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 131, 1993

*Iwona Konarzewska**

THE PROBLEM OF OBSERVATION MATRIX CONDITIONING AND SENSITIVITY OF THE LEAST SQUARES ESTIMATES - MONTE CARLO STUDY**

1. INTRODUCTION

The paper is devoted to the problem of sample sensitivity of least squares (LS) estimates in the case of bad-conditioned data for linear econometric model. Sensitivity is understood as the response of estimates to marginal changes in observation matrix and is measured by the values of first derivatives of estimates with respect to values of observations. As representative value characterizing LS sensitivity for given set of data the sensitivity coefficient with highest magnitude is chosen. This approach described in details in K o n a r z e w s k a (1986), K o n a r z e w s k a, M i l o (1987), K o n a r z e w s k a (1988) is useful for the purpose of regression diagnostics. Sensitivity of parameter estimates on small changes in data matrix is closely connected with three types of problems occuring in sample data:

- multicollinearity of explanatory variables resulting in bad conditioning of data matrix X;
- existence of influential observations, that is, rows of data matrix X situated far from other rows in a space of observations on explanatory variables;

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

** Paper presented at the conference on Multivariate Statistical Analysis WAS´89, Podklasztorze n.Sulejów.

- existence of so called outliers, that is, points of observations on the dependent variable Y situated far from other.

All three situations mentioned above can be potential reasons of sensitivity of regression results - estimates of parameters, predictions.

We would like to present shortly the idea of measuring sensitivity and the results of Monte Carlo experiment conducted to give a review of sensitivity coefficients behaviour in different model conditions.

2. DEFINITION OF THE MODEL AND SOME IMPORTANT CHARACTERISTICS

The following linear model is considered:

$$\begin{cases} Y = \beta'\mathbf{X} + \Xi \\ \mathbf{X} = (X_1, \ldots, X_k), \quad \mathcal{E}(\mathbf{X}) = \mu \in \mathcal{R}^{kx1}, \quad \mathcal{D}(\mathbf{X}) = \Sigma \in \mathcal{R}^{kxk} \\ \mathcal{P}_\Xi = \mathcal{N}(0, \sigma_\Xi^2), \text{ cov } (\mathbf{X}, \Xi) = \mathbb{O} \in \mathcal{R}^{kx1} \\ \beta \in \mathcal{R}^{kx1} \text{ - a vector of parameters} \end{cases} \quad (1)$$

where:

$\mathcal{E}, \mathcal{D}$ - operators of expected value and variance-covariance matrix respectively,

cov - operator of covariance between random variables or random variable with random vector,

$\mathcal{R}^{nxk}$ - n x k space of real numbers.

The model formulated above is theoretical one - for the purpose of parameter estimation we need sample observations on explanatory variables X_i, $i = \overline{1, k}$ and on the dependent variable Y. The model (2) given below is a sample version of (1) when we assume that the matrix X of observations on X_i is nonstochastic.

$$\begin{cases} Y = X\beta + \Xi \\ X \in \mathcal{R}^{nxk}, \quad rz(X) = k \\ \beta \in \mathcal{R}^{kx1}, \quad \mathcal{P}_\Xi = \mathcal{N}(\mathbb{O}, \sigma_\Xi^2 I) \end{cases} \quad (2)$$

where $\mathcal{P}_\Xi$ denotes probability distribution of the random vector Ξ.

The quality of the model (1) can be measured by the *explanation level of the model* defined as

$$\rho^2 = \frac{[\mathrm{cov}(Y, \beta'\mathbb{X}]^2}{\mathcal{D}(Y)\,\mathcal{D}(\beta'\mathbb{X})} \tag{3}$$

The measure ρ^2 is equal to the squared multiple correlation coefficient between Y and a set of explanatory variables X_i. Under the assumptions of the model (1) we obtain

$$\rho^2 = 1 - \frac{\mathcal{D}(\Xi)}{\mathcal{D}(Y)} \tag{4}$$

The equivalent to ρ^2 for the model (2) is, so called, determination coefficient R^2, which is a squared sample correlation coefficient between variable Y and $\hat{Y} = \mathbb{X}'B$, where B is an estimator of vector β. When B is an LS estimator, that is $B = B_{LS}$, where

$$B_{LS} = (X'X)^{-1}X'Y \tag{5}$$

then

$$\hat{Y} = X(X'X)^{-1}X'Y = HY$$

$$H = X(X'X)^{-1}X', \quad H \in \mathcal{R}^{nxn} \tag{6}$$

Taking for Y its sample realization $y \in \mathcal{R}^{nx1}$, the sample analog of explanation level ρ^2 is

$$R^2 = \frac{(y'M\,Hy)^2}{y'M\,yy'H\,M\,Hy}$$

$$M = I_n - 1/n\,jj', \quad j = [1, 1, \ldots, 1], \quad M \in \mathcal{R}^{nxn} \tag{7}$$

Sample vectors for $\hat{Y}$ and B_{LS} are denoted by $\hat{y} \in \mathcal{R}^{nx1}$ and $b_{LS} \in \mathcal{R}^{kx1}$. A vector $\hat{y} = Xb_{LS}$ is called LS predictor and e = = $y - \hat{y}$ is a vector of LS residuals.

We accept also the following notation:

$$X'X = \mathcal{X} = [\chi_{ij}]$$

$$(X'X)^{-1} = \mathcal{X}^{-1} = [\chi^{ij}].$$

The problem of multicollinearity is closely connected with sensitivity of estimates; sensitivity on small changes in data is one of its possible effects. The useful measure applied in multicollinearity diagnostics is the *condition level* of the observation matrix defined as

$$\kappa(X) = \frac{\mu_{max}}{\mu_{min}} \tag{8}$$

where μ_{max} and μ_{min} denote maximal and minimal singular values of X (see: G o l u b, R e i n s c h (1971) on Singular Value Decomposition - SVD)[1]. This measure is being used in examining the solutions of linear equations systems. The condition level $\kappa(X)$ is strictly dependent on scale - variables should be scaled identically to have unit lengths, for instance, standardized. Following B e l s l e y, K u h, W e l s c h (1980) we say that matrix is bad-conditioned if $\kappa(X) \geq 15$.

Other important consequencies of bad conditioning for LS estimator properties are the following:

a) high variances and high TMSE (Total Mean Square Error) regulated by a value of explanation level,

b) estimator of squared length of parameter vector - squared length of LS estimates - is always biased and the amount of this bias depends also on $\kappa(X)$ regulated by explanation level or σ^2_{Ξ}.

3. LEAST SQUARES ESTIMATES SAMPLE SENSITIVITY

We propose to measure sample sensitivity of estimates by values of first derivatives of estimates over changes in data matrix X. Applying some known theorems from matrix differential calculus it was shown in K o n a r z e w s k a, M i l o (1987), K o n a r z e w s k a (1988) that in the case of LS estimation method we obtain

$$\frac{\partial b_{LS}}{\partial x_t} = \mathcal{X}^{-1}[e_t j_l - x_{t*} b_l], \quad t = \overline{1,n}, \quad l = \overline{1,k} \tag{9}$$

where:

x'_{t*} - t-th row of X matrix,

b_l - l-th element of b_{LS} vector,

e_t - t-th residual,

$j_l \in \mathcal{R}^{kx1}$ - a vector of zeros except l-th element equal to one.

---

[1] The theorem which connects a value of condition level of X with sensitivity of diagonals of $\mathcal{X}^{-1}$ given in B e l s l e y, K u h, W e l s c h (1980, p. 174-176) shows that $\kappa(X)$ constitutes an upper limit for the ratio of relative change of χ^{ii} caused by marginal relative change of x_{rs}, i, $s = \overline{1,k}$, $r = \overline{1,n}$.

Elementwise representation for each b_i $i = \overline{1,k}$ is given as

$$\frac{\partial b_i}{\partial x_{tl}} = (\chi^{i*})' \, m_{tl}, \quad i = \overline{1,k}, \quad t = \overline{1,n}, \quad l = \overline{1,k} \qquad (10)$$

where

$m'_{tl} = [-x_{t1}b_l, \; -x_{t2}b_l, \; \ldots, \; e_t - x_{tl}b_l, \; \ldots, \; -x_{tk}b_l]$,

$(\chi^{i*})'$ - i-th row of $\mathcal{X}^{-1}$.

The elements of the vector multiplier $m_{tl} \in \mathcal{R}^{kx1}$ depend on position (t, l) and do not depend on b_i. The formulas (9) and (10) show strong dependence between sensitivity and values of $\mathcal{X}^{-1}$ matrix diagonals and the way in which multicollinearity can result in sensitivity of estimates. The influence of residuals with high magnitudes (symptoms of outlier existence), and of large values of H matrix diagonals (see (6) - symptoms of existence of influential observations) on estimates sensitivity is not clear in general and depends on individual sample.

4. THE CONSTRUCTION OF MONTE CARLO EXPERIMENT

The aim of Monte Carlo study was to check a behaviour of LS estimator when small changes in observation matrix were introduced. The following model characteristics were considered:

a) condition level of X; moreover, three schemes of location of its singular values in the range (μ_{min}, μ_{max}) were distinguished;

b) explanation level ρ^2 of the model;

c) the parameter vector β in k-dimensional space - two cases were considered - parallel to eigenvector connected with smallest and largest eigenvalue of $X'X$.

Experiment was undertaken to give a review, not detailed numerical information.

4.1. GENERATION OF X MATRIX

The X matrices of dimension 20 x 4 were generated according to the following algorithm.

1. First we generated a matrix X^* of dimension 20 x 4. Each of four columns of the matrix X^* was generated using random number

generator from uniform distribution on the range [-1, 1]. Three different X^* matrices were generated.

2. Each of X^* matrices was decomposed using SVD decomposition (G o l u b, R e i n s c h (1971))

$$X^* = UD^*V',$$

where:

$U \in \mathcal{R}^{nxk}$ - a matrix of normalized eigenvectors of XX' connected with its k nonzero eigenvalues;

$D^* = \mathrm{diag}(\mu_1^*, \ldots, \mu_k^*)$ - diagonal matrix with singular values of X (equal to nonnegative square roots of eigenvalues) on the main diagonal;

$V \in \mathcal{R}^{kxk}$ - orthogonal matrix of normalized eigen vectors of $X'X$, $V'V = VV' = I_k$.

3. Let κ be the desired value of condition level of X and $\mu_1 \geq \mu_2 \geq \mu_3 \geq \mu_4$ be singular values of X. Then μ_i are calculated according to one of the methods:

$$1^\circ \quad \mu_1 = \kappa\, \mu_4^*, \quad \mu_i = \mu_i^*, \quad i = 2, 3, 4;$$

$$2^\circ \quad \mu_1 = \kappa\, \mu_4^*, \quad \mu_2 = \mu_1 \frac{\mu_2^*}{\mu_1^*}, \quad \mu_i = \mu_i^*, \quad i = 3, 4;$$

$$3^\circ \quad \mu_1 = \kappa\, \mu_4^*, \quad \mu_2 = \mu_1 \frac{\mu_2^*}{\mu_1^*}, \quad \mu_3 = \mu_1 \frac{\mu_3^*}{\mu_1^*}, \quad \mu_4 = \mu_4^*.$$

4. The elements of X matrix were calculated from the formula

$$X = U\, D\, V',$$

where $D = \mathrm{diag}\,(\mu_1, \ldots, \mu_4)$.

According to the algorithm described above we generated 54 matrices X for 6 condition levels (κ = 5, 10, 30, 50, 100, 500). Three ways of calculating μ_i allows us to obtain matrices with the same condition levels but differing by the relations among singular values. These different relations can be interpreted in extreme situations of singularity of X as different ranks - in the case 1° r(X) = 1, 2° r(X) = 2, 3° r(X) = 3.

4.2. GENERATION OF OTHER MODEL ELEMENTS

Parameter vector β is taken as:

1° $\beta = 10\ v_{min}$

2° $\beta = 10\ v_{max}$,

where v_{min} and v_{max} are normalized eigenvectors connected with smallest and highest eigen value of $X'X$.

These two orthogonal choices of β's are extreme cases for biased estimators statistical characteristics proposed when sample matrix is bad conditioned, as was shown in K o n a r z e w s k a (1988) for the case of ridge estimator.

Hundred sample realizations of Ξ were generated using random number generator from multivariate normal distribution, with covariance matrix given as:

$$\mathcal{D}(\Xi) = \sigma_{\Xi}^2\ I$$

where

$$\sigma_{\Xi}^2 = s_{y*}^2\ (1 - \rho^2),\quad s_{y*}^2 = \frac{1}{20}\sum_{t=1}^{20}(y_t^* - \bar{y}^*).$$

We chose 3 explanation levels othe model $\rho^2 = 0.8, 0.9, 0.95$.

5. THE RESULTS OF THE EXPERIMENT

The main results of Monte Carlo experiment are presented on Figures 1-6. We present only absolute values of calculated maximal sensitivity coefficients for each set of data. Three matrices X^* are denoted by Q, T, X; resulting X matrices are denoted by QA, QB, QC, TA, TB, TC, XA, XB, XC where letter A corresponds to the case 1° of evaluating singular values, B - to the case 2°, C - to the case 3°.

The following tendencies in behaviour of maximal sensitivity coefficients are observed:

1° Strong dependence of sensitivity on condition level.

2° These dependence is approximately linear (on graphs we accepted logarithmic scale for condition level axis).

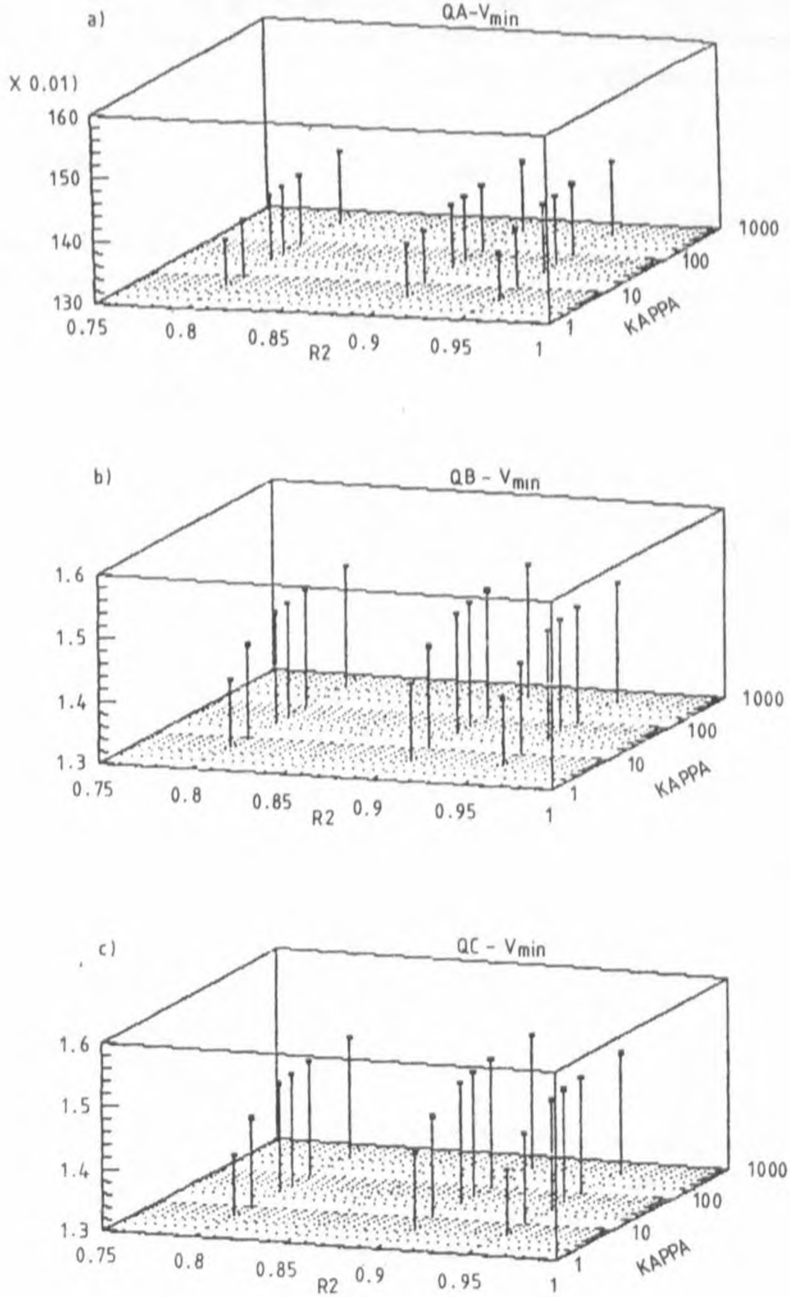

Fig. 1. Maximal sensitivity coefficients for the matrix Q and $\beta^{-} v_{min}$: a) QA, b) QB, c) QC

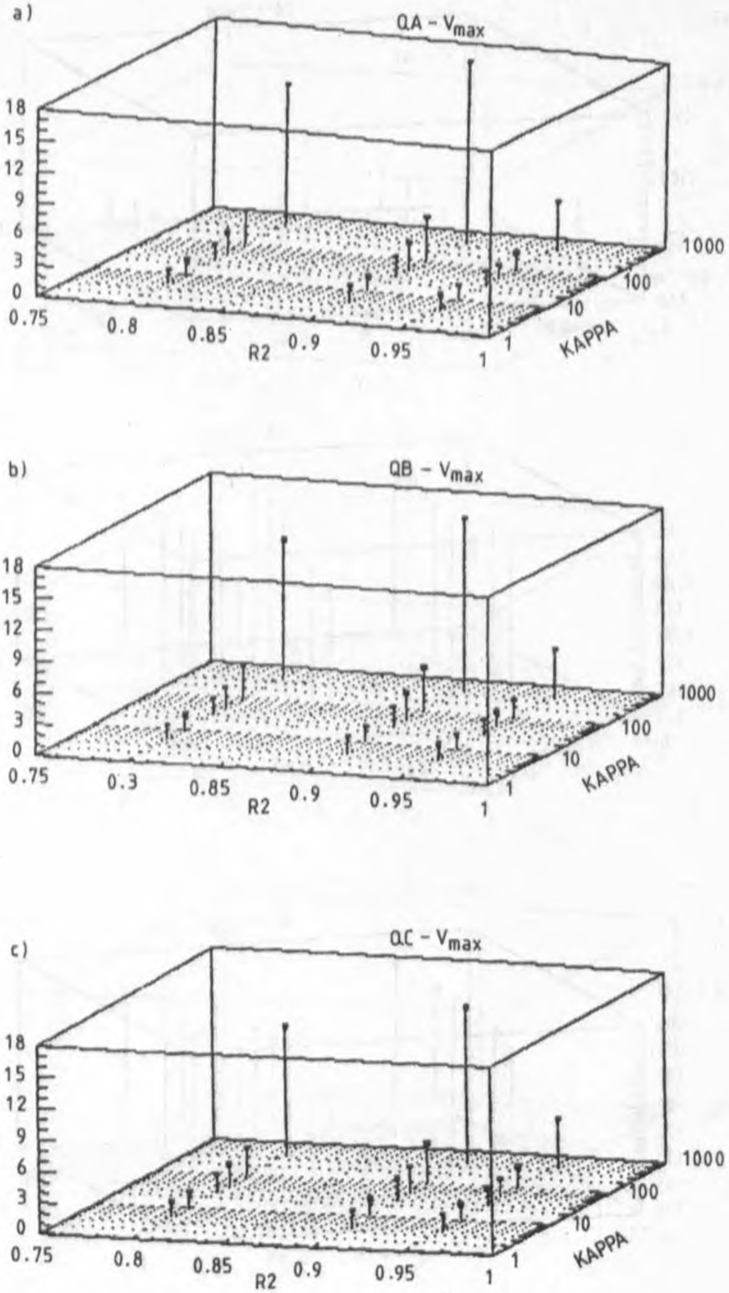

Fig. 2. Maximal sensitivity coefficients for the matrix Q and $\beta^{\sim} v_{max}$: a) QA, b) QB, c) QC

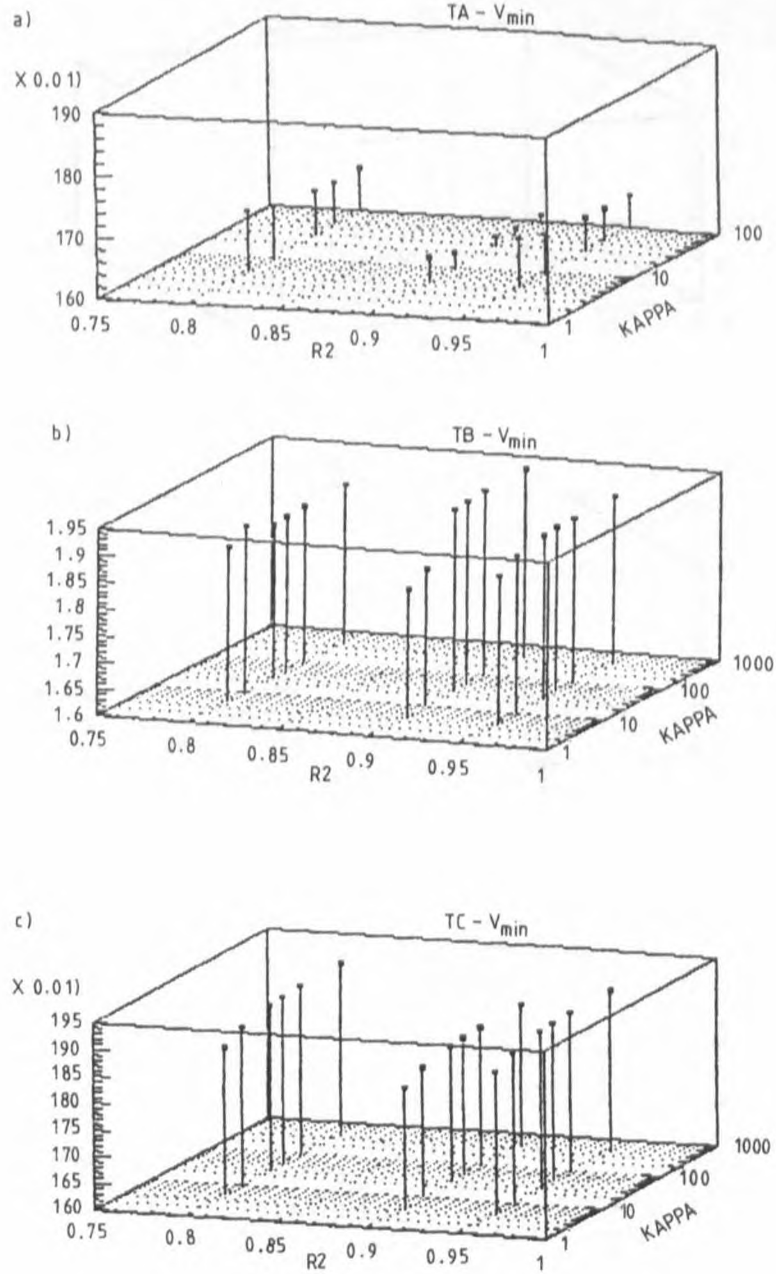

Fig. 3. Maximal sensitivity coefficients for the matrix T and $\beta^{\sim} v_{min}$:
a) TA, b) TB, c) TC

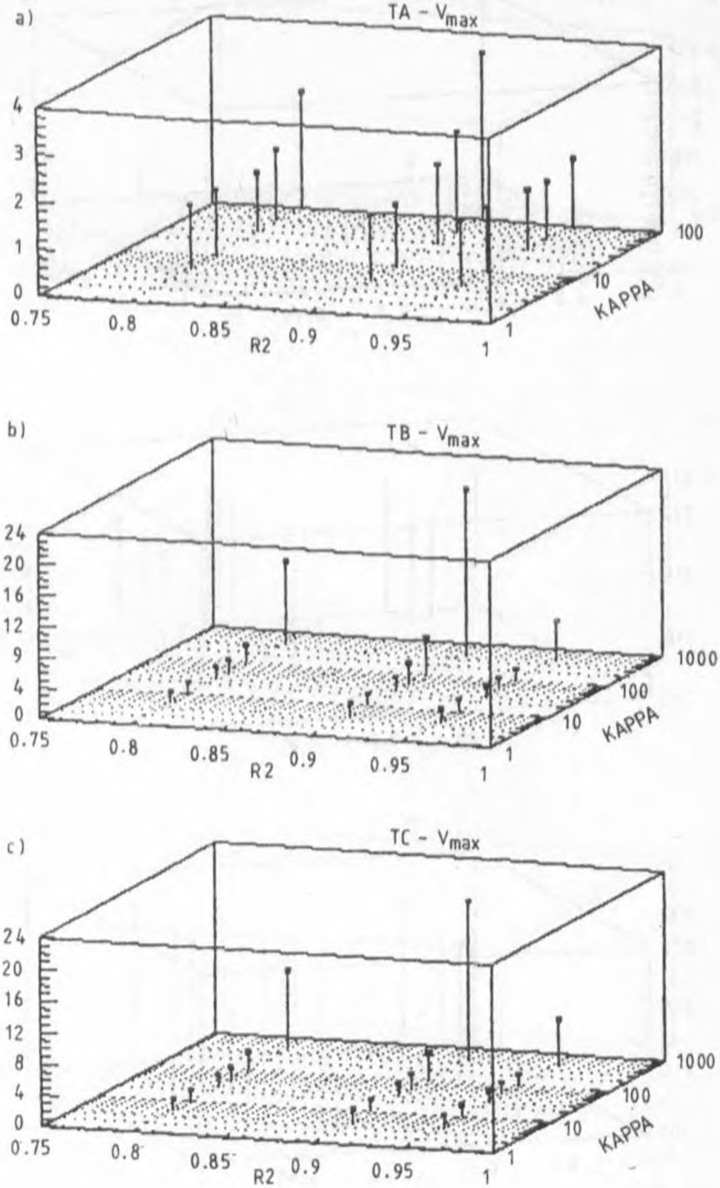

Fig. 4. Maximal sensitivity coefficients for the matrix T and $\beta^{-} v_{max}$:
a) TA, b) TB, c) TC

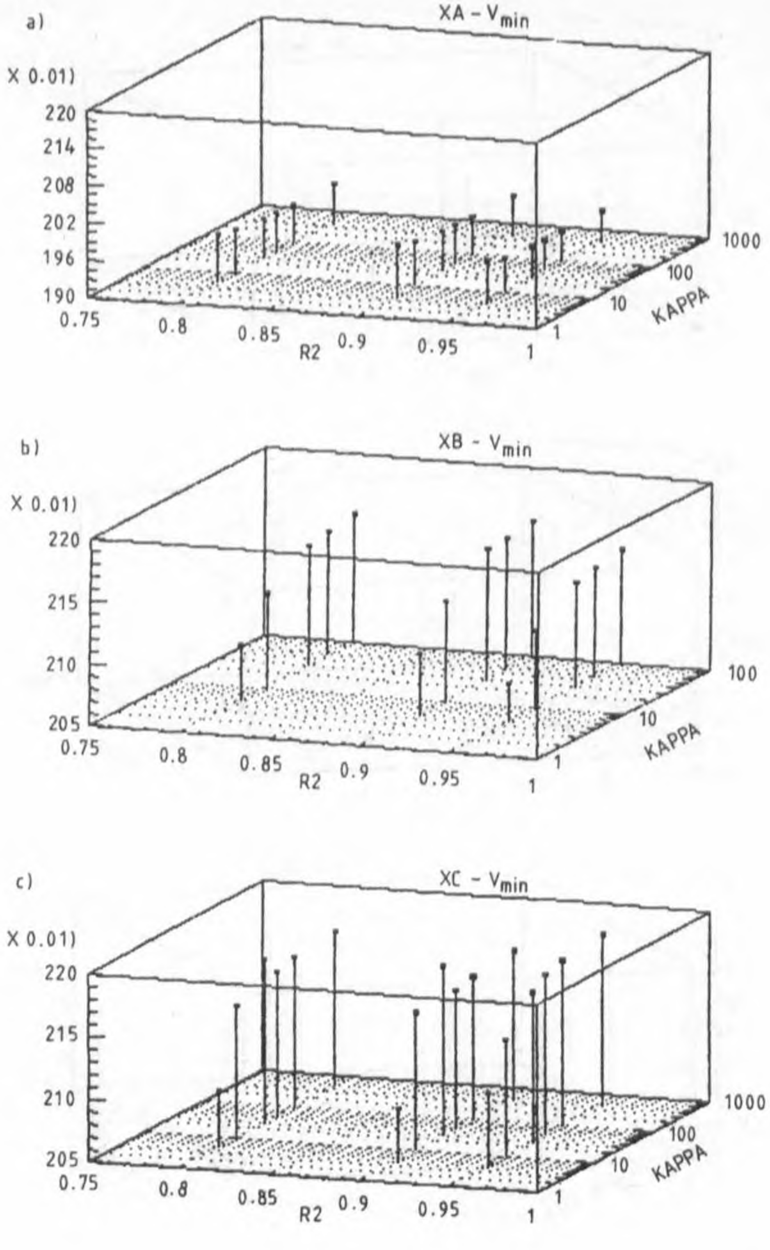

Fig. 5. Maximal sensitivity coefficients for the matrix X and $\beta^{\sim} v_{min}$:
a) XA, b) XB, c) XC

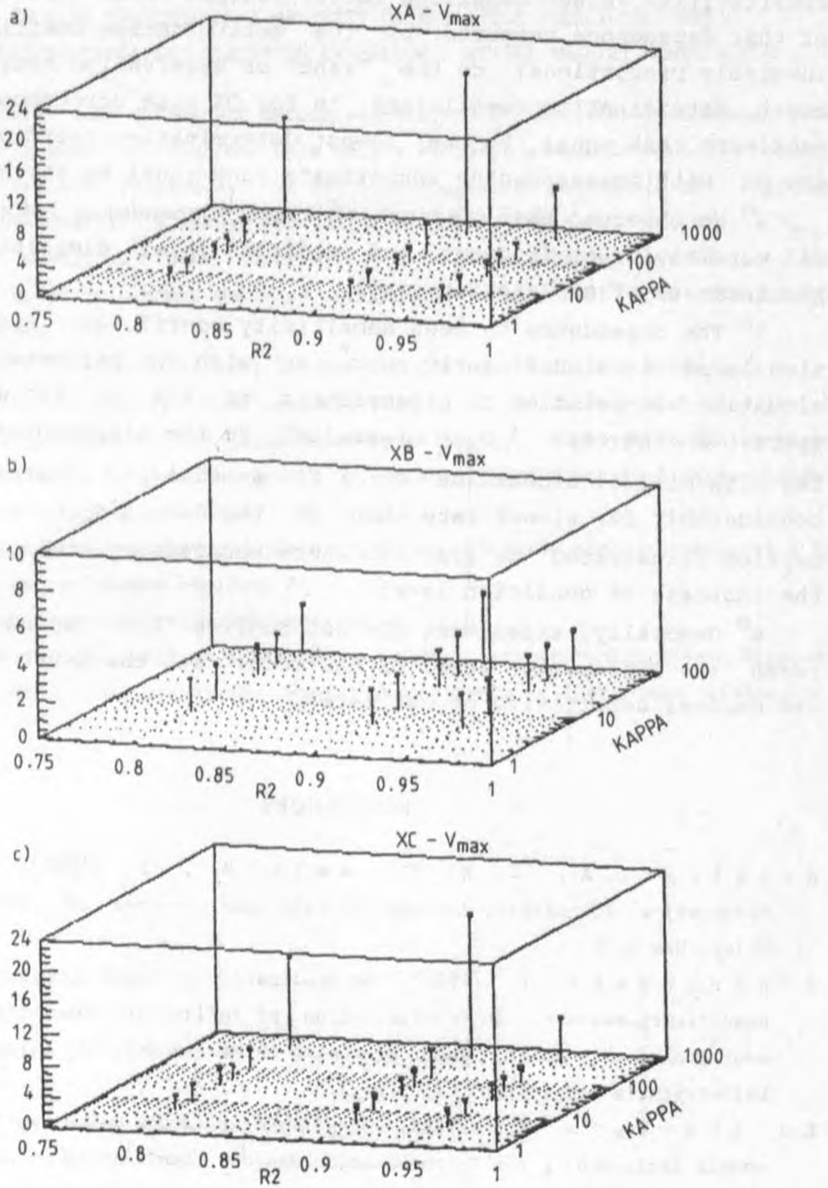

Fig. 6. Maximal sensitivity coefficients for the matrix X and $\beta^{-} v_{max}$: a) XA, b) XB, c) XC

3° Performed estimation of the linear regression of maximal sensitivities versus condition levels showed that the strenght of that dependence measured by the determination coefficient is inversely proportional to the "rank" of observation matrix (greatest determination coefficient is for QA with corresponding approximate rank equal to one; lowest determination coefficient was for QC with corresponding approximate rank equal to three).

4° We observed that strength of linear dependence between maximal sensitivity coefficient and condition level diminishes with the increase of explanation level.

5° The dependence between sensitivity coefficient and condition level is significantly connected with the parameter vector situation in relation to eigenvectors of $X'X$ in k-dimensional space. In the case $\beta^{-}v_{min}$ (parallel to the eigenvector connected with minimal eigenvalue of $X'X$) sensitivity increases in considerably far slower rate than in the case $\beta^{-}v_{max}$, with exception illustrated on graph 5a where sensitivity diminishes with the increase of condition level.

6° Generally, experiment did not confirm the dependence between the explanation level of the model and the level of observed maximal sensitivity of estimates.

REFERENCES

B e l s l e y D. A., K u h E., W e l s c h R. E. (1980), *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, Wiley, New York.

K o n a r z e w s k a I. (1986), *On Application of Least Squares Estimates Sensitivity Measures in Identification of Influential Observation in Linear Econometric Model*, paper presented on Macromodels'86, Szczyrk, "Acta Universitatis Lodziensis" (in print).

K o n a r z e w s k a I. (1988), *Wrażliwość wybranych metod estymacji dla modeli liniowych i źle uwarunkowanych danych* [Sensitivity of chosen estimation methods for linear models and bad-conditioned data], Ph.D thesis, University of Łódź.

K o n a r z e w s k a I., M i l o W. (1987), *Notes on Sensitivity of Least Squares Estimates and Their Chosen Functions*, "Acta Universitatis Lodziensis", Folia oeconomica nr 68, p. 89-100.

Iwona Konarzewska

PROBLEM UWARUNKOWANIA MACIERZY OBSERWACJI A WRAŻLIWOŚĆ OCEN UZYSKANYCH METODĄ NAJMNIEJSZYCH KWADRATÓW - WYNIKI ANALIZY MONTE CARLO

Wrażliwość ocen uzyskanych metodą najmniejszych kwadratów rozumiana jest jako reakcja ocen na krańcowo małe zmiany wartości elementów macierzy obserwacji. Miernikami wrażliwości są wartości pierwszych pochodnych ocen względem wartości obserwacji. W wyniku otrzymuje się trójwymiarowe macierze $[\Theta_{itl}]$ (i - numer oceny, t - numer obserwacji, l - numer zmiennej objaśniającej modelu). Maksymalny element macierzy wskaźników wrażliwości ocen dla danego zbioru danych.

Przyjęte zostało założenie, że wrażliwości ocen m.n.k. są funkcyjnie związane ze stopniem uwarunkowania macierzy obserwacji na zmiennych objaśniających X. Celem eksperymentu Monte Carlo było sprawdzenie, czy przyjęcie takiej hipotezy jest zasadne przy zmieniających się następujących warunkach eksperymentu:

- wariancja zakłóceń modelu;
- kierunek wektora parametrów modelu względem wektorów własnych macierzy $X'X$;
- stopień uwarunkowania macierzy X;
- struktura wartości osobliwych macierzy X.

Eksperyment pozwolił na wskazanie warunków, przy których przyjęta hipoteza może być uznana za prawdziwą. Wyniki eksperymentu zilustrowano wieloma wykresami.