ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 90, 1989

*Władysław Milo**

REGRESSION DIAGNOSTIC OF ILL-CONDITIONED STATISTICAL DATA

1. INTRODUCTION

B e l s l e y, K u h, W e l s c h [2] have presented a survey of methods of detecting and assesing the degree of collinearity and its effects. They have been examining effects manifested in the behaviour of least square estimator's sample values $b = \mathfrak{X}^{-1}x'y$, $\mathfrak{X} = x'x$, $x \in R^{n\times k}$, $b \in R^k$, $y \in R^n$, where $R^{n\times k}$ is the vector-space of $n \times k$ real matrices, R^l is the vector-space of l-tuples of real numbers over R. The term "collinearity" has its origin in geometry and in fact is most often used for two vectors (points) only. Two vectors $x_{.1} = (x_{11}, x_{21})'$, $x_{.2} = (x_{21}, x_{22})'$ are said to be collinear if they lie on the same straight line $\mathcal{L}$ (so $x_{.1}$ and x_2 are co-line points). The collinear points $x_{.1}$, $x_{.2}$ need not point in the same direction. Thus if $x_{.1}$ is collinear with $x_{.2}$, then $(-x_{.1})$ is also collinear with $x_{.2}$. In the case of $x_{.1}, x_{.2} \in \mathcal{L} \subset R^2$ the term collinearity is well placed in our geometric intuition. Even in this simple case it has its algebraic counterpart, i.e. two vectors $x_{.1}, x_{.2} \in R^2$ are said to be collinear iff there exist two reals $\beta_1, \beta_2 \in R$ such that $\beta_1 x_{.1} + \beta_2 x_{.2} = 0$ under $\beta_1^2 + \beta_2^2 > 0$ or if $x_{.1}$ can be expressed in terms of $x_{.2}$ in the form $x_{.1} = -\frac{\beta_2}{\beta_1} x_{.2}$ under $\beta_1 \neq 0$. It comes from the state-

* Lecturer at the Institute of Econometrics and Statistics of the University of Łódź.

ment that $x_{.1}$ is collinear with $x_{.2}$ iff $x\beta = 0$ where $x \equiv (x_{.1}, x_{.2}) \in R^{2x2}$, $\beta = (\beta_1, \beta_2)' \in R^{2x1} \equiv R^2$, i.e. iff $x_{1.} \perp \beta$, $x_{2.} \perp \beta$, where $x'_{1.} = (x_{11}\, x_{12})$, $x'_{2.} = (x_{21}\, x_{22})$. The equation $x\beta = 0$ can also be written as $\sum_{i=1}^{2} \beta_i x_{.i} = 0$ under $\beta_1^2 + \beta_2^2 > 0$. It defines linear dependence of vectors $x_{.1}, x_{.2}$. In R^n a set $\{x_{.j} \in R^n,\ j = \overline{1, k}\}$, $x_{.j} = (x_{1j}, \ldots, x_{nj})'$ is linearly dependent if there exist scalars β_j, $j = \overline{1, k}$ not all zero such that $\sum_{j=1}^{k} \beta_j x_{.j} = 0$, i.e. the null vector is a linear combination of the vectors $x_{.j}$, $j = \overline{1, k}$. In general a n-dimensional vector y is said to be lineary dependent on a set $\{x_{.j},\ j = \overline{1, k}\}$ of n-dimensional vectors $x_{.1}, \ldots, x_{.k}$ if $y = \sum_{1}^{k} \beta_j x_{.j}$. A set $\{x_{.j},\ j = \overline{1, k}\}$ of distinct n-dimensional vectors lie on the same straight line if there exist k non-zero coefficients $\beta_1, \ldots, \beta_k$ such that $\sum_{j=1}^{k} \beta_j x_{.j} = 0$ and $\sum_{j=1}^{k} \beta_j = 0$. One can also look at the equation $\sum_{j=1}^{k} \beta_j x_{.j} = x\beta = 0$ from another point of view. Denoting $x'_{i.} = (x_{i1}, \ldots, x_{ik})$, $i = \overline{1, n}$ we can write $x\beta = 0$ as

$$x\beta = \begin{pmatrix} x'_{1.}\ \beta \\ \ldots\ldots \\ x'_{n.}\ \beta \end{pmatrix} = \begin{pmatrix} 0 \\ \ldots\ldots \\ 0 \end{pmatrix},$$

where each i-th scalar equation $x'_{i.}\beta = 0$, $i = \overline{1, k}$, defines an i-th hyperplane. These equations also tell us that $x_{i.} \perp \beta$, $i = \overline{1, k}$. However, if $x_{i.} \perp \beta$, $i = \overline{1, k}$, where β is the vector of coefficients of linear combination of vectors $x_{.j} \in \{x_{.j}\}_{j=1}^{k}$ that defines the vector $0 = \sum \beta_j x_{.j}$, then in the case of equation $Y = x\,\beta + \Xi$ we obtain $Y = \Xi$.

So for those values of β that are the values of coefficients of linear dependence between $x_{.1}, \ldots, x_{.k}$ we obtain null-explanation (zero effect in equation) of Y in $Y = x\,\beta + \Xi$

by factors $x_{.1}, \ldots, x_{.k}$. This means that β for which $x\beta = 0$ or, putting it in other words, $x_{i.} \perp \beta$ for $i = \overline{1, k}$ is a pathological point of parameter space from the model specification point of view. Therefore, if the sample value of b is approximately orthogonal to each row $x'_{i.}$ of the matrix x, then the point b is suspected to be a pathological point with the null explanation degree. The point $\beta \in R^k$; $x\beta = 0$ can always be regained from the matrix $(x = x_{.1}, \ldots, x_{.k})$.

One can distinguish two ways of regaining $\beta : x\beta = 0$ from x. The first way is based on the definition of singular value $\lambda_i^{1/2}$ if matrix x, i.e.

$$x\beta^{(i)} = \lambda_i^{1/2}$$

where $\beta = \beta^{(i)}$ is the singular (eigen) vector attached to the singular value $\lambda_i^{1/2}$. It is known that for $\lambda_i^{1/2} = 0$ we have $\lambda_i^{1/2}\, \beta^{(i)} = 0 = x\, \beta^{(i)} \equiv x\, \beta$. So the components of eigen-vector of the matrix x (or $x'x$) are the coefficients of linear combination defining the null vector.

The second way (see par. 2) of regaining is based on a generalisation of angle (cosine) definition of dependence between two vectors.

Both ways (see par. 2) are enabling diagnostics of existence and degree of bad conditioning of the data x and (y, x). In this place we should explain what we mean by the term "bad-conditioning of x" in the context of linear models. We assume that the data $y \in R^n$ are generated according to the following linear model

$$NMgxH_o^{bc} \equiv (R^{nxk},\ \mathcal{S},\ \mathcal{S}_y,\ Y = x\,\beta + \Xi,\ k_o \leqslant k,\ \nu_x \geqslant \nu_x^*,\ n_o = n,$$

$$P_y = N_y\,(x\,\beta,\ \sigma^2 I)),$$

where

$\mathcal{S} = (U, F, P)$, $P(U) = 1$ is a complete probability space with a space U of elementary events, a borel σ-field F of subsets of U, a probability measure P;

$\mathcal{S}_y = (R^n, F_{R^n}, P_y)$, $P_y(R^n) = 1$, F_{R^n} is a borel σ-field of subsets of R^n and $\mathcal{B}y$ is a complete probability space induced by $Y : \mathcal{B} \to \mathcal{B}_Y$;

k_o = rank x, n = rank ($DY \equiv \mathcal{E}(Y - \mathcal{E}Y)(Y - \mathcal{E}Y)'$, $\mathcal{E}$ is an expectation-value operator;

$P_Y = N_Y(x\beta, \sigma^2 I)$ is the gaussian measure of probability of random vector Y with moments $\mathcal{E}Y = x\beta$, $DY = \sigma^2 I$;

$\nu_x = \frac{\lambda_{max}^{1/2}(x)}{\lambda_{min}^{1/2}(x)} = \frac{\lambda_k^{1/2}(x)}{\lambda_1^{1/2}(x)}$ is an index of bad conditioning of matrix x equal to the ratio of the largest singular value $\lambda_k^{1/2}$ of x to the smallest singular value $\lambda_1^{1/2}$ of x, ν_x^* is a threshold value of ν_x enabling the distinction of bad and well conditioned matrices x.

For semiorthonormal matrices $\nu_x = 1$. For matrices with $k_o < k$, $\nu_x = \infty$. These two situations are extreme ones concerning extremely well and extremely badconditioned matrices. For usual cases $\nu_x \in]1, \infty[$. However, the author does not know how to fix the value ν_x^* in a proper way. One should derive this value from a given optimisation criteria for estimators or predictors or tests. Problems of unique determining of ν_x^* with respect to a given criteria of statistics evaluations are open. The presented above ideas of detecting and measuring bad-conditioned data from the point of view of estimation and prediction theory have been treated in the works of B e l s l e y, K u h, W e l s c h [2], G u n s t [5], M i l o [7], K e n d a l l [6]; S i l v e y [10].

2. DETECTION OF LINEAR DEPENDENCIES IN DATA

Extremely bad conditioned data x will be called further collinear data or multicollinear data (or data with linearly dependent columns (vectors) $\{x_{.j}\}_{j=1}^{k}$ of x).

The main goal of this section is the studying of the linear dependencies between the vectors from $\left\{x_{.j},\ j = \overline{1, k}\right\}$ in the sense of linear algebra or in the sense of multidimensional geometry and detection of the number and structure of these linear dependencies.

2.1. Detection Based on the Definition of Singular Values and Vectors

We recall that

$$xv_{\cdot j} = \lambda_j^{1/2} u_{\cdot j}, \qquad x = \sum_{j=1}^{k_o} \lambda_j^{1/2} u_{\cdot j} v'_{\cdot j},$$

where $\lambda_j^{1/2}$ is a singular value of x, λ_j is the j-th eigen-value of $x'x$ (or xx'), $u_{\cdot j}$ is the j-th eigen-vector of xx' connected with λ_j. For $\lambda_j^{1/2} = 0$ we have $\sum_{i=1}^{k} v_{ij} x_{\cdot i} = 0$. This equation displays the linear dependence between vectors from $\{x_{i'}\}$ $i = \overline{1, k}$. The vectors involved in this dependence should be displayed by the nonzero elements of $v_{\cdot j}$. Suppose that singular values have been increasingly ordered as $\lambda_1^{1/2} \ldots \lambda_k^{1/2}$. If $\lambda_1^{1/2} = \lambda_2^{1/2} = 0$, then there are two linear combinations of vectors (two linear dependencies) connected with two zero singular values.

The indices of nonzero elements of vectors corresponding to zero eigen-values should display the indices of vectors $\{x_{\cdot j}\}$ that are involved in these linear dependencies.

The following examples are instructive in this respect.

E x a m p l e 1.

$$x = \begin{pmatrix} 1 & 1 & 3 & 2 & 1 & 1 \\ 0 & 2 & 2 & 3 & 3 & 0 \\ 0 & 3 & 3 & -1 & 4 & -5 \\ 0 & 4 & 4 & 0 & 1 & -1 \\ 0 & 5 & 5 & 0 & 0 & 0 \\ 0 & 6 & 6 & 0 & 0 & 0 \end{pmatrix}; \quad x_{\cdot 3} = 2x_{\cdot 1} + x_{\cdot 2}, \quad x_{\cdot 6} = x_{\cdot 4} - x_{\cdot 5}$$

Using Jacobi method or singular value decomposition method one can find eigen values $\lambda = (\lambda^{1/2})^2$ equal squares of singular values,

$\lambda_1 = 0,\ \lambda_2 = 1.978,\ \lambda_3 = 199.9,\ \lambda_4 = 1.289,\ \lambda_5 = 18.72$

$\lambda_6 = 37.11.$

So singular values for λ_2 and λ_4 are equal

$$\lambda_1^{1/2} = 0, \quad \lambda_2^{1/2} = 1.006, \quad \lambda_4^{1/2} = 1.13.$$

Eigen vectors corresponding to these singular values are

$$v_{.1} = (0.816, \ 0.408, \ -0.408, \ 0, \ 0, \ 0)'$$

$$v_{.2} = (-0.556, \ 0.564, \ -0.547, \ 0.272, \ -0.006, \ -0.003)'$$

$$v_{.4} = (0.089, \ -0.092, \ 0.085, \ 0.524, \ -0.578, \ -0.607)'.$$

By knowledge of ways of constructing x we knew a priori that there are two linear dependencies. They define $x_{.3}$ and $x_{.6}$ so rank x = 6 - 2 = 4. There is no clear cut confirmation that there are two zero singular values ($\lambda_1^{1/2}$ is obvious case, but $\lambda_2^{1/2}$ and $\lambda_4^{1/2}$ are almost equal and different from zero). The conclusions that follow from λ_1 and $v_{.1}$ are confirming underlying scheme because $0.8x_{.1} + 0.4x_{.2} - 0.4x_{.3} + 0x_{.4} + 0x_{.5} + 0x_{.6} = 0 \Longleftrightarrow 2x_{.1} + x_{.2} = x_{.3}$. From the entities of vector $x_{.4}$ it is seen that more probable is a diagnostic based on $v_{.4}$ i.e. the first three coordinates are approximately zero, but we can do it on the ground of a priori knowledge of $x_{.6} = x_{.4} - x_{.5}$. Lack of this knowledge makes this step very undecisive. Even with the a priori information we can only say that the underlying scheme defining $x_{.6}$ is fulfilled approximately ($5.2x_{.4} - 5.8x_{.5} \cong 6.1x_{.6}$).

E x a m p l e 2.

$$x = \begin{pmatrix} 1 & 1 & 1 & 1 & 3 & 2 \\ 1 & 2 & 0 & -1 & 2 & 0 \\ 1 & 3 & -1 & 2 & 1 & 1 \\ 1 & 4 & -2 & -2 & 0 & 0 \\ 1 & 5 & -3 & 3 & -1 & 3 \\ 1 & 6 & -4 & -3 & -2 & 0 \end{pmatrix}; \qquad \begin{aligned} x_{.3} &= 2x_{.1} - x_{.2} \\ x_{.4} &= x_{.2} + 2x_{.3}. \end{aligned}$$

$$\lambda_1^{1/2} = 128.6, \quad \lambda_2^{1/2} = 40.01, \quad \lambda_3^{1/2} = 19.05, \quad \lambda_4^{1/2} = 0,$$

$$\lambda_5^{1/2} = 1.33, \quad \lambda_6^{1/2} = 0;$$

$$v_{.4} = (0.943, \ -0.236, \ 0, \ 0, \ -0{,}236, \ 0)';$$

$$v_{.6} = (0, \ 0.41, \ 0.82, \ 0, \ -0{,}41, \ 0)';$$

The results obtained in Example 2 are correct as far as singular values are concerned (there are two zero values as it should be). However, the corresponding eigen-vectors $v_{.4}$ and $v_{.6}$ for $\lambda_4^{1/2}$ and $\lambda_6^{1/2}$ are introducing doubts whether in the case of matrices with two or more linear dependencies between some columns there is a possibility to detect the appropriate columns entering the original linear dependencies underlying the scheme of generation for x. For instance the nonzero elements of $v_{.4}$ (i.e. v_{14}, v_{24}, v_{54}) are only partially correct. Fully correct vector should have the elements v_{14}, v_{24} and v_{34} different from zero. Similarly uncorrect results have been obtained in the case of $v_{.6}$.

Is the malfunction of code of programming responsible for these results? It needs further exploration. However, similar results have been obtained for other examples with two linear dependencies. In the case of one linear dependency the identification of original vectors entering the original linear dependence was correct. However, this diagnostic of number and the structure of linear dependencies can be spoiled by not extreme though severe bad-conditioning in x (or x'x). For small range vectors almost parallel between themselves it is likely that we will suffer in regaining the original schemes of linear dependencies between columns.

2.2. Detection Based on an Extension of an Angle Measure of Dependence Between Two Vectors

From analytic geometry and linear algebra we know that the following scalar function

$$c_{ij} \equiv \cos \measuredangle (x_{.i}, x_{.j}) \equiv \frac{< x_{.i}, x_{.j} >}{\|x_{.i}\| \ \|x_{.j}\|}, \quad i, j = \overline{1, k}, \qquad (1)$$

$$< x_{.i}, x_{.j} > = \sum_{l=1}^{n} x_{li} x_{lj}, \quad \|x_{.j}\| = \sqrt{\sum_{i=1}^{n} x_{ij}^2},$$

defines cosine of angle between two given vectors $x_{.i}$, $x_{.j}$. It is obvious that

$c_{ij} \in [-1, 1]$;

$x_{.i} \perp x_{.j} \iff c_{ij} = 0$;

$x_{.i} \parallel x_{.j} \iff c_{ij} = 1$ (or $c_{ij} = -1$).

In angular terms $x_{.i} \perp x_{.j} \iff \alpha = l \frac{\pi^{o}}{2}, l = \pm 1, \pm 3,$

$x_{.i} \parallel x_{.j} \iff \alpha = l \pi^{o}, l = 0, \pm 1, \pm 2.$

For applied econometrician or statistician the measure c_{ij} is also attractive when there are linear relationships encompassing only two vectors in each relation.

Now we extend this approach to the case when there are linear dependencies (possibly more than one for given x) encompassing in each relationship more than two vectors of matrix X.

As before let $X = (x_{.1}, \ldots, x_{.k})$ be an n x k matrix or a linearly dependent system of vectors with the vector $x_{.j}$ defined as $x_{.j} = \sum_{l=1}^{k} \alpha_l x_{.l}$. We remember that $c_{ij} = \cos(x_{.i}, x_{.j}) = \frac{\langle x_{.i}, x_{.j} \rangle}{\|x_{.i}\| \, \|x_{.j}\|}, \quad i \neq j = \overline{1, k}$.

By definition of $x_{.j}$ we have

$$c_{ij} = \frac{\langle x_{.i}, x_{.j} \rangle}{\|x_{.i}\| \, \|x_{.j}\|} = \frac{\langle x_{.i}, \sum_{l=1}^{k} \alpha_l x_{.l} \rangle}{\|x_{.i}\| \, \|x_{.j}\|} \tag{2}$$

and hence

$$c_{ij} \|x_{.j}\| = \frac{\langle x_{.i}, \alpha_1 x_{.1} + \ldots + \alpha_{j-1} x_{.j-1}, \alpha_{j+1} x_{.j+1} \ldots \alpha_k x_{.k} \rangle}{\|x_{.i}\|} \tag{3}$$

Introducing the following notation and formulae

$$d_i^{(j)} = c_{ij} \; \| x_{.j} \| , \quad d_{i,l}^{(j)} = c_{il} \; \| x_{.l} \| ,$$

$$l \neq j, \quad l = \overline{1, \ldots, j-1, \; j+1, \; \ldots, \; k} ;$$

$$c_{il} = \frac{\langle x_{.i} \cdot x_{.l} \rangle}{\| x_{.i} \| \, \| x_{.l} \|}, \quad l \neq j, \quad l = \overline{1, \ldots, j-1, j+1, \ldots, k},$$

one can transform (3) to

$$d_i^{(j)} = \alpha_1^{(j)} d_{i1}^{(j)} + \ldots + \alpha_{j-1}^{(j)} d_{i,j-1}^{(j)} + \alpha_{j+1}^{(j)} d_{i,j+1}^{(j)} + \ldots +$$

$$+ \alpha_k^{(j)} d_{i,k}^{(j)} \tag{4}$$

Putting

$$d^{(j)} = \begin{pmatrix} d_1^{(j)} \\ \cdot \\ \cdot \\ \cdot \\ d_{j-1}^{(j)} \\ d_{j+1}^{(j)} \\ \cdot \\ \cdot \\ \cdot \\ d_k^{(j)} \end{pmatrix}, \quad \alpha^{(j)} = \begin{pmatrix} \alpha_1^{(j)} \\ \cdot \\ \cdot \\ \cdot \\ \alpha_{j-1}^{(j)} \\ \alpha_{j+1}^{(j)} \\ \cdot \\ \cdot \\ \cdot \\ \alpha_k^{(j)} \end{pmatrix}$$

$$D^{(j)} = \begin{pmatrix} d_{11}^{(j)} & \ldots & d_{1,j-1}^{(j)}, & d_{1,j+1}^{(j)} & \ldots & d_{1,k}^{(j)} \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ d_{j-1,1}^{(j)} & \ldots & d_{j-1,j-1}^{(j)}, & d_{j-1,j+1}^{(j)} & \ldots & d_{j-1,k}^{(j)} \\ d_{j+1,1}^{(j)} & \ldots & d_{j+1,j-1}^{(j)} & d_{j+1,j+1}^{(j)} & \ldots & d_{j+1,k}^{(j)} \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ d_{k,1}^{(j)} & & d_{k,j-1}^{(j)} & d_{k,j+1}^{(j)} & & d_{k,k}^{(j)} \end{pmatrix} \tag{5}$$

we can write the system of (k-1) equations in (k-1) unknown $\alpha^{(j)}$ for given $x_{.j} = \sum_{l=1}^{k} \alpha_l x_l$, $l \neq j$, i.e.

$$D^{(j)}\alpha^{(j)} = d^{(j)}, \quad j = \overline{1, k} \tag{6}$$

This linear system has solution

$$\alpha^{(j)} = (D^{(j)})^{-1} d^{(j)} \tag{7}$$

if rank $D^{(j)} = k - 1$, $j = \overline{1, k}$.

For numerical efficiency reasons, if k is fixed (so one does not need to study different model specification), then it is sufficient to consider $j = k$.

So it suffices to find

$$\alpha^{(k)} = (D^{/k/})^{-1} d^{(k)} \tag{7a}$$

From $\alpha^{(k)}$ we know that the index k (for which the linear combination $x_{.k} = \sum_{l=1}^{k} \alpha_l x_{.l}$, $k \neq l$ can be regained) is the index of $x_{.k}$ for which the linear combination was constructed.

It can be seen from the following examples

E x a m p l e 3.

$$x = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 3 & 2 & 1 \end{pmatrix}, \quad \lambda_1^{1/2} = 0, \quad \lambda_2^{1/2} = \sqrt{2.4}, \quad \lambda_3^{1/2} = \sqrt{41.6};$$

$$x_{.3} = 2^{-1} x_{.2}, \quad v'_{.1} = (0, 1, -2).$$

Hence $0x_{.1} + 1x_{.2} - 2x_{.3} = 0 \Longrightarrow x_{.2} = 2x_{.3}$ or $x_{.3} = 2^{-1}x_{.2}$.

So zero-eigen-value and corresponding eigen-vector approach, if there is no bad conditioning in the data x, enables us to identify existing equivalent linear combination of vectors defining given vector $x_{.j}$ (i.e. $x_{.2} = 2x_{.3}$ or equivalently $x_{.3} = 2^{-1}x_{.2}$).

The minus sign of the third component of $v_{.1}$ indicates that $x_{.3}$ is linearly related to $x_{.2}$.

Using our angle approach we obtain

$$C = \begin{pmatrix} 1 & 0.872871 & 0.872876 \\ 0.872871 & 1 & 1 \\ 0.872876 & 1 & 1 \end{pmatrix} = \begin{pmatrix} \cos 0^{\circ} & \cos 29^{\circ} & \cos 29^{\circ} \\ \cos 29^{\circ} & \cos 0^{\circ} & \cos 0^{\circ} \\ \cos 29^{\circ} & \cos 0^{\circ} & \cos 0^{\circ} \end{pmatrix},$$

$$\arg C = \begin{pmatrix} 0^{\circ} & 29^{\circ} & 29^{\circ} \\ 29^{\circ} & 0^{\circ} & 0^{\circ} \\ 29^{\circ} & 0^{\circ} & 0^{\circ} \end{pmatrix},$$

By using (7), C, x we can calculate

$$\alpha^{(3)} = \frac{1}{4.3650} \begin{pmatrix} 4.8989 & -4.2763 \\ -3.2661 & 3.7417 \end{pmatrix} \begin{pmatrix} 2.1382 \\ 2.4495 \end{pmatrix} = \begin{pmatrix} 0 \\ 2^{-1} \end{pmatrix}$$

or

$$\alpha^{(2)} = \frac{1}{2.1825} \begin{pmatrix} 2.4495 & -2.1381 \\ -3.2659 & 3.7417 \end{pmatrix} \begin{pmatrix} 4.2762 \\ 4.8989 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}.$$

Hence, due to $\alpha^{(3)}$ we can write down

$$x_{\cdot 3} = 0x_{\cdot 1} + 2^{-1}x_{\cdot 2} = 2^{-1}x_{\cdot 2}$$

or due to $\alpha^{(2)}$

$$x_{\cdot 2} = 0x_{\cdot 1} + 2x_{\cdot 3} = 2x_{\cdot 3}.$$

By definition $x_{\cdot j} = \sum_{\substack{l=1 \\ l \neq j}}^{k} \alpha_1 x_{\cdot 1}$ we have $\alpha_1 x_{\cdot 1} + \ldots + \alpha_{j-1} x_{\cdot j-1} + - 1x_{\cdot j} + \alpha_{j+1} x_{\cdot, j+1} + \ldots + \alpha_k x_{\cdot k} = 0$ and hence the extended (full) form of $\alpha_f^{(j)}$ calculated according to $\alpha^{(j)}$ from (7) is equal (in the last example) $\alpha_f^{(3)} = (0,\ 2^{-1},\ -1)'$ or according to $\alpha^{(2)}$, $\alpha_f^{(2)} = (0,\ -1,\ 2)'$.

It is seen that both singular values and vectors approach and extended angle approach enable us to detect linearly dependent columns (vectors) from x.

E x a m p l e 4.

$$x = \begin{pmatrix} 2 & 1 & 1 \\ 4 & 1 & 3 \\ 2 & 1 & 1 \\ 3 & 1 & 2 \\ 1 & 0 & 1 \end{pmatrix} \quad \begin{array}{l} x_{\cdot 3} = x_{\cdot 1} - x_{\cdot 2} \\ \lambda_1^{1/2} = \sqrt{53.15}, \quad \lambda_2^{1/2} = 0, \quad \lambda_3^{1/2} = \sqrt{0.847} \\ v_{\cdot 2} = (-0.577, \ 0.577, \ 0.577)'. \end{array}$$

According to singular vectors approach we have $x_{\cdot 1} = x_{\cdot 2} + x_{\cdot 3}$ or equivalently $x_{\cdot 3} = x_{\cdot 1} - x_{\cdot 2}$. According to our angle approach we have

$$C = \begin{pmatrix} 1 & 0.9432 & 0.9861 \\ 0.9432 & 1 & 0.8750 \\ 0.9861 & 0.8750 & 1 \end{pmatrix}, \quad \arg C = \begin{pmatrix} 0^{\circ} & 19^{\circ}18' & 9^{\circ}32 \\ 19^{\circ}18' & 0^{\circ} & 28^{\circ}58' \\ 9^{\circ}32' & 28^{\circ}58' & 0^{\circ} \end{pmatrix}.$$

Using (7) we have

$$\alpha^{(3)} = \begin{pmatrix} 5.831 & 1.886 \\ 5.499 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 3.95 \\ 3.50 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

and hence

$$x_{\cdot 3} = x_{\cdot 1} - x_{\cdot 2}$$

So the results of detections made by singular vectors approach and our approach are correct.

Due to simplicity of calculations within our angle method (it needs only stable inversion method used in finding the inverse matrix $(D^{(j)})^{-1}$) and very imaginative angle evaluation of degree of dependency on the ground of C and argC, our method seems to be more preferable in practice with respect to singular vectors method (where numerical complexity of calculations is greater).

It is obvious that deeper formal and numerical studies of relative efficiency of these methods are needed, i.e.

a) for what special structure matrices x the relative numerical efficiency of both methods are the same,

b) whether $k = 3$ is enough for the relative numerical efficiency of angle method to be greater (if x is not a special matrix).

2.3. Detection of Two Linear Combinations of Vectors in the System $x = (x_{.1}, \dots, x_{.k})$

In the situation of two linear combinations in x we have

$$x_{.j} = \sum_{l=1}^{j-1} \alpha_l x_{.l}, \quad x_{.k} = \sum_{\nu=j+1}^{k-1} \alpha_\nu x_{.\nu}$$

From the definitions of $x_{.j}$ and c_{ij} we have

$$d_i^{(j)} = \|x_{.j}\| c_{ij} = \alpha_1^{(j)} c_{i1} \|x_{.1}\| + \alpha_2^{(j)} c_{i2} \|x_{.2}\| + \dots +$$

$$+ \alpha_{j-1}^{(j)} \|x_{.,j-1}\| c_{i,j-1}, \quad i = \overline{1, j-1}$$

and similarly for $x_{.k}$, c_{ik} we have

$$d_i^{(k)} = \|x_{.k}\| c_{ik} = \alpha_{j+1}^{(k)} c_{i,j+1} \|x_{.,j+1}\| + \dots +$$

$$+ \alpha_{k-1}^{j} c_{i,k-1} \|x_{.,k-1}\|, \quad i = \overline{j+1, k-1}$$

Denoting

$$d^{(j)} = (d_1^{(j)}, d_2^{(j)}, \dots, d_{j-1}^{(j)})', \quad (d^{(k)} = d_{j+1}^{(k)}, \dots, d_{k-1}^{(k)})',$$

$$D^{(j)} = \begin{pmatrix} d_{11}^{(j)} & \dots & d_{1,j-1}^{(j)} \\ \vdots & & \vdots \\ d_{j-1,1}^{(j)} & \dots & d_{j-1,j-1}^{(j)} \end{pmatrix} \quad D^{(k)} = \begin{pmatrix} d_{j+1,j+1}^{(k)} & \dots & d_{j+1,k-1}^{(k)} \\ \vdots & & \vdots \\ d_{k-1,j+1}^{(k)} & \dots & d_{k-1,k-1}^{(k)} \end{pmatrix}$$

where $d_{il}^{(j)} = c_{il} \|x_{.l}\|$, $i,l = \overline{1, j-1}$, $l = \overline{j+1, k-1}$, we can write

$$D^{(j)} \alpha^{(j)} = d^{(j)}$$

$$D^{(k)} \alpha^{(k)} = d^{(k)} \quad (8)$$

If det $D^{(j)} \neq 0$ and det $D^{(k)} \neq 0$, then

$$\alpha^{(j)} = \left(D^{(j)}\right)^{-1} d^{(j)}$$
$$\alpha^{(k)} = \left(D^{(k)}\right)^{-1} d^{(k)} \tag{8a}$$

where $\alpha^{(j)}$ and $\alpha^{(k)}$ are the vectors of coefficients of linear combinations defining $x_{.j}$ and $x_{.k}$.

It is obvious how to extend (8a) on more than two linear combinations in x (if k is very large).

From (8a) it follows that if $\alpha^{(j)}$, $\alpha^{(k)} \neq 0$, then there are two linearly dependent relationships (combinations) with coefficients given by the components of $\alpha^{(j)}$, $\alpha^{(k)}$. If only $\alpha^{(j)}$ (or $\alpha^{(k)}$) is different from the zero vector then there exists only one linear combination.

E x a m p l e 4.

$$x = \begin{pmatrix} 2 & 1 & 1 & 2 & 4 \\ 0 & 1 & -1 & -3 & -1 \\ 3 & 1 & 2 & 6 & 8 \\ -1 & 1 & -2 & 4 & 6 \\ 4 & 1 & 3 & 0 & 2 \\ -2 & 1 & -3 & 5 & 7 \end{pmatrix}, \qquad \begin{matrix} x_{.1} = x_{.2} + x_{.3} \\ x_{.5} = 2x_{.2} + x_{.4} \end{matrix}$$

$$\lambda_1^{1/2} = \sqrt{59.27}, \quad \lambda_2^{1/2} = \sqrt{8.55}, \quad \lambda_3^{1/2} = 0, \quad \lambda_4^{1/2} = 0,$$

$$\lambda_5^{1/2} = \sqrt{260.18}$$

$$v_{.3} = (-0.620,\ 0.468,\ 0.620,\ -0.076,\ +0.076)'$$

$$v_{.4} = (0.209,\ 0.703,\ -0.209,\ 0.457,\ -0.457)'.$$

It is obvious that in this case detection of first linear combination is easy. But the second, although still is seen, is only seen relatively.

The matrix C is given in lower diagonal form as

$$C = \begin{pmatrix} 1 & & & & \\ 0.4201 & 1 & & & \\ 0.9075 & 0 & 1 & & \\ 0.1446 & 0.6025 & -0.1195 & 1 & \\ 0.2631 & 0.8141 & -0.0869 & 0.9539 & 1 \end{pmatrix}.$$

and

$$\arg C = \begin{pmatrix} 0^{o} & & & & \\ 65^{o}16' & 0^{o} & & & \\ 24^{o}84' & 90^{o} & 0^{o} & & \\ 81^{o}69' & 52^{o}95' & 96^{o}86' & 0^{o} & \\ 74^{o}75' & 35^{o}50' & 94^{o}98' & 17^{o}46' & 0^{o} \end{pmatrix}$$

We leave to the readers further simple calculations connected with finding solutions according to (8a).

3. SENSITIVITY OF SOME FUNCTIONS TO LEVELS OF BAD CONDITIONING

For running linear regression empirical models as some generations of given statistical data (y, x) it is good to know what is the influence of bad conditioning of data on some statistics (estimators, predictors, tests).

We limit out attention to the most popular linear models with the main stochastic equation

$$Y = x\beta + \Xi \tag{9}$$

and assumptions

$$x \in R^{nxk}, \quad \beta \in R^{k} \tag{9a}$$

$$P_{\Xi} = N_{\Xi}(0, \sigma^2 I) \tag{9b}$$

where

Ξ has normal distribution with the first moment $\mathcal{E}\Xi = 0$ and second $\mathcal{D}\Xi = \sigma^2 I$. Sample values of Y, Ξ are denoted further by y, ξ. It is interesting to measure local behaviour of some functions by the use of classical derivatives of these functions with respect to their arguments.

First let us propose some working quantities. These are

$$\frac{\partial \lambda_i}{\partial x_{rs}} = 2v_{si} x'_{r\cdot} V_{\cdot i} \tag{10}$$

where

λ_i is i-th eigen value of $x'x$,

$V_{\cdot i}$ is i-th eigen vector of $x'x$ corresponding to λ_i,

v_{si} is (s, i)-th element of $V = (V_{\cdot 1}, \dots, V_{\cdot k})$, $V'V = VV' = I_{(k)}$,

$x'_{r\cdot}$ is r-th row of x;

$$\nu \equiv \nu_{x'x} = \frac{\lambda_k}{\lambda_1}, \quad \lambda_1 = \lambda_{min}(x'x), \quad \lambda_k = \lambda_{max}(x'x) \tag{11}$$

$$\frac{\partial \nu}{\partial x_{rs}} = 2\lambda_1^{-2}(\lambda_1 v_{sk} x'_{r\cdot} V_{\cdot k} - \lambda_k v_{s1} x'_{r\cdot} V_{\cdot 1}) =$$

$$= 2\lambda_1^{-2}(\lambda_1 v_{sk} V'_{\cdot k} - \lambda_k v_{s1} V'_{\cdot 1})\, x_{r\cdot} \quad \text{or due to (11)} \tag{12}$$

$$= 2\lambda_1^{-1}(v_{sk} V'_{\cdot k} x_{r\cdot} - \nu\, v_{s1} V'_{\cdot 1} x_{r\cdot})$$

$$\frac{\partial \nu}{\partial x_{\cdot s}} = 2\lambda_1^{-1} x(v_{sk} V_{\cdot k} - \nu v_{s1} V_{\cdot 1}) \tag{13}$$

$$\frac{\partial \nu}{\partial x} = 2\lambda_1^{-1} x (V_{\cdot k} V'_{\cdot k} - \nu V_{\cdot 1} V'_{\cdot 1}) \tag{14}$$

The above formulae have been derived by the author for the purpose of lecture notes on "Time Series Analysis" during the years 1982-1984 at the University of Łódź.

They have been derived thanks to other indispensible formulae given by Graham [4], Dwyer [3], Neudecker [6], Balestra [1].

It is interesting to know that

$$\frac{\partial \lambda_i}{\partial x_{rs}} = 0 \iff v_{si} = 0 \quad \text{or} \quad x_{r\cdot} \perp V_{\cdot i} \quad \text{or}$$

$$(v_{si} = 0, x_{r\cdot} \perp V_{\cdot i}) \begin{cases} s, i = \overline{1, k} \\ r = \overline{1, n} \end{cases},$$

$$\frac{\partial\lambda_i}{\partial x_{.s}} = 0 \iff V_{si} = 0 \text{ or } (\forall r: x_{r.} \perp V_{.i}) \text{ or}$$

$(V_{si} = 0, \ \forall r : x_{r.} \perp v_{.i})$,

$s, i = \overline{1, k}$

$$\frac{\partial\lambda_i}{\partial x} = 0 \iff (\forall r: x_{r.} \perp V_{.i}) \text{ or } (V_{.i}V'_{.i} = 0),$$

$$x \neq 0 \begin{cases} r = \overline{1, n} \\ i = \overline{1, k} \end{cases};$$

$$\frac{\partial\nu}{\partial x} = 0 \iff V_{\cdot k} V'_{\cdot k} = \nu V_{.1} V'_{.1};$$

$$\frac{\partial\nu}{\partial x_{.s}} = 0 \iff \forall r: x_{r.} \perp (v_{sk}V_{\cdot k} - \nu v_{s1}V_{.1})) \text{ or}$$

$(V_{\cdot k} = \nu v_{sk}^{-1} v_{s1}V_{\cdot 1}$ if $k_o < k)$;

$$\frac{\partial\nu}{\partial x_{rs}} = 0 \iff (v_{sk} = v_{s1} = 0) \text{ or } \left(\nu = \frac{v_{sk}}{v_{s1}} \frac{x'_{r.}V_{\cdot k}}{x'_{r.}V_{.1}}\right) \text{ or}$$

$(x_{r.} \perp V_{\cdot k}, \ x_{r.} \perp V_{.1})$.

For regression diagnostics one of the most important results is that for each x, $\nu \in [1, \infty[$ we have

$$\frac{\partial E'E}{\partial\nu} = 0, \quad E = Y - xB, \quad B = (x'x)^{-1}x'Y \tag{15}$$

The relation (15) tells us that the behaviour of $E'E$, $e'e$, $(E'E)(n-k)^{-1}$, $e'e(n-k)^{-1}$ does not depend locally in the neighbourhood of ν on the degree (level) of bad conditioning (measured by $\nu = \lambda_k\lambda_1^{-1}$). The same is true for each particular λ_i. i.e.

$$\frac{\partial E'E}{\partial\lambda_i} = 0 \quad \forall i = \overline{1, k} \tag{16}$$

Another very important diagnostic measure is defined by

$$\frac{\partial MSEB}{\partial \lambda_i} = -\sigma^2 \lambda_i^{-2}, \quad i = \overline{1, k} \tag{17}$$

where MSEB is the mean square error of estimator B.

The following formula relates the local sensitivity of MSEB to the level ν of bad conditioning of $x'x$:

$$\frac{dMSEB}{d\nu} = \sigma^2 \lambda_k^{-1} - \sigma^2 \lambda_1^{-1} \nu^{-2} = \sigma^2(\lambda_k^{-1} - \lambda_1^{-1} \nu^{-2}) \tag{18}$$

The quantities $\frac{\partial MSEB}{\partial \lambda_i}$, $\frac{dMSEB}{d\nu}$ are attractive for the purposes of simulation studies of the influence of bad conditioning level on the precision of B. In practice we do not know σ^2 and the only knowledge we have is $S_E^2 = (n - k)^{-1} E'E$, i.e. an estimator of σ^2. Replacing σ^2 by this estimator we have sample estimates of above measures (see (17), (18)). These can be used as approximate characteristic of sensitivity of least squares estimator's precision to the local small changes of level ν of bad conditioning.

It is obvious that for ideally conditioned data (i.e. $\nu = 1$) we have $\frac{dMSEB}{d\nu} = 0$ and that for each $\sigma^2 > 0$ and $\nu > 1$, the function MSEB is increasing with respect to ν (i.e. $\frac{dMSEB}{d\nu} > 0$ always if $\nu > 1$, $\sigma^2 > 0$).

In the case of predictor $\hat{Y} = xB = x\, \mathfrak{X}^{-1} x'Y$, $\mathfrak{X} = x'x$, by using SVD ideas we have two cases.

1. The case of $\hat{Y} = x\, \mathfrak{X}^{-1} x'Y$, Y not treated as function of x,

$$\hat{Y} = \sum_{i=1}^{k} U_{\cdot i} U'_{\cdot i} Y \tag{19}$$

$U_{\cdot i}$ = the i-th column of U, i.e. the nxn matrix of eigen-vectors of $x\,x'$.

It means that

$$\frac{d\hat{Y}}{d\nu} = 0, \quad \forall \nu, x \tag{19a}$$

2. The case of $\hat{Y} = xB = x\beta + x\mathfrak{X}^{-1} x'E$, where $Y = x\beta + E$. In this case

$$\hat{Y} = \sum_{i=1}^{k} \lambda_i^{1/2} u_{\cdot i} v'_{\cdot i} \beta + \sum_{i=1}^{k} u_{\cdot i} u'_{\cdot i} \Xi \quad (20)$$

$$\hat{Y} = \lambda_k^{1/2} \nu^{-1/2} u_{\cdot 1} v'_{\cdot 1} \beta + \lambda_1^{1/2} \nu^{1/2} u_{\cdot k} v'_{\cdot k} \beta + \Sigma u_{\cdot i} u'_{\cdot i} \Xi$$

Hence

$$\frac{d\hat{Y}}{d\nu} = - \frac{1}{2} \nu^{-3/2} \lambda_k^{1/2} u_{\cdot 1} v'_{\cdot 1} \beta + \frac{1}{2} \nu^{-1/2} \lambda_1^{1/2} u_{\cdot k} v'_{\cdot k} \beta \quad (21)$$

It is obvious that $\frac{d\hat{Y}}{d\nu} = 0$ (the predictor Y is insensitive to small local changes in the neighbourhood of value ν) iff $v_{\cdot 1} \perp \beta$ and simultaneously $v_{\cdot k} \perp \beta$. The predictor $\hat{Y}$ is increasing componentwise function of ν iff $\frac{d\hat{Y}}{d\nu} > 0$, i.e. if $u_{\cdot k} > \nu^{-1/2} \cdot (v'_{\cdot 1} \beta)^{-1} (v'_{\cdot k} \beta) u_{\cdot 1}$.

4. FINAL REMARKS

Due to space limitations we have not presented here details of derivations for particular formulae.

They can be easily completed by readers from the assumed forms of functions defining Y, B, E, $E'E$, $\hat{Y}$, λ_i, ν, MSEB.

Our scope of sensitivity analysis is by no means complete. Outside of this scope we have left, among others, such topics as sensitivity of moments of l-s ridge estimators, predictors, test statistics with respect to λ_i, ν.

REFERENCES

[1] B a l e s t r a P. (1976), *La derivation matricielle*, Dijon IME, 115.

[2] B e l s l e y D., K u h E., W e l s c h E. (1980), *Regression Diagnostics*, Wiley, New York, 292.

[3] D w y e r P. (1967), *Some Applications of Matrix Derivatives*, J. Amer. Statist. Assoc., 607-625.

[4] G r a h a m A. (1981), *Cronecker Products*, Ellis Horwood Chichester, 130.

[5] G u n s t R. (1983), *Regression Analysis with Multicollinear Predictor Variables*, Commun. in Statist. Theory and Meth, 12, 2217-2260.

[6] K e n d a l l M. (1957), *A Course in Multivariate Analysis*, London Griffin.

[7] M i l o W. (1985), *Liniowe modele statystyczne*, University Press, Łódź, 364.

[8] N e u d e c k e r H. (1969), *Some Theorems on Matrix Differentiation*, J. Amer. Statist. Assoc., 953-693.

[9] P o l l o c k D. S. G. (1985), *Tensor Products and Matrix Differential Calculus*, Linear Algebra and its Appl., 169-194.

[10] S i l v e y S. (1969), *Multicollinearity and Imprecise Estimation*, J. Roy. Statist. Soc., Ser. B, 539-552.

Władysław Milo

REGRESYJNA DIAGNOSTYKA ŹLE UWARUNKOWANYCH DANYCH STATYSTYCZNYCH

Celem artykułu jest opis wykrywania liniowych zależności wg:

a) współrzędnych wektora własnego macierzy x'x,

b) uogólnionej macierzy kątów między wektorami macierzy x.

Podano przykłady liczbowej analizy zależności za pomocą metod (a) - (b) oraz wskazano na wady i zalety obu metod.

W paragrafie 3 wyprowadzono nowe wzory uzależniające wrażliwość wskaźnika złego uwarunkowania od wpływowych elementów z macierzy x, wpływowych kolumn x oraz nowe wzory uzależniające wrażliwość statystyk B, $\hat{Y}$, E, E'E, E'E $(n-k)^{-1}$ oraz ich charakterystyk na poziom złego uwarunkowania.

Podano warunki wystarczające i konieczne bezwrażliwości tych statystyk.