A NECESSARY AND SUFFICIENT CONDITION FOR CONSISTENCY
OF THE LS-ESTIMATES IN LINEAR REGRESSION

by

Friedhelm Eicker
University of North Carolina

This research was supported by the Office of Naval Research under Contract No. Nonr-855(06) for research in probability and statistics at Chapel Hill. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Institute of Statistics
Mimeograph Series No. 266
October, 1960
A NECESSARY AND SUFFICIENT CONDITION FOR CONSISTENCY
OF THE LS-ESTIMATES IN LINEAR REGRESSION

by

Friedhelm Eicker, University of North Carolina*

THE
LIBRARY
UNIV
NO

1. Notations and definitions.

The model used for linear regression is described by the equation

\[ y_t = x_{1t} \beta_1 + \ldots + x_{qt} \beta_q + \varepsilon_t, \quad t = 1, 2, \ldots , \]

or in vector notation (which is that used in the analysis of stochastic difference equations \( \sum_{5}^{7}, \sum_{6}^{7} \))

\[ y = X \beta + \varepsilon, \]

where the vector of observed values is for any \( N \) \( y = (y_1, y_2, \ldots , y_N)^\prime \).**

The vectors of the known regression variables are \( x_t = (x_{1t}, \ldots , x_{qt})^\prime \), their matrix is \( X = (x_1, \ldots , x_q) \). Further \( \beta = (\beta_1, \ldots , \beta_q)^\prime \) and \( \varepsilon = (\varepsilon_1, \ldots , \varepsilon_N)^\prime \) where the \( \varepsilon \)'s are independent and have zero mean. The \( \varepsilon \)'s need not be distributed identically. Let us denote by \( f \) the \( N \)-dimensional vector of their distribution functions (d.f.'s), and by \( F \) the space formed by all those vectors, subjected possibly to

*On leave from the University of Mainz.

**Prime denotes the transpose of a vector or matrix.

This research was supported by the Office of Naval Research under Contract No. Nonr-855(06) for research in probability and statistics at Chapel Hill. Reproduction in whole or in part is permitted for any purpose of the United States Government.
a general condition; we will need later on the restriction that all variances, \( \text{var } e_i \), of d.f.'s in \( F \) are uniformly bounded with respect to \( i \). Let \( A \) be the \( q \)-dimensional space of all admissible parameter vectors \( \beta \); we later on allow \( A \) to be the entire \( q \)-dimensional real space. Let \( A \times F \) be the product space of \( A \) and \( F \).

The least squares (LS) estimators \( \hat{\beta} \) of \( \beta \) which are considered here are determined from the normal equations

\[
(3) \quad X'Y = P\hat{\beta}, \quad P = X'X.
\]

Under estimators consistent on a set of parametric quantities and functions we understand an estimator which is consistent for any particular choice of parameters and functions out of this set.

2. A condition for consistent estimators.

The asymptotic behaviour of the estimators in linear regression are governed by the following:

Theorem: The LS-estimates \( \hat{\beta} \) of \( \beta \) in (1), where \( E\varepsilon_t = 0 \),
\[ E\varepsilon_t^2 < \text{const}^* \] for all \( t \), and the \( \varepsilon \)'s are independently distributed,

*The symbol const is used in formulae to denote any constant; if several occur in an equation or inequality they are usually different.
are consistent on $A \times F$ if and only if $\lambda_{\text{min}}(P) \rightarrow \infty$. Here $\lambda_{\text{min}}(P)$ is the minimum characteristic value of $P = X'X$.

Remarks. (I) If $X$ is not of full rank, but if the linear relations between its columns are known, then this case after a simple transformation is subjected to the previous theorem.

(II) In the theorem, $F$ is defined as the d.f. space of the $\epsilon$'s with countable many dimensions. The variances of all d.f.'s are uniformly bounded. Actually the theorem holds also for any subspace of $F$ which contains a d.f. vector $\{N(0,\sigma^2_1), N(0,\sigma^2_2), \ldots\}$ where all variances $\sigma^2_1$ lie between two positive constants.

Proof: From (2) and (3) comes $X'\tilde{\epsilon} = P(b-\beta)$, $b$ can be uniquely determined if and only if for almost all $N$, $|P| \neq 0$. Hence for $N > N_0$

$$P^{-1}X'\tilde{\epsilon} = b - \beta.$$  

(I) Sufficiency: Clearly $E(P^{-1}X'\tilde{\epsilon}) = 0$. The variance of each component of the vector $P^{-1}X'\tilde{\epsilon}$ tends to zero if and only if $E(\epsilon'XP^{-2}X'\epsilon) \rightarrow 0$. Because of $E\epsilon_t^2 < \text{const}$ the left side is $o(twXP^{-2}X'\epsilon) = o(tr P^{-1}) = o(1/\lambda_{\text{min}}(P)) \rightarrow 0$. Hence with the Tchebycheff inequality $b \rightarrow \beta$ uniformly on $A \times F$.

(II) Necessity: If $b \rightarrow \beta$ uniformly on $A \times F$ then this holds
especially for $\varepsilon_t \sim N(0, \sigma_t^2)$, $t = 1, 2, \ldots$, the $\sigma_t$ lying between two positive constants. If $\mathbf{z}_i$ denotes the $i$-th row in $\mathbf{P}^{-1}$ then

$$b_i - \beta_i = \mathbf{z}_i^\top \mathbf{X} \varepsilon \sim N(0, \mathbf{z}_i^\top \mathbf{X} \Sigma \mathbf{X}^\top \mathbf{z}_i),$$

where $\Sigma$ is a diagonal matrix having $\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2$ as its diagonal. Now with two positive constants

$$\text{const } (\mathbf{P}^{-1})_{ii} < \text{var } (b_i - \beta_i) < \text{const } (\mathbf{P}^{-1})_{ii},$$

$(\mathbf{P}^{-1})_{ii}$ being the $i$-th diagonal element of $\mathbf{P}^{-1}$. This term as well as their sum for $i = 1, 2, \ldots, q$ which equals $\sum_i 1/\lambda_i(P)$, must tend to zero because of the consistency of $\mathbf{b}$. Hence $\lambda_{\min}(P) \to \infty$.

3. Some applications and remarks.

(I) As $\mathbf{P}_{ii} \geq \lambda_{\min}(P)$, $\mathbf{x}_j^2 \to \infty$ for all $j = 1, \ldots, q$, as $N \to \infty$. $\lambda_{\min}(P)$ is a non-negative, non-decreasing sequence in $N$.

(II) From the Gershgorin method we have $\lambda_{\min}(P) \to \infty$ if for all $i$

$$\mathbf{x}_i^2 - \sum_{j=1, j\neq i}^q |\mathbf{x}_j| \to \infty.$$
(II) The definition:

\[ \lambda_{\text{min}}(P) = \min_u (Xu)^2, \quad u^2 = 1, \]

which is helpful in many applications in this context, is used in the following example:

If \( x_{it} = t^c_i, \quad c_i = c_{i+1}, \) \textbf{(Polynomial regression)} then because of

\[ \sum_{t=1}^{N} t^c = o(N^{c+1}), \]

the order of \( Xu \) is determined by the first non-zero component of \( u \). The slowest growth of \( Xu \) is therefore \( \text{order} \)

\[ \min_u (Xu)^2 = o(N^{2c+1}), \quad u^2 = 1. \]

Hence consistent estimates are obtained, also for non integers \( c_i \), if and only if \[ \min_i c_i = c_q > \frac{1}{2}. \] Similarly, regression vectors may be treated in which exponentials, or exponentials plus polynomials occur.

(IV) One obtains consistent estimators in \textbf{trigonometric regression} of the kind

\[ x_{2i-1}, t = \cos \, w_{it}, \quad x_{2i}, t = \sin \, w_{it}, \quad i = 1, 2, \ldots, q, \]
if \( w_i \neq w_j \) and \( w_i \neq 2n - w_j \) for \( i \neq j \), further \( w_i < 2n \) for all \( i \). \( P \) tends in this case to a diagonal matrix whose non-zero elements are \( o(N) \).

(V) In the theorem it is required that \( F \) contains a d.f. vector of certain normal distributions. It is an interesting open question to ask how the restrictions for a subclass of \( F \) can be narrowed such that \( \lambda_{\min}(P) \rightarrow \infty \) is still a necessary condition. On the other hand a generalization of the theorem for certain classes of dependent errors seems to be easily possible.

(VI) It is also possible to give sufficient conditions for consistent LS-estimates in a linear model of time series analysis where regression with regard to lagged variables is included according to the equation:

\[
y_t = \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} + \beta_1 x_{1t} + \cdots + \beta_q x_{qt} + \varepsilon_t.
\]

The \( \varepsilon \)'s are here independently distributed, but need not be identical. The \( x_{it} \) are the exogenous regression variables. The theorem given in \( 2 \) is a certain specialization of the result for this general equation, from where our notation originates.

Acknowledgement: The author wishes to express his thanks to Professor Wassily Hoeffding and to Professor Harold Hotelling for critical remarks and suggestions.
Literature: