EKONOMETRIA ECONOMETRICS 3(41) • 2013
ISSN 1507-3866

Joanna Plebaniak
Warsaw School of Economics

ON THE QUALITY OF REGULAR CORRELATED PAIRS (R(*K*), R_0(*K*)), MEASURED BY THE VALUE OF THE COEFFICIENT R^2(*K*)

Abstract: If a correlated pair ($\mathbf{R}(k)$, $\mathbf{R}_0(k)$) is a regular correlated pair, then the coefficient measuring the quality of such a pair satisfies the inequality:

$$\mathrm{r}^2(k) \geq F(k).$$

If the correlation matrix $\mathbf{R}(k)$ is majorized by the universal matrix, then:

$$F(k) = \frac{\mathbf{R}_0^{\mathrm{T}}(k)\mathbf{R}_0(k)}{1 + \mathbf{R}_0^{\mathrm{T}}(-1)\mathbf{R}_0(-1)},$$

where $\mathbf{R}_0^{\mathrm{T}}(-1) = \begin{bmatrix} r_2 & r_3 & \ldots & r_k \end{bmatrix}$.

If a regular correlated pair ($\mathbf{R}(k)$, $\mathbf{R}_0(k)$) is calculated using the method of maximum value of the integral information volume, then:

$$F(k) = \frac{\mathbf{R}_0^{\mathrm{T}}(k)\mathbf{R}_0(k)}{\max S_i},$$

where S_i signifies the sum of modules of the elements located in the *i* column of the $\mathbf{R}(k)$ matrix.

In each of those cases it is easy to calculate the value of $F(k)$ as far as the arithmetic is concerned.

Keywords: correlation matrix, universal matrix, Hellwig's inequality, the coefficient $\mathrm{r}^2(k)$.

The literature (see: [Hellwig 1976; Kolupa 1977; Kolupa 1993; Kolupa; Plebaniak 2011]) presents a number of different inequalities, the solution of which signifies that the quality of a regular correlated pair ($\mathbf{R}(k)$, $\mathbf{R}_0(k)$) (and therefore the quality of the uni-equational linear econometric model defined by this pair), measured by the value of the coefficient $\mathrm{r}^2(k)$ is at least equal to the quantity $F(k)$, for which:

$$\mathrm{r}^2(k) \geq F(k). \tag{1}$$

In practice, the quantity $F(k)$ needs to be relatively easy to calculate, because we need a criterion upon which the indicator $\mathrm{r}^2(k)$ can be estimated.

This paper presents a review of type (1) inequalities.

Let $\mathbf{R}(k) = [r_{ij}]$ signify a k-degree correlation matrix, the elements r_{ij} of which denote indicators of correlation between the pairs of individual explanatory variables Z_i and Z_j ($r_{ij} = r(Z_i, Z_j)$, $i, j = 1, 2, \ldots, k$) of a uni-equational linear econometric model of the following form:

$$Y = \alpha_1 Z_1 + \alpha_2 Z_2 + \ldots + \alpha_k Z_k + e. \quad (2)$$

And let us have a k-dimensional column vector $\mathbf{R}_0(k) = [r_i]$ with elements r_i constituting the coefficients of correlation between the endogenous variable Y of the model (2) and the individual explanatory variables Z_i ($r_i = r(Y, Z_i)$, $i = 1, 2, \ldots, k$) of the same model.

Quoting after Z. Hellwig (see: [Hellwig 1976]), if:

$$0 < r_1 \le r_2 \le \ldots \le r_k < 1, \quad (3)$$

then the correlated pair $(\mathbf{R}(k), \mathbf{R}_0(k))$ is called a regular correlated pair.

The quality of each correlated pair $(\mathbf{R}(k), \mathbf{R}_0(k))$, and in particular a regular pair of this type, is measured by the value of the coefficient $\mathrm{r}^2(k)$ defined by the formula:

$$\mathrm{r}^2(k) = \mathbf{R}_0^{\mathrm{T}}(k)\mathbf{R}^{-1}(k)\mathbf{R}_0(k). \quad (4)$$

Let us remember that the correlated pair $(\mathbf{R}(k), \mathbf{R}_0(k))$ exists only when:

$$0 < \mathrm{r}^2(k) < 1 \quad (5)$$

(see: [Hauke, Pomianowski 1984]).

Let us note that if the matrix inequality:

$$\mathbf{0}(k) \le \mathbf{R}(k) \le \mathbf{G}(k) \quad (6)$$

is satisfied, where $\mathbf{0}(k)$, $\mathbf{R}(k)$, $\mathbf{G}(k)$ signify a k-degree matrix constituting, consequently, the zero matrix, the correlation matrix and the universal matrix, then the model defined by the regular correlated pair is coincident.

The above is the content of the famous hypothesis presented by Hellwig in 1976 and proven by Kolupa in 1993 (see: [Kolupa 1993]).

The inequality (6) contains the condition $\mathbf{R}(k) \le \mathbf{G}(k)$, which means that the correlation matrix $\mathbf{R}(k)$ is majorized by the universal matrix $\mathbf{G}(k)$. The matrix $\mathbf{G}(k)$ is defined below (see (7)).

Let us remember that the elements of the matrix $\mathbf{G}(k)$ are the elements g_{ij} defined as follows (see: [Hellwig 1976]):

$$g_{ij} = \begin{cases} 1 & \text{dla} \quad i = j \\ r_i r_j & \text{dla} \quad i \ne j \end{cases}, \quad (7)$$

where the coefficients r_i and r_j are the correlation coefficients between the pairs of variables (Y, Z_i) and (Y, Z_j), $i, j = 1, 2, \ldots, k$.

The paper (see: [Kolupa, Plebaniak 2011]) proves that if ($\mathbf{R}(k)$, $\mathbf{R}_0(k)$) is a regular correlated pair and the correlation matrix $\mathbf{R}(k)$ is majorized by the universal matrix $\mathbf{G}(k)$, then:

$$\mathrm{r}^2(k) \geq \frac{Q}{1+Q}, \tag{8}$$

where:

$$Q = \sum_{i=1}^{k} \frac{r_i^2}{1-r_i^2}. \tag{9}$$

Considering that the pair ($\mathbf{R}(k)$, $\mathbf{R}_0(k)$) is a regular correlated pair (see (3)), we have:

$$Q = \frac{\mathbf{R}_0^{\mathrm{T}}(k)\mathbf{R}_0(k)}{1-r_1^2} \tag{10}$$

and:

$$1+Q = \frac{1-r_1^2+\mathbf{R}_0^{\mathrm{T}}(k)\mathbf{R}_0(k)}{1-r_1^2} + \frac{1+(r_2^2+\ldots+r_k^2)}{1-r_1^2} = \frac{\mathbf{R}_0^{\mathrm{T}}(-1)\mathbf{R}_0(-1)}{1-r_1^2}, \tag{11}$$

therefore:

$$\frac{Q}{1+Q} = \frac{\mathbf{R}_0^{\mathrm{T}}(k)\mathbf{R}_0(k)}{1-r_1^2}\,\frac{1-r_1^2}{1+\mathbf{R}_0^{\mathrm{T}}(-1)\mathbf{R}_0(-1)} = \frac{\mathbf{R}_0^{\mathrm{T}}(k)\mathbf{R}_0(k)}{1+\mathbf{R}_0^{\mathrm{T}}(-1)\mathbf{R}_0(-1)}, \tag{12}$$

where:

$$\mathbf{R}_0^{\mathrm{T}}(-1) = \begin{bmatrix} r_2 & r_3 & \ldots & r_k \end{bmatrix}. \tag{13}$$

Finally, on the basis of (8) and (12) we obtain:

$$\mathrm{r}^2(k) \geq \frac{\mathbf{R}_0^{\mathrm{T}}(k)\mathbf{R}_0(k)}{1+\mathbf{R}_0^{\mathrm{T}}(-1)\mathbf{R}_0(-1)}. \tag{14}$$

Note that it is easy to calculate the expression constituting the right side of the inequality (14). Thus we can state that the quality of a correlated pair expressed by the coefficient $\mathrm{r}^2(k)$ at least equals the expression on the right side of the inequality (14).

The formula (14) shows that the value of the expression on the right side of the formula (14) is less than one. Its numerator is the square of the vector $\mathbf{R}_0(k)$, while the denominator is the sum of 1 and the square of the vector $\mathbf{R}_0(-1)$ created from the vector $\mathbf{R}_0(k)$ by excluding its first element r_1.

Hellwig proved (see: [Hellwig 1976]) that the coefficient $\mathrm{r}^2(k)$ calculated for a regular correlated pair ($\mathbf{R}(k)$, $\mathbf{R}_0(k)$) determined by the method of maximum value of the integral information volume indicator, satisfies the inequality:

$$\mathrm{r}^2(k) \geq H(k), \tag{15}$$

where:

$$H(k) = \sum_{i=1}^{k} \frac{r_i^2}{S_i} \tag{16}$$

and S_i, $i = 1, 2, \ldots, k$ is the sum of modules of the elements of the matrix **R**(k) located in the i column of the matrix.

Note that for the coefficient $H(k)$, the following inequality will be satisfied:

$$\frac{\mathbf{R}_0^{\mathrm{T}}(k)\mathbf{R}_0(k)}{\max S_i} \le H(k) \le \frac{\mathbf{R}_0^{\mathrm{T}}(k)\mathbf{R}_0(k)}{\min S_i} \tag{17}$$

hence the inequality (15) can be substituted with the inequality:

$$\mathrm{r}^2(k) \ge \frac{\mathbf{R}_0^{\mathrm{T}}(k)\mathbf{R}_0(k)}{\max S_i}. \tag{18}$$

Therefore the inequality (18) signifies that the quality of a regular correlated pair (**R**(k), $\mathbf{R}_0(k)$) can be estimated without the need to calculate the value of $H(k)$.

Let us now notice that if the sum of modules of the elements matrix **R**(k) located in each of its columns is constant and equal to q, then:

$$H(k) = \frac{1}{q}\mathbf{R}_0^{\mathrm{T}}(k)\mathbf{R}_0(k). \tag{19}$$

For example, this is the case with any matrix of the second degree of correlation $\left(H(2) = \frac{r_1^2 + r_2^2}{1 + r_{12}} \right)$.

Generally therefore, the inequality (18) signifies that the quality of a regular correlated pair (**R**(k), $\mathbf{R}_0(k)$) measured by the value of the coefficient $\mathrm{r}^2(k)$ is no less than the square of the vector $\mathbf{R}_0(k)$ multiplied by the inverse of the largest sum of modules of the elements located in every column of the matrix **R**(k).

Let us conclude with an illustration of the formulas (14) and (18).

Example 1 (illustrating formula (14))

We are analyzing a regular correlated pair (**R**(3), $\mathbf{R}_0(3)$) (see: [Kolupa, Plebaniak 2011]), where:

$$\mathbf{R}(3) = \begin{bmatrix} 1 & 0{,}01 & 0{,}02 \\ 0{,}01 & 1 & 0{,}04 \\ 0{,}02 & 0{,}04 & 1 \end{bmatrix}, \qquad \mathbf{R}_0(3) = \begin{bmatrix} 0{,}1 \\ 0{,}2 \\ 0{,}3 \end{bmatrix}. \tag{20}$$

Coefficient $\mathrm{r}_{\mathbf{R}}^2(3)$ for the model defined by the correlated pair described by the formula (20) equals (see: [Kolupa 1982]):

$$r_{\mathbf{R}}^2(3) = \mathbf{R}_0^{\mathrm{T}}(3)\mathbf{R}^{-1}(3)\mathbf{R}_0(3) = \begin{bmatrix} 0{,}1 & 0{,}2 & 0{,}3 \end{bmatrix} \begin{bmatrix} 1 & 0{,}01 & 0{,}02 \\ 0{,}01 & 1 & 0{,}04 \\ 0{,}02 & 0{,}04 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0{,}1 \\ 0{,}2 \\ 0{,}3 \end{bmatrix} \approx 0{,}1339. \tag{21}$$

The right side of the formula (14) equals:

$$\frac{\mathbf{R}_0^{\mathrm{T}}(3)\mathbf{R}_0(3)}{1+\mathbf{R}_0^{\mathrm{T}}(-1)\mathbf{R}_0(-1)} = \frac{\begin{bmatrix} 0{,}1 & 0{,}2 & 0{,}3 \end{bmatrix}\begin{bmatrix} 0{,}1 \\ 0{,}2 \\ 0{,}3 \end{bmatrix}}{1+\begin{bmatrix} 0{,}2 & 0{,}3 \end{bmatrix}\begin{bmatrix} 0{,}2 \\ 0{,}3 \end{bmatrix}} \approx 0{,}12 \tag{22}$$

therefore the inequality (14) is satisfied.

Example 2 (illustrating the formula (18))

Once again, we are analyzing the regular correlated pair ($\mathbf{R}(3)$, $\mathbf{R}_0(3)$), described by the formula (20). The coefficient $r_{\mathbf{R}}^2(3)$ of this model equals 0,1339 (see: (22)).

The right hand side of the formula (18) equals:

$$\frac{\mathbf{R}_0^{\mathrm{T}}(3)\mathbf{R}_0(3)}{\max S_i} = \frac{\begin{bmatrix} 0{,}1 & 0{,}2 & 0{,}3 \end{bmatrix}\begin{bmatrix} 0{,}1 \\ 0{,}2 \\ 0{,}3 \end{bmatrix}}{1+|0{,}02|+|0{,}04|} \approx 0{,}1320 \tag{23}$$

therefore the inequality (18) is satisfied.

Literature

Hauke J., Pomianowski J., *Związki korelacyjne i kryterium nieujemnej nieokreśloności macierzy*, „Przegląd Statystyczny" 1984, no. 3,

Hellwig Z., *Przechodniość relacji skorelowania zmiennych losowych i płynące stąd wnioski ekonometryczne,* „Przegląd Statystyczny" 1976, no. 1,

Kolupa M., *Dowód hipotezy Z. Hellwiga,* „Przegląd Statystyczny" 1993, no. 2,

Kolupa M., *Dowód pewnego twierdzenia Z. Hellwiga i pewne własności macierzy uniwersalnej,* „Przegląd Statystyczny" 1977, no. 3,

Kolupa M., *Macierze brzegowe w badaniach ekonometrycznych,* PWE, Warszawa 1982.

Kolupa M., Plebaniak J., *Konsekwencje majoryzowania macierzy korelacji przez macierz uniwersalną na tle nierówności Hellwiga,* „Przegląd Statystyczny" 2011, no. 3-4.

O JAKOŚCI REGULARNYCH PAR KORELACYJNYCH $(\mathbf{R}(K), \mathbf{R}_0(K))$ MIERZONEJ WARTOŚCIĄ WSPÓŁCZYNNIKA $\mathbf{R}^2(K)$

Streszczenie: Jeżeli para korelacyjna $(\mathbf{R}(k), \mathbf{R}_0(k))$ jest regularną parą korelacyjną, to współczynnik mierzący jakość takiej pary spełnia nierówność: $r^2(k) \geq F(k)$. Jeżeli macierz korelacji $\mathbf{R}(k)$ jest majoryzowana przez macierz uniwersalną, to:

$$F(k) = \frac{\mathbf{R}_0^{\mathrm{T}}(k)\mathbf{R}_0(k)}{1 + \mathbf{R}_0^{\mathrm{T}}(-1)\mathbf{R}_0(-1)},$$

gdzie $\mathbf{R}_0^{\mathrm{T}}(-1) = \begin{bmatrix} r_2 & r_3 & \ldots & r_k \end{bmatrix}$. Jeżeli regularna para korelacyjna $(\mathbf{R}(k), \mathbf{R}_0(k))$ jest wyznaczona metodą maksymalnej pojemności integracyjnej, to:

$$F(k) = \frac{\mathbf{R}_0^{\mathrm{T}}(k)\mathbf{R}_0(k)}{\max S_i},$$

gdzie S_i oznacza sumę modułów elementów położonych w i-tej kolumnie macierzy $\mathbf{R}(k)$. W każdym z tych przypadków wyznaczenie wielkości $F(k)$ jest łatwe pod względem rachunkowym.

Słowa kluczowe: macierz korelacji, macierz uniwersalna, nierówność Hellwiga, współczynnik $r^2(k)$.