Applications of Mathematics

František Štulajter

Consistency of linear and quadratic least squares estimators in regression models with covariance stationary errors

Applications of Mathematics, Vol. 36 (1991), No. 2, 149–155

Persistent URL: `http://dml.cz/dmlcz/104452`

CONSISTENCY OF LINEAR AND QUADRATIC LEAST SQUARES ESTIMATORS IN REGRESSION MODELS WITH COVARIANCE STATIONARY ERRORS

František Štulajter

(Received October 2, 1989)

Summary. The least squres invariant quadratic estimator of an unknown covariance function of a stochastic process is defined and a sufficient condition for consistency of this estimator is derived. The mean value of the observed process is assumed to fulfil a linear regression model. A sufficient condition for consistency of the least squares estimator of the regression parameters is derived, too.

Keywords: Stochastic process, least squares estimators, quadratic invariant estimators, consistency.

AMS Classification: 62M10, 62J05.

0. INTRODUCTION

The problem of consistency of the linear least squares estimators of regression coefficients in linear regression models belongs to the classical problems of the asymptotic statistics and was studied for example in [1], [2], [3], [4]. Only a few results are known about the consistency of quadratic estimators, see [5], [6], in the mixed linear models. The aim of this article is to define the invariant quadratic least squares estimators of a covariance function of a covariance stationary random process, mean value of which fulfils the linear regression model, and to find conditions under which these estimators are consistent. It will be shown that these conditions do not depend on the design matrix of the regression model; they depend only on some limit properties of the estimated covariance function and are rather weak.

It will be assumed in the sequel that the $n \times 1$ random vector $\boldsymbol{X}$ is a finite observation of a random process $X = \{X(t);\ t \in Z\}$, $Z = \{0, \pm 1, \pm 2, \ldots\}$ with $E[X(t)] = \sum_{i=1}^{k} \beta_i f_i(t);\ t \in Z$, where $f_1, \ldots, f_k$ are known functions and $\operatorname{Cov}\{X(s); X(t)\} =$

$= R(|s - t|)$; $s, t \in Z$. Thus we can write

$$X = F\beta + \varepsilon\,, \quad \text{where} \quad E[\varepsilon] = 0\,, \quad E[\varepsilon\varepsilon'] = \Sigma \tag{0.1}$$

with $\Sigma_{ij} = R(|i - j|)$; $i, j = 1, 2, \ldots, n$ and F is a known $n \times k$ matrix of the full rank k.

1. CONSISTENCY OF THE LEAST SQUARES ESTIMATOR OF REGRESSION PARAMETERS

Let

$$\hat{\beta} = (F'F)^{-1} F'X \tag{1.1}$$

be the LSE of β in (0.1). Then it is clear that the covariance matrix $\Sigma_{\hat{\beta}}$ of the estimator $\hat{\beta}$ is given by

$$\Sigma_{\hat{\beta}} = (F'F)^{-1} F'\Sigma F(F'F)^{-1}\,,$$

and for the dispersion of any linear function $f'\hat{\beta}$; $f \in E^k$ of $\hat{\beta}$ we get

$$D_\Sigma[f'\hat{\beta}] = f'(F'F)^{-1} F'\Sigma F(F'F)^{-1} f\,.$$

Using the Schwarz inequality we can write

$$D_\Sigma[f'\hat{\beta}] = (\Sigma F(F'F)^{-1} f, F(F'F)^{-1} f)_{E^n} \leqq$$
$$\leqq \|\Sigma\| \,.\, \|F(F'F)^{-1} f\|^2_{E^n} = \|\Sigma\| \,.\, f'(F'F)^{-1} f\,,$$

where $(a, b)_{E^n} = a'b = \sum_{i=1}^{n} a_i b_i$; $a, b \in E^n$ is the usual Euclidean inner product in E^n and $\|A\|$ denotes the Euclidean norm of a matrix A generated by the inner product $(A, B) = \sum_{j,i=1}^{n} A_{ij} B_{ij}$ defined in the space of $n \times n$ matrices.

Since $\Sigma_{ij} = R(|i - j|)$, we can write

$$\|\Sigma\| = (n\, R^2(0) + 2 \sum_{t=1}^{n-1} (n - t)\, R^2(t))^{1/2}\,. \tag{1.2}$$

F depends on n and let us write $F = F_n$ in the sequel. Now let $(F_n'F_n) = O(1/n)$, by which we mean that $\lim_{n\to\infty} (F_n'F_n)^{-1} \,.\, n = G$, where G is a nonnegative definite matrix. Then for any $f \in E^k$ we have

$$\lim_{n\to\infty} D[f'\hat{\beta}] \leqq \lim_{n\to\infty} \left(\frac{R^2(0)}{n} + \frac{2}{n} \sum_{t=1}^{n} \left(1 - \frac{t}{n}\right) R^2(t)\right)^{1/2} .\, f'Gf\,. \tag{1.3}$$

The following theorem is a consequence of the derived inequality.

Theorem 1.1. *Let X fulfil the model* (0.1), *and let* $(F_n'F_n)^{-1} = O(1/n)$. *Let*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \left(1 - \frac{t}{n}\right) R^2(t) = 0\,.$$

Then for any $f \in E^k$, $f'\hat{\beta}$ converges in probability to $f'\beta$ as $n \to \infty$.

Proof. It follows from the Tchebyshev inequality and from the inequality (1.3).

Corollary 1.1. *Let* $(\mathbf{F}_n'\mathbf{F}_n)^{-1} = O(1/n)$ *and let* $\lim_{t\to\infty} R(t) = 0$. *Then Theorem* 1.1 *holds.*

Proof. It follows from the inequality

$$\frac{1}{n}\sum_{t=1}^{n}\left(1-\frac{t}{n}\right)R^2(t) \leqq \frac{1}{n}\sum_{t=1}^{n} R^2(t)$$

and from the fact that

$$\lim_{t\to\infty} R^2(t) = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} R^2(t) = 0 \quad \text{of} \quad \lim_{t\to\infty} R(t) = 0\,.$$

Example. Let $E[X(t)] = \beta_1 + \beta_2 t$ be a linear trend of X. Then

$$\mathbf{F}_n = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 2 & \dots & n \end{pmatrix}' \quad \text{and} \quad (\mathbf{F}_n'\mathbf{F}_n)^{-1} = \begin{pmatrix} \dfrac{2(2n+1)}{n(n-1)} & -\dfrac{6}{n(n-1)} \\ -\dfrac{6}{n(n-1)} & \dfrac{12}{n(n^2-1)} \end{pmatrix} = O(1/n)$$

and the first condition of Theorem 1.1 is fulfilled. Thus the claim of Theorem 1.1 is true if the second condition of the theorem (or its corollary) is fulfilled.

2. CONSISTENCY OF THE LEAST SQUARES ESTIMATOR OF A COVARIANCE FUNCTION

Let us consider again the model (0.1). Then the assumption $\Sigma_{ij} = R(|i-j|)$; $i, j = 1, 2, \dots, n$ can be written in the form $\boldsymbol{\Sigma} = \sum_{t=0}^{n-1} R(t)\,\mathbf{U}(t)$, where $\mathbf{U}(0) = \mathbf{I}$; $\mathbf{U}(t) = \mathbf{K}(t) + \mathbf{K}(t)'$, with $\mathbf{K}(t)$ being the block matrix

$$\mathbf{K}(t) = \begin{pmatrix} \mathcal{O} & \mathbf{I}(t) \\ \mathcal{O} & \mathcal{O} \end{pmatrix},$$

where $\mathbf{I}(t)$ is the $(n-t)\times(n-t)$ identity matrix for $t = 1, 2, \dots, n-1$. Thus the model (0.1) can be regarded as a mixed linear model with unknown parameters $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)'$ and $\mathbf{R} = (R(0), \dots, R(n-1))'$.

Let $\hat{\boldsymbol{\beta}}$ be the least squares estimator of $\boldsymbol{\beta}$ and let

$$\tilde{\Sigma} = (\mathbf{X} - \mathbf{F}\hat{\boldsymbol{\beta}})(\mathbf{X} - \mathbf{F}\hat{\boldsymbol{\beta}})' = (\mathbf{I} - \mathbf{P})\,\mathbf{X}\mathbf{X}'(\mathbf{I} - \mathbf{P})\,,$$

where $\mathbf{P} = \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\,\mathbf{F}'$ is a preliminary estimate of Σ.

Definition. *The estimator* $\hat{\mathbf{R}} = (\hat{R}(0), \dots, \hat{R}(n-1))'$ *for which the equality*

$$\Big\|\sum_{i=0}^{n-1}\hat{R}(i)\,\mathbf{U}_i - \tilde{\Sigma}\Big\|^2 = \min_{c}\Big\|\sum_{i=0}^{n-1} c_i\mathbf{U}_i - \tilde{\Sigma}\Big\|^2$$

holds will be called the least squares estimator of $\mathbf{R}$.

Lemma 2.1. *The estimator* $\mathbf{R}$ *given by*

$$\hat{R}(t) = \frac{1}{n-t} \sum_{i=1}^{n-t} (X(i+t) - \mathbf{F}\hat{\boldsymbol{\beta}})(i+t))(X(i) - (\mathbf{F}\hat{\boldsymbol{\beta}})(i)); \tag{2.1}$$

$$t = 0, 1, \ldots, n-1,$$

where $\mathbf{F}\hat{\boldsymbol{\beta}} = ((\mathbf{F}\hat{\boldsymbol{\beta}})(1), \ldots, (\mathbf{F}\hat{\boldsymbol{\beta}})(n))'$, $\hat{\boldsymbol{\beta}}$ *given by* (1.1), *is the least squares estimator of* $\mathbf{R}$.

Proof. We are looking for the projection $\hat{\Sigma} = \sum_{i=0}^{n-1} \hat{R}(i)\, \mathbf{U}(i)$ of the matrix $\tilde{\Sigma}$ on the subspace generated by matrices $\mathbf{U}_0, \ldots, \mathbf{U}_{n-1}$. These matrices are orthogonal, $(\mathbf{U}_s, \mathbf{U}_t) = 0$ for $s \neq t$ and their Gramm's matrix $\mathbf{G}$ is diagonal with $G_{00} = n$ and $G_{tt} = \|\mathbf{U}_t\|^2 = 2(n-t)$; $t = 1, 2, \ldots, n-1$. Thus the vector $\hat{\mathbf{R}}$ which determines the projection $\hat{\Sigma}$ of $\tilde{\Sigma}$ is given by

$$\hat{\mathbf{R}} = \mathbf{G}^{-1}(\tilde{\Sigma}, \mathbf{U}), \tag{2.2}$$

where $(\tilde{\Sigma}, \mathbf{U}) = ((\tilde{\Sigma}, \mathbf{U}_0), \ldots, (\tilde{\Sigma}, \mathbf{U}_{n-1}))'$. We see that $\hat{\mathbf{R}}$ given by (2.2) is the same estimator as that given by (2.1). ∎

Now let $\mathbf{B}(0) = \mathbf{I}$ and let $\mathbf{B}(t) = 1/2\, \mathbf{U}(t) = 1/2\, (\mathbf{K}(t) + \mathbf{K}(t)')$, $t = 1, 2, \ldots, n-1$. Then we can write

$$\hat{R}(t) = \frac{1}{n-t} (\tilde{\Sigma}, \mathbf{B}(t)) = \frac{1}{n-t} \operatorname{tr} (\tilde{\Sigma}\, \mathbf{B}(t)) =$$

$$= \frac{1}{n-t} \mathbf{X}'(\mathbf{I} - \mathbf{P})\, \mathbf{B}(t)\, (\mathbf{I} - \mathbf{P})\, \mathbf{X}; \quad t = 0, 1, \ldots, n-1.$$

Using the theory of invariant quadratic estimators given in [5] we see that estimators $\hat{R}(t)$; $t = 0, 1, \ldots, n-1$ are quadratic (in $\mathbf{X}$) estimators, which are invariant with respect to the regression parameters $\boldsymbol{\beta}$. According to this theory we can write

$$E_\Sigma[\hat{R}(t)] = \frac{1}{n-t} \operatorname{tr} ((\mathbf{I} - \mathbf{P})\, \mathbf{B}(t)\, (\mathbf{I} - \mathbf{P})\, \Sigma) \tag{2.3}$$

and, for a normally distributed random vector $\mathbf{X}$,

$$D_\Sigma[\hat{R}(t)] = \frac{2}{(n-t)^2} \operatorname{tr} (((\mathbf{I} - \mathbf{P})\, \mathbf{B}(t)\, (\mathbf{I} - \mathbf{P})\, \Sigma)^2) \tag{2.4}$$

for the mean value and the dispersion of the estimators $\hat{R}(t)$; $t = 0, 1, \ldots, n-1$, where $\Sigma = \sum_{i=0}^{n-1} R(i)\, \mathbf{U}(i)$.

From (2.3) we get

$$E_\Sigma[\hat{R}(t)] = R(t) - \frac{1}{n-t} (\operatorname{tr} (\mathbf{B}(t)\, \mathbf{P}\Sigma) + \tag{2.5}$$

$$+ \operatorname{tr} (\mathbf{B}(t)\, \Sigma\mathbf{P}) - \operatorname{tr} (\mathbf{B}(t)\, \mathbf{P}\Sigma\mathbf{P})),$$

from which it follows that the estimators $\hat{R}(t)$; $t = 0, 1, \dots, n-1$ are biased.

Using the Schwarz inequality we get

$$|\mathrm{tr}\,(\mathbf{B}(t)\,\mathbf{P}\Sigma)| \leqq \|\mathbf{B}(t)\,\mathbf{P}\| \,.\, \|\Sigma\| \,.$$

The same inequality holds with $|\mathrm{tr}\,(\mathbf{P}\,\mathbf{B}(t)\,\Sigma)|$ on the left hand side and

$$|\mathrm{tr}\,(\mathbf{B}(t)\,\mathbf{P}\Sigma\mathbf{P})| \leqq \|\mathbf{B}(t)\,\mathbf{P}\| \,.\, \|\Sigma\| \,,\, \|\mathbf{P}\| \,,$$

since $\|\mathbf{AB}\| \leqq \|\mathbf{A}\| \,.\, \|\mathbf{B}\|$ for any matrices $\mathbf{A}$ and $\mathbf{B}$.

Now, let $\mathbf{A}$ be any $n \times n$ matrix. Then we can write

$$\mathbf{K}(t)\,\mathbf{A} = \begin{pmatrix} \mathcal{O} & \mathbf{I}(t) \\ \mathcal{O} & \mathcal{O} \end{pmatrix} \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{21} & \mathbf{A}_{22} \\ \mathcal{O} & \mathcal{O} \end{pmatrix},$$

where $\mathbf{A}_{21}$ is $(n-t) \times t$ matrix and $\mathbf{A}_{22}$ is $(n-t) \times (n-t)$ matrix, from which it can be seen that $\|\mathbf{K}(t)\,\mathbf{A}\| \leqq \|\mathbf{A}\|$. By analogy $\|\mathbf{K}(t)'\,\mathbf{A}\| \leqq \|\mathbf{A}\|$. Thus we can write

$$\|\mathbf{B}(t)\,\mathbf{P}\| = \|\tfrac{1}{2}\,(\mathbf{K}(t) + \mathbf{K}(t)')\,\mathbf{P}\| \leqq \|\mathbf{P}\| \,. \tag{2.6}$$

Using this inequality we get

$$\begin{aligned} &\frac{1}{n-t}\,|\mathrm{tr}\,(\mathbf{B}(t)\,\mathbf{P}\Sigma)| \leqq \|\mathbf{P}\| \,.\, \|\Sigma\| \frac{1}{n-t}\,, \\ &\frac{1}{n-t}\,|\mathrm{tr}\,(\mathbf{P}\,\mathbf{B}(t)\,\Sigma)| \leqq \frac{1}{n-t}\,\|\mathbf{P}\| \,.\, \|\Sigma\| \quad \text{and} \\ &\frac{1}{n-t}\,|\mathrm{tr}\,(\mathbf{B}(t)\,\mathbf{P}\Sigma\mathbf{P})| \leqq \frac{1}{n-t}\,\|\mathbf{P}\|^2 \,.\, \|\Sigma\| \,. \end{aligned} \tag{2.7}$$

Since $\|\mathbf{P}\|^2 = \mathrm{tr}\,(\mathbf{P}^2) = \mathrm{tr}\,(\mathbf{P}) = \mathit{rank}\,(\mathbf{P}) = k$ we can prove the following theorem.

Theorem 2.1. *Let for a covariance function $R(\cdot)$ of a random process X the condition*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \left(1 - \frac{t}{n}\right) R^2(t) = 0$$

hold. Then the estimators $\hat{R}(t)$ given (2.1) are asymptotically unbiased estimators of $R(t)$ for every fixed t.

Proof. From (2.5) and (2.7) we get, using (1.2), the inequality

$$|E_\Sigma[\hat{R}(t)] - R(t)| \leqq \frac{n}{n-t}\,(2k^{1/2} + k)\,.$$

$$.\left(\frac{R^2(0)}{n} + \frac{2}{n} \sum_{t=1}^{n} \left(1 - \frac{t}{n}\right) R^2(t)\right)^{1/2}.$$

Corollary 2.1. *Let $\lim_{t\to\infty} R(t) = 0$. Then the claim of the preceding theorem holds, too.*

Proof. The same as that of Corollary 1.1. ∎

Now let us assume that the random process X is *Gaussian*. Then we have the expression (2.4) for the variances of the estimators $\hat{R}(t)$, which can be written in the form

$$D_\Sigma[\hat{R}(t)] = \frac{2}{(n-t)^2}\left[\operatorname{tr}\left((\mathbf{B}(t)\,\Sigma)^2\right) + \operatorname{tr}\left((\mathbf{B}(t)\,\mathbf{P}\Sigma)^2\right) + \ldots\right.$$

$$\left.\ldots + (-1)\operatorname{tr}(\mathbf{P}\,\mathbf{B}(t)\,\Sigma\mathbf{P}\,\mathbf{B}(t)\,\mathbf{P}\Sigma)\right].$$

Using the inequalities $|\operatorname{tr}(\mathbf{AC})| \leqq \|\mathbf{A}\| \,.\, \|\mathbf{C}\|$, $\|\mathbf{AC}\| \leqq \|\mathbf{A}\| \,.\, \|\mathbf{C}\|$, $\|\mathbf{B}(t)\,\mathbf{A}\| \leqq \|\mathbf{A}\|$, which hold for any matrices $\mathbf{A}$ and $\mathbf{C}$ we get the following inequality

$$D_\Sigma[\hat{R}(t)] \leqq \frac{c}{(n-t)^2}\,\|\Sigma\|^2\,, \tag{2.8}$$

where the constant c does not depend on n but depends only on k, since, for example,

$$|\operatorname{tr}((\mathbf{B}(t)\,\Sigma)^2)| \leqq \|\mathbf{B}(t)\,\Sigma\|^2 \leqq \|\Sigma\|^2\,, \quad \text{or}$$

$$|\operatorname{tr}(\mathbf{P}\,\mathbf{B}(t)\,\Sigma\mathbf{P}\,\mathbf{B}(t)\,\mathbf{P}\Sigma)| \leqq \|\mathbf{P}\|^3 \,.\, \|\Sigma\|^2 = k^{3/2}\|\Sigma\|^2\,.$$

The following theorem holds.

Theorem 2.2. *Let for a covariance function $R(\cdot)$ of the Gaussian random process X the condition*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n}\left(1-\frac{t}{n}\right)R^2(t) = 0$$

hold. Let $\hat{R}(t)$ be the estimators of $R(t)$ given by (2.1). Then $\lim_{n\to\infty} E_R[\hat{R}(t) - R(t)]^2 = 0$ for every fixed t.

Proof. Since

$$E_R[\hat{R}(t) - R(t)]^2 = D_\Sigma[\hat{R}(t)] + (E_\Sigma[\hat{R}(t)] - R(t))^2\,,$$

the proof follows from Theorem 2.1 and from the inequality (2.8).

Corollary 2.2. *Let $\lim_{t\to\infty} R(t) = 0$. Then Theorem 2.2 holds, too.*

Proof. The same as that of Corollary 1.1.

Remarks. 1. Using the Tchebyshev inequality we can prove that $\hat{R}(t)$ converges in probability to $R(t)$ for every fixed t as $n \to \infty$ under the assumptions of Theorem 2.2 or its corollary.

2. The required assumptions for consistency of the estimators $\hat{R}(t)$ are rather weak and they depend neither on the shape of the regression model for the mean value, nor on the number of regression parameters.

Theorem 2.3. *Let $\hat{\Sigma} = \sum_{t=0}^{m} \hat{R}(t)\,\mathbf{U}(t)$, where $\hat{\mathbf{R}}$ is the least square estimator of* $\mathbf{R}$

given by (2.1) *and m is any positive integer. Let* $\Sigma = \sum_{t=0}^{m} R(t)\, \mathbf{U}(t)$ *and let*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \left(1 - \frac{t}{n}\right) R^2(t) = 0\,,$$

where $R(\cdot)$ *is an unknown covyriance function of a Gaussian random process X. Then*

$$\lim_{n\to\infty} E[\|\hat{\Sigma} - \Sigma\|^2] = 0\,.$$

Proof. We have

$$\|\hat{\Sigma} - \Sigma\|^2 = \|\sum_{t=0}^{m}(\hat{R}(t) - R(t))\, \mathbf{U}(t)\|^2 = \sum_{t=0}^{m}(\hat{R}(t) - R(t))^2 \,.\, \|\mathbf{U}(t)\|^2$$

since the matrices $\mathbf{U}(t)$; $t = 0, 1, \ldots, m$ are orthogonal, and the proof is completed by applying the results of Theorem 2.2.

References

[1] *T. W. Anderson, J. B. Taylor:* Strong consistency of least squares estimates in normal linear regression. Ann. Stat. 4 (1976), 788–790.

[2] *E. Z. Demidenko:* Linear and nonlinear regression. (Russian) Finansy i statistika, Moscow 1981.

[3] *E. J. Hannan:* Rates of convergence for time series regression. Advances Appl.. Prob. 10 (1978), 740–743.

[4] *V. Solo:* Strong consistency of least squares estimators in regression with correlated disturbances. Ann. Stat. 9 (1981), 689–693.

[5] *F. Štulajter:* Estimators in random processes. (Slovak). Alfa, Bratislava 1989.

[6] *R. Thrum, J. Kleffe:* Inequalities for moments of quadratic forms with applications to almost sure convergence. Math. Oper. Stat. Ser. Stat. 14 (1983), 211–216.

Súhrn

KONZISTENCIA LINEÁRNYCH A KVADRATICKÝCH ODHADOV V REGRESNÝCH MODELOCH S KOVARIANČNE STACIONÁRNYMI CHYBAMI

FRANTIŠEK ŠTULAJTER

V článku je definovaný invariantný kvadratický odhad získaný metódou najmenších štvorcov pre neznámu kovariančnú funkciu náhodného procesu. Je odvodená postačujúca podmienka pre konzistenciu tohoto odhadu. O strednej hodnote pozorovaného náhodného procesu sa pritom predpokladá, že spĺňa lineárny regresný model. Opäť je odvodená postačujúca podmienka pre konzistenciu odhadu regresných parametrov získaných metódou najmenších štvorcov.

Author's address: doc. RNDr. *František Štulajter,* CSc., Katedra teórie pravdepod. a mat. štat. MFF UK, Mlynská dolina, 842 15 Bratislava.