Jackknife covariance matrix estimation for observations from mixture

Rostyslav Maiboroda*, **Olena Sugakova**

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

mre@univ.kiev.ua (R. Maiboroda), sugak@univ.kiev.ua (O. Sugakova)

Abstract   A general jackknife estimator for the asymptotic covariance of moment estimators is considered in the case when the sample is taken from a mixture with varying concentrations of components. Consistency of the estimator is demonstrated. A fast algorithm for its calculation is described. The estimator is applied to construction of confidence sets for regression parameters in the linear regression with errors in variables. An application to sociological data analysis is considered.

Keywords   Finite mixture model, orthogonal regression, mixture with varying concentrations, nonparametric estimation, asymptotic covariance matrix estimation, confidence ellipsoid, jackknife, errors-in-variables model

2010 MSC   62J05, 62G20

1   Introduction

Finite Mixture Models (FMM) are widely used in the analysis of biological, economic and sociological data. For a comprehensive survey of different statistical techniques based on FMMs, see [9]. Mixtures with Varying Concentrations (MVC) is a subclass of these models in which the mixing probabilities are not constant, but vary for different observations (see [4, 5]).

In this paper we consider application of the jackknife technique to the estimation of asymptotic covariance matrix (the covariance matrix for asymptotically normal

---

*Corresponding author.

VTeX

estimator, ACM) in the case when the data are described by the MVC model. The jackknife is a well-known resampling technique usually applied to i.i.d. samples (see Section 5.5.2 in [11], Chapter 4 in [15], Chapter 2 in [12]). On the jackknife estimates of variance for censored and dependent data, see [14]. Its modification to the case of the MVC model in which the observations are still independent but not identically distributed needs some efforts.

We obtained a general theorem on consistency of the jackknife estimators for ACM for moment estimators in the MVC models and apply this result to construct confidence sets for regression coefficients in linear errors-in-variables models for MVC data. On general errors-in-variables models, see [2, 3, 8]. The model and the estimators for the regression coefficients considered in this paper was proposed in [6], where the asymptotic normality of these estimates is shown.

The rest of the paper is organized as follows. In Section 2 we introduce the MVC model and describe the estimation technique for these models based on weighted moments. In Section 3 the jackknife estimates for the ACM are introduced and conditions of their consistency formulated. Section 4 is devoted to the algorithm of fast computation of the jackknife estimates. In Section 5 we apply the previous results to construct confidence sets for linear regression coefficients in errors-in-variables models with MVC. In Section 6 results of simulations are presented. In Section 7 we present results of application of the proposed technique to analyze sociological data. Proofs are placed in Section 8. Section 9 contains concluding remarks.

2   Mixtures with varying concentrations

We consider a dataset in which each observed subject O belongs to one of M subpopulations (mixture components). The number $\kappa(O)$ of the population which O belongs to is unknown. We observe d numeric characteristics of O which form the vector $\xi(O) = (\xi^1(O), \ldots, \xi^d(O))^T \in \mathbb{R}^d$ of observable variables. The distribution of $\xi(O)$ may depend on the component $\kappa(O)$:

$$F_\xi^{(m)}(A) = \mathsf{P}\{\xi(O) \in A \mid \kappa(O) = m\}, \quad m = 1, \ldots, M,$$

where A is any Borel subset of $\mathbb{R}^d$.

We observe variables of n independent subjects $\xi_j = \xi(O_j)$. The probability to obtain j-th subject from m-th component

$$p_j^{(m)} = \mathsf{P}\{\kappa(O_j) = m\}$$

can be considered as the concentration of the m-th component in the mixture when the j-th observation was made. The concentrations are known and can vary for different observations.

So, the distribution of ξ_j is described by the model of mixture with varying concentrations:

$$\mathsf{P}\{\xi_j \in A\} = \sum_{m=1}^{M} p_j^{(m)} F_\xi^{(m)}(A) \tag{1}$$

We will denote by

$$\mu^{(m)} = \mathsf{E}^{(m)}[\xi] = \mathsf{E}[\xi(O) \mid \kappa(O) = m] = \int_{\mathbb{R}^d} x F_\xi^{(m)}(dx)$$

the vector of theoretical first moments of the m-th component distribution. In what follows, $\text{Cov}^{(m)}[\xi]$ means the covariance of $\xi(O)$ for the m-th component, $\text{Var}^{(m)}[\xi^l]$ means the variance of $\xi^l(O)$ for this component and so on.

To estimate $\mu^{(k)}$ by observations $\xi_1, \ldots, \xi_n$ one can use the weighted sample mean

$$\bar{\xi}_{;n}^{(k)} = \bar{\xi}^{(k)} = \sum_{j=1}^{n} a_j^{(k)} \xi_j, \tag{2}$$

where $a_j^{(k)} = a_{j;n}^{(k)}$ are some weights dependent on components' concentrations, but not on the observed $\xi_j = \xi_{j;n}$. (In what follows we denote by the subscript $; n$ that the corresponding quantity is considered for the sample size n. In most cases this subscript is dropped to simplify notations.)

To obtain unbiased estimates in (2) one needs to select the weights satisfying the assumption

$$\sum_{j=1}^{n} a_j^{(k)} p_j^{(m)} = \begin{cases} 1 & \text{if } k = m, \\ 0 & \text{if } k \neq m. \end{cases} \tag{3}$$

Let us denote

$$\mathbf{\Xi} = (\xi_1^T, \ldots, \xi_n^T)^T = \begin{pmatrix} \xi_1^1 & \cdots & \xi_1^d \\ \vdots & \ddots & \vdots \\ \xi_n^1 & \cdots & \xi_n^d \end{pmatrix},$$

$$\mathbf{a} = \begin{pmatrix} a_1^{(1)} & \cdots & a_1^{(M)} \\ \vdots & \ddots & \vdots \\ a_n^{(1)} & \cdots & a_n^{(M)} \end{pmatrix} \quad \text{and} \quad \mathbf{p} = \begin{pmatrix} p_1^{(1)} & \cdots & p_1^{(M)} \\ \vdots & \ddots & \vdots \\ p_n^{(1)} & \cdots & p_n^{(M)} \end{pmatrix}.$$

Then $\mathbf{p}_{\boldsymbol{\cdot}}^{(m)} = (p_1^{(m)}, \ldots, p_n^{(m)})^T$, $\mathbf{p}_j^{\boldsymbol{\cdot}} = (p_j^{(1)}, \ldots, p_j^{(M)})^T$ and the same notation is used for the matrix $\mathbf{a}$.

In this notation the unbiasedness condition (3) reads

$$\mathbf{a}^T \mathbf{p} = \mathbb{E}, \tag{4}$$

where $\mathbb{E}$ means the $M \times M$ unit matrix.

There can be many choices of $\mathbf{a}$ satisfying (4). In [4, 5] minimax weights are considered defined by[2]

$$\mathbf{a} = \mathbf{p}\mathbf{\Gamma}^{-1}, \tag{5}$$

---

[2]In fact, in [4] and [5] the weights are defined as $\tilde{\mathbf{a}} = n\mathbf{a}$ and $\bar{\xi}^{(k)} = \frac{1}{n}\sum_{j=1}^{n} \tilde{a}_j^{(k)} \xi_j$. In this paper we adopt notation which allows to simplify formulas for fast estimator calculation in Section 4.

where

$$\mathbf{\Gamma} = \mathbf{\Gamma}_{;n} = \mathbf{p}^T \mathbf{p}$$

is the Gram matrix of the set of concentration vectors $\mathbf{p}_{\cdot}^{(1)}, \ldots, \mathbf{p}_{\cdot}^{(M)}$. In what follows, we assume that these vectors are linearly independent, so $\det \mathbf{\Gamma} > 0$ and $\mathbf{\Gamma}^{-1}$ exists. See [5] on the optimal properties of the estimates for concentration distributions based on the minimax weights (5).

To describe the asymptotic behavior of $\bar{\xi}_{;n}^{(k)}$ as $n \to \infty$, we will calculate its covariance matrix.

Notice that

$$\mathrm{Cov}[\xi_j] = \mathsf{E}[\xi_j \xi_j^T] - \mathsf{E}[\xi_j] \mathsf{E}[\xi_j]^T = \sum_{m=1}^{M} p_j^{(m)} \mathbf{\Sigma}^{(m)} - \sum_{m,l=1}^{M} p_j^{(m)} p_j^{(l)} \mu^{(m)} (\mu^{(l)})^T,$$

where $\mathbf{\Sigma}^{(m)} = \mathrm{Cov}^{(m)}[\xi] = \mathrm{Cov}[\xi(O) \mid \kappa(O) = m]$. So,

$$n \, \mathrm{Cov} \, \bar{\xi}^{(k)} = \sum_{m=1}^{M} \langle (\mathbf{a}^{(k)})^2 \mathbf{p}^{(m)} \rangle_{;n} \mathbf{\Sigma}^{(m)} - \sum_{m,l=1}^{M} \langle (\mathbf{a}^{(k)})^2 \mathbf{p}^{(m)} \mathbf{p}^{(l)} \rangle_{;n} \mu^{(m)} (\mu^{(l)})^T,$$

where

$$\langle (\mathbf{a}^{(k)})^2 \mathbf{p}^{(m)} \rangle_{;n} = n \sum_{j=1}^{n} (a_j^{(k)})^2 p_j^{(m)}, \qquad \langle (\mathbf{a}^{(k)})^2 \mathbf{p}^{(m)} \mathbf{p}^{(l)} \rangle_{;n} = n \sum_{j=1}^{n} (a_j^{(k)})^2 p_j^{(m)} p_j^{(l)}.$$

Assume that the limits

$$\langle (\mathbf{a}^{(k)})^2 \mathbf{p}^{(m)} \mathbf{p}^{(l)} \rangle_{\infty} \overset{\mathrm{def}}{=} \lim_{n \to \infty} \langle (\mathbf{a}^{(k)})^2 \mathbf{p}^{(m)} \mathbf{p}^{(l)} \rangle_{;n} \tag{6}$$

exist. Then the limits

$$\langle (\mathbf{a}^{(k)})^2 \mathbf{p}^{(m)} \rangle_{\infty} \overset{\mathrm{def}}{=} \lim_{n \to \infty} \langle (\mathbf{a}^{(k)})^2 \mathbf{p}^{(m)} \rangle_{;n}$$

exist also, since due to (3) we have $\sum_{l=1}^{M} p_j^l = 1$ for all j.

So, under this assumption,

$$n \, \mathrm{Cov}[\bar{\xi}^{(k)}] \to \mathbf{\Sigma}_{\infty} \quad \text{as } n \to \infty, \tag{7}$$

where

$$\mathbf{\Sigma}_{\infty} \overset{\mathrm{def}}{=} \sum_{m=1}^{M} \langle (\mathbf{a}^{(k)})^2 \mathbf{p}^{(m)} \rangle_{\infty} \mathbf{\Sigma}^{(m)} - \sum_{m,l=1}^{M} \langle (\mathbf{a}^{(k)})^2 \mathbf{p}^{(m)} \mathbf{p}^{(l)} \rangle_{\infty} \mu^{(m)} (\mu^{(l)})^T.$$

Theorem 1. *Assume that:*

1. $\frac{1}{n} \mathbf{\Gamma}_{;n} \to \mathbf{\Gamma}_{\infty}$ *as $n \to \infty$ and $\det \mathbf{\Gamma}_{\infty} > 0$.*

2. *Assumption (6) holds.*

3. $\mathsf{E}^{(m)}[\|\xi\|^2] < \infty$ *for all $m = 1, \ldots, M$.*

Then

$$\sqrt{n}(\bar{\xi}_{;n}^{(k)} - \mu^{(k)}) \overset{\mathrm{W}}{\longrightarrow} N(0, \mathbf{\Sigma}_{\infty}^{(k)}).$$

This theorem is a simple corollary of Theorem 4.3 in [5].

3   Jackknife estimation of ACM of moment estimators

In what follows, we will consider unknown parameters of the component distribution $F_\xi^{(k)}$, which can be represented in the form

$$\vartheta = \vartheta^{(k)} = H(\mu^{(k)}), \tag{8}$$

where $H : \mathbb{R}^d \to \mathbb{R}^q$ is some known function. A natural estimator for such parameter by the sample $\xi_1, \ldots, \xi_n$ is

$$\hat{\vartheta} = \hat{\vartheta}_{;n}^{(k)} = H(\bar{\xi}_{;n}^{(k)}). \tag{9}$$

Then asymptotic behavior of this estimator is described by the following theorem.

Theorem 2. *In assumptions of Theorem 1, if H is continuously differentiable in some neighborhood of $\mu^{(k)}$, then*

$$\sqrt{n}(\hat{\vartheta}_{;n}^{(k)} - \vartheta^{(k)}) \xrightarrow{\text{W}} N(0, \mathbf{V}_\infty),$$

where

$$\mathbf{V}_\infty = \mathbf{V}_\infty^{(k)} = \mathbf{H}'(\mu^{(k)})\mathbf{\Sigma}_\infty^{(k)}(\mathbf{H}'(\mu^{(k)}))^T, \tag{10}$$

$$\mathbf{H}' = \begin{pmatrix} \frac{\partial H^1}{\partial \mu^1} & \cdots & \frac{\partial H^1}{\partial \mu^d} \\ \vdots & \ddots & \vdots \\ \frac{\partial H^q}{\partial \mu^1} & \cdots & \frac{\partial H^q}{\partial \mu^d} \end{pmatrix}.$$

This theorem is a simple implication from our Theorem 1 and Theorem 3 in Section 5, Chapter 1 of [1].

So, $\mathbf{V}_\infty$ defined by (10) is the ACM of the estimator $\hat{\vartheta}^{(k)}$ (the covariance matrix of the limit normal distribution of the normalized difference between the estimator and the estimated parameter). If it was known one could use it to construct tests for hypotheses on $\vartheta^{(k)}$ or to derive confidence set for $\vartheta^{(k)}$. In fact, for most estimators the ACM is unknown. Usually some estimate of $\mathbf{V}_\infty$ is used to replace its true value in statistical algorithms.

The jackknife is one of most general techniques of ACM estimation. Let $\hat{\vartheta}$ be any estimator of ϑ by the data $\xi_1, \ldots, \xi_n$:

$$\hat{\vartheta} = \hat{\vartheta}(\xi_1, \ldots, \xi_n).$$

Consider estimates of the same form which are calculated by all observations without one

$$\hat{\vartheta}_{i-} = \hat{\vartheta}(\xi_1, \ldots, \xi_{i-1}, \xi_{i+1}, \ldots, \xi_n).$$

Then the jackknife estimator for $\mathbf{V}_\infty$ is defined by

$$\hat{\mathbf{V}}_{;n} = \hat{\mathbf{V}}_{;n}^{(k)} = n \sum_{i=1}^{n} (\hat{\vartheta}_{i-} - \hat{\vartheta})(\hat{\vartheta}_{i-} - \hat{\vartheta})^T. \tag{11}$$

In our case $\hat{\vartheta} = H(\bar{\xi}^{(k)})$, so

$$\hat{\vartheta}_{i-} = H(\bar{\xi}_{i-}^{(k)}), \tag{12}$$

where

$$\bar{\xi}_{j-}^{(k)} = \sum_{j \neq i} a_{ji-}^{(k)} \xi_j. \tag{13}$$

Here $\mathbf{a}_{i-} = (a_{ji-}^{(m)}, j = 1, \dots, n, m = 1, \dots, M)$ is the minimax weights matrix calculated by the matrix $\mathbf{p}_{i-}$ of concentrations of all observations except i-th one. That is, $\mathbf{p}_{i-} = (p_1^{\bullet}, \dots, p_{i-1}^{\bullet}, 0, p_{i+1}^{\bullet}, \dots, p_n^{\bullet})^T$,

$$\mathbf{a}_{i-} = \mathbf{p}_{i-} \mathbf{\Gamma}_{i-}^{-1}, \tag{14}$$

$$\mathbf{\Gamma}_{i-} = \mathbf{p}_{i-}^T \mathbf{p}_{i-}. \tag{15}$$

Notice that 0 is placed at the i-th row of $\mathbf{p}_{i-}$ as a placeholder only, to preserve the numbering of the rows in $\mathbf{p}_{i-}$ and $\mathbf{a}_{i-}$, which corresponds to the numbering of subjects in the sample.

Theorem 3. *Let ϑ be defined by (8), $\hat{\vartheta}$ by (9), $\mathbf{V}_\infty$ by (10) and $\hat{V}_{;n}^{(k)}$ by (11)–(15). Assume that:*

1. *H is twice continuously differentiable in some neighborhood of $\mu^{(k)}$.*

2. *There exists some $\alpha > 4$, such that $\mathsf{E}[\|\xi(O)\|^\alpha \mid \kappa(O) = m] < \infty$ for all $m = 1, \dots, M$.*

3. *$\frac{1}{n}\mathbf{\Gamma}_{;n} \to \mathbf{\Gamma}_\infty$ as $n \to \infty$ and $\det \mathbf{\Gamma}_\infty > 0$.*

4. *Assumption (6) holds.*

Then $\hat{\mathbf{V}}_{;n}^{(k)} \to \mathbf{V}_\infty^{(k)}$.

For **proof** see Section 8.

4 Fast calculation algorithm for jackknife estimator

Direct calculation of $\hat{\mathbf{V}}_{;n}$ by (11)–(15) needs $\sim Cn^2$ elementary operations. Here we consider an algorithm which reduces the computational complexity to $\sim Cn$ operations (linear complexity).

Notice that $\mathbf{\Gamma}_{i-} = \mathbf{\Gamma} - \mathbf{p}_i^{\bullet}(\mathbf{p}_i^{\bullet})^T$. So

$$(\mathbf{\Gamma}_{i-})^{-1} = \mathbf{\Gamma}^{-1} + \frac{1}{1 - h_i}\mathbf{\Gamma}^{-1}\mathbf{p}_i^{\bullet}(\mathbf{p}_i^{\bullet})^T\mathbf{\Gamma}^{-1}, \tag{16}$$

where

$$h_i = (\mathbf{p}_i^{\bullet})^T \mathbf{\Gamma}^{-1}\mathbf{p}_i^{\bullet}. \tag{17}$$

(Formula (16) can be demonstrated directly by checking $\mathbf{\Gamma}_{i-}^{-1}\mathbf{\Gamma}_{i-} = \mathbb{E}$. It is also a corollary to the Serman–Morrison–Woodbury formula, see A.9.4 in [13].)

Let us denote

$$\bar{\boldsymbol{\xi}}_{i-} = \begin{pmatrix} \bar{\xi}_{i-}^{1(1)} & \cdots & \bar{\xi}_{i-}^{d(1)} \\ \vdots & \ddots & \vdots \\ \bar{\xi}_{i-}^{1(M)} & \cdots & \bar{\xi}_{i-}^{d(M)} \end{pmatrix} = ((\bar{\xi}_{i-}^{(1)})^T, \ldots, (\bar{\xi}_{i-}^{(M)})^T)^T$$

and

$$\bar{\boldsymbol{\xi}} = ((\bar{\xi}^{(1)})^T, \ldots, (\bar{\xi}^{(M)})^T)^T. \tag{18}$$

Then $\bar{\boldsymbol{\xi}}_{i-} = (\boldsymbol{\Gamma}_{i-})^{-1}\mathbf{p}_{i-}\boldsymbol{\Xi}_{i-}$, where $\boldsymbol{\Xi}_{i-} = (\xi_1, \ldots, \xi_{i-1}, 0, \xi_{i+1}, \ldots, \xi_n)^T$. (Zero at the i-th row is a placeholder as in the matrix $\mathbf{p}_{i-}$.) Applying (16) one obtains

$$\bar{\boldsymbol{\xi}}_{i-} = \boldsymbol{\Gamma}^{-1}(\mathbf{p}_{i-})^T \boldsymbol{\Xi}_{i-} + \frac{1}{1-h_i}\boldsymbol{\Gamma}^{-1}\mathbf{p}_i^{\boldsymbol{\cdot}}(\mathbf{p}_i^{\boldsymbol{\cdot}})^T \boldsymbol{\Gamma}^{-1}(\mathbf{p}_i^{\boldsymbol{\cdot}})^T \boldsymbol{\Xi}_{i-}.$$

This together with $(\mathbf{p}_{i-})^T \boldsymbol{\Xi}_{i-} = \mathbf{p}^T \boldsymbol{\Xi} - \mathbf{p}_i^{\boldsymbol{\cdot}}\xi_i^T$ implies

$$\bar{\boldsymbol{\xi}}_{i-} = \bar{\boldsymbol{\xi}} + \frac{1}{1-h_i}\mathbf{a}_i^{\boldsymbol{\cdot}}((\mathbf{p}_i^{\boldsymbol{\cdot}})^T\bar{\boldsymbol{\xi}} - \xi_i^T). \tag{19}$$

Then the following algorithm allows one to calculate $\hat{\mathbf{V}}^{(m)}$ for all $m = 1, \ldots, M$ at once by $\sim Cn$ operations.

Algorithm

1. Calculate $\boldsymbol{\Gamma}$ and $\mathbf{a}$ by (5).
2. Calculate h_i, $i = 1, \ldots, n$ by (17).
3. Calculate $\bar{\boldsymbol{\xi}}$ by (2) and (18).
4. Calculate $\bar{\boldsymbol{\xi}}_{i-}$ by (19).
5. Calculate $\hat{\vartheta}_{i-} = \hat{\vartheta}_{i-}^{(m)} = H(\bar{\xi}_{i-1}^{(m)})$ for $i = 1, \ldots, n, m = 1, \ldots, M$.
6. Calculate $\hat{\mathbf{V}}_{;n}^{(m)}$ by (11) for all $m = 1, \ldots, M$.

5 Regression with errors in variables

In this section we consider a mixture of simple linear regressions with errors in variables. A modification of orthogonal regression estimation technique was proposed for the regression coefficients estimation in [6]. We will show how the jackknife ACM estimators from Section 3 can be applied in this case to construct confidence sets for the regression coefficients.

Recall the errors-in-variables regression model in the context of mixture with varying concentrations.

We consider the case when each subject O has two variables of interest: $x(O)$ and $y(O)$. These variables are related by a strict linear dependence with coefficients depending on the component that O belongs to:

$$y(O) = b_0^{(\kappa(O))} + b_1^{(\kappa(O))}x(O), \tag{20}$$

where $b_0^{(m)}$, $b_1^{(m)}$ are the regression coefficients for the m-th component.

The true values of $x(O)$ and $y(O)$ are unobservable. These variables are observed with measurement errors

$$X(O) = x(O) + \varepsilon_X(O), \qquad Y(O) = y(O) + \varepsilon_Y(O). \tag{21}$$

Here we assume that the errors $\varepsilon_X(O)$ and $\varepsilon_Y(O)$ are conditionally independent given $\kappa(O) = m$,

$$\mathsf{E}^{(m)}\,\varepsilon_X = \mathsf{E}^{(m)}\,\varepsilon_Y = 0 \quad \text{and} \quad \text{Var}^{(m)}\,\varepsilon_X = \text{Var}^{(m)}\,\varepsilon_Y = \sigma_{(m)}^2 \tag{22}$$

for all $m = 1, \ldots, M$. So the distributions of $\varepsilon_X(O)$ and $\varepsilon_Y(O)$ can be different, but their variances are the same for a given subject. We assume that $\sigma_{(m)}^2 > 0$, $m = 1, \ldots, M$, and are unknown.

As in Section 2 we observe a sample $(X(O_j), Y(O_j))^T = (X_j, Y_j)^T$, $j = 1, \ldots, n$, from the mixture with known concentrations $p_j^{(m)} = \mathsf{P}\{\kappa(O_j) = m\}$.

In the case of homogeneous sample, when there is no mixture, the classical way to estimate b_0 and b_1 is orthogonal regression. That is, the estimator is taken as the minimizer of the total least squares functional which is the sum of squares of minimal Euclidean distances from the observation points to the regression line. The modification of this technique for mixtures with varying concentrations proposed in [6] leads to the following estimators for $b_0^{(k)}$ and $b_1^{(k)}$:

$$\hat{b}_1^{(k)} = \frac{\hat{S}_{XX}^{(k)} - \hat{S}_{YY}^{(k)} + \sqrt{(\hat{S}_{XX}^{(k)} - \hat{S}_{YY}^{(k)})^2 + 4(\hat{S}_{XY}^{(k)})^2}}{2\hat{S}_{XY}^{(k)}}, \tag{23}$$

$$\hat{b}_0^{(k)} = \bar{Y}^k - \hat{b}_1^{(k)}\bar{X}^{(k)},$$

where

$$\bar{X}^{(k)} = \sum_{j=1}^n a_j^{(k)} X_j, \qquad \bar{Y}^{(k)} = \sum_{j=1}^n a_j^{(k)} Y_j,$$

$$\hat{S}_{XX}^{(k)} = \sum_{j=1}^n a_j^{(k)}(X_j - \bar{X}^{(k)})^2, \qquad \hat{S}_{YY}^{(k)} = \sum_{j=1}^n a_j^{(k)}(Y_j - \bar{Y}^{(k)})^2,$$

$$\hat{S}_{XY}^{(k)} = \sum_{j=1}^n a_j^{(k)}(X_j - \bar{X}^{(k)})(Y_j - \bar{Y}^{(k)}).$$

Conditions of consistency and asymptotic normality of these estimators are given in Theorems 5.1 and 5.2 from [6]. For example, under the assumptions of Theorem 4 we obtain

$$\sqrt{n}(\hat{\vartheta}_{;n}^{(k)} - \vartheta^{(k)}) \to N(0, \mathbf{V}_\infty^{(k)}),$$

where $\vartheta^{(k)} = (b_0^{(k)}, b_1^{(k)})^T$.

The ACM $\mathbf{V}_\infty$ of the estimator is given by formula (21) in [6]. This formula is rather complicated and involves theoretical moments of unobservable variables $x(O)$, $\varepsilon_X(O)$ and $\varepsilon_Y(O)$. So it is natural to estimate $\mathbf{V}_\infty$ by the jackknife technique, which doesn't need to know or estimate these moments.

Notice that the estimator $(\hat{b}_0^{(k)}, \hat{b}_1^{(k)})^T$ can be represented in terms of Section 3 if we expand the space of observable variables including quadratic terms. That is, we consider the sample

$$\xi_j = (X_j, Y_j, (X_j)^2, (Y_j)^2, (X_j Y_j)^2)^T, \quad j = 1, \ldots, n.$$

Then the estimator $\hat{\vartheta}_{;n} = (\hat{\vartheta}_{;n}^1, \hat{\vartheta}_{;n}^2)^T = (\hat{b}_0^{(k)}, \hat{b}_1^{(k)})^T$ defined by (23) can be represented in form (9) with twice continuously differentiable function H if $\text{Var}^{(k)}[x] \neq 0$ and $b_1^{(k)} \neq 0$.

So we can apply the technique developed in Sections 3–4. Let us define the estimator $\hat{\mathbf{V}}_{;n}^{(k)}$ for $\mathbf{V}_\infty^{(k)}$ by (11).

Theorem 4. *Assume that the following conditions hold.*

1. *$\frac{1}{n}\boldsymbol{\Gamma}_{;n} \to \boldsymbol{\Gamma}_\infty$ as $n \to \infty$ and $\det \boldsymbol{\Gamma}_\infty > 0$.*

2. *Assumption (6) holds.*

3. *$\mathsf{E}^{(m)}[(x)^{12}] < \infty$, $\mathsf{E}^{(m)}[(\varepsilon_X)^{12}] < \infty$, $\mathsf{E}^{(m)}[(\varepsilon_Y)^{12}] < \infty$ for all $m = 1, \ldots, M$.*

4. *$\text{Var}^{(k)} x(0) \neq 0$ and $b_1^{(k)} \neq 0$.*

Then $\hat{\mathbf{V}}_{;n}^{(k)} \to \mathbf{V}_\infty^{(k)}$ in probability as $n \to \infty$.

This theorem is a simple combination of Theorem 3 and Theorem 5.2 from [6].

In what follows we assume that $\mathbf{V}_\infty^{(k)}$ is nonsingular. This assumption holds, e.g. if for all m the distributions of $x(O)$, $\varepsilon_X(O)$ and $\varepsilon_Y(O)$ given $\kappa(O) = m$ are absolutely continuous with continuous PDFs. (The proof of this fact is rather technical, so we do not present it here.)

We can construct a confidence set (ellipsoid) for the unknown parameter $\vartheta^{(k)}$ applying the Theorem 4 by the usual way. Namely, let for any $\mathbf{t} \in \mathbb{R}^2$

$$T_{;n}(\mathbf{t}) = (\mathbf{t} - \hat{\vartheta}_{;n}^{(k)})^T (\hat{\mathbf{V}}_{;n}^{(k)})^{-1} (\mathbf{t} - \hat{\vartheta}_{;n}^{(k)}).$$

Then in the assumptions of Theorem 4, if $\det \mathbf{V}_\infty^{(k)} \neq 0$,

$$T_{;n}(\vartheta^{(k)}) \xrightarrow{\text{W}} \eta, \quad \text{as } n \to \infty, \tag{24}$$

where η is a random variable (r.v.) with chi-square distribution with 2 degrees of freedom.

Consider $B_{\alpha;n} = \{\mathbf{t} \in \mathbb{R}^2 : T_{;n}(\mathbf{t}) \leq Q^\eta(1-\alpha)\}$, where $Q^\eta(\alpha)$ means the quantile of level α for the r.v. η. By (24)

$$\mathsf{P}\{\vartheta^{(k)} \in B_{\alpha;n}\} \to 1 - \alpha, \tag{25}$$

so $B_{\alpha;n}$ is an asymptotic confidence set for $\vartheta^{(k)}$ of level α.

6   Results of simulation

To assess performance of the proposed technique we performed a small simulation study. In the following three experiments we calculated covering frequencies of confidence sets for regression coefficients in the model (20)–(22) constructed by (25) and corresponding one-dimensional confidence intervals.

In all experiments for sample size $n = 100$ through 5000 we generated $B = 1000$ samples and calculated estimates for the parameters and corresponding confidence sets. The one-dimensional confidence intervals for $b_i^{(k)}$ were calculated by the standard formula

$$\left[\hat{b}_{i;n}^{(k)} - \lambda_{\alpha/2} \sqrt{\frac{\hat{v}_{ii;n}^{(k)}}{n}}, \, \hat{b}_{i;n}^{(k)} + \lambda_{\alpha/2} \sqrt{\frac{\hat{v}_{ii;n}^{(k)}}{n}} \right],$$

where $\hat{v}_{ii;n}^{(k)}$ is the i-th diagonal entry of the matrix $\hat{\mathbf{V}}_{;n}^{(k)}$, $\lambda_{\alpha/2}$ is the quantile of level $1 - \alpha/2$ for the standard normal distribution. The confidence level for the sets and intervals was taken $\alpha = 0.05$.

Then the numbers of cases when the confidence set covers the true value of the estimated parameter were calculated and divided by B. These are the covering frequencies reported in the tables below.

In all the experiments we considered two-component mixture ($M = 2$) with the concentrations of components

$$p_{j;n}^1 = j/n, \qquad p_{j;n}^2 = 1 - j/n.$$

The regression coefficients were taken as

$$b_0^{(1)} = 1/2, \qquad b_1^{(1)} = 2, \qquad b_0^{(2)} = -1/2, \qquad b_1^{(2)} = -1/3,$$

and the distribution of the true (unobservable) regressor $x(O)$ was $N(0, 2)$ for $\kappa(O) = 1$ and $N(1, 2)$ for $\kappa(O) = 2$.

Experiment 1. In this experiment we let ε_X and $\varepsilon_Y \sim N(0, 0.25)$. The variance of the errors is so small that the regression coefficients can be estimated with no difficulties even for small sample sizes.

The covering frequencies for confidence sets are presented in Table 1. It seems that they approach the nominal covering probability 0.95 with satisfactory accuracy for sample sizes $n \geq 1000$.

Table 1. Covering frequencies for confidence sets in Experiment 1

n	$b_0^{(1)}$	$b_1^{(1)}$	$(b_0^{(1)}, b_1^{(1)})$	$b_0^{(2)}$	$b_1^{(2)}$	$(b_0^{(2)}, b_1^{(2)})$
100	0.935	0.961	0.948	0.936	0.987	0.957
250	0.953	0.960	0.950	0.964	0.980	0.950
500	0.940	0.954	0.939	0.958	0.973	0.962
1000	0.946	0.949	0.943	0.954	0.971	0.935
2500	0.961	0.949	0.948	0.937	0.953	0.947
5000	0.947	0.949	0.948	0.954	0.956	0.958

Experiment 2. In this experiment we enlarged the variance of the error terms taking it as $\sigma^2 = 2$. All other parameters were the same as in Experiment 1. The results are presented in Table 2.

Table 2. Covering frequencies for confidence sets in Experiment 2

n	$b_0^{(1)}$	$b_1^{(1)}$	$(b_0^{(1)}, b_1^{(1)})$	$b_0^{(2)}$	$b_1^{(2)}$	$(b_0^{(2)}, b_1^{(2)})$
100	0.969	0.942	0.918	0.950	0.974	0.958
250	0.958	0.956	0.945	0.946	0.962	0.959
500	0.949	0.945	0.936	0.953	0.966	0.960
1000	0.959	0.946	0.954	0.947	0.958	0.942
2500	0.956	0.949	0.950	0.947	0.961	0.958
5000	0.953	0.941	0.952	0.955	0.955	0.968

It seems that the increase of errors dispersion doesn't deteriorate covering accuracy of the confidence sets.

Experiment 3. Here we consider the case when the errors distributions are heavy tailed. We generate the data with ε_X and ε_Y having Student-T distribution with df $= 14$ degrees of freedom. (This is the smallest df for which assumptions of Theorem 4 hold.) Covering frequencies are presented in Table 3.

Table 3. Covering frequencies for confidence sets in Experiment 3

n	$b_0^{(1)}$	$b_1^{(1)}$	$(b_0^{(1)}, b_1^{(1)})$	$b_0^{(2)}$	$b_1^{(2)}$	$(b_0^{(2)}, b_1^{(2)})$
100	0.935	0.961	0.948	0.936	0.987	0.957
250	0.953	0.960	0.950	0.964	0.980	0.950
500	0.940	0.954	0.939	0.958	0.973	0.962
1000	0.946	0.949	0.943	0.954	0.971	0.935
2500	0.961	0.949	0.948	0.937	0.953	0.947
5000	0.947	0.949	0.948	0.954	0.956	0.958

It seems that the accuracy of covering slightly decreased but this decrease is insignificant for practical purposes.

7 Sociologic analysis of EIT data

We would like to demonstrate advantages of the proposed technique by application to the analysis of the External Independent Testing (EIT) data (see [7]). EIT is a set of exams for high school graduates in Ukraine which must be passed for admission to universities. We use data on EIT-2016 from the official site of *Ukrainian Center for Educational Quality Assessment*.[3]

In this presentation we consider only the data on scores on two subjects: *Ukrainian language and literature* (Ukr) and on *Mathematics* (Math). The scores range from 100 to 200 points. (We have excluded the data on persons who failed on one of the exams

---

[3]https://zno.testportal.com.ua/stat/2016

or didn't pass these exams at all.) EIT-2016 contain such data on 246 thousands of examinees. The information on the region (Oblast) of Ukraine in which each examinee attended the high school is also available in EIT-2016.

Our aim is to investigate how dependence between Ukr and Math scores differs for examinees grown up in different environments There can be, e.g. an environment of adherents of Ukrainian culture and Ukrainian state, or in the environment of persons critical toward the Ukrainan independence. EIT-2016 doesn't contain information on such issues. So we use data on Ukrainian Parliament (Verhovna Rada) election results to deduce approximate proportions of adherents of different political choices in different regions of Ukraine.

We divided adherents of 29 parties and blocks that took part in the elections into three large groups, which are the components of our mixture:

(1) Pro-Ukrainian persons, voting for the parties that then created the ruling coalition (BPP, Batkivschyna, Narodny Front, Radicals and Samopomich)

(2) Contra-Ukrainian persons who voted for the Opposition block, voted against all or voted for small parties which where under 5% threshold on these elections.

(3) Neutral persons who did not took part in the voting.

Combining these data with EIT-2016 we obtain the sample (X_j, Y_j), $j = 1, \ldots, n$, where X_j is the Math score of the j-th examinee, Y_j is his/her Ukr score. The concentrations of components (p_j^1, p_j^2, p_j^3) are taken as frequencies of adherents of corresponding political choice at the region where the j-th examinee attended the high school.

In [7] the authors propose to use classical linear regression model (in which the error appears in the response only) to describe dependence between X_j and Y_j in these data. But the errors-in-variables model can be more adequate since the causes which deteriorate ideal functional dependence Ukr $= b_0 + b_1$ Math can affect both Math and Ukr scores causing random deviations of, maybe, the same dispersion for each variable.

So, in this presentation, we assumed that the data are described by the model (20)–(22), where $\kappa(O) = 1, 2, 3$ means the component (environment at which the person O was grown up) corresponding to one of three political choices given above. (Lower and upper endpoints of confidence intervals are given in columns named low and upp correspondingly.)

In this model we calculated the confidence intervals of the level $\alpha = 0.05/3 \approx 0.0167$ to derive the unilateral level $\alpha = 0.05$ in comparisons of three intervals derived for three different components. The results are presented in Table 4. We observe that the obtained intervals are rather narrow. They don't intersect for different

Table 4. Confidence sets for coefficients of regression between Math and Ukr

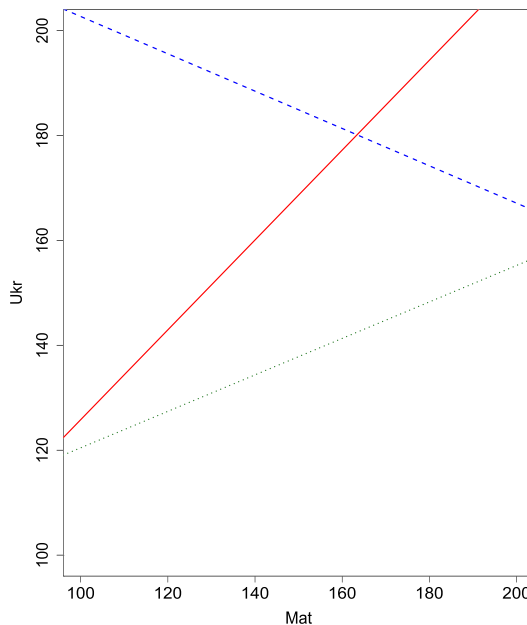|	Pro		Contra		Neutral	
	low	upp	low	upp	low	upp
$b_0^{(k)}$	40.12	40.22	236.3	240.1	84.21	87.19
$b_1^{(k)}$	0.8562	-0.366	-0.345	-2.80	0.335	0.359

Fig. 1. Estimated orthogonal regression lines for EIT-2016 data

components. So, the regression coefficients for different components are significantly different. (Of course, it is so only if our theoretical model of the data distribution is adequate.)

The orthogonal regression lines corresponding to different components are presented on Fig 1. The solid line corresponds to the Pro-component, the dashed line for the Contra-component and the dotted line for the Neutrals.

These results have simple and plausible explanation. Say, in the Pro-component the success in Ukr positively correlates with the general school successes, so with Math scores, too. It is natural for persons who are interested in Ukrainian culture and literature. In the Contra-component the correlation is negative. Why? The persons with high Math grades in this component do not feel the need to learn Ukrainian. But the persons with less success in Math try to improve their average score (by which the admission to universities is made) by increasing their Ukr score. The Neutral component shows positive correlation between Math and Ukr, but it is less then the correlation in the Pro-component.

Surely, these explanations are too simple to be absolutely correct. We consider them only as examples of hypotheses which can be deduced from the data by the proposed technique.

8 Proofs

To demonstrate Theorem 3 we need three lemmas. Below the symbols C and c mean finite constants, maybe different.

Lemma 1. *Assume that* $\det \Gamma_\infty > 0$. *Then:*

1. $\sup_{j=1,\ldots,n;m=1,\ldots,M} |a_{j;n}^m| = O(n^{-1})$.

2. $\sup_{i,j=1,\ldots,n;i\neq j;m=1,\ldots,M} |a_{j;n}^m - a_{ji-;n}^m| = O(n^{-2})$.

Proof. By definition, $\frac{1}{n}\Gamma_{;n} \to \Gamma_\infty$, so there exists $c > 0$ such that $\det \Gamma_{;n} > c$ for all n large enough. This together with $|p_{j;n}^m| < 1$ imply

$$\|\Gamma_{;n}^{-1}\| \leq \frac{C}{n}. \tag{26}$$

(Here $\| \cdot \|$ means the operator norm.) Taking into account that $\mathbf{a}_{;n} = \mathbf{p}_{;n}\Gamma_{;n}^{-1}$, we obtain the first statement of the lemma.

Then by (16)–(17),

$$\mathbf{a}_j^\cdot - \mathbf{a}_{ji-}^\cdot = \Gamma^{-1}\mathbf{p}_j^\cdot - \Gamma_{i-}^{-1}\mathbf{p}_j^\cdot = \frac{1}{h_i}\Gamma^{-1}\mathbf{p}_i^\cdot(\mathbf{p}_i^\cdot)^T\Gamma^{-1}\mathbf{p}_j^\cdot.$$

This together with (26) yields the second statement. $\qquad\square$

Lemma 2. *Assume that for* $m = 1, \ldots, M$:

1. $\mathsf{E}^{(m)}[\xi] = 0$;

2. *For some* $\delta > 0$, $\mathsf{E}^{(m)}(\xi)^2|\log|\xi||^{(1+\delta)} < \infty$;

3. $\det \Gamma_\infty > 0$.

Then for some $C < \infty$,

$$\mathsf{P}\left\{|\bar{\xi}^{(k)}| > C\sqrt{\frac{\log\log n}{n}}\right\} \to 0 \quad \text{as } n \to \infty.$$

Proof. Let $\eta_1, \ldots, \eta_n$ be independent random variables with $\mathsf{E}\,\eta_i = 0$. Let us denote $B_n = \sum_{j=1}^n \mathsf{E}(\eta_j)^2$. Then the last formula in the proof of Theorem 7.2 and Theorem 7.3 in [10] imply the following proposition.

Proposition 1. *If*

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \sum_{j=1}^n \mathsf{E}\,\eta_j^2|\log|\eta_j||^{1+\delta} < \infty,$$

then for any b *such that* $0 < b < \sqrt{1+\delta}$,

$$\mathsf{P}\left\{\sum_{j=1}^n \eta_j \geq b\sqrt{2B_n \log\log B_n}\right\} \leq (\log B_n)^{-b^2}$$

for n large enough.

Let $\eta_j = \pm n a_{j;n}^{(k)} \xi_j^l$. Then $B_n = n^2 \sum_{j=1}^{n} (a_{j;n}^{(k)})^2 \operatorname{Var} \xi_j^l \sim Cn$ by Lemma 1. Assumption 2 implies that the assumption of Proposition 1 holds. So,

$$\mathsf{P}\left\{ \sum_{j=1}^{n} n a_{j;n}^{(k)} \xi_j^l > b\sqrt{2Cn \log \log Cn} \right\} \to 0.$$

This implies the statement of the lemma. □

Lemma 3. *Assume that for some $\alpha > 0$*

$$\mathsf{E}^{(m)}[|\xi|^\alpha] < \infty \quad \text{for all } m = 1, \dots, M.$$

Then for any $\beta > 1/\alpha$ there exists $C < \infty$ such that

$$\mathsf{P}\{ \sup_{j=1,\dots,n} |\xi_j| > Cn^\beta \} \to 0 \quad \text{as } n \to \infty.$$

Proof. By the Chebyshov inequality we obtain that for some $0 < R < \infty$,

$$\mathsf{P}\{|\xi_j| > x\} \le \frac{R}{x^\alpha}.$$

Then for $\alpha\beta > 1$,

$$\mathsf{P}\{ \sup_{j=1,\dots,n} |\xi_j| > cn^\beta \} = 1 - \mathsf{P}\{ \sup_{j=1,\dots,n} |\xi_j| \le cn^\beta \}$$

$$= 1 - \prod_{j=1}^{n} \mathsf{P}\{|\xi_j| \le Cn^\beta\} = 1 - \prod_{j=1}^{n} (1 - \mathsf{P}\{|\xi_j| > Cn^\beta\})$$

$$\le 1 - \left(1 - \frac{R}{C^\alpha n^{\alpha\beta}}\right)^n = 1 - \exp\left(n \log\left(1 - \frac{R}{C^\alpha n^{\alpha\beta}}\right)\right)$$

$$\sim 1 - \exp\left(\frac{-nR}{C^\alpha n^{\alpha\beta}}\right) \to 0 \quad \text{as } n \to \infty,$$

if $\alpha\beta > 1$.

Lemma is proved. □

Proof of Theorem 3. Let $\xi_j' = \xi_j - \mathsf{E}\,\xi_j$. Then

$$\bar{\xi}^{(k)} = \sum_{j=1}^{n} a_j^{(k)} \xi_j = \sum_{j=1}^{n} a_j^{(k)} \xi_j' + \sum_{j=1}^{n} \sum_{m=1}^{M} a_j^{(k)} p_j^{(m)} \mu^{(m)} = \sum_{j=1}^{n} a_j^{(k)} \xi_j' + \mu^{(k)},$$

due to (3). Similarly,

$$\bar{\xi}_{i-}^{(k)} = \sum_{j \ne i} a_{ji-}^{(k)} \xi_j' + \mu^{(k)}.$$

Let us denote $\mathbf{U}_i = (U_1, \dots, U_d)^T = \bar{\xi}^{(k)} - \bar{\xi}_{i-}^{(k)}$. Then

$$\mathbf{U}_i = \sum_{j=1}^{n} a_j^{(k)} \xi_j' - \sum_{j \ne i} a_{ji-}^{(k)} \xi_j' = a_i^{(k)} \xi_i' + \sum_{j \ne i} (a_j^{(k)} - a_{ji-}^{(k)}) \xi_j' \qquad (27)$$

and
$$\hat{\vartheta} - \hat{\vartheta}_{i-} = H(\bar{\xi}^{(k)}) - H(\bar{\xi}_{i-}^{(k)}) = \mathbf{H}'(\zeta_i)\mathbf{U}_i,$$

where ζ_i is some intermediate point between $\bar{\xi}^{(k)}$ and $\bar{\xi}_{i-}^{(k)}$. So,

$$\hat{\mathbf{V}}_{;n} = n \sum_{i=1}^{n} \mathbf{H}'(\zeta_i)\mathbf{U}_i\mathbf{U}_i^T (\mathbf{H}'(\zeta_i))^T. \tag{28}$$

Let us denote

$$\tilde{\mathbf{V}}_{;n} = n \sum_{i=1}^{n} \mathbf{H}'(\mu^{(k)})\mathbf{U}_i\mathbf{U}_i^T (\mathbf{H}'(\mu^{(k)}))^T. \tag{29}$$

We will show that

$$\tilde{\mathbf{V}}_{;n} \to \mathbf{V}_{\infty} \quad \text{as } n \to \infty \text{ in probability} \tag{30}$$

and

$$n\|\hat{\mathbf{V}}_{;n} - \tilde{\mathbf{V}}_{;n}\| \to 0 \quad \text{as } n \to \infty \text{ in probability.} \tag{31}$$

These two equations imply the statement of the theorem.

We start from (30). Let us calculate $\mathsf{E}\,\tilde{\mathbf{V}}_{;n}$. Notice that

$$\mathsf{E}\,\mathbf{U}_i\mathbf{U}_i^T = (a_i^{(k)})^2 \,\mathsf{E}\,\xi_i'(\xi_i')^T + \sum_{j \neq i}(a_j^{(k)} - a_{ji-}^{(k)})^2 \,\mathsf{E}\,\xi_i'(\xi_i')^T.$$

By Assumption 2 of the theorem, $\sup_i \|\,\mathsf{E}\,\xi_i'(\xi_i')^T\| < C$, and by Lemma 1,

$$\sup_{j=1,\dots,n} (a_j^{(k)} - a_{ji-}^{(k)})^2 = O(n^{-4}).$$

So,

$$\mathsf{E}\,\tilde{\mathbf{V}}_{;n} = n\mathbf{H}'(\mu^{(k)}) \sum_{i=1}^{n}(a_i^{(k)})^2 \,\mathsf{E}\,\xi_i'(\xi_i')^T (\mathbf{H}'(\mu^{(k)}))^T + O(n^{-1}).$$

By the same way as in (9), we obtain

$$\mathsf{E}\,\tilde{\mathbf{V}}_{;n} \to \mathbf{H}'(\mu^{(k)})\boldsymbol{\Sigma}_{\infty}^{(k)}(\mathbf{H}'(\mu^{(k)}))^T = \mathbf{V}_{\infty}. \tag{32}$$

Now, let us estimate

$$\mathsf{E}\,\|\tilde{\mathbf{V}}_{;n} - \mathsf{E}\,\tilde{\mathbf{V}}_{;n}\|^2 \leq Cn^2 \sum_{l_1,l_2=1}^{d} \sum_{i=1}^{n} \mathsf{E}(U_i^{l_1}U_i^{l_2} - \mathsf{E}\,U_i^{l_1}U_i^{l_2})^2$$

$$\leq Cn^2 \sum_{l=1}^{n} \sum_{i=1}^{n} \mathsf{E}(U_i^l)^4. \tag{33}$$

Notice that

$$\mathsf{E}(U_i^l)^4 = \mathsf{E}\left(a_i^{(k)}\xi_i^{l'} + \sum_{j \neq i}(a_i^{(k)} - a_{ji-}^{(k)})\xi_j^{l'}\right)^4$$

$$= (a_i^{(k)})^4 \, \mathsf{E}(\xi_j^{l'})^4 + 6(a_i^{(k)})^2 \, \mathsf{E}\left(\sum_{j\neq i}(a_i^{(k)} - a_{ji-}^{(k)})\xi_j^{l'}\right)^2$$

$$+ \mathsf{E}\left(\sum_{j\neq i}(a_i^{(k)} - a_{ji-}^{(k)})\xi_j^{l'}\right)^4 = O(n^{-4})$$

due to Lemma 1 and Assumption 2 of the theorem. So by (33) we obtain

$$\mathsf{E}\,\|\tilde{\mathbf{V}}_{;n} - \mathsf{E}\,\tilde{\mathbf{V}}_{;n}\|^2 = O(n^{-1}).$$

This and (32) imply (30).

Let us show (31). Notice that

$$n(\hat{\mathbf{V}}_{;n} - \tilde{\mathbf{V}}_{;n}) = n\sum_{i=1}^{n}(\mathbf{H}'(\zeta_i) - \mathbf{H}'(\mu^{(k)}))\mathbf{U}_i(\mathbf{U}_i)^T(\mathbf{H}'(\zeta_i))^T$$

$$+ n\sum_{i=1}^{n}(\mathbf{H}'(\mu^{(k)}))\mathbf{U}_i(\mathbf{U}_i)^T(\mathbf{H}'(\zeta_i) - \mathbf{H}'(\mu^{(k)}))^T. \tag{34}$$

By Lemma 1, $\sup_i \|\mathbf{U}_i(\mathbf{U}_i)^T\| \leq Cn^{-2}\sup_i \|\xi_i'\|^2$. By Lemma 3 and Assumption 2 of the theorem,

$$\sup_i \|\xi_i'\|^2 = O_P(n^\beta) \quad \text{for } \beta > \frac{2}{\alpha}.$$

Since $\alpha > 2$ we may take here $\beta < 1/2$. Let us estimate $\sup_i \|\mathbf{H}'(\zeta_i) - \mathbf{H}'(\mu^{(k)})\|$. Notice that ζ_i is an intermediate point between $\bar{\zeta}^{(k)}$ and $\bar{\zeta}_{i-}^{(k)}$. By Lemma 2,

$$\|\bar{\xi}^{(k)} - \mu^{(k)}\| = O_P\left(\sqrt{\frac{\log\log n}{n}}\right).$$

Then, by Lemma 1,

$$\sup_i \|\bar{\xi}^{(k)} - \bar{\xi}_{i-}^{(k)}\| \leq Cn^{-1}\sup_i |\xi_i'| = O_P(n^{-1+\beta})$$

and

$$\sup_i \|\zeta_i - \mu^{(k)}\| = O_P\left(\sqrt{\frac{\log\log n}{n}}\right).$$

Due to Assumption 1 of the theorem this implies

$$\sup_i \|\mathbf{H}'(\zeta_i) - \mathbf{H}'(\mu^{(k)})\| = O_P\left(\sqrt{\frac{\log\log n}{n}}\right)$$

and $\sup_i \|\mathbf{H}'(\zeta_i)\| = O_P(1)$.

Combining these estimates with (34), we obtain

$$n(\hat{\mathbf{V}}_{;n} - \tilde{\mathbf{V}}_{;n}) = n \sum_{i=1}^{n} O_P \left(\sqrt{\frac{\log \log n}{n}} \right) \frac{C}{n^2} n^\beta O_P(1) = O_P(1),$$

since $\beta < 1/2$. This is (31).

Combining (31) and (30), we obtain the statement of theorem.   □

9 Conclusions

We introduced a modification of the jackknife technique for the ACM estimation for moment estimators by observations from mixtures with varying concentrations. A fast algorithm is proposed which implements this technique. Consistency of derived estimator is demonstrated. Results of simulations demonstrate its practical applicability for sample sizes $n > 1000$.

References

[1] Borovkov, A.A.: Mathematical Statistics. Gordon and Breach Science Publishers, Amsterdam (1998). MR1712750

[2] Branham, R.: Total Least Squares in Astronomy. In: Total Least Squares and Errors-in-Variables Modeling, pp. 375–384. Springer, Dordrecht (2002). MR1952962. https://doi.org/10.1007/978-94-017-3552-0_33

[3] Cheng, C.-L., Van Ness, J.: Statistical Regression with Measurement Error, Kendall's Library of Statistics 6. Arnold, London (1999). MR1719513

[4] Maiboroda, R.: Statistical analysis of mixtures. Kyiv University Publishers, Kyiv (in Ukrainian) (2003)

[5] Maiboroda, R., Sugakova, O.: Statistics of mixtures with varying concentrations with application to DNA microarray data analysis. J. Nonparametr. Stat. **24**(1), 201–215 (2012). MR2885834. https://doi.org/10.1080/10485252.2011.630076

[6] Maiboroda, R., Navara, H., Sugakova, O.: Orthogonal regression for observations from mixtures. Teor. Imovir. Mat. Stat. **99**, 152–167 (2018)

[7] Miroshnichenko V, M.R.: Confidence ellipsoids for regression coefficients by observations from a mixture. Mod. Stoch. Theory Appl. **5**(2), 225–245 (2018). MR3813093. https://doi.org/10.15559/18-vmsta105

[8] Masiuk, S., Kukush, A., Shklyar, S., Chepurny, M., Likhtarov, I. (eds.): Radiation Risk Estimation: Based on Measurement Error Models, 2nd edn. de Gruyter series in Mathematics and Life Sciences, vol. 5. de Gruyter (2017). MR3726857

[9] McLachlan, G.J., Lee, S.X., Rathnayake, S.I.: Finite Mixture Models. Ann. Rev. Stat. Appl. **6**, 355–378 (2019). MR3939525. https://doi.org/10.1146/annurev-statistics-031017-100325

[10] Petrov, V.: Limit theorems of probability theory: sequences of independent random variables. Clarendon Press

[11] Shao, J.: Mathematical statistics. Springer, New York (2007). MR2002723. https://doi.org/10.1007/b97553

[12] Shao, J., Tu, D.: The Jackknife and Bootstrap. Springer (2012). MR1351010. https://doi.org/10.1007/978-1-4612-0795-5

[13] Seber, G., Lee, A.: Linear regression analysis. Wiley (2003). MR1958247. https://doi.org/10.1002/9780471722199

[14] Therneau Terry, M., Grambsch Patricia, M.: Modeling Survival Data Extending the Cox Model. Springer (2000). MR1774977. https://doi.org/10.1007/978-1-4757-3294-8

[15] Wolter, K.: Introduction To Variance Estimation. Springer (2007). MR2288542