptomatically, i.e. as the cause of wrong signs or other symptoms of nonsense regressions (D. Ferrar, R. Glauber, 1967). The rule of thumb to be respected in order to avoid the occurrence of “harmful collinearity” is that the values of the coefficient of Pearson correlation between the explanatory variables has to be smaller than 0.8.

---

[1] *It is important to notice that in techinical sciences and engineering “the signal to noise ratio” is an indicator used to measure how much of the signal was corrupted by the environment features. In some statistical papers the signal to noise ratio is seen as the inverse of the coefficient of variation.*

Belsey (1976) distinguished between **"harmful collinearity" and "degrading collinearity".** The distinction is made considering that strong correlations between the explanatory variables are a potential problem for the quality of estimation results. In other words, in conditions of high correlated explanatory variables, collinearity is potentially harmful, but in fact it has to be tested if the estimated results **are really harmed or just degraded.** It is admitted that the problem of collinearity is generated by the ill-conditioned data and two diagnosis tests used for determining the degrading and harmful collinearity were proposed (Belsey, 1991).

Having in mind the types of collinearity presented above and the values of the coefficients of alignment to collinearity hazard we propose a rule of thumb for detecting the collinearity features, as follows:

- ✓ **weak (acceptable) collinearity** if in a multiple linear regression all the coefficients of alignment to collinearity hazard are bigger than 0.5.
- ✓ **degrading collinearity** if all the coefficients of alignment to collinearity hazard are positive and at least one of them is smaller than 0.5.
- ✓ **harmful collinearity** if at least one of the coefficients of alignment to collinearity hazard is negative.

The reasons for the proposed rule of thumb are:

a) The limit between the weak (acceptable) and the degrading collinearity was chosen having in view a value of the coefficient of alignment to collinearity hazard, which can be easily kept in mind and was inspired by Belsey (1976) according to which a rule of thumb estimate establish that "estimated values are degraded when two or more variances have at least half of their magnitude associated with a singular value". In our case, an estimation is considered degraded by collinearity when at least one of the estimated parameters maintain the right sign and have the derived value diminished by more than half its proper value. Consequently, a **collinearity is considered weak** (acceptable) if the all the derived estimated values have the right signs and are bigger than the half of the proper estimated parameter values.

b) **Harmful collinearity** is considered to occur if at least one of the derived estimated parameters values has **a wrong sign** (contrary to the proper estimated parameter)[1]. In other words, the harmful collinearity is associated with the negative values of the coefficient of alignment to collinearity hazard.

---

[1] *In some papers, for example, Conrad (2006) the situation when the sign of an estimated parameter of a multiple linear regression is contrary to that of proper estimated value is defined as "unexpected sign of the estimated parameter".*

3. Intensity of correlation with the resultative variable as a ranking factor for explanatory variables in a linear regression with three explanatory variables

The formula (6) shows that in a linear regression equation with three explanatory variables, the size of coefficients of alignment to collinearity hazard is determined by the values of three Pearson coefficients of correlation between explanatory variables, on the one hand, and the size of the six ratios r_{jk}. As a rule, Pearson coefficients of correlation between explanatory variables and r_{jk} ratios are positive, or their positivity can be ensured through algebraic transformations.

Given the assumptions mentioned above, in relation to the absolute value of Pearson coefficients of correlation between explanatory variables and resultative variable, **the rank of explanatory variables in the respective linear regression may be determined.**

Thus, ratios r_{jk} can be computed for two successive explanatory variables (r_{21} and r_{32}), and also for the explanatory variables which are the least and most strongly correlated with resultative variable (r_{3min}), that is:

$$r_{21} = \frac{R(x_2;y)}{R(x_1;y)}, \tag{11}$$

$$r_{32} = \frac{R(x_3;y)}{R(x_2;y)} \tag{12}$$

$$r_{31} = r_{3\min} = \frac{R(x_3;y)}{R(x_1;y)} \tag{13}$$

From now on the ratios r_{jk} are defined as “coefficients of mediated, by resultative variable, correlation between explanatory variable x_{k+i} and explanatory variable x_k [1] and will be written “coefficients of m.r.v. correlation between explanatory variable x_{k+i} and explanatory variable x_k”. The explanatory variable which absolutely is most correlated with the resultative variable may be defined as primordial explanatory variable. The other explanatory variables are of rank 2 or 3 depending on the intensity of correlation with the resultative variable.

---

[1] *Consequently, the absolute values of coefficients of m.r.v. correlation between the explanatory variables x_{k+i} and x_k.range between 0 and 1.*

If we consider that x_1 is the primordial explanatory variable, x_2 is the explanatory variable of rank 2 and x_3 is the explanatory variable of rank 3, the computation formulae for the three coefficients of alignment to collinearity hazard are:

$$T_{31} = \frac{\begin{matrix} 1 & R(x_1;x_2) & R(x_1;x_3) \\ r_{21} & 1 & R(x_2;x_3) \\ r_{3\min} & R(x_2;x_3) & 1 \end{matrix}}{(R_{jk})_3} \tag{14}$$

$$T_{32} = \frac{\begin{matrix} 1 & \frac{1}{r_{21}} & R(x_1;x_3) \\ R(x_1;x_2) & 1 & R(x_2;x_3) \\ R(x_1;x_3) & r_{32} & 1 \end{matrix}}{(R_{jk})_3} \tag{15}$$

$$T_{33} = \frac{\begin{matrix} 1 & R(x_1;x_2) & \frac{1}{r_{3\min}} \\ R(x_1;x_2) & 1 & \frac{1}{r_{32}} \\ R(x_1;x_3) & R(x_2;x_3) & 1 \end{matrix}}{(R_{jk})_3} \tag{16}$$

The impact of coefficients of m.r.v correlation between explanatory variables on the values of the coefficients of alignment to collinearity hazard can be highlighted by taking into consideration particular cases of coefficients alignment to collinearity hazard in conditions of non-differentiation of Pearson coefficients of correlation between explanatory variables. In other words, the hypothesis is that all Pearson coefficients of correlation between the explanatory variables are equal to R_{3max}. It has to be noticed that R_{3max} is the Pearson coefficient of correlation between explanatory variables with the highest absolute value.

In these circumstances, we can compute the following indicators, namely:

a) arithmetical mean of coefficients of alignment to collinearity hazard in conditions of non-differentiation of coefficients of Pearson and m.r.v. correlation between explanatory variables (T_{3R}). In this situation, all $r_{k+1,k}$=1, and the indicator T_{3R} may be computed by the formula:

$$T_{3R} = \frac{1}{1 + 2 \cdot R_{3\max}} \tag{17}$$

The indicator T_{3R} is a very important one, because it detects the first premises for the type of collinearity. Depending on its values we can speak about a “potential weak collinearity”, if R_{3max} <0.5, or about a “potential degrading collinearity” if R_{3max} >0.5.

b) **arithmetical mean of coefficients of alignment to collinearity hazard in conditions of Pearson coefficients of correlation non-differentiation and of coefficients of m.r.v. correlation particular differentiations ($(T_{3Rrpdk})_{med}$).**

The respective indicator is computed in order to highlight the impact of the ratios between the absolute minimum and maximum values of coefficients of m.r.v correlation of explanatory variables on the values of coefficients of alignment to collinearity hazard. In order to achieve this objective, it is necessary to establish the differentiation of particular importance for coefficients r_{3k} and their arithmetical mean.

In conditions of non-differentiation of Pearson coefficients of correlation between explanatory variables, the arithmetical mean of coefficients of alignment to collinearity hazard ($(T_{3Rrefk})_{med}$) is:

$$(T_{3Rrefk})_{med} = \frac{1 + R_{3\max} - (\frac{1}{3}) \cdot R_{3\max}(r_{21} + r_{3\min} + \frac{1}{r_{21}} + r_{32} + \frac{1}{r_{32}} + \frac{1}{r_{3\min}})}{1 + R_{3\max} - 2R_{3\max}^2} \tag{18}$$

If r_{3min} and R_{3max} are given, the maximum value of expression (18) is obtained if: $r_{21} = r_{32} = \sqrt{r_{3\min}}$. We call the respective situation as an "ordered differentiation of coefficients of m. r. v. correlation between explanatory variables” $(T_{3Rrordk})_{med}$.

$$(T_{3Rrordk})_{med} = \frac{1 + 2 \cdot R_{3\max} - (\frac{1}{3}) \cdot (\frac{1 - r_{3\min} \cdot \sqrt{r_{3\min}}}{1 - \sqrt{r_{3\min}}})^2 \cdot \frac{R_{3\max}}{r_{3\min}}}{1 + R_{3\max} - 2R_{3\max}^2} \tag{19}$$

The minimum value of the arithmetical mean of the coefficients of alignment to collinearity hazard in the context of non-differentiation of Pearson coefficients correlation between explanatory variables appears in two cases, namely:

● $r_{21} = r_{3\min}$ and $r_{32} = 1$. Thus, $(T_{3Rrefk})_{med} = \dfrac{1+\frac{1}{3}\cdot R_{3\max} - \frac{2}{3}\cdot R_{3\max}\cdot\frac{(1+r_{3\min}^2)}{r_{3\min}}}{1+R_{3\max}-2R_{3\max}^2}$ (20)

We define this situation as "standard differentiation of coefficients of m.r.v. correlation between explicative variables" (T_{3Rrstk}). In fact, this is the fastest differentiation of the coefficients of m.r.v. correlation between explanatory variables.

● $r_{21} = 1$ and $r_{32} = r_{3\min}$. This the case of "the most postponed differentiation of the coefficients of m.r.v. correlation between explanatory variables". In this situation, the arithmetical mean of coefficient of alignment to collinearity hazard is equal to that is of the standard differentiation of coefficients of m.r.v. correlation between explanatory variables.

● arithmetical mean of coefficients of alignment to collinearity hazard in conditions of non-differentiation of Pearson coefficients of correlation between explanatory variables, and of real differentiation of coefficients of m.r.v. correlation between explanatory variables (($T_{3Rrefk})_{med}$.

$$(T_{3Rrefk})_{med} = \frac{1+2\cdot R_{3\max} - 3\cdot R_{3\max}\cdot\frac{r_{1medaritm}}{r_{1medharm}}}{1+R_{3\max}-2\cdot R_{3\max}^2}, \quad (21)$$

where: $r_{medaritm}$, $r_{medharm}$= arithmetical and harmonic means of the coefficients of m.r.v. correlation between the explanatory variables and primordial explanatory variable.

Analogously, we can define the indicators mentioned above for each individual coefficients of alignment to collinearity hazard.

It is very important to notice that differentiation of the m.r.v. coefficients of correlation between explanatory variables is the first factor that make possible the occurrence of harmful collinearity and harden the constraints imposed for the existence of degrading and weak collinearity.

The standard differentiation of coefficients of m. r. v. correlation between explanatory variables imposes the **softest constraints i**n order to avoid the occurrence of harmful collinearity and to achieve the weak or degrading collinearity.

The conditions that define the type of collinearity in a standard differentiation of m.r.v. coefficients of correlation between explanatory variables are:

For weak collinearity: $r_{3\min} > \dfrac{2 \cdot R_{3\max}}{1 - R_{3\max} + 2 \cdot R_{3\max}^2}$ (22)

For degrading collinearity: $R_{3\max} \le r_{3\min} \le \dfrac{2 \cdot R_{3\max}}{1 - R_{3\max} + 2 \cdot R_{3\max}^2}$ (23)

For harmful collinearity: $r_{3\min} < R_{3\max}$ (24)

The hardest constraints that have to be fulfilled in order to obtain the degrading or weak collinearity appear in case of the most postponed differentiation of the coefficients of correlation between explanatory variables. In other words, the conditions defining the type of collinearity in the above-mentioned differentiations of coefficients of m.r.v. correlation between explanatory variables are:

For weak collinearity: $r_{3\min} > \dfrac{4 \cdot R_{3\max}}{1 + R_{3\max} + 2 \cdot R_{3\max}^2}$ (25)

For degrading collinearity: $\dfrac{2 \cdot R_{3\max}}{1 + R_{3\max}} \le r_{3\min} \le \dfrac{4 \cdot R_{3\max}}{1 + R_{3\max} + 2 \cdot R_{3\max}^2}$ (26)

For harmful collinearity: $r_{3\min} < \dfrac{2 \cdot R_{3\max}}{1 + R_{3\max}}$ (27)

4. Explanatory variables as attractors of collinearity

In a linear regression with three explanatory variables, it should be kept in mind that the determinant $(R_{jk})_3$ value is influenced by the size of three Pearson coefficients of correlation between explanatory variables. For this reason, explanatory variables should also be viewed as "attractors of collinearity" and is very important to detect their order. The three Pearson coefficients of correlation between explanatory variables can be ordered according to their absolute values and the determinant $(R_{jk})_3$ can be expressed as:

$$(R_{jk})_3 = 2 \cdot {R_{3\max}}^3 \cdot K_1^2 \cdot K_2 - R_{3\max}^2 \cdot (1 + K_1^2 + K_1^2 \cdot K_2^2) + 1 \quad (28)$$

where:

R_{3max}= Pearson coefficient of correlation between explanatory variables with the highest absolute value.

K_1 = ratio between the Pearson coefficient of correlation between explanatory variables ranked two in terms of absolute values and R_{3max}

K_2 = ratio between the Pearson coefficient correlation ranked three and two in terms of absolute values.

As we noted before, usually all the coefficients of Pearson correlation between explanatory variables are positive or this condition may be obtained through algebraic transformations. Consequently, the values of ratios K_1 and K_2 are positive and smaller than one.

An exception may occur if there are only two positive Pearson coefficients of correlation between explanatory variables, the other coefficient being negative. In such situations, the ratio K_1 and K_2 are computed having in view the real values of coefficients of Pearson correlation between explanatory variables. Therefore, K_1 is positive and smaller than unit while K_2 is negative.

The ratios K_1 and K_2 play an important role in defining explanatory variables as "attractors of collinearity".

Thus, **the main attractor of collinearity** is the explanatory variable that is present in both Pearson coefficients of correlation, which enable the computation of the ratio K_1. In fact, the ratio K_1 represents the coefficient of correlation mediated by the main attractor of collinearity between the attractor of collinearity of order three and of order two.

Attractor of collinearity of order 2 is the explanatory variable that is present in the Pearson coefficients of correlation between explanatory variables having the highest and lowest absolute value.

Attractor of collinearity of order 3 is the explanatory variable that makes possible to compute ratio K_2. In other words, the ratio K_2 represents the coefficient of correlation mediated by attractor of order 3 between the attractor of order of order 2 and the main attractor of collinearity.

On the other hand, if we denote by $K_2 = L \cdot R_{3max}$ $(L > 0)$, we obtain:

$$(R_{jk})_3 = (1 - R_{3\max}^2)(1 - R_{3\max}^2 \cdot K_1^2) - R_{3\max}^4 \cdot K_1^2 \cdot (1-L)^2 \tag{29}$$

The maximum value of determinant $(R_{jk})_3$ is obtained if R_{3max}=0, i.e. if all the three explanatory variables are strictly independent, and implicitly the collinearity does not occur at all.

If K_1=K_2 =0 it results: $(R_{jk})_3$ = 1- $R^2_{3\max}$ (30)

and this is the case when the third explanatory variable is independent of the previous two.

If K_1=1 it results: $(R_{jk})_3$ = $(1-R^2_{3\max})^2 - R^4_{3\max}\cdot(1-L)^2$ (31).

If L = 1, a local maximum of the determinant $(Rjk)_3$ is obtained. Determinant $(Rjk)_3$ can also be expressed in three ways, i.e., in relation to the three attractors of collinearity:

a) the main attractor of collinearity (x_{c1})

$(R_{jk})_3$ = $1-R^2_{3\max}\cdot K_1^2 - R_{3\max}\cdot A - R_{3\max}\cdot K_1\cdot K_2\cdot C$ (32)

b) the attractor of collinearity of order 2 (x_{c2})

$(R_{jk})_3$ = $1-R^2_{3\max}\cdot K_1^2\cdot K_2^2 - R_{3\max}\cdot A - R_{3\max}\cdot K_1\cdot B$ (33)

c) the attractor of collinearity of order 3 (x_{c3})

$(R_{jk})_3$ = $1-R^2_{3\max} - R_{3\max}\cdot K_1\cdot B - R_{3\max}\cdot K_1\cdot K_2\cdot C$, (34)

where:

A= $R_{3\max} - R^2_{3\max}\cdot K_1^2\cdot K_2$ (35)

B= $R_{3\max}\cdot K_1 - R^2_{3\max}\cdot K_1\cdot K_2$ (36)

C= $R_{3\max}\cdot K_1\cdot K_2 - R^2_{3\max}\cdot K_1$ (37)

We notice that if L>1, we obtain A> B>C>0. If L<1, we obtain A>B>0 and C<0.

If there is no differentiation of the coefficients of m. r. v. correlation between explanatory variables, i.e. r_{12} = r_{23} = 1, the computation formulae for the three coefficients of alignment to collinearity hazard corresponding to attractors of collinearity are:

$$T_{xc1} = \frac{(1-R_{3\max}\cdot K_1\cdot K_2)\cdot(1+R_{3\max}\cdot K_1\cdot K_2 - R_{3\max} - R_{3\max}\cdot K_1)}{(R_{jk})_3} \quad (38)$$

$$T_{xc2} = \frac{(1-R_{3\max}\cdot K_1)\cdot(1+R_{3\max}\cdot K_1 - R_{3\max} - R_{3\max}\cdot K_1\cdot K_2)}{(R_{jk})_3} \quad (39)$$

$$T_{xc3} = \frac{(1 - R_{3\max}) \cdot (1 + R_{3\max} - R_{3\max} \cdot K_1 - R_{3\max} \cdot K_1 \cdot K_2)}{(R_{jk})_3} \tag{40}$$

In the context of non-differentiation of coefficients of m. r. v. correlation between the explanatory variables, negative values of the coefficients of alignment to collinearity hazard appear only with attractor of collinearity of order 1, if L<1.

Therefore, parameter *L* allows a new facet of collinearity, i.e. **the** collinearity between two explanatory variables related to a third explanatory variable. In fact, this concept is an extension of the situation that arises in computing coefficients of alignment to collinearity hazard in case of a linear regression with two explanatory variable, where we can talk about the collinearity between two explanatory variables related to the resultative variable[1]. Shortly, in this paper we name this situation as "essential mediated collinearity". **By means of the values of L we can identify three types of essential mediated collinearity.**

a) **non-harmful essential mediated collinearity** if $L \geq 1$. In this case, as we mentioned before, the negative coefficients of alignment to collinearity hazard are not determined by the distribution of the coefficients of Pearson correlation between explanatory variables. If the harmful collinearity occurs, this is only a result of the values of m.r.v. coefficients of correlation between the explanatory variables.

b) **potential harmful essential mediated collinearity** if: $0 \leq L < 1$. In this case, it is possible to appear a negative value of the coefficient of alignment to collinearity hazard at the main attractor of collinearity, due to the absolute values and distribution of the coefficients of correlation between explanatory variables. Of course, harmful collinearity may also occur with the values of m.r.v. coefficients of correlation between explanatory variables, and consequently in this situation the harmfulness of the collinearity is potentially higher.

---

[1] *We should note that the value of coefficient of m. r. v. correlation between explanatory variables has to be related to the value of Pearson coefficient of correlation between the respective explanatory variables. Taking into account the conditions determining negative values for the coefficients of alignment to collinearity hazard in case of a linear regression with two explanatory variables F.M. Pavelescu (1986) we may conclude that if the absolute value of coefficient of m.r.v. correlation is smaller than that of the Pearson coefficient of correlation between explanatory variables, a critical threshold is exceeded and collinearity becomes harmful.*

c) **essential mediated anticollinearity** if: $L < 0$. In this case, the negative values of L[1] enable a sensible increase of the values in coefficients of alignment to collinearity hazard, if the values of m.r.v. coefficients of correlation between explanatory variables create conditions for a weak or degrading collinearity. If the premises created by m.r.v. coefficients of correlation between explanatory variables are for harmful collinearity, in case of a linear regression with three explanatory variables, mediated essential anticollinearity can avoid or sensibly diminish the respective form of collinearity.

The above considerations show that the values of coefficients of alignment to collinearity are modeled by a series of factors with contradictory influences. For these reasons, building a model of factorial analysis for the values of coefficients of alignment to collinearity hazard is very useful both for individual cases and for the arithmetical mean of those indicators. Therefore, we are able to determine the initial type of collinearity, generated by R_{3max}, and afterwards to reveal the factor contributions in obtaining the arithmetical mean and individual values of coefficients of alignment to collinearity hazard and consequently the type of collinearity in the regression equation taken into account.

5. Quantification of factorial contributions to obtain the arithmetical mean of coefficients of alignment to collinearity hazard

Having in view the methodology of factorial analysis, we can identify three main categories of influence on the arithmetical mean of coefficients of alignment to collinearity hazard:

● *the influence of the Pearson coefficient of correlation between the explanatory variables with he highest absolute value (T_{3R});*

---

[1] *We should note that mediated anticollinearity usualy reffers to a situation when two of the Pearson coefficients of correlation between explanatory variables have different signs and the third has the absolute value close to zero. In fact, the concept of mediated anticollinearity is an extention of the concept of "anticollinearity between to explanatory variables related to a resultative variable. If in a linear regression with two explanatory variables there is anticollinearity, both coefficients of alignment to collinearity hazard are positive and bigger than unit (F.M. Pavelescu, 2009).*

$$T_{3R} = \frac{1}{1 + 2 \cdot R_{3\max}} \tag{41}$$

As we stated before, the value of this indicator provides the initial premises for the type of collinearity.

● *the influence of the values and distribution of coefficients of m.r.v. correlation between explanatory variables differentiation ((Δ r)$_{med}$)*

$$(\Delta\, r)_{med} = \frac{1 + 2 \cdot R_{3\max} - 3 \cdot R_{3\max} \cdot \frac{r_{1medaritm}}{r_{1medharm}}}{(1 - R_{3\max}) \cdot (1 + 2 \cdot R_{3\max})} - \frac{1}{1 + 2 \cdot R_{3\max}} \tag{42}$$

We notice that $(\Delta\, r)_{med} < 0$.

● *influence of of the differentiation Pearson coefficients of correlation related to $R_{max}((\Delta\,(k_1\text{-}k_2)_{med})$*

$$(\Delta k_1 - k_2)_{med} = T_{3kmed} - \frac{1 + 2 \cdot R_{3\max} - 3 \cdot R^2{}_{3\max} \cdot \frac{r_{1medaritm}}{r_{1medharm}}}{1 + R_{3\max} - 2 \cdot R^2_{3\max}} \tag{43}$$

We notice that $(\Delta\,(k_1\text{-}k_2)_{med} > 0$.

Within the last two major categories of influence, the factorial analysis can be deepened. Thus, **in the case of the differentiation of coefficients of m.r.v. correlation in the context of non-differentiation of Pearson coefficients of correlation between explanatory variables (Δ r) the following influences[1] can be identified:**

[1] *It is important to mention that the respective influences may be computed if all the coefficients of m.r.v. correlations between explanatory variables are positive. If this condition is not fullfilled, and this may be the case when there are anticollinearity additional steps and changes in the computation has to be made.*

Firstly, we have to take into account that in a linear regression with three explanatory variables anticollinearity occours in three situations, namely: a) the coeficients of m.r.v. correlation are positive and one of the Pearson coefficients of correlation between explanatory variables is negative; b) there is one negative coefficient of m.r.v. correlation, while all Pearson coefficients of correlation between explanatory variables are positive; c) there are two negative coefficients of m.r.v. correlation and one negative Pearson coefficient of correlation between explanatory variables.

I) The influence of standard differentiation of coefficients of m.r.v. correlation between explanatory variables $(\Delta\, r_{st})_{med}$

$$(\Delta r_{st})_{med} = \frac{1+R_{3\max} - (\frac{2}{3})\cdot(1+r_{3\min}^2)\cdot\frac{R_{3\max}}{r_{3\min}}}{(1-R_{3\max})\cdot(1+2\cdot R_{3\max})} - \frac{1}{1+2\cdot R_{3\max}} \tag{44}$$

We notice that $(\Delta\, r_{st})_{med}<0$

II) *The influence of the ordered differentiation of coefficients of m.r.v. correlation between explanatory variables* $((\Delta\, r_{ord})_{med}.)$

$$(\Delta r_{ord})_{med} = \frac{1+2\cdot R_{3\max} - (\frac{1}{3})\cdot(\frac{1-r_{3\min}\cdot\sqrt{r_{3\min}}}{1-\sqrt{r_{3\min}}})^2\cdot\frac{R_{3\max}}{r_{3\min}}}{1+R_{3\max}+2\cdot R_{3\max}^2} - \frac{1+R_{\max} - (\frac{2}{3})\cdot(1+r_{3\min}^2)\cdot\frac{R_{3\max}}{r_{3\min}}}{1+R_{3\max}+2\cdot R_{3\max}^2} \tag{45}$$

We notice that $(\Delta\, r_{ord})_{med}.>0$

In this context, the implementaton of factorial analysis methodology is necessary to take some additional steps:

1) *If there is one negative Pearson coefficient of correlation between the explanatory variables, algebraic transformations are made so that the Pearson coefficient of correlation with the largest absolute value become positive. Thus, this indicator will further act as* R_{max}*. If, after algebrical transformations, there is a negative Pearson correlation coefficient, the coefficient* K_1 *is computed as a ratio between the two positive Pearson coefficients of correlation. Coefficient* K_2 *represents the ratio of the negative Pearson coefficient of correlation to the positive Pearson coefficient of correlation with smaller absolute value.*
2) *If coeficient* r_{21} *is negative, the methodology presented above will be used, but taking into account the sign of the respective coefficient. Regarding the ranks of the explanatory variables, they will be shown, as in the standard methodology, by the absolute value of the Pearson coefficients of correlation between explanatory variables and resultative variable.*
3) *If coeficient* r_{32} *is negative, the detailed factorial analysis cannot be applied for the differentiation of coefficients of m.r.v. correlation. In these circumstances, we can compute only syntetically the influence of the differentiation of coefficients of m.r.v. correlation and in a second stage we can the detailed factorial analysis concerning the influence of the differentiation of Pearson coefficients of correlation between explanatory variables.*

III) The influence of the real differentiation of coefficients of m.r.v. correlation between explanatory variables ((Δ $r_{ef})_{med}$).

$$\Delta r_{ef} = \frac{1+2\cdot R_{3\max}+3\cdot R_{3\max}\cdot\dfrac{r_{1medarm}}{r_{1medharm}}}{1+R_{3\max}+2R_{3\max}^2} - \frac{1+2\cdot R_{3\max}-(\dfrac{1}{3})\cdot(\dfrac{1-r_{3\min}\cdot\sqrt{r_{3\min}}}{1-\sqrt{r_{3\min}}})^2\cdot\dfrac{R_{3\max}}{r_{3\min}}}{1+R_{3\max}+2\cdot R_{3\max}^2}$$

(46)

We notice that (Δ $r_{ef})_{med}<0$.

In order to determine the **influence of the differentiation of Pearson coefficients of correlation related to R_{3max}** (Δ (k1, k2)) it is firstly necessary to compute an additional indicator, ($T_{3refk1ef}$). The respective indicator is related to the real values of R_{3max}, m.r.v. coefficients of correlation and of coefficient K_1, while K_2=1.

On this base, we can compute two influences:

I) The influence of coefficient K_1 differentiation related to R_{3max} ((Δ $k1)_{med}$)

$$(\Delta k_1)_{med} = (T_{3Rrefk1ef})_{med} - \frac{1+2\cdot R_{3\max}-3\cdot R^2{}_{3\max}\cdot\dfrac{r_{1medaritm}}{r_{1medharm}}}{1+R_{3\max}-2\cdot R_{3\max}^2} \tag{47}$$

We notice that (Δ $k1)_{med}>0$.

II) The influence of the real value of the coefficient K_2 ((Δ $k_{2ef})_{med}$)

$$\Delta k_{2ef} = (T_{3ef})_{med} - (T_{3Rrefk1ef})_{med} \tag{48}$$

Usually, (Δ $k_{2ef})_{med}>0$.

Also, we can compute the **cumulative influence of the real differentiation of Pearson coefficients of correlation and coefficients of m.r.v. correlation between the explanatory variables (Δ (r, K_1, $K_2)_{med}$).**

$$(\Delta(r,k_{1,}k_2))_{med} = (T_{3k})_{med} - \frac{1}{1+2\cdot R_{3\max}} \tag{49}$$

The sign of (Δ (r, K_1, $K_2)_{med}$) depends on the features of the differentiation of coefficients of m.r.v. correlation between explanatory variables and on the type of essential mediated collinearity.

Similarly, the above influences can be computed on each individual coefficient of alignment to collinearity hazard.

Two numerical examples. Factorial analysis of coefficients of alignment to collinearity hazard in case of estimation of the Kmenta production function related to non-manual and manual professious in Spain and the United Kingdom.

In order to illustrate the possibility of practical use of the proposed factorial analysis model for the coefficients of alignment to collinearity hazard we have chosen Kmenta function applied to non-manual and manual segments of total employment estimated parameters in case of Spain and U.K. economy during 1995-2002 (Pavelescu F.M. coordinator, 2007)[1].

This function is defined as: $\ln Y = \ln A_3 + \alpha_3 \cdot \ln L_w + \beta_3 \cdot \ln L_b + \chi_3 \cdot \ln^2 (L_w/L_b)$, where:

$\ln Y$= natural logarithm of gross domestic product indices.

$\ln L_w$= natural logarithm of indices of non-manual professions.

$\ln L_{b\,/}$= natural logarithm of indices of manual professions.

$\ln A_3$, α_3, β_3 and χ_3= estimated parameters.

As for Spain the estimation result is:

$$\ln Y = -0.0048 + 0.0396 \cdot \ln L_w + 0.0541 \cdot \ln L_b - 1.7303 \cdot \ln^2 (L_w/L_b), \quad R^2_3 = 0.9972$$

(-0.4381) (4.0794) (0.2499) (-1.6771)

N.B. R^2_3 is the coefficient of determination and the computed values of Student test are shown in brackets.

The proper value of the estimated parameters are: α_1=0.7257, β_1=1.1328 and χ_1=7.6494, while coefficients of alignment to collinearity hazard are: $T_{3\alpha}$=1.1570, $T_{3\beta}$=0.0478 and $T_{3\chi}$=-0.2262. It is important to observe that the

[1]*A criticism of the numerical examples presented here is the fact that time series used for estimations are very short. This is mainly a consequence of the changes in the international statistical methodology concerning the proffesional structure of total employment. But, having in view that first target of these numerical examples is to emphasize the modelling factors of coefficients of alignment to collinearity hazard and to explore the of occurence of regularities and exceptions to the rule in the context of linear regression with three explanatory variables regression, we may consider that the examples presented further are relevant.*

coefficient of alignment to collinearity hazard related to parameter χ_3 is negative. Therefore, in this case we face a harmful collinearity.

In order to detect the causes of the occurrence of the negative coefficient of alignment to collinearity hazard, we first computed the Pearson coefficients of correlation between each explanatory variable and the resultative variables. On this basis, we have computed the coefficients of m. r. v. correlation between the explanatory variables related to the primordial explanatory variable of the regression model (r_{k1}). This way, the ranks of the explanatory variables could be established (Table 1).

Table 1. The coefficients of m.r.v. correlation between explanatory variables for the function $\ln Y = \ln A_3 + \alpha_3 \cdot \ln L_w + \beta_3 \cdot \ln L_b + \chi_3 \cdot \ln^2 (L_w/L_b)$, in Spain, 1995-2002

(1994=100)

Explanatory variable	R(ln Y, ln (X_k))	r_{k1}	$r_{k+1,k}$
Ln L_w	0.9963	1.0000	-
Ln L_w	0.9874	0.9911	0.9911
$\ln^2 (L_w/L_b)$	0.9352	0.9386	0.9471

The values of Pearson coefficient of correlation between explanatory variables are bigger than 0.9350, while r_{3min} = 0.9471. The differentiation trajectory of the coefficients of m.r.v. correlation is overexponential. The ranks of the explanatory variables are: ln L_w = primordial explanatory variable, ln L_b= explanatory variable of rank 2, $\ln^2 (L_w/L_b)$ = explanatory variable of rank 3. All the Pearson coefficients of correlation between the explanatory variable are positive (Table 2).

Table 2. Matrix of Pearson coefficients of correlation between explanatory variables for the function $\ln Y = \ln A_3 + \alpha_3 \cdot \ln L_w + \beta_3 \cdot \ln L_b + \chi_3 \cdot \ln^2 (L_w/L_b)$, in Spain, 1995-2002

(1994=100)

Explanatory variable	Ln L_w	ln L_b	$\ln^2 (L_w/L_b)$
Ln L_w	1.0000	0.9814	0.9579
Ln L_b	0.9814	1.0000	0.9028
$\ln^2 (L_w/L_b)$	0.9579	0.9028	1.0000

The explanatory variable L_b represents the main attractor of collinearity; explanatory variable L_w, the attractor of collinearity of order 2; and explanatory variable $\ln^2 (L_w/L_b)$, the attractor of collinearity of order 3. R_{3max} = 0.9814, K_1 =0.9761, and K_2 = 0.9424. It is important to notice the very high value of the coefficients R_{3max}. and K_1. Also, there is a potential harmful essential mediated collinearity, because $K_2 < R_{3max}$.

Therefore, the coefficient of alignment to collinearity hazard, in conditions of non-differentiation of coefficients of Pearson and m.r.v. correlation between explanatory variables (T_{3R}), is 0.3375, (Table 3). In this context, the first premises are for a degrading collinearity.

Table 3. The coefficients of alignment to collinearity hazard depending on the differentiation of the coefficients of correlation between explanatory variables for the function $\ln Y = \ln A_3 + \alpha_3 \cdot \ln L_w + \beta_3 \cdot \ln L_b + \chi_3 \cdot \ln^2 (L_w/L_b)$, in Spain, 1995-2002

(1994=100)

Explanatory variable	Ln L_w	Ln L_b	$\ln^2 (L_w/L_b)$	Arithmetical mean
T_{3R}	0.3375	0.3375	0.3375	0.3375
T_{3Rrst}	2.5227	-0.8265	-0.8265	0.2899
T_{3Rrord}	1.9851	0.3197	-1.3993	0.3018
T_{3Rref}	1.5893	1.1187	-1.8210	0.2957
$T_{3Rrefk1ef}$	0.8995	0.4227	-0.3455	0.3256
T_{3k}	1.1570	0.0477	-0.2262	0.3262

Because $r_{3\min} < R_{3\max}$, in case of non-differentiation of the Pearson coefficients of correlation and of the standard differentiation of coefficients of m.r.v. correlation between explanatory variables, two of the coefficients of alignment to collinearity hazard (T_{3Rrst}) are negative. In case of ordered differentiation of coefficients of m.r.v. correlation between explanatory variables, one may observe an increased polarization of the absolute values and a change in sign at one of the above-mentioned indicators. Consequently, in the respective differentiation of the coefficients of m.r.v. correlation between explanatory variables, there are two positive coefficients of alignment to collinearity hazard, the third being negative.

In the context of real differentiation of coefficients of m.r.v. correlation between explanatory variables the negativity of one of the coefficients of alignment to collinearity hazard becomes more pronounced.

The values and distributions of coefficients K_1 and K_2 determine an increase in the arithmetical mean of the coefficients of alignment to collinearity hazard. Although the factors mentioned above contribute to a significant increase in the value of the coefficient of alignment to collinearity related to the explanatory variable of rank 3, the respective indicator remains negative. Under these conditions, the collinearity in this linear regression is a harmful one. Also, it is important to remark that the order of the coefficients of alignment to collinearity hazard is similar to the order established by the ranks of explanatory variables.

The computation of factorial influence shows that the Pearson coefficient of correlation between explanatory variables with maximum value (R_{3max}) has a major contribution to obtaining the arithmetical mean of coefficients of alignment to collinearity hazard. The ratio of T_{3R} to $(T_{3k})_{med}$ is 96.63%% (Table 4).

Table 4. Factorial contribution to the determination of coefficients of alignment to collinearity hazard for the function lnY=ln $A_3+\alpha_3 \cdot \ln L_w + \beta_3 \cdot \ln L_b + \chi_3 \cdot \ln^2 (L_w/L_b)$, in Spain during 1995-2002

(1994=100)

Explanatory variable	Ln L_w	Ln L_b	$\ln^2 (L_w/L_b)$	Arithmetical mean
Δr_{st}	2.1852	-1.164	-1.164	-0.0476
Δr_{ord}	-0.5376	1.1462	-0.5728	0.0119
Δr_{ef}	-0.3958	0.799	-0.4217	-0.0061
Δk_1	-0.6898	-0.696	1.4755	0.0299
Δk_2	0.2575	-0.375	0.1193	0.0006
Δ r	**1.2518**	**0.7812**	**-2.1585**	**-0.0418**
$\Delta (k_1,k_2)$	**-0.4323**	**-1.071**	**1.5948**	**0.0305**
$\Delta (r, k_1, k_2)$	**0.8195**	**-0.2898**	**-0.5637**	**-0.0113**

As a whole, coefficients $r_{k+i,\,k}$ determinate a decrease by 0.0418 in the arithmetical mean of the coefficients of alignment to collinearity hazard. The differentiation of Pearson coefficients of correlation between explanatory variables results in an increase in the arithmetical mean of the coefficients of alignment to collinearity hazard, by 0.0305 as the main contribution is made by coefficient K_1.

The computations we have made represent a validation of theoretical assumptions presented before referring to the situation when the differentiation of coefficients of m.r.v. correlation between explanatory variables is overexponential under the conditions of potential harmful essential mediated collinearity.

In case of the United Kingdom, the estimated function is:

$\ln Y=0.0069+0.2791 \ln L_w +0.7477 \ln L_b +1.4668 \ln^2 (L_w/L_b)$, $R^2_3= 0.9981$
(1.2818) (8.7967) (13.4001) (2.1559)

The proper estimated values are: α_1=1.6064, β_1=0.6560 and χ_1=2.2546, and all coefficients of alignment to collinearity hazard are positive: $T_{3\alpha}$=0.7962, $T_{3\beta}$=1.1398 and $T_{3\chi}$=0.6506.

The positivity of all coefficients of alignment to collinearity hazard is obtained in the context of a sensible differentiation of the intensity of correlation between the explanatory variables and resultative variable. The Pearson coefficients of correlation between the resultative variables and explanatory variables are dispersed, ranging between 0.2849 and 0.7906, and thus r_{3min}= 0.3594. The ranks of explanatory variables are: ln L_w = primordial explanatory variable, ln L_b= explanatory variable of rank 2 $\ln^2 (L_w/L_b)$ = explanatory variable of rank 3 (table 5).

Table 5. The coefficients of m.r.v. correlation between explanatory variables for the function $\ln Y=\ln A_3+\alpha_3 \ln L_w +\beta_3 \cdot \ln L_b +\chi_3 \cdot \ln^2 (L_w/L_b)$, in the United Kingdom, 1995-2002

(1994=100)

Explanatory variable	R(ln Y, ln (X_k))	r_{k1}	$R_{k+1,k}$
Ln L_w	0,7906	1.0000	-
Ln L_b	0,6269	0,7929	0,7929
$\ln^2 (L_w/L_b)$	0,2842	0,3594	0,4533

The matrix of Pearson coefficients of correlation between explanatory variables reveals a negative value of R (ln L_b; $\ln^2(L_w/L_b)$) signaling the existence of anticollinearity in the linear regression equation (Table 6).

Table 6. Matrix of Pearson coefficients of correlation between explanatory variables for the function $\ln Y=\ln A_3+\alpha_3 \cdot \ln L_w +\beta_3 \cdot \ln L_b +\chi_3 \cdot \ln^2 (L_w/L_b)$, in United Kingdom, 1995-2002

(1994=100)

Explanatory variable	Ln L_w	ln L_b	$\ln^2 (L_w/L_b)$
Ln L_w	1,0000	0.0227	0,7834
Ln L_b	0,0227	1,0000	-0,5513
$\ln^2 (L_w/L_b)$	0.7834	-0,5513	1,0000

As a consequence, explanatory variable $\ln^2 (L_w/L_b)$ represents the main attractor of collinearity, explanatory variable $\ln L_w$ is the attractor of collinearity of order 2, and explanatory variable $\ln L_b$ is the attractor of collinearity of order 3. R_{3max} = 0.7834, K_1 = 0.0290, and K_2 = -24.2662.

The maximum absolute value of Pearson coefficient of correlation between the explanatory variables (R_{3max}) determines that T_{3R}= 0.3896 (Table 7). Therefore, the initial premises are for a degrading collinearity.

Because $r_{3\min} < R_{3\max}$, under the conditions of non-differentiation of the Pearson coefficients of correlation and of standard differentiation of coefficients of m.v.r. correlation between explanatory variables, two of the coefficients of alignment to collinearity hazard (T_{3Rrst}) are negative. Also, the arithmetical mean of the respective indicator is negative.

Table 7. The coefficients of alignment to collinearity hazard depending on the differentiation of the coefficients of correlation between explanatory variables for the function $\ln Y = \ln A_3 + \alpha_3 \cdot \ln L_w + \beta_3 \cdot \ln L_b + \chi_3 \cdot \ln^2 (L_w/L_b)$, in the United Kingdom, 1995-2002

(1994=100)

Explanatory variable	ln L_w	Ln L_b	$\ln^2 (L_w/L_b)$	Arithmetical mean
T_{3R}	0.3896	0.3896	0.3896	0.3896
T_{3Rrst}	2.1944	-2.1209	-2.1209	-0.6825
T_{3Rrord}	1.8561	0.0128	-3.0619	-0.3977
T_{3Rref}	1.5838	0.7917	-3.8197	-0.4814
$T_{3Rrefk1ef}$	1.8496	0.9787	-3.0801	-0.0839
T_{3k}	0.7963	1.1398	0.6506	0.8622

Under the conditions of ordered distribution of coefficients of m.r.v. correlation between explanatory variables there are two positive coefficients of alignment to collinearity hazard. The third coefficient of alignment to collinearity hazard takes a very strong negative value. The real differentiation of coefficients of m.r.v. correlation between explanatory variables leads to an increase in the negativity of the coefficient of alignment to collinearity hazard related to explanatory variable of rank 3. This is a consequence of an underexponential differentiation of the coefficients of m.r.v. correlation between explanatory variables.

Coefficient K_1 acts to decrease the polarization of individual coefficients of alignment to collinearity hazard and for the increase in the arithmetical mean of the above-mentioned indicator. Coefficient K_2, through its negative value contributes essentially to the occurrence of positive values for all the coefficients of alignment to collinearity hazard. The presence of anticollinearity determines an increase by 0.9461 in the arithmetical mean of coefficients T_{3k} $(T_{3k})_{med}$ as against the level obtained in case of T_{3R}, respectively from 0.3896 to 0.8622. (table 8).

Table 8. Factorial contribution to the determination of coefficients of alignment to collinearity hazard for the function $\ln Y = \ln A_3 + \alpha_3 \cdot \ln L_w + \beta_3 \cdot \ln L_b + \chi_3 \cdot \ln^2 (L_w/L_b)$, în the United Kingdom, 1995-2002

(1994=100)

Explanatory variable	Ln L_w	ln L_b	$\ln^2 (L_w/L_b)$	Arithmetical mean
Δ r_{st}	1.8048	-2.5105	-2.5105	-1.0721
Δ r_{ord}	-0.3383	2.1337	-0.941	0.2848
Δ r_{ef}	-0.2723	0.7789	-0.7578	-0.0837
Δ k_1	0.2658	0.187	0.7396	0.3975
Δ k_2	-1.0533	0.1611	3.7307	0.9461
Δ r	1.1942	0.4021	-4.2093	-0.8710
Δ (k_1,k_2)	-0.7875	0.3481	4.4703	1.3436
Δ (r, k_1, k_2)	0.4067	0.7502	0.261	0.4726

At the same time, the coefficient K_2 determines a sensible increase in the value of the coefficient of alignment to collinearity hazard related to the explanatory variable of rank. 3. This way the respective indicator becomes positive. The impact on the values of the other coefficients of alignment to collinearity the hazard is a moderate growth in the case of the pointer related to the explanatory variable of rank 2 and a decrease in case of the indicator related to the explanatory variable of rank 1. Consequently, all individual values of the coefficients of alignment to collinearity hazard are higher than 0.65. Therefore, we can identify the existence of a weak collinearity, essentially generated by the anticollinearity between the explanatory variables of rank 2 and rank 3. It is important to notice that anticollinearity between the respective explanatory variables determined not only a weak collinearity at the level of the linear regression equation, but also a change in the order of values of coefficients of alignment to collinearity hazard, which is different from the order established by ranks of explanatory variables.

6. Conclusions. Types of collinearity and the relationship between the ranks of explanatory variables and the order of attractors of collinearity

In the case of linear regression with three explanatory variables, estimating parameters by ordinary least squares method is influenced by the existence of collinearity. The respective phenomenon is revealed by the individual values of coefficients of alignment to collinearity hazard. In order to provide feasible estimation results and to avoid harmful collinearity, it is necessary that all three coefficients of alignment to collinearity hazard be positive. Also, an indicator of feasibility for the linear regression model as a whole is the arithmetical mean of coefficients of alignment to collinearity hazard. In the context of positivity of all individual values, the higher arithmetical mean of respective indicator is, the lower the distortions generated by collinearity in a linear regression model with three explanatory variables could be considered.

In the analysis of the coefficients of alignment to collinearity hazard it is important to consider that the modeling factors of the indicator are:

- maximum absolute value of Pearson coefficients of correlation between explanatory variables (R_{3max})[1];
- the minimum value of the coefficients of m.r.v. between the explanatory variables in relation to the primordial explanatory variable in the linear regression model (r_{3min})[2];
- trajectory of differentiation of the coefficients of m.r.v. correlation between the explanatory variables;
- type of essential mediated collinearity (non-harmful, potential harmful, anticollinearity);
- attractors of collinearity distribution as against explanatory variables.

[1] *If R_{3max} is high, there are prerequisites for a low arithmetical mean of the coefficients of alignment to collinearity hazard and to a degrading or harmful collinearity.*

[2] *r_{3min} has a direct influence on the accomplishment of the positivity condition for all the coefficients of alignment to collinearity hazard. Low values of this indicator favour a sensible decrease in the arithmetical mean and the occurrence of harmful collinearity.*

If the last three modeling factors are taken into account, we may identify in a linear regression with three explanatory variables a number of 36 cases of coefficients of alignment to collinearity hazard (Table 9)[1].

Table 9. Possible cases for the coefficients of alignment to collinearity hazard in a linear regression with three explanatory variables

Cases	Differentiation trajectory form. $r_{k,k+1}$	Type of essential mediated collinearity	Ranks of explanatory variables			Order of attractor of collinearity		
A1	over exponential	non-harmful	1	2	3	1	2	3
A2	overexponential	potential harmful	1	2	3	1	2	3
A3	overexponential	anticollinearity	1	2	3	1	2	3
A4	underexponential	non-harmful	1	2	3	1	2	3
A5	underexponential	potential harmful	1	2	3	1	2	3
A6	underexponential	anticollinearity	1	2	3	1	2	3
B1	over exponential	non-harmful	1	3	2	1	2	3
B2	overexponential	potential harmful	1	3	2	1	2	3
B3	overexponential	anticollinearity	1	3	2	1	2	3
B4	underexponential	non-harmful	1	3	2	1	2	3
B5	underexponential	potential harmful	1	3	2	1	2	3
B6	underexponential	anticollinearity	1	3	2	1	2	3
C1	over exponential	non-harmful	2	1	3	1	2	3
C2	overexponential	potential harmful	2	1	3	1	2	3
C3	overexponential	anticollinearity	2	1	3	1	2	3
C4	underexponential	non-harmful	2	1	3	1	2	3
C5	underexponential	potential harmful	2	1	3	1	2	3
C6	underexponential	anticollinearity	2	1	3	1	2	3
D1	over exponential	non-harmful	2	3	1	1	2	3
D2	overexponential	potential harmful	2	3	1	1	2	3
D3	overexponential	anticollinearity	2	3	1	1	2	3
D4	underexponential	non-harmful	2	3	1	1	2	3
D5	underexponential	potential harmful	2	3	1	1	2	3
D6	underexponential	anticollinearity	2	3	1	1	2	3
E1	over exponential	non-harmful	3	1	2	1	2	3
E2	overexponential	potential harmful	3	1	2	1	2	3

[1] *The number of 36 cases represents the product between the number of differentiation trajectories of coefficients of m.r.v. (2), the number of types of essential mediated collinearity (3), and the number of permutations of atractors of collinearity as against ranks of explanatory variables (6).*

Cases	Differentiation trajectory form. $r_{k,k+1}$	Type of essential mediated collinearity	Ranks of explanatory variables			Order of attractor of collinearity		
E3	overexponential	anticollinearity	3	1	2	1	2	3
E4	underexponential	non-harmful	3	1	2	1	2	3
E5	underexponential	potential harmful	3	1	2	1	2	3
E6	underexponential	anticollinearity	3	1	2	1	2	3
F1	over exponential	non-harmful	3	2	1	1	2	3
F2	overexponential	potential harmful	3	2	1	1	2	3
F3	overexponential	anticollinearity	3	2	1	1	2	3
F4	underexponential	non-harmful	3	2	1	1	2	3
F5	underexponential	potential harmful	3	2	1	1	2	3
F6	underexponential	anticollinearity	3	2	1	1	2	3

The review of all possible situations reveals that in a linear regression with three explanatory variables, the attractor s of collinearity distribution as against ranks of explanatory variables has an important influence in obtaining the values of coefficients of alignment to collinearity hazard.

In order to avoid the occurrence of harmful collinearity, the most favorable situations are when the primordial variable is the main attractor of collinearity and the most unfavorable one is when the main attractor of collinearity is the explanatory variable of rank 3. This is a consequence of fact that the main attractor of collinearity put the highest pressure for the occurrence of negative coefficients of alignment to collinearity hazard. The respective pressure can be alleviated by the presence of under-unit coefficients of m.r.v. correlation between the explanatory variables, as is the case of the primordial explanatory variable.

We should note that if the main attractor of collinearity is the primordial explanatory variable, the attractor of collinearity of order 2 is the explanatory variable of rank 2 and the attractor of collinearity of order 3 is the explanatory variable of rank 3, it occurs a symmetry between the differentiation of the Pearson coefficients of correlation between explanatory variables and the differentiation of the coefficients of m.r.v. correlation between explanatory variables. In this case, it is possible to prevent the occurrence of negative coefficients of alignment to collinearity hazard and consequently the harmful collinearity. But this favorable situation is obtained at the price of the “contestation” of the order of importance (ranks) of the explanatory variables established by the coefficients of Pearson correlation between the resultative variables and each of the explanatory variables.

The above-mentioned "contestation "[1] causes a rebalance of the coefficients of alignment to collinearity hazard and consequently to a decreasing probability for the occurrence of the harmful collinearity.

The situations when the ranks of explanatory variables and the order of attractors of collinearity are not the same represent a form of a "disorder" in the relationship between the resultative variable and the explanatory variables, taken as a whole. In fact, we may admit that the "standard model of the relationship between the explanatory variables and attractors of collinearity" is considered when the ranks of the explanatory variables are the same with the order of attractor s of collinearity.

The non-harmful and potential harmful essential mediated collinearity may appear both in the standard model of relationship between the explanatory variables and attractors of collinearity and also in the context of departures from the respective model.

In case of an essential mediated anticollinearity we may find only the situation of departure from the standard model of relationship between the explanatory variables and the attractors of collinearity. This is a consequence of the fact that anticollinearity represents a "disorder" in relation to explanatory variables.

Therefore, in order to improve the feasibility of the estimated parameters of a linear regression model when using least squares method is important to consider the following rules:

a) before estimating parameters of a linear regression model it is necessary to compute the values of the Pearson coefficients of correlation between explanatory variables;

b) avoid the use of explanatory variables highly correlated or explanatory variables that strongly differ in terms of their absolute correlation with the resultative variable;

c) test if $r_{3\min} > R_{3\max}$.

If the respective condition is accomplished one may begin to estimate the linear regression model. If $r_{3\min} < R_{3\max}$ we have to check the existence of

---

[1] The respective "contestation" is the factor that determines in a lot of cases the values of the coefficients T_{nk}, which at first sight appear to be hazardous. This is the reason why we call the coefficients mentioned above "coefficients of alignment to colinearity hazard".

anticollinearity, revealed the divergence of signs of the coefficients of Pearson and m.r.v correlation between explanatory variables. If the anticollinearity is not detected parameter estimation of the linear regression model with three explanatory variables is no longer paid, because at least one negative coefficient of alignment to collinearity hazard will occur.

References

1. D. Belsey, "Multicollinearity: Diagnosing its presence and assessing the potential damage it causes least square estimation", NBER Working Paper no. 154, Cambridge, Massachusetts, October 1976.
2. D. Belsey, "Conditional diagnostics: collinearity and weak data in regression", *Wiley Series in Probability*, John Wiley, New York, 1991.
3. C. Conrad, "Applied Regression Analysis. Lectures Notes", Department of Economics, Pomona College, Claremont, California, Spring 2006.
4. C. Dougherty, Introduction *to econometrics*, Second Edition, Oxford University Press, 2002.
5. D. Douglas, D. Clader, J. Cristy, P. Michaelis, D. Belsey, "Test about harmful collinearity among predictor variables used in modeling global temperature", *Climate research,* vol. 24, June 2003.
6. D. Farrar, R. Glauger, "Multicollinearity in regression analysis: The problem revisited", *Review of Economics and Statistics*, February 1967.
7. W. Greene, *Econometric Analysis*, Mc Millan Publishing Company, New York, 1993.
8. A. Isaic-Maniu, C. Mitrut, V. Voineagu, *Statistics for Business Management*, Economic Publishing House, Bucharest 1996.
9. D. Jula, *Introduction to Econometrics*, Editure Professional Consulting, Bucharest, 2003.
10. L. Klein, An introduction to econometrics, Prentice Hall, 1962.
11. F. M. Pavelescu, "Some considerations regarding the Cobb-Douglas production function estimated parameters. A new approach", *Revue Roumaine des Sciences Sociales,* Tome 30, No. 1-2/1986.
12. F. M. Pavelescu, "Impact of collinearity on the estimated parameters and values of classical Statistical tests in conditions of multifactorial linear OLS regressions" in *Romanian Journal of Economic Forecast,* No. 2/2005.
13. F. M. Pavelescu (coord.), C.Ciutacu, L. Chivu, A.C. Munteanu, *Capitalul uman şi creşterea economică*, Editura Expert, Bucureşti, 2007 (Human capital and economic growth, Expert Publishing House, Bucharest, 2007).
14. F.M. Pavelescu, "A review of Student Test properties in condition of multifactorial linear regression", în *Romanian Journal of Economic Forecast,* No.1/2009.
15. E. Pecican, *Macroeconometries. Government economic policy and econometrics*. Editura Economică, Bucureşti, 1996 (Macroeconometrics. Gouvernamental econometrics and economic policies. Economic Publishing House, Bucharest, 1996).
16. A.Tudorel, S. Stancu, *Statistică. Teorie and aplicații*, Editura All, Bucureşti, 1994. (Statistics. Theory and applications, ALL Publishing House, Bucharest, 1994).
17. R. Wiliams, *Sociology Graduate Statistics*, University of Notre Dame, Indiana, 2008.