APPLICATIONES MATHEMATICAE
25,4 (1999), pp. 401–410

A. CZAPKIEWICZ (Kraków)

ON ESTIMATION OF PARAMETERS IN THE BIVARIATE LINEAR ERRORS-IN-VARIABLES MODEL

Abstract. We discuss some methods of estimation in bivariate errors-in-variables linear models. We also suggest a method of constructing consistent estimators in the case when the error disturbances have the normal distribution with unknown parameters. It is based on the theory of estimating variance components in linear models. A simulation study is presented which compares this estimator with the maximum likelihood one.

1. Introduction. Simple linear regression models describe linear functional relationships when there is an observation error in only the dependent variable Y. The X (independent variable) is assumed to be measured precisely. The parameters in this model are estimated by the classical ordinary least squares (OLS) method, which gives unbiased estimators.

In situations where both variables are subject to error the errors-in-variables models are applied. We assume that the true linear relationship is given by $y = as+b$. The actual observed values are $X = s+\varepsilon$ and $Y = y+\delta$, where the symbols ε and δ denote the corresponding error disturbances. The measurement error in X sometimes happens to be overlooked and the OLS estimation of the parameters is chosen because of its familiarity and ease of use. It is well known that in the errors-in-variables model, the simple regression estimator (OLS) is inconsistent.

Cochran (1968) has given a general discussion of the consequences of using the OLS estimator in errors-in-variables models. However, there exist cases when this estimator has mean squared error smaller than other esti-

1991 *Mathematics Subject Classification*: Primary 62J05.
Key words and phrases: regression line, replication model, variance component, estimator, maximum likelihood method, simulation study.

mators described in the literature (Kettelapper (1983), Penev and Raykov (1993)).

Different aspects, such as estimation, identifiability, robustness, asymptotic and small-sample properties in linear errors-in-variables models have aroused considerable interest (e.g., Kendall and Stuart (1979), Fuller (1987), Bunke and Bunke (1989)).

The parameters a and b are often not identifiable (Reiersol (1950)). For example, they are not identifiable in the case where errors have a joint normal distribution with unknown variances. However, a has been shown to be identifiable under various sets of assumptions, for example: the variance of one disturbance is known or, more commonly, the variance ratio λ is known. There are several methods of estimating a and b in such cases. From the practical point of view, we are most interested in an approach which leads to estimating the parameters in the case when all $a, b, s, \sigma_\varepsilon^2, \sigma_\delta^2$ have to be considered as unknown. In that case, the replication of measurement of each pair of observations (X_i, Y_i) m_i times overcomes the nonidentifiability (Bunke and Bunke (1989)) and enables us to construct consistent estimators.

In this paper an approach is proposed which enables us to obtain identifiability by repeating measurements of only one variable, for example Y_i. Useful estimators of regression slopes are constructed. The theory of estimation of variance components in linear models (Rao and Kleffe (1988), Gnot (1991)) is applied. A simulation study is presented which compares this estimator with the maximum likelihood one.

2. Methodology. Consider the model

$$X_i = s_i + \varepsilon_i, \quad Y_{ij} = as_i + b + \delta_{ij}, \quad i = 1, \ldots, n, \ j = 1, \ldots, m_i, \tag{1}$$

where n is the sample size. Assume that s_i is an unknown constant and ε_i and δ_{ij} are random variables with mean zero and variances σ_ε^2 and σ_δ^2, respectively. We consider the situation when

$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2), \qquad \delta_{ij} \sim N(0, \sigma_\delta^2).$$

2.1. *The maximum likelihood estimator of* $\widehat{a}$. In this model the obvious method of estimating the unknown parameters is the maximum likelihood procedure. This provides estimators which, under quite general regularity conditions, are consistent, asymptotically efficient and asymptotically normal.

The likelihood method for model (1) requires finding a maximum for a function of $n + 4$ variables.

Let us describe the likelihood function. Let $z_{ij} = (X_i, Y_{ij})'$ and

$$\Sigma = \begin{bmatrix} \sigma_\varepsilon^2 & 0 \\ 0 & \sigma_\delta^2 \end{bmatrix}.$$

Further, let

$$\mu_i = (s_i, as_i + b)'.$$

The log-likelihood function is then

$$l(a, b, s_1, \ldots, s_n, \Sigma) = -\frac{m}{2} \log \det[\Sigma] - \frac{1}{2} \sum_i \sum_j (z_{ij} - \mu_i)' \Sigma^{-1} (z_{ij} - \mu_i).$$

This model is a particular case of the situation when both X_i and Y_i are repeated m_i times. In this case $z_{ij} = (X_{ij}, Y_{ij})$. The problem of finding a maximum is hard both from the theoretical and computational point of view. It was discussed by Cox (1976). In the more general case, i.e. when both observables X_i, Y_i are repeated m_i times, the maximum is realized by a solution of an equation of the fourth degree in a.

Let

$$W_Z = \sum_i \sum_j (z_{ij} - z_{i.})(z_{ij} - z_{i.})'$$

and

$$B_Z = \sum_i m_i (z_{i.} - \bar{z}_{..})(z_{i.} - \bar{z}_{..})'.$$

Using the notations

$$W_Z = \begin{bmatrix} w_x & w_{xy} \\ w_{xy} & w_y \end{bmatrix} \quad \text{and} \quad B_Z = \begin{bmatrix} b_x & b_{xy} \\ b_{xy} & b_y \end{bmatrix}$$

we can write the equation for $\widehat{a}$ as

$$\text{(2)} \quad [aw_x(b_y - ab_{xy}) - w_y(ab_x - b_{xy})]B(a) \\ -(b_y - ab_{xy})(ab_{xy})(b_y - a^2 b_x) = 0$$

where $B(a) = b_y - 2ab_{xy} + a^2 b_x$.

The estimator of $\widehat{b}$ is

$$\widehat{b} = \bar{Y}_{..} - \widehat{a}\bar{X}_{..}.$$

The form of $\widehat{a}$ is not simple. If we consider the situation where $X_{ij} = X_i$ for each j, we also have a cumbersome equation of the fourth degree in a.

2.2. *Estimation of the parameters of linear regression using the variance components estimation method.* In this section we suggest another approach to estimating the regression slopes.

Consider a model (1),

$$X_i = s_i + \varepsilon_i, \quad Y_{ij} = as_i + b + \delta_{ij}.$$

To simplify the calculations we assume that $m_i = m$ for each i. By substituting s_i in the last relation we obtain

$$Y_{ij} = aX_i + b + \delta_{ij} + \gamma_i \tag{3}$$

where $\gamma_i = -a\varepsilon_i$.

Replacement of the distribution of (X, Y) by the conditional distribution of Y with respect to X enables us to use a different model (treating X_i as a constant) to estimate the same parameters $a, b, \sigma_\varepsilon^2, \sigma_\delta^2$. The technique of variance components can be applied for this purpose.

The general linear model with two variance components has the form

$$Y = X\beta + U_1\Phi_1 + U_2\Phi_2, \tag{4}$$

where Y is a k-dimensional vector of observation, X is a known $k\times p$ matrix with $\operatorname{rank}(X) = p$, β is a vector of unknown constants, U_1, U_2 are known matrices and Φ_1, Φ_2 are non-observable random vectors satisfying

$$E(\Phi_i) = 0, \quad E(\Phi_i\Phi_j') = 0, \quad E(\Phi_i\Phi_i') = \sigma_i^2 I_{t_i}, \quad i, j = 1, 2.$$

The expectation vector and the variance matrix of Y are

$$E(Y) = X\beta, \quad \operatorname{Var}(Y) = \sigma_1^2 V_1 + \sigma_2^2 V_2, \quad V_i = U_iU_i', \quad i = 1, 2. \tag{5}$$

In this section we are basing on the following theorems.

Let $\sigma = (\sigma_1, \sigma_2)$ and $f'\sigma = f_1\sigma_1^2 + f_2\sigma_2^2$ be a linear combination of variance components and let A be a symmetric $k\times k$ matrix. The quadratic form $y'Ay$ is an unbiased estimator of $f'\sigma$ if

$$E(y'Ay) = f'\sigma$$

for each β and σ.

Consider the situation when V_2 in (5) is the $k \times k$ unit matrix. Let B be a $(k - p) \times k$ matrix satisfying

$$BB' = I_{k-p}, \tag{6}$$

$$B'B = I - XX^+ = M. \tag{7}$$

Such a B always exists, although it is not unique.

Define a matrix W as

$$W = BVB'.$$

THEOREM 1. *In the linear model with two variance components σ_1^2, σ_2^2 all the functions $f'\sigma$ are invariantly estimated if the number of different eigenvalues of W is greater than one.*

A specially interesting case is when W has exactly two distinct eigenvalues and is singular. Then we have the following theorem:

THEOREM 2. *In the model with two components where the matrix W has two different eigenvalues and W is singular the best local unbiased and invariant estimator of $f'\sigma$ has the form*

$$\left(\frac{f_1}{\alpha_1^2\nu_1} - \frac{\alpha_1 f_2 - f_1}{\alpha_1^2\nu_2}\right)Y'MVMY + \frac{\alpha_1 f_2 - f_1}{\alpha_1\nu_2}Y'MY, \tag{8}$$

where α_1 is the unique nonzero eigenvalue of W with multiplicity ν_1 and $\nu_2 = k - p - \nu_1$ is the multiplicity of the zero eigenvalue of W.

In particular, for the variance components σ_1^2 and σ_2^2 the uniformly best estimators are

$$\widehat{\sigma_1^2} = \frac{\nu_1 + \nu_2}{\alpha_1^2\nu_1\nu_2}Y'MVMY - \frac{1}{\alpha_1\nu_2}Y'MY,$$
$$\widehat{\sigma_2^2} = \frac{1}{\nu_2}Y'MY - \frac{1}{\alpha_1\nu_2}Y'MVMY.$$

The proofs of these theorems are given by Gnot (1991).

In the model (3) we obtain two variance components:

$$\sigma_1^2 = a^2\sigma_\varepsilon^2, \qquad \sigma_2^2 = \sigma_\delta^2. \tag{9}$$

The vector Y has the form

$$Y = [y_1', \ldots, y_n']', \qquad y_i = [Y_{i1}, \ldots Y_{im}]',$$

and the matrix X is

$$X = \begin{bmatrix} X_1 & 1 \\ \vdots & \vdots \\ X_n & 1 \end{bmatrix},$$

where each X_i is repeated m times.

The matrix U_1 is

$$U_1 = I_n \otimes \Delta \tag{10}$$

where Δ' is an m-dimensional vector of 1's: $\Delta = (1, \ldots, 1)$. The matrix U_2 is the $nm \times nm$ unit matrix, and

$$\Phi_1 = [\gamma_1, \ldots, \gamma_n]', \qquad \Phi_2 = [\phi_1^{*\prime}, \ldots, \phi_n^{*\prime}]'$$

where $\phi_i^* = [\delta_{i1}, \ldots \delta_{im}]'$.

From the considerations in the previous sections we have

$$E(\Phi_i) = 0, \quad E(\Phi_i \Phi_j') = 0, \quad E(\Phi_i \Phi_i') = \sigma_i^2 I_{mn}, \quad i, j = 1, 2.$$

With the above notations, we have the following theorem.

THEOREM 3. *The best local, invariant, unbiased estimators of* σ_1^2 *and* σ_2^2 *in* (3) *are*

$$\widehat{\sigma_1^2} = \frac{nm-2}{m^2(n-2)(m-1)n} Y'MVMY - \frac{1}{mn(m-1)} Y'MY,$$
$$\widehat{\sigma_2^2} = \frac{1}{n(m-1)} Y'MY - \frac{1}{mn(m-1)} Y'MVMY,$$

where

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

To prove this theorem we use the following simple lemmas:

LEMMA 1. *Let* A *and* B *be* $n \times m$ *and* $m \times n$ *matrices, respectively. Then* AB *and* BA *have the same nonzero eingenvalues.*

LEMMA 2. *Let* A *be a symmetric matrix. If* $A^2 = mA$, *then* m *is a unique nonzero eigenvalue of* A.

LEMMA 3. *Let* A *and* B *be matrices such that* AB *and* BA *are symmetric. If* λ *is a unique nonzero eigenvalue of* AB, *of multiplicity* k, *then* λ *is an eigenvalue of* BA *of multiplicity* k.

Proof. We can view the matrices A and B as linear maps $A : \mathbb{R}^m \to \mathbb{R}^n$ and $B : \mathbb{R}^n \to \mathbb{R}^m$. By assumption,

$$\mathbb{R}^n = V_\lambda + V_0$$

where V_λ and V_0 are some subspaces and

$$ABv = \lambda v, \quad v \in V_\lambda,$$
$$ABv = 0, \quad v \in V_0.$$

Then $\mathbb{R}^m = B(V_\lambda) + B(V_0) + B(\mathbb{R}^n)^\perp$ and

$$BAw = \lambda w, \quad w \in B(V_\lambda),$$
$$BAw = 0, \quad w \in B(V_0).$$

If $w \in B(\mathbb{R}^n)^\perp$ then $B^T w = 0$ and $A^T B^T w = 0$ so $BAw = 0$. This implies that

$$BAw = \lambda w \quad \text{if } w \in B(V_\lambda),$$
$$BAw = 0 \quad \text{if } w \in B(V_\lambda)^\perp.$$

Let k' be the dimension of $B(V_\lambda)$. Then λ is a unique nonzero eigenvalue of BA of multiplicity k' and $k' \le k$.

We repeat this reasoning for AB and conclude that λ is an eigenvalue of AB of multiplicity $k'' \le k' \le k$. But the assumption $k'' = k$ implies that $k' = k$. ∎

Proof of Theorem 3. Let B be an $(nm-2) \times nm$ matrix satisfying (6) and (7), and let $W = BVB'$ and $V = U_1 U_1'$ where U_1 is given by (10). Then

$$W = BVB' = BU_1U_1'B = BU_1(BU_1)'.$$

Let

$$W^* = (BU_1)'BU_1 = U_1'B'BU_1 = U_1'MU_1.$$

Lemma 1 implies that W and W^* have the same nonzero eigenvalues.

Let X have the decomposition

$$X = C\Lambda D' \tag{11}$$

where Λ is a diagonal matrix, and C and D are matrices such that $C'C = I$ and $D'D = I$ (Rao (1973)). Let α_1 and α_2 be nonzero eigenvalues of XX'. From (11) and from the definition of the Moore–Penrose inverse matrix we have $XX^+ = CC'$. The matrix C is formed by normalized eigenvectors corresponding to α_1 and α_2.

It is easy to see that $W^{*2} = mW^*$ (the proof is algebraic). From Lemma 2 we know that m is a unique nonzero eigenvalue of W^*. The trace of W^* is $m(n-2)$ so we have the multiplicity of m. From Lemma 3 we see that m is a unique nonzero eigenvalue of W of multiplicity $n-2$. So W is singular and has two distinct eigenvalues: zero with multiplicity $n(m-1)$ and m with multiplicity $n-2$. The conditions of Theorem 2 are satisfied so we have the desired formulas on the estimators of the variance components σ_1^2 and σ_2^2.

The expression for M follows from the fact that X has full rank.

Theorem 4. *Let $\beta = [a, b]$. Then*

$$\tilde{\beta} = [X'\widetilde{Z}^{-1}X]^{-1}X'\widetilde{Z}^{-1}Y$$

where $\widetilde{Z} = \widehat{\sigma_1^2}V + \widehat{\sigma_2^2}I_{nm}$ is an estimator of regression slopes consistent in the quadratic mean.

Proof. Let us write the model (3) as

$$Y_{ij} = aX_i + b + z_{ij} \quad \text{where} \quad z_{ij} = \delta_{ij} - a\varepsilon_i.$$

We note that the random variables z_{ij} are correlated. It is easy to notice that the variance and covariance matrix has the form

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

Now, we can estimate the parameters a and b by

$$\widetilde{\beta} = [X'Z^{-1}X]^{-1}X'Z^{-1}Y. \tag{12}$$

For Z we take the estimator $\widetilde{Z} = \widehat{\sigma_1^2}V + \widehat{\sigma_2^2}I_{nm}$.

The estimators $\widehat{\sigma_1^2}$ and $\widehat{\sigma_2^2}$, having minimal variance, are convergent in the quadratic mean. The relation (12) can be approximated by a linear one. Since $\widehat{\sigma_1^2}$ and $\widehat{\sigma_2^2}$ are convergent in the mean, so are linear functions of them. So we have

$$\lim_{n,m\to\infty} E(|\beta - \widehat{\beta}|^2) = 0.$$

The estimators of the variances σ_ε^2 and σ_δ^2 can be calculated from (9).

3. Simulation studies. Let $\widetilde{a}$ be an estimator constructed using variance components theory (VCE) and $\widehat{a}$ be a maximal likelihood estimator (MLE). Simulation studies have been carried out to illustrate the properties of $\widetilde{a}$ by comparison with the mean squared errors of the above estimators.

We fix theoretical values of the regression coefficients a and b. Using a standard simulator of normal distribution we have

$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2), \qquad \delta_{ij} \sim N(0, \sigma_\delta^2).$$

These were used to generate the "observables" (X_i, Y_{ij}) according to (1). We assumed that the values of s_i are some random real variables with values in $(i, i+1)$, where i are successive natural numbers. Then $\widetilde{a}$ is estimated as in Theorem 4 (where $\widetilde{\beta} = [\widetilde{a}, \widetilde{b}]$) and $\widehat{a}$ is estimated as in (2).

Repeating the procedure 2000 times, we were able to generate 2000 values for $\widehat{a}$ and $\widetilde{a}$. Since the theoretical a is fixed and known we can find the "exact" values of the mean squared error (MSE) of $\widetilde{a}$ and $\widehat{a}$. Simulations have been carried out for various variances of measurement errors, various a and various n and m. The results are quite satisfactory. The following general conclusions can be drawn:

(a) Both estimators have the same asymptotic variance.

(b) When the slope a is large, and other parameters are held fixed, the maximum likelihood estimator (MLE) is to be preferred. On the contrary, if a is smaller than 1 then the estimator based on variance components (VCE) has smaller mean squared error.

(c) When σ_ε^2 is greater than σ_δ^2 then MLE is better. In the opposite situation VCE is to be preferred.

(d) When the variances of both disturbances are equal and the regression slopes a tend to 1 very small differences between the mean squared errors occur.

In Table 1 we present some results of the simulations.

TABLE 1. Comparison of mean squared errors for MLE and VCE

n	m	σ_ε	σ_δ	a	MSE($\widetilde{a}$)	MSE($\widehat{a}$)	Preferred estimator
6	3	0.3	1	1	0.0230	0.0257	VCE
6	6				0.0151	0.0167	
8	3				0.0096	0.0109	
8	6				0.0062	0.0064	
12	3				0.0031	0.0034	
17	3				0.0011	0.0012	
6	3	1	0.3	1	0.0654	0.0571	MLE
6	6				0.0612	0.0561	
8	3				0.0300	0.0247	
8	6				0.0296	0.0241	
12	3				0.0097	0.0076	
17	3				0.0033	0.0025	
6	3	0.5	0.5	0.5	0.0082	0.0101	VCE
6	6				0.0059	0.0073	
8	3				0.0033	0.0034	
8	6				0.0025	0.0026	
12	3				0.0010	0.0011	
17	3				0.0004	0.0004	
10	3	0.5	0.5	0.3	0.0013	0.0020	
12	3				0.0007	0.0009	
17	3				0.0003	0.0003	
10	3	0.5	1	0.5	0.0042	0.0068	
17	3				0.0008	0.0010	
6	3	0.5	0.5	5	0.3799	0.3745	MLE
8	3				0.1573	0.1497	
8	6				0.1518	0.1455	
12	3				0.0444	0.0398	
17	3				0.0161	0.0146	
10	3	1	0.5	5	0.3976	0.3144	
12	3				0.2211	0.1636	
17	3				0.0857	0.0617	
6	3	0.5	0.5	1	0.0188	0.0199	both
6	6				0.0182	0.0186	
8	3				0.0088	0.0085	
8	6				0.0070	0.0071	
12	3				0.0025	0.0024	
17	3				0.0008	0.0008	

Acknowledgements. The author thanks Dr. A. L. Dawidowicz for his encouragement and help in the preparation of this paper, and Prof. R. Zieliński for his constructive suggestions.

References

O. Bunke and H. Bunke (1989), *Non-Linear Regression, Functional Relationships, and Robust Methods*, Wiley, New York.

C. G. Chow (1983), *Econometrics*, Mc-Graw-Hill, New York.

W. G. Cochran (1968), *Errors of measurement in statistics*, Technometrics 10, 637–666.

N. R. Cox (1976), *The linear structural relation for several groups of data*, Biometrika 63, 231–237.

W. A. Fuller (1987), *Measurement Error Models*, Wiley, New York.

L. J. Gleser, R. J. Carroll and P. P. Gallo (1987), *The limiting distribution of least squares in an errors-in-variables regression model*, Ann. Statist. 15, 220–233.

S. Gnot (1991), *Estimation of Variance Components in Linear Models*, Wyd. Naukowo-Techniczne, Warszawa (in Polish).

M. G. Kendall and A. Stuart (1979), *The Advanced Theory of Statistics*, Vol. 2, Griffin, London.

R. Kettelapper (1983), *On estimating parameters in a simple error-in-variables model*, Technometrics 25, 43–47.

A. Olsen, J. Seely and D. Birkes (1976), *Invariant quadratic unbiased estimation for two variance components*, Ann. Statist. 4, 878–890.

S. Penev and T. Raykov (1993), *On choosing estimators in a simple linear errors-in-variables model*, Comm. Statist. Theory Methods 22, 2429–2444.

C. R. Rao (1973), *Linear Statistical Inference and its Applications*, Wiley, New York.

C. R. Rao and J. Kleffe (1988), *Estimation of Variance Components and Applications*, North-Holland Ser. Statist. Probab. 3, North-Holland, Amsterdam.

O. Reiersol (1950), *Identifiability of a linear relation between variables which are subject to error*, Econometrica 18, 575–589.

Anna Czapkiewicz
Faculty of Management
University of Mining and Metallurgy
Gramatyka 10
30-067 Kraków, Poland
E-mail: gzrembie@cyf-kr.edu.pl

Received on 13.10.1997;
revised version on 21.4.1998 and 6.7.1998