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 90, 1989

*Halina Klepacz**

THE BIAS OF ESTIMATORS OF MODELS WITH ERRORS IN VARIABLES

1. INTRODUCTION

The main subject of this paper are estimators of parameters of one-equation linear models with errors in exogenous variables. The known analyses were concentrated on, as for the class of models under investigation, the comparison of asymptotic mean square errors of chosen estimators according to different estimation methods.

From theoretical and practical points of view the small sample features of estimators for the errors in variables models are also important. A practicioner is interested in the conditions of applying special estimation methods. If we study these methods in terms of mean square error equal to the sum of bias square and variance then, knowing the size of the bias of parameter estimates of the model describing the results of observations with measurement errors is absolutely necessary.

The paper contains the analysis of biases of the chosen estimates of the three versions of the model[1]:

1) with two exogenous variables, one of which is measured with error;

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

[1] For the model with one explanatory variable measured with error, the analysis of results of the Monte Carlo experiment was presented in [1], [3].

2) with three explanatory variables, one of which has the measurement error;

3) with three explanatory variables, two of which are observed with error.

The analysis is conducted on the basis of the results of Monte Carlo experiments[2] for chosen: methods, the sample sizes, coefficient of determination, the errors of the measurement of variables and the correlation coefficient between the explanatory variables.

2. STEERING VARIABLES AND PARAMETERS ACCEPTED IN THE EXPERIMENTS

We will analyse the biases of the estimators of the modified least squares method (OLS), of the maximum likelihood method (MLM), of the instrumental variables methods: by Wald (IV WAL), Bartlett (IV BAR) and Durbin (IV DUR). These biases were calculated and compared for three versions of one-equation, linear model [1] with errors in explanatory variables, i.e.

$$Y_i = \alpha_0 + \alpha_1 + \tilde{X}_{1i} + \alpha_3 X_{3i} + (\varepsilon_i - \alpha_3 V_{3i}) \tag{1}$$

$$Y_i = \alpha_0 + \alpha_1 \tilde{X}_{1i} + \alpha_2 \tilde{X}_{2i} + \alpha_3 X_{3i} + (\varepsilon_i - \alpha_3 V_{3i}) \tag{2}$$

$$Y_i = \alpha_0 + \alpha_1 \tilde{X}_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{3i} + (\varepsilon_i - \alpha_2 V_{2i} - \alpha_3 V_{3i}) \tag{3}$$

Instead of unobservable explanatory variable $\tilde{X}_j$ the observable $X_j = \tilde{X}_j + V_j$ is introduced in these models. Random variable V_j expressing the measurement errors of the X_j, is normally distributed with expected value $E(V_j) = 0$ and variance $D^2(V_j) = \sigma_j^2$, $j = 2, 3$. Moreover, it is assumed that: $E(\varepsilon/V) = E(\varepsilon) = 0$, $E(Y/V) = X\alpha$ and X is the matrix of observations $E(\varepsilon\varepsilon^T/V) = \sigma^2 I$.

---

[2] The construction of the sample space was discussed in [3] the review of estimation methods for the models with measurement errors are given in [1], [2].

Variables[3] X_j are treated in experiments as fixed ones, being repeated in all experiment repetitions.

In the experiments that have been carried out, the number of observations in a sample was fixed at the levels n = 20, 30, 40, 50 (observations for X_j were chosen as the first n elements from the chosen column of a table of random numbers). The values of structural parameters of the models (1) - (3) were fixed at the level: $\alpha_0 = 1000$, $\alpha_1 = 3$, $\alpha_2 = 1$, $\alpha_3 = 2$. The choice of a given vector of structural parameters is limiting the range of reasoning to the given structure of linear model with errors-in-variables. The directions of changes of estimation biases and the quantities of these biases depend on the size of the sample and stochastic structure of the chosen models.

The following levels were accepted: $R^2 =$ 0.50; 0.60; 0.70; 0.80; 0.90; 0.95; 0.99 and the RB levels, i.e. the quotient of the error variance and the sample variance of a given explanatory variable: 0.01; 0.05; 0.10; 0.15.

3. THE BIASES OF ESTIMATORS OF CHOSEN METHODS IN DEPENDENCE ON THE SAMPLE SIZE AND THE ERRORS OF MEASUREMENT

We shall analyse now the influence if the enlarging of explanatory variables in the model on the biases of the parameter estimations of the models with errors in the variables.

It should be remembered that the increase of values of RB error in the model with one explanatory variable[4] measured with error caused almost proportional increase of the average biases. It was observed in relation to the method which we may call "basic" as further we shall compare the results obtained by this method with others. The enlarging of the model by adding an additional variable measured without error, causes similar tendency, but, for each tested method, except the LSM and MLM, we observe the decrease in biases of the average estimates of para-

---

[3] The sequences of the variable values taken from [4].

[4] The analysis of Monte Carlo experiment for the model $Y_i = \alpha_0 + \alpha_3\tilde{X}_{3i} + (\varepsilon_i - \alpha_3 V_{3i})$ is given in [1] and [3].

meters, for the models with variable measured with errors (see Table 1).

T a b l e 1

Average biases of the parameter estimates of the models (1)-(2) in relation to the average estimates of the basic method for n = 20 and R^2 = 0.99, IP = 500

RB	Methods	Average biases (in %) for the model with					
		one[a] variable	two[b] variables		three[b] variables		
		(X_3)	(X_3)	$(\tilde{X}_1)$	(X_3)	$(\tilde{X}_1)$	$(\tilde{X}_2)$
0.05	LSM	-4.4	-3.5	0.6	-3.3	0.6	-0.7
	OLS	0.6	1.5	-0.5	1.8	-0.2	0.5
	MLM	0.7	1.8	-0.2	2.1	-0.3	0.6
	IV DUR	-3.7	-3.7	0.6	-3.5	0.7	-0.8
0.10	LSM	-8.2	-7.1	1.1	-6.8	1.3	-1.6
	OLS	1.4	2.8	-0.2	3.4	-0.4	0.2
	MLM	1.8	3.4	-0.4	4.0	-0.6	1.2
	IV DUR	-7.6	-7.1	1.1	-6.8	1.3	-1.6
0.15	LSM	-11.4	-10.6	-10.6	-10.0	1.9	-2.4
	OLS	2.2	3.7	-0.3	5.3	-0.5	0.9
	MLM	2.7	5.0	-0.6	6.0	-1.0	1.7
	IV DUR	-10.3	-10.1	-1.6	-9.7	1.8	-2.3

[a] The results given in this column are taken from [3].

[b] The construction of the sample space was discussed in [3] the review of estimation methods for the models with measurement errors are given in [1], [2].

Average estimates of the parameters for such models are also biased although the change of the bias direction is in the small degree, a characteristic feature of these biases, i.e. if in a given method we underestimate parameters with the variable measured with error, then the average value of estimator with "no--error" variable is overestimating parameter and vice versa.

The introduction of another "no-error" variable to the model (1) causes rather small increase in the absolute average biases of the parameter estimates with the first "no-error" variable and the diminishing of the average biases of estimators in the same direction, with the variable measured with error (except OLS and LSM methods, where the tendency is the opposite one).

If we introduce additionally to the model (1) a variable measured with error (instead of "no-error" one), then the increase in biases can be observed for the estimates of both parameters. The bigger RB for the introduced variable is, the greater the increase in biases will be.

The increase in this bias is not very big for the parameters estimate with the first variable measured with error and big for parameter estimates with the variable measured without error. This bias increase quickly together with the increase of the RB level (see Table 2), and the increase of each measurement error of RB causes the increase of all the biases.

Table 2

Average biases of the parameter estimates of the model (3) in relation to the average estimates of the basic method for n = 20 and $R^2 = 0.99$

RB2	Methods	Average biases of the parameter estimates for					
		RB1 = 0.05			RB1 = 0.15		
		α_1	α_2	α_3	α_1	α_2	α_3
0.05	LSM	1.3	-5.1	-3.9	2.5	-13.6	-4.3
	OLS	-0.4	2.2	1.6	-0.8	5.4	1.7
	MLM	-0.6	2.9	1.7	-1.3	7.2	1.9
	IV DUR	1.5	-6.2	-4.0	2.5	-13.0	-4.3
0.10	LSM	1.9	-6.0	-7.5	3.2	-14.5	-7.9
	OLS	-0.6	2.7	2.8	-1.5	7.3	3.2
	MLM	-0.9	3.3	3.4	-1.6	7.7	3.7
	IV DUR	2.1	-7.3	-7.4	3.2	-14.9	-7.7
0.15	LSM	2.5	-6.8	-10.8	3.8	-15.3	-11.2
	OLS	-0.9	3.3	3.4	-1.6	7.6	3.8
	MLM	-1.2	4.0	5.3	-1.9	8.3	5.7
	IV DUR	2.7	-8.4	-10.3	3.7	-15.8	-10.6

In LSM IV DUR methods the introduction of an additional variable (with or without error) causes the underestimation of the estimate when a variable is measured with error as well as for the introduced variable; while in OLS and MLM methods there can be observed the overestimation of average parameter estimates and its absolute value is smaller than for LSM and IV DUR.

T a b l e 3

Average biases (in %) of the parameter estimates of the model (3) in relation to the average estimates of the basic method in relation to the measurement errors of the explanatory variables for n = 20, R^2 = 0.95, IP = 500

Estimated parameter	RB1	RB2	Methods of estimation			
			LSM	OLS	MLM	IV DUR
α_1		0.01	1.3	-0.6	-0.8	1.3
		0.05	1.8	-0.9	-1.0	1.9
	0.10	0.10	2.5	-1.2	-1.4	2.6
		0.15	3.2	-1.4	-1.7	3.3
	0.01		1.3	-0.6	-0.7	1.4
	0.05		1.9	-0.9	-1.0	2.0
	0.10	0.10	2.5	-1.3	-1.4	2.6
	0.15		3.1	-1.5	-1.7	3.1
α_2		0.01	-7.9	3.4	5.5	-8.9
		0.05	-8.7	3.9	5.8	-9.8
	0.10	0.10	-9.5	4.2	6.3	-10.9
		0.15	-12.0	5.6	7.3	-13.7
	0.01		-2.3	1.1	1.6	-3.7
	0.05		-5.5	2.8	3.8	-7.4
	0.10	0.10	-9.5	4.9	6.3	-10.9
	0.15		-13.3	7.6	8.9	-13.5
α_3		0.01	-1.1	0.3	-0.6	-1.1
		0.05	-4.1	1.6	1.9	-4.1
	0.10	0.10	-7.6	3.1	3.6	-7.5
		0.15	-11.0	4.9	5.6	-10.4
	0.01		-7.2	2.8	3.4	-7.4
	0.05		-7.4	2.9	3.5	-7.3
	0.10	0.10	-7.6	3.1	3.6	-7.5
	0.15		-7.8	3.2	3.8	-7.6

Average biases of the parameter estimates for the model (3) in relation to the average parameter estimates of the basic method depend more on RB than on the changes of determination coefficient. This is illustrated by the Tables 3 and 4. However, we observe that together with the increase of R^2 the biases of parameter estimates OLS and MLM decrease and the biases of the estimates for other method increase.

T a b l e 4

Average biases of the parameter estimates for the model (3) in relation to the average estimates of the basic method for RB1 = 0.05, RB2 = 0.10, n = 20, TP = 500

Estimated parameter	Method of estimation	Assumed values of correlation coefficient						
		0.5	0.6	0.7	0.8	0.9	0.95	0.99
α_1	LSM	1.3	1.4	1.6	1.7	1.8	1.9	1.9
	OLS	-1.3	-1.3	-1.2	-1.1	1.0	-0.9	-0.8
	MLM	-1.5	-1.4	-1.3	-1.3	-1.1	-1.0	-0.9
	IV DUR	1.4	1.5	1.7	1.9	2.0	2.0	2.1
	IV WAL	0.4	0.5	0.7	0.7	0.8	0.9	0.9
	IV BAR	0.8	1.0	1.2	1.4	1.5	1.6	1.6
α_2	LSM	-3.7	-4.2	-4.8	-4.8	-5.1	-5.5	-6.0
	OLS	5.1	4.8	4.2	3.8	3.3	2.8	2.7
	MLM	5.3	5.1	4.7	4.3	4.1	3.8	3.3
	IV DUR	-4.9	-5.8	-6.1	-6.3	-6.7	-7.4	-7.3
	IV WAL	1.6	0.9	0.5	0.2	0.1	-0.2	-0.5
	IV BAR	-3.2	-4.3	-5.1	-5.7	-6.2	-6.4	-6.8
α_3	LSM	-6.9	-7.1	-7.2	-7.1	-7.3	-7.4	-7.5
	OLS	3.6	3.5	3.4	3.4	3.1	2.9	2.8
	MLM	3.8	3.7	3.7	3.7	3.5	3.5	3.4
	IV DUR	-6.7	-6.8	-6.9	-7.1	-7.2	-7.3	-7.4
	IV WAL	-6.3	-6.1	-5.8	-5.6	-5.3	-5.1	-4.8
	IV BAR	-4.8	-5.1	-5.3	-5.6	-5.8	-6.0	-6.3

The increase of the sample size (see: Table 1 and 5) does not cause the decrease in biases either in LSM or in IV DUR. This decrease is observed for OLS and MLM methods.

Table 5

Average biases of the parameter estimates
of the model (2) in relation
to the average estimates of the basic method
for RB1 = 0.10 and n = 50, IP = 500

RB2	Method	Assumed value of the determination coefficient								
		0.90			0.95			0.99		
		$\tilde{X}_1$	$\tilde{X}_2$	$\tilde{X}_3$	$\tilde{X}_1$	$\tilde{X}_2$	$\tilde{X}_3$	$\tilde{X}_1$	$\tilde{X}_2$	$\tilde{X}_3$
	LSM	0.8	-1.4	-4.6	0.8	-1.3	-4.6	0.8	-1.3	-4.6
	OLS	0.0	0.2	0.5	0.0	0.2	0.5	0.0	0.2	0.6
0.05	MLM	0.0	0.2	0.5	0.0	0.2	0.5	0.0	0.2	0.6
	IV DUR	0.7	-1.7	-4.4	0.7	-1.3	-4.3	0.7	-1.2	-4.3
	LSM	1.5	-2.7	-8.9	1.4	-2.6	-8.8	1.5	-2.5	-8.8
	OLS	-0.1	0.3	1.1	-0.1	0.3	1.1	-0.1	0.4	1.1
0.10	MLM	-0.1	0.3	1.1	-0.1	0.4	1.1	-0.1	0.4	1.2
	IV DUR	1.4	-2.5	-8.2	1.4	-2.4	-8.2	1.4	-2.3	-8.1
	LSM	2.1	-3.9	-12.8	2.1	-3.8	-12.7	2.1	-3.7	-12.6
	OLS	-0.2	0.5	1.6	-0.2	0.5	1.6	0.2	0.5	1.6
0.15	MLM	-0.2	0.5	1.6	-0.2	0.6	1.7	-0.2	0.6	1.8
	IV DUR	2.0	-3.6	-11.8	1.9	-3.5	-11.7	1.9	-3.3	-11.6

We can observe significant influence of the degree of correlation between the explanatory variables measured without any error on the biases of parameter estimates of the models (see Table 6). The increase of the degree of variable correlation causes the increase of biases for both the parameter estimates measured with and without error. The absolute value of estimators bias is greater when the variable is measured with error. Considering the biases in different methods of estimation we can state that OLS and MLM are far better than the others and IV-methods are not "better" than oridinary least squares method if in samples of 20 elements the correlation coefficients increase. Quick reaction of the IV WAL method on the increase of the values of the correlation coefficient leading to large biases when correlation coefficient ϱ^2 is large is very alarming.

Table 6

Average biases of the parameter estimates
of the model (1) in relation
to average estimates of the basic method
for n = 20, R^2 = 0.95, RB = 0.5
in dependence on the degree of correlation
of the explanatory variables IP = 500

Estimated parameter	Correlation coefficient	Method of estimation					
		LSM	OLS	MLM	IV DUR	IWAL	IV BAR
α_2	0.03	0.1	-0.0	0.0	0.1	0.1	0.1
	0.45	0.5	-0.1	-0.2	0.6	0.6	0.7
	0.70	1.1	-0.4	-0.6	1.4	2.7	1.6
	0.88	3.1	-1.7	-2.0	3.3	21.8	3.7
α_3	0.03	-3.3	1.4	1.7	-3.5	-2.7	-3.3
	0.45	-4.0	1.9	2.2	-4.6	-4.1	-5.2
	0.70	-6.8	3.1	3.5	-7.9	-15.7	-9.1
	0.88	-13.3	7.9	8.7	-13,6	-26.5	-16.4

On the basis of the carried out experiments we can conclude that the more "no-error" variables we have in the model, the smaller the average biases of the estimates covering parameters standing nearby a variable measured with error are in LSM and IV DUR. If the variable is measured without error, there exist hardly observable biases. The opposite situation takes place in the case of enlarging of the model with variables measured with error.

There can also be noticed insignificant influence of R^2 and the sample size on average biases of the parameter estimates in relation to the basis method, i.e. to the model where all the variables are observed without error.

4. FURTHER INVESTIGATIONS

The presented analysis of biases of parameter estimates on the basis of Monte Carlo experiment for the chosen linear models has a narrow scope. Only a picture of behaviour of the chosen estimation methods under given experiment conditions was obtained. To evaluate the efficiency of these methods in a broad sense, it is necessary to enlarge the investigations, first of all, to consider different numerical structures of parameters for the chosen models, given different degrees of correlation of explanatory variables. It would be advisable to carry out experiments given lower levels of R^2 coefficient and higher levels of RB. In the models with at least two variables with measurement errors, there is an additional problem caused by the necessity of taking into account different variants of relationships between the errors for separate variables while analysing the properties of these methods.

REFERENCES

[1] K l e p a c z H. (1984), *Efektywność metod estymacji modeli jednorównaniowych z błędami w zmiennych*, doctor thesis, Łódź.

[2] K l e p a c z H. (1984), *Przegląd metod estymacji modeli jednorównaniowych z błędami w zmiennych*, Zesz. Nauk. Akad. Ekon., 181 (Kraków).

[3] K l e p a c z H., *The Efficiency of Estimation Methods for Models with Errors in Explanatory Variables*, forthcoming.

[4] Z i e l i ń s k i R. (1972), *Tablice statystyczne*, PWE, Warszawa.

Halina Klepacz

OBCIĄŻENIE ESTYMATORÓW PARAMETRÓW MODELI Z BŁĘDAMI W ZMIENNYCH OBJAŚNIAJĄCYCH

W artykule analizuje się wielkości obciążeń parametrów strukturalnych modeli z dwiema zmiennymi objaśniającymi, w tym jedna jest obarczona błędem pomiaru oraz z trzema zmiennymi, z których jedna lub dwie są mierzone z błędem. Analizę przeprowadza się ze względu na metody estymacji, wielkość próby, poziom: współczynnika determinacji, błędu pomiaru i współczynnika korelacji między zmiennymi.