

Jackknife for nonlinear estimating equations

Rostyslav Maiboroda, Vitalii Miroshnychenko, Olena Sugakova

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

rostmaiboroda@knu.ua (R. Maiboroda), vitaliy.miroshnychenko@gmail.com (V. Miroshnychenko),
sugak@univ.kiev.ua (O. Sugakova)

Received: 23 January 2022, Revised: 26 May 2022, Accepted: 26 May 2022,
Published online: 14 June 2022

Abstract In mixture with varying concentrations model (MVC) one deals with a nonhomogeneous sample which consists of subjects belonging to a fixed number of different populations (mixture components). The population which a subject belongs to is unknown, but the probabilities to belong to a given component are known and vary from observation to observation. The distribution of subjects' observed features depends on the component which it belongs to.

Generalized estimating equations (GEE) for Euclidean parameters in MVC models are considered. Under suitable assumptions the obtained estimators are asymptotically normal. A jackknife (JK) technique for the estimation of their asymptotic covariance matrices is described. Consistency of JK-estimators is demonstrated. An application to a model of mixture of nonlinear regressions and a real life example are presented.

Keywords Finite mixture model, nonlinear regression, mixture with varying concentrations, generalized estimating equations, jackknife, confidence ellipsoid

MSC2020 MSC 62F40, 62G10

1 Introduction

This paper continues studies of the jackknife (JK) technique application for statistical inference based on the model of mixture with varying concentrations (MVC). JK is a powerful tool of asymptotic covariance estimation of asymptotically normal statistics introduced by Quenouille (1949) and Tukey (1958). On its applications for homogeneous samples, see [11] and [1]. The JK technique was applied to heteroscedastic nonlinear regression models in [9]. Applications to errors-in-variables models are considered in [12, 13].

© 2022 The Author(s). Published by VTeX. Open access article under the [CC BY](https://creativecommons.org/licenses/by/4.0/) license.

In MVC models one deals with a nonhomogeneous sample which consists of subjects belonging to M different subpopulations (mixture components). One knows the probabilities with which a subject belongs to the mixture components and these probabilities are different for different subjects. So the considered observations are independent but not identically distributed. Modification of JK to such data analysis is a challenging problem.

On parametric inference in regression MVC models, see [2]. Estimation in non-parametric MVC models is discussed in [3]. In [5] a jackknife application to MVC of linear regressions models with errors in variables is considered. It is shown that the JK-estimators are consistent and allow to construct asymptotic confidence intervals for regression coefficients based on orthogonal regression estimators. In [7] a general result on asymptotic normality of generalized estimating equation (GEE) estimators for MVC is obtained which is applied to derive asymptotic normality of a modification of least squares (LS) estimators for MVC of nonlinear regressions models. A JK estimator for the asymptotic covariance was introduced in [7] also, but its properties were not investigated analytically.

In this paper we consider JK estimation of asymptotic covariance of GEE estimators in MVC models and show its consistency. The MVC model and the GEE estimator are discussed in Section 2. A version of JK for MVC is described in Section 3. Main results on consistency, asymptotic normality of GEE estimator and consistency of JK estimator of asymptotic covariance are presented in Section 4. Here we also consider an application to some nonlinear regression model. In Section 5 the developed statistical techniques are applied to a real life sociological data. Conclusive remarks are placed in Section 6. Section 7 contains technical proofs.

2 MVC model and GEE estimation

In the MVC model we assume that each observed subject O belongs to one of M different mixture components (subpopulations) $\mathcal{P}_k, k = 1, \dots, M$. The sample contains n subjects O_1, \dots, O_n . Let $\kappa_j = k$ iff $O_j \in \mathcal{P}_k$. The true κ_j are unknown, but one knows the mixing probabilities

$$p_{j;n}^m = \mathbf{P}\{\kappa_j = k\}.$$

These probabilities are also called the concentrations of the k -th component at j -th observation.

The D -dimensional vector of observed variables of O will be denoted by $\xi(O) = (\xi^1(O), \dots, \xi^D(O))^T \in \mathbb{R}^D$, $\xi_j = \xi_{j;n} = \xi(O_j)$.

Let $F^{(k)}$ be the distribution of $\xi(O)$ for $O \in \mathcal{P}_k$, i.e.

$$F^{(k)}(A) = \mathbf{P}\{\xi(O) \in A \mid O \in \mathcal{P}_k\}$$

for all Borel sets $A \subseteq \mathbb{R}^D$. Then

$$\mathbf{P}\{\xi_j \in A\} = \sum_{k=1}^M p_{j;n}^k F^{(k)}(A). \quad (1)$$

So, in the MVC model one observes independent ξ_j , $j = 1, \dots, n$, with the distribution defined by (1). In this paper we adopt a semiparametric model of components' distributions

$$F^{(k)}(A) = F(A, \boldsymbol{\vartheta}^{(k)}, \boldsymbol{\nu}^{(k)}), \quad k = 1, \dots, M, \quad (2)$$

where F is some known function of its arguments, $\boldsymbol{\vartheta}^{(k)} \in \Theta \subseteq \mathbb{R}^d$ are unknown Euclidean parameters of interest, $\boldsymbol{\nu}^{(k)}$ are some nonparametric nuisance parameters.

In what follows we will denote by $\xi_{(k)}$ a random vector with distribution $F^{(k)}$ which can be considered as the value of $\xi(O)$ for a subject O selected at random from the component \mathcal{P}_k .

Example. Consider the model of mixture of regressions from [7]. In this model the observations are $\xi_j = (Y_j, X_j^1, \dots, X_j^m)^T$, where Y_j is the response and $\mathbf{X}_j = (X_j^1, \dots, X_j^m)^T$ is the vector of regressors in the regression model

$$Y_j = g(\mathbf{X}_j; \boldsymbol{\vartheta}^{(\kappa_j)}) + \varepsilon_j, \quad (3)$$

where g is a known regression function, $\boldsymbol{\vartheta}^{(k)}$ is a vector of unknown regression coefficients in the k -th mixture component, ε_j are regression error terms. (In [7] somewhat more general model is considered, in which the regression functions and parameter spaces can be different for different components. In this presentation we restrict ourselves to simplify notation. The main result on JK-consistency can be extended to the general case considered in [7]).

We assume that ε_j are independent for different j and, for each j , ε_j and \mathbf{X}_j are conditionally independent given κ_j . Let $F_X^{(k)}$ and $F_\varepsilon^{(k)}$ be the conditional distributions of \mathbf{X}_j and ε_j given $O_j \in \mathcal{P}_k$. We assume that $\mathbb{E}[\varepsilon_j \mid \kappa_j = k] = \int x F_X^{(k)}(dx) = 0$ for all $k = 1, \dots, M$.

Model (3) is a partial case of model (1)–(2) in which the nuisance parameters are the distributions of regressors and errors for all components, i.e., $\boldsymbol{\nu}^{(k)} = (F_X^{(k)}, F_\varepsilon^{(k)})$.

To estimate $\boldsymbol{\vartheta}^{(k)}$ in (1)–(2) we apply the technique of generalized estimating equations (GEE) considered in [7]. (On GEE estimation technique and its relations to least squares, maximum likelihood and M-estimators in context of i.i.d. observations, see Section 5.4 in [10].) Let us choose an elementary estimating function $\mathbf{s} : \mathbb{R}^D \times \Theta \rightarrow \mathbb{R}^d$ such that

$$\mathbb{E} \mathbf{s}(\xi_{(k)}, \boldsymbol{\gamma}) = 0 \text{ iff } \boldsymbol{\gamma} = \boldsymbol{\vartheta}^{(k)}. \quad (4)$$

So, considering (4) as an equation of $\boldsymbol{\gamma} = (\gamma^1, \dots, \gamma^d)^T \in \Theta \subseteq \mathbb{R}^d$, we observe that its unique solution is $\boldsymbol{\vartheta}^{(k)}$. To obtain an estimator for $\boldsymbol{\vartheta}^{(k)}$ we replace $\mathbb{E} \mathbf{s}(\xi_{(k)}, \boldsymbol{\gamma})$ by its estimator

$$\mathbf{S}_n^{(k)}(\boldsymbol{\gamma}) = \sum_{j=1}^n a_{j;n}^k \mathbf{s}(\xi_{j;n}; \boldsymbol{\gamma}), \quad (5)$$

where $a_{j;n}^k$ are nonrandom weights satisfying the assumption

$$\sum_{j=1}^n a_{j;n}^k p_{j;n}^m = \begin{cases} 1 & \text{if } k = m, \\ 0 & \text{if } k \neq m, \end{cases} \quad \text{for all } m = 1, \dots, M. \quad (6)$$

Observe that $\mathbf{S}_n^{(k)}(\boldsymbol{\gamma})$ is an unbiased estimator for $\mathbf{E} \mathbf{s}(\boldsymbol{\xi}_{(k)}, \boldsymbol{\gamma})$ under (6), i.e., $\mathbf{E} \mathbf{S}_n^{(k)}(\boldsymbol{\gamma}) = \mathbf{E} \mathbf{s}(\boldsymbol{\xi}_{(k)}, \boldsymbol{\gamma})$. The GEE estimator to $\boldsymbol{\vartheta}^{(k)}$ is any statistics $\hat{\boldsymbol{\vartheta}}_n^{(k)}$ such that

$$\mathbf{S}_n^{(k)}(\hat{\boldsymbol{\vartheta}}_n^{(k)}) = 0, \text{ a.s.} \quad (7)$$

E.g., in the model (3) differentiation of the least squares functional yields the elementary estimating function

$$\mathbf{s}(\boldsymbol{\gamma}, Y, \mathbf{X}) = (Y - g(\mathbf{X}, \boldsymbol{\gamma})) \dot{\mathbf{g}}(\mathbf{X}, \boldsymbol{\gamma}), \quad (8)$$

where

$$\dot{\mathbf{g}}(\mathbf{X}, \boldsymbol{\gamma}) = \left(\frac{\partial g(\mathbf{X}, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^1}, \dots, \frac{\partial g(\mathbf{X}, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^d} \right)^T.$$

In this paper we consider only the minimax weights $a_{j;n}^k$, which can be defined as follows. Let $\mathbf{p}_{;n}$ be the matrix of all concentrations for all components of the mixture:

$$\mathbf{p}_{;n} = \begin{pmatrix} p_{1;n}^1 & \cdots & p_{1;n}^M \\ \vdots & \ddots & \vdots \\ p_{n;n}^1 & \cdots & p_{n;n}^M \end{pmatrix}.$$

Then the matrix of all weights $\mathbf{a}_{;n} = (a_{j;n}^m)_{j=1,\dots,n,m=1,\dots,M}$ is defined as

$$\mathbf{a}_{;n} = \mathbf{p}_{;n} \boldsymbol{\Gamma}_{;n}^{-1}, \quad (9)$$

where $\boldsymbol{\Gamma}_{;n} = \mathbf{p}_{;n}^T \mathbf{p}_{;n}$. (We assume that $\det \boldsymbol{\Gamma}_{;n} \neq 0$). Minimax properties of $\mathbf{a}_{;n}$ were discussed in [3]. For one alternative approach to weighting in GEE for MVC, see [4].

3 Jackknife for MVC

Consider the set of parameters $\boldsymbol{\vartheta}^{(k)}$, $k = 1, \dots, M$, for different components as one long vector parameter $\boldsymbol{\vartheta} = ((\boldsymbol{\vartheta}^{(1)})^T, \dots, (\boldsymbol{\vartheta}^{(M)})^T)^T$ and similarly for the set of estimators $\hat{\boldsymbol{\vartheta}}_n = ((\hat{\boldsymbol{\vartheta}}_n^{(1)})^T, \dots, (\hat{\boldsymbol{\vartheta}}_n^{(M)})^T)^T$. (Recall that $\boldsymbol{\vartheta}^{(k)}, \hat{\boldsymbol{\vartheta}}_n^{(k)} \in \mathbb{R}^d$.) It was shown in [7] that under suitable assumptions (see Theorem 3 below) the estimator $\hat{\boldsymbol{\vartheta}}_{;n}$ is asymptotically normal, i.e.

$$\sqrt{n}(\hat{\boldsymbol{\vartheta}}_{;n} - \boldsymbol{\vartheta}) \xrightarrow{W} N(0, \mathbf{V}).$$

To apply asymptotic normality for hypotheses testing one needs an estimator for the dispersion matrix (asymptotic covariance) \mathbf{V} given by (16). Jackknife (JK) is a powerful tool for constructing such estimators. We now consider its modification for the

MVC models (cf. [5]). Let

$$\mathbf{p}_{;-i,n} = \begin{pmatrix} p_{1;n}^1 & \cdots & p_{1;n}^M \\ \vdots & \ddots & \vdots \\ p_{i-1;n}^1 & \cdots & p_{i-1;n}^M \\ 0 & \cdots & 0 \\ p_{i+1;n}^1 & \cdots & p_{i+1;n}^M \\ \vdots & \ddots & \vdots \\ p_{n;n}^1 & \cdots & p_{n;n}^M \end{pmatrix},$$

i.e. $\mathbf{p}_{;-i,n}$ is the matrix $\mathbf{p}_{;n}$ with the i -th row replaced by the zero row. Then $\mathbf{\Gamma}_{;-i,n} = \mathbf{p}_{;-i,n}^T \mathbf{p}_{;-i,n}$ and

$$\mathbf{a}_{;-i,n} = \mathbf{p}_{;-i,n} \mathbf{\Gamma}_{;-i,n}^{-1}. \quad (10)$$

Let

$$\mathbf{S}_{-in}^{(k)}(\boldsymbol{\gamma}) = \sum_{j \neq i} a_{j;-i,n}^k \mathbf{s}(\boldsymbol{\xi}_{j;n}; \boldsymbol{\gamma}) \quad (11)$$

and define $\hat{\boldsymbol{\vartheta}}_{-in}^{(k)}$ as a statistics which satisfy

$$\mathbf{S}_{-in}^{(k)}(\hat{\boldsymbol{\vartheta}}_{-in}^{(k)}) = 0, \text{ a.s.}, \quad (12)$$

$\hat{\boldsymbol{\vartheta}}_{-in} = ((\hat{\boldsymbol{\vartheta}}_{-in}^{(1)})^T, \dots, (\hat{\boldsymbol{\vartheta}}_{-in}^{(M)})^T)^T$. In fact, $\hat{\boldsymbol{\vartheta}}_{-in}$ is the GEE estimator for $\boldsymbol{\vartheta}$ calculated by the sample which contains all the observed subjects O_j , except the i -th one. Then the JK estimator for \mathbf{V} is

$$\hat{\mathbf{V}}_n = n \sum_{i=1}^n (\hat{\boldsymbol{\vartheta}}_{-in} - \hat{\boldsymbol{\vartheta}}_n)(\hat{\boldsymbol{\vartheta}}_{-in} - \hat{\boldsymbol{\vartheta}}_n)^T. \quad (13)$$

(On some efficient algorithms for calculation of $\mathbf{a}_{;-i,n}$ and $\hat{\mathbf{V}}_n$ see [5]).

4 Main theorems

In this section we consider asymptotic behavior of $\hat{\boldsymbol{\vartheta}}_n$ and $\hat{\mathbf{V}}_n$ as $n \rightarrow \infty$. Note that we do not assume any relationship between the samples $\{\boldsymbol{\xi}_{j;n}, j = 1, \dots, n\}$ for different n . They can be independent or dependent, or a smaller sample can be a part of a larger one. The concentration arrays $\mathbf{p}_{;n}$ are also unrelated for different n .

To formulate the theorems we need some notations and assumptions. In what follows we assume that the limit

$$\mathbf{\Gamma}_\infty = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{p}_{;n}^T \mathbf{p}_{;n} \quad (14)$$

exists and $\det \mathbf{\Gamma}_\infty \neq 0$.

For a vector \mathbf{x} , the symbol $|\mathbf{x}|$ means the Euclidean norm. For a matrix \mathbf{A} , $|\mathbf{A}|$ is the operator norm. Let $\boldsymbol{\psi}(\mathbf{x}, \boldsymbol{\gamma})$ be any function of $\mathbf{x} \in \mathbb{R}^D$, $\boldsymbol{\gamma} \in \Theta$, maybe vector-

matrix-valued, $h : \mathbb{R}^D \rightarrow \mathbb{R}$, $\rho : \Theta \times \Theta \rightarrow \mathbb{R}$. We say that ψ satisfy **Condition $\Psi_{h,\rho}$** iff, for all $\gamma_1, \gamma_2 \in \Theta$ and all $\mathbf{x} \in \mathbb{R}^D$,

$$|\psi(\mathbf{x}, \gamma_1) - \psi(\mathbf{x}, \gamma_2)| \leq h(\mathbf{x})\rho(\gamma_1, \gamma_2).$$

A set of functions $\{\psi_i, i \in I\}$ satisfy Condition **$\Psi_{h,\rho}$** if ψ_i satisfy **$\Psi_{h,\rho}$** for each $i \in I$.

The following theorem states conditions of consistency of $\hat{\boldsymbol{\vartheta}}_n$ and consistency of $\hat{\boldsymbol{\vartheta}}_{-in}$ uniform by i .

Theorem 1 (Consistency). *Let the following assumptions hold.*

(C1) Θ is a compact set in \mathbb{R}^d .

(C2) Condition **$\Psi_{h,\rho}$** holds for the elementary estimating functions \mathbf{s} with some functions ρ and h .

(C3) ρ is a continuous function on $\Theta \times \Theta$ with $\rho(\boldsymbol{\gamma}, \boldsymbol{\gamma}) = 0$ for all $\boldsymbol{\gamma} \in \Theta$.

(C4) For all $l = 1, \dots, M$, $\mathbf{E} |\mathbf{s}(\boldsymbol{\xi}_{(l)}, \boldsymbol{\vartheta}^{(l)})|^2 < \infty$ and $\mathbf{E}(h(\boldsymbol{\xi}_{(l)}))^2 < \infty$.

(C5) $\mathbf{E} \mathbf{s}(\boldsymbol{\xi}_{(k)}, \boldsymbol{\gamma}) = 0$ if and only if $\boldsymbol{\gamma} = \boldsymbol{\vartheta}^{(k)}$.

(C6) $\det \boldsymbol{\Gamma}_\infty > 0$.

(C7) $\mathbf{P}\{\exists \boldsymbol{\gamma} \in \Theta, \text{ such that } \mathbf{S}_n^{(k)}(\boldsymbol{\gamma}) = 0\} \rightarrow 1$ as $n \rightarrow \infty$.

(C7') $\mathbf{P}\{\forall i = 1, \dots, n, \exists \boldsymbol{\gamma}_i \in \Theta, \text{ such that } \mathbf{S}_{-in}^{(k)}(\boldsymbol{\gamma}_i) = 0\} \rightarrow 1$ as $n \rightarrow \infty$.

Then, under the assumptions (C1)–(C7), $\hat{\boldsymbol{\vartheta}}_n^{(k)} \xrightarrow{\mathbf{P}} \boldsymbol{\vartheta}^{(k)}$ as $n \rightarrow \infty$, and under the assumptions (C1)–(C6), (C7'),

$$\sup_{i=1, \dots, n} |\hat{\boldsymbol{\vartheta}}_{-in}^{(k)} - \boldsymbol{\vartheta}^{(k)}| \xrightarrow{\mathbf{P}} 0 \quad (15)$$

as $n \rightarrow \infty$.

Assumptions (C7) and (C7') claim the existence of GEE solutions with probability tending to 1 as $n \rightarrow \infty$. They seem rather imperfect. The following theorem provides conditions under which they hold.

Let $\dot{\mathbf{S}}(\boldsymbol{\gamma})$ be the Jacobian of a vector-valued function \mathbf{S} , $\mathbf{S}_\infty(\boldsymbol{\gamma}) = \mathbf{E} \mathbf{s}(\boldsymbol{\xi}_{(k)}, \boldsymbol{\gamma})$.

Theorem 2 (Existence). *Let the assumptions (C1)–(C6) of Theorem 1 hold and, moreover,*

(E1) $\boldsymbol{\vartheta}^{(k)}$ is an inner point of Θ .

(E2) $\dot{\mathbf{S}}_\infty^{(k)}(\boldsymbol{\vartheta}^{(k)})$ exists and $\det \dot{\mathbf{S}}_\infty^{(k)}(\boldsymbol{\vartheta}^{(k)}) \neq 0$.

Then assumptions (C7) and (C7') of Theorem 1 hold.

To formulate the asymptotic normality result we need some additional notations. Let

$$\mathbf{M}^{(k)}(\boldsymbol{\gamma}) = \mathbf{E} \dot{\mathbf{s}}(\boldsymbol{\xi}_{(k)}, \boldsymbol{\gamma}), \quad \mathbf{M}^{(k)} = \mathbf{M}^{(k)}(\boldsymbol{\vartheta}^{(k)}) = \mathbf{E} \dot{\mathbf{s}}(\boldsymbol{\xi}_{(k)}, \boldsymbol{\vartheta}^{(k)}),$$

$$\langle \mathbf{a}^k \mathbf{a}^m \mathbf{p}^l \mathbf{p}^i \rangle = \lim_{n \rightarrow \infty} n \sum_{j=1}^n a_{j;n}^k a_{j;n}^m p_{j;n}^l p_{j;n}^i,$$

$$\langle \mathbf{a}^k \mathbf{a}^m \mathbf{p}^l \rangle = \lim_{n \rightarrow \infty} n \sum_{j=1}^n a_{j;n}^k a_{j;n}^m p_{j;n}^l,$$

(existence of these limits is a condition in the following Theorem 3). Now

$$\begin{aligned} \mathbf{Z}^{(m,l)} &= \sum_{i=1}^M \langle \mathbf{a}^m \mathbf{a}^l \mathbf{p}^i \rangle \mathbf{E} \mathbf{s}(\boldsymbol{\xi}_{(i)}, \boldsymbol{\vartheta}^{(m)}) \mathbf{s}(\boldsymbol{\xi}_{(i)}, \boldsymbol{\vartheta}^{(l)})^T \\ &\quad - \sum_{i_1, i_2=1}^M \langle \mathbf{a}^m \mathbf{a}^l \mathbf{p}^{i_1} \mathbf{p}^{i_2} \rangle \mathbf{E} \mathbf{s}(\boldsymbol{\xi}_{(i_1)}, \boldsymbol{\vartheta}^{(m)}) \mathbf{E} \mathbf{s}(\boldsymbol{\xi}_{(i_2)}, \boldsymbol{\vartheta}^{(l)})^T, \\ \mathbf{V}^{(m,l)} &= (\mathbf{M}^{(m)})^{-1} \mathbf{Z}^{(m,l)} (\mathbf{M}^{(l)})^{-T} \end{aligned}$$

(here and below $\mathbf{M}^{-T} = (\mathbf{M}^{-1})^T$). Let's pack all the matrices $\mathbf{V}^{(m,l)}$ into one $(Md) \times (Md)$ matrix

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}^{(1,1)} & \dots & \mathbf{V}^{(1,M)} \\ \vdots & \ddots & \vdots \\ \mathbf{V}^{(M,1)} & \dots & \mathbf{V}^{(M,M)} \end{pmatrix}. \quad (16)$$

Let $\mathbf{s}(\mathbf{x}, \boldsymbol{\gamma}) = (s^1(\mathbf{x}, \boldsymbol{\gamma}), \dots, s^d(\mathbf{x}, \boldsymbol{\gamma}))^T$.

Theorem 3 (Asymptotic normality). *Let the following assumptions hold.*

(AN1) $\boldsymbol{\vartheta}$ is an inner point of $\Theta^M = \Theta \times \dots \times \Theta$.

(AN2) There exists an open ball B centered in $\boldsymbol{\vartheta}$, such that the derivatives

$$\frac{\partial^2 s^l(\mathbf{x}, \boldsymbol{\gamma})}{\partial \gamma^i \partial \gamma^j}$$

exist for all $\boldsymbol{\gamma} = (\gamma^1, \dots, \gamma^d)^T \in B$, all $l, i, j = 1, \dots, d$, and almost all \mathbf{x} (w.r.t. all $F^{(k)}$, $k = 1, \dots, M$).

(AN3) There exists a function $h : \mathbb{R}^D \rightarrow \mathbb{R}$ such that

$$\max_{l,i,j} \sup_{\boldsymbol{\gamma} \in B} \left| \frac{\partial^2 s^l(\mathbf{x}, \boldsymbol{\gamma})}{\partial \gamma^i \partial \gamma^j} \right| \leq h(\mathbf{x})$$

and $\mathbf{E}(h(\boldsymbol{\xi}_{(k)}))^\alpha < \infty$ for some $\alpha > 1$ for all $k = 1, \dots, M$.

(AN4) $\mathbf{E} |\mathbf{s}(\boldsymbol{\xi}_{(k)}, \boldsymbol{\vartheta}^{(k)})|^2 < \infty$ for all $k = 1, \dots, M$.

(AN5) $\mathbf{M}^{(k)}$ are finite and nonsingular for all $k = 1, \dots, M$.

(AN6) The limits $\langle \mathbf{a}^k \mathbf{a}^m \mathbf{p}^i \mathbf{p}^l \rangle$ exist for all $k, m, i, l = 1, \dots, M$.

(AN7) Matrix $\mathbf{\Gamma}_\infty$ exists and is nonsingular.

(AN8) $\hat{\boldsymbol{\vartheta}}_n$ exists and is a consistent estimator for $\boldsymbol{\vartheta}$.

Then

$$\sqrt{n}(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}) \xrightarrow{W} N(0, \mathbf{V})$$

as $n \rightarrow \infty$.

(Note that in Theorems 1–4 and the lemmas in Section 7 below the functions h can be different).

In fact, Theorem 3 is just Theorem 2 from [7] reformulated in terms of the present paper.

Now we are ready to formulate the theorem on consistency of the JK estimator of \mathbf{V} .

Theorem 4. Assume that assumptions (AN1), (AN5), (AN6), (AN7) of Theorem 3 hold and, moreover:

(JK1) There exists a function $h : \mathbb{R}^D \rightarrow \mathbb{R}$ such that

$$\sup_{\boldsymbol{\gamma} \in \Theta} |\mathbf{s}(\mathbf{x}, \boldsymbol{\gamma})| \leq h(\mathbf{x}), \quad \sup_{\boldsymbol{\gamma} \in \Theta} |\dot{\mathbf{s}}(\mathbf{x}, \boldsymbol{\gamma})| \leq h(\mathbf{x}),$$

$$\max_{l,i,j} \sup_{\boldsymbol{\gamma} \in B} \left| \frac{\partial^2 s^l(\mathbf{x}, \boldsymbol{\gamma})}{\partial \gamma^i \partial \gamma^j} \right| \leq h(x)$$

and for some $\alpha > 4$,

$$\mathbf{E}(h(\boldsymbol{\xi}_{(l)}))^\alpha < \infty,$$

(JK2) $\hat{\boldsymbol{\vartheta}}_n$ is a \sqrt{n} -consistent estimator of $\boldsymbol{\vartheta}$,

(JK3) $\sup_{i=1, \dots, n} |\hat{\boldsymbol{\vartheta}}_{-in} - \boldsymbol{\vartheta}| \xrightarrow{P} 0$ as $n \rightarrow \infty$.

Then $\hat{\mathbf{V}}_n \xrightarrow{P} \mathbf{V}$ as $n \rightarrow \infty$.

Example. Let the observed data be $\boldsymbol{\xi}_j = (Y_j, X_j)^T$, $j = 1, \dots, n$, where dependence between X_j and Y_j is described by the regression model (3) with

$$g(X_j, \boldsymbol{\vartheta}^{(k)}) = \frac{1}{1 + \exp(-\boldsymbol{\vartheta}_0^{(k)} - \boldsymbol{\vartheta}_1^{(k)} X_j)}, \quad (17)$$

where $\boldsymbol{\vartheta}^{(k)} = (\boldsymbol{\vartheta}_0^{(k)}, \boldsymbol{\vartheta}_1^{(k)})^T$ are the vectors of regression coefficients for the k -th mixture component.

Assume that $\boldsymbol{\gamma} \in \Theta$, where Θ is a compact set in \mathbb{R}^2 . Then for the elementary estimating function \mathbf{s} defined by (8) we obtain

$$|\mathbf{s}(\boldsymbol{\xi}_j, \boldsymbol{\gamma})| \leq C(1 + |X_j|)(1 + |\varepsilon_j|) \text{ and } \left| \frac{\partial^2}{\partial \gamma^i \gamma^i} s^l(\boldsymbol{\xi}_j, \boldsymbol{\gamma}) \right| \leq C(1 + |X_j|^3)(1 + |\varepsilon_j|),$$

where $C < \infty$ is some constant. So Assumption (JK1) holds if $\mathbf{E}[(\varepsilon_j)^4 | \kappa_j = k] < \infty$ and $\mathbf{E}[|X_j|^{12} | \kappa_j = k] < \infty$ for all $k = 1, \dots, M$. Assumption (AN5) holds if

$$\text{Var}[X_j | \kappa_j = k] > 0. \quad (18)$$

Assumption (C5) also holds under (18), see Theorem 2 in [8]. So, under rather mild assumptions Theorems 1–4 hold for generalized least squares estimator in this model. In [7] confidence sets for $\boldsymbol{\vartheta}^{(k)}$ are constructed based on the asymptotic normality of $\hat{\boldsymbol{\vartheta}}_n^{(k)}$ and consistency of $\hat{\mathbf{V}}_n^{(k,k)}$. Namely, the confidence ellipsoid for $\boldsymbol{\vartheta}^{(k)}$ is defined as

$$B_{\alpha,n} = \{\boldsymbol{\gamma} \in \mathbb{R}^d : (\boldsymbol{\gamma} - \boldsymbol{\vartheta}^{(k)})^T (\hat{\mathbf{V}}_n^{(k,k)})^{-1} (\boldsymbol{\gamma} - \boldsymbol{\vartheta}^{(k)}) \leq Q^\eta(1 - \alpha)\},$$

where $Q^\eta(1 - \alpha)$ is the quantile of level $1 - \alpha$ of the χ^2 -distribution with d degrees of freedom. Then under the assumptions of Theorems 1–4, if $\det \mathbf{Z}^{(k,k)} \neq 0$,

$$\lim_{n \rightarrow \infty} \mathbf{P}\{\boldsymbol{\vartheta}^{(k)} \in B_{\alpha,n}\} = 1 - \alpha.$$

In [7] results of simulations are presented which show that these ellipsoids can be used for samples large enough.

5 Application to sociological data

In this section we show how the considered technique can be applied to the statistical analysis of real life data. In many sociological problems one deals with two sets of data from two different sources. The first (I) set consists of individual records with values of variables which present personal information of investigated persons. The second (A) set contain some averaged information on large groups of a variable of investigated persons which is not presented in the I-set. The problem is how to merge information from A- and I-sets to infer on the model involving variables from both sets.

We consider as the I-set a data of results of Ukrainian External Independent Testing (EIT) in 2016 from the official site of *Ukrainian Center for Educational Quality Assessment*. EIT exams are to be passed by high school graduates for admission to universities. Information on scores in *Ukrainian language and literature* (Ukr) and on *Mathematics* (Math) for nearly 246 000 examinees of EIT-2016 is available.¹ In [5] linear regression dependence between Ukr and Math is assumed. In this paper we consider the model

$$\text{Ukr} = \frac{1}{1 + \exp(-\boldsymbol{\vartheta}_0^{(k)} - \boldsymbol{\vartheta}_1^{(k)} \text{Math})} + \varepsilon, \quad (19)$$

¹Original EIT scores range from 100 to 200. In this presentation we rescale them onto [0, 1].

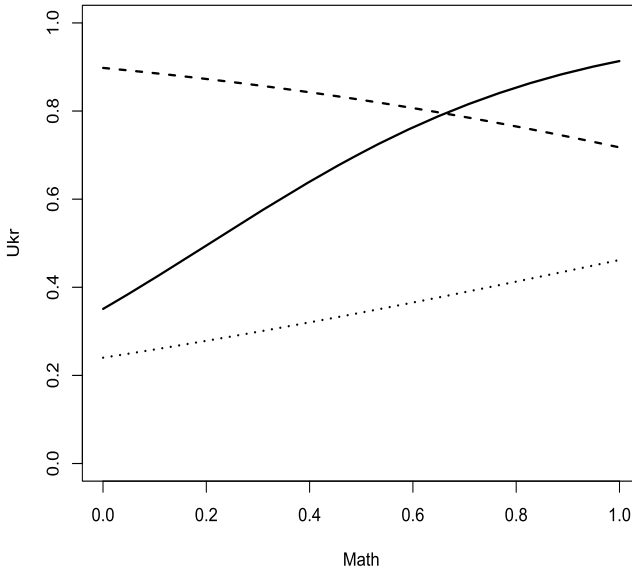


Fig. 1. Estimated logistic regression lines by EIT-2016 data. Solid line for 1st component, dashed line for 2nd, dotted line for 3rd one

in which the coefficients $\vartheta_0^{(k)}$ and $\vartheta_1^{(k)}$ depend on the political attitudes of the adult environment in which the student was brought up. There can be a family of Ukrainian independence adherents or an environment critical to existence of Ukrainian state and culture.

EIT-2016 does not contain information on political issues. But for each examinee the region of Ukraine is recorded where he/she graduated. So we used data on results of Ukrainian Parliament (Verhovna Rada) elections-2014 to get approximate proportions of adherents of different political choices in regions of Ukraine (A-set). All possible electoral choices at these elections (voting for one of parties, voting against all or not to take part in the voting) were divided into three groups (components): (1) pro-Ukrainian, (2) contra-Ukrainian and (3) neutral (see [5] for details). The concentrations of components are taken as frequencies of adherents of corresponding electoral choice at the region where j -th examinee attended high school. On Fig. 1 the fitted regression lines are presented. The dependence between Math and Ukr on this picture seems significantly different in the three components. Say, in the pro component it is increasing and seemingly nonlinear, in the contra component it is decreasing, in the neutral one it is increasing and quite near to linear dependence.

To verify significance of these differences we constructed the confidence ellipsoids for the parameters as described in Section 4. By the Bonferroni rule, to infer with the significance level $\alpha_0 = 0.05$, we took the levels of the ellipsoids $\alpha = \alpha_0/3 = 0.01666$. Obtained ellipsoids are presented on Fig. 2. Since they are not intersecting, we conclude that the differences between the parameters are significant for all the components.

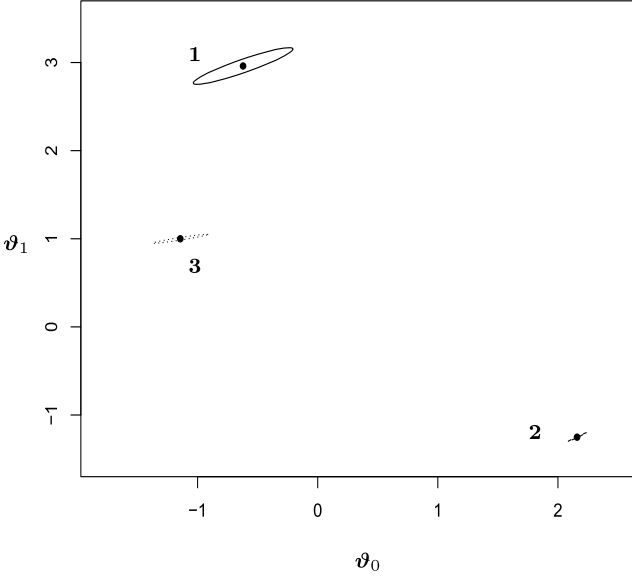


Fig. 2. Confidence ellipsoids for logistic regression parameters by EIT-2016 data

6 Conclusion

So, we obtained conditions under which the JK estimator for asymptotic variance is valid for nonlinear GEE estimators in MVC models. The presented example of sociological data analysis demonstrates possibilities of practical applications of this estimator.

7 Proofs

We start from some auxiliary lemmas.

Lemma 1. Assume that Γ_∞ exists and is nonsingular. Then for some constant $C_a < \infty$,

$$|a_{j;n}^k| \leq \frac{C_a}{n}, \quad |a_{j;n}^k - a_{j;-in}^k| \leq \frac{C_a}{n^2}$$

for all $k = 1, \dots, M$, $1 \leq i \neq j \leq n$, $n = 1, 2, \dots$

For the **proof**, see [5], lemma 1.

Let $\psi_{j;n}$, $j = 1, \dots, n$, $n = 1, 2, \dots$, be any set of functions with domain \mathbb{R}^D and values in a set of scalars, or vectors, or matrices. We use the following notation

$$\Psi_n^{(k)}(\boldsymbol{\gamma}) = \sum_{j=1}^n a_{j;n}^k \psi_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\gamma}),$$

$$\Psi_{-in}^{(k)}(\boldsymbol{\gamma}) = \sum_{j \neq i} a_{j;-in}^k \psi_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\gamma}).$$

In this notation ψ can be replaced by any other symbol, e.g., $\tilde{\psi}$, \mathbf{s} or $\dot{\mathbf{s}}$.

Lemma 2. Let $H : \mathbb{R}^D \rightarrow \mathbb{R}$ be some function, such that

$$(i) \max_{j,n} \sup_{\boldsymbol{\gamma} \in \Theta} |\psi_{j;n}(x, \boldsymbol{\gamma})| \leq H(x),$$

$$(ii) \text{ for some } \alpha > 1 \text{ and all } k = 1, \dots, M, \mathbf{E}(H(\boldsymbol{\xi}_{(k)}))^\alpha < \infty.$$

Then

$$\max_{j=1, \dots, n} \sup_{\boldsymbol{\gamma} \in \Theta} |\psi_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\gamma})| = o_P(n^{1/\alpha}).$$

Proof. Let

$$\tilde{\psi}_{j;n} = \sup_{\boldsymbol{\gamma} \in \Theta} |\psi_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\gamma})|.$$

Then

$$\mathbb{P}_n(C) = \mathbf{P}\left\{ \max_{j=1, \dots, n} \tilde{\psi}_{j;n} > C \right\} = 1 - \prod_{j=1}^n (1 - \mathbf{P}\{\tilde{\psi}_{j;n} > C\}). \quad (20)$$

Put $C_n = C_0 n^{1/\alpha}$ for some $C_0 > 0$. Then

$$\tilde{p}_n = \sup_{j=1, \dots, n} \mathbf{P}\{\tilde{\psi}_{j;n} > C_n\} = o(1/n). \quad (21)$$

Really,

$$\begin{aligned} \mathbf{P}\{\tilde{\psi}_{j;n} > C_n\} &\leq \mathbf{P}\{H(\boldsymbol{\xi}_{j;n}) > C_n\} \\ &\leq \sum_{k=1}^M p_{j;n}^k \mathbf{P}\{H(\boldsymbol{\xi}_{(k)}) > C_n\} \leq \sum_{k=1}^M \mathbf{P}\{H(\boldsymbol{\xi}_{(k)}) > C_n\}. \end{aligned}$$

So, to show (21) one needs only to observe that

$$n \mathbf{P}\{H(\boldsymbol{\xi}_{(k)}) > C_n\} = o(1) \text{ for all } k = 1, \dots, M. \quad (22)$$

But

$$\begin{aligned} n \mathbf{P}\{H(\boldsymbol{\xi}_{(k)}) > C_n\} &= n \mathbf{E} \mathbf{1}\{H(\boldsymbol{\xi}_{(k)}) > C_n\} \\ &\leq n \mathbf{E} \mathbf{1}\{H(\boldsymbol{\xi}_{(k)}) > C_n\} \frac{H(\boldsymbol{\xi}_{(k)})^\alpha}{C_n^\alpha} \leq \frac{1}{C_0^\alpha} \mathbf{E} H(\boldsymbol{\xi}_{(k)})^\alpha \mathbf{1}\{H(\boldsymbol{\xi}_{(k)}) > C_n\} = o(1) \end{aligned}$$

due to the assumption (ii) of the lemma. So (21) holds. From (20) and (21) we obtain

$$\mathbb{P}_n(C) \leq 1 - \exp(n \log(1 - \tilde{p}_n)) = 1 - \exp(-n \cdot o(1/n)) = o(1)$$

for any $C_0 > 0$. □

Lemma 3. Let the following assumptions hold.

$$(i) \det \Gamma_\infty \neq 0.$$

(ii) A set of functions $\psi_{j;n}$, $j = 1, \dots, n$, $n \in \mathbb{N}$, satisfy Condition $\Psi_{h,\rho}$ with a function h such that $\mathbf{E}(h(\boldsymbol{\xi}_{(k)}))^2 < \infty$ for all $k = 1, \dots, M$.

Then there exists a sequence of random variables $\zeta_n = O_p(1)$, such that for all $n = 1, 2, \dots$, all $\gamma_1, \gamma_2 \in \Theta$ and all $i = 1, \dots, n$ the following inequalities hold:

$$|\Psi_n^{(k)}(\gamma_1) - \Psi_n^{(k)}(\gamma_2)| \leq \zeta_n \rho(\gamma_1, \gamma_2), \quad (23)$$

$$|\Psi_{-in}^{(k)}(\gamma_1) - \Psi_{-in}^{(k)}(\gamma_2)| \leq \zeta_n \rho(\gamma_1, \gamma_2). \quad (24)$$

Proof. Let us start with (23). Observe that by Condition $\Psi_{h,\rho}$ and Lemma 1,

$$|\Psi_n^{(k)}(\gamma_1) - \Psi_n^{(k)}(\gamma_2)| \leq \sum_{j=1}^n |a_{j;n}^k| h(\xi_{j;n}) \rho(\gamma_1, \gamma_2) \leq \frac{C_a \rho(\gamma_1, \gamma_2)}{n} \sum_{j=1}^n h(\xi_{j;n}).$$

By assumption (ii) of the lemma

$$A_1 = \max_{j,n} \mathbf{E} h(\xi_{j;n}) \leq \max_{k=1, \dots, M} \mathbf{E} h(\xi^{(k)}) < \infty$$

and $A_2 = \mathbf{E}(h(\xi_{j;n}))^2 < \infty$. So, for any $\lambda > A_1$,

$$\mathbf{P} \left\{ \frac{1}{n} \sum_{j=1}^n h(\xi_{j;n}) > \lambda \right\} \leq \frac{\text{Var} \frac{1}{n} \sum_{j=1}^n h(\xi_{j;n})}{\left(\lambda - \mathbf{E} \frac{1}{n} \sum_{j=1}^n h(\xi_{j;n}) \right)^2} \leq \frac{A_2}{n(\lambda - A_1)^2} \rightarrow 0$$

as $n \rightarrow \infty$. So (23) holds with $\zeta_n = \frac{C_a}{n} \sum_{j=1}^n h(\xi_{j;n})$.

To show (24) observe that

$$|\Psi_{-in}^{(k)}(\gamma_1) - \Psi_{-in}^{(k)}(\gamma_2)| \leq \sum_{j=1}^n |a_{j;-in}^k| h(\xi_{j;n}) \rho(\gamma_1, \gamma_2) \leq \frac{C_a \rho(\gamma_1, \gamma_2)}{n} \sum_{j=1}^n h(\xi_{j;n})$$

and

$$|a_{j;-in}^k| \leq |a_{j;n}^k| + |a_{j;n}^k - a_{j;-in}^k| \leq \frac{C'_a}{n}$$

for some $C'_a < \infty$ due to Lemma 1. The rest of the proof is the same as for (23). \square

Lemma 4. *Let the following assumptions hold.*

- (i) Θ is a compact in \mathbb{R}^d .
- (ii) Condition $\Psi_{h,\rho}$ holds for $\{\psi_{j;n}, j = 1, \dots, n, n = 1, 2, \dots\}$.
- (iii) ρ is a continuous function on $\Theta \times \Theta$ and $\rho(\gamma, \gamma) = 0$ for all $\gamma \in \Theta$.
- (iv) For all $k = 1, \dots, M$, $\mathbf{E}(h(\xi^{(k)}))^2 < \infty$, $\max_{j,n} \mathbf{E} |\psi_{j;n}(\xi^{(k)}, \vartheta^{(k)})|^2 < \infty$.
- (v) $\det \Gamma_\infty \neq 0$.

Then

$$\sup_{\gamma \in \Theta} |\Psi_n^{(k)}(\gamma) - \mathbf{E} \Psi_n^{(k)}(\gamma)| \xrightarrow{\mathbf{P}} 0, \quad (25)$$

$$\sup_{\gamma \in \Theta} |\Psi_{-in}^{(k)}(\gamma) - \mathbf{E} \Psi_{-in}^{(k)}(\gamma)| \xrightarrow{\mathbf{P}} 0 \quad (26)$$

as $n \rightarrow \infty$.

Proof. Let $\bar{\psi}_{j;n}(\boldsymbol{\gamma}) = \mathbf{E} \psi_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\gamma})$,

$$\tilde{\psi}_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\gamma}) = \psi_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\gamma}) - \bar{\psi}_{j;n}(\boldsymbol{\gamma}).$$

Then $\tilde{\psi}_{j;n}$ satisfy Condition $\Psi_{\tilde{h}, \rho}$ with $\tilde{h}(\mathbf{x}) = h(\mathbf{x}) + C_\Psi$, where C_Ψ is some constant, e.g.,

$$C_\Psi = \max_{k=1, \dots, M} \mathbf{E} h(\boldsymbol{\xi}^{(k)}) < \infty.$$

By the assumption (iv),

$$\mathbf{E}(\tilde{h}(\boldsymbol{\xi}^{(k)}))^2 < \infty \quad (27)$$

and

$$\max_{j,n} \mathbf{E} |\tilde{\psi}_{j;n}(\boldsymbol{\xi}^{(k)}, \boldsymbol{\vartheta}^{(k)})|^2 < \infty, \text{ for all } k = 1, \dots, M. \quad (28)$$

To prove the lemma is sufficient to show that

$$\sup_{\boldsymbol{\gamma} \in \Theta} |\tilde{\Psi}_n^{(k)}(\boldsymbol{\gamma})| \xrightarrow{\mathbf{P}} 0 \quad (29)$$

and

$$\sup_{\boldsymbol{\gamma} \in \Theta} |\tilde{\Psi}_{-in}^{(k)}(\boldsymbol{\gamma})| \xrightarrow{\mathbf{P}} 0 \quad (30)$$

as $n \rightarrow \infty$.

Let us show (29). Consider the case when $\psi_{j;n}$ are scalar-valued. Then

$$\mathbf{E} \tilde{\Psi}_n^{(k)}(\boldsymbol{\gamma}) = \sum_{j=1}^n a_{j;n}^k \mathbf{E} \tilde{\psi}_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\gamma}) = 0 \quad (31)$$

and

$$\text{Var} \tilde{\Psi}_n^{(k)}(\boldsymbol{\gamma}) = \sum_{j=1}^n (a_{j;n}^k)^2 \text{Var} \tilde{\psi}_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\gamma}) \leq \frac{C_a^2}{n} \max_{j,n} \text{Var} \tilde{\psi}_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\gamma}). \quad (32)$$

Assumptions (iii) and (iv) with inequalities (27) and (28) imply

$$\max_{j,n} \text{Var} \tilde{\psi}_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\gamma}) < \infty.$$

So, by (31) and (32), we get

$$\tilde{\Psi}_n^{(k)}(\boldsymbol{\gamma}) \xrightarrow{\mathbf{P}} 0 \text{ as } n \rightarrow \infty \quad (33)$$

for any $\boldsymbol{\gamma} \in \Theta$.

It is obvious that if $\psi_{j;n}$ are vector- or matrix-valued, then (33) holds coordinate-wise.

Applying Lemma 3 to $\tilde{\psi}_{j;n}$ one obtains

$$|\tilde{\Psi}_n^{(k)}(\boldsymbol{\gamma}_1) - \tilde{\Psi}_n^{(k)}(\boldsymbol{\gamma}_2)| \leq \zeta_n \rho(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) \text{ for all } \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \Theta \quad (34)$$

with some $\zeta_n = O_P(1)$. To prove (29) we have to show that for any $\delta > 0$ and $\varepsilon > 0$ there exists such n_0 that for all $n > n_0$

$$\mathbf{P}\{\sup_{\boldsymbol{\gamma} \in \Theta} |\tilde{\Psi}_n^{(k)}(\boldsymbol{\gamma})| > \delta\} < \varepsilon. \quad (35)$$

Fix ε and δ . Choose a nonrandom $C_\zeta < \infty$ such that $\mathbf{P}\{\zeta_n > C_\zeta\} \leq \varepsilon/2$ for all n .

Since Θ is compact by assumption (iii) of the lemma there exists a finite set $T = \{\mathbf{t}_1, \dots, \mathbf{t}_L\} \subset \Theta$ such that for each $\boldsymbol{\gamma} \in \Theta$ one can choose $l(\boldsymbol{\gamma}) \in \{1, \dots, L\}$ with

$$\rho(\boldsymbol{\gamma}, \mathbf{t}_{l(\boldsymbol{\gamma})}) < \frac{\delta}{2C_\zeta}. \quad (36)$$

Note that

$$|\tilde{\Psi}_n^{(k)}(\boldsymbol{\gamma})| \leq |\tilde{\Psi}_n^{(k)}(\boldsymbol{\gamma}) - \tilde{\Psi}_n^{(k)}(\mathbf{t}_{l(\boldsymbol{\gamma})})| + |\tilde{\Psi}_n^{(k)}(\mathbf{t}_{l(\boldsymbol{\gamma})})|,$$

so

$$\begin{aligned} \sup_{\boldsymbol{\gamma} \in \Theta} |\tilde{\Psi}_n^{(k)}(\boldsymbol{\gamma})| &\leq \sup_{\boldsymbol{\gamma} \in \Theta} |\tilde{\Psi}_n^{(k)}(\boldsymbol{\gamma}) - \tilde{\Psi}_n^{(k)}(\mathbf{t}_{l(\boldsymbol{\gamma})})| + \max_{l=1, \dots, L} |\tilde{\Psi}_n^{(k)}(\mathbf{t}_l)| \\ &\leq \zeta_n \sup_{\boldsymbol{\gamma} \in \Theta} \rho(\boldsymbol{\gamma}, \mathbf{t}_{l(\boldsymbol{\gamma})}) + \max_{l=1, \dots, L} |\tilde{\Psi}_n^{(k)}(\mathbf{t}_l)|. \end{aligned}$$

Therefore

$$\mathbf{P}\{\sup_{\boldsymbol{\gamma} \in \Theta} |\tilde{\Psi}_n^{(k)}(\boldsymbol{\gamma})| > \delta\} \leq \mathbf{P}\{\zeta_n \sup_{\boldsymbol{\gamma} \in \Theta} \rho(\boldsymbol{\gamma}, \mathbf{t}_{l(\boldsymbol{\gamma})}) > \delta/2\} + \mathbf{P}\{\max_{l=1, \dots, L} |\tilde{\Psi}_n^{(k)}(\mathbf{t}_l)| > \delta/2\}. \quad (37)$$

The second term in the RHS of (37) tends to 0 as $n \rightarrow \infty$ due to (33). So it is less than $\varepsilon/2$ for n large enough. By (36) the first term is less than $\mathbf{P}\{\zeta_n > C_\zeta\} < \varepsilon/2$.

So (35) holds and (25) is shown.

Let us show (26). Observe that

$$\tilde{\Psi}_n^{(k)}(\boldsymbol{\gamma}) - \tilde{\Psi}_{-in}^{(k)}(\boldsymbol{\gamma}) = a_{i;n}^k \tilde{\psi}_i(\boldsymbol{\xi}_{i;n}, \boldsymbol{\gamma}) + \sum_{j \neq i} (a_{j;n}^k - a_{j;-in}^k) \tilde{\psi}_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\gamma}).$$

To estimate $\max_{j,n} \sup_{\boldsymbol{\gamma} \in \Theta} |\tilde{\psi}_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\gamma})|$ we apply Lemma 2 with $H(x) = |x|$. Assumption (iv) of Lemma 4 implies that the assertion of Lemma 2 holds with $\alpha = 2$. So

$$\max_{j,n} \sup_{\boldsymbol{\gamma} \in \Theta} |\tilde{\psi}_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\gamma})| = O_P(n^{1/2}).$$

From Lemma 1 we get

$$|a_{i;n}^k| \leq C_a/n, \quad |a_{j;n}^k - a_{j;-in}^k| \leq C_a/n^2,$$

so

$$\max_{i=1, \dots, n} \sup_{\boldsymbol{\gamma} \in \Theta} |\tilde{\Psi}_n^{(k)}(\boldsymbol{\gamma}) - \tilde{\Psi}_{-in}^{(k)}(\boldsymbol{\gamma})| = O_P(n^{-1/2}).$$

This with (25) implies (26). \square

Proof of Theorem 1. 1. We will show that $\hat{\boldsymbol{\vartheta}}_n^{(k)} \xrightarrow{P} \boldsymbol{\vartheta}^{(k)}$ as $n \rightarrow \infty$.

Let $\mathbf{S}_\infty^{(k)}(\boldsymbol{\gamma}) = \mathbf{E} \mathbf{s}(\boldsymbol{\xi}^{(k)}, \boldsymbol{\gamma})$. Assumptions (C3) and (C4) imply that $\mathbf{S}_\infty^{(k)}(\boldsymbol{\gamma})$ is continuous on $\boldsymbol{\gamma} \in \Theta$. By (C5), $|\mathbf{S}_\infty^{(k)}(\boldsymbol{\gamma})| > 0$ for all $\boldsymbol{\gamma} \neq \boldsymbol{\vartheta}^{(k)}$.

Fix any $\varepsilon > 0$ and consider $\mathcal{N}_\varepsilon = \{\boldsymbol{\gamma} \in \Theta : |\boldsymbol{\gamma} - \boldsymbol{\vartheta}^{(k)}| \geq \varepsilon\}$. Then

$$s_{\min} = \inf_{\boldsymbol{\gamma} \in \mathcal{N}_\varepsilon} |\mathbf{S}_\infty^{(k)}(\boldsymbol{\gamma})| > 0.$$

For $\boldsymbol{\gamma} \in \mathcal{N}_\varepsilon$

$$\begin{aligned} |\mathbf{S}_n^{(k)}(\boldsymbol{\gamma})| &\geq |\mathbf{S}_\infty^{(k)}(\boldsymbol{\gamma})| - |\mathbf{S}_\infty^{(k)}(\boldsymbol{\gamma}) - \mathbf{S}_n^{(k)}(\boldsymbol{\gamma})| \\ &\geq s_{\min} - \sup_{\boldsymbol{\gamma} \in \Theta} |\mathbf{S}_\infty^{(k)}(\boldsymbol{\gamma}) - \mathbf{S}_n^{(k)}(\boldsymbol{\gamma})|. \end{aligned}$$

So

$$\mathbf{P} \left\{ \inf_{\boldsymbol{\gamma} \in \mathcal{N}_\varepsilon} |\mathbf{S}_n^{(k)}(\boldsymbol{\gamma})| \leq \frac{s_{\min}}{2} \right\} \leq \mathbf{P} \left\{ \sup_{\boldsymbol{\gamma} \in \Theta} |\mathbf{S}_\infty^{(k)}(\boldsymbol{\gamma}) - \mathbf{S}_n^{(k)}(\boldsymbol{\gamma})| \geq \frac{s_{\min}}{2} \right\}.$$

Applying Lemma 4 with $\psi_{j;n} = \mathbf{s}$ one obtains that

$$\sup_{\boldsymbol{\gamma} \in \Theta} |\mathbf{S}_\infty^{(k)}(\boldsymbol{\gamma}) - \mathbf{S}_n^{(k)}(\boldsymbol{\gamma})| \xrightarrow{P} 0 \quad (38)$$

as $n \rightarrow \infty$. Therefore

$$\mathbf{P} \left\{ \inf_{\boldsymbol{\gamma} \in \mathcal{N}_\varepsilon} |\mathbf{S}_n^{(k)}(\boldsymbol{\gamma})| \leq \frac{s_{\min}}{2} \right\} \rightarrow 0.$$

Since $\mathbf{S}_n^{(k)}(\hat{\boldsymbol{\vartheta}}_n^{(k)}) = 0$, this implies $\mathbf{P}\{\hat{\boldsymbol{\vartheta}}_n^{(k)} \in \mathcal{N}_\varepsilon\} \rightarrow 0$ as $n \rightarrow \infty$, i.e. $\hat{\boldsymbol{\vartheta}}_n^{(k)} \xrightarrow{P} \boldsymbol{\vartheta}^{(k)}$.

2. Let us show (15).

By Lemma 4,

$$\max_{i=1, \dots, n} \sup_{\boldsymbol{\gamma} \in \Theta} |\mathbf{S}_{-in}^{(k)} - \mathbf{S}_\infty^{(k)}(\boldsymbol{\gamma})| \xrightarrow{P} 0. \quad (39)$$

Then

$$\begin{aligned} &\mathbf{P} \left\{ \min_{i=1, \dots, n} \inf_{\boldsymbol{\gamma} \in \mathcal{N}_\varepsilon} |\mathbf{S}_{-in}^{(k)}(\boldsymbol{\gamma})| \leq \frac{s_{\min}}{2} \right\} \\ &\leq \mathbf{P} \left\{ \max_{i=1, \dots, n} \sup_{\boldsymbol{\gamma} \in \Theta} |\mathbf{S}_\infty^{(k)}(\boldsymbol{\gamma}) - \mathbf{S}_{-in}^{(k)}(\boldsymbol{\gamma})| \geq \frac{s_{\min}}{2} \right\} \xrightarrow{P} 0. \end{aligned}$$

From this we obtain (15) by the same way as in the first part of the proof. \square

Proof of Theorem 2. The proof follows the lines of the proof of Theorem A.10 in [6] with the use of (38) and (39) instead of the law of large numbers. \square

Proof of Theorem 4. Let

$$\mathbf{s}_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\gamma}) = \mathbf{s}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\gamma}) - \mathbf{E} \mathbf{s}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\gamma}).$$

Then, due to (6),

$$\begin{aligned} \sum_{j=1}^n a_{j;n}^l \mathbf{s}_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\gamma}) &= \sum_{j=1}^n a_{j;n}^l \mathbf{s}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\gamma}) - \sum_{m=1}^M \sum_{j=1}^n a_{j;n}^l \mathcal{P}_{j;n}^m \mathbf{E} \mathbf{s}(\boldsymbol{\xi}_{(m)}, \boldsymbol{\gamma}) \\ &= \sum_{j=1}^n a_{j;n}^l \mathbf{s}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\gamma}) = \mathbf{S}_n^{(l)}(\boldsymbol{\gamma}). \end{aligned}$$

So

$$\mathbf{S}_n^{(l)}(\hat{\boldsymbol{\vartheta}}_n^{(l)}) = \sum_{j=1}^n a_{j;n}^l \mathbf{s}_{j;n}(\boldsymbol{\xi}_{j;n}, \hat{\boldsymbol{\vartheta}}_n^{(l)}) = 0. \quad (40)$$

Similarly,

$$\mathbf{S}_{-in}^{(l)}(\hat{\boldsymbol{\vartheta}}_{-in}^{(l)}) = \sum_{j \neq i} a_{j;-in}^l \mathbf{s}_{j;n}(\boldsymbol{\xi}_{j;n}, \hat{\boldsymbol{\vartheta}}_{-in}^{(l)}) = 0. \quad (41)$$

By the Mean Value theorem, there exists $t_{ni}^l \in [0, 1]$ such that

$$-\mathbf{S}_{-in}^{(l)}(\hat{\boldsymbol{\vartheta}}_n^{(l)}) = \mathbf{S}_{-in}^{(l)}(\hat{\boldsymbol{\vartheta}}_{-in}^{(l)}) - \mathbf{S}_{-in}^{(l)}(\hat{\boldsymbol{\vartheta}}_n^{(l)}) = \dot{\mathbf{S}}_{-in}^{(l)}(\boldsymbol{\zeta}_{-in}^{(l)})(\hat{\boldsymbol{\vartheta}}_{-in}^{(l)} - \hat{\boldsymbol{\vartheta}}_n^{(l)}), \quad (42)$$

where

$$\boldsymbol{\zeta}_{-in}^{(l)} = (1 - t_{ni}^l) \hat{\boldsymbol{\vartheta}}_{-in}^{(l)} + t_{ni}^l \hat{\boldsymbol{\vartheta}}_n^{(l)}.$$

Observe that, by Assumption (JK1),

$$|\dot{\mathbf{s}}_{j;n}(\mathbf{x}, \boldsymbol{\gamma}_1) - \dot{\mathbf{s}}_{j;n}(\mathbf{x}, \boldsymbol{\gamma}_2)| \leq 2h(\mathbf{x})|\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2|.$$

Applying Lemma 4 with $\psi_{j;n} = \dot{\mathbf{s}}_{j;n}^l$, $\rho(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$, we obtain

$$\max_{i=1, \dots, n} \sup_{\boldsymbol{\gamma} \in \Theta} |\dot{\mathbf{S}}_{-in}^{(l)}(\boldsymbol{\gamma}) - \mathbf{M}^{(l)}(\boldsymbol{\gamma})| \xrightarrow{\mathbb{P}} 0 \quad (43)$$

as $n \rightarrow \infty$.

By Assumption (JK1), applying the Lebesgue Dominated Convergence theorem, we obtain $\mathbf{M}^{(l)}(\boldsymbol{\gamma}) \rightarrow \mathbf{M}^{(l)}(\boldsymbol{\vartheta}^{(l)})$ as $\boldsymbol{\gamma} \rightarrow \boldsymbol{\vartheta}^{(l)}$. So, by Assumptions (JK2) and (JK3), we obtain

$$\max_{i=1, \dots, n} |\dot{\mathbf{S}}_{-in}^{(l)}(\boldsymbol{\zeta}_{-in}^{(l)}) - \mathbf{M}^{(l)}(\boldsymbol{\vartheta}^{(l)})| \xrightarrow{\mathbb{P}} 0.$$

By (AN5) $\mathbf{M}^{(l)}(\boldsymbol{\vartheta}^{(l)}) = \mathbf{M}^{(l)}$ is nonsingular, so

$$\mathbb{P}\{\det \dot{\mathbf{S}}_{-in}^{(l)}(\boldsymbol{\zeta}_{-in}^{(l)}) \neq 0, \forall i = 1, \dots, n\} \rightarrow 1$$

and

$$\Lambda_n^l = \max_{i=1, \dots, n} |(\dot{\mathbf{S}}_{-in}^{(l)}(\boldsymbol{\zeta}_{-in}^{(l)}))^{-1}| = O_p(1). \quad (44)$$

So, with probability which tends to 1 as $n \rightarrow \infty$,

$$\begin{aligned} |\hat{\boldsymbol{\vartheta}}_{-in}^{(l)} - \hat{\boldsymbol{\vartheta}}_n^{(l)}| &= |(\dot{\mathbf{S}}_{-in}^{(l)}(\boldsymbol{\zeta}_{-in}^{(l)}))^{-1}(-\mathbf{S}_{-in}^{(l)}(\hat{\boldsymbol{\vartheta}}_n^{(l)}))| \leq \Lambda_n^l |\mathbf{S}_{-in}^{(l)}(\hat{\boldsymbol{\vartheta}}_n^{(l)})| \\ &= \Lambda_n^l |\mathbf{S}_{-in}^{(l)}(\boldsymbol{\vartheta}^{(l)}) + \dot{\mathbf{S}}_{-in}^{(l)}(\tilde{\boldsymbol{\zeta}}_{-in}^{(l)})(\hat{\boldsymbol{\vartheta}}_n^{(l)} - \boldsymbol{\vartheta}^{(l)})|, \end{aligned} \quad (45)$$

where $\tilde{\boldsymbol{\zeta}}_{-in}^{(l)}$ are some intermediate points between $\hat{\boldsymbol{\vartheta}}_n^{(l)}$ and $\boldsymbol{\vartheta}^{(l)}$. From (43) and Assumption (JK2) we obtain

$$\dot{\mathbf{S}}_{-in}^{(l)}(\tilde{\boldsymbol{\zeta}}_{ni}^{(l)})(\hat{\boldsymbol{\vartheta}}_n^{(l)} - \boldsymbol{\vartheta}^{(l)}) = O_P(n^{-1/2}). \quad (46)$$

Then

$$\mathbf{S}_{-in}^{(l)}(\boldsymbol{\vartheta}^{(l)}) = \mathbf{S}_n^{(l)}(\boldsymbol{\vartheta}^{(l)}) - a_{i;n}^l \mathbf{s}_{i;n}(\boldsymbol{\xi}_{i;n}, \boldsymbol{\vartheta}^{(l)}) - \sum_{j \neq i} (a_{j;n}^l - a_{j;-in}^l) \mathbf{s}_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\vartheta}^{(l)}).$$

Observe that $\mathbb{E} \mathbf{s}_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\vartheta}^{(l)}) = 0$ and

$$\mathbb{E} |\mathbf{S}_n^{(l)}(\boldsymbol{\vartheta}^{(l)})|^2 = \sum_{j=1}^n (a_{j;n}^{(l)})^2 \mathbb{E} |\mathbf{s}_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\vartheta}^{(l)})|^2 = O(n^{-1})$$

due to Lemma 1. So $\mathbf{S}_n^{(l)}(\boldsymbol{\vartheta}^{(l)}) = O_P(n^{-1/2})$.

By Lemmas 1, 2 and Assumption (JK1),

$$\max_{i=1, \dots, n} |a_{i;n}^l \mathbf{s}_{i;n}(\boldsymbol{\xi}_{i;n}, \boldsymbol{\vartheta}^{(l)}) + \sum_{j \neq i} (a_{j;n}^l - a_{j;-in}^l) \mathbf{s}_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\vartheta}^{(l)})| = O_P(n^{\beta-1}) \quad (47)$$

for any $\beta \geq 1/\alpha$. With $\beta = 1/2$ we get

$$\max_{i=1, \dots, n} |\mathbf{S}_{-in}^{(l)}(\boldsymbol{\vartheta}^{(l)})| = O_P(n^{-1/2}).$$

This with (44)–(46) yields

$$\max_{i=1, \dots, n} |\hat{\boldsymbol{\vartheta}}_{-in}^{(l)} - \hat{\boldsymbol{\vartheta}}_n^{(l)}| = O_P(n^{-1/2}). \quad (48)$$

Let $\mathbf{M}_{-in}^{(l)} = \dot{\mathbf{S}}_{-in}^{(l)}(\boldsymbol{\zeta}_{-in}^{(l)})$. By (40)–(42), we obtain

$$\begin{aligned} \hat{\boldsymbol{\vartheta}}_{-in}^{(l)} - \hat{\boldsymbol{\vartheta}}_n^{(l)} &= (\mathbf{M}_{-in}^{(l)})^{-1} (\mathbf{S}_n^{(l)}(\hat{\boldsymbol{\vartheta}}_n^{(l)}) - \mathbf{S}_{-in}^{(l)}(\hat{\boldsymbol{\vartheta}}_n^{(l)})) \\ &= (\mathbf{M}_{-in}^{(l)})^{-1} \left(a_{i;n}^l \mathbf{s}_{i;n}(\boldsymbol{\xi}_{i;n}, \hat{\boldsymbol{\vartheta}}_n^{(l)}) - \sum_{j \neq i} (a_{j;n}^l - a_{j;-in}^l) \mathbf{s}_{j;n}(\boldsymbol{\xi}_{j;n}, \hat{\boldsymbol{\vartheta}}_n^{(l)}) \right). \end{aligned} \quad (49)$$

Put

$$\mathbf{U}_i^{(l)} = a_{i;n}^l \mathbf{s}_{i;n}(\boldsymbol{\xi}_{i;n}, \boldsymbol{\vartheta}^{(l)}) + \sum_{j \neq i} (a_{j;n}^l - a_{j;-in}^l) \mathbf{s}_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\vartheta}^{(l)}), \quad (50)$$

$$\begin{aligned} \Delta_i^{(l)U} &= a_{i;n}^l (\mathbf{s}_{i;n}(\boldsymbol{\xi}_{i;n}, \hat{\boldsymbol{\vartheta}}_n^{(l)}) - \mathbf{s}_{i;n}(\boldsymbol{\xi}_{i;n}, \boldsymbol{\vartheta}^{(l)})) \\ &\quad + \sum_{j \neq i} (a_{j;n}^l - a_{j;-in}^l) (\mathbf{s}_{j;n}(\boldsymbol{\xi}_{j;n}, \hat{\boldsymbol{\vartheta}}_n^{(l)}) - \mathbf{s}_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\vartheta}^{(l)})), \end{aligned} \quad (51)$$

$$\Delta_i^{(l)M} = (\mathbf{M}_{-in}^{(l)})^{-1} - (\mathbf{M}^{(l)})^{-1}. \quad (52)$$

Then, by (49),

$$\hat{\boldsymbol{\theta}}_{-in}^{(l)} - \hat{\boldsymbol{\theta}}_n^{(l)} = ((\mathbf{M}^{(l)})^{-1} + \Delta_i^{(l)M})(\mathbf{U}_i^{(l)} + \Delta_i^{(l)U}).$$

So

$$\hat{\mathbf{V}}_n^{(k,m)} = n \sum_{i=1}^n (\hat{\boldsymbol{\theta}}_{-in}^{(k)} - \hat{\boldsymbol{\theta}}_n^{(k)})(\hat{\boldsymbol{\theta}}_{-in}^{(m)} - \hat{\boldsymbol{\theta}}_n^{(m)})^T = n \sum_{i=1}^n v_i, \quad (53)$$

where

$$\mathbf{v}_i = ((\mathbf{M}^{(k)})^{-1} + \Delta_i^{(k)M})(\mathbf{U}_i^{(k)} + \Delta_i^{(k)U})(\mathbf{U}_i^{(m)} + \Delta_i^{(m)U})^T ((\mathbf{M}^{(m)})^{-1} + \Delta_i^{(m)M})^{-T}. \quad (54)$$

Consider

$$\tilde{\mathbf{V}}_n^{(k,m)} = n \sum_{i=1}^n \tilde{\mathbf{v}}_i, \quad \tilde{\mathbf{v}}_i = (\mathbf{M}^{(k)})^{-1} \mathbf{U}_i^{(k)} (\mathbf{U}_i^{(m)})^T (\mathbf{M}^{(m)})^{-T}. \quad (55)$$

We will show that

$$|\hat{\mathbf{V}}_n^{(k,m)} - \tilde{\mathbf{V}}_n^{(k,m)}| \xrightarrow{P} 0, \text{ as } n \rightarrow \infty, \quad (56)$$

and

$$\tilde{\mathbf{V}}_n^{(k,m)} \xrightarrow{P} \mathbf{V}^{(k,m)}, \text{ as } n \rightarrow \infty. \quad (57)$$

Convergences (56) and (57) for all $k, m = 1, \dots, M$ imply the statement of the theorem.

To show (56) consider the following expansion

$$\mathbf{v}_i - \tilde{\mathbf{v}}_i = \mathbf{v}_i^1 + \mathbf{v}_i^2 + \mathbf{v}_i^3 + \mathbf{v}_i^4,$$

where

$$\begin{aligned} \mathbf{v}_i^1 &= \Delta_i^{(k)M} (\mathbf{U}_i^{(k)} + \Delta_i^{(k)U})(\mathbf{U}_i^{(m)} + \Delta_i^{(m)U})^T ((\mathbf{M}^{(m)})^{-1} + \Delta_i^{(m)M})^T, \\ \mathbf{v}_i^2 &= (\mathbf{M}^{(k)})^{-1} \Delta_i^{(k)U} (\mathbf{U}_i^{(m)} + \Delta_i^{(m)U})^T ((\mathbf{M}^{(m)})^{-1} + \Delta_i^{(m)M})^T, \\ \mathbf{v}_i^3 &= (\mathbf{M}^{(k)})^{-1} \mathbf{U}_i^{(k)} (\Delta_i^{(m)U})^T ((\mathbf{M}^{(m)})^{-1} + \Delta_i^{(m)M})^T, \\ \mathbf{v}_i^4 &= (\mathbf{M}^{(k)})^{-1} \mathbf{U}_i^{(k)} (\mathbf{U}_i^{(m)})^T ((\Delta_i^{(m)M})^T). \end{aligned} \quad (58)$$

Let us estimate each \mathbf{v}_i^l separately. At first we bound $\Delta_i^{(k)M}$.

Applying Lemma 4 to $\psi_{j;n}(\mathbf{x}, \boldsymbol{\gamma}) = \frac{\partial}{\partial \gamma^l} \dot{\mathbf{s}}_{j;n}(\mathbf{x}, \boldsymbol{\gamma})$, $l = 1, \dots, d$, by the same way as for $\mathbf{S}_{-in}^{(k)}$ one obtains

$$\max_{i=1, \dots, n} \sup_{\boldsymbol{\gamma} \in \Theta} \left| \frac{\partial}{\partial \gamma^l} \mathbf{S}_{-in}^{(k)}(\boldsymbol{\gamma}) \right| = O_P(1). \quad (59)$$

Then, by the Mean Value theorem and Assumption (JK2), we get

$$\max_{i=1, \dots, n} |\dot{\mathbf{S}}_{-in}^{(k)}(\boldsymbol{\zeta}_{-in}^{(k)}) - \dot{\mathbf{S}}_{-in}^{(k)}(\boldsymbol{\theta}^{(k)})| = O_P(1) O_P(n^{-1/2}) = O_P(n^{-1/2}). \quad (60)$$

Applying Lemmas 1 and 2 by the same way as in (47) to $\dot{\mathbf{S}}_{j;n}^{(l)}$ we obtain

$$\max_{i=1,\dots,n} |\dot{\mathbf{S}}_{-in}^{(k)}(\boldsymbol{\vartheta}^{(k)}) - \dot{\mathbf{S}}^{(k)}(\boldsymbol{\vartheta}^{(k)})| = O_P(n^{\beta-1}). \quad (61)$$

Variances of each entry of $\dot{\mathbf{S}}_n^{(k)}(\boldsymbol{\vartheta})$ can be estimated as $O(n^{-1})$, so

$$|\dot{\mathbf{S}}_n^{(k)}(\boldsymbol{\vartheta}^{(k)}) - \mathbf{M}^{(k)}| = O_P(n^{-1/2}). \quad (62)$$

Since $\mathbf{M}_{-in}^{(k)} = \dot{\mathbf{S}}_{-in}^{(k)}(\boldsymbol{\zeta}_{-in}^{(k)})$, formulas (60)–(62) yield

$$\max_{i=1,\dots,n} |\mathbf{M}_{-in}^{(k)} - \mathbf{M}^{(k)}| = O_P(n^{-1/2}),$$

So, due to Assumption (AN5),

$$\max_{i=1,\dots,n} |\Delta_i^{(k)M}| = O_P(n^{-1/2}). \quad (63)$$

By Lemmas 1 and 2 and Assumption (JK2),

$$\max_{i=1,\dots,n} |\mathbf{U}_i^{(k)}| = O_P(n^{\beta-1}). \quad (64)$$

Let us bound

$$\begin{aligned} \max_{i=1,\dots,n} |\Delta_i^{(k)U}| &\leq \max_{i=1,\dots,n} (|a_{i;n}^k| \cdot |\dot{\mathbf{S}}_{i;n}(\boldsymbol{\zeta}_i)| \cdot |\hat{\boldsymbol{\vartheta}}_n^{(k)} - \boldsymbol{\vartheta}^{(k)}| \\ &\quad + \sum_{j \neq i} |a_{j;n}^k - a_{j;-in}^k| \cdot |\dot{\mathbf{S}}_{i;n}(\boldsymbol{\zeta}_j)| \cdot |\hat{\boldsymbol{\vartheta}}_n^{(k)} - \boldsymbol{\vartheta}^{(k)}|. \end{aligned} \quad (65)$$

(here $\boldsymbol{\zeta}_j$ are some intermediate points between $\hat{\boldsymbol{\vartheta}}_n^{(k)}$ and $\boldsymbol{\vartheta}^{(k)}$).

By Lemma 2, $\max_{j=1,\dots,n} \sup_{\boldsymbol{\gamma} \in \Theta} |\dot{\mathbf{S}}_{j;n}(\boldsymbol{\gamma})| = O_P(n^\beta)$. So (65), Lemma 1 and (JK2) imply

$$\max_{i=1,\dots,n} |\Delta_i^{(k)U}| = O_P(n^{\beta-3/2}). \quad (66)$$

Now we bound \mathbf{v}_i^l defined by (58).

By (63), (64) and (66),

$$\begin{aligned} \max_{i=1,\dots,n} |\mathbf{v}_i^1| &\leq \max_{i=1,\dots,n} |\Delta_i^{(k)M}| \cdot (|\mathbf{U}_i^{(k)}| + |\Delta_i^{(k)U}|) \\ &\quad \cdot (|\mathbf{U}_i^{(m)}| + |\Delta_i^{(m)U}|) (|\mathbf{M}^{(m)-1}| + |\Delta_i^{(m)M}|) \\ &= O_P(n^{-1/2}) O_P(n^{\beta-1})^2 O_P(1) = O_P(n^{2\beta-5/2}). \end{aligned}$$

Similarly,

$$\max_{i=1,\dots,n} |\mathbf{v}_i^4| = O_P(n^{2\beta-5/2}).$$

For \mathbf{v}_i^2 (and, similarly, \mathbf{v}_i^3), we have

$$\begin{aligned} \max_{i=1, \dots, n} |\mathbf{v}_i^2| &\leq \max_{i=1, \dots, n} |(\mathbf{M}^{(k)})^{-1}| \cdot |\Delta_i^{(k)U}| \\ &\quad \cdot (|\mathbf{U}_i^{(m)}| + |\Delta_i^{(m)U}|)(|\mathbf{M}^{(m)}| + |\Delta_i^{(m)M}|) \\ &= O_P(n^{\beta-3/2})O_P(n^{\beta-1})O_P(1) = O_P(n^{2\beta-5/2}). \end{aligned}$$

Therefore

$$|\hat{\mathbf{V}}_n^{(m,k)} - \tilde{\mathbf{V}}_n^{(m,k)}| \leq n \sum_{i=1}^n |\mathbf{v}_i - \tilde{\mathbf{v}}_i| \leq n^2 \max_{i=1, \dots, n} \sum_{l=1}^4 |\mathbf{v}_i^l| = O_P(n^{2\beta-1/2}) = o_P(1)$$

for $1/\alpha \leq \beta < 1/4$. (Recall that we can take any $\beta \geq 1/\alpha$ and $\alpha > 4$).

So (56) holds. To prove the theorem we need only to verify (57).

Consider

$$\hat{\mathbf{Z}}_n^{(k,m)} = n \sum_{i=1}^n \mathbf{U}_i^{(k)} (\mathbf{U}_i^{(m)})^T.$$

Then $\mathbb{E} \hat{\mathbf{Z}}_n^{(k,m)} = \bar{\mathbf{Z}}_{1,n} + \bar{\mathbf{Z}}_{2,n}$, where

$$\bar{\mathbf{Z}}_{1,n} = n \sum_{j=1}^n a_{j;n}^k a_{j;n}^m \mathbb{E} \mathbf{s}_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\vartheta}^{(k)}) \mathbf{s}_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\vartheta}^{(m)}),$$

$$\bar{\mathbf{Z}}_{2,n} = n \sum_{j=1}^n \sum_{j \neq i} (a_{j;n}^k - a_{j;-in}^k)(a_{j;n}^m - a_{j;-in}^m) \mathbb{E} \mathbf{s}_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\vartheta}^{(k)}) \mathbf{s}_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\vartheta}^{(m)}).$$

By Lemma 1,

$$\max_{i=1, \dots, n} \left| \sum_{j \neq i} (a_{j;n}^k - a_{j;-in}^k)(a_{j;n}^m - a_{j;-in}^m) \mathbb{E} \mathbf{s}_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\vartheta}^{(k)}) \mathbf{s}_{j;n}(\boldsymbol{\xi}_{j;n}, \boldsymbol{\vartheta}^{(m)}) \right| = O(n^{-3}),$$

so $\bar{\mathbf{Z}}_{2,n} = n^2 O(n^{-3}) = O(n^{-1})$. This implies

$$\mathbb{E} \hat{\mathbf{Z}}_n^{(k,m)} \sim \bar{\mathbf{Z}}_{1,n} \rightarrow \mathbf{Z}^{(k,m)}. \quad (67)$$

Let us bound

$$\mathbb{E} |\hat{\mathbf{Z}}_n^{(k,m)} - \mathbb{E} \hat{\mathbf{Z}}_n^{(k,m)}|^2 \leq \sum_{p,q=1}^d \text{Var}(\hat{Z}_n^{pq(k,m)}) \quad (68)$$

(here and below $\hat{Z}_n^{pq(k,m)}$ is the (p, q) entry of the matrix $\hat{\mathbf{Z}}_n^{(k,m)}$, $U_i^{p(k)}$ is the p -th entry of the vector $\mathbf{U}^{(k)}$).

Consider

$$\begin{aligned} \text{Var}(\hat{Z}_n^{pq(k,m)}) &= n^2 \sum_{i=1}^n \text{Var}(U_i^{p(k)} U_i^{q(m)}) \leq n^3 \max_{i,p,q} \mathbb{E}(U_i^{p(k)} U_i^{q(m)})^2 \\ &\leq n^3 \sqrt{\max_{i,p,q} \mathbb{E}(U_i^{p(k)})^4 \mathbb{E}(U_i^{q(m)})^4} \leq n^3 \max_{i,p,l} \mathbb{E}(U_i^{p(l)})^4. \end{aligned} \quad (69)$$

(We applied the Cauchy–Schwarz inequality here.)

Let $\eta_j = s_{j;n}^p(\boldsymbol{\xi}_{j;n}, \boldsymbol{\vartheta}^{(k)})$, $b_{ij} = a_{j;n}^k - a_{j;-in}^k$. Then $\mathbf{E} \eta_i = 0$, so

$$\mathbf{E}(U_i^{p(k)})^4 = \mathbf{E}(a_i^k \eta_i + \sum_{j \neq i} b_{ij} \eta_j)^4 = J_{1;n} + J_{2;n} + J_{3;n}, \quad (70)$$

where

$$J_{1;n} = (a_{i;n}^k)^4 \mathbf{E}(\eta_i)^4, \quad J_{2;n} = 6(a_{i;n}^k)^2 \mathbf{E}(\eta_i)^2 \mathbf{E}(\sum_{j \neq i} b_{ij} \eta_j)^2,$$

$$J_{3;n} = \mathbf{E}(\sum_{j \neq i} b_{ij} \eta_j)^4.$$

By Assumption (JK1) and Lemma 1, $J_{1;n} = O(n^{-4})$,

$$J_{2;n} \leq \frac{C}{n^2} \sum_{j \neq i} (b_{ij})^2 \mathbf{E}(\eta_j)^2 = O(n^{-5}),$$

$$\begin{aligned} J_{3;n} &\leq \sum_{j \neq i} (b_{ij})^4 \mathbf{E}(\eta_j)^4 + 6 \sum_{j_1, j_2 \neq i} (b_{ij_1})^2 (b_{ij_2})^2 \mathbf{E}(\eta_{j_1})^2 \mathbf{E}(\eta_{j_2})^2 \\ &= O(n^{-7}) + n^2 O(n^{-8}) = O(n^{-6}). \end{aligned}$$

So, from (70) we obtain

$$\sup_{i,n} \mathbf{E}(U_i^{p(l)})^4 = O(n^{-4})$$

and (69), (68) yield

$$\mathbf{E} |\hat{Z}_n^{(k,m)} - \mathbf{E} \hat{Z}_n^{(k,m)}|^2 = O(n^{-4})n^3 = O(n^{-1}).$$

This with (67) imply (57). □

References

- [1] Efron, B., Stein, C.: The jackknife estimate of variance. *Ann. Appl. Stat.* **9**, 586–596 (1981). [MR0615434](#)
- [2] Grün, B., Leisch, F.: Fitting finite mixtures of linear regression models with varying & fixed effects in R. In: Rizzi, A., Vichi, M. (eds.) *Compstat 2006 – Proceedings in Computational Statistics*, pp. 853–860. Physica Verlag, Heidelberg, Germany (2006). [MR2173118](#)
- [3] Maiboroda, R., Sugakova, O.: Statistics of mixtures with varying concentrations with application to DNA microarray data analysis. *J. Nonparametr. Stat.* **24**(1), 201–205 (2012). [MR2885834](#). <https://doi.org/10.1080/10485252.2011.630076>
- [4] Maiboroda, R., Sugakova, O., Doronin, A.: Generalized estimating equations for mixtures with varying concentrations. *Can. J. Stat.* **41**(2), 217–236 (2013). [MR3061876](#). <https://doi.org/10.1002/cjs.11170>
- [5] Maiboroda, R., Sugakova, O.: Jackknife covariance matrix estimation for observations from mixture. *Mod. Stoch. Theory Appl.* **6**(4), 495–513 (2019). [MR4047396](#). <https://doi.org/10.15559/19-vmsta145>

- [6] Masiuk, S., Kukush, A., Shklyar, S., Chepurny, M., Likhtarov, I.: Radiation Risk Estimation: Based on Measurement Error Models. De Gruyter, Berlin, (2017). [MR3726857](#)
- [7] Miroschnichenko, V., Maiboroda, R.: Asymptotic normality of modified LS estimator for mixture of nonlinear regressions. *Mod. Stoch. Theory Appl.* **7**(4), 435–448 (2020). [MR4195645](#)
- [8] Miroschnychenko, V.O.: Generalized least squares estimates for mixture of nonlinear regressions. *Bulletin of Taras Shevchenko National University of Kyiv Series: Physics & Mathematics* **5**, 25–29 (2019)
- [9] Shao, J.: Consistency of least-squares estimator and its jackknife variance estimator in nonlinear models. *Can. J. Stat.* **20**(4), 415–428 (1992). [MR1208353](#). <https://doi.org/10.2307/3315611>
- [10] Shao, J.: *Mathematical statistics*. Springer, (2003). [MR2002723](#). <https://doi.org/10.1007/b97553>
- [11] Shao, J., Tu, D.: *The Jackknife and Bootstrap*. Springer, (2012). [MR1351010](#). <https://doi.org/10.1007/978-1-4612-0795-5>
- [12] Wang, L., Yu, F. Jackknife resample method for precision estimation of weighted total least squares. *Commun. Stat., Simul. Comput.* (2019). [MR4253814](#). <https://doi.org/10.1080/03610918.2019.1580727>
- [13] Wang, L., Yu, F., Li, Z., Zou, C. Jackknife method for variance components estimation of partial EIV model. *J. Surv. Eng.* **146**(4), (2020). [https://doi.org/10.1061/\(ASCE\)SU.1943-5428.0000327](https://doi.org/10.1061/(ASCE)SU.1943-5428.0000327)