Jaromír Kudeláš

The linear model with variance-covariance components and jackknife estimation

Applications of Mathematics, Vol. 39 (1994), No. 2, 111–125

Persistent URL: `http://dml.cz/dmlcz/134248`

THE LINEAR MODEL WITH VARIANCE-COVARIANCE COMPONENTS AND JACKKNIFE ESTIMATION

JAROMÍR KUDELÁŠ, Bratislava

(Received November 20, 1991)

Summary. Let θ^* be a biased estimate of the parameter ϑ based on all observations $x_1, \ldots, x_n$ and let θ^*_{-i} $(i = 1, 2, \ldots, n)$ be the same estimate of the parameter ϑ obtained after deletion of the i-th observation. If the expectation of the estimators θ^* and θ^*_{-i} are expressed as

$$\begin{aligned} \mathrm{E}(\theta^*) &= \vartheta + a(n)b(\vartheta) \\ \mathrm{E}(\theta^*_{-i}) &= \vartheta + a(n-1)b(\vartheta) \qquad i = 1, 2, \ldots, n, \end{aligned}$$

where $a(n)$ is a known sequence of real numbers and $b(\vartheta)$ is a function of ϑ, then this system of equations can be regarded as a linear model. The least squares method gives the generalized jackknife estimator. Using this method, it is possible to obtain the unbiased estimator of the parameter ϑ.

Keywords: Jackknife estimator, least squares estimator, linear model, estimator of variance-covariance components, consistency

AMS classification: 62F10

1. INTRODUCTION

The jackknife method was originally proposed by Quenouille (1949), and later expanded by the same author in 1956. Let ϑ be an unknown parameter, and let $X_1, \ldots, X_n$ be a sample of n independent identically distributed observations from the probability distribution P_ϑ. The essence of the jackknife method is to divide the n observations into k groups of size l $(n \doteq kl)$. Let ϑ^*_{-0} be an estimate of the parameter ϑ based on all n observations, and let ϑ^*_{-i}, $i = 1, \ldots, k$ denote the estimate of ϑ obtained after deletion of the i-th group of observations. ϑ^*_{-i} is the estimate of ϑ from the remaining $(k-1)l$ observations. Let

$$\hat{\vartheta}_i = k\vartheta^*_{-0} - (k-1)\vartheta^*_{-i} \qquad i = 1, \ldots, k. \tag{1.1}$$

These quantities are called pseudo-values by Tukey [8]. The jackknife estimate of ϑ is the average of the $\hat{\vartheta}_i$, $i = 1, \ldots, k$:

$$\hat{\vartheta} = \frac{1}{k}\sum_{i=1}^{k} \hat{\vartheta}_i = k\vartheta^*_{-0} - \frac{k-1}{k}\sum_{i=1}^{k} \vartheta^*_{-i}. \tag{1.2}$$

The jackknife eliminates exactly the n^{-1} bias term. This result holds for all values of k if

$$\mathrm{E}(\vartheta^*_{-0}) = \vartheta + a(kl)^{-1} + b(kl)^{-2} + \ldots, \tag{1.3}$$

where a, b are real constants.

One can show that

$$\mathrm{E}(\hat{\vartheta}) = \vartheta - b\big(k(k-1)l^2\big)^{-1} + \ldots\ . \tag{1.4}$$

Tukey has proposed that in many instances $\hat{\vartheta}_i$ are approximately independently, identically distributed. If this assumption is correct, then

$$k(k-1)\sum_{i=1}^{k}(\hat{\vartheta}_i - \hat{\vartheta})^2 \tag{1.5}$$

should be an approximate estimate of $\mathrm{Var}(\hat{\vartheta})$ and

$$(\hat{\vartheta} - \vartheta)\Big\{k(k-1)\sum_{i=1}^{k}(\hat{\vartheta}_i - \hat{\vartheta})^2\Big\}^{-\frac{1}{2}} \tag{1.6}$$

should be approximately distributed as a Student t-random variable with $k-1$ degrees of freedom.

2. System of linear equations and unbiased estimate

The object of our investigation is the random variable $\hat{\theta} = \hat{\theta}(x_1, \ldots, x_n)$ independent of $\vartheta \in \Theta \subset \mathbf{R}^1$ such that

$$\hat{\theta}(\cdot): (\mathbf{R}^n, \mathscr{B}^n) \to (\mathbf{R}^1, \mathscr{B}^1)$$

where $X_1, \ldots, X_n$ is a random sample from the probability distribution P_ϑ, $\vartheta \in \Theta \subset \mathbf{R}^1$, and ϑ is an unknown parameter.

We will denote a biased estimator by θ^* and the unbiased one by $\hat{\theta}$. Let $\mathscr{U}$ denote the class of unbiased estimators of the parameter ϑ with the property $\forall \hat{\theta} \in \mathscr{U}$; $\operatorname{Var}(\hat{\theta}) < \infty$, and let $\mathscr{T}$ denote the class of biased estimators of the parameter ϑ with the same property $\forall \theta^* \in \Theta \subset \mathbf{R}^1$; $\operatorname{Var}(\theta^*) < \infty$. Let $\mathscr{U} \neq \emptyset$ and $\mathscr{T} \neq \emptyset$. We assume that it is possible to get a biased estimator $\theta^* \in \mathscr{T}$ only, for which the equality

$$\mathrm{E}(\theta^*) = \vartheta + a(n)b(\vartheta) \tag{2.1}$$

holds, where $a(n)$ is a known sequence of real numbers, $b(\vartheta)$ is a function defined in ϑ, and $\operatorname{Var}(\theta^*) < \infty$.

Let there exist two estimators $\theta_1^*, \theta_2^* \in \mathscr{T}$ which are uncorrelated,

$$\begin{aligned} \mathrm{E}(\theta_1^*) &= \vartheta + a(n)b(\vartheta) \\ \mathrm{E}(\theta_2^*) &= \vartheta + c(n)b(\vartheta) \end{aligned} \tag{2.2}$$

where $a(n) \neq c(n)$.

Therefore we can look at (2.2) as at a system of linear equations with variables ϑ and $b(\vartheta)$. The solution of (2.2) is

$$\vartheta = \frac{\begin{vmatrix} \mathrm{E}(\theta_1^*) & a(n) \\ \mathrm{E}(\theta_2^*) & c(n) \end{vmatrix}}{\begin{vmatrix} 1 & a(n) \\ 1 & c(n) \end{vmatrix}}, \qquad b(\vartheta) = \frac{\begin{vmatrix} 1 & \mathrm{E}(\theta_1^*) \\ 1 & \mathrm{E}(\theta_2^*) \end{vmatrix}}{\begin{vmatrix} 1 & a(n) \\ 1 & c(n) \end{vmatrix}}. \tag{2.3}$$

After removing the expectation from the result (2.3) we obtain the unbiased estimator of the parameter ϑ and the unbiased estimator of the function $b(\vartheta)$:

$$\hat{\theta} = \left[c(n) - a(n)\right]^{-1}\left[c(n)\theta_1^* - a(n)\theta_2^*\right], \tag{2.4}$$

$$\hat{b}(\vartheta) = \left[c(n) - a(n)\right]^{-1}(\theta_2^* - \theta_1^*). \tag{2.5}$$

Obviously, $\hat{\theta} \in \mathscr{U}$.

This result holds for the following case, too:

$$\mathrm{E}(\theta_i^*) = \vartheta + \sum_{j=1}^{s} a_{ij}(n) b_j(\vartheta) \qquad i = 1, 2, \ldots, s+1$$

where $\theta_i^* \in \mathcal{T}$. The sequences $\{a_{ij}(n)\}$ are known for $i = 1, \ldots, s+1$; $j = 1, \ldots, s$. Let $a_{ij}(n) \neq a_{kj}(n)$ for $k \neq i$. Then the estimator

$$\text{(2.6)} \qquad T = \frac{\begin{vmatrix} \theta_1^* & a_{11}(n) & \ldots & a_{1s}(n) \\ \theta_2^* & a_{21}(n) & \ldots & a_{2s}(n) \\ \ldots & \ldots & \ldots & \ldots \\ \theta_{s+1}^* & a_{s+1,1}(n) & \ldots & a_{s+1,s}(n) \end{vmatrix}}{\begin{vmatrix} 1 & a_{11}(n) & \ldots & a_{1s}(n) \\ 1 & a_{21}(n) & \ldots & a_{2s}(n) \\ \ldots & \ldots & \ldots & \ldots \\ 1 & a_{s+1,1}(n) & \ldots & a_{s+1,s}(n) \end{vmatrix}}$$

is unbiased.

In the case that the number of estimators m is less than $s+1$ ($m < s+1$), the estimator (2.6) eliminates only $m-1$ terms from the bias of the biased estimators θ_i^*, $i = 1, 2, \ldots, s+1$.

How to obtain so many estimators with the same bias, this is the question. Let θ_{-i}^*, $i = 1, \ldots, n$ denote the estimates obtained after deletion of the ith observation, and let θ_{-ij}^* denote the estimates obtained after deletions of the ith and jth observations. θ_{-ijk}^*, θ_{-ijkl}^*, ... are obtained similarly.

Now, taking into account (2.4) and the above facts, let us construct the following estimator and denote it $\hat{\theta}_i$:

$$\text{(2.7)} \qquad \hat{\theta}_i = \left[a(n-1) - a(n)\right]^{-1}\left[a(n-1)\theta_{-0}^* - a(n)\theta_{-i}^*\right]$$

$i = 1, 2, \ldots, n$, where $a(n) \neq a(n-1)$ and θ_{-0}^* is the estimator of ϑ based on all observations. Obviously, $\hat{\theta}_i \in \mathcal{U}$ if $\theta_{-0}^* \in \mathcal{T}$. We will call the estimator (2.7) a pseudo-value. Obviously, the number of pseudo-values is n. We define the jackknife estimator (JE) as the average of the pseudo-values:

$$\text{(2.8)} \qquad \hat{\theta} = n^{-1}\sum_{i=1}^{n} \hat{\theta}_i = \left[a(n-1) - a(n)\right]^{-1}\left[a(n-1)\theta_{-0}^* - a(n)n^{-1}\sum_{i=1}^{n}\theta_{-i}^*\right].$$

This estimator is a generalized jackknife estimator from (1.2).

It is easy to show that $\mathrm{Var}(\hat{\theta}) < \mathrm{Var}(\hat{\theta}_i)$.

Theorem 1. *Let the sequence $a(n)$ satisfy the condition*

$$\frac{a(n)}{a(n-1)-a(n)} = O(n) \tag{2.9}$$

an let $\lim_{n\to\infty} n(\theta^*_{-0} - \vartheta) = k$ hold with probability 1, where k is an arbitrary real constant. Then the jackknife estimator given by (2.8) converges to the parameter ϑ for $n \to \infty$ (where n denotes the size of the random sample) with probability 1.

Remark 1. For instance, the sequence $a(n) = \frac{R(n)}{Q(n)}$, where $R(x)$ is a polynomial of degree l_1, $Q(x)$ is a polynomial of degree l_2 and $l_1 < l_2$, has the property (2.9).

Proof of Theorem 1. The condition $P\left[\lim_{n\to\infty} n(\theta^*_{-0} - \vartheta) = k\right] = 1$ implies

$$P\left[\lim_{n\to\infty} n\left(n^{-1}\sum_{i=1}^{n}\theta^*_{-i} - \vartheta\right) = k\right] = 1.$$

Let A and B denote the sets

$$A = \left\{\omega;\ \forall\varepsilon > 0\ \exists n_0,\ \forall n > n_0\colon \left|n(\theta^*_{-0} - \vartheta) - k\right| < \frac{\varepsilon}{2}\right\},$$
$$B = \left\{\omega;\ \forall\varepsilon > 0\ \exists n_1,\ \forall n > n_1\colon \left|n\left(n^{-1}\sum_{i=1}^{n}\theta^*_{-i} - \vartheta\right) - k\right| < \frac{\varepsilon}{2}\right\}$$

and let A^c, B^c denote their complements. $P(A) = P(B) = 1$.

It is easy to see that

$$\begin{aligned}
1 &= 1 - P(A^c) - P(B^c) \leqslant 1 - P(A^c \cup B^c) = P(A \cap B) \\
&\leqslant P\Big(\Big\{\omega;\ \forall\varepsilon > 0\ \exists n_2 = \max(n_0, n_1), \\
&\qquad \forall n > n_2\colon \left|n(\theta^*_{-0} - \vartheta) - k\right| + \left|n(n^{-1}\sum_{i=1}^{n}\theta^*_{-i} - \vartheta) - k\right| < \varepsilon\Big\}\Big) \\
&\leqslant P\Big(\Big\{\omega;\ \forall\varepsilon > 0\ \exists n_2,\ \forall n > n_2\colon \left|n(\theta^*_{-0} - \vartheta) - n\left(n^{-1}\sum_{i=1}^{n}\theta^*_{-i} - \vartheta\right)\right| < \varepsilon\Big\}\Big) \\
&= P\Big(\Big\{\omega;\ \forall\varepsilon > 0\ \exists n_2,\ \forall n > n_2\colon \left|n\left(\theta^*_{-0} - n^{-1}\sum_{i=1}^{n}\theta^*_{-i}\right)\right| < \varepsilon\Big\}\Big).
\end{aligned}$$

Hence

$$P\left[\lim_{n\to\infty} n\left(\theta^*_{-0} - n^{-1}\sum_{i=1}^{n}\theta^*_{-i}\right) = 0\right] = 1. \tag{2.10}$$

Now we can express the JE (2.8) in the form,

$$\hat{\theta} = \theta^*_{-0} + \frac{a(n)}{a(n-1) - a(n)} \left(\theta^*_{-0} - n^{-1} \sum_{i=1}^{n} \theta^*_{-i} \right). \tag{2.11}$$

Hence the assumptions and the property (2.9) prove the theorem. □

3. Unbiased estimator obtained from the least square estimator

Let us return to the system of linear equations (2.2) at the beginning of Section 2. (2.2) can be regarded as the linear model

$$\begin{aligned} \mathrm{E}(Y) &= \mathbf{X}\beta \\ \mathrm{Var}(Y) &= \mathbf{U} \end{aligned}$$

where $Y' = (\theta^*_1, \theta^*_2)$, $\beta' = (\vartheta, b(\vartheta))$, $\mathbf{X} = \begin{pmatrix} 1 & a(n) \\ 1 & c(n) \end{pmatrix}$ and $\mathrm{Var}(\theta^*_1) > 0$, $\mathrm{Var}(\theta^*_2) > 0$. The matrix $\mathbf{U}$ is nonsingular, because θ^*_1, θ^*_2 are uncorrelated. The least square estimator (LSE) of β is

$$\hat{\beta} = \left[c(n) - a(n)\right]^{-1} \begin{pmatrix} c(n)\theta^*_1 - a(n)\theta^*_2 \\ \theta^*_2 - \theta^*_1 \end{pmatrix}.$$

It is the same result as in (2.4), (2.5). If we substitute $c(n) = a(n-1)$, we obtain the pseudo-values from (2.7).

Now it will be shown that the jackknife estimator (2.8) is the best linear unbiased estimator obtained as a linear combination of the estimators θ^*_{-i}, $i = 0, 1, 2, \ldots, n$.

Let us consider a linear model

$$\begin{aligned} \mathrm{E}(Y) &= \mathbf{A}\beta \\ \mathrm{Var}(Y) &= \mathbf{V} \end{aligned} \tag{3.1}$$

where $Y' = (\theta^*_{-0}, \theta^*_{-1}, \ldots, \theta^*_{-n}) \in \mathbf{R}^{n+1}$, $\beta' = (\vartheta, b(\vartheta))$, $\mathbf{A}$ is the $(n+1) \times 2$ matrix

$$\mathbf{A} = \begin{pmatrix} 1 & a(n) \\ 1 & a(n-1) \\ \multicolumn{2}{c}{\cdots\cdots\cdots} \\ 1 & a(n-1) \end{pmatrix} \qquad a(n) \neq a(n-1)$$

and $\mathbf{V}$ is a square symmetric matrix of order $(n+1)$

$$\mathbf{V} = \begin{pmatrix} p & v & v & v & \dots & v \\ & r & z & z & \dots & z \\ & & r & z & \dots & z \\ & & & \dots & \dots & \dots \\ & & & & \ddots & z \\ & & & & & r \end{pmatrix},$$

where $p = \operatorname{Var}(\theta^*_{-0}) < \infty$, $r = \operatorname{Var}(\theta^*_{-i}) < \infty$, $v = \operatorname{cov}(\theta^*_{-0}, \theta^*_{-i})$, $z = \operatorname{cov}(\theta^*_{-i}, \theta^*_{-j})$, $i, j = 1, 2, \dots, n$.

Theorem 2. *The Gauss-Markov estimator of the parameter ϑ in the linear model (3.1) is the jackknife estimator (2.8):*

$$\hat{\theta} = \Big[a(n-1) - a(n)\Big]^{-1} \Big[a(n-1)\theta^*_{-0} - a(n)n \sum_{i=1}^{n} \theta^*_{-i}\Big]. \tag{3.2}$$

P r o o f. When the covariance matrix of Y is nonsingular then the proof involves few difficulties. The inverse $\mathbf{V}^{-1}$ of the matrix $\mathbf{V}$ is

$$\mathbf{V}^{-1} = m^{-1} \begin{pmatrix} \big[r + (n-1)z\big](r-z) & -v(r-z) & -v(r-z) & \dots \\ & \begin{matrix} pr + (n-2)pz \\ -(n-1)v^2 \end{matrix} & v^2 - pz & \dots \\ & & \begin{matrix} pr + (n-2)pz \\ -(n-1)v^2 \end{matrix} & \dots \\ & & & \ddots \end{pmatrix}$$

where $m = \big[pr + (n-1)pz - nv^2\big](r-z)$.

In the case the equation,

$$\hat{\beta} = (\mathbf{A}^\bullet \mathbf{V}^{-1} \mathbf{A})^{-1} \mathbf{A}^\bullet \mathbf{V}^{-1} Y \tag{3.3}$$

holds.

Substituting $\mathbf{V}^{-1}$ into (3.3), the result can be obtained immediately. □

R e m a r k 2. The previous theorem is proved in [2] for the case in which the covariance matrix is singular. Lavergne and Mathieu used the statement proved by Kruskal in [1]:

$$(\mathbf{A}^\bullet \mathbf{V}^{-1} \mathbf{A})^{-1} \mathbf{A}^\bullet \mathbf{V}^{-1} Y = (\mathbf{A}^\bullet \mathbf{A})^{-1} \mathbf{A}^\bullet Y \Leftrightarrow \mu(\mathbf{VA}) \subseteq \mu(\mathbf{A}), \tag{3.4}$$

where $\mu(\mathbf{A})$ denotes the subspace of $\mathbf{R}^{n+1}$ spanned by the columns of the matrix $\mathbf{A}$. This equation holds even if $\mathbf{V}$ is singular.

Let us investigate the situation when the covariance matrix $\mathbf{V}$ is nonsingular. We have

$$\begin{vmatrix} p & W' \\ W & \mathbf{D} \end{vmatrix} = |\mathbf{D}||p - W'\mathbf{D}^{-1}W|,$$

where the vector W is given by $W' = (v, v, \ldots, v)$ and $\mathbf{D}$ is the $(n \times n)$ symmetric matrix of the form

$$\mathbf{D} = \begin{pmatrix} r & z & z & \ldots & z \\ & r & z & \ldots & z \\ & & \cdots\cdots & & \\ & & & \ddots & z \\ & & & & r \end{pmatrix}.$$

We get the following conditions for nonsingularity of the matrix $\mathbf{V}$: $r \neq z$, $r \neq (1-n)z$ and $p \neq nv^2\left[r + (n-1)z\right]^{-1}$.

Corollary 1. *The jackknife estimator (JE) from* (3.2) *is the best of all estimators which are obtained from the linear combination of the estimators* θ^*_{-i}, $i = 0, 1, 2, \ldots, n$.

Corollary 2. *Let* S^* *and* T^* *be the estimators of the parameter* ϑ_1 *and* ϑ_2, *respectively. Let*

$$\mathrm{E}(S^*) = \vartheta_1 + a(n)b(\vartheta_1)$$
$$\mathrm{E}(T^*) = \vartheta_2 + a(n)b(\vartheta_2).$$

Then $\hat{S} + \hat{T}$ *is the JE of the parameter* $\vartheta_1 + \vartheta_2$, *where* $\hat{S}$ *and* $\hat{T}$ *are the JE of the parameter* ϑ_1 *and* ϑ_2, *respectively.*

Corollary 3. *The JE is independent of the covariance matrix* $\mathbf{V}$ *of the vector* $Y' = (\theta^*_{-0}, \theta^*_{-1}, \ldots, \theta^*_{-n})$.

Corollary 4. *Let the random vector* Y *be normally distributed with the mean* $\mathbf{A}\beta$ *and a nonsingular covariance matrix* $\mathbf{V}$. *Denote* $S^2 = (Y - \mathbf{A}\hat{\beta})'\mathbf{V}^{-1}(Y - \mathbf{A}\hat{\beta})$. *Then the random variable*

$$\frac{\hat{\theta} - \vartheta}{\left[\frac{S^2 \operatorname{Var}(\hat{\theta})}{n+1}\right]^{\frac{1}{2}}} \tag{3.5}$$

is distributed as a t-random variable with $n + 1$ *degrees of freedom.*

P r o o f. The assumption yields

$$\hat{\beta} \sim N(\beta, (\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}).$$

Since $(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{V}\mathbf{V}^{-1}[\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'] = 0$, it may be asserted that $\hat{\beta}$ and S^2 are independent. Obviously $S^2 \sim X^2_{n+1}$. This completes the proof. □

R e m a r k 3. The following assertion holds:

$$\mu(\mathbf{VA}) \subseteq \mu(\mathbf{A}) \Leftrightarrow \mathrm{Ker}(\mathbf{A}') \subseteq \mathrm{Ker}(\mathbf{A}'\mathbf{V})$$

where $\mu(\mathbf{A})$ denotes the subspace of $\mathbf{R}^{n+1}$ spanned by the columns of the matrix $\mathbf{A}$, and $\mathrm{Ker}(\mathbf{A}')$ denotes the null space of $\mathbf{A}'$. Then

$$\mathbf{A}'[\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'] = 0 \Rightarrow \mathbf{A}'\mathbf{V}[\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'] = 0.$$

It is easy to see that

$$(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{V}[\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}']\mathbf{V}^{-}[\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'] = 0.$$

Therefore $\hat{\beta}$ and S^2 are independent also in the case when the matrix $\mathbf{V}$ is singular, and then the following result holds.

The random variable

$$\frac{\hat{\theta} - \vartheta}{\left[\frac{S^2 \,\mathrm{Var}(\hat{\theta})}{k}\right]^{\frac{1}{2}}}$$

is distributed as a t-random variable with k degrees of freedom, where $k = R(\mathbf{V})$ is the rank of the covariance matrix $\mathbf{V}$, under the normality condition of the vector Y.

R e m a r k 4. Quenouille's JE (see (1.2)) is a particular case of the estimator (3.2).

R e m a r k 5. If the bias of the estimator has more terms, i.e.

$$\mathrm{E}(\theta^*) = \vartheta + \sum_{i=1}^{p} a_i(n) b_i(\vartheta),$$

then the JE can be obtained analogously.

4. Estimator of covariance components in the linear model (3.1)

Let us express the covariance matrix $\mathbf{V}$ of the vector Y as a decomposition of the form

$$\mathbf{V} = p\mathbf{V}_1 + r\mathbf{V}_2 + v\mathbf{V}_3 + z\mathbf{V}_4 \tag{4.1}$$

where p, r, v, z are unknown parameters and the matrices $\mathbf{V}_i$, $i = 1,2,3,4$ are given by

$$\mathbf{V}_1 = \begin{pmatrix} 1 & 0' \\ 0 & \mathbf{O} \end{pmatrix}, \quad \mathbf{V}_2 = \begin{pmatrix} 0 & 0' \\ 0 & \mathbf{I}_n \end{pmatrix}, \quad \mathbf{V}_3 = \begin{pmatrix} 0 & e' \\ e & \mathbf{O} \end{pmatrix}, \quad \mathbf{V}_4 = \begin{pmatrix} 0 & 0' \\ 0 & ee' - \mathbf{I}_n \end{pmatrix},$$

where $e' = (1, \dots, 1)$ is a vector of order n, $\mathbf{I}_n$ is the identity matrix of order n. The order of the matrices (4.1) is $(n+1) \times (n+1)$.

Thus the model is expressed as a linear model with variance-covariance components (see Štulajter in [7]).

Let us have an arbitrary linear model with variance-covariance components:

$$\begin{aligned} &\mathrm{E}(Y) = \mathbf{X}\beta \\ &\mathrm{Var}(Y) = = \mathbf{V} = \sum_{i=1}^{k} \nu_i \mathbf{V}_i, \end{aligned} \tag{4.2}$$

where ν_i, $i = 1,2,\dots,k$, are unknown parameters and $\mathbf{V}_i$, $i = 1,2,\dots,k$, are known matrices.

Let us use the Natural Least Squares Estimator (NLSE) of the covariance parameters ν_i, $i = 1,2,\dots,k$ given by Štulajter in [7]:

$$\hat{\nu}_i = \frac{1}{\|\mathbf{V}_i\|^2} \operatorname{tr}(\tilde{\mathbf{V}}\mathbf{V}_i) \tag{4.3}$$

if the matrices $\mathbf{V}_1$, ..., $\mathbf{V}_k$ are mutually orthogonal. $\|\mathbf{C}\|$ denotes the Euclidean norm of a matrix $\mathbf{C}$ generated by the inner product

$$(\mathbf{C}, \mathbf{B}) = \operatorname{tr}(\mathbf{C}\mathbf{B}') = \sum_{i,j=1}^{n} \mathbf{C}_{ij}\mathbf{B}_{ij},$$

where $\mathbf{C}_{ij}$ denotes the (i,j)-th element of the matrix $\mathbf{C}$.

$\tilde{\mathbf{V}} = (Y - \mathbf{X}\hat{\beta})(Y - \mathbf{X}\hat{\beta})'$, where $\hat{\beta}$ is the best linear unbiased estimator of the parameter β in the linear model (4.2).

In our particular case the form of the estimators is as follows:

$$\hat{p} = 0$$
$$\hat{v} = 0$$
$$\hat{r} = \frac{1}{n}\sum_{i=1}^{n}(\theta^*_{-i} - \bar{\theta}^*)^2 \tag{4.4}$$
$$\hat{z} = \frac{2}{n(n-1)}\sum_{i<j}^{n}(\theta^*_{-i} - \bar{\theta}^*)(\theta^*_{-j} - \bar{\theta}^*)$$
$$= \frac{1}{n(n-1)}\sum_{i=1}^{n}(\theta^*_{-i} - \bar{\theta}^*)^2, \tag{4.5}$$

where $\bar{\theta}^* = \frac{1}{n}\sum_{i=1}^{n}\theta^*_{-i}$ and $p = \mathrm{Var}(\theta^*_{-0})$, $r = \mathrm{Var}(\theta^*_{-1})$, $v = \mathrm{cov}(\theta^*_{-0}, \theta^*_{-1})$, $z = \mathrm{cov}(\theta^*_{-1}, \theta^*_{-2})$.

It is the logical conclusion that $\hat{p} = 0$, $\hat{v} = 0$, because we have only one estimate θ^*_{-0} and the moment of the 2nd order cannot be calculated from one realization of θ^*_{-0}.

Theorem 3. *Let $p = \mathrm{Var}(\theta^*_{-0})$, $r = \mathrm{Var}(\theta^*_{-1})$, $v = \mathrm{cov}(\theta^*_{-0}, \theta^*_{-1})$ and $z = \mathrm{cov}(\theta^*_{-1}, \theta^*_{-2})$ be the finite second moments for all n. If z converges to zero for $n \to \infty$, then the estimator $\hat{r}$ is an asymptotically unbiased estimator of the parameter r.*

Proof. It is clear that

$$\mathrm{E}(\hat{r}) = \mathrm{E}\Big\{\frac{1}{n}\sum_{i=1}^{n}\big[(\theta^*_{-i} - \mathrm{E}\theta^*_{-i}) - (\bar{\theta}^* - \mathrm{E}\bar{\theta}^*)\big]^2\Big\} =$$
$$= \frac{1}{n}\sum_{i=1}^{n}\big[\mathrm{Var}(\theta^*_{-i}) + \mathrm{Var}(\bar{\theta}^*) - 2\,\mathrm{cov}(\bar{\theta}^*, \theta^*_{-i})\big] =$$
$$= r - \mathrm{Var}(\bar{\theta}^*) = r - \frac{1}{n}r - \frac{n-1}{n}z.$$

□

Theorem 4. *If the random vector $Y' = (\theta^*_{-0}, \theta^*_{-1}, \ldots, \theta^*_{-n})$ is normally distributed and if z converges to zero for $n \to \infty$, then the estimator $\hat{r}$ converges to r in quadratic mean.*

P r o o f. First, let us think about the form of the matrix $\tilde{\mathbf{V}}$:

$$\tilde{\mathbf{V}} = \begin{pmatrix} 0 & 0 & 0 & \dots & \\ & (\theta^*_{-1} - \bar{\theta}^*)^2 & (\theta^*_{-1} - \bar{\theta}^*)(\theta^*_{-2} - \bar{\theta}^*) & \dots & (\theta^*_{-1} - \bar{\theta}^*)(\theta^*_{-n} - \bar{\theta}^*) \\ & & (\theta^*_{-2} - \bar{\theta}^*)^2 & \dots & (\theta^*_{-2} - \bar{\theta}^*)(\theta^*_{-n} - \bar{\theta}^*) \\ & & & & \dots \\ & & & & (\theta^*_{-n} - \bar{\theta}^*)^2 \end{pmatrix}$$

Let us denote $\mathbf{P} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$, $\mathbf{M} = \mathbf{I} - \mathbf{P}$. Then $\tilde{\mathbf{V}} = \mathbf{M}Y(\mathbf{M}Y)'$ and

$$\hat{r} = \frac{1}{\|\mathbf{V}_2\|^2} \operatorname{tr}(\tilde{\mathbf{V}}\mathbf{V}_2) = \frac{1}{n} \operatorname{tr}(\tilde{\mathbf{V}}) = \frac{1}{n} Y'\mathbf{M}Y$$

because $\tilde{\mathbf{V}}\mathbf{V}_2 = \tilde{\mathbf{V}}$.

Further,

$$\operatorname{Var}(\hat{r}) = \frac{1}{n^2} \operatorname{Var}(Y'\mathbf{M}Y) = \frac{1}{n^2} 2 \operatorname{tr}(\mathbf{MVMV})$$

because Y is normally distributed.

By virtue of the equation

$$\mathbf{MV} = -\frac{1}{n} \begin{pmatrix} 0 & 0 & 0 & 0 & \dots & 0 \\ & 1-n & 1 & 1 & \dots & 1 \\ & & 1-n & 1 & \dots & 1 \\ & & & \dots & \dots & \dots \\ & & & & \ddots & 1 \\ & & & & & 1-n \end{pmatrix} \begin{pmatrix} p & v & v & v & \dots & v \\ & r & z & z & \dots & z \\ & & r & z & \dots & z \\ & & & \dots & \dots & \dots \\ & & & & \ddots & z \\ & & & & & r \end{pmatrix}$$
$$= (r - z)\mathbf{M}$$

we obtain

$$\operatorname{Var}(\hat{r}) = \frac{1}{n^2} 2 \operatorname{tr}\left[(r-z)^2\mathbf{M}\right] = \frac{1}{n^2} 2(r-z)^2 R(\mathbf{M})$$
$$= 2\frac{n-1}{n^2}(r-z)^2 \to 0 \qquad \text{for } n \to \infty,$$

where $R(\mathbf{M})$ is the rank of the matrix $\mathbf{M}$.

It is clear that $\mathrm{E}(\hat{r} - r)^2 = \operatorname{Var}(\hat{r}) + \left[\operatorname{bias}(\hat{r})\right]^2 \to 0$ for $n \to \infty$. □

In this part, one application of the previous theory will be shown. Let $X_1, \ldots, X_n$ be the random sample from the probability distribution P_γ, where γ is an unknown parameter $\gamma \in \Gamma \subset \mathbf{R}^1$. Let $\hat{\gamma}$ be an unbiased estimator of the parameter γ. Consider an arbitrary nonlinear function $g\colon \Gamma \to \mathbf{R}^1$, which in a neighbourhood of γ has all uniformly bounded derivatives. The aim is to estimate the parameter $\theta = g(\gamma)$ with a lower bias than the bias of the estimator $g(\hat{\gamma})$. $\theta^* = g(\hat{\gamma})$ is in general a biased estimator.

We assume the following two conditions:

1. The estimate $\hat{\gamma}$ is an element of the neighbourhood of the point γ.
2. All moments of the estimator $\hat{\gamma}$ are finite.

We can now write the Taylor expansion

$$g(x) = g(\gamma) + (x-\gamma)g'(\gamma) + \frac{1}{2}(x-\gamma)^2 g''(\gamma) + \frac{1}{3!}(x-\gamma)^3 g'''(\gamma) + \ldots .$$

Substituting $x = \hat{\gamma}$ we get

$$g(\hat{\gamma}) = g(\gamma) + (\hat{\gamma}-\gamma)g'(\gamma) + \frac{1}{2}(\hat{\gamma}-\gamma)^2 g''(\gamma) + \frac{1}{3!}(\hat{\gamma}-\gamma)^3 g'''(\gamma) + \ldots .$$

Then

$$\mathrm{E}\big[g(\hat{\gamma})\big] = g(\gamma) + \mathrm{Var}(\hat{\gamma})\frac{1}{2}g''(\gamma) + \mu_3(\hat{\gamma})\frac{1}{3!}g'''(\gamma) + \ldots$$

where $\mu_k(\hat{\gamma})$ denotes the central moment of order k of the random variable $\hat{\gamma}$.

If we remove the term with the highest order, then the bias diminishes properly. In many cases the first term of the bias has a higher order than the subsequent terms. The pseudo-values are given by (2.4) as

$$T_i = \frac{\mathrm{Var}(\hat{\gamma}_{-1})g(\hat{\gamma}) - \mathrm{Var}(\hat{\gamma})g(\hat{\gamma}_{-i})}{\mathrm{Var}(\hat{\gamma}_{-1}) - \mathrm{Var}(\hat{\gamma})} \tag{5.1}$$

and the JE of the parameter $g(\gamma)$ is

$$T = \frac{\mathrm{Var}(\hat{\gamma}_{-1})g(\hat{\gamma}) - \mathrm{Var}(\hat{\gamma})\,\frac{1}{n}\sum_{i=1}^{n} g(\hat{\gamma}_{-i})}{\mathrm{Var}(\hat{\gamma}_{-1}) - \mathrm{Var}(\hat{\gamma})} \tag{5.2}$$

where $\hat{\gamma}_{-i}$ is the estimate obtained after the deletion of the ith observation.

In this case, the estimators T_i, $i = 1, \ldots, n$, and T are not unbiased. The estimators (5.1) and (5.2) remove the term $\frac{1}{2}\,\mathrm{Var}(\hat{\gamma})g''(\gamma)$ from the bias. In this way we can remove an arbitrary term of the bias.

At this point, we need to emphasize that this is only a formal consideration as to how to reduce the bias of estimators.

Theorem 5. *Let $X_1, \ldots, X_n$ be a random sample from the probability distribution P_γ, $\gamma \in \Gamma \subset \mathbf{R}^1$. Further, let g be any polynomial function of the kth order, where $1 < k < n-1$, and let $\mu_{k+1}(\hat{\gamma})$ exist, where $\mathrm{E}(\hat{\gamma}) = \gamma$. Then there exists an unbiased estimator of the parameter $\theta = g(\gamma)$.*

P r o o f. The proof follows immediately from the expression (2.6) and Remark 5. □

The following examples can be given as an illustration of the previous theoretical considerations.

E x a m p l e 1. Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$; μ is an unknown parameter.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{5.3}$$

is an unbiased estimator of the parameter μ.

$$\overline{X}_{-i} = \frac{1}{n-1}\sum_{j=1}^{n} X_j \qquad j \neq i,\ i = 1, 2, \ldots, n.$$

The estimator $\overline{X}^2$ is a biased estimator of the parameter μ^2.

$$\mathrm{E}(\overline{X}^2) = \mu^2 + \frac{1}{n}\sigma^2.$$

The JE of the parameter $\theta = \mu^2$ is

$$\hat{\theta} = n\overline{X}^2 - \frac{n-1}{n}\sum_{i=1}^{n} \overline{X}^2_{-i}. \tag{5.4}$$

This expression is the same as the jackknife estimator given by Miller (1964).

Let us calculate the mean value and the variance of (5.4). Obviously,

$$\mathrm{E}(\hat{\theta}) = \mu^2$$

$$\mathrm{Var}(\overline{X}^2) = \frac{2}{n^2}\sigma^4 + \frac{4}{n}\mu^2\sigma^2 \tag{5.5}$$

$$\mathrm{Var}(\hat{\theta}) = \frac{2}{n(n-1)}\sigma^4 + \frac{4}{n}\mu^2\sigma^2 \tag{5.6}$$

and hence

$$\mathrm{MSE}(\overline{X}^2) = \mathrm{Var}(\overline{X}^2) + n^{-2}\sigma^4 = \frac{3}{n^2}\sigma^4 + \frac{4}{n}\mu^2\sigma^2. \tag{5.7}$$

Thus it becomes clear that $\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) < \mathrm{MSE}(\overline{X}^2)$ for $n > 3$.

In this case, the JE is more appropriate than the original biased estimator $\overline{X}^2$.

Example 2. Let us consider the random sample from the preceding example. Let us find an unbiased estimator of the parameter μ^4.

Because of the relation,

$$\mathrm{E}(\overline{X}^4) = \mu^4 + \mathrm{Var}(\overline{X})6\mu^2 + \mu_4(\overline{X})$$

the JE is derived in the form

$$\hat{\theta} = \frac{1}{2}\left[n^2\overline{X}^4 - \frac{2(n-1)^2}{n}\sum_{i=1}^{n}\overline{X}^4_{-i} + \frac{2(n-2)^2}{n(n-1)}\sum_{i<j}^{n}\overline{X}^4_{-ij}\right] \tag{5.8}$$

where

$$\overline{X}_{-ij} = \frac{1}{n-2}\sum_{\substack{k=1\\k\neq i\\k\neq j}}^{n} X_k.$$

The bias is totally removed by JE (5.8) totally. It is not necessary to know the values of μ and σ^2.

References

[1] *W. Kruskal*: When are Gauss-Markov and least square estimators identical? A coordinate free approach. Ann. Math. Statistics *39* (1968), 70–75.

[2] *G. H. Lavergne, J. R. Mathieu*: The jackknife method and the Gauss-Markov estimation. Probability and Math. Statistics *8* (1987), 111–116.

[3] *R. G. Miller, jr.*: An unbalanced jackknife. Ann. Math. Statistics *2* (1974), 880–891.

[4] *M. H. Quenouille*: Approximate test of correlation in time-series. J. Roy. Statist. Soc. Ser. B *11* (1949), 68–84.

[5] *M. H. Quenouille*: Notes on bias in estimation. Biometrika *43* (1956), 353–360.

[6] *C. R. Rao*: Linear statistical inference and its applications. J. Wiley, 1973.

[7] *F. Štulajter*: Consistency of linear and quadratic least squares estimators in regression models with covariance stationary errors. Applications of Mathematics *36(2)* (1991), 149–155.

[8] *J. W. Tukey*: Variances of variance components: II. The unbalanced single classification. Ann. Math. Statist. *28* (1957), 43–56.

Author's address: *Jaromír Kudeláš*, University of Economics, Department of Mathematics, Odbojárov 10, 832 20 Bratislava, SR.