On weak consistency in linear models with equi-correlated random errors

By: Xinwei Jia, M. Bhaskara Rao, Haimeng Zhang

Jia, X., Rao, M. B., and Zhang, H. (2003). On weak consistency in linear models with equi-correlated random errors. *Statistics*, *37(6)*, 467-477

Abstract:

A new estimator in linear models with equi-correlated random errors is postulated. Consistency properties of the proposed estimator and the ordinary least squares estimator are studied. It is shown that the new estimator has smaller variance than the usual least squares estimator under some mild conditions. In addition, it is observed that the new estimator tends to be weakly consistent in many cases where the usual least squares estimator is not.

Keywords: Weak consistency | Equi-correlated errors | Least squares estimator | Linear models

Article:

1 INTRODUCTION

Let Y_i, $i \geq 1$ be a sequence of random variables satisfying

$$Y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{ip}\beta_p + \varepsilon_i, \quad i \geq 1. \tag{1}$$

Here x_{ij}s are *known constants*, $\beta = (\beta_1, \beta_2, \ldots, \beta_p)^T \in R^p$ is the unknown parameter vector (T denoting matrix transpose), where p is a fixed positive integer, ϵ_i, $i \geq 1$ is a sequence of random variables with mean zero. Assume that Y_ns are square-integrable. For each $n \geq 1$, let $\epsilon_{(n)} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)^T$, and $Q_{(n)}$ = dispersion matrix of $\epsilon_{(n)}$. We further denote $\mathbf{Y}_{(n)} = (Y_1, Y_2, \ldots, Y_n)^T$, $\mathbf{X}_{(n)} = (x_{ij})_{n\times p}$. Suppose for some $n \geq p$, rank$(\mathbf{X}_{(n)}) = p$. The main focus of this paper is to examine the issue of weak consistency when the random errors ϵ_is are equi-correlated.

One prime candidate for examining the issue of consistency is the least squares estimator (LSE) $\hat{\beta}_{(n)} = (\mathbf{X}_{(n)}^T \mathbf{X}_{(n)})^{-1} \mathbf{X}_{(n)}^T \mathbf{Y}_{(n)}$ of β based on the first n random variables $\mathbf{Y}_{(n)}$, where $\mathbf{Y}_{(n)}$ is given by

$$\mathbf{Y}_{(n)} = \mathbf{X}_{(n)}\beta + \varepsilon_{(n)}. \tag{2}$$

The least squares estimator has been considered as the standard estimator of the regression parameter for the linear model. It is simple and more importantly, it is always computable. Moreover, it is well-known that in the case of equi-correlated random errors, the LSE is actually the best linear unbiased estimator for the linear model (2) with intercept, *i.e.* $x_{i1} = 1$ for all $i \geq 1$, (McElroy, 1967).

The weak consistency of $\hat{\beta}_{(n)}$, *i.e.* the convergence of $\hat{\beta}_{(n)}$ to β in probability, has been extensively discussed in the literature. Eicker (1963) was probably the first researcher to study the weak consistency of the LSE for the linear model. Drygas (1976) showed that, in the general case of correlated random errors, if

$$0 < \inf_n \{\text{minimum eigenvalue of } Q_{(n)}\} \leq \sup_n \{\text{maximum eigenvalue of } Q_{(n)}\} < \infty, \tag{3}$$

then $\hat{\beta}_{(n)}$ converges to β in probability if and only if the minimum eigenvalue of $(\mathbf{X}_{(n)}^{\mathrm{T}} \mathbf{X}_{(n)})$ tends to infinity, or equivalently $(\mathbf{X}_{(n)}^{\mathrm{T}} \mathbf{X}_{(n)})^{-1}$ converges to the zero matrix. The result of Drygas (1976) is more general than the one quoted above. The general result is concerned with the weak consistency of estimable linear functions. Kaffes and Rao (1982) examined the issue of weak consistency under the minimal condition that Y_ns are integrable.

Now, let us consider the case of equi-correlated random errors, *i.e.* $E\epsilon_i^2 = \sigma^2 > 0$ for every $i \geq 1$, and $E\epsilon_i \epsilon_j = \rho\sigma^2$ for every $i \neq j$and for some $0 \leq \rho < 1$. Note that in this case, the dispersion matrix $Q_{(n)} = D_n(\epsilon_{(n)}) = \sigma^2(1-\rho)I_n + \sigma^2 \rho J_n$, where I_n is the $n \times n$ identity matrix and $J_n = \mathbf{1}_{(n)} \mathbf{1}_{(n)}^{\mathrm{T}}$ for $\mathbf{1}_{(n)} = (1, 1, \ldots, 1)^{\mathrm{T}}$, the column vector of order $n \times 1$ with each entry equal to unity. The maximum eigenvalue of $Q_{(n)} = \sigma^2(1 + (n-1)\rho)$. If $\rho > 0$, $\sup_n$ {maximum eigenvalue of $Q_{(n)}$} is equal to infinity, *i.e.*condition (3) will never be satisfied. In such a case, Kaffes and Rao (1982) imposed the following sufficient condition,

$$n(\mathbf{X}_{(n)}^{\mathrm{T}} \mathbf{X}_{(n)})^{-1} \text{ converges to the zero matrix,} \tag{4}$$

to guarantee the weak consistency of $\hat{\beta}_{(n)}$. Although (4) is a stronger condition, there are examples in which (4) fails to hold. For instance, if $x_{i1} = 1$, $i \geq 1$ in Eq. (2), *i.e.* the linear model with intercept, then the (1, 1)th element of $n(\mathbf{X}_{(n)}^{\mathrm{T}} \mathbf{X}_{(n)})^{-1}$ has lower bound 1. Therefore, it is worthwhile to seek some other conditions under which the LSE is weakly consistent and/or some other estimators which are better than the LSE in the case of equi-correlated random errors.

In this paper we propose a new estimator $\hat{\beta}_{(n)}^*$ given by Eq. (8), and then consider the weak consistency of the estimator $\hat{\beta}_{(n)}^*$ and LSE $\hat{\beta}_{(n)}$ in the case of equi-correlated random errors. It turns out that the estimator $\hat{\beta}_{(n)}^*$ is better than the LSE $\hat{\beta}_{(n)}$ in the sense of smaller dispersion matrix (Theorem 1 in Sec. 2), and it is weakly consistent under some mild conditions on the design matrix (Theorems 2 and 3 in Section 2).

To motivate the introduction of the new estimator $\hat{\beta}_{(n)}^*$, we consider the following example with $p = 1$ in Eq. (2).

Example 1 Let $Y_i = x_i \beta + \epsilon_i$, $i \geq 1$. The LSE is given by

$$\hat{\beta}_{(n)} = \frac{(\sum_{i=1}^{n} x_i Y_i)}{(\sum_{i=1}^{n} x_i^2)}$$

with variance

$$\mathrm{var}(\hat{\beta}_{(n)}) = \frac{\sigma^2(1-\rho)}{(\sum_{i=1}^{n} x_i^2)} + \frac{\sigma^2 \rho n^2 \bar{x}_{(n)}^2}{(\sum_{i=1}^{n} x_i^2)^2},$$

where $\bar{x}_{(n)} = (1/n) \sum_{i=1}^{n} x_i$. Now, we consider the "centered" estimator for β

$$\hat{\beta}_{(n)}^* = \frac{[\sum_{i=1}^{n} (x_i - \bar{x}_{(n)}) Y_i]}{[\sum_{i=1}^{n} (x_i - \bar{x}_{(n)})^2]}$$

with variance

$$\mathrm{var}(\hat{\beta}_{(n)}^*) = \frac{\sigma^2(1-\rho)}{[\sum_{i=1}^{n} (x_i - \bar{x}_{(n)})^2]} = \frac{\sigma^2(1-\rho)}{(\sum_{i=1}^{n} x_i^2)} + \frac{\sigma^2(1-\rho) n \bar{x}_{(n)}^2}{\{(\sum_{i=1}^{n} x_i^2)[\sum_{i=1}^{n} (x_i - \bar{x}_{(n)})^2]\}}.$$

Note that $\hat{\beta}_{(n)}^*$ and $\hat{\beta}_{(n)}$ are both weighted averages of Y_i s but with different weights. Therefore, from the difference of the variances

$$\mathrm{var}(\hat{\beta}_{(n)}) - \mathrm{var}(\hat{\beta}_{(n)}^*) = \frac{n^2 \sigma^2 (1-\rho) \bar{x}_{(n)}^2}{(\sum_{i=1}^{n} x_i^2)^2} \left\{ \rho - \frac{1-\rho}{n[1 - n\bar{x}_{(n)}^2 / \sum_{i=1}^{n} x_i^2]} \right\},$$

one can see that $\mathrm{var}(\hat{\beta}_{(n)}) \geq \mathrm{var}(\hat{\beta}_{(n)}^*)$ for $\rho > 0$ if $n \bar{x}_{(n)}^2 / (\sum_{i=1}^{n} x_i^2)$ tends to a constant that is less than 1, as n goes to ∞. Consequently, the weak consistency of $\hat{\beta}_{(n)}$ implies the weak consistency of $\hat{\beta}_{(n)}^*$, but the converse is not necessarily true. We would like to point out that in this example, there is no minimum variance linear unbiased estimator of β.

In the general case of p regressors, we derive the new estimator $\hat{\beta}_{(n)}^*$ of β in the same manner as shown above. More explicitly, for $n \geq 2$, let $P_{(n)}$ be the matrix of order $(n-1) \times n$ such that the matrix with order of $n \times n$

$$C_{(n)} = \begin{pmatrix} (1/\sqrt{n}) \mathbf{1}_{(n)}^{\mathrm{T}} \\ P_{(n)} \end{pmatrix} \tag{5}$$

is orthogonal, *i.e.* $C_{(n)}^{\mathrm{T}} C_{(n)} = C_{(n)} C_{(n)}^{\mathrm{T}} = I_n$. Here $\mathbf{1}_{(n)} = (1, 1, \ldots, 1)^{\mathrm{T}}$ and I_n is the identity matrix of order $n \times n$. Obviously, $P_{(n)}$, $n \geq 2$ satisfies the following properties:

$$P_{(n)} P_{(n)}^{\mathrm{T}} = I_{n-1};$$

$$P_{(n)} Q_{(n)} P_{(n)}^{\mathrm{T}} = \sigma^2 (1 - \rho) I_{n-1};$$

$$P_{(n)}^{\mathrm{T}} P_{(n)} = I_n - (1/n) J_n.$$

One example of $P_{(n)}$ is the Helmert matrix [*e.g.* see Press (1982), pages 13 and 14], which can be constructed as follows. The ith row of $P_{(n)}$ is given by

$$\left(\frac{1}{\sqrt{i(i+1)}}, \frac{1}{\sqrt{i(i+1)}}, \ldots, \frac{1}{\sqrt{i(i+1)}}, \frac{-i}{\sqrt{i(i+1)}}, 0, \ldots, 0 \right),$$

where the entry

$$-i/\sqrt{i(i+1)}$$

occurs at the $(i+1)$th position, *i.e.*

$$P_{(n)} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 0 & \cdots & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & 0 & \cdots & 0 \\ \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{-3}{\sqrt{12}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \frac{1}{\sqrt{(n-1)n}} & \frac{1}{\sqrt{(n-1)n}} & \frac{1}{\sqrt{(n-1)n}} & \frac{1}{\sqrt{(n-1)n}} & \cdots & \frac{-(n-1)}{\sqrt{(n-1)n}} \end{pmatrix}.$$

Introducing a column vector of random variables by

$$\mathbf{Z}_{(n)} = \begin{pmatrix} Z_{1n} \\ Z_{2n} \\ \vdots \\ Z_{(n-1)n} \end{pmatrix} = P_{(n)}\mathbf{Y}_{(n)}, \quad n \geq 2,$$

one can notice that

$$E\mathbf{Z}_{(n)} = P_{(n)}\mathbf{X}_{(n)}\beta \equiv \mathbf{U}_{(n-1)}\beta, \tag{6}$$

say, and the dispersion matrix of $\mathbf{Z}_{(n)}$ is given by

$$D_{n-1}(\mathbf{Z}_{(n)}) = \sigma^2(1-\rho)I_{n-1}. \tag{7}$$

Note that the components of $\mathbf{Z}_{(n)}$ are pairwise uncorrelated with common variance $\sigma^2(1-\rho)$. Therefore, the best linear unbiased estimator (BLUE) of β based on $\mathbf{Z}_{(n)}$ is

$$\hat{\beta}^*_{(n)} = (\mathbf{U}^{\mathrm{T}}_{(n-1)}\mathbf{U}_{(n-1)})^{-1}\mathbf{U}^{\mathrm{T}}_{(n-1)}\mathbf{Z}_{(n)}. \tag{8}$$

Here we assume $\mathrm{rank}(\mathbf{U}_{(n-1)}) = p$. A comment is in order on the rank of $\mathbf{U}_{(n-1)}$. If $\mathbf{1}_{(n)}$ belongs to the vector space spanned by the columns of $\mathbf{X}_{(n)}$, $\mathrm{rank}(\mathbf{U}_{(n-1)}) < p$. From now on, we will assume that $\mathbf{U}_{(n-1)}^{\mathrm{T}}\mathbf{U}_{(n-1)}$ is invertible for every $n \geq N$ for some integer N. In Section 4, we will comment on the case when one of the columns of $\mathbf{X}_{(n)}$ is $\mathbf{1}_{(n)}$. Now

$$\mathbf{U}^{\mathrm{T}}_{(n-1)}\mathbf{U}_{(n-1)} = \mathbf{X}^{\mathrm{T}}_{(n)}\mathbf{X}_{(n)} - n\bar{\mathbf{X}}_{(n)}\bar{\mathbf{X}}^{\mathrm{T}}_{(n)},$$

and

$$\mathbf{U}^{\mathrm{T}}_{(n-1)}\mathbf{Z}_{(n)} = \mathbf{X}^{\mathrm{T}}_{(n)}\mathbf{Y}_{(n)} - \frac{1}{n}\mathbf{X}^{\mathrm{T}}_{(n)}J_n\mathbf{Y}_{(n)},$$

where $\bar{\mathbf{X}}_{(n)}^{\mathrm{T}} = (1/n)\mathbf{1}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)}$. Therefore, the “centered” estimator can be rewritten as

$$\hat{\beta}^*_{(n)} = (\mathbf{X}^{\mathrm{T}}_{(n)}\mathbf{X}_{(n)} - n\bar{\mathbf{X}}_{(n)}\bar{\mathbf{X}}^{\mathrm{T}}_{(n)})^{-1}\mathbf{X}^{\mathrm{T}}_{(n)}\left(I_n - \frac{1}{n}J_n\right)\mathbf{Y}_{(n)}. \tag{9}$$

The referee suggested that we could look at the transformed model

$$
\begin{aligned}
\mathbf{W}_{(n)} &= \left(I_n - \frac{1}{n}J_n\right)\mathbf{Y}_{(n)} \\
&= \left(I_n - \frac{1}{n}J_n\right)\mathbf{X}_{(n)}\beta + \left(I_n - \frac{1}{n}J_n\right)\varepsilon_{(n)}, \quad n \geq 1.
\end{aligned}
$$

Note that $\mathrm{Disp}(\mathbf{W}_{(n)}) = \sigma^2(1-\rho)(I_n - (1/n)J_n)$, and the LSE of β based on $\mathbf{W}_{(n)}$ is precisely the estimator $\hat{\beta}_{(n)}^{*}$. In this case, the dispersion matrix of $\mathbf{W}_{(n)}$ is known except for the factor $\sigma^2(1-\rho)$. The matrix $(I_n - (1/n)J_n)$ has rank of $n-1$, which means that the support of the distribution of $\mathbf{W}_{(n)}$ is some $(n-1)$ dimensional subspace of R^n. We make an additional transformation of $\mathbf{W}_{(n)}$ so that the transformed vector has a non-singular dispersion matrix. One choice of the transformation is $P_{(n)}$ in Eq. (5). Let $\mathbf{T}_{(n)} = P_{(n)}\mathbf{W}_{(n)}$. It turns out that $\mathbf{T}_{(n)} = \mathbf{Z}_{(n)}$, which reduces to the model (6).

Our paper is organized into four sections. In Section 2, the weak consistency of $\hat{\beta}_{(n)}^{*}$ and $\hat{\beta}_{(n)}$ is discussed. In Section 3, we examine the consistency properties under the normality assumption of random errors. Some concluding remarks are mentioned in Section 4.

2 WEAK CONSISTENCY OF THE ESTIMATORS

In this section, we consider the weak consistency of $\hat{\beta}_{(n)}^{*}$ and $\hat{\beta}_{(n)}$. First, it is easily seen that the dispersion matrices of $\hat{\beta}_{(n)}$ and $\hat{\beta}_{(n)}^{*}$ are given by

$$
\begin{aligned}
D_p[\hat{\beta}_{(n)}] &= (\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}\mathbf{X}_{(n)}^{\mathrm{T}}Q_{(n)}\mathbf{X}_{(n)}(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1} \\
&= \sigma^2(1-\rho)(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1} + \sigma^2\rho(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}\mathbf{X}_{(n)}^{\mathrm{T}}J_n\mathbf{X}_{(n)}(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}, \qquad (10) \\
D_p[\hat{\beta}_{(n)}^{*}] &= \sigma^2(1-\rho)(\mathbf{U}_{(n-1)}^{\mathrm{T}}\mathbf{U}_{(n-1)})^{-1} \\
&= \sigma^2(1-\rho)[\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)} - n\bar{\mathbf{X}}_{(n)}\bar{\mathbf{X}}_{(n)}^{\mathrm{T}}]^{-1} \\
&= \sigma^2(1-\rho)(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1} + \sigma^2(1-\rho)\frac{(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}\mathbf{X}_{(n)}^{\mathrm{T}}J_n\mathbf{X}_{(n)}(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}}{n - n^2\bar{\mathbf{X}}_{(n)}^{\mathrm{T}}(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}\bar{\mathbf{X}}_{(n)}}, \qquad (11)
\end{aligned}
$$

where $D_p[\cdot]$ represents a $p \times p$ dispersion matrix. The last equality holds due to the standard result, which can be found in Rao (1973, page 33) or Rao and Rao (1998, page 281).
Since $\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)}$ and $\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)} - n\bar{\mathbf{X}}_{(n)}\bar{\mathbf{X}}_{(n)}^{\mathrm{T}}$ are invertible, $n - n^2\bar{\mathbf{X}}_{(n)}^{\mathrm{T}}(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}\bar{\mathbf{X}}_{(n)}$ is positive. By comparing (10) and (11), we have the following theorem.

Theorerm 1

Suppose that $\lim_{n\to\infty} n\bar{\mathbf{X}}_{(n)}^{\mathrm{T}}(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}\bar{\mathbf{X}}_{(n)} = c < 1$. *Then, for sufficiently large n, we have*

$$D_p[\hat{\beta}^*_{(n)}] \leq D_p[\hat{\beta}_{(n)}].$$

■

Now we consider the weak consistency of $\hat{\beta}_{(n)}^*$ *and* $\hat{\beta}_{(n)}$. *We denote the eigenvalues of* $\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)}$ *in the following increasing order:*

$$0 < \lambda_1^{(n)} \leq \lambda_2^{(n)} \leq \lambda_3^{(n)} \leq \cdots \leq \lambda_p^{(n)}.$$

First, we have the following lemmas.

Lemma 1

$$\mathrm{Tr}(D_p[\hat{\beta}^*_{(n)}]) \leq \frac{\sigma^2(1-\rho)p}{\lambda_1^{(n)}} + \frac{\sigma^2(1-\rho)(\lambda_1^{(n)})^{-1}}{1 - n\bar{\mathbf{X}}_{(n)}^{\mathrm{T}}(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}\bar{\mathbf{X}}_{(n)}},$$

where Tr(A) *denotes the trace of a square matrix* A.

Proof It is trivial from Theorem 3.1 in Kaffes and Rao (1982). ▪

Lemma 2

For every integer $1 \leq k < \infty$, *we have*

$$\frac{\|\bar{\mathbf{X}}_{(n)}\|^2}{(\lambda_p^{(n)})^k} \leq \bar{\mathbf{X}}_{(n)}^{\mathrm{T}}(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-k}\bar{\mathbf{X}}_{(n)} \leq \frac{\|\bar{\mathbf{X}}_{(n)}\|^2}{(\lambda_1^{(n)})^k}$$

where $\|\cdot\|$ *is the usual Euclidean norm in* R^p.

Proof It is a trivial consequence of

$$\lambda_{\min}[A]I_p \leq A \leq \lambda_{\max}[A]I_p,$$

for any square matrix A of order $p \times p$. ▪

Lemma 3

The following statements are true:

$$\lim_{n\to\infty} \frac{\|\bar{\mathbf{X}}_{(n)}\|^2}{(\lambda_p^{(n)})} = 0;$$

If, furthermore, for some $1 \le j \le p$, $x_{ij} = c$, *a non-zero constant, for every* $i \ge N$ *and* $N \le n$ *fixed, then*

$$\lim_{n\to\infty} n\|\bar{\mathbf{X}}_{(n)}\| = \infty \quad and \quad \lim_{n\to\infty}(\lambda_p^{(n)}) = \infty.$$

Proof Let $\bar{\mathbf{X}}_{(n)} = (\bar{x}_{n1}, \bar{x}_{n2}, \ldots, \bar{x}_{np})$. Notice that

$$\sum_{k=1}^{p} \lambda_k^{(n)} = \sum_{k=1}^{p}\sum_{i=1}^{n} x_{ik}^2 \ge \sum_{k=1}^{p} n\bar{x}_{np}^2 = n\|\bar{\mathbf{X}}_{(n)}\|^2. \tag{12}$$

Hence,

$$\frac{\|\bar{\mathbf{X}}_{(n)}\|^2}{(p\lambda_p^{(n)})} \le \frac{\|\bar{\mathbf{X}}_{(n)}\|^2}{\sum_{k=1}^{p}\lambda_k^{(n)}} \le \frac{1}{n},$$

which is part 1.

Now look at part 2. The average of the jth column of $\mathbf{X}_{(n)}$ is $\bar{x}_{nj} = (1/n)\sum_{i=1}^{N} x_{ij} + (n-N)c/n$. Thus

$$\begin{aligned} n\|\bar{\mathbf{X}}_{(n)}\| &= n\sqrt{\bar{x}_{n1}^2 + \bar{x}_{n2}^2 + \cdots + \bar{x}_{np}^2} \ge n\sqrt{\bar{x}_{nj}^2} \\ &= n\sqrt{\left(\frac{1}{n}\right)\sum_{i=1}^{N} x_{ij} + \frac{(n-N)c}{n}}. \end{aligned}$$

Therefore, part 2 is trivial from above and Eq. (12). ▪

Now we present the main theorems for our paper.

Theorem 2

For the estimator $\hat{\beta}_{(n)}^{*}$, *we have the following.*

$\hat{\beta}_{(n)}^{*}$ *is a sequence of weakly consistent estimators for* β *if and only if*

$$\lim_{n\to\infty} \lambda_{\max}[\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)} - n\bar{\mathbf{X}}_{(n)}\bar{\mathbf{X}}_{(n)}^{\mathrm{T}}]^{-1} = 0.$$

Here $\lambda_{\max}[A]$ *denotes the maximum eigenvalue of the matrix A.*

A necessary condition for the consistency of $\hat{\beta}_{(n)}^{*}$ *is that*

$$\lim_{n\to\infty}(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}=\mathbf{0},$$

where **0** *stands for the zero matrix.*

If $n\|\bar{\mathbf{X}}_{(n)}\|^2 \le M$ *for some positive constant M, then* $\hat{\beta}_{(n)}^{*}$ *is a sequence of weakly consistent estimators if and only if*

$$\lim_{n\to\infty}(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}=\mathbf{0}.$$

Proof Part 1 is trivial from Theorem 3.1 (c) of Drygas (1976) by observing Eq. (7). From the inequality

$$\lambda_{\min}[\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)}-n\bar{\mathbf{X}}_{(n)}\bar{\mathbf{X}}_{(n)}^{\mathrm{T}}]\le\lambda_{\min}[\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)}],$$

it follows that

$$\lambda_{\min}[\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)}]\longrightarrow\infty,$$

i.e.

$$\lim_{n\to\infty}(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}=\mathbf{0}.$$

Thus part 2 follows. For part 3, it suffices to prove the sufficiency. Let $\lim_{n\to\infty}(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}=\mathbf{0}$. Suppose now that $n\|\bar{\mathbf{X}}_{(n)}\|^2\le M$ for all sufficiently large n. Then from Lemma 2,

$$\lim_{n\to\infty}\bar{\mathbf{X}}_{(n)}^{\mathrm{T}}(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}\bar{\mathbf{X}}_{(n)}=0.$$

Therefore, $\mathrm{Tr}(D_p[\hat{\beta}_{(n)}^{*}])$ tends to zero, which gives the weak consistency of $\hat{\beta}_{(n)}^{*}$. ▪

The next theorem gives conditions for the consistency of both the usual LSE and the “centered” estimator $\hat{\beta}_{(n)}^{*}$.

Theorem 3

Suppose that $\lim_{n\to\infty}(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}=\mathbf{0}$. *The following are valid.*

If $n\|\bar{\mathbf{X}}_{(n)}\|=o(\lambda_1^{(n)})$, *then* $\hat{\beta}_{(n)}$ *is weakly consistent for* β.

If $n\|\bar{\mathbf{X}}_{(n)}\|^2 = o(\lambda_1^{(n)})$, *then* $\hat{\beta}_{(n)}^*$ *is weakly consistent for* β.

Proof From Lemma 2, noting that

$$\begin{aligned}\mathrm{Tr}(D_p[\hat{\beta}_{(n)}]) &= \sigma^2(1-\rho)\mathrm{Tr}((\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}) + \sigma^2\rho\mathrm{Tr}[(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}\mathbf{X}_{(n)}^{\mathrm{T}}J_n\mathbf{X}_{(n)}(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}] \\ &= \sigma^2(1-\rho)\mathrm{Tr}((\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}) + \sigma^2\rho\mathrm{Tr}[(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}n^2\bar{\mathbf{X}}_{(n)}\bar{\mathbf{X}}_{(n)}^{\mathrm{T}}(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}] \\ &\le \frac{\sigma^2(1-\rho)p}{\lambda_1^{(n)}} + \sigma^2\rho n^2\bar{\mathbf{X}}_{(n)}^{\mathrm{T}}(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-2}\bar{\mathbf{X}}_{(n)} \\ &\le \frac{\sigma^2(1-\rho)p}{\lambda_1^{(n)}} + \frac{\sigma^2\rho n^2\|\bar{\mathbf{X}}_{(n)}\|^2}{(\lambda_1^{(n)})^2},\end{aligned}$$

part 1 is trivial from our assumption. For part 2, one has $\lim_{n\to\infty} n\|\bar{\mathbf{X}}_{(n)}\|^2/\lambda_1^{(n)} = 0$. Using Lemma 2, we then have

$$\lim_{n\to\infty}(1 - n\bar{\mathbf{X}}_{(n)}^{\mathrm{T}}(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-2}\bar{\mathbf{X}}_{(n)}) = 1.$$

It follows from Lemma 1 that $\mathrm{Tr}(D_p[\hat{\beta}_{(n)}^*])$ tends to zero, which is sufficient for $\hat{\beta}_{(n)}^*$ to converge to β in probability as $n \to \infty$. ▪

Remark From Theorem 3, one can see that if $n^2\|\bar{\mathbf{X}}_{(n)}\|^2 = o(\lambda_1^{(n)})$, then $\lim_{n\to\infty}(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1} = \mathbf{0}$ is the sufficient condition for the weak consistency of both $\hat{\beta}_{(n)}$ and $\hat{\beta}_{(n)}^*$.

The following example indicates that in many situations, $\hat{\beta}_{(n)}^*$ is weakly consistent but $\hat{\beta}_{(n)}$ is not. It should also be noted that condition (4) imposed by Kaffes and Rao (1982) is not satisfied here.

Example 2 Let

$$Y_i = \begin{cases}\beta_1 + i\beta_2 + \varepsilon_i, & \text{if } i = \text{odd number}, \\ i\beta_2 + \varepsilon_i, & \text{if } i = \text{even number},\end{cases}$$

for $i = 1, 2, \ldots$. ϵ_is are equi-corrleated with variance σ^2 and positive correlation coefficient ρ. First suppose n is an even number. Simple calculation gives

$$\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)} = \left(\frac{n}{2}\right)\begin{pmatrix}1 & \frac{n}{2} \\ \frac{n}{2} & \frac{(n+1)(2n+1)}{3}\end{pmatrix},$$

and

$$(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1} = \frac{24}{n(5n^2+12n+4)}\begin{pmatrix} \frac{(n+1)(2n+1)}{3} & \frac{-n}{2} \\ \frac{-n}{2} & 1 \end{pmatrix} \longrightarrow \mathbf{0}.$$

One can see that $n(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}$ does not tend to zero matrix. The mean vector is $\bar{\mathbf{X}}_{(n)} = (1/2, (n+1)/2)^{\mathrm{T}}$. Simple calculation leads to

$$\begin{aligned}(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}\mathbf{X}_{(n)}^{\mathrm{T}}J_n\mathbf{X}_{(n)}(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1} &= n^2(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}\bar{\mathbf{X}}_{(n)}\bar{\mathbf{X}}_{(n)}^{\mathrm{T}}(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1} \\ &= \frac{144}{(5n^2+12n+4)^2}\begin{pmatrix} \frac{P_4(n)}{36} & 0 \\ 0 & 0 \end{pmatrix},\end{aligned}$$

where $P_4(n)$ is a polynomial of n with a degree of 4. Obviously, the above does not tend to zero matrix as n goes to infinity. Therefore, if ϵ_is are further normally distributed, the usual LSE $\hat{\beta}_{(n)}$ is not consistent.

One can easily verify that

$$n\bar{\mathbf{X}}_{(n)}^{\mathrm{T}}(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}\bar{\mathbf{X}}_{(n)} = \frac{4n^2+12n+8}{5n^2+12n+4}$$

tends to $c = 4/5 < 1$. The consistency of $\hat{\beta}_{(n)}^*$ cannot be concluded from Theorem 1. However, it can be found that

$$[\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)} - n\bar{\mathbf{X}}_{(n)}\bar{\mathbf{X}}_{(n)}^{\mathrm{T}}]^{-1} = \begin{pmatrix} \frac{n}{4} & \frac{-n}{4} \\ \frac{-n}{4} & \frac{n(n^2-1)}{12} \end{pmatrix}^{-1} = \frac{12}{n(n^2-1)}\begin{pmatrix} n^2-1 & 1 \\ 1 & 1 \end{pmatrix} \longrightarrow \mathbf{0},$$

as n goes to infinity. The case of odd number for n gives the same result. Theorem 2 shows that $\hat{\beta}_{(n)}^*$ is a sequence of weakly consistent estimators for β.

3 WEAK CONSISTENCY FOR NORMAL RANDOM ERRORS

Substantial discussions on estimates and significance tests on the parameter, under the additional assumption that ϵ_is are multivariate-normally distributed, can be found in Halperin (1951) and Arnold (1981). In this section, we consider the weak consistency of estimators $\hat{\beta}_{(n)}^*$ and $\hat{\beta}_{(n)}$ under this assumption. From Theorem 1, one has the following immediate corollary.

Corollary 1

Suppose that

$$\lim_{n\to\infty} n\bar{\mathbf{X}}_{(n)}^{\mathrm{T}}(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}\bar{\mathbf{X}}_{(n)} = c < 1,$$

then the weak consistency of $\hat\beta_{(n)}$ *implies the weak consistency of* $\hat\beta_{(n)}^{*}$. ▪

It should be noted that in the case of equi-correlated multivariate-normally distributed random errors with $\rho > 0$, the condition that $\lim_{n\to\infty} (\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1} = \mathbf{0}$ is not enough to guarantee the weak consistency of both $\hat\beta_{(n)}$ and $\hat\beta_{(n)}^{*}$. As an example, we consider the following linear model.

Example 3 Suppose the linear model

$$Y_i = \begin{cases} k\beta, & \text{if } i = 1, \\ \beta, & \text{if } i \geq 2, \end{cases}$$

where $k > 1$ is fixed. Obviously, $(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1} = 1/(k^2 + n - 1) \to 0$. The estimators $\hat\beta_{(n)}$ and $\hat\beta_{(n)}^{*}$, however, are not weakly consistent if the correlated coefficient $\rho > 0$, since $\mathrm{var}(\hat\beta_{(n)}^{*}) \to \sigma^2(1-\rho)/(k-1)^2$, and $\mathrm{var}(\hat\beta_{(n)}) \to \sigma^2\rho$.

The following corollary from Theorem 3 provides the answer.

Corollary 2

Assume that $n\|\bar{\mathbf{X}}_n\|$ *is uniformly bounded for sufficiently large n, then a sufficient condition for the weak consistency of* $\hat\beta_{(n)}$ *and* $\hat\beta_{(n)}^{*}$ *is* $\lim_{n\to\infty}(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1} = \mathbf{0}$. *In addition, it is also the necessary condition for the weak consistency of* $\hat\beta_{(n)}$.

Proof The sufficiency follows from the remark right after Theorem 3. Now if $\hat\beta_{(n)}$ is consistent, it is necessary that $\mathrm{Tr}(D_p[\hat\beta_{(n)}]) = \sum_{j=1}^{p} \mathrm{var}(\hat\beta_{j(n)})$ tends to zero, where $\hat\beta_{j(n)}$ is the usual LSE of the jth parameter β_j in the linear model (2). From Eq. (10), we know that $\mathrm{Tr}((\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}) \to 0$, which implies that $(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}$ tends to the zero matrix as n tends to infinity.

It is natural to consider the weighted LSE of β based on the following two-stage estimation.

$$\tilde\beta_{(n)} = (\mathbf{X}_{(n)}^{\mathrm{T}}\hat{Q}_{(n)}^{-1}\mathbf{X}_{(n)})^{-1}\mathbf{X}_{(n)}^{\mathrm{T}}\hat{Q}_{(n)}^{-1}\mathbf{Y}_{(n)},$$

with

$$\hat{Q}_{(n)} = \hat\sigma^2(1-\hat\rho)I_n + \hat\sigma^2\hat\rho J_n,$$

where $\hat\sigma^2$ and $\hat\rho$ are some estimators for σ^2 and ρ, respectively. One could choose maximum likelihood estimators for σ^2 and ρ. However, it appears impossible to have closed form expressions for maximum likelihood estimators of σ^2 and ρ in the case of normally distributed random errors. This conclusion was also mentioned by Halperin (1951). Furthermore, as pointed out by Arnold (1979, 1981), there are no maximum likelihood estimators for the linear model (2) with intercept. In such a case, one could also transform the given model into a variance components model and then pursue the MINUQE procedure. [For MINUQE procedure, see Rao and Rao (1998).] We will not pursue these ideas in this paper.

4 CONCLUDING REMARKS

This paper is devoted to the weak consistency of both LSE $\hat{\beta}_{(n)}$ and the "centered" estimator $\hat{\beta}_{(n)}^{*}$ in the case of equi-correlated random errors. It turns out that under some mild conditions, the "centered" estimator is weakly consistent. In addition, the weak consistency under the assumption of multivariate normality is also discussed.

It should be pointed out that in the usual linear regression model (2) with intercept and equi-correlated random errors, the LSE for the parameter β is never consistent if $\rho > 0$. It is easy to verify that

$$(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1}\mathbf{X}_{(n)}^{\mathrm{T}}J_n\mathbf{X}_{(n)}(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Therefore,

$$D_p[\hat{\beta}_{(n)}] = \sigma^2(1-\rho)(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1} + \sigma^2\rho \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Obviously, if $\rho > 0$, $\hat{\beta}_{(n)}$ is not consistent. However, one can observe that the inconsistency is actually due to the inconsistency of $\hat{\beta}_{1(n)}$, the LSE for the intercept parameter β_1. If $\lim_{n\to\infty} (\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)})^{-1} = 0$, $(\hat{\beta}_{2(n)}, \hat{\beta}_{3(n)}, \ldots, \hat{\beta}_{p(n)})$ is a weakly consistent estimator of $(\beta_2, \beta_3, \ldots, \beta_p)$. In addition, the proposed estimator $\hat{\beta}_{(n)}^{*}$ does not exist because $\mathbf{U}_{(n-1)}^{\mathrm{T}}\mathbf{U}_{(n-1)} = \mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{X}_{(n)} - n\,\bar{\mathbf{X}}_{(n)}\bar{\mathbf{X}}_{(n)}^{\mathrm{T}}$ is now a singular matrix.

Acknowledgements

The authors thank the referee for constructive comments which have led to an improvement of the paper.

References

1. Arnold, S. F. (1979). Linear models with exchangeably distributed errors. *Journal of the American Statistical Association*, 74: 194–199.

2. Arnold S. F. (1981) *The Theory of Linear Models and Multivariate Analysis* Wiley New York

3. Drygas, H. (1976). Weak and strong consistency of the least squares estimator in regression models. *Z. Wahrsch. Verw. Gebiete*, 34: 119–127.

4. Eicker, F. (1963). Asymptotic normality and consistency of least squares estimators for families of linear regressions. *Annals of Mathematical Statistics*, 34: 447–456.

5. Halperin, M. (1951). Normal regression theory in the presence of intra-class correlation. *Annals of Mathematical Statistics*, 22: 573–580.

6. Kaffes, D. and Rao, M. B. (1982). Weak consistency of least squares estimators in linear models. *Journal of Multivariate Analysis*, 12: 186–198.

7. McElroy, F. W. (1967). A necessary and sufficient condition that ordinary least-squares estimators be best linear unbiased.*Journal of the American Statistical Association*, 62: 1302–1304.

8. Press S. J. (1982) *Applied Multivariate Analysis* 2nd ed. E. Krieger Publishing Company Florida

9. Rao C. R. (1973) *Linear Statistical Inference and Its Applications* 2nd ed. Wiley New York

10. Rao C. R. Rao M. B. (1998) *Matrix Algebra and Its Applications to Statistics and Econometrics* World Scientific New Jersey