ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 196, 2006

*Wiesław Wagner**

MATRIX H AND ITS APPLICATIONS IN ECONOMIC AND TOURIST RESEARCH

Abstract. The paper characterizes and gives some basic properties of matrix **H** that is the operator of orthogonal projection on the space of columns of matrix **X**, which is the hat matrix in linear regression model. The author enumerates some properties of diagonal elements of the matrix, which were defined as lever points and high-lever points. The paper deals also with the form of matrix **H** with repeating observations relating to independent variables. Two examples show some applications of matrix **H**.

Key words: linear regression model, hat matrix, prediction matrix, lever points.

1. INTRODUCTION

The hat matrix **X** essentially parts in statistical analysis of the linear regression model. It is defined as the set of n observable p-dimensional column vectors made on interpretative variables $x_0, x_1, ..., x_{p-1}$, where x_0 is the variable which is identically equal to 1 for each observation (case).

Matrix **X** is characterized by many algebraical properties and it is the basis to define matrix **H** that is the operator of the orthogonal projection on its column space. These properties will be enumerated without their derivations. That is because we can easily find them in many available literature positions and among them such authors as: Anderson (1958), Arnold (1981), Kolupa and Witkowski (1981), Rao (1982), Oktaba (1986), Rousseeuw and Leroy (1987).

* Professor, University of Information Technology and Management in Rzeszów.

2. DESIGNATION AND PROPERTIES OF MATRIX H

Showing the properties of matrix **H** we will use the abridged notation. We will mark given concept connected with matrix **X** or matrix **H** with the symbol (D). Further on, in the subsections (i), (ii),... we enumerate the analytical results referring to the given definition and if a need of giving some further properties arises then we will use denotations (1), (2),...

(D1) $\mathbf{X}: n \times p,\ n > p,\ r(\mathbf{X}) = p$ – the assumption about the rank of the matrix **X**:
(i) $\mathbf{X} = [\mathbf{x}_1', \mathbf{x}_2', \ldots, \mathbf{x}_n']$ – the row presentation of matrix **X**,
(ii) $\mathbf{x}_i' : 1 \times p$ – i-th row vector.

(D2) $\mathbf{X}'\mathbf{X}: p \times p$ – the matrix of sums of squares and products (the moments matrix):
(i) $r(\mathbf{X}'\mathbf{X}) = p$ – the rank of the matrix,
(ii) $\mathbf{X}'\mathbf{X}$ – symmetry,
(iii) $|\mathbf{X}'\mathbf{X}| \neq 0$ non-zero value of the matrix determinant,

(D3) $\mathbf{G} = (\mathbf{X}'\mathbf{X})^{-1} : p \times p$ – the matrix inverse to $\mathbf{X}'\mathbf{X}$,

(D4) $\mathbf{A} = \mathbf{GH}' : p \times n$ – the catcher matrix sygnifying the linear transformation of row vectors of the hat matrix **X** by the matrix of the transformation **G**:
(i) $\mathbf{AX} = \mathbf{I}_p$ – the property of the orthonormality of the system p of row vectors of the matrix **A** and the column matrixes **X**,
(ii) $\mathbf{AA}' = \mathbf{G}$ – the product of the catcher matrix and its transposed one is equal to the matrix of the transformation **G**.

(D5) $\mathbf{H} = \mathbf{XA} = \mathbf{XGX}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' : n \times n$ – the matrix of the orthogonal projection:
(i) $\mathbf{H} = \mathbf{X}'$ – symmetry,
(ii) $\mathbf{HH} = \mathbf{H}$ – idempotency,
(iii) $r(H) = tr(\mathbf{H}) = p$ – the equality of the rank and the trace of matrix **H**,
(iv) $\mathbf{HX} = \mathbf{X}$ – the orthogonal projection on columns of matrix **X**,
(v) $\mathbf{H1} = \mathbf{1}$ – the property of the one-value vector which means that the vector 1 is one of the column vectors of matrix **X**,
(vi) $\mathbf{H} = \tilde{\mathbf{H}}$, if $\tilde{\mathbf{X}} = \mathbf{X}\tilde{\mathbf{B}}$, with $\tilde{\mathbf{B}} : p \times p,\ r(\tilde{\mathbf{B}}) = p$,
(vii) $ch(\mathbf{H}) = 1$, the multiplication factor of p, $ch(\mathbf{H}) = 0$, the multiplication factor of $n - p$ of characteristic roots,
(viii) $\mathbf{H} = (h_{ij}),\ h_{ij} = \mathbf{x}_i'\mathbf{G}\mathbf{x}_j,\ i, j = 1, 2, \ldots, n$:
(1) $h_{ii} = \mathbf{x}_i'\mathbf{G}\mathbf{x}_i$ i-th diagonal element of matrix **H**,
(2) $1/n \leqslant h_{ii} \leqslant 1$ – for all diagonal elements in the one-value range $(0, 1\rangle$,

(3) $\sum_{i=1}^{n} h_{ii} = p < n$ – the sum of diagonal elements,

(4) $h_{ii} = \sum_{j=1}^{n} h_{ij}h_{ij} = \sum_{j=1}^{n} h_{ij}^2 = h_{ij}^2 + \sum_{\substack{j=1 \\ j \neq i}}^{n} h_{ij}^2$ – the property of the symmetry and the idempotency,

(5) $\sum_{j=1}^{n} h_{ij} = 1,\ i = 1, 2, ..., n$ – the sum of row (column) elements,

(6) $\sum_{j=1}^{n} h_{ij}\mathbf{x}_j' = \mathbf{x}_i'$ – the linear combination for i-th row vector of matrix $\mathbf{X}$,

(7) $h_{ii}h_{jj} - h_{ij}^2 \geqslant 0,\ i\ j = 1, 2, ..., n$ – the property of the tetrad of elements of i, j-th row and the column of matrix $\mathbf{H}$,

(8) $(1 - h_{ii})(1 - h_{jj}) - h_{ij}^2 \geqslant 0,\ i,\ j = 1, 2, ..., n$ – the property of the tetrad of elements i, j-th row and the column of matrix $\mathbf{I} - \mathbf{H}$,

(9) $\mathbf{x}'\mathbf{G}\mathbf{x} \leqslant \max_i h_{ii}$ for any $\mathbf{x} \neq 0$ – the condition of the upper bound of diagonal elements,

(10) $MD_i^2 = (\mathbf{x}_i - \overline{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x}_i - \overline{\mathbf{x}}) = (n-1)\left(h_{ii} - \frac{1}{n}\right)$ – Mahalanobis distance where adequately $\overline{\mathbf{x}}$ is the vector of means while $\mathbf{S}$ is the covariance matrix in the form of $\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})'$,

(ix) $\mathbf{H} = \mathbf{1}\mathbf{1}'/n + \mathbf{Y}(\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'$, where $\mathbf{Y}$ is the matrix with the aligned row vectors $\mathbf{y}_i' = \mathbf{v}_i' - \overline{\mathbf{v}}'$, while $\mathbf{x}_i' = (1, \mathbf{v}_u')$ is i-th row of the matrix $\mathbf{X}$ and $\overline{\mathbf{v}}'$ is the vector of means determined from observable variables $x_1, x_2, ..., x_{p-1}$:

(1) $h_{ij} = \frac{1}{n} + \mathbf{y}_i'(\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{y}_j$,

(2) $h_{ii} = \frac{1}{n} + \frac{(x_{i1} - \overline{x}_1)^2}{\sum_{j=1}^{n}(x_{j1} - \overline{x}_1)^2}$ for $p = 1$, in that case the matrix $\mathbf{Y}$ takes the form of the column vector $\mathbf{x}_1$, that is $\mathbf{Y} = \mathbf{x}_1 = (x_{11}, x_{21}, ..., x_{n1})'$,

(x) $h_{ii} = \frac{1}{n} + \sum_{k=1}^{p-1} \frac{(\mathbf{a}_i'\mathbf{y}_i)^2}{\lambda_k}$, where $\lambda_1 \leqslant \lambda_2 \leqslant ... \leqslant \lambda_{p-1}$ are the eigenvalues of matrix $\mathbf{Y}'\mathbf{Y}$, while matrix $\mathbf{Y}$ is given in the point (*ix*), however $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_{p-1}$ are orthonormal eigenvectors relating to roots $\lambda_1, \lambda_2, ..., \lambda_{p-1}$ (Cook, Weisberg 1980),

(xi) $h_{ii} = \frac{1}{n} + \mathbf{y}_i'\mathbf{y}_i \sum_{k=1}^{p-1} \frac{\cos^2\gamma_{ki}}{\lambda_k}$ where the same denotations as in the point (x) were assumed while $\cos\gamma_{ki} = \frac{\mathbf{a}_k'\mathbf{y}_i}{\sqrt{\mathbf{y}_i'\mathbf{y}_i}}$ is the cosine of the angle between vectors $\mathbf{a}_k$ and $\mathbf{y}_i$ (Cook, Weisberg 1980),

(xii) $\mathbf{H} = \mathbf{U}\mathbf{U}'$, where $\mathbf{X} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}'$ signifies the distribution of singular values of matrix $\mathbf{X}$ while $\mathbf{U}'\mathbf{U} = \mathbf{I}$, $\mathbf{\Lambda}$ – the diagonal matrix with eigenvalues of matrix $\mathbf{X}'\mathbf{X}$ (or $\mathbf{X}\mathbf{X}'$), $\mathbf{V}$ – orthogonal (Haoglin, Welsch 1978).

In D5 (vi) we showed one of the properties of the invariance of matrix $\mathbf{H}$ with non-singular transformation which was defined by the matrix $\tilde{\mathbf{B}}$. The standarization of the observation $x_{ij} \to (x_{ij} - a_j)/b_j$, $j = 1, 2, ..., p-1$ is one of the often used transformations on observable variables $x_1, x_2, ..., x_{p-1}$ is. For the advisable standarized transformation the lemma proceeds.

Lemma. If there is defined matrix $\mathbf{H}$ for matrix $\mathbf{X} = (x_{ij})$, $i = 1, 2, ..., n$; $j = 0, 1, ..., p-1$ it stays invariant after applying the standarized transformation $x_{ij} \to (x_{ij} - a_j)/b_j$, where a_j, b_j are adequately the position and variation measures from samples for variables $x_1, x_2, ..., x_{p-1}$.

Given lemma is directly connected with D5 (vi) by the matrix

$$\tilde{\mathbf{B}} = \begin{bmatrix} 1 & -a_1/b_1 & -a_2/b_2 & \dots & -a_{p-1}/b_{p-1} \\ 0 & 1/b_1 & 0 & \dots & 0 \\ 0 & 0 & 1/b_2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1/b_{p-1} \end{bmatrix}.$$

Let us emphasize that the property D5 (vi) will be always satisfied for the matrix $\tilde{\mathbf{B}}$ of the diagonal form $\tilde{\mathbf{B}} = \text{diag}/1, \tilde{b}_1, \tilde{b}_2, ..., \tilde{b}_{p-1})$.

We will show the derivation of matrix $\mathbf{H}$ in the particular case if $p = 2$:

- $\mathbf{X} = [\mathbf{1}, \mathbf{x}]$, $\mathbf{X}: n \times 2$, $\mathbf{1}: n \times 1$, $x: n \times 1$, $\mathbf{1}$ – the one-value vector,
- $\mathbf{X}'\mathbf{X} = \begin{bmatrix} \mathbf{1}'\mathbf{1} & \mathbf{1}'\mathbf{x} \\ \mathbf{x}'\mathbf{1} & \mathbf{x}'\mathbf{x} \end{bmatrix} = \begin{bmatrix} n & SX \\ SX & SX2 \end{bmatrix}$,

where $n = \mathbf{1}'\mathbf{1}$, $SX = \mathbf{1}'\mathbf{x} = \mathbf{x}'\mathbf{1}$ $SX2 = \mathbf{x}'\mathbf{x}$,

- $\mathbf{G} = (\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{W}\begin{bmatrix} SX2 & -SX \\ -SX & n \end{bmatrix}$,

where $W = |\mathbf{X}'\mathbf{X}| = n \cdot SX2 - SX^2 = n \cdot SSX$, $SSX = SX2 - n\bar{x}^2$,

- $\mathbf{H} = \mathbf{X}\mathbf{G}\mathbf{X}' = \frac{1}{W}[SX2\mathbf{1}\mathbf{1}' - SX(\mathbf{1}\mathbf{x}' + \mathbf{x}\mathbf{1}') + n\mathbf{x}\mathbf{x}']$,
- $\mathbf{H} = (h_{ij})$, $h_{ij} = (SX2 - SX(x_i + x_j) + nx_ix_j)/W$, i, $j = 1, 2, ..., n$,

- $h_{ii} = (SX2 - 2SXx_i + nx_i^2)/W = [(SX2 - n\overline{x}^2) + n(\overline{x}^2 - 2\overline{x}x_i + x_i^2)]/W =$
$= [SSX + n(\overline{x} - x_i)^2]/(n\,SSX) = 1/n + (x_i - \overline{x})^2/SSX = \frac{1}{n} + \frac{(x_i - \overline{x})^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}.$

Matrix **H** signifies the connection between the estimated observations vector $\hat{\mathbf{y}}$ the observable interpretative variable **y** and observations vector **y** on this variable in the form of $\hat{\mathbf{y}} = \mathbf{Hy}$ which is characterized in the analysis of the linear regression model. In so called diagnostic analysis diagonal elements of matrix **H** are defined as lever points playing an important part in assigning of the influential points. We assume high values of lever points as significant ones. These values exceed some assigned threshold value h_0. Most often at this point we assume the average value $h_0 = p/n$. That means that this is the average value for lever points. In diagnostic analysis cases which satisfy the inequality $h_{ii} > h_0$ are commonly called high lever points. They play an important part in identification of influential points.

3. MATRIX WITH REPEATED ROW VECTORS OF MATRIX X

Using notations applied in Section 2 we derivate matrix **H** but only in case if in hat matrix **X** repeated row vectors occur. Such situation often appears in research of the cause-and-effect relation if we assume the principle of the replication for non-changed values of interpretative variables. We discuss the indicated problem in the following points:

1. We assume that the matrix **X** contains q groups of repeated row vectors (cases).
2. Repeated vectors create the system of vectors $\mathbf{x}_1', \mathbf{x}_2', ..., \mathbf{x}_q'$ with the multiplication factor adequately $k_1, k_2, ..., k_q$ so that $\sum_{r=1}^{q} k_r = n$.
3. Vectors $\mathbf{x}_1', \mathbf{x}_2', ..., \mathbf{x}_q'$ were metrically moved in matrix **X**, making $(k_r \times p)$-dimensional submatrixes $\mathbf{X}_r$ with all k_r identical vectors $\mathbf{x}_r'$.
4. The matrix **X** with repeated row vectors takes the form of $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2', ..., \mathbf{X}_q')'$, where each of submatrixes is signified as the Kronecker's product of k_r-dimensional one-value vector and the observations vector, that is $\mathbf{X}_r = \mathbf{1}_{k_r} \otimes \mathbf{x}_r'$, $r = 1, 2, ..., q$.
5. The relation for the product of matrix **X** occurs as follows:

$$\mathbf{X}'\mathbf{X} = \sum_{r=1}^{q} \mathbf{X}_r'\mathbf{X}_r = \sum_{r=1}^{q} (\mathbf{1}_{k_r} \otimes \mathbf{x}_r')'(\mathbf{1}_{k_r} \otimes\otimes \mathbf{x}_r') = \sum_{r=1}^{q} (\mathbf{1}_{k_r}' \mathbf{1}_{k_r} \otimes \mathbf{x}_r \mathbf{x}_r') = \sum_{r=1}^{q} k_r \mathbf{x}_r \mathbf{x}_r'.$$

6. For matrix $\mathbf{Z}_r = \mathrm{k}_r \mathbf{x}_r \mathbf{x}_r'$ the following properties occur:

- $r(\mathbf{Z}_r) = 1$ – the rank of the matrix,
- $\mathbf{Z}_r' = \mathbf{Z}_r$ – symmetry,
- $\mathbf{Z}_r \mathbf{Z}_r' = \lambda_r \mathbf{Z}_r$ where $\lambda_r = k_r \mathbf{x}_r' \mathbf{x}_r$ is the non-zero root of matrix $\mathbf{X}_r' \mathbf{Z}_r$, while the remaining roots are equal to zero.

The matrix $\mathbf{H}$ in the case of repeated cases can be signified in the form of block matrixes $\mathbf{H} = (\mathbf{H}_{rs})$ while $r,\ s = 1, 2, \ldots, q$ occurs in the form of

$$\mathbf{H}_{rs} = \mathbf{X}_r(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_s' = (\mathbf{1}_{k_r} \otimes \mathbf{x}_r')(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{1}_{k_r} \otimes \mathbf{x}_s')' = \mathbf{1}_{k_r}\mathbf{1}_{k_s}' \otimes \mathbf{x}_r'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_s =$$
$$= h_{rs}\mathbf{1}_{k_r}\mathbf{1}_{k_s}' = h_{rs}\mathbf{J}_{k_r k_s},$$

where $\mathbf{J}_{k_r k_s}$ is $(k_r \times k_s)$-dimensional matrix of all one-value elements.

For the matrix $\mathbf{H}$ expressed by block matrixes with repeated vectors in matrix $\mathbf{X}$ we can claim that:

- all its elements in block submatrixes are identical,
- if $q = p - 1$ elements of matrix $\mathbf{H}$ do not depend on the value of vectors $\mathbf{x}_1', \mathbf{x}_2', \ldots, \mathbf{x}_q'$, but only on their multiplication factors and, to be more precise, on their converse.
- If $q > p - 1$, then values of the matrix $\mathbf{H}$ elements depend on vectors: $\mathbf{x}_1', \mathbf{x}_2', \ldots, \mathbf{x}_q'$.

All the mentioned problems will be exemplified as follows.

Example 1. We have hat matrix $\mathbf{X} = \begin{bmatrix} 1 & 2 & 5 \\ 1 & 2 & 5 \\ 1 & 3 & 2 \\ 1 & 3 & 2 \\ 1 & 3 & 2 \\ 1 & 5 & 7 \end{bmatrix}$. We also have $q = 3$

and row vectors $\mathbf{x}_1' = (1\,2\,5)$, $\mathbf{x}_2' = (1\,3\,2)$, $\mathbf{x}_3' = (1\,5\,7)$ with multiplication factors $k_1 = 2$, $k_2 = 3$, $k_3 = 1$. The matrix $\mathbf{H}$ takes the following form here

$\mathbf{H} = \begin{bmatrix} \frac{1}{2}\mathbf{J}_{22} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{1}{3}\mathbf{J}_{33} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 1 \end{bmatrix}$. It is directly signified from the multiplication

factors 2, 3 and 1 of row vectors appearing. In given example the number of linear-independent row vectors is equal to the rank of matrix $\mathbf{X}$.

Example 2. We assume the form of the hat matrix $\mathbf{X} = \begin{bmatrix} 1 & 2 & 5 \\ 1 & 2 & 5 \\ 1 & 3 & 2 \\ 1 & 3 & 2 \\ 1 & 3 & 2 \\ 1 & 5 & 7 \\ 1 & 7 & 6 \end{bmatrix}$. We have $q = 4$ and row vectors $\mathbf{x}_1' = (1\,2\,5)$, $\mathbf{x}_2' = (1\,3\,2)$, $\mathbf{x}_3' = (1\,5\,7)$, $\mathbf{x}_4' = (1\,7\,6)$ adequately with multiplication factors: $k_1 = 2$, $k_2 = 3$, $k_3 = 1$, $k_4 = 1$. The following case takes place here: $q = 4 > 3 = p$. The matrix H with 10×10 measurements in the block form was given, to make things easier and for the sake of symmetry, in the form of the low-triangular matrix.

$$\mathbf{H} = \begin{bmatrix} 0{,}4226\mathbf{J}_{22} & & & \\ 0{,}0301\mathbf{J}_{32} & 0{,}3216\mathbf{J}_{33} & & \\ 0{,}2063\mathbf{J}_{12} & -0{,}0802\mathbf{J}_{13} & 0{,}4498 & \\ -0{,}1418\mathbf{J}_{12} & 0{,}0552\mathbf{J}_{13} & 0{,}3782 & 0{,}7399 \end{bmatrix}.$$

The matrix **H** can be used in economic research with the emerging set of cases, which have the essential significance for the estimation of the linear regression model and for the consideration of them from the side of protruding observations within independent variables. The possibility of the introduction of ordering of diagonal elements of matrix **H** let us treat their values as the ones which depart very little from some concentration of the multivariate sample and remote significantly from the centre of the sample. Such emerged cases can become particularly interesting and always demand making some additional monographic investigations.

4. APPLICATION OF MATRIX H IN ECONOMIC AND TOURIST RESEARCH

Here we will present general hypothetical conceptions of investigations of various authors where, among independent variables, there were also considered such of them which refer directly to tourist research. In the first example there will be only one variable while in the second example there will be all variables. In presented examples we do not give any numerical results because it would lengthen the paper significantly. It will be done this way in another study because it is very difficult to appoint diagonal elements of matrix **H** with the very big n.

Example 1. The paper of Sobczyk (1998) dealt with the characterizing the synthetic meter for investigating of the spatial demand for services. Among many possible diagnostic characteristics the investigation took into consideration:

x_1 – the population size in general in thousands,

x_2 – the urban population size in thousands,

x_3 – the employment in the industry per 1000 people,

x_4 – the inhabited apartments in thousands,

x_5 – the number of TV subscribers per 1000 people,

x_6 – the number of the conduit telephony per 1000 people,

x_7 – passenger cars per 1000 people,

x_8 – lodging places in tourist lodging facilities per 10 thousand people,

x_9 – the purchase of the farm produce per 1 hectare of the farm arable lands per cereals units in dt,

x_{10} – the average gross salary in zloty,

x_{11} – the retail of goods per 1 resident in zloty.

All the above mentioned variables inform us about the intensity of the potential demand for the services. The investigation itself can concern various spatial units i.e. communes, districts, provinces, cities in the segmentation form of the population size (for example cities with the population which totals 20–50 thousands inhabitants), regions and so on.

The paper of G. Sobczyk shows the applications of her own synthetic meter to investigate the diversity of the demand for services in provinces. It would be very interesting to notice what would be such ranking for lever points like and at the same time whether the high lever points occur.

Example 4. In tourist research (Mudambi, Baum 1996) of perception of visiting particular country by tourists from various countries authors investigated the total guest's expenses per one day (variable y) depending on the following independent variables:

x_1 – substitute price, that is the percentage of guests who confirmed that the price was the main factor which influenced the decision about arriving in the particular country,

x_2 – the substitute income, that is the national income per one resident published by World Bank,

x_3 – the average number of nights spent in investigated country during the visit,

x_4 – the percentage of guests who stayed in registered hotels,

x_5 – the percentage of people working in enterprises,

x_6 – the percentage of people working in private companies

x_7 – the percentage of tourists on leave,

x_8 – the percentage of guests in business,

x_9 – the average age of tourists,

x_{10} – the artificial variables for Islamic countries.

In such investigation we can consider every country which has well-organized statistical service which can provide numerical data relating to the mentioned variables. Quoted authors made such investigations on the tourist market in Turkey in years 1988–1993 with the application of personal data forms.

Using matrix H in this problem we could identify which countries make possible sets of protruding observations in the set of independent variables. At the same time, we can also notice whether high lever cases could influence the quality of the estimation of the regression model, and as the further consequence, making decision about the tourist motion to Turkey.

REFERENCES

Anderson T. W. (1958), *An Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, New York.

Arnold S. F. (1981), *The Theory of Linear Models and Multivariate Analysis*, John Wiley and Sons, New York.

Cook R. D., Weisberg S. (1982), *Residuals and Influence in Regression*, Chapman and Hall, New York.

Haoglin D. C., Welsch R. E. (1978), "The Hat Matrix in Regression and ANOVA", *American Statistician*, **32**, 17–22.

Kolupa M., Witkowski J. M. (1981), *Wybrane metody numeryczne algebry liniowej w ekonometrii*, Państwowe Wydawnictwo Naukowe, Warszawa.

Mudambi R., Baum T. (1996), "Strategic Segmentation: An Empirical Analysis of Tourist Expenditure in Turkey", *Journal of Travel Research.*

Oktaba W. (1986), *Metody matematyczne w doświadczalnictwie*, Państwowe Wydawnictwo Naukowe, Warszawa.

Rao C. R. (1982), *Modele liniowe statystyki matematycznej*, Państwowe Wydawnictwo Naukowe, Warszawa.

Rousseuw P. J., Leroy A. M. (1987), *Robust Regression and Outlier Detection*, John Wiley and Sons, New York.

Sobczyk G. (1998), "Syntetyczny miernik w badaniach przestrzennego zróżnicowania popytu na usługi", *Wiadomości Statystyczne*, **6**(445), 48–56.

Wiesław Wagner

MACIERZ H I JEJ ZASTOSOWANIA W BADANIACH EKONOMICZNYCH I TURYSTYCZNYCH

(Streszczenie)

W artykule określono i podano podstawowe własności macierzy **H**, będącej operatorem ortogonalnego rzutu na przestrzeń kolumn macierzy **X**, która jest macierzą układu w modelu regresji liniowej. Wymienione zostały własności elementów diagonalnych tej macierzy, które zostały określone jako punkty dźwigniowe i wysoko dźwigniowe punkty. Także zajęto się postacią macierzy **H** przy powtarzających się obserwacjach odnoszących się do zmiennych niezależnych. Pewne zastosowania macierzy **H** podano na dwóch przykładach.