A C T A   U N I V E R S I T A T I S   L O D Z I E N S I S

FOLIA OECONOMICA 68, 1987

Iwona Konarzewska*, Władysław Milo**

NOTES ON SENSITIVITY OF LEAST SQUARES ESTIMATES AND THEIR CHOSEN FUNCTIONS

1. Introduction

In this paper we will analyze possibilities of measuring a sensitivity of sample values of least squares estimator $B = (x'x)^{-1}x'Y$ of the model $\mathcal{NM}$ parameter vector $\beta \in R^k$ on small changes in values of observation results referring to the elements of a matrix x and a vector y, where

$$\mathcal{NM} = (R^{nxk}, S, Y = x\beta + \Xi, P_Y = \mathcal{N}_Y(x\beta, \sigma^2 I)),$$

and

R^{nxk} - a set of real nxk matrices,

$S = (\mathcal{U}, \mathcal{F}, P)$ - a probability space with measure P defined on σ - field $\mathcal{F}$ of Borel subsets of $\mathcal{U}$ with $P(\mathcal{U}) = 1$,

$Y, \Xi : (\mathcal{U}, \mathcal{F}) \to (R^n, \mathcal{F}_{R^n})$; $y, \xi \in R^n$, $y = (y_1, ..., y_n)'$,

$$x \in R^{nxk}, x = [x_{tp}]\ t = \overline{1, n},\ p = \overline{1, k},\ \mathcal{E}(Y) = x\beta,\ \mathcal{D}(Y) = \sigma^2 I,$$

$P_Y = \mathcal{N}_Y(x\beta, \sigma^2 I)$ - to be read: "a probability distribution of Y is n-dimensional normal distribution with a mean $\mathcal{E}(Y) = x\beta$ and dispersion $\mathcal{D}(Y) = \sigma^2 I$".

We propose some numerical characteristics (indicators, measures) of components of the vector b or of the whole vector b sensitivity on small changes in the matrices x, y. Proposed indicators enable to detect those elements of the matrices x, y which

* Senior Assistant, Institute of Econometrics and Statistics, University of Łódź.

** Lecturer, Institute of Econometrics and Statistics, University of Łódź.

strongly pertrub numerically the solution $b = (x'x)^{-1}x'y$. It seems that indicators proposed in this paper can be easily used as a base for some respecifications of the elements of statistical models.

We will accept the following notations:

$$x = [x_{tl}] \quad t = \overline{1, n}, \quad l = \overline{1, k}, \quad y = (y_1, \ldots, y_n)'$$

$$x'x = [x_{ij}] \; i, j = \overline{1, k} \qquad (x'x)^{-1} = [x^{ij}] \; i, j = \overline{1, k}$$

$$x'y = (z_1, \ldots, z_k)' \qquad x^{+} = (x'x)^{-1}x' = [x^{+}_{ij}] \; i = \overline{1, k}, \; j = \overline{1, n}$$

$$b = (x'x)^{-1}x'y = x^{+}y.$$

In § 2 we remind some statements, partly known from literature (see D w y e r [2], W r o b l e s k i [7], B a l e s t r a [1], T h e i l [6], B e l s l e y, K u h, W e l s c h [3], P o l a s e k [5]), referring to matrix derivatives calculus. These statements are used in defining measures useful in detecting the most influential observations on model variables (in a sense of their influence on LS estimates numerical sensitivity). In § 3 we derive a formula for a derivative of a vector of LS estimates with regard to a fixed element of the matrix x. This formula was used in formulating a measure of LS estimates sensitivity on small changes in a fixed value of observation as well as in formulating measures of sensitivity of chosen functions of these estimates (see § 4). As it was examined in literature (see e.g. one of the most recent formulations in B e l s l e y, K u h, W e l s c h [3]), the existence of multicollinearity causes numerical instability of the elements of inverse matrix to the matrix $x'x$. Consequently, this can induce instability of a solution of $x'xb = x'y$ system of normal equations, that is, sensitivity of estimates b on small changes in data. On the other side the estimator values depend on values of elements of vector $x'y$. Therefore without detailed numerical analysis one cannot state that in individual case obtained estimates are really contaminated by multicollinearity. The problem concerns the estimation of LS estimator variance-covariance matrix as well.

It is obvious that the proposed sensitivity measures require further studies concerning:

- the interpretation of results obtainable by their usage,
- analytical behaviour for data sets generated according to different models,
- their applicability in detecting influential observations, model specification searches.

2. Chosen properties of matrix derivatives

The following properties of derivatives of different functions of the matrices **x** and **y** will be used:

Property 2.1:

$$\frac{\partial \mathbf{x}'\mathbf{x}}{\partial x_{tl}} = S_{(t,l)},$$

where $S_{(t,\,l)} = Q_{(t,\,l)} + Q'_{(t,\,l)}$,

$$Q_{(t,\,l)} = \Theta_{(l,\,t)}\mathbf{x} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ x_{t1} & x_{t2} & \cdots & x_{tk} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 \end{bmatrix},\quad Q_{(t,\,l)} \in R^{k\times k},$$

and $\Theta_{l,\,t}$ is a null kxn matrix besides the element (l, t) which is equal to one. ◆

(A proof results immediately from the matrix **x'x** definition).
We can introduce Prop. 2.1 in a different form as follows:

Property 2.1.a:

$$\frac{\partial x_{rs}}{\partial x_{tl}} = \begin{cases} 0, & r, s \neq l \\ x_{tl}, & ((r = l) \wedge (s \neq l)) \vee ((r \neq l) \wedge (s = l)) \\ 2x_{tl}, & (r = l) \wedge (s = l). \end{cases}$$ ◆

Property 2.2 (see Theil [6]):

$$\frac{\partial \mathfrak{X}^{rs}}{\partial \mathfrak{X}_{ij}} = - \mathfrak{X}^{ri}\, \mathfrak{X}^{js}, \quad r, s, i, j = \overline{1, k}.$$

Property 2.3 (see Dwyer [2]):

$$\frac{\partial (\mathbf{x}'\mathbf{x})^{-1}}{\partial x_{tl}} = -(\mathbf{x}'\mathbf{x})^{-1} S_{(t,l)} (\mathbf{x}'\mathbf{x})^{-1}, \quad t = \overline{1, n}, \quad l = \overline{1, k}. \blacklozenge$$

We can also present Prop. 2.3 in the following way:

Property 2.3.a:

$$\frac{\partial \mathfrak{X}^{rs}}{\partial x_{tl}} = - 2\, \mathfrak{X}^{rl} x^{+}_{ts}, \quad t = \overline{1, n}, \quad r, s, l = \overline{1, k}. \quad \blacklozenge$$

Proof:

$$\frac{\partial \mathfrak{X}^{rs}}{\partial x_{tl}} = \sum_{i,j} \frac{\partial \mathfrak{X}^{rs}}{\partial \mathfrak{X}_{ij}} \frac{\partial \mathfrak{X}_{ij}}{\partial x_{tl}} = \sum\nolimits_1 + \sum\nolimits_2 + \sum\nolimits_3.$$

Considering Prop. 2.1.a we decompose the first sum in the above expression into three sums, where

$$\sum\nolimits_1 = \sum_{i,j \neq l} \frac{\partial \mathfrak{X}^{rs}}{\partial \mathfrak{X}_{ij}} \frac{\partial \mathfrak{X}_{ij}}{\partial x_{tl}}, \quad \sum\nolimits_2 = \sum_{\substack{(i=l \wedge j \neq l) \\ \vee (i \neq l \wedge j = l)}} \frac{\partial \mathfrak{X}^{rs}}{\partial \mathfrak{X}_{ij}} \frac{\partial \mathfrak{X}_{ij}}{\partial x_{tl}},$$

$$\sum\nolimits_3 = \sum_{(i=l) \wedge (j=l)} \frac{\partial \mathfrak{X}^{rs}}{\partial \mathfrak{X}_{ij}} \frac{\partial \mathfrak{X}_{ij}}{\partial x_{tl}}.$$

Hence, by virtue of Prop. 2.1.a. we have

$$\frac{\partial \mathfrak{X}^{rs}}{\partial x_{tl}} = 0 + \sum_{\substack{(i=l \wedge j \neq l) \\ \vee (i \neq l \wedge j = l)}} \frac{\partial \mathfrak{X}^{rs}}{\partial \mathfrak{X}_{ij}} x_{tl} + 2 \frac{\partial \mathfrak{X}^{rs}}{\partial \mathfrak{X}_{ll}} x_{tl}.$$

Due to symmetry of the matrix $x'x$ and Prop. 2.2 we can write the above expression as follows:

$$\frac{\partial x^{rs}}{\partial x_{tl}} = 2\sum_{j=1}^{k}\frac{\partial x^{rs}}{\partial x_{lj}}x_{tj} = -2\sum_{j=1}^{k}x^{rl}x^{js}x_{tj} =$$

$$= -2x^{rl}\sum_{j=1}^{k}x^{js}x_{tj} = -2x^{rl}x^{+}_{ts}$$

which ends the proof.

Property 2.4:

$$\frac{\partial x'y}{\partial x_{tl}} = y_t\, j_l$$

where $j_l = (j_{l_1}, \dots, j_{l_k})'$, $j_{l_i} = \begin{cases} 0 & i \neq l \\ 1 & i = l \end{cases}$, $i = \overline{1, k}$. ◆

3. Measures of least squares estimates' sensitivity

By sensitivity of least squares estimates we understand a reaction of these estimates values on small changes in values of elements of the matrix x and vector y. We define sensitivity measures using derivatives of b with respect to x or y. The vector $\frac{\partial b}{\partial x_{tl}}$ may have different analytical forms. One of them is given in M i l o, W a s i l e w s k i [4]. Here we introduce its another form.

Property 3.1:

$$\frac{\partial b}{\partial x_{tl}} = (x'x)^{-1}[y_t j_l - S_{(t,l)}b]$$

$$= (x'x)^{-1}[e_t j_l - x_{t.}b_l],\ t = \overline{1, n},\ l = \overline{1, k},$$

where

x'_t - the t row of the matrix x,

$$e_t = y_t - \sum_{j=1}^{k} x_{tj}b_j = y_t - x'_{t.}\, b. \blacklozenge$$

Proof:

$$\frac{\partial b}{\partial x_{tl}} = (x'x)^{-1}\frac{\partial x'y}{\partial x_{tl}} + \frac{\partial (x'x)^{-1}}{\partial x_{tl}}x'y$$

$$= (x'x)^{-1}\left[y_t j_l - S_{(t,l)}(x'x)^{-1}x'y\right]$$

$$= (x'x)^{-1}\left[y_t j_l - S_{(t,l)}b\right].$$

Considering

$$S_{(t,l)} = Q_{(t,l)} + Q'_{(t,l)}$$

we can write next

$$\frac{\partial b}{\partial x_{tl}} = (x'x)^{-1}\left[y_t j_l - Q_{(t,l)}b - Q'_{(t,l)}b\right]$$

$$= (x'x)^{-1}\left[y_t j_l - x'_{t.}b\; j_l - x_{t.}b_l\right]$$

$$= (x'x)^{-1}\left[e_t j_l - x_{t.}b_l\right]$$

which concludes the proof.

Using Prop. 3.1 one can construct different sample sensitivity measures of the estimator B. We propose here the following ones:

$$\omega_{il} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial b_i}{\partial x_{tl}},\ i,\ l = \overline{1,\ k}$$

ω_{il} - mean sample sensitivity of B_i with respect to results of observations on a X_l variable;

$$w_{it} = \frac{1}{k}\sum_{l=1}^{k}\frac{\partial b_i}{\partial x_{tl}},\quad i = \overline{1,\ k},\quad t = \overline{1,\ n}$$

w_{it} - averaged sample sensitivity of B_i with respect to results of observations for the period t on all variables $X_1, \ldots, X_k$ (for models in which explanatory variables are in the same scale or standarized): $\bar{\omega}_l = \|\omega_l\|$, where $\omega_l = (\omega_{1l}, \omega_{2l}, \ldots, \omega_{kl})$ and $\|d\|$ is euclidean norm of d;

$\bar{\omega}_l$ - indicator of sample sensitivity of B with respect to results of observations on X_l variable;

$$\bar{w}_t = \|w_t\|, \text{ where } w_t = (w_{1t}, w_{2t}, \ldots, w_{kt})$$

$\bar{w}_t$ - an indicator of sample sensitivity of B with respect to results of observations on all explanatory variables from the period t (for models in which explanatory variables are in the same scale or standardized);

$$\tilde{w}_{it} = \frac{1}{k} \sum_{l=1}^{k} \frac{\partial b_i}{\partial x_{tl}} \frac{x_{tl}}{b_i}$$

w_{it} - weighted mean sensitivity of b_i on results of observations for explanatory variables from the period t;

$$\bar{\tilde{w}}_t = \|\tilde{w}_t\|, \text{ where } \tilde{w}_t = (\tilde{w}_{1t}, \tilde{w}_{2t}, \ldots, \tilde{w}_{kt})$$

$\bar{\tilde{w}}_t$ - an indicator of weighted sensitivity of b on results of observations for all explanatory variables from the period t.

In Belsley, Kuh, Welsch [3] we can find an alternative to w_{it} measure. It is based also on matrix derivatives. We introduce shortly its construction.

Let $D \in R^{n \times n}$ be a matrix which is defined as follows:

$$D = \begin{bmatrix} 1 & & & & 0 \\ & \ddots & & & \\ & & 1 \; d_{t_1} & & \\ & & & \ddots & \\ 0 & & & & 1 \end{bmatrix}, \quad d_t \in R_+,$$

that is D is a unit matrix except element d_t as a t-element on main diagonal. We define estimate b (d_t) as a weighted LS estimate with D as a weight matrix, that is

$$b(d_t) = (x'Dx)^{-1}x'Dy.$$

It is proved by the authors that

$$\frac{\partial b(d_t)}{\partial d_t} = (x'x)^{-1} \frac{x_{t.}\, e_t}{[1 - (1 - d_t)\, h_t]^2},$$

where

e_t - t-th LS residual,

h_t - t-th diagonal element of a hat matrix $H = xx^+$.

Evaluating this derivative in a point $d_t = 1$ we obtain

$$\delta_t = \frac{\partial b(d_t)}{\partial d_t}\bigg|_{d_t = 1} = (x'x)^{-1}x_{t.}e_t = x^{+\prime}_{t.}e_t.$$

The elements of a vector $\delta_t \in R^{k\times 1}$ are measures of influence that the observations on explanatory variables from the period t can have on LS estimates. From this reason the elements δ_{it} of the vector δ_t serve the same purpose as w_{it}. Without detailed analysis it is not obvious what is the relation between w_{it} (or $\tilde{w}_{it}$) and δ_{it} for a concrete case of data. It can be proved that

$$w_{it} - \delta_{it} = e_t(\bar{x}^{i.} - x^+_{it}) - x^+_{it}\,\bar{b},$$

where

$\bar{x}^{i.}$ - the mean value of elements in i-th row of $(x'x)^{-1}$ matrix

$\bar{b}$ - the mean value of LS estimates.

Similarly we can investigate the sample sensitivity of B on results of observations for the explained variable Y.

Property 3.2:

$$\frac{\partial b}{\partial y_t} = x^+_{.t},$$

where

$x^+_{.t}$ - the t-th column of the matrix x^+, $t = \overline{1, n}$.

The proof of Prop. 3.2 results immediately from the definition of b.

A measure of the mean sample sensitivity of B_i on the results of observations on Y we define as follows:

$$\omega_{i,y} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial b_i}{\partial y_t}, \quad i = \overline{1,\ k}$$

or in the case when variables are in different scales, for ability to compare

$$\tilde{\omega}_{i,y} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial b_i}{\partial y_t}\frac{y_t}{b_i}, \quad i = \overline{1,\ k}.$$

Mean sample sensitivity of B on small changes in **y** is following

$$\bar{\omega}_y = \|\omega_y\|, \quad \omega_y = (\omega_{1,y}, \ldots, \omega_{k,y})$$

or

$$\tilde{\omega}_y = \|\tilde{\omega}_y\| \quad \tilde{\omega}_y = (\tilde{\omega}_{1,y}, \ldots, \tilde{\omega}_{k,y}).$$

It was not possible to construct similar sensitivity measures of **b** with respect to changes of the model correlation structure in the sense of $\frac{\partial b_i}{\partial \mathfrak{X}_{rs}}$ i, r, s = $\overline{1,\ k}$ and $\frac{\partial b_i}{\partial z_r}$ i, r = $\overline{1,\ k}$ because of non unique evaluation of the elements of **x** by **x'x** and **x'y**.

4. Measures of sensitivity of chosen functions of LS estimates

We will find expressions defining a sensitivity on small changes in matrix **x** of such functions of LS estimates as

$$\hat{y} = xb, \quad e'e, \quad R^2 = 1 - \frac{e'e}{(n-k)S_y^2},$$

where

S_y^2 - sample variance of Y, $S_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$.

Property 4.1:

$$\frac{\partial \hat{y}}{\partial x_{tl}} = b_l(j_t - (x^+)' x_{t.}) + e_t x^+_{l.}$$
$$= b_l(j_t - h_{t.}) + e_t x^+_{l.},$$

where

$x^{+'}_{l.}$ - l-th row of $x^+ \in R^{kxn}$;

$h'_{t.}$ - t-th row of $H = xx^+$,

$j_t \in R^{nx1}$, $j_t = (j_1, \ldots, j_t, \ldots, j_n)'$, $j_i = \begin{cases} 0 & i \neq t \\ 1 & i = t \end{cases}$

$i = \overline{1, n}$. ◆

The proof comes immediately from definition of $\hat{y}$ and properties 2.3 and 3.1.

Property 4.2:

$$\frac{\partial e'e}{\partial x_{tl}} = 2\, b_l e'[j_t - h_{t.}].$$ ◆

Proof:

$$\frac{\partial e'e}{\partial x_{tl}} = \frac{\partial(\hat{y} - y)'(\hat{y} - y)}{\partial x_{tl}} = \frac{\partial \hat{y}'\hat{y}}{\partial x_{tl}} - 2\,\frac{\partial y'\hat{y}}{\partial x_{tl}} + \frac{\partial y'y}{\partial x_{tl}} =$$

$$= 2\,\hat{y}'\,\frac{\partial \hat{y}}{\partial x_{tl}} - 2\,y'\,\frac{\partial \hat{y}}{\partial x_{tl}} =$$

$$= 2(\hat{y} - y)'\,\frac{\partial \hat{y}}{\partial x_{tl}} = 2\,e'[b_l(j_t - h_{t.}) + e_t x^{+'} j_l]$$

$$= 2\,b_l e'[j_t - h_{t.}] = 2\,b_l e' j_t.$$

The last equality comes from the fact that $e'x^{+'} = 0$.

Property 4.3:

$$\frac{\partial R^2}{\partial x_{tl}} = \frac{2b_l}{(n - k)\,S_y^2}\; e'\, j_t.$$ ◆

Proof:

We define R^2 - sample estimate of a square multiple correlation coefficient as:

$$R^2 = 1 - \frac{e'e}{(n-k)S_y^2}.$$

Hence

$$\frac{\partial R^2}{\partial x_{tl}} = \frac{1}{(n-k)S_y^2}\frac{\partial e'e}{\partial x_{tl}} = \frac{2b_1}{(n-k)S_y^2} e' \mathfrak{J}_t.$$

what ends the proof.

5. Final remarks

In this paper we have introduced some propositions of measuring the sensitivity of LS estimates and their chosen functions on small changes in data(these changes can be understood as measurement errors of model variables or rounding errors). The described measures can be used in detecting the most influential observations (in the sense of their influence on solution of a set of normal equations) and in variable selection techniques. It seems that the measures can be useful especially for models with multicollinearity. More detailed analysis of their behaviour and derivation of cut off levels for these measures is planned by the authors.

Bibliography

[1] B a l e s t r a P. (1976): La Derivation Matricelle, Collection de L'Institute de Mathematiques Economiques, 12, Paris, Sirey.

[2] D w y e r P. S. (1962): Some Applications of Matrix Derivatives in Multivariate Analysis, "Journal of the American Statistical Association, 62, p. 607-625.

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

[4] M i l o W., W a s i l e w s k i Z. (1982): Estymacja odporna parametrów modelu liniowego. Cz. II, typescript within R. III.9.4.2, University of Łódź.

[5] P o l a s e k W. (1982): Local Sensitivity Analysis and Matrix Derivatives [in:] Operations Research in Progress, D. Reidel Publishing Co, p. 425-443.

[6] T h e i l H. (1971): Principles of Econometrics, New York, Wiley.

[7] W r o b l e s k i W. (1963): Extensions of the Dwyer-Mac Phail Matrix Derivative Calculus, Ann Arbor, University of Michigan.

Iwona Konarzewska, Władysław Milo

UWAGI O WRAŻLIWOŚCI OCEN WEDŁUG METODY NAJMNIEJSZYCH KWADRATÓW I ICH WYBRANYCH FUNKCJI

W artykule analizujemy możliwość pomiaru wrażliwości próbkowych wartości ocen według m.n.k. $b = (x'x)^{-1}x'y$ na małe zmiany wartości wyników obserwacji, tzn. małe zmiany elementów macierzy **x** i wektora **y**. Proponowane wskaźniki, oparte na pochodnych macierzy i wektorów, pozwalają na wykrycie tych elementów macierzy **x**, **y**, które silnie numerycznie zaburzają rozwiązanie układu równań normalnych. Wydaje się, że mogą one stanowić punkt wyjścia do pewnych respecyfikacji elementów modelu statystycznego.