⊠연구논문

소표본 errors-in-variables 모형에서의 통계 추론*

소병수
이화여자대학교 통계학과

Small-Sample Inference in the Errors-in-Variables Model

Beong-Soo So
Dept. of statistics, Ewha Womans University

Abstract

We consider the semiparametric linear errors-in-variables model: $y_i = \alpha + \beta u_i + \varepsilon_i$, $x_i = u_i + \delta_i$ $i=1,\cdots,n$ where (x_i, y_i) stands for an observation vector, (u_i) denotes a set of incidental nuisance parameters, (α, β) is a vector of regression parameters and $(\varepsilon_i, \delta_i)$ are mutually uncorrelated measurement errors with zero mean and finite variances but otherwise unknown distributions. On the basis of a simple small-sample low-noise approximation, we propose a new method of comparing the mean squared errors(MSE) of the various competing estimators of the true regression parameters (α, β). Then we show that a class of estimators including the classical least squares estimator and the maximum likelihood estimator are consistent and first-order efficient within the class of all regular consistent estimators irrespective of type of measurement errors.

* This research was supported by the Ewha research fund, 1994.

1. Introduction

We consider the following linear semiparametric errors-in-variables model:

$$\begin{aligned} y_i &= \alpha+\beta u_i+\ \varepsilon_i \\ x_i &= u_i+\delta_i, \qquad i=1,\cdots,n \end{aligned} \tag{1.1}$$

where (x_i, y_i) represent observations of the unknown error free true values (u_i, v_i) which are connected by the linear relations: $v_i=\alpha+\beta u_i$, with unknown regression parameters (α, β) and $(\delta_i, \varepsilon_i)$ are mutually uncorrelated measurement errors with zero mean and finite variances σ_x^2, σ_y^2 respectively but otherwise *unknown* distributions. In this model, we are mainly interested in the problem of efficient estimation of the unknown functional relations between the true error-free variables (u_i, v_i) on the basis of noisy observations (x_i, y_i), $i=1,\cdots,n$.

When the independent variables u_i are subject to measurement errors, the ordinary least squares estimator (OLSE) of the regression parameters (α, β) are known to be generally biased and inconsistent as $n\rightarrow\infty$. See Anderson(1976) for a discussion of this problem and some of its generalizations. Historically, most of the research efforts have been directed toward finding alternative estimators which have desirable properties such as large sample consistency and efficiency under various assumptions on the variables u_i and measurement errors. For example, see Fuller (1987) for more extensive review on this approach. Under usual normal error model, Bickel and Ritov(1987) considered the problem of efficient estimation of the regression parameters (α, β) when the error-free variables u_i are i.i.d. random variables with unknown mixture distribution $G(\cdot)$.

But most of the previous works on this important problem have a serious limitation in that they are heavily based on the large-sample properties such as asymptotic biases and asymptotic MSEs of the various competing estimators. Thus these results cannot be directly applied to the important class of problems with relatively small sample sizes. This is typically the case in the experimental set-up where high cost of the each experiment does not allow large number of replications as is commonly required in the usual observational study.

In view of its practical importance, the problem of efficient estimation in the small-sample errors-in-variables model received relatively little attention in the literature. As related works in this area, Villegas(1969) considered asymptotic properties of the least squares estimator when there are repeated observations.

In this paper we will focus on the problem of efficient estimation in the small-sample errors-in-variables model and develop new optimality theory based on the small-sample low-noise asymptotics. Specifically, we introduce appropriate definitions of small-sample consistency and efficiency of the regular estimators of the regression parameters (α, β) and then we show that the both maximum likelihood estimator and the least squares estimator have a justification in its own right as low-noise consistent and first-order efficient estimator without any reference to the specific distributional assumptions on the measurement errors such as normality.

This paper is organized as follows: In section 2. we first introduce the new definitions of the small-sample consistency and low-noise efficiency of the estimators of (α, β). Then we derive an important lower-bound for the AMSE of the arbitrary regular consitent estimators which depends only on the 2-nd order moments of the measurement errors but is independent of the type of error distributions. In section 3, we show that the maximum likelihood estimator and the least squares estimator are both consistent and first-order efficient irrespective of the type of measurement errors. Finally in section 4, we provide some simulation results which illustrate the practical relevancy of the small-sample low-noise approximations.

2. Main Results

In this paper we always assume that the sample size n is a fixed finite number and we use the following notations. Let $z=(z_1, \cdots, z_n)^t$, $z_i=(x_i, y_i)$, $i=1, \cdots, n$ be the 2n-dimensional vector of observations and let $\mu=(\mu_1, \cdots, \mu_n)^t$ be the mean vector $E(z)$ of z defined by $E(z_i)=\mu_i=(u_i, \alpha+\beta u_i)$, $i=1, \cdots, n$. We also note that the mean vector μ of z depends on the parameter vector $\theta=(\alpha, \beta, u_1, \cdots, u_n)^t$ and thus we may denote it by $\mu(\theta)$ or by $\mu(\alpha, \beta, u_i)$ showing the explicit dependence on the relevant parameters. We also assume that vaiance ratio $r=(\sigma_x/\sigma_y)^2$ is a known constant as is usually assumed in the errors-in-variables literature in order to ensure the identifiability of the regression parameters (α, β).

We now introduce the following definition of the regular consisteny of the estimators of the arbitrary estimand $g(\alpha, \beta)$.

Definition 1. An estimator $h(z)$ of $g(\alpha, \beta)$ is called *regular consistent* if $h(\cdot)$ is a continuously differentiable function of z and satisfies the condition:

$$h(\mu(\theta)) = h((u_i, \alpha + \beta u_i)) = g(\alpha, \beta) \quad \text{holds}$$
$$\text{for all } \theta = (\alpha, \beta, u_1, \cdots, u_n) \tag{2.1}$$

Remark 1. Note that our definition of consistency (2.1) is completely different from the usual large-sample definition of consistency which is used in most of previous works as in Fuller (1987) and Bickel and Ritov (1987) because we do not consider the behaviour of the estimator as the sample size n gets large but study the performance of the estimator as the variances σ_x^2 , σ_y^2 of the measurement errors get small for finite sample size n.

Remark 2. For regular estimators, we note that condition (2.1) is equivalent to the more familiar concept of *asymptotic unbiasedness* as $\sigma_x, \sigma_y \to 0$ which is defined by:

$$\lim_{\sigma_x, \sigma_y \to 0} E[h(z)] = g(\alpha, \beta) \text{ for all } \theta = (\alpha, \beta, u_1, \cdots, u_n). \tag{2.2}$$

Similarly we note that the condition (2.1) is equivalent to that of the *consistency in probability* as $\sigma_x, \sigma_y \to 0$:

$$\lim_{\sigma_x, \sigma_y \to 0} h(z) = g(\alpha, \beta) \text{ for all } \theta = (\alpha, \beta, u_1, \cdots, u_n). \tag{2.3}$$

Remark 3. Any reasonable estimator of $g(\alpha, \beta)$ must be consistent in the sense of (2.1) because when there is no measurement error it seems perfectly reasonable to require that we should be able to recover the true values (α, β) exactly no matter what they are. In fact every estimator considered in the literature satisfies this requirement including the maximun likelihood estimator and least squares estimator.

In order to compare the performances of various regular consistent estimators, we introduce the definition of the AMSE (Asymptotic Mean Squared Error) of the estimator as follows.

Definition 2. AMSE (Asymptotic Mean Squared Error) of the regular consistent estimator $h(z)$ of $g(\alpha, \beta)$ is the quantity defined by:

$$AMSE[h(z)] = \sigma_y^2 \lim_{\sigma_y \to 0} (E[h(z) - g(\alpha, \beta)]^2 / \sigma_y^2). \tag{2.4}$$

Now we are ready to establish the fundamental lower bound for the AMSEs of the regular consistent estimators.

Theorem 1. If $h(z)$ is a regular consistent estimator of $g(\alpha, \beta)$, then we have the lower bound:

$$AMSE[h(z)] \geq \sigma_y^2(1 + r\beta^2)(g_\alpha^2 - 2\bar{u}g_\alpha g_\beta + g_\beta^2\overline{u^2})/s_{uu} \tag{2.5}$$

$$\text{for all } (\alpha, \beta, u_1, \cdots, u_n)$$

where $\bar{u} = \sum_{i=1}^{n} u_i/n$, $\overline{u^2} = \sum_{i=1}^{n} u_i^2/n$, $s_{uu} = \sum_{i=1}^{n}(u_i - \bar{u})^2$,

and $(g_\alpha, g_\beta) = (\partial g/\partial \alpha, \partial g/\partial \beta)$.

Proof. By the regular consistency of the estimator $h(\cdot)$, we get the identity:

$$h(\mu(\alpha, \beta, u_1, \cdots, u_n)) = g(\alpha, \beta) \quad \text{for all } (\alpha, \beta, u_i).$$

Differentiating above identity partially with respect to α, β, u_i respectively, we get the following series of identities:

$$\sum_{i=1}^{n} \partial h / \partial y_i = \partial g / \partial \alpha \tag{2.6}$$

$$\sum_{i=1}^{n} (\partial h / \partial y_i) u_i = \partial g / \partial \beta \tag{2.7}$$

$$\partial h/\partial x_i + (\partial h/\partial y_i)\beta \; = \; 0, \;\; i=1,\cdots,n \qquad (2.8)$$

Multiplying (2.6), (2.7) and (2.8) by k_1, k_2 and $c_i \;\; i=1,\cdots,n$ respectively and adding them, we have the identity:

$$\sum_{i=1}^{n} \partial h/\partial y_i(k_1 + k_2 u_i) + \sum_{i=1}^{n} c_i(\partial h/\partial x_i + \partial h/\partial y_i \, \beta) = k_1 g_\alpha + k_2 g_\beta \qquad (2.9)$$

Now by the Cauchy-Schwartz inequality, we have the inequality:

$$\begin{aligned} &[\sum_{i=1}^{n}(\partial h/\partial x_i)^2\sigma_x^2 + \sum_{i=1}^{n}(\partial h/\partial y_i)^2\sigma_y^2 \;] \geq \\ &[\sum_{i=1}^{n} c_i^2/\sigma_x^2 + \sum_{i=1}^{n}(k_1 + k_2 u_i + c_i\,\beta)^2/\sigma_y^2]^{-1} \cdot (k_1 g_\alpha + k_2 g_\beta)^2 \end{aligned} \qquad (2.10)$$

Taking supremum of the lower bound (2.10) with respect to $(c_1,\cdots,c_n)$, we obtain the inequality:

$$\begin{aligned} &\sum_{i=1}^{n}((\partial h/\partial x_i)^2\sigma_x^2 + (\partial h/\partial y_i)^2\sigma_y^2) \geq \\ &\sigma_y^2(k_i g_\alpha + k_2 g_\beta)^2/[\sum_{i=1}^{n}(k_1 + k_2 u_i)^2/(1+\beta^2 r)] \end{aligned} \qquad (2.11)$$

Again taking supremum of the lower bound (2.11) with respect to (k_1, k_2), we get the inequality:

$$(\partial h/\partial x_i)^2\sigma_x^2 + (\partial h/\partial y_i)^2\sigma_y^2 \geq \sigma_y^2(1+\beta^2 r)(g_\alpha^2 - 2g_\alpha g a_\beta \overline{u} + \overline{u^2} g_\beta^2)/s_{uu} \qquad (2.12)$$

Now we note that the regular consistency of the estimator $h(\cdot)$ implies that:

$$\begin{aligned} h(z) - g(\alpha,\beta) \; &= \; h(z) - h(\mu) \\ &= \; \sum_{i=1}^{n}(\partial h/\partial x_i)\,\delta_i + (\partial h/\partial y_i)\,\varepsilon_i + o(\sigma_y)\,. \end{aligned} \qquad (2.13)$$

and thus

$$AMSE[h(z)] = \sum_{i=1}^{n}[(\partial h/\partial x_i)^2\sigma_x^2 + (\partial h/\partial y_i)^2\sigma_y^2]. \tag{2.14}$$

immediately. (2.12) and (2.14) finish the proof.

Remark 4. We note that the lower bound (2.5) is semiparametric lower bound for the AMSE of the regular consistent estimator because it depends on the 2-nd order moment of the measurement errors $(\delta_i, \varepsilon_i)$ and independent of the type of the distributions of the measurement errors.

Now we will consider the problem of identifying a regular consistent estimator whose AMSE attains the lower bound (2.5) for all $\theta = (\alpha, \beta, u_i)$.

3. Optimal Estimators

Motivated by the universal lower bound (2.5), we first define the optimality of the regular consistent estimator as follows.

Definition 3. A regular consistent estimator $h(z)$ of $g(\alpha, \beta)$ is called *optimal* if we have the equality:

$$AMSE[h(z)] = \sigma_y^2(1+\beta^2 r)[g_\alpha^2 - 2g_\alpha g_\beta \overline{u} + \overline{u^2} g_\beta^2]/s_{uu} \tag{3.1}$$

$$\text{for all } (\alpha, \beta, u_1, \cdots, u_n).$$

Now we are ready to establish the main result of the paper which claims the optimality of the normal maximum likelihood estimator and the ordinary least squares estimator. First we introduce the definition of the normal maximum likelihood estimator.

Definition 4. Normal maximum likelihood estimator(MLE) of (α, β) is defined by

$$(\widehat{\alpha_m}, \widehat{\beta_m}) = \arg\min_{\alpha,\beta} \sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2/(1+\beta^2 r). \tag{3.2}$$

It is well known in the literature, see p405 of Kendall and Stuart (1979), that

$$\widehat{\beta_m} = [\,(s_{yy} - r^{-1}s_{xx}) + ((s_{yy} - r^{-1}s_{xx})^2 + 4r^{-1}s_{xy}^2)^{1/2}]/2s_{xy} \tag{3.3}$$

$$\widehat{\alpha_m} = \bar{y} - \widehat{\beta_m}\bar{x}$$

where $r = (\sigma_x/\sigma_y)^2$. We also define ordinary least squares estimators (OLSE) of (α, β) as follows;

$$\widehat{\beta_o} = s_{xy}/s_{xx}, \quad \widehat{\alpha_o} = \bar{y} - \widehat{\beta_o}\bar{x}. \tag{3.4}$$

Here we use the standard notation; $s_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2$, $s_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$, $s_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2$ and $\bar{x} = \sum_{i=1}^{n} x_i/n$, $\bar{y} = \sum_{i=1}^{n} y_i/n$.

Theorem 2. Both $g(\widehat{\alpha_m}, \widehat{\beta_m})$ and $g(\widehat{\alpha_o}, \widehat{\beta_o})$ are regular consistent and *optimal* estimators of $g(\alpha, \beta)$.

Proof. By the direct substitution, we can check the regular consistency of the two estimators immediately. As for the proof of the optimality, it suffices to show that AMSE matrix of the OLSE and the MLE is given by

$$AMSE(\hat{\alpha}, \hat{\beta}) = \sigma_y^2(1 + \beta^2 r)\begin{bmatrix} 1 & -\bar{u} \\ -\bar{u} & \bar{u}^2 \end{bmatrix}/s_{uu} \tag{3.5}$$

First we note that direct Taylor expansions yield the identities:

$$\begin{aligned} s_{xx} &= s_{uu} + 2\sum_{i=1}^{n}(u_i - \bar{u})\delta_i + o(\sigma_y) \\ s_{xy} &= \beta s_{uu} + \sum(u_i - \bar{u})(\beta\delta_i + \varepsilon_i) + o(\sigma_y) \\ s_{yy} &= \beta^2 s_{uu} + 2\sum(u_i - \bar{u})\beta\varepsilon_i + o(\sigma_y) \end{aligned} \tag{3.6}$$

Above identities imply immediately that

$$\begin{aligned}\widehat{\beta_o}-\beta&=\delta s_{xy}/s_{xx}-\ s_{xy}\delta s_{xx}/s_{xx}^2+\ o(\sigma_y)\\&=\sum(u_i-\bar{u})(\varepsilon_i-\beta\delta_i)/s_{uu}+o(\sigma_y)\end{aligned}\tag{3.7}$$

$$\begin{aligned}\widehat{\beta_m}-\beta&=(\partial h/\partial s_{uu})\delta s_{xx}+(\partial h/\partial s_{xy})\delta s_{xy}+(\partial h/\partial s_{yy})\delta s_{yy}+o(\sigma_y)\\&=\sum(u_i-\bar{u})(\varepsilon_i-\beta\delta_i)/s_{uu}+o(\sigma_y)\end{aligned}\tag{3.8}$$

and

$$\hat{\alpha}-\alpha\ =\bar{\delta}-\beta\bar{\varepsilon}-\bar{u}(\hat{\beta}-\beta)+o(\sigma_y)\tag{3.9}$$

where $\delta s_{xx}=s_{ss}-s_{uu}$, $\delta s_{xy}=s_{xy}-\beta s_{uu}$, $\delta s_{yy}=s_{yy}-\beta^2 s_{uu}$.

Above identities (3.7), (3.8), (3.9) imply (3.5) immediately.
This completes the proof.

Remark 5. In order to further discriminate the so-called second-order efficient estimator among the various first-order efficient estimators, we have to take into account more terms in the Taylor expansion and should also assume the knowledge of the third and forth-order moments of the measurement errors which is typically not available in the small-sample experiment. Therefore this topic will not be considered in this paper.

4. Simulation and Discussions

In order to illustrate the practical applicability of the small-sample asymptotics in the real data analysis, we have done simple simulation experiment which compares the approximate results of our work with the exact results given by the Monte-Carlo simulation of the exact MSEs of the normal MLE and OLSE.

We conducted Monte-Carlo simulation for the computation of MSEs of classical OLSE and normal MLE. For simpe comparison, we used sample size n=3 with design $(u_1, u_2, u_3)=(-1, 0, +1)$ and tried three different configurations depending on the size of the common standard deviation $\sigma_x=\sigma_y$: 0.1, 0.2, 0.5 with true parameter values; $\alpha=0, \beta=1$. We used normal random numbers for the measurement errors and computed MSEs of the estimators with 1000 replications. On close examination, the simulated values of the MSEs of the both estimators are generally in good agreement with the values given by the approximate formula (3.1) for AMSEs derived in Section 3. This results seems to justify the validity of the small-sample low-noise approximation developed in this paper. Following table provides typical comparison between our results and those from the Monte-Carlo simulation study.

Simulated MSEs and AMSE for n=3 with design $(u_1, u_2, u_3)=(-1, 0, +1)$

	MSE		
σ	OLSE	MLE	AMSE
0.1	.011	.011	.01
0.2	.039	.043	.04
0.3	.225	.273	.25

Finally we discuss the problems of extending our results to non-linear and multivariate errors-in-variables models.

Remark 6. If we consider general *non-linear* errors-in-variables model:

$$y_i = f(u_i; \beta) + \varepsilon_i$$

$$x_i = u_i + \delta_i \quad i=1, \cdots, n$$

where $f(\cdot\,;\beta)$ represents an arbitrary non-linear function of u_i and β denotes vector of regression parameters. We can establish similar optimality results for the non-linear least squares estimators of β within the small-sample low-noise framework.

Remark 7. Instead of the univariate errors-in-variables model (1.1), we can consider *multivariate* errors-in-variables model where each of the observations in (x_i) $i=1, \cdots, n$ and the error-free true values (u_i) are m-dimensional vectors. Then we can extend our results to this model without difficulty by considering appropriate small-sample low-noise approximation.

References

[1] Anderson, T.W.(1976) "Estimation of linear functional relation: approximate distributions and connections with simultaneous equations in economics," *Journal of the Royal Statistical Society ,(B)* 38, pp. 1-19.

[2] Bickel, P.J. and Ritov, Y.(1987), "Efficient estimation in the errors in the variables model," *Annals of Statistics,* 3, pp. 1038-1069.

[3] Fuller, W.A.(1987), *Measurement error models*, Wiley, New York.

[4] Kendall, M. and Stuart, A.(1979), *The Advanced Theory of Statistics* ,Vol. 2, 4-th ed. MacMillan, New York.

[5] Villegas, C.(1969), "On the least squares estimation of non-linear relations," *Annals of Mathematical Staistics,* 40, pp. 462-466.