

Estimation in a linear errors-in-variables model under a mixture of classical and Berkson errors

Mykyta Yakovliev*, Alexander Kukush

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

mykyta.yakovliev@gmail.com (M. Yakovliev), alexander_kukush@univ.keiv.ua (A. Kukush)

Received: 10 March 2021, Revised: 16 June 2021, Accepted: 28 June 2021,
Published online: 26 July 2021

Abstract A linear structural regression model is studied, where the covariate is observed with a mixture of the classical and Berkson measurement errors. Both variances of the classical and Berkson errors are assumed known. Without normality assumptions, consistent estimators of model parameters are constructed and conditions for their asymptotic normality are given. The estimators are divided into two asymptotically independent groups.

Keywords Linear errors-in-variables model, mixture of the classical and Berkson errors, consistent estimators, asymptotically independent estimators

2010 MSC 62505, 62H12

1 Introduction

Regression models with measurement errors in covariates are quite popular nowadays [1, 2, 4], see also [5] for the comparison of various estimation methods in such models.

We consider a linear regression model under the presence of the classical and Berkson errors in the covariate:

$$y = \beta_0 + \beta_1\xi + \varepsilon, \quad w = x + \delta, \quad \xi = x + u. \quad (1.1)$$

*Corresponding author.

Here, y is the observable response variable, ξ and x are unobservable latent variables, w is the observable surrogate variable; ε , δ and u are centred errors, ε is error in response, δ is the classical measurement error, and u is Berkson measurement error; random variables x , ε , δ and u are independent.

In model (1.1), we have a mixture of the classical and Berkson errors. Let D stand for the variance. Indicate two extreme cases.

- (a) $D\delta = 0$, then $\delta = 0$, and (1.1) yields a linear model with Berkson error [1, 2]

$$y = \beta_0 + \beta_1\xi + \varepsilon, \quad \xi = w + u. \tag{1.2}$$

- (b) $Du = 0$, then $u = 0$, and (1.1) yields a linear model with the classical error [1, 2]

$$y = \beta_0 + \beta_1\xi + \varepsilon, \quad w = \xi + \delta. \tag{1.3}$$

Thus, the model (1.1) combines seminal models (1.2) and (1.3).

Models with a mixture of the classical and Berkson errors appear in radio-epidemiology. In [4, Section 7.2] the following measurement error model is considered:

$$D_i^{mes} = \bar{D}_i^{tr} + \sigma_i\gamma_i, \quad D_i^{tr} = \bar{D}_i^{tr} \delta_{F,i}. \tag{1.4}$$

Here, D_i^{mes} is the measured individual instrumental absorbed thyroid dose for the i th person of a cohort of persons residing in Ukrainian regions that suffered from the Chernobyl accident, D_i^{tr} is the corresponding true absorbed thyroid dose (i.e., the first latent variable), \bar{D}_i^{tr} is the second latent variable; $\sigma_i\gamma_i$ is the additive classical error, $\delta_{F,i}$ is the multiplicative Berkson error, σ_i is the standard deviation of the heteroscedastic classical measurement error, γ_i is standard normal and $\delta_{F,i}$ is a lognormal random variable; \bar{D}_i^{tr} , γ_i and $\delta_{F,i}$ are independent random variables.

In [4], the model (1.4) is combined with the binary model which resembles a logistic one:

$$P(Y_i = 1|D_i^{tr}) = \frac{\lambda_i}{1 + \lambda_i}, \quad P(Y_i = 0|D_i^{tr}) = \frac{1}{1 + \lambda_i}, \tag{1.5}$$

where λ_i is the total incidence rate related to cases of thyroid cancer,

$$\lambda_i = \lambda_0 + EAR \cdot D_i^{tr}. \tag{1.6}$$

Here, positive regression coefficients λ_0 and EAR are the background incidence rate and the excess absolute risk, respectively. In the binary model (1.5), (1.6), (1.4), the observed sample consists of pairs (Y_i, D_i^{mes}) , $i = 1, \dots, N$, where $Y_i = 1$ in the case of detected disease, and $Y_i = 0$ in the absence of disease within some time interval.

The presented linear model (1.1) is a simplified analogue of the binary measurement error model, where ξ , w and x are counterparts of D_i^{tr} , D_i^{mes} and \bar{D}_i^{tr} , respectively, and the binary model (1.5), (1.6) is replaced with the linear regression, and the multiplicative Berkson error $\delta_{F,i}$ is replaced with the additive Berkson error u .

The goal of the present paper is to study asymptotic properties of estimators of model parameters in the linear regression (1.1). The modest aim is to have a better understanding of the binary model (1.5), (1.6), (1.4) and similar models.

The paper is organized as follows. In Section 2, we present the observation model in more detail, and under the normality of x and u , derive from the underlying model the one like (1.3) with the classical error only. At that we obtain consistent estimators for β_0 and β_1 which unexpectedly coincide with the adjusted least squares estimators [2, 4], constructed by ignoring Berkson error u . The proposed estimators remain consistent without the normality of x and u . Section 3 gives conditions for the asymptotic normality of the estimators, and we divide them into two asymptotically independent groups. In doing so, we reparametrize the model similarly to [3], where the basic model (1.3) was studied. Section 4 concludes our findings.

We use the following notation. The symbol \mathbf{E} denotes expectation and acts as an operator on the total product of quantities, \mathbf{cov} stands for the covariance of two random variables and for the covariance matrix of a random vector. The upper index \top denotes transposition. In the paper, all the vectors are column ones. The bar means averaging over $i = 1, \dots, n$, e.g., $\bar{a} := n^{-1} \sum_{i=1}^n a_i$, $\overline{ab^\top} := n^{-1} \sum_{i=1}^n a_i b_i^\top$. Sample covariance of random variables $\{a_i, b_i, i = 1, \dots, n\}$ is denoted as S_{ab} , i.e. $S_{ab} = n^{-1} \sum_{i=1}^n (a_i - \bar{a})(b_i - \bar{b})$. Convergence with probability 1 and in distribution are denoted as $\xrightarrow{P1}$ and \xrightarrow{d} , respectively. A sequence of random variables that converges to zero in probability is denoted as $o_p(1)$, and a sequence of bounded in probability random variables is denoted as $O_p(1)$. I_p stands for the identity $p \times p$ matrix.

2 Construction of estimators for the normal latent variable and the normal Berkson error

2.1 Model and assumptions

We consider the structural model (1.1). Denote $\mu = \mathbf{E}\xi$ and let $\sigma_y^2, \sigma_\xi^2, \sigma_\varepsilon^2, \sigma_w^2, \sigma_x^2, \sigma_\delta^2$ and σ_u^2 be the variances of $y, \xi, \varepsilon, w, x, \delta$ and u , respectively. We need the following conditions for the consistency of the proposed estimators of model parameters.

- (i) Random variables x, ε, δ and u are independent.
- (ii) Random variables ε, δ and u have zero expectations and finite variances, and x has a finite and positive variance σ_x^2 .
- (iii) Variances of σ_δ^2 and σ_u^2 are positive and known, and other model parameters $\beta_0, \beta_1, \mu, \sigma_\varepsilon^2, \sigma_x^2$ are unknown.

Consider independent copies of model (1.1):

$$y_i = \beta_0 + \beta_1 \xi_i + \varepsilon_i, \quad w_i = x_i + \delta_i, \quad \xi_i = x_i + u_i, \quad i = 1, 2, \dots$$

Under assumption (i), this means that random vectors $(x_i, \varepsilon_i, \delta_i, u_i)^\top, i = 1, 2, \dots$, are i.i.d. and have the same distribution as $(x, \varepsilon, \delta, u)^\top$. Based on observations $(y_i, w_i), i = 1, \dots, n$, we want to estimate the unknown model parameters.

Remark 1. We allow $\sigma_\varepsilon^2 = 0$. The corresponding model (with $\varepsilon = 0$) is called data model.

Now, we explain why we impose condition (iii). The classical errors-in-variables model (1.3), with normally distributed ξ, ε and δ , and unknown 6 parameters $\beta_0, \beta_1, \mu, \sigma_\xi^2, \sigma_\varepsilon^2, \sigma_\delta^2$ is not identifiable [2]. Hence for the model (1.1), condition (iii) assumes σ_δ^2 to be known. The next statement explains why we suppose that σ_u^2 is known as well.

Lemma 1. *Consider the model (1.1) under conditions (i) and (ii). Let σ_δ^2 be known and random variables x, ε, δ and u be Gaussian. Then this model with 6 unknown parameters $\beta_0, \beta_1, \mu, \sigma_x^2, \sigma_\varepsilon^2, \sigma_u^2$ is not identifiable.*

Proof. The distribution of the observed Gaussian vector $Z := (y, w)^\top$ is uniquely defined by EZ and $C := \text{cov}(Z)$. Introduce two different collections of model parameters:

- (a) $\beta_0 = 0, \beta_1 = 1, \mu = 0, \sigma_x^2 = 1, \sigma_\varepsilon^2 = 1, \sigma_u^2 = 1$, and
- (b) $\beta_0 = 0, \beta_1 = 1, \mu = 0, \sigma_x^2 = 1, \sigma_\varepsilon^2 = 0.5, \sigma_u^2 = 1.5$.

In both cases it holds

$$EZ = 0, C = \begin{bmatrix} 3 & 1 \\ 1 & 3 + \sigma_\delta^2 \end{bmatrix}.$$

Therefore, the distribution of Z is the same for both collections of parameters, and the model is not identifiable. □

Notice that under conditions of Lemma 1, the parameters β_0 and β_1 are identifiable (see [2] for the definition of an identifiable parameter). Moreover, in the next subsection we will construct consistent estimators, as $n \rightarrow \infty$, for β_0 and β_1 under the only known parameter σ_δ^2 .

2.2 Consistent estimators of model parameters

Now, besides conditions (i) to (iii), we assume the following.

- (iv) Random variables x and u are Gaussian.

Now, out of (1.1) we derive a linear model with the classical error only. The conditional distribution of x given ξ is as follows [1, 4]:

$$L(x|\xi) = N(K\xi + (1 - K)\mu, K\sigma_u^2),$$

$K := \sigma_x^2/\sigma_\xi^2$ is the reliability ratio [2], $0 \leq K \leq 1$. Moreover, x can be decomposed as

$$x = K\xi + (1 - K)\mu + \sqrt{K}\sigma_u\gamma, \gamma \sim N(0, 1),$$

where $\xi, \gamma, \varepsilon, \delta$ are mutually independent. Then

$$\begin{aligned} w &= K\xi + (1 - K)\mu + \sqrt{K}\sigma_u\gamma + \delta. \\ \frac{w}{K} - \frac{1 - K}{K}\mu &= \xi + \frac{\sigma_u}{\sqrt{K}}\gamma + \frac{\delta}{K}. \end{aligned}$$

Introduce new variables

$$z := \frac{w}{K} - \frac{1-K}{K}\mu, \quad v := \frac{\sigma_u}{\sqrt{K}}\gamma + \frac{\delta}{K}.$$

We derived a linear model with the classical error:

$$y = \beta_0 + \beta_1\xi + \varepsilon, \quad z = \xi + v, \tag{2.1}$$

with independent ξ, ε, v and $\sigma_v^2 := Dv = \sigma_u^2/K + \sigma_\delta^2/K^2$.

Suppose at the moment that K is known. Then the adjusted least squares (ALS) estimator $\tilde{\beta}_1$ of β_1 is consistent and given as [2, 4]:

$$\tilde{\beta}_1 := \frac{S_{zy}}{S_{zz} - \sigma_v^2} = \frac{\frac{1}{K}S_{wy}}{\frac{1}{K^2}S_{ww} - \sigma_v^2} = \frac{S_{wy}}{\frac{S_{ww} - \sigma_\delta^2}{K} - \sigma_u^2}. \tag{2.2}$$

When K is unknown, we can estimate it consistently as

$$\hat{K} = \frac{\hat{\sigma}_x^2}{\hat{\sigma}_x^2 + \sigma_u^2} = \frac{S_{ww} - \sigma_\delta^2}{S_{ww} - \sigma_\delta^2 + \sigma_u^2}. \tag{2.3}$$

Now, we insert (2.3) into (2.2) instead of K and obtain the desired estimator

$$\hat{\beta}_1 = \frac{S_{wy}}{S_{ww} - \sigma_\delta^2}. \tag{2.4}$$

Next, in model (2.1) the ALS estimator of β_0 is as follows [2, 4]:

$$\tilde{\beta}_0 := \bar{y} - \tilde{\beta}_1\bar{z} = \bar{y} - \tilde{\beta}_1\left(\frac{\bar{w}}{K} - \frac{1-K}{K}\mu\right).$$

But K and μ are unknown, and instead of them we substitute the corresponding consistent estimators (2.3) and

$$\hat{\mu} := \bar{w}. \tag{2.5}$$

Then $\tilde{\beta}_1$ changes to $\hat{\beta}_1$, and we obtain the desired estimator

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\left(\frac{\bar{w}}{\hat{K}} - \frac{1-\hat{K}}{\hat{K}}\bar{w}\right) = \bar{y} - \hat{\beta}_1\bar{w}. \tag{2.6}$$

It is remarkable that $\hat{\beta}_0$ and $\hat{\beta}_1$ are the so-called naive ALS estimators in the model (1.1), where we neglected the presence of the Berkson error u . To be precise, $\hat{\beta}_0$ and $\hat{\beta}_1$ are the ALS estimators for the classical model (1.3). The estimators (2.4), (2.6) use σ_δ^2 but not σ_u^2 .

In our model, we have to estimate 5 parameters $\beta_0, \beta_1, \mu, \sigma_x^2, \sigma_\varepsilon^2$. We possess already 3 estimators (2.6), (2.4) and (2.5). Moreover, we used the estimator

$$\hat{\sigma}_x^2 = S_{ww} - \sigma_\delta^2. \tag{2.7}$$

Finally, in the model (2.1) the ALS estimator of σ_ε^2 is as follows [4]:

$$\tilde{\sigma}_\varepsilon^2 = S_{yy} - \tilde{\beta}_1 S_{zy} = S_{yy} - \frac{\tilde{\beta}_1}{K} S_{wy}.$$

Instead of unknown K , we substitute (2.3) and get the final estimator

$$\hat{\sigma}_\varepsilon^2 := S_{yy} - \frac{\hat{\beta}_1 S_{wy} (S_{ww} - \sigma_\delta^2 + \sigma_u^2)}{S_{ww} - \sigma_\delta^2}. \tag{2.8}$$

Though we derived the estimators under the normality assumption (iv), they remain consistent without this restriction.

Theorem 1. *In model (1.1), assume conditions (i)–(iii). Then there exists a random number n_0 such that expressions (2.4), (2.6), (2.5), (2.7), (2.8) are well defined with probability 1 for all $n \geq n_0$ and yield strongly consistent estimators of $\beta_1, \beta_0, \mu, \sigma_x^2, \sigma_\varepsilon^2$, respectively, i.e., they converge a.s. to the corresponding true values as $n \rightarrow \infty$.*

Proof. Here, we check the strong consistency of $\hat{\beta}_1$ only. We have

$$\begin{aligned} \mathbf{cov}(w, y) &= \mathbf{cov}(x, \beta_1 \xi) = \beta_1 \mathbf{cov}(x, x + u) = \beta_1 \sigma_x^2, \\ \hat{\beta}_1 &\xrightarrow{P1} \frac{\mathbf{cov}(w, y)}{Dw - \sigma_\delta^2} = \frac{\beta_1 \sigma_x^2}{\sigma_x^2} = \beta_1, \end{aligned}$$

where $\xrightarrow{P1}$ denotes the convergence with probability 1 and indicates the strong consistency of the estimator. □

3 Asymptotic normality of the estimators

3.1 Asymptotic variance of the estimator of slope coefficient

We need the following moment assumption.

(v) $E\delta^4 < \infty$.

Theorem 2. *Assume conditions (i), (ii), (v), and suppose that σ_δ^2 is known and positive. Then the estimator (2.4) is asymptotically normal, in more detail,*

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} N(0, \sigma_{\beta_1}^2), \tag{3.1}$$

where

$$\sigma_{\beta_1}^2 := \frac{\beta_1^2(\sigma_\delta^2 \sigma_x^2 + D(\delta^2) + \sigma_u^2 \sigma_w^2) + \sigma_\varepsilon^2 \sigma_w^2}{\sigma_x^4}. \tag{3.2}$$

Proof. We follow the line of the proof of Theorem 2.22 [4], and use expansions of sample covariances and Slutsky’s lemma [4, p. 44]. We centralize x as

$$\rho := x - \mu.$$

Then

$$y = \beta_0 + \beta_1\mu + \beta_1\rho + \varepsilon + \beta_1u, \quad w = \mu + \rho + \delta,$$

$$S_{wy} = S_{\rho+\delta, \beta_1\rho+\varepsilon+\beta_1u} = \beta_1 S_{\rho\rho} + S_{\rho\varepsilon} + \beta_1 S_{\rho u} + \beta_1 S_{\delta\rho} + S_{\delta\varepsilon} + \beta_1 S_{\delta u}, \quad (3.3)$$

$$S_{ww} = S_{\rho+\delta, \rho+\delta} = S_{\rho\rho} + 2S_{\rho\delta} + S_{\delta\delta}, \quad (3.4)$$

$$S_{ww} - \sigma_\delta^2 = \sigma_x^2 + o_p(1). \quad (3.5)$$

Using (2.4) and expansions (3.3)–(3.5), we obtain

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) = \frac{-\beta_1\sqrt{n}(S_{\delta\delta} - \sigma_\delta^2 - S_{u\rho} + S_{\delta\rho} - S_{u\delta}) + \sqrt{n}(S_{\rho\varepsilon} + S_{\delta\varepsilon})}{\sigma_x^2 + o_p(1)}. \quad (3.6)$$

Next,

$$S_{\delta\delta} - \sigma_\delta^2 = \bar{\delta}^2 - \sigma_\delta^2 + \frac{o_p(1)}{\sqrt{n}},$$

$$S_{u\rho} = \bar{u}\bar{\rho} + \frac{o_p(1)}{\sqrt{n}}, \quad S_{\delta\rho} = \bar{\delta}\bar{\rho} + \frac{o_p(1)}{\sqrt{n}},$$

$$S_{u\delta} = \bar{u}\bar{\delta} + \frac{o_p(1)}{\sqrt{n}}, \quad S_{\rho\varepsilon} = \bar{\rho}\bar{\varepsilon} + \frac{o_p(1)}{\sqrt{n}}, \quad S_{\delta\varepsilon} = \bar{\delta}\bar{\varepsilon} + \frac{o_p(1)}{\sqrt{n}}.$$

We insert these relations into (3.6) and get

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) = o_p(1) + \frac{-\beta_1\sqrt{n}(\bar{\delta}^2 - \sigma_\delta^2 - \bar{u}\bar{\rho} + \bar{\delta}\bar{\rho} - \bar{u}\bar{\delta}) + \sqrt{n}(\bar{\rho}\bar{\varepsilon} + \bar{\delta}\bar{\varepsilon})}{\sigma_x^2}. \quad (3.7)$$

Using condition (v) and Central Limit Theorem, we get

$$\begin{aligned} \sqrt{n}(\bar{\delta}^2 - \sigma_\delta^2, \bar{u}\bar{\rho}, \bar{\delta}\bar{\rho}, \bar{u}\bar{\delta}, \bar{\rho}\bar{\varepsilon}, \bar{\delta}\bar{\varepsilon})^\top &\xrightarrow{d} \gamma = (\gamma_i)_1^6 \sim N_6(0, S), \\ S &= \text{diag}(D(\delta^2), \sigma_u^2\sigma_x^2, \sigma_\delta^2\sigma_x^2, \sigma_\delta^2\sigma_u^2, \sigma_x^2\sigma_\varepsilon^2, \sigma_\delta^2\sigma_\varepsilon^2). \end{aligned}$$

The diagonal of S contains variances of averaged random variables, e.g., $S_{22} = D(u\rho) = E(u\rho)^2 = \sigma_u^2\sigma_x^2$, and off-diagonal entries of S are vanishing because $\delta, \rho, \varepsilon, u$ are independent. Then the numerator in (3.7) converges in distribution to

$$-\beta_1(\gamma_1 - \gamma_2 + \gamma_3 - \gamma_4) + \gamma_5 + \gamma_6 \sim N(0, \beta_1^2(S_{11} + S_{22} + S_{33} + S_{44}) + S_{55} + S_{66}).$$

Relation (3.7) and Slutsky’s lemma imply (3.1) with

$$\sigma_{\hat{\beta}_1}^2 = \frac{\beta_1^2(D(\delta^2) + \sigma_u^2\sigma_x^2 + \sigma_\delta^2\sigma_x^2 + \sigma_u^2\sigma_\delta^2) + \sigma_x^2\sigma_\varepsilon^2 + \sigma_\delta^2\sigma_\varepsilon^2}{\sigma_x^4}. \quad (3.8)$$

Since $\sigma_w^2 = \sigma_x^2 + \sigma_\delta^2$, the right-hand sides of (3.8) and (3.2) coincide. □

Remark 2. If condition (v) is replaced with the assumption $\delta \sim N(0, \sigma_\delta^2)$, then $D(\delta^2) = 2\sigma_\delta^4$ and (3.2) is simplified as

$$\sigma_{\hat{\beta}_1}^2 = \frac{\beta_1^2(\sigma_\delta^4 + \sigma_\delta^2\sigma_w^2 + \sigma_u^2\sigma_w^2) + \sigma_w^2\sigma_\varepsilon^2}{\sigma_x^4}. \quad (3.9)$$

The convergence (3.1), (3.2) can be applied to construct the asymptotic confidence interval for β_1 . For this purpose we have to ensure that $\sigma_{\hat{\beta}_1}^2 > 0$ (this holds if either $\sigma_\delta^2 > 0$ or $\beta_1 \neq 0$) and to estimate $\sigma_{\hat{\beta}_1}^2$ consistently. The latter is possible for normal δ , since all the parameters on the right-hand side of (3.9) are estimated consistently due to Theorem 1. Without the normality of δ , it is problematic to estimate the 4th moment $D(\delta^2)$ in (3.8). If $E\delta^4$ is assumed known, then $\sigma_{\hat{\beta}_1}^2$ can be estimated consistently as well.

Analysis of formula (3.8) allows to find out in which proportion the classical error and Berkson one affect the quality of the slope estimation. Denote

$$\Lambda_u = |\beta_1| \sigma_u \sigma_x, \quad \Lambda_\delta = \sqrt{\beta_1^2 (\sigma_\delta^2 \sigma_x^2 + D(\delta^2)) + \sigma_\delta^2 \sigma_\varepsilon^2}, \quad \Lambda_{u\delta} = |\beta_1| \sigma_u \sigma_\delta.$$

Then

$$\sigma_{\hat{\beta}_1}^2 = \sigma_x^{-4} (\Lambda_\delta^2 + \Lambda_u^2 + \Lambda_{u\delta}^2 + \sigma_x^2 \sigma_\varepsilon^2).$$

The normalized slope estimator $\sqrt{n}(\hat{\beta}_1 - \beta_1)$ can be approximated in distribution by a random variable

$$\sigma_x^{-2} (\Lambda_\delta \gamma_1 + \Lambda_u \gamma_2 + \Lambda_{u\delta} \gamma_3 + \sigma_x \sigma_\varepsilon \gamma_4),$$

with i.i.d. standard normal $\gamma_1, \dots, \gamma_4$. Given σ_x^2 and σ_ε^2 , terms $\Lambda_\delta \gamma_1$, $\Lambda_u \gamma_2$ and $\Lambda_{u\delta} \gamma_3$ distinguish the influence of the classical error, Berkson error and of the cumulative effect from both errors, respectively, on the precision of the slope estimator. Thus, this influence can be evaluated in proportion $\Lambda_\delta : \Lambda_u : \Lambda_{u\delta}$. Suppose that $\beta_1 \neq 0$ and $E\delta^4$ is known. The influence can be estimated in proportion $\hat{\Lambda}_\delta : \hat{\Lambda}_u : \hat{\Lambda}_{u\delta}$, with

$$\hat{\Lambda}_u := |\hat{\beta}_1| \sigma_u \hat{\sigma}_x, \quad \hat{\Lambda}_\delta := \sqrt{\hat{\beta}_1^2 (\sigma_\delta^2 \hat{\sigma}_x^2 + D(\delta^2)) + \sigma_\delta^2 \hat{\sigma}_\varepsilon^2}, \quad \hat{\Lambda}_{u\delta} := |\hat{\beta}_1| \sigma_u \sigma_\delta.$$

3.2 Asymptotic independence of groups of estimators

We slightly reparametrize the model (1.1) to a form

$$y = \mu_y + \beta_1(\xi - \mu) + \varepsilon, \quad w = x + \delta, \quad \xi = x + u. \tag{3.10}$$

This model is obtained from (1.1) after introducing a new parameter $\mu_y = \beta_0 + \beta_1 \mu$ in place of β_0 . Based on independent copies of the model

$$y_i = \mu_y + \beta_1(\xi_i - \mu) + \varepsilon_i, \quad w_i = x_i + \delta_i, \quad \xi_i = x_i + u_i$$

(here, independent random vectors $(x_i, \varepsilon_i, \delta_i, u_i)$, $i \geq 1$, are distributed as a random vector $(x, \varepsilon, \delta, u)$ in (3.10)) and on observations (y_i, w_i) , $i = 1, \dots, n$, we estimate a vector of unknown parameters

$$\theta = (\mu, \mu_y, \sigma_w^2, \beta_1, \sigma_\varepsilon^2)^\top,$$

assuming condition (iii) which states that σ_δ^2 and σ_u^2 are known. For model (3.10), we assume also (i), (ii) and impose two additional assumptions.

(vi) x, δ, u and ε have zero skewness, i.e., their centered third moments are zeros.

(vii) $\sigma_\varepsilon^2 > 0$, $E\varepsilon^4 < \infty$ and the distribution of ε is not concentrated at two points.

Theorem 1 implies that a strongly consistent estimator

$$\widehat{\theta} = (\widehat{\mu}, \widehat{\mu}_y, \widehat{\sigma}_w^2, \widehat{\beta}_1, \widehat{\sigma}_\varepsilon^2)^\top$$

of θ can be defined explicitly as

$$\widehat{\theta} = (\overline{w}, \overline{y}, S_{ww}, \frac{S_{wy}}{S_{ww} - \sigma_\delta^2}, S_{yy} - \frac{S_{wy}^2(S_{ww} - \sigma_\delta^2 + \sigma_u^2)}{(S_{ww} - \sigma_\delta^2)^2})^\top. \tag{3.11}$$

Introduce the corresponding estimation function

$$\begin{aligned} s &= s(\theta; w, y) = (s^\mu, s^{\mu_y}, s^{\sigma_w^2}, s^{\beta_1}, s^{\sigma_\varepsilon^2})^\top, \tag{3.12} \\ s^\mu &:= w - \mu, \quad s^{\mu_y} := y - \mu_y, \quad s^{\sigma_w^2} := (w - \mu)^2 - \sigma_w^2, \\ s^{\beta_1} &:= \beta_1(w - \mu)^2 - \beta_1\sigma_\delta^2 - (w - \mu)(y - \mu_y), \\ s^{\sigma_\varepsilon^2} &:= (y - \mu_y)^2 - \sigma_\varepsilon^2 - \beta_1^2(w - \mu)^2 + \beta_1^2(\sigma_\delta^2 - \sigma_u^2). \end{aligned}$$

With probability one, the estimator (3.11) satisfies the estimating equation

$$\sum_{i=1}^n s(\widehat{\theta}; w_i, y_i) = 0.$$

Definition 1. Let $\widehat{\alpha}$ and $\widehat{\beta}$ be asymptotically normal estimators of $\alpha \in \mathbb{R}^p$ and $\beta \in \mathbb{R}^q$, respectively, such that

$$\sqrt{n} \begin{bmatrix} \widehat{\alpha} - \alpha \\ \widehat{\beta} - \beta \end{bmatrix} \xrightarrow{d} N_{p+q}(0, \Sigma) \text{ as } n \rightarrow \infty,$$

with a nonsingular asymptotic covariance matrix Σ . The estimators $\widehat{\alpha}$ and $\widehat{\beta}$ are called asymptotically independent if Σ can be partitioned as

$$\Sigma = \text{block-diag}(\Sigma_\alpha, \Sigma_\beta) = \begin{bmatrix} \Sigma_\alpha & 0 \\ 0 & \Sigma_\beta \end{bmatrix},$$

with $\Sigma_\alpha \in \mathbb{R}^{p \times p}$ and $\Sigma_\beta \in \mathbb{R}^{q \times q}$.

It is convenient to deal with asymptotically independent estimators $\widehat{\alpha}$ and $\widehat{\beta}$, because asymptotic confidence region for the augmented parameter $(\alpha^\top \beta^\top)^\top$ can be constructed as the Cartesian product of asymptotic confidence ellipsoids for α and β .

Theorem 3. Assume conditions (i)–(iii), (vii) and that x, δ, u have finite 4th moments. Then:

(a) the estimator (3.11) in model (3.10) is asymptotically normal, in more detail,

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} N_5(0, \Sigma^\theta) \tag{3.13}$$

with a nonsingular asymptotic covariance matrix Σ^θ ;

(b) under additional assumption (vi), groups of estimators $(\widehat{\mu}, \widehat{\mu}_y)^\top$ and $(\widehat{\sigma}_w^2, \widehat{\beta}_1, \widehat{\sigma}_\varepsilon^2)^\top$ are asymptotically independent.

Proof. (a) We prove (3.13) with a nonsingular Σ^θ .

1. Since all the variances in the underlying model are assumed positive, the true vector θ is an inner point of the parameter set $\Theta = \mathbb{R}^2 \times (0, \infty) \times \mathbb{R} \times (0, \infty)$.

As was mentioned above, $\widehat{\theta}$ is strongly consistent. The estimating function (3.12) is unbiased, i.e. $\mathbf{E}_\theta s(\theta; w, y) = 0$. Introduce two matrices

$$V := -\mathbf{E}_\theta \frac{\partial s(\theta; w, y)}{\partial \theta^\top} = \text{block-diag}(I_2, V_2),$$

with

$$V_2 := \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\sigma_x^2 & 0 \\ 0 & 2\beta_1(\sigma_x^2 + \sigma_u^2) & 1 \end{bmatrix},$$

and

$$B := \mathbf{cov}_\theta(s(\theta; w, y)).$$

Since $\varepsilon, x, \delta, u$ have finite 4th moments, B is well defined.

The unbiasedness of $s(\theta; w, y)$, consistency of $\widehat{\theta}$ and nonsingularity of V imply (3.13) by Theorem A.26 from [4], and Σ^θ can be found by the sandwich formula

$$\Sigma^\theta = V^{-1} B (V^{-1})^\top.$$

2. It remains to prove that B is nonsingular. For this purpose, we have to show that the five random variables

$$\begin{aligned} s^\mu &= s^\mu(\theta; w, y), \quad s^{\mu_y} = s^{\mu_y}(\theta; w, y), \quad s^{\sigma_w^2} = s^{\sigma_w^2}(\theta; w, y), \\ s^{\beta_1} &= s^{\beta_1}(\theta; w, y), \quad s^{\sigma_\varepsilon^2} = s^{\sigma_\varepsilon^2}(\theta; w, y) \end{aligned}$$

are linearly independent for the true value of θ .

Consider a random vector

$$h := (w - \mu, y - \mu_y, (w - \mu)^2, (w - \mu)(y - \mu_y), (y - \mu_y)^2)^\top. \quad (3.14)$$

It holds

$$(s^\mu, s^{\mu_y}, s^{\sigma_w^2}, s^{\beta_1}, s^{\sigma_\varepsilon^2})^\top = Th + a,$$

where $T = T(\theta)$ is a nonsingular square matrix and $a = a(\theta)$ is a nonrandom vector. The matrix T is nonsingular, hence it is enough to show that neither nontrivial linear combination of the components of h is a constant.

We use the centralization $\rho = x - \mu$. Suppose that for some real numbers $a_{11}, a_{12}, a_{22}, a_1, a_2$ and a_3 , the following holds with probability one:

$$\begin{aligned} F &:= a_{11}(w - \mu)^2 + a_{12}(w - \mu)(y - \mu_y) + a_{22}(y - \mu_y)^2 + \\ &\quad + a_1(w - \mu) + a_2(y - \mu_y) + a_3 = 0, \\ F &= a_{11}(\rho + \delta)^2 + a_{12}(\rho + \delta)(\beta_1\rho + \beta_1u + \varepsilon) + a_{22}(\beta_1\rho + \beta_1u + \varepsilon)^2 + \\ &\quad + a_1(\rho + \delta) + a_2(\beta_1\rho + \beta_1u + \varepsilon) + a_3 = 0. \end{aligned}$$

Then a.s.

$$0 = \mathbf{E}[F|\varepsilon] = a_{22}\varepsilon^2 + a_2\varepsilon + b_3, \quad b_3 \in \mathbb{R},$$

hence due to condition (vii), $a_{22} = a_2 = 0$. And we have a.s.

$$0 = E[F|\delta] = a_{11}\delta^2 + a_1\delta + c_3, \quad c_3 \in \mathbb{R}. \tag{3.15}$$

Consider two cases about the support of δ .

2.1. Here we suppose that δ is not concentrated at two points. Then (3.15) implies $a_{11} = a_1 = 0$. Next, a.s.

$$\begin{aligned} 0 &= E[F|\varepsilon, \delta] = a_{12}\delta\varepsilon + d_3, \quad d_3 \in \mathbb{R}, \\ 0 &= D(a_{12}\delta\varepsilon) = a_{12}^2\sigma_\delta^2\sigma_\varepsilon^2, \quad a_{12} = 0, \end{aligned}$$

and we get the desired

$$a_{11} = a_{12} = a_{22} = a_1 = a_2 = 0. \tag{3.16}$$

2.2. Now, we suppose that for some $\delta_0 \neq 0$, it holds

$$P(\delta = \delta_0) = P(\delta = -\delta_0) = \frac{1}{2}.$$

Then with probability one,

$$\begin{aligned} F(\rho, \varepsilon, \delta_0, u) &= F(\rho, \varepsilon, -\delta_0, u) = 0, \\ 0 &= F(\rho, \varepsilon, \delta_0, u) - F(\rho, \varepsilon, -\delta_0, u) = 2\delta_0 G, \\ 0 &= G = 2a_{11}\rho + a_{12}(\beta_1\rho + \beta_1u + \varepsilon) + a_1, \\ 0 &= E[G|\varepsilon] = a_{12}\varepsilon + a_1, \quad a_{12} = a_1 = 0; \\ a_{11}\rho &= 0 \text{ a.s., } a_{11} = 0. \end{aligned}$$

Thus, in this case (3.16) holds as well. Statement (a) of Theorem 3 is proven.

(b) Now, we rely additionally on the assumption (vi) about vanishing centered third moments. By statement (a), B is nonsingular. We have to show that it has a block-diagonal structure

$$B = \text{block-diag}(B_1, B_2) \tag{3.17}$$

with some matrices $B_1 \in \mathbb{R}^{2 \times 2}$ and $B_2 \in \mathbb{R}^{3 \times 3}$, then Σ^θ will be block-diagonal as well, with nonsingular blocks:

$$\Sigma^\theta = \text{block-diag}(\Sigma_1, \Sigma_2), \quad \Sigma_1 = B_1, \quad \Sigma_2 = V_2^{-1}B_2(V_2^{-1})^T,$$

and statement (b) of Theorem 3 will be proven.

Using assumption (vi), we have:

$$\begin{aligned} \mathbf{cov}(s^\mu, s^{\sigma_w^2}) &= E(x - \mu)^3 + E\delta^3 = 0; \\ \mathbf{cov}(s^\mu, s^{\beta_1}) &= \mathbf{cov}(w - \mu, \beta_1(w - \mu)^2) - \mathbf{cov}(w - \mu, (w - \mu)(y - \mu_y)) = \\ &= -E(\rho + \delta)^2(\beta_1\rho + \beta_1u + \varepsilon) = -\beta_1 E\rho^3 = 0; \\ \mathbf{cov}(s^\mu, s^{\sigma_\varepsilon^2}) &= \mathbf{cov}(w - \mu, (y - \mu_y)^2) - \beta_1^2 \mathbf{cov}(w - \mu, (w - \mu)^2) = \\ &= E(w - \mu)(y - \mu_y)^2 = E(\rho + \delta)(\beta_1\rho + \beta_1u + \varepsilon)^2 = \beta_1^2 E\rho^3 = 0; \end{aligned}$$

$$\begin{aligned} \mathbf{cov}(s^{\mu_y}, s^{\sigma_w^2}) &= \beta_1 \mathbf{E}\rho^3 = 0; \\ \mathbf{cov}(s^{\mu_y}, s^{\beta_1}) &= \beta_1 \mathbf{E}(w - \mu)^2 (y - \mu_y) - \mathbf{E}(w - \mu)(y - \mu_y)^2 = \\ &= \beta_1^2 \mathbf{E}\rho^3 - \beta_1^2 \mathbf{E}\rho^3 = 0; \\ \mathbf{cov}(s^{\mu_y}, s^{\sigma_\varepsilon^2}) &= \mathbf{E}(y - \mu_y)^3 - \beta_1^2 \mathbf{E}(w - \mu)^2 (y - \mu_y) = \\ &= \mathbf{E}(\beta_1 \rho + \beta_1 u + \varepsilon)^3 - \beta_1^3 \mathbf{E}\rho^3 = \beta_1^3 \mathbf{E}u^3 + \mathbf{E}\varepsilon^3 = 0. \end{aligned}$$

This proves relation (3.17). □

Remark 3. Theorem 3 is not valid without condition (vii). Indeed, suppose that for some $\varepsilon_0 \neq 0$,

$$\mathbf{P}(\varepsilon = \varepsilon_0) = \mathbf{P}(\varepsilon = -\varepsilon_0) = \frac{1}{2}.$$

If additionally $\beta_1 = 0$ then $(y - \mu_y)^2 = \varepsilon^2 = \varepsilon_0^2$ a.s., and

$$F_0 := (y - \mu_y)^2 - \varepsilon_0^2 = 0 \text{ a.s.}$$

Thus, certain nontrivial linear combination of components of vector (3.14) is a constant, hence the block B_2 in (3.17) is singular, and the asymptotic covariance matrix Σ^θ is degenerate in this specific case.

4 Simulation study

We simulated test data in order to evaluate the cover probability for the asymptotic confidence interval of the slope parameter, which is constructed based on Theorem 2. Observations in model (1.1) were generated as follows: $x \sim N(-1, 1)$, $u \sim N(0, \sigma_u^2)$ with $\sigma_u^2 \in \{10i : i = 1, \dots, 15\}$, $\delta \sim N(0, 1)$, $\varepsilon \sim N(0, 1)$, $\beta_1 = 2$, $\beta_0 = 1$, with the sample size $n \in \{10i : i = 1, \dots, 10\}$. For each collection of model parameters, $N = 10,000$ realizations were generated. For each realization, the slope estimate and the estimate of its asymptotic variance were computed (here, we inserted into (3.2) the estimates of all unknown model parameters). For an ensemble of N realizations, the cover probability was calculated for constructed 95% asymptotic confidence intervals for the slope parameter. We briefly report the obtained results.

Figure 1 shows how the cover probability deviation decreases from 0.95 with increase of the sample size. This effect is stable for different values of the Berkson error variance. Figure 2 illustrates how the cover probability deviation increases from 0.95 with increase of the Berkson error variance. As can be seen in Figure 1, the latter effect is getting weaker with increase of the sample size.

5 Conclusion

We dealt with a linear observation model (1.1) with a mixture of the classical and Berkson errors in the covariate. Surprisingly enough, we constructed consistent estimators for the regression parameters without the knowledge of the variance of the Berkson error. Nevertheless, the size of the Berkson error makes influence on the asymptotic variances of $\hat{\beta}_0$ and $\hat{\beta}_1$.

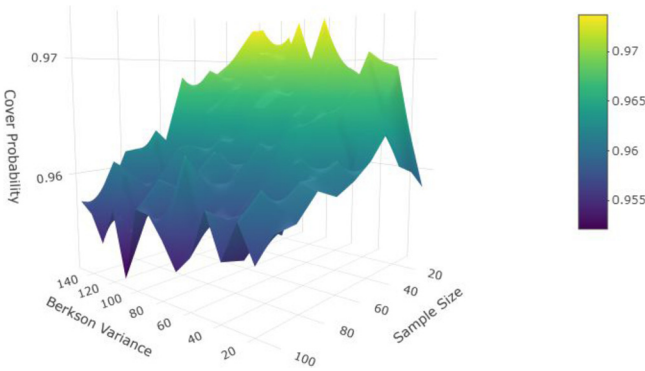


Fig. 1. Cover Probability Plot - Sample Size effect perspective

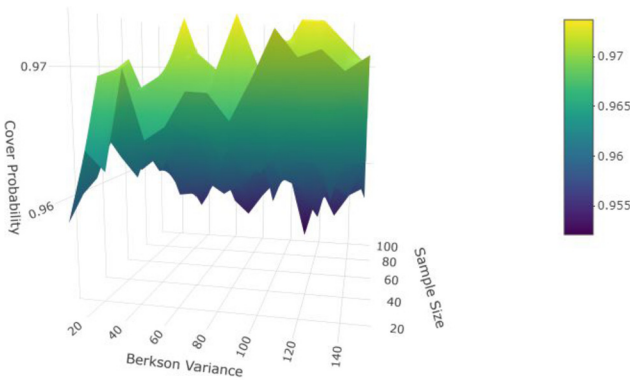


Fig. 2. Cover Probability Plot - Berkson effect perspective

Then we modified the model to an equivalent centralized form (3.10). This made possible to divide estimators of all unknown model parameters into two asymptotically independent groups.

In future we intend to consider the prediction problem for the model (1.1), like it was done in [6] for various measurement error models. Also it would be interesting to consider a polynomial model with a mixture of the classical and Berkson errors, as well as a version of linear model with a vector response and vector covariate.

Funding

Research is supported by the National Research Fund of Ukraine grant 2020.02/0026.

References

[1] Carroll, R.J., Ruppert, D., Stefanski, L.A., Crainiceanu, C.M.: Measurement Error in Non-linear Models: A Modern Perspective, 2nd edn. Monogr. Stat. Appl. Probab., vol. 105,

- p. 455. Chapman & Hall/CRC, Boca Raton, FL (2006). [MR2243417](#). <https://doi.org/10.1201/9781420010138>
- [2] Cheng, C.-L., Ness, J.W.V.: *Statistical Regression with Measurement Error*. Kendall's Library of Statistics, vol. 6. Arnold, London; co-published by Oxford University Press, New York (1999). [MR1719513](#)
- [3] Kukush, O.G., Tsaregorodtsev, Y.V., Shklyar, S.V.: Asymptotically independent estimators in a structural linear model with measurement errors. *Ukr. Math. J.* **68**(11), 1741–1755 (2017). [MR3621452](#). <https://doi.org/10.1007/s11253-017-1324-8>
- [4] Masiuk, S.V., Kukush, A.G., Shklyar, S.V., Chepurny, M.I., Likhtarov, I.A.: *Radiation Risk Estimation: Based on Measurement Error Models*, 2nd edn. De Gruyter Series in Mathematics and Life Sciences, vol. 5, p. 238. De Gruyter, Berlin (2017). [MR3726857](#). <https://doi.org/10.1515/9783110433661>
- [5] Schneeweiss, H., Kukush, A.: Comparing the efficiency of structural and functional methods in measurement error models. *Theory Probab. Math. Stat.* **80**, 131–142 (2010). [MR2541958](#). <https://doi.org/10.1090/S0094-9000-2010-00800-3>
- [6] Senko, I., Kukush, A.: Prediction in polynomial errors-in-variables models. *Mod. Stoch. Theory Appl.* **7**(2), 203–219 (2020). [MR4120615](#). <https://doi.org/10.15559/20-vmsta154>