STRONG CONSISTENCY OF ESTIMATION OF NUMBER OF
REGRESSION VARIABLES WHEN T (U) PITTSBURGH UNIV PA
CENTER FOR MULTIVARIATE ANALYSIS Y WU JUN 87 TR-87-15
WU JUN 87 TR-87-15
Strong consistency of estimation of number of regression variables when the errors are independent and their expectations are not equal to each other.

Yuehua Wu

Consider the linear regression model

\[ y_i = x_i'\beta + e_i, \quad i = 1, 2, \ldots \]

where \( \{x_i\} \) is a sequence of known p-vectors, \( \beta' = (\beta_1, \ldots, \beta_p) \) is an unknown p-vector, known as regression coefficients, \( \{e_i\} \) is a sequence of random errors. It is of interest to test the hypothesis

\[ H_k: \beta_{k+1} = \ldots = \beta_p = 0, \quad k = 0, 1, \ldots, p. \]

We do not assume that the random errors are identically distributed and have zero means, since it is sometimes unrealistic. As a compensation for this relaxation, we assume the errors have a common bounded support \([a, a]\). Under certain
conditions, we obtain the strongly consistent estimate of the number of \( k \) for which \( \beta_k \neq 0 \) and \( \beta_{k+1} = \ldots = \beta_p = 0 \), by using the information theoretical criteria.
STRONG CONSISTENCY OF ESTIMATION OF NUMBER OF REGRESSION VARIABLES WHEN THE ERRORS ARE INDEPENDENT AND THEIR EXPECTATIONS ARE NOT EQUAL TO EACH OTHER*

Yuehua Wu
Center for Multivariate Analysis
University of Pittsburgh
STRONG CONSISTENCY OF ESTIMATION OF NUMBER OF REGRESSION VARIABLES WHEN THE ERRORS ARE INDEPENDENT AND THEIR EXPECTATIONS ARE NOT EQUAL TO EACH OTHER*

Yuehua Wu
Center for Multivariate Analysis
University of Pittsburgh

June 1987
Technical Report No. 87-15

*Research partially supported by the Air Force Office of Scientific Research (AFSC) under contract F49620-85-C-0008. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.
Abstract

Consider the linear regression model \( y_i = x_i' \beta + e_i, \quad i = 1, 2, \ldots \), where \( \{x_i\} \) is a sequence of known p-vectors, \( \beta = (\beta_1, \ldots, \beta_p) \) is an unknown p-vector, known as regression coefficients, \( \{e_i\} \) is a sequence of random errors. It is of interest to test the hypothesis \( H_k: \beta_{k+1} = \ldots = \beta_p = 0, \quad k = 0, 1, \ldots, p \).

We do not assume that the random errors are identically distributed and have zero means, since it is sometimes unrealistic. As a compensation for this relaxation, we assume the errors have a common bounded support \([a_1, a_2]\). Under certain conditions, we obtain the strongly consistent estimate of the number \( k \) for which \( \beta_k \neq 0 \) and \( \beta_{k+1} = \ldots = \beta_p = 0 \), by using the information theoretical criteria.
1. Introduction

Consider the linear model
\[ y_i = x_i \beta + e_i, \quad i = 1, 2, \ldots, n, \]
where \( x_i \)'s are experiment points, \( \beta = (\beta_1, \ldots, \beta_n)' \) is the regression coefficient vector to be estimated, and \( e_i \)'s are random errors. In the usual linear regression model it is assumed that the random errors have vanishing expectations and common variance. In this case, the famous least square estimation (LSE) method plays an important role in making statistical inference upon the regression coefficient vector \( \beta \). In the literature, there are a lot of papers concerning with the LSE and many important results are obtained (a part of work refers to [1],[2] and [3]). However the unbiasedness and consistency (even the weak one) of LSE strongly depend on the assumption that the expectations of errors are zero, and this assumption is not realistic sometimes. It is of interest to find a consistent estimates of the regression coefficients when the expectations of errors are not equal to each other. In [4] two methods for finding consistent estimates of the regression coefficient vector \( \beta \) are proposed.

The first method is to use the measure
\[ Q_n(\beta) = \max_{1 \leq i \leq n} (y_i - x_i \beta) - \min_{1 \leq i \leq n} (y_i - x_i \beta) \]
The estimator \( \hat{\beta}_n \) of \( \beta \) is defined as the vector which minimizes \( Q_n(\beta) \). The estimate \( \hat{\beta}_n \) is temporarily called MD estimate of \( \beta \) in [4] (the estimate based on the Maximum Difference between residuals).

The second method is to use the measure
\[ \tilde{Q}_n(\beta) = \max_{1 \leq i \leq n} |y_i - x_i \beta| \]
Denote by \( \tilde{\beta}_n \) the value of \( \beta \) which minimizes \( \tilde{Q}_n(\beta) \). Also, \( \tilde{\beta}_n \) is temporarily called MA estimate of \( \beta \) (the estimate based on the Maximum Absolute values of residuals).
Under certain conditions, both $\hat{\beta}_n$ and $\tilde{\beta}_n$ are shown to be strongly consistent in [4].

Now let us consider the hypotheses

\[ H_k: \beta_{k+1} = \beta_{k+2} = \ldots = \beta_p = 0 \text{ and } \beta_k \neq 0 \]
\[ k = 0, 1, \ldots, p-1. \]

It is of interest to determine the true model $H_k$ by using the model selection criteria. Denote by $\hat{\beta}_{kn} = (\hat{\beta}_{k1n}, \ldots, \hat{\beta}_{kkn}, 0, \ldots, 0)'$ the vector which minimizes $Q_n(\beta)$ under the restriction $\beta_{k+1} = \ldots = \beta_p = 0$ and denoted by $\tilde{\beta}_{kn} = (\tilde{\beta}_{k1n}, \ldots, \tilde{\beta}_{kkn}, 0, \ldots, 0)'$ the vector which minimizes $\tilde{Q}_n(\beta)$ under the restriction $\beta_{k+1} = \ldots = \beta_p = 0$. Write

\[ \hat{Q}_k = Q_n(\hat{\beta}_{kn}) \]

and

\[ \tilde{Q}_k = \tilde{Q}_n(\tilde{\beta}_{kn}) \]

Choose a sequence of constants $C_n$, satisfying certain conditions which will be specified later, and define

\[ \hat{R}_k = \hat{Q}_k + kC_n \]

and

\[ \tilde{R}_k = \tilde{Q}_k + kC_n \]

Choose
\[ \hat{k} = \text{ArgMin}\{R_k : k \in \{0, \ldots, p\}\} \]

and

\[ \tilde{k} = \text{ArgMin}\{R_k : k \in \{0, \ldots, p\}\} \]

where ArgMin denote the index which minimizes the quantities following the symbol ArgMin.

In this paper we shall consider the consistency of \( \hat{k} \) and \( \tilde{k} \) to the true model \( k_0 \).

2. Consistency of \( \hat{k} \)

In this section, we make the following general assumptions:

Assumption 1. The errors \( e_i, i = 1, 2, \ldots \) are independent.

Assumption 2. \( P\{e_n \in [a_1, a_2]\} = 0 \) and there is a positive constant \( \Delta \) such that for any \( \epsilon > 0 \) and any \( n \), we have

\[ P\{e_n \in [a_1, a_1 + \epsilon]\} \geq \Delta \epsilon \]

and

\[ P\{e_n \in [a_2 - \epsilon, a_2]\} \geq \Delta \epsilon. \]

Assumption 3. For any \( a > 0 \), there exists a positive constant \( C \) such that for any vector \( \alpha \neq 0 \) it follows that

\[ \#\{i \leq n, |\ell(x_i) - \ell(\alpha)| < a\} \geq Cn \]

for large \( n \), hereafter \( \ell(\alpha) = \alpha' |\alpha| \).
Assumption 4. There exists a positive constant m such that

\[ |x_i| > m, \quad \text{for } i = 1, 2, \ldots \]

Now let us estimate \( Q_n(\hat{\beta}) \). Define

\[ E_n^{(1)} = \{ i \leq n, -x'_i (\hat{\beta}_n - \beta) > 0 \} \]

\[ E_n^{(2)} = \{ i \leq n, x'_i (\hat{\beta}_n - \beta) > 0 \} \]

Split \( S_n = \{ x \in R^d : |x| = 1 \} \) into \( d \) disjoint parts \( \Sigma_1, \ldots, \Sigma_d \) such that \( \forall \ x, y \in \Sigma_j, xy > 3/4 \). Let \( \gamma_j \in \Sigma_j, j = 1, \ldots, d \). Define \( E_n^j = \{ i \leq n, x_i^j (\hat{\beta}_n - \beta) > 3/4 \}, j = 1, \ldots, d \). By Assumption 3, there exists \( \delta_n > 0 \) such that

\[ \#(E_n^j) \geq \delta_n n, \quad j = 1, 2, \ldots, d. \]

It is easy to see that \( -x_i' (\hat{\beta}_n - \beta) \in \Sigma_j \) and \( i \in E_n^j \) implies that

\[ -x_i' (\hat{\beta}_n - \beta) > 0, \quad \text{i.e. } i \in E_n^{(1)} \]

and that \( x_i' (\hat{\beta}_n - \beta) \in \Sigma_j \) and \( i \in E_n^j \) implies that

\[ x_i' (\hat{\beta}_n - \beta) > 0, \quad \text{i.e. } i \in E_n^{(2)} \]

Take \( r_n \) satisfying

\[ r_n \to 0 \quad \text{and} \quad nr_n / \log n \to \infty, \]

we have
\[ P \left( Q_n \left( \hat{\beta}_n \right) \leq a_2 - a_1 - 2r_n \right) \]

\[ \leq P \left( \max_{1 \leq j \leq d} \left( e_i \leq a_2 - r_n \right) \right) + P \left( \min_{1 \leq j \leq d} \left( e_i \geq a_1 + r_n \right) \right) \]

\[ \leq \sum_{j=1}^{d} P \left( \max_{1 \leq j \leq d} \left( e_i \leq a_2 - r_n, -\delta (\hat{\beta}_n - \beta) \in \Sigma_j \right) \right) \]

\[ + \sum_{j=1}^{d} P \left( \min_{1 \leq j \leq d} \left( e_i \geq a_1 + r_n, -\delta (\hat{\beta}_n - \beta) \in \Sigma_j \right) \right) \]

\[ \leq \sum_{j=1}^{d} P \left( \max_{1 \leq j \leq d} \left( e_i \leq a_2 - r_n \right) \right) \]

\[ + \sum_{j=1}^{d} P \left( \min_{1 \leq j \leq d} \left( e_i \geq a_1 + r_n \right) \right) \]

\[ \leq 2d (1-\Delta r_n) \delta \ln n \leq 2d e^{-\Delta r_n} \delta \ln n \leq 2d/n^2 \]

for large \( n \). By Borel-Cantelli Lemma we have

\[ Q_n \left( \hat{\beta}_n \right) \geq a_2 - a_1 - 2r_n, \quad a.s. \]

when \( n \) is large enough.

Let \( k_0 \) be the index of the true model and let \( \beta_0 \) be the true parameter. Then obviously we have \( \text{for } p \geq k > k_0 \).
\[ Q_n(\hat{\beta}_n) = Q_n(\hat{\beta}_{p_n}) \leq Q_n(\hat{\beta}_{kn}) \]
\[ \leq Q_n(\hat{\beta}_{k0n}) \leq Q_n(\hat{\beta}_0) \leq a_2 - a_1 \]

Thus

\[ 0 \leq Q_n(\hat{\beta}_{k0n}) - Q_n(\hat{\beta}_{kn}) \leq 2r_n, \quad p \geq k \geq k_0 \]

If we take \( C_n \) such that \( C_n \to 0, C_n/r \to \infty \), then for \( k > k_0 \)

\[ \hat{R}_k - \hat{R}_{k0} = (k - k_0)C_n + Q_n(\hat{\beta}_{kn}) - Q_n(\hat{\beta}_{k0n}) > 0, \quad (2) \]

for all large \( n \).

Next, we consider the case of \( k < k_0 \). Denote

\[ \eta = |\beta_{k0}| > 0 \]

and define

\[ \mathcal{E}_n^+ = \{ i \leq n, |\ell(x_i) + \ell(\hat{\beta}_{kn} - \beta_0) | < 1/2 \} \]
\[ \mathcal{E}_n^- = \{ i \leq n, |\ell(x_i) - \ell(\hat{\beta}_{kn} - \beta_0) | < 1/2 \} \]

Split \( S_p \) into \( b \) disjoint parts \( \Pi_1, \ldots, \Pi_b \) such that \( \forall x, y \in \Pi_i, |x - y| < 1/4 \). Let \( \xi_j \in \Pi_i, j = 1, \ldots, b \). Define
\[ f^j_n = \{ i \leq n, |\ell(x_i) - \xi_j | < 1/4 \}, \quad j = 1, \ldots, b. \]

By Assumption 3, there exists \( \delta_2 > 0 \) such that
\[ \#(F^j_n) \geq \delta_2 n, \quad j = 1, 2, \ldots, b. \]

It is easy to see that
\[ -\ell(\hat{\beta}_{kn} - \beta_0) \in \Pi_j \quad \text{and} \quad i \in F^j_n \]

which implies that
\[ |\ell(x_i) + \ell(\hat{\beta}_{kn} - \beta_0) | < 1/2, \quad \text{i.e.} \quad i \in E^+ \]

Also,
\[ \ell(\hat{\beta}_{kn} - \beta_0) \in \Pi_j \quad \text{and} \quad i \in F^j_n, \]

which implies that
\[ |\ell(x_i) + \ell(\hat{\beta}_{kn} - \beta_0) | < 1/2, \quad \text{i.e.} \quad i \in E^- \]

For \( i \in E^- \), we have
\[ x_i' (\hat{\beta}_{kn} - \beta_0) = |x_i| |\hat{\beta}_{kn} - \beta_0| \ell(x_i) \ell(\hat{\beta}_{kn} - \beta_0) \]
Similarly for \( i \in E_n^+ \), we have

\[ x'_i (\hat{\beta}_{kn} - \beta_0) \geq m\eta/2. \]

Hence

\[ Q_n (\hat{\beta}_{kn}) = \max_{i \in E_n^+} e_i - \min_{i \in E_n^+} e_i + m\eta \]

Thus

\[ P(Q_n (\hat{\beta}_{kn}) \leq a_2 - a_1 + m\eta/2) \]

\[ \leq P(\max_{i \in E_n^+} e_i \leq a_2 - m\eta/4) + P(\min_{i \in E_n^+} e_i \geq a_1 + m\eta/4) \]

\[ \leq \sum_{j=1}^{b} P(\max_{i \in E_n^+} e_i \leq a_2 - m\eta/4, -\lambda (\hat{\beta}_{kn} - \beta_0) \in \Pi_j) \]

\[ + \sum_{j=1}^{b} P(\min_{i \in E_n^+} e_i \geq a_1 + m\eta/4, \lambda (\hat{\beta}_{kn} - \beta_0) \in \Pi_j) \]

\[ \leq \sum_{j=1}^{b} P(\max_{i \in E_n^+} e_i \leq a_2 - m\eta/4, -\lambda (\hat{\beta}_{kn} - \beta_0) \in \Pi_j) \]

\[ + \sum_{j=1}^{b} P(\min_{i \in E_n^+} e_i \geq a_1 + m\eta/4, \lambda (\hat{\beta}_{kn} - \beta_0) \in \Pi_j) \]
\[ + \sum_{j=1}^{b} \mathbb{P}(\min_{i \in F_n^j} e_i \geq a_1 + m\eta/4, \ell (\hat{\beta}_{kn} - \beta_0) \in \Pi_j) \]
\[ \leq \sum_{j=1}^{b} \mathbb{P}(\max_{i \in F_n^j} e_i \leq a_2 - m\eta/4) \]
\[ + \sum_{j=1}^{b} \mathbb{P}(\min_{i \in F_n^j} e_i \geq a_1 + m\eta/4) \]
\[ \leq 2b(1 - \Delta m\eta/4) \delta_{2n} \leq 2b e^{-\Delta m\eta \delta_{2n}/4} \leq 2b/n^2 \]

for large \( n \). By Borel–Cantelli Lemma, we have, with probability one,

\[ Q_n(\hat{\beta}_n) \geq a_2 - a_1 + m\eta/2, \quad \text{for all large } n. \]

Thus for \( k < k_0 \), we have

\[ \hat{R}_k - \hat{R}_{k_0} = Q_n(\hat{\beta}_{kn}) - Q_n(\hat{\beta}_{k_0n}) - (k_0 - k) C_n \]
\[ \geq m\eta/2 - (k_0 - k) C_n > 0, \quad (3) \]

for large \( n \), since \( C_n \to 0 \).

(2) and (3) imply that \( \hat{k} \) is strongly consistent. Summarize the above arguments, we get the following theorem.

**Theorem 1.** Choose \( C_n \) satisfying
Suppose the four Assumptions given at the beginning of this section are true, then \( k \to k \), a.s.

Proof. Use the arguments given before. We only need to note that for any sequence of \( C_n \) satisfying (i) and (ii), we can always choose \( r_n \) such that

(i) \( \frac{r}{C_n} \to 0 \).

(ii) \( \frac{n r}{\log n} \to \infty \).

Q.E.D.

3. Consistency of \( \tilde{k} \)

In this section, we shall make the following general assumptions:

Assumption 1. The error \( e_i, i = 1, 2, \ldots \), are independent.

Assumption 2. \( |a_1| < a_2 \). \( \forall n \Rightarrow P(e_n \in [a_1, a_2]) = 0 \) there is a positive constant \( \Delta \) such that for any \( \epsilon > 0 \) and for any \( n \), we have

\[ P(e_n \in [a_2 - \epsilon, a_2]) \geq \Delta \epsilon. \]

Assumption 3: Same as Assumption 3 in Section 2

Assumption 4: There exists a positive constant \( \tilde{m} \) such that

\[ |x_i| > \tilde{m}, \quad \text{for } i = 1, 2, \ldots \]

Now let us estimate \( \bar{Q}_n(\tilde{S}_n) \). Define
\[ E_n = \{ i \leq n, x_i (\beta - \bar{\beta}_n) > 0 \} \]

Split \( S \) into \( d \) disjoint parts \( \tilde{\Sigma}_1, \ldots, \tilde{\Sigma}_d \) such that \( \forall x, y \in \tilde{\Sigma}_j, xy > 3/4 \).

Let \( \gamma_j \in \tilde{\Sigma}_j, j = 1, \ldots, d \). Define \( E^j_n = \{ i \leq n, \lambda(x_i) \gamma_j > 3/4 \}, j = 1, \ldots, d \). By Assumption 3, there exists \( \delta_1 > 0 \) such that

\[ \#(E^j_n) \geq \delta_1 n, \quad j = 1, \ldots, d \]

It is easy to see that

\[ -\lambda (\bar{\beta}_n - \beta) \in \tilde{\Sigma}_j \quad \text{and} \quad i \in E^j_n \]

imply that \( x_i \lambda (\beta - \bar{\beta}_n) > 0 \), i.e., \( i \in E_n \).

Take \( r_n \) satisfying

\[ r_n \to 0, \quad nr_n / \log n \to \infty \]

We have

\[
P(\tilde{Q}_n (\bar{\beta}_n) \leq a_2 - r_n) \leq P(\max_{i \in E_n} e_i \leq a_2 - r_n) - \sum_{j=1}^{d} P(\max_{i \in E_n} e_i \leq a_2 - r_n, \lambda (\beta - \bar{\beta}_n) \in \tilde{\Sigma}_j)
\]

\[
\leq \sum_{j=1}^{d} P(\max_{i \in E_n} e_i \leq a_2 - r_n, \lambda (\beta - \bar{\beta}_n) \in \tilde{\Sigma}_j)
\]

\[
\leq \sum_{j=1}^{d} P(\max_{i \in E_n^j} e_i \leq a_2 - r_n, \lambda (\beta - \bar{\beta}_n) \in \tilde{\Sigma}_j)
\]
\[
\begin{align*}
\sum_{j=1}^{d} P(\max_{i \in E_j} e_i \leq a_2 - r_n) \\
\leq d (1 - \delta) \delta n \leq d / n^2
\end{align*}
\]

for large \( n \). By Borel-Cantelli Lemma we have

\[
\tilde{Q}_n(\tilde{\beta}_n) \geq a_2 - r_n', \quad a.s.
\]

when \( n \) is large enough.

Let \( k_0 \) be the index of the true model and let \( \beta_0 \) be the true parameter. Then obviously we have for \( p > k > k_0 \)

\[
\tilde{Q}_n(\tilde{\beta}_n) = \tilde{Q}_n(\tilde{\beta}_{k_n}) \leq \tilde{Q}_n(\tilde{\beta}_{k_0}) \leq \tilde{Q}_n(\beta_0) \leq a_2
\]

Thus

\[
0 \leq \tilde{Q}_n(\tilde{\beta}_{k_0}) - \tilde{Q}_n(\tilde{\beta}_{k_n}) \leq r_n', \quad p > k \geq k_0.
\]

If we take \( C_n \) such that

\[
C_n \to 0, \quad C_n / r_n \to \infty
\]

then for \( k > k_0 \)
\[ \tilde{R}_k - \tilde{R}_{k_0} = (k - k_0) \mathcal{C}_n + Q_n(\tilde{\beta}_{kn}) - Q_n(\tilde{\beta}_{k_0n}) > 0 \]  

(4)

for all large \( n \).

Next, we consider the case of \( k < k_0 \). Denote

\[ \tilde{\eta} = |\tilde{\beta}_{k_0}| > 0 \]

and define

\[ \tilde{E}_n = \{ i \leq n, \quad \mathcal{L}(x_i) \mathcal{L}(\tilde{\beta}_{kn} - \beta_0) \leq -1/2 \} \]

Split \( S_p \) into \( \tilde{\Pi}_1, \ldots, \tilde{\Pi}_b \), such that \( \forall x, y \in \tilde{\Pi}_j, x'y \geq 1/2 \).

Let \( \tilde{\xi}_j \in \tilde{\Pi}_j, j = 1, \ldots, b \). Define \( \tilde{F}_n^j \) as \( \tilde{F}_n^j = \{ i \leq n, \quad \mathcal{L}(x_i) \tilde{\xi}_j \geq 275/280 \} \), \( j = 1, \ldots, b \). By Assumption 3, there exists \( \tilde{\delta}_2 > 0 \) such that

\[ \#(\tilde{F}_n^j) \geq \tilde{\delta}_2 n, \quad j = 1, \ldots, b \]

It is easy to see that \( -\mathcal{L}(\tilde{\beta}_{kn} - \beta_0) \in \tilde{\Pi}_j \) and \( i \in \tilde{F}_n^j \) imply that

\[ \mathcal{L}(x_i) \mathcal{L}(\tilde{\beta}_{kn} - \beta_0) \leq -1/2, \quad i.e. \ i \in \tilde{E}_n \]

For \( i \in \tilde{E}_n \), we have
\[ |x_i^{(i)} (\bar{\beta}_{kn} - \beta_0) | = |x_i| |\bar{\beta}_{kn} - \beta_0| |\ell(x_i) | \ell(\bar{\beta}_{kn} - \beta_0) | \geq \bar{m}\eta/2 \]

Hence

\[ \bar{Q}(\bar{\beta}_{kn}) \geq \max_{i \in \mathcal{E}_n} \bar{m}\eta/2 \]

Thus

\[ P(\bar{Q}(\bar{\beta}_{kn}) \leq a_2 + \bar{m}\eta/4) \]

\[ \leq P(\max_{i \in \mathcal{E}_n} \bar{m}\eta/4) \]

\[ \leq \sum_{j=1}^{b} P(\max_{i \in \mathcal{E}_n} \bar{m}\eta/4, -\ell(\bar{\beta}_{kn} - \beta_0) \in \bar{\Pi}_j) \]

\[ \leq \sum_{j=1}^{b} P(\max_{i \in \mathcal{E}_n} \bar{m}\eta/4, -\ell(\bar{\beta}_{kn} - \beta_0) \in \bar{\Pi}_j) \]

\[ \leq \sum_{j=1}^{b} P(\max_{i \in \mathcal{E}_n} \bar{m}\eta/4) \]

\[ \leq b \left( 1 - \Delta \bar{m}\eta/4 \right) \delta_2^n \leq b/n^2 \]

for large \( n \). By Borel-Cantelli Lemma, we have with probability one, when \( n \) large enough
\[ \tilde{Q}_n(\tilde{\beta}_{kn}) \geq a_2 + \bar{m} \bar{n}/4. \]

Thus for \( k < k_0 \), we have

\[ \tilde{R}_k - \tilde{R}_{k_0} = \tilde{Q}_n(\tilde{\beta}_{kn}) - \tilde{Q}_n(\tilde{\beta}_{k_0n}) - (k_0 - k)C_n \]

\[ \leq \bar{m} \bar{n}/4 - (k_0 - k)C_n > 0, \]

for large \( n \), since \( C_n \to 0 \).

(4) and (5) proves \( \bar{k} \) is consistent. Summarize the above arguments, we get the following theorem.

**Theorem 2.** Choose \( C_n \) satisfying

(i) \( C_n \to 0 \).

(ii) \( nC_n / \log n \to \infty \)

Suppose the four assumptions given at the beginning of this section are true, then \( \bar{k} \to k \), a.s.

Proof. Use the arguments given before, we only need to notice that for any sequence of \( C_n \) satisfying (i) and (ii), we can always choose \( r_n \) such that

(i) \( r_n / C_n \to 0 \)

(ii) \( n r_n / \log n \to \infty \)

Q.E.D.
4. General Case

In this section we consider the same regression model (1) but the problem we are going to solve is to determine the subset (or the model) \( J = \{1 \leq j_1 < \ldots < j_k \leq p\} \) such that \( \beta_j \neq 0 \) if and only if \( j \in J \) We make the same assumptions as given in previous sections.

Of course we can use the procedure described in section 2 to determine the model \( J \) as follows. For each permutation \( \pi \) of \( \beta = (\beta_1, \ldots, \beta_p) \), similarly rearranging \( (x_1, \ldots, x_p) \), we get a new model \( M_{\pi} \). Under this model, using the approach given in section 2 and 3, we obtain estimates \( \hat{k} = \hat{k}_{\pi} = \min \hat{k}_{\pi} \) and \( \bar{k} = \bar{k}_{\pi} = \min \bar{k}_{\pi} \) and let \( \hat{J}_1 = \{\hat{\pi}(1), \ldots, \hat{\pi}(k)\} \) and \( \bar{J}_1 = \{\bar{\pi}(1), \ldots, \bar{\pi}(k)\} \), we can easily prove that, by using Theorem 1 and 2, \( \hat{J}_1 \rightarrow J, \text{ a.s.} \) and \( \bar{J}_1 \rightarrow J, \text{ a.s.} \)

An alternative method to estimate \( J \) is given as follows. Suppose \( T \) is a subset of \( \{1, \ldots, p\} \). Consider the model \( T \):

\[
\gamma_n = x_n(T)'\beta(T) + e_n,
\]

where \( x_n(T) = (x_{i,T})_{j \in T} \) and \( \beta(T) = (\beta_j, j \in T) \). Let

\[
Q_n(T) = \min \left( \max \left( y_i - x_i(T)'\beta(T) \right) \right)_{\beta(T)} 1 \leq i \leq n
\]

\[
- \min \left( y_i - x_i(T)'\beta(T) \right) \left( y_i - x_i(T)'\beta(T) \right)_{\beta(T)} 1 \leq i \leq n
\]

and

\[
\bar{Q}_n(T) = \min \max |y_i - x_i(T)'\beta(T)|_{\beta(T)} 1 \leq i \leq n
\]

Define
\[ \hat{R}_T = Q_n(T) + \#(T)C_n \]

and

\[ \hat{R}_{T_2}^\sim = \tilde{Q}(T) + \#(T)C_n \]

Choose \( \hat{J}_2 \) such that

\[ \hat{R}_{T_2}^\sim = \min_{\hat{J}_2} \hat{R}_T \]

and choose \( \tilde{J}_2 \) such that

\[ \tilde{R}_{T_2}^\sim = \min_{\tilde{J}_2} \tilde{R}_T \]

We can also prove that \( \hat{J}_2 \rightarrow J \), a.s. and \( \tilde{J}_2 \rightarrow J \), a.s. However, there would be too much computation involved when \( p \) is relatively large. In the first case, there are totally \( p! \) permutations while in the second there are \( 2^p \) subsets of \( \{1, \ldots, p\} \).

In light of this, we propose another approach to estimate \( J \) which only involves \( p + 1 \) quantities to be computed.

Now let

\[ B(j) = (B_1, \ldots, B_{j-1}, 0, B_{j+1}, \ldots, B_p) \]

and define

\[ Q_n(j) = \min \left\{ \max_{\beta(j)} (y_i - x_i' \beta(j)) \right\}_{1 \leq i \leq n} \]

\[ - \min_{1 \leq i \leq n} (y_i - x_i' \beta(j)) \]

and
\[ \tilde{Q}_n(j) = \min_{\beta(j)} \max_{1 \leq i \leq n} |y_i - x_i^\top \beta(j)|. \]

Write

\[ \tilde{R}(n,j) = \tilde{Q}_n(j) - \tilde{Q}_p - C_n \]

and

\[ \tilde{R}(n,j) = \tilde{Q}_n(j) - \tilde{Q}_p - C_n. \]

We choose

\[ \hat{J}_n = \{j_1, \ldots, j_k\} = \{j: \tilde{R}(n,j) > 0\} \]

and

\[ \tilde{J}_n = \{\tilde{j}_1, \ldots, \tilde{j}_k\} = \{j: \tilde{R}(n,j) > 0\} \]

Then we have the following theorems.

**Theorem 3.** Under the conditions of theorem 1, we have that

\[ \hat{J}_n \to J, \text{ a. s.} \]

where model \( J = \{i_1, \ldots, i_k\} \) is the true one.

**Proof** If \( j \in J \), by (3) with the replacement that \( k_0 = p \) and \( k = p - 1 \), we have that with probability one, \( \tilde{R}(n,j) > 0 \) for all large \( n \). I.e., \( j \in \hat{J}_n \). Hence, when \( n \)
large enough, \( \hat{J}_n \not\in J \). Conversely, if \( j \not\in J \), using the same argument as proving theorem 1, we have

\[
\hat{R}(n,j) = Q_n(j) - \hat{Q}_p - C_n
\]

\[
\leq O(\log n/n) - C_n \quad \text{a.s.}
\]

which together with (iii) implies that

\[
\hat{R}(n,j) < 0, \quad \text{for large } n,
\]

i.e. \( j \not\in J \) when \( n \) large enough. Therefore \( \hat{J}_n = J \) which completes the proof of Theorem 3.

**Theorem 4.** Under the conditions of theorem 2, we have that

\[
\bar{J}_n = J, \quad \text{a.s.}
\]

where model \( J = \{j_1, \ldots, j_k\} \) is the true one.

**Proof.** If \( j \in J \), by (5) with the replacement that \( k_0 = p \) and \( k = p-1 \), we have that with probability one, \( \tilde{R}(n,j) > 0 \) for all large \( n \), i.e., \( j \in \bar{J}_n \). Hence, when \( n \) large enough, \( \bar{J}_n \not\in J \). Conversely, if \( i \not\in J \), using the same argument as proving theorem 2, we have

\[
\tilde{R}(n,j) = \tilde{Q}_n(j) - \tilde{Q}_p - C_n
\]

\[
\leq O(\log n/n) - C_n \quad \text{a.s.}
\]

which together with (ii) implies that
\[
\tilde{R}(n,j) < 0, \quad \text{for large } n,
\]

i.e., \( j \in \tilde{J}_n \) when \( n \) large enough. Therefore \( \tilde{J}_n \supseteq J \) which completes the proof of Theorem 4.
REFERENCES


REPORT DOCUMENTATION PAGE

<table>
<thead>
<tr>
<th>1. REPORT NUMBER</th>
<th>2. GOVT ACCESSION NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>87-15</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4. TITLE (and Subtitle)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong consistency of estimation of number of regression variables when the errors are independent and their expectations are not equal to each other.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>7. AUTHOR(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yuehua Wu</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>9. PERFORMING ORGANIZATION NAME AND ADDRESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Center for Multivariate Analysis</td>
</tr>
<tr>
<td>University of Pittsburgh, 515 Thackeray Hall</td>
</tr>
<tr>
<td>Pittsburgh, PA 15260</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>10. PROGRAM ELEMENT PROJECT, TASK AREA &amp; WORK UNIT NUMBERS</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>11. CONTROLLING OFFICE NAME AND ADDRESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air Force Office of Scientific Research</td>
</tr>
<tr>
<td>Department of the Air Force</td>
</tr>
<tr>
<td>Bolling Air Force Base, DC 20332</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>12. REPORT DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>June 1987</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>13. NUMBER OF PAGES</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>16. DISTRIBUTION STATEMENT (of this Report)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approved for public release; distribution unlimited</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>19. KEY WORDS (Continue on reverse side if necessary and identify by block number)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear model, model selection, regression coefficient, strong consistency</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>20. ABSTRACT (Continue on reverse side if necessary and identify by block number)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consider the linear regression model $y_i = x_i'\beta + e_i$, $i = 1, 2, \ldots$, where ${x_i}$ is a sequence of known p-vectors, $\beta' = (\beta_1, \ldots, \beta_p)$ is an unknown p-vector, known as regression coefficients, ${e_i}$ is a sequence of random errors. It is of interest to test the hypothesis $H_k: \beta_{k+1} = \ldots = \beta_p = 0$, $k = 0, 1, \ldots, p$. We do not assume that the random errors are identically distributed and have zero means, since it is sometimes unrealistic. As a compensation for this relaxation, we assume the errors have a common bounded support $[a_l, a_u]$. Under certain conditions, the estimates of the regression coefficients are strongly consistent. The conditions for this consistence are given in the report.</td>
</tr>
</tbody>
</table>
conditions, we obtain the strongly consistent estimate of the number of $k$ for which $\beta_k \neq 0$ and $\beta_{k+1} = \ldots = \beta_p = 0$, by using the information theoretical criteria.
END
DATE
FILMED
DEC.
1987