UNIVERSITATIS IAGELLONICAE ACTA MATHEMATICA, FASCICULUS XLIII
2005

THE MULTIPLE REGRESSION MODEL WHERE INDEPENDENT VARIABLES ARE MEASURED UNPRECISELY

by Anna Czapkiewicz and Antoni L. Dawidowicz

Abstract. We present a multiple regression model where all independent variables are subject to error. We suggest a method to the estimation of unknown parameter. The estimators constructed by this method are unbiased and its covariance matrix is minimal in the class of linear unbiased estimators.

Introduction. In the classical Gauss–Markov linear regression model it is assumed, that independent variables are measured precisely. The error of measurement is connected with the dependent variable only. The unknown parameters in this model are estimated by the ordinary least squares method, which gives unbiased estimators. In this paper we discuss the situation where the dependent variable as well as the independent variables are perturbed by errors of measure.

We discuss the model

$$X^t = S^t + \varepsilon^t \;\; t = 1, \dots, p,$$

$$Y = S\beta + \delta,$$

where $X^t = [x_1^t, \dots, x_n^t]^T$, $Y = [y_1, \dots, y_n]^T$ are vectors of observables, the matrix S has the form

$$S = \left[S^1, S^2, \dots, S^p, 1_n\right],$$

$$S^t = [s_1^t, \dots s_n^t]^T \quad t = 1, \dots, p,$$

s_i^t is unknown deterministic variable,
$\beta = [\beta_1, \dots, \beta_p, \beta_{p+1}]^T$ is a vector of unknown parameters.

Furthermore

$$\varepsilon^t = [\varepsilon_1^t, \ldots, \varepsilon_n^t]^T \quad \varepsilon_i^t \sim N(0, \sigma_{\varepsilon_t}^2) \quad E(\varepsilon_i^t \varepsilon_j^t) = 0 \quad i \neq j,$$

$$\delta = [\delta_1, \ldots, \delta_n]^T \quad \delta_i \sim N(0, \sigma_\delta^2) \quad E(\delta_i \delta_j) = 0 \quad i \neq j,$$

$$E(\varepsilon_i^t \delta_i) = 0, \quad t = 1, \ldots, p.$$

It is easy to show that this model, with errors having normal distributions with unknown variances, is nonidentifiable. This fact may be proved by a method analogous to that described in the Reiersol's paper [**11**]. To overcome this difficulty it is assumed that either one error variance is known or the ratio of the variances is known. Such situations are described, for example, in Kendall and Stuart [**8**] or Fuller [**6**].

In this paper we present another approach to the construction of consistent estimators of regression slopes. This method enables us to create estimator of β which is unbiased with minimal covariance matrix in the class of linear unbiased estimators, i.e. for another linear unbiased estimator its covariance matrix is greater in the sense of order defined by positive definity. In the case when parameters of normal distribution are unknown, such defined models are identifiable.

We discuss a model

$$\begin{aligned} X_i^t &= s_i^t + \varepsilon_i^t, \quad \varepsilon_i \sim N(0, \sigma_{\varepsilon_t}^2), \\ Y_{ij} &= S_i \beta + \delta_{ij}, \quad \delta_{ij} \sim N(0, \sigma_\delta^2), \\ S_i &= [s_i^1, \ldots, s_i^p, 1]^T, \quad i = 1 \ldots n, \ \ j = 1, \ldots m. \end{aligned} \tag{1}$$

In the case of the simple regression model, the literature presents models where both independent and dependent variables are repeated m times (Cox [**2**], Dolby [**5**], Bunke and Bunke [**1**]). In this model the maximum likelihood method gives desired properties of estimators. Applying the maximum likelihood method to model (1) poses technical problems. We construct the estimators based on the variance components theory.

The matter of paper is presented in the following phases. The method of variance components is used to the multiple regression model, where only one independent variable is measured unprecisely. Finally the same procedure is used for the model where all independent variables are perturbed. Unfortunately, in this model it is impossible to estimate the variance of each independent variable separately.

1. Variance components estimation method. Let $Y = X\beta + U_1\Phi_1 + U_2\Phi_2$ be the general linear model with two variance components, where Y is a vector of observables, the matrix X is known, β is an unknown vector of parameters and Φ_1, Φ_2 are random vectors such that

$$E(\Phi_1) = 0, \quad E(\Phi_2) = 0, \quad E(\Phi_1 \Phi_2^T) = 0.$$

Furthermore

$$U_1^T U_1 = V, \quad U_2^T U_2 = I.$$

Let us define the matrix W as

$$W = BVB^T,$$

where

$$BB^T = I, \;\; B^T B = M,$$
$$M = I - XX^+,$$

where X^+ denotes the Moore–Penrose inverse matrix. We recall one of general theorem (Gnot [**7**])

Theorem 1. *In the model with two components where the matrix W has two different eigenvalues and W is singular, the best unbiased and invariant estimator of linear combination of variance components $f_1\sigma_1 + f_2\sigma_2$ has the form*

$$[f_1/\alpha_1^2\nu_1 - (\alpha_1 f_2 - f_1)/\alpha_1^2\nu_2]Y^T MVMY + [(\alpha_1 f_2 - f_1)/\alpha_1\nu_2]Y^T MY,$$

where α_1 is the unique nonzero eigenvalue of W with multiplicity ν_1 and ν_2 is the multiplicity of the zero eigenvalue of W.

We use the result of this theorem to the following model.

Let have

$$Z_i = s_i + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma_\varepsilon^2),$$
$$Y_{ij} = \beta_o s_i + X_i^T \beta^* + \delta_{ij} \;\; \delta_{ij} \sim N(0, \sigma_\delta^2).$$

Z_i is an perturbed observable, the vector

$$X_i = [x_i^1, \ldots, x_i^{p-1}, 1], \;\; i = 1, \ldots, n$$

is a deterministic vector of observables.

If we substitute s_i in the last formula we obtain

$$Y_{ij} = \beta_o Z_i + X_i^T \beta^* + \gamma_i + \delta_{ij},$$

where $\gamma_i = -\beta_o \varepsilon_i$.

Replacement of the distribution of (Z_i, Y_{ij}) with the conditional distribution of Y_{ij} with respect to Z_i enables us to use a different model (treating Z_i as a nonrandom value) to estimate the same parameters β. We obtain the model

$$Y = X\beta + U_1\Phi_1 + U_2\Phi_2, \tag{2}$$

where $\beta = [\beta_o, \beta^*]^T$ is a vector of unknown parameters,

$$Y = [y_1^T \ldots, y_n^T]^T \quad y_i = [Y_{i1}, \ldots Y_{im}]^T,$$

$$X = \begin{bmatrix} Z_1 1_m & x_1^1 1_m & \dots & x_1^{p-2} & 1_m \\ Z_2 1_m & x_2^1 1_m & \dots & x_2^{p-2} & 1_m \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Z_n 1_m & x_n^1 1_m & \dots & x_n^{p-2} & 1_m \end{bmatrix},$$

where $1_m = [1, \dots, 1]^T$.
The matrices U_1 , U_2 have the forms

$$U_1 = I_n \otimes 1_m, \quad U_2 = I_{mn}, \tag{3}$$

where $A \otimes B$ denotes the Kronecker product of the matrices A and B. The vectors Φ_1, Φ_2 are given as follows

$$\Phi_1 = [\gamma_1, \dots, \gamma_n]^T,$$

$$\Phi_2 = [\phi_1^T, \dots, \phi_n^T]^T \text{ where } \phi_i = [\delta_{i1}, \dots \delta_{im}]^T.$$

Furthermore, we may notice that

$$E(\Phi_i) = 0, \quad E(\Phi_i \Phi_j^T) = 0, \quad E(\Phi_i \Phi_j^T) = \sigma_i^2 I, \quad i, j = 1, 2.$$

The variance components are

$$\sigma_1^2 = a^2 \sigma_\varepsilon^2 \text{ and } \sigma_2^2 = \sigma_\delta^2. \tag{4}$$

THEOREM 2. *The uniformly best invariant unbiased estimators of σ_1^2 and σ_2^2 in this model are*

$$\widetilde{\sigma_1^2} = \frac{nm - (p+1)}{m^2(n-p-1)(m-1)n} Y^T MVMY - \frac{1}{mn(m-1)} Y^T MY,$$

$$\widetilde{\sigma_2^2} = \frac{1}{n(m-1)} Y^T MY - \frac{1}{mn(m-1)} Y^T MVMY,$$

where

$$M = I - X(X^T X)^{-1} X^T.$$

PROOF. Let B be $(nm - p - 1) \times nm$-dimensional matrix defined so that

$$BB^T = I_{nm-p-1} \quad B^T B = I - XX^+ = M$$

and

$$W = BVB^T \quad \text{where} \quad V = U_1 U_1^T.$$

Because $VV = mV$ and $X^T V = mX^T$ we may notice that

$$W^2 = BVB^T BVB^T = BVMVB^T = BV(I - X(X^T X)^{-1} X^T) VB^T$$
$$= mBVMB^T = mBVB^T BB^T = mBVB^T.$$

Thence and because W is a symmetric matrix, W has the unique nonzero eigenvalue $-m$.

The calculation of the trace of the matrix MV gives us the multiplicity of this eigenvalue. The matrix X has a decomposition

$$X = C\Lambda D^T,$$

where Λ is a diagonal matrix, and C and D are such matrices that $C^T C = I$ and $D^T D$=I. Let $\alpha_1, \dots \alpha_{p+1}$ be eigenvalues of XX^T. From the definition of the Moore–Penrose inverse matrix, we infer that $XX^+ = CC^T$, where C is formed by normalized eigenvectors corresponding to $\alpha_1, \dots \alpha_{p+1}$.

Applying those remarks, we can calculate the trace of MV. It equals $m(n-p-1)$.

Matrix W is singular with two eigenvalues: 0 with multiplicity $n(m-1)$ and m with multiplicity $(n-p-1)$. The conditions of Theorem 2 are satisfied, so we have the desired formulas for the estimators of the variance components σ_1^2 i σ_2^2. The expression for M follows from the fact that X has full rank. So

$$M = I - X(X^T X)^{-1} X^T.$$

□

Having calculated formulas for $\widetilde{\sigma_1^2}$ and $\widetilde{\sigma_2^2}$, we may construct an estimator for β. It is natural to take

$$\tilde{\beta} = [X^T \tilde{Z}^{-1} X]^{-1} X^T \tilde{Z}^{-1} Y, \tag{5}$$

where $\tilde{Z} = \widetilde{\sigma_1^2} V + \widetilde{\sigma_2^2} I_{nm}$.

Theorem 3. *The estimators of unknown parameters β based on variance components theory have the following properties:*
(i) The estimator of $\tilde{\beta}$ in (5) does not depend on values of $\tilde{\sigma}_1$ and $\tilde{\sigma}_2$ and has the form

$$\tilde{\beta} = (X^T X)^{-1} X^T Y.$$

(ii) The estimator $\tilde{\beta}$ has the normal distribution with expectation β and with covariance matrix

$$(m\sigma_1 + \sigma_2)(X^T X)^{-1} = (ma^2\sigma_\epsilon^2 + \sigma_\delta^2)(X^T X)^{-1}. \tag{6}$$

(iii) The estimator $\tilde{\beta}$ is unbiased with minimal covariance matrix in the class of linear unbiased estimators, the estimator $\tilde{\sigma_\delta}$ is the uniformly best unbiased estimator of σ_δ and the estimator

$$\tilde{\sigma}_\epsilon = \sqrt{\frac{\tilde{\sigma}_2^2}{\tilde{\beta_o}^2}}$$

is weakly consistent.

Proof. We can notice, after simple calculation, that for every p and q there is

$$X^T(pV + qI_{mn}) = (mp + q)X^T. \tag{7}$$

(i) From (7) there follows that

$$X^T\tilde{Z} = X^T(\tilde{\sigma_1}^2 V + \tilde{\sigma}_2^2 I_{mn}) = (m\tilde{\sigma}_1^2 + \tilde{\sigma}_2^2)X^T,$$

thus

$$X^T = (m\tilde{\sigma}_1^2 + \tilde{\sigma}_2^2)X^T(\tilde{Z})^{-1}$$

and

$$\tilde{\beta} = (m\tilde{\sigma}_1^2 + \tilde{\sigma^2}_2)(X^TX)^{-1}\frac{1}{m\tilde{\sigma}_1^2 + \tilde{\sigma}_2^2}X^TY = (X^TX)^{-1}X^TY.$$

(ii) As we have assumed that Y has a normal distribution, then the estimator $\tilde{\beta}$, as a linear function of Y, also has a normal distribution. The expectation of $\tilde{\beta}$ is

$$E\left((X^TX)^{-1}X^TY\right) = (X^TX)^{-1}X^TE(Y) = (X^TX)^{-1}X^TX\beta = \beta.$$

The covariance matrix of $\tilde{\beta}$, var$(\tilde{\beta})$ is:

$$\begin{aligned} Var\left(\tilde{\beta}\right) &= E\left((X^TX)^{-1}X^TYY^TX(X^TX)^{-1}\right) - \beta\beta^T \\ &= (X^TX)^{-1}X^TE(YY^T)X(X^TX)^{-1} - \beta\beta^T. \end{aligned}$$

Because

$$E(YY^T) = \sigma_1^2V + \sigma_2^2I + (X\beta)(X\beta)^T,$$

from (7), we receive that

$$Var(\tilde{\beta}) = (m\sigma_1^2 + \sigma_2^2)(X^TX)^{-1}.$$

Let us consider another linear unbiased estimator L^TY of β. We can prove that covariance matrix of estimator L^TY is not smaller than the covariance matrix of estimator $(X^TX)^{-1}X^TY$ (i.e., the difference between these matrices is non-negative defined). Let us put A^T as

$$A^T = L^T - (X^TX)^{-1}X^T.$$

We can notice that $E(A^TY) = 0$. It implies that also $A^TX = 0$. Due to this fact (7) there is

$$\begin{aligned} &E\left(A^TY((X^TX)^{-1}X^TY)^T\right) = A^TE(YY^T)X(X^TX)^{-1} \\ &\quad = A^T(\sigma_1^2V + \sigma_2^2I_{mn})X(X^TX)^{-1} = (m\sigma_1^2 + \sigma_2^2)A^TX(X^TX)^{-1} = 0. \end{aligned} \tag{8}$$

Now we shall present $Var(L^TY)$ as

$$Var\left(L^TY\right) = Var\left(L^TY - (X^TX)^{-1}X^TY + (X^TX)^{-1}X^TY\right).$$

By equation (8), we may write $Var(L^TY)$ as

$$Var(L^TY) = Var(L^TY - (X^TX)^{-1}X^TY) + Var((X^TX)^{-1}X^TY).$$

The first component is a non-negative defined matrix, so

$$Var\left(L^TY\right) \geq Var\left((X^TX)^{-1}X^TY\right).$$

The properties of $\tilde{\sigma_\delta}$ follow from Theorem 2, while the properties of $\tilde{\sigma_\epsilon}$ follow from the Słucki Theorem. □

REMARK 1. For a simple regression model we can notice that variances of estimators $\tilde{\beta_o}$ and $\tilde{\beta_p}$ using variance components theory have the forms

$$\mathrm{var}(\tilde{\beta_o}) = \frac{m\beta_o^2\sigma_\varepsilon^2 + \sigma_\delta^2}{m\sum_{i=1}^n (Z_i - \bar{Z})^2},$$

$$\mathrm{var}(\tilde{\beta_p}) = \frac{m\beta_o^2\sigma_\varepsilon^2 + \sigma_\delta^2}{mn} \frac{\sum_{i=1}^n Z_i^2}{\sum_{i=1}^n (Z_i - \bar{Z})^2}.$$

For simple regression variances of β_o and β_p, the maximum likelihood method and the theory of variance components are comparable.

2. The model where all independent variables are perturbed. This last method allow us to generalize our results to a multiple regression model where independent variables are subject to error. The model has the following form

$$x_i^t = s_i^t + \varepsilon_i^t \quad \varepsilon_i \sim N(0, \sigma_{\varepsilon_t}^2),$$
$$Y_{ij} = S_i\beta + \delta_{ij} \quad \delta_{ij} \sim N(0, \sigma_\delta^2),$$
$$S_i = [s_i^1, \ldots, s_i^p, 1]^T, \quad i = 1 \ldots n, \;\; j = 1, \ldots m.$$

If we substitute s_i in the last formula, we obtain

$$Y_{ij} = X_i^T\beta + \gamma_i + \delta_{ij},$$

where $\gamma_i = -\varepsilon_i^T\beta$.

$$X_i = [x_i^1, \ldots, x_i^p], \quad \varepsilon_i = [\varepsilon_i^1, \ldots, \varepsilon_i^p].$$

We received two components $\sigma_1^2 = \sum_{t=1}^p \beta_t\sigma_{\varepsilon_t}^2$ and $\sigma_2^2 = \sigma_\delta^2$. We can use the results of previous section to model thus defined. In the case when all independent variable are subject to error, repeating the dependent variable allows us to estimate unknown parameters β and unknown variance σ_δ^2 and the sum of unknown variances $\sigma_{\varepsilon_t}^2$. It is impossible to estimate each $\sigma_{\varepsilon_t}^2$ separately.

References

1. Bunke O., Bunke H., *Non-Linear Regression, Functional Relationships and Robust Methods*, Wiley, New York, 1989.
2. Cox N.R., *The linear structural relation for several groups of data*, Biometrika, **63** (1976), 231–237.
3. Czapkiewicz A., Dawidowicz A.L., *On estimation of parameters in the bivariate errors-in-variables model*, Appl. Math., **4** (1999), 401–410.
4. Czapkiewicz A., *The role of variance components in the model of dependence between random variables. Application of selected models in economy,* Uniwersytet Łódzki, Ph.D. Dissertation (in Polish) (2001).
5. Dolby G.R., *The ultrastructural relation a synthesis of the functioanal and structural relations*, Biometrika, **63** (1976), 39–50.
6. Fuller W.A., *Measurement Error Models*, Wiley, New York, 1987.
7. Gnot S., *Estimation of Variance Components in Linear Models,* Wyd. Naukowo–Techniczne, Warszawa, 1991 (in Polish).
8. Kendall M.G., Stuart A., *The advanced theory of statistics,* Vol. 2, Griffin, London, 1979.
9. Lehmann E.L., *Theory of Point Estimaion*, Wiley, New York, 1983.
10. Rao C.R., Kleffe R., *Estimation of Variance Components*, North-Holland Ser. Statist. Probab. 3, North-Holland, Amsterdam, 1988.
11. Reiersol O., *Identifiability of a linear relation between variables which are subject to error,* Econometrica, **18** (1950), 575–589.

Received June 2, 2002

AGH University of Science and Technology
Institute of Managing
ul. Gramatyka 10
30-067 Kraków
Poland
e-mail: gzrembie@cyf-kr.edu.pl

Jagiellonian University
Institute of Mathematics
ul. Reymonta 4
30-059 Kraków
Poland
e-mail: Antoni.Leon.Dawidowicz@im.uj.edu.pl