Asymptotic normality of minimum $L_1$-norm estimates in linear models

Consider the standard linear model $Y_i = x_i^T \beta + e_i, i = 1, \ldots, n, \ldots$ where $x_1, x_2, \ldots$ are assumed to be known $p$-vectors, $\beta$ the unknown $p$-vector of regression coefficients, and $e_1, e_2, \ldots$ the independent random error sequence each having a median zero. Define the Minimum $L_1$-Norm estimator.
20. (continued)

\( \hat{\beta}_n \) as the solution of the minimization problem

\[
\sum_{i=1}^{n} | Y_i - x_i^T \hat{\beta}_n | = \inf \sum_{i=1}^{n} | Y_i - x_i^T \beta | ; \beta \in \mathbb{R}^p
\]

It is proved in this paper that \( \hat{\beta}_n \) is asymptotically normal under very weak conditions. In particular, the condition imposed on \( \{x_i\} \) is exactly the same which ensuring the asymptotic normality of Least Squares estimate:

\[
\lim_{n \to \infty} \max_{1 \leq i \leq n} x_i (\sum_{j=1}^{n} x_j x_j^T)^{-1} x_i = 0.
\]
Center for Multivariate Analysis
University of Pittsburgh
ASYMPTOTIC NORMALITY OF MINIMUM L_1-NORM ESTIMATES IN LINEAR MODELS*

Z.D. Bai, X.R. Chen, Y. Wu and L.C. Zhao

Center for Multivariate Analysis
University of Pittsburgh
Technical Report No. 87-35

September 1987

Center for Multivariate Analysis
Fifth Floor Thackeray Hall
University of Pittsburgh
Pittsburgh, PA 15260

* This work is supported by Contract N00014-85-K-0292 of the Office of Naval Research and Contract F49620-85-C-0008 of the Air Force Office of Scientific Research. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.
ASYMPTOTIC NORMALITY OF MINIMUM L_1-NORM ESTIMATES IN LINEAR MODELS*

Z.D. Bai, X.R. Chen, Y. Wu and L.C. Zhao

ABSTRACT

Consider the standard linear model \( Y_i = x_i^* \beta + e_i, \ i = 1, \ldots, n, \ldots \) where \( x_1, x_2, \ldots \) are assumed to be known \( p \)-vectors, \( \beta \) the unknown \( p \)-vector of regression coefficients, and \( e_1, e_2, \ldots \) the independent random error sequence each having a median zero. Define the Minimum L_1-Norm estimator \( \hat{\beta}_n \) as the solution of the minimization problem \( \inf \{ \sum_{i=1}^{n} |Y_i - x_i^* \hat{\beta}_n| : \beta \in \mathbb{R}^p \} \). It is proved in this paper that \( \hat{\beta}_n \) is asymptotically normal under very weak conditions. In particular, the condition imposed on \( \{x_i\} \) is exactly the same which ensuring the asymptotic normality of Least Squares estimate: \( \lim_{n \to \infty} \max_{1 \leq i \leq n} x_i (\sum_{j=1}^{n} x_j^t x_j)^{-1} x_i = 0. \)

AMS 1980 subject classifications: Primary 62J05; secondary 62F12.

Key words and phrases: linear model, Minimum L_1-Norm estimate, consistency, asymptotic normality.

* This work is supported by Contract N00014-85-K-0292 of the Office of Naval Research and Contract F49620-85-C-0008 of the Air Force Office of Scientific Research. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.
1. INTRODUCTION AND SUMMARY

Consider the standard linear regression model

\[ Y_i = x_i' \beta_0 + e_i, \quad i = 1, \ldots, n \quad (1.1) \]

where \( x_1, x_2, \ldots \) are assumed to be known p-vectors, \( \beta_0 \) the unknown p-vector of regression coefficients, and \( e_1, e_2, \ldots \) the i.i.d. random errors with a common density function \( f \) and median zero, \( f \) is continuous at 0 and \( f(0) > 0 \).

The Minimum \( L_1 \)-Norm (ML\( _1 \)N) estimate \( \hat{\beta}_n \) of \( \beta_0 \) is defined as a solution of the

\[
\sum_{i=1}^{n} |Y_i - x_i' \hat{\beta}_n| = \inf \{ \sum_{i=1}^{n} |Y_i - x_i' \beta| : \beta \in \mathbb{R}^p \}.
\]

Here we assume that the parameter space is the whole p-dimensional Euclidean space \( \mathbb{R}^p \). It will be indicated (see Corollary 3 below) that no change in the argument is needed when \( \mathbb{R}^p \) is replaced by any of its subset containing the true parameter \( \beta_0 \) as an inner point.

The ML\( _1 \)N estimate, whose usefulness is by now universally recognized, dates back to Laplace. But for a long time in history it never attracted much attention. One reason is in the difficulty of its computation, which has now been resolved with the advent of modern computing facilities, and the paper of Charnes et al (1955) linking the computation of \( \hat{\beta}_n \) to the solution of a linear programming problem. Another reason is the lack of an adequate asymptotic theory. It is well known that in the problem of estimating the median of a univariate population, the sample median is (under certain conditions) asymptotic normal. Motivated by this simple case, write

\[
S_n = \sum_{i=1}^{n} x_i x_i'.
\]
It is naturally expected that \((\xrightarrow{D}\) means convergence in distribution)

\[
2f(0)S_n^{1/2}(\hat{\beta}_n - \beta_0) \xrightarrow{D} N(0, I_p), \quad \text{as } n \to \infty
\]  

(1.4)

under reasonable conditions. Here \(I_p\) is the identity matrix of order \(p\).

The first attempt to give a proof of (1.4) was made by Bassett et al (1978). They assumed that \(\{e_i\}\) satisfies the conditions stated earlier, the solution of (1.2) is unique (a condition difficult to justify), and that

\[
\frac{S_n}{n} \to Q, \quad \text{a positive definite matrix.}
\]  

(1.5)

Unfortunately their argument contains serious mathematical gaps which do not seem easy to resolve. For one thing, they overlooked the fact that the \(o(1)\) at the right-hand side of the equation above (3.10) of their paper should be \(o_h(1)\), and it is by no means clear that the convergence (as \(T \to \infty\)) \(o_h(1) \to 0\) should be uniform over \(h \in H\). Moreover, the assertion (3.9) is not generally valid. A simple counter-example is (in notations of their paper):

\[
Y_t = x_t \beta + u_t, \quad t = 1, 2, ..., \quad (\beta: \text{one-dimensional})
\]

\[
x_1 = 1/\sqrt{2}, \quad x_2 = 1 + \sqrt{2}/10, \quad x_3 = x_4 = ... = 1
\]

\[
u_1, u_2, ... \text{ i.i.d.}, \quad u_1 \sim N(0,1).
\]

It is easy to verify that all conditions, including the uniqueness assumption and nonlattice condition, are satisfied. But it can easily be shown that

\[
P_T\left(Z_T(S, h) \in C[0,1]\right) = 0
\]

for \(h = 1\) (which belongs to \(H = \{1, 2, 3, ..., T\}\) when \(T = 2, 4, 6, ..., \)) and (3.9) breaks down (see Appendix 1).

Bloomfield and Steiger (1983) advanced a proof of (1.4) under the assumption that \(x_1, x_2, ...\) are observations of a random vector \(x\) with a positive
definite covariance matrix, and \((x_1, Y_1), (x_2, Y_2), \ldots\) are stationary and ergodic. Unfortunately they failed to notice that for \(\{g_n(c)\}\) (defined by (6) on p.45 of their book) to be equicontinuous, \(g_n(c)\) must be defined as \(\sum_{i=1}^{n} h_n(r_i(c))/n\), and not \(\sum_{i=1}^{n} h_n(r_i(c))\) as in their book. But if \(g_n(c)\) is defined as \(\sum_{i=1}^{n} h_n(r_i(c))/n\), the assertion \(n^{-1}\left[D_n(c_n)\right] \to 2f(0)\zeta\) on p.47 should be \(\left[D_n(c_n)\right] \to 2f(0)\zeta\), and one can only obtain \(c_n - a_n \to 0\) in probability, not the crucial assertion (8) on p.46 of their book, and the proof breaks down. Besides, they made the mistake that the function \(h_n(t)\) defined on p.45 of their book has no second order derivative at \(t = \pm n^{-p}\), making the relation (12) on p.47 invalid.

Meanwhile Amemiya (1982) gave a proof of (1.4) by approximating the absolute value function with a twice-differentiable one. He made, in addition to (1.5), the assumption that \(\{x_i\}\) is a bounded sequence, and that \(\varepsilon_0\) is confined in a compact region. Unfortunately his proof, too, is invalid. One problem is that his assertions (in notations of his paper) \(A_1 \to 0\) in (3.12) and \(B_1 \to 0\) (in (3.22)) are both incorrect. Quite contrary, we have shown by simple arguments that actually \(A_1 \to 0\), \(B_1 \to 0\) in probability (see Appendix 2). Another crucial point is that in order to show \(T(\hat{g} - \hat{g}^*)' (\hat{g} - \hat{g}^*) \to 0\) in probability as \(T \to \infty\), (3.11) should be understood as sup\(|S(\hat{g}) - S(\hat{g})|: \hat{g} \in D\) \(\to 0\) in probability (\(D\) is the parameter space), while his argument, even freed from the error indicated above, is obviously not sufficient for this.

When the regression model contains a constant term: \(Y_i = \alpha_0 + x_i \beta_0 + \epsilon_i\), \(i = 1, 2, \ldots\), Bloomfield and Steiger (1983, p.62 Lemma 1) noticed the interesting fact that the ML\(_n\) estimator \(\hat{\beta}_n\) of \(\beta_0\) is in fact a special case of a class of rank estimators introduced by Jaeckel (1972). Jackel showed that his
estimator is asymptotically equivalent to an estimator introduced by Jurečková (1971). From this a proof of asymptotic normality of the ML₁N estimator \( \hat{\theta}_n \) can be obtained by using the theorem proved by Jurečková (1971). However, this does not give a satisfactory solution of the problem for the following two reasons: First, Jurečková’s theorem imposes very cumbersome conditions on the sequence \( \{x_i\} \) which are difficult to verify. Her theorem also requires the existence of Fisher information of the density \( f \) of the error, so \( f \) must be positive and absolute continuous on \( \mathbb{R} \). Even the simple uniform distribution \( \mathbb{R}(-1,1) \) does not meet this condition. Second, the theorem so obtained cannot deal with the case of (1.1) in which no constant term is present. If such a constant is present, the theorem cannot deal with this term.

Dupačová (1987) proved a theorem concerning the asymptotic normality of possible-constrained ML₁N estimates in case that \( \{x_i\} \) is a random sequence. Her theorem, when applied to the unconstrained case, gives roughly the result stated by Bloomfield and Steiger (1983), as mentioned earlier. There is a mathematically undesirable condition in her theorem: \( \|x_i\| \) possesses a finite moment of third order.

It is the purpose of this paper to give a rigorous proof of (1.4) under minimum conditions. First, in the i.i.d. case, we have the following theorem:

**THEOREM 1.** Suppose that in model (1.1), \( e_1, e_2, ... \) are independent and identically distributed with a common distribution function \( F \), and the following two conditions are satisfied.

1. There exists \( \Delta > 0 \) such that \( f(u) = F'(u) \) exists when \( |u| < \Delta \), \( f \) is continuous at 0, \( f(0) > 0 \) and \( F(0) = 1/2 \).
2. $S_n$ is nonsingular for some $n$, and

$$\lim_{n \to \infty} \max_{1 \leq i < n} \frac{1}{S_n} S_{n-1}^{-1} x_i = 0.$$  \hspace{1cm} (1.6)

Then (1.4) is true.

**Remark.** The condition (1.6) is exactly the same as that which guarantees the asymptotic normality of the Least Squares estimate of $\beta$ in case that $\{e_i\}$ is i.i.d. and $Ee_1 = 0, 0 < Ee_1^2 < \infty$. It was expected that the conditions ensuring the asymptotic normality of Minimum $L_1$-Norm estimate might be more stringent (as compared with the LS case), as the Minimum $L_1$-Norm estimate is nonlinear while the LS estimate is linear.

**COROLLARY 1.** If $\{e_i\}$ is i.i.d., condition 1 is satisfied, and there exists constant sequence $\{g_n\}$ such that $g_n \to \infty, g_{n+1}/g_n \to 1$, and

$$S_n/g_n \to A \text{ positive definite.}$$  \hspace{1cm} (1.7)

Then (1.4) is true.

Wu (1981) mentioned this condition in connection with the problem of consistency of LS estimates.

This corollary contains, as a special case, the result stated in Bassett and Koenker (1978). In turn it implies the following result:

**COROLLARY 2.** Suppose that $\{e_i\}$ is i.i.d., condition 1 is satisfied, and $x_1, x_2, \ldots$ are i.i.d. observations of a random vector $X$ such that $E(XX')$ is positive definite, $\{x_i\}$ and $\{e_i\}$ are independent. Then with probability one (for almost every sample sequence $\{x_i\}$), (1.4) is true.

Of course, we need not consider (1.4) as a conditional statement: it is also true unconditionally. Thus we reach the conclusion stated in Bloomfield and Steiger (1983).
COROLLARY 3. The Minimum $L_1$-Norm estimate $\hat{\beta}_n$ is weak consistent under the conditions of Theorem 1 (see the remark after Lemma 4).

This assertion follows from (1.4), and the fact that

$$S_n^{-1} \to 0.$$  \hfill (1.8)

For a proof of (1.8), fix $m$ such that $S_m$ is positive definite. For any positive integer $N$, denote by $\rho_{N1} \leq \rho_{N2} \leq \cdots \leq \rho_{NP}$ the eigenvalues of $S_N$. Then by a result of von Neumann (1937), we have

$$\text{tr}(S_mS_n^{-1}) \geq \sum_{i=1}^{p} \rho_{mi}/\rho_{ni}, \quad m \leq n.$$  \hfill (1.9)

But, by (1.6)

$$\text{tr}(S_mS_n^{-1}) = \sum_{j=1}^{m} \text{tr}(x_j x_j' S_n^{-1}) = \sum_{j=1}^{m} \text{tr}(x_j' S_n^{-1} x_j)$$

$$= \sum_{j=1}^{m} x_j' S_n^{-1} x_j \leq m \max_{1 \leq i \leq m} x_i' S_n^{-1} x_i \to 0, \quad \text{as } n \to \infty.$$  \hfill (1.10)

Since $\rho_{m1} > 0$, from (1.9) and (1.10), we have $\lim_{n \to \infty} \rho_{n1} = \infty$, and (1.8) is proved.

From Corollary 3 it follows that if we use a subset $G$ containing $\hat{\beta}_0$ as an inner point to replace $R^P$ in (1.2), and denote the resulting solution by $\hat{\beta}_n(G)$, we shall have $P(\hat{\beta}_n(G) \neq \hat{\beta}_n) \to 0$ as $n \to \infty$. Hence (1.4) is still true if $\hat{\beta}_n$ is replaced by $\hat{\beta}_n(G)$.

In passing we note that Y. Wu (1987) proved the strong consistency of $\hat{\beta}_n$ under conditions slightly stronger than those of Theorem 1. It does not seem possible to give a proof under the conditions of Theorem 1.

In practical applications there is usually a constant term in the re-
gression function, and instead of (1.1) we have the form
\[ Y_i = \alpha_0 + x_i^T \hat{\beta}_0 + e_i, \quad i = 1, \ldots, n, \ldots \quad (1.11) \]

Although (1.11), as a special case of (1.1), can be dealt with by Theorem 1, for inference purpose it will be convenient to have a theorem formulated in the following manner.

**THEOREM 2.** Write \((\hat{\alpha}_n, \hat{\beta}_n')\) the Minimum L1-Norm estimate of \((\alpha_0, \beta_0')\), and
\[ \bar{x}_n = (x_1 + \ldots + x_n)/n, \quad T_n = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}_n)(x_i - \bar{x}_n)', \quad (1.12) \]

Suppose that \(\{e_i\}\) is i.i.d., condition 1 of Theorem 1 is satisfied, \(T_n\) is nonsingular for some \(n\), and that
\[ \lim_{n \to \infty} \max_{1 \leq i < n} (x_i - \bar{x}_n)' T_n^{-1} (x_i - \bar{x}_n) = 0. \quad (1.13) \]

Then as \(n \to \infty\), we have
\[ 2f(0) T_n^{1/2} (\hat{\alpha}_n - \alpha_0) \xrightarrow{\mathcal{D}} N(0, 1) \quad (1.14) \]
\[ \frac{2f(0) \sqrt{n}}{\sqrt{1 + n \bar{x}' T_n^{-1} \bar{x}_n}} (\hat{\beta}_n - \beta_0) \xrightarrow{\mathcal{D}} N(0, 1). \quad (1.15) \]

Also, the two variables \(2f(0) T_n^{1/2} (\hat{\beta}_n - \beta_0)\) and \(2f(0) \sqrt{n} (\hat{\alpha}_n - \alpha_0) + \bar{x}' T_n^{-1} \bar{x}_n (\hat{\beta}_n - \beta_0)\) are asymptotically independent.

We note that the weak consistency of \(\hat{\alpha}_n\) and \(\hat{\beta}_n\) still holds true. For \(\hat{\beta}_n\) the assertion follows from (1.14) and \(T_n^{-1} \to 0\), which is a consequence of (1.13), in much the same way as (1.8) is a consequence of (1.6). For \(\hat{\alpha}_n\) the assertion follows from (1.15) and \(\bar{x}' T_n^{-1} \bar{x}_n \to 0\), which is a trivial consequence of (1.13) and the fact that \(T_n^{-1} \to 0\).
Corollary 2 can trivially be modified to accommodate the model (1.11). All we need to do is to replace the matrix \( E(XX') \) by the covariance matrix \( E(X - EX)(X - EX)' \).

The above two theorems can easily be extended to the case that \( e_1, e_2, ... \) are independent but not necessarily identically distributed. We shall only give a formulation of the following result.

**Theorem 3.** Suppose that in model (1.1), \( e_1, e_2, ... \) are independent, the distribution function \( F_i(x) \) is differentiable over an interval \((-\Delta, \Delta)\), \( F_i(0) = 1/2, i = 1, 2, ... \) and \( \Delta > 0 \) does not depend on \( i \). Write \( f_i(x) = F'_i(x) \).

Suppose that \( \{f_i(x)\} \) is equicontinuous at \( x = 0 \) and \( 0 < \inf_i f_i(0) \leq \sup_i f_i(0) < \infty \). Finally, suppose that (1.6) is true. Then as \( n \to \infty \), we have

\[
2S^{-1/2} \sum_{i=1}^{n} f_i(0)x_i x_i' (\hat{\beta}_n - \beta_0) \xrightarrow{d} N(0, I_n).
\]  

(1.16)

Our main task is to prove Theorem 1. Once this is achieved, only some trivial modifications are needed in proving Theorem 3, and much the same can be said about Theorem 2. To prove Theorem 1, it will be found convenient to reformulate the original problem in the following manner. Write

\[
\sum_{i=1}^{n} x_i x_i' = \sum_{i=1}^{n} x_{ni} x_{ni}' = n_I,
\]

\[
\sum_{i=1}^{n} x_{ni} = \sum_{i=1}^{n} x_{ni} = n_x,
\]

Then (1.1) has the form

\[
Y_{ni} = \beta_n + e_{ni}, \quad i = 1, ..., n, \quad n = 1, 2, ...
\]

with

\[
\sum_{i=1}^{n} x_{ni} x_{ni}' = I_p, \quad n = 1, 2, ...
\]  

(1.18)
Denote by $\hat{\beta}_n$ the Minimum $L_1$-Norm estimate of $\beta_{n0}$ in (1.17). Then we have the following general theorem, which includes Theorem 1 as a special case.

**THEOREM 1**'. Suppose that in model (1.17), $x_{n1}, \ldots, x_{nn}$ are known $p$-vectors satisfying (1.18), $e_{n1}, \ldots, e_{nn}$ are i.i.d. variables whose common distribution function $F$ does not depend on $n$ and satisfies the condition 1 of Theorem 1. Also, assume that

$$d_n \equiv \max_{1 \leq i \leq n} \|x_{ni}\| \to 0, \text{ as } n \to \infty$$

(1.19)

where $\| \cdot \|$ is the Euclidean norm in $\mathbb{R}^p$. Then as $n \to \infty$ we have

$$2f(0)(\hat{\beta}_n - \beta_{n0}) \xrightarrow{P} N(0, I_p).$$

(1.20)

This theorem will be proved in Section 3. Part of the reasoning is contained in Section 2 in the form of several preliminary lemmas.
2. SOME LEMMAS

LEMMA 1 (Bennett). Suppose that \( \xi_1, \ldots, \xi_n \) are independent, \( E\xi_i = 0, \)
\( |\xi_i| \leq b < \infty, \) \( i = 1, \ldots, n, \) where \( b \) is a constant. Write \( V = \sum_{i=1}^{n} \text{Var}(\xi_i)/n. \)
Then for any \( \varepsilon > 0, \) we have

\[
P(\sum_{i=1}^{n} \frac{\xi_i}{n} \geq \varepsilon) \leq 2 \exp\left(-n\varepsilon^2/(2V+2b)\right). \tag{2.1}
\]

For a proof, see Bennett (1962).

In model (1.17), we can assume that

\( \beta_{n0} = 0 \) \hspace{1cm} (2.2)

without any loss of generality. This we shall always do in the sequel, and we have \( Y_{ni} = e_{ni}. \) For any vector \( a = (a_1, \ldots, a_p)' \) in \( \mathbb{R}^p, \) write \( |a| = \max_{1 \leq i \leq p} |a_i|. \) \( I(A) \) will be used to denote the indicator of the set \( A. \)

LEMMA 2. Suppose that in model (1.17) the conditions of Theorem 1' are satisfied, with the possible exception of (1.19). Then we have

\[
I\left(\sum_{i=1}^{n} \text{sgn}(Y_{ni} - x_i ' \hat{\beta}_n) x_{ni} \mid \geq (p+1)d_n\right) \leq I(\hat{\beta}_n \mid \geq \Delta/\sqrt{pd_n}) \tag{2.3}
\]

with probability one, where \( \text{sgn}(0) = 0, \text{sgn}(a) = a/|a| \) for \( a \neq 0. \)

Proof. Since by definition \( \hat{\beta}_n \) is a minimization point of \( \sum_{i=1}^{n} |Y_{ni} - x_i ' \hat{\beta}_n| \) as a function of \( \beta \in \mathbb{R}^p, \) for any unit vector \( \varrho \in \mathbb{R}^p, \) we have by taking directional derivative

\[
- \sum_{i=1}^{n} \text{sgn}(Y_{ni} - x_i ' \hat{\beta}_n) x_i ' \varrho I(Y_{ni} \neq x_i ' \hat{\beta}_n) + \sum_{i=1}^{n} |x_i ' \varrho| I(Y_{ni} = x_i ' \hat{\beta}_n) \geq 0.
\]

This implies, in view of the arbitrariness of \( \varrho, \) that
\[ \sum_{i=1}^{n} \text{sgn}(Y_{ni} - x_{nk}^i \hat{\delta}_{n}) x_{ni} \leq \sum_{i=1}^{n} |x_{ni}| I(Y_{ni} = x_{ni}^i \hat{\delta}_{n}). \]  

(2.4)

Now suppose that \(|\hat{\delta}_{n}| < \Delta / (\sqrt{p} d_n)\). If \(1 \leq i_1 < i_2 < \ldots < i_{p+1} \leq n\) and

\[ Y_{nk} = x_{nk}^i \hat{\delta}_{n}, \quad k = i_1, \ldots, i_{p+1}. \]

Then find real constants \(c_1, \ldots, c_{p+1}\) not equaling to zero simultaneously such that

\[ c_1 x_{ni_1} + \ldots + c_{p+1} x_{ni_{p+1}} = 0. \]

We have

\[ c_1 Y_{ni_1} + \ldots + c_{p+1} Y_{ni_{p+1}} = 0. \]

(2.5)

Considering this, and the fact that \(|\hat{\delta}_{n}| < \Delta / (\sqrt{p} d_n)\) implies \(|x_{ni}^i \hat{\delta}_{n}| \leq \Delta\), we reach the following conclusion:

The event \( \left\{ \sum_{i=1}^{n} |x_{ni}| I(Y_{ni} = x_{ni}^i \hat{\delta}_{n}) \geq (p+1)d_n, \text{ and } |\hat{\delta}_{n}| < \Delta / (\sqrt{p} d_n) \right\} \)

\(\subset\) the event \( \bigcup_{1 \leq i_1 < \ldots < i_{p+1} \leq n} \left\{ \text{there exists constants } c_1, \ldots, c_{p+1} \right. \)

\(\text{depending only on } x_{ni_1}, \ldots, x_{ni_{p+1}} \), not all zero, such that

\[ \sum_{j=1}^{p+1} c_j Y_{ni_j} = 0, \text{ and } |Y_{ni_j}| < \Delta, \; j = 1, \ldots, p+1. \]

(2.6)

Since \(Y_{ni} = e_{ni_1}, e_{ni_2}, \ldots, e_{nn}\) are i.i.d. variables whose common distribution function is continuous over \((-\Delta, \Delta)\), it is seen that the probability of the event on the right-hand side of (2.6) is zero. This fact, taken together with (2.4), implies (2.3), and the lemma is proved.

In order to introduce the crucial Lemma 3, we have to define some notations.

By (1.19) it follows that there exists constant sequence \(\{v_n\}\) of positive
even integers such that as $n \to \infty$,
\[ u_n \to \infty, \quad u_n^2 d_n \to 0, \quad \sqrt{p} u_n^2 d_n < 1/2, \quad \text{for } n \text{ large.} \quad (2.7) \]

In the following we assume that the conditions of Theorem 1' are met. Write
\[ M = M_n = \lfloor \log \frac{n}{\log u_n} \rfloor, \quad r_m = u_n^m, \quad m = 1, \ldots, M, \quad r_{M+1} = n. \quad (2.8) \]
\[ D = D_n = \{ \beta = (\beta_1, \ldots, \beta_p)^\top: -u_n \leq \beta_i < u_n, \quad i = 1, \ldots, p \}. \quad (2.9) \]

Note that $M \geq 1$ for $n$ large. This is true because by (1.18) we have
\[ n d_n^2 \geq 1, \quad \text{therefore } u_n < n^{1/4}. \]
Partition $D$ into a number of intervals $\tilde{D}_1, \tilde{D}_2, \ldots, \tilde{D}_j$, each having the form \( \{x = (x(1), \ldots, x(p))^\top: a_i \leq x(i) < b_i, \quad i = 1, \ldots, p\} \), such that
\[ J_1 \leq u_n^{2p}, \quad L(\tilde{D}_i) \leq 1, \quad \tilde{D}_i \cap \tilde{D}_j = \emptyset, \quad i, j = 1, \ldots, J_1, \quad i \neq j \]
where \( L(A) \) is defined as \( \sup\{ |u - v|: u \in A, v \in A \} \). Now each subset $\tilde{D}_k$ is
again partitioned into a number of disjoint intervals $\tilde{D}_{k1}, \tilde{D}_{k2}, \ldots, \tilde{D}_{kJ_2}$
which can further be partitioned. This process is defined inductively as
follows: Suppose that after the $(m-1)$-th round we have partitioned $D$ into
\[ \{\tilde{D}_{j_1} \ldots j_{m-1}\} \]
then in the $m$-th round we take each $\tilde{D}_{j_1} \ldots j_{m-1}$ and partition
it into a number of disjoint intervals $\{\tilde{D}_{j_1} \ldots j_{m-1}^\varepsilon: \varepsilon = 1, 2, \ldots, J_m\}$, such that
\[ J_m \leq u_n^{2p}, \quad L(\tilde{D}_{j_1} \ldots j_{m-1}^\varepsilon) \leq u_n^{-2(m-1)}, \quad \varepsilon = 1, \ldots, J_m \quad (2.10) \]
The process ends with the completion of the $(M+1)$-th round. Denote
\[ G_m = \{\tilde{j}_1 \ldots j_m\} = \text{the partitioning of } D \text{ after the } m \text{-th round} \]
\[ m = 1, 2, \ldots, M+1. \quad (2.11) \]
A typical element of $G_m$ is denoted by $B$ or $\tilde{D}_{j_1 \ldots j_m}$, and a chosen point in it is denoted by $b$ or $b_{j_1 \ldots j_m}$. Put

$$
\psi_{ni}(\theta) = \text{sgn}(e_{ni} - x_{ni}' \theta) - \text{sgn}(e_{ni}) \tag{2.12}
$$

$$
\lambda_{ni}(\theta) = 2(F(x_{ni}' \theta) - 1/2) = -\psi_{ni}(\theta) \tag{2.13}
$$

$$
t_{ni}(\theta) = \psi_{ni}(\theta) + \lambda_{ni}(\theta). \tag{2.14}
$$

By (2.7), when $\theta \in D$ and $i \leq n$, we have, for large $n$,

$$
|x_{ni}' \theta| \leq \sqrt{p} d_n |\theta| < (2\nu^{-1}n) \tag{2.15}
$$

and for $b \in B$, $b^* \in B$ where $B \in G_m$, we have, for large $n$,

$$
|x_{ni}'(b-b^*)| \leq \sqrt{p} d_n |b-b^*| < (2\nu_n^{-2m})^{-1}, \quad 1 \leq i \leq n. \tag{2.16}
$$

From (2.15) and the conditions imposed on the distribution function $F$ (see condition 1 of Theorem 1), it follows that we can find a constant $C > 1$ such that for $\theta \in D$ and $1 \leq i \leq n$:

$$
2(F(x_{ni}' \theta + \mu_n^{-2m}) - F(x_{ni}' \theta - \mu_n^{-2m})) \leq C \mu_n^{-2m} \tag{2.17}
$$

$$
|\lambda_{ni}(\theta)| \leq C \nu_n^{-1} \tag{2.18}
$$

$$
P(|e_{ni}| \leq \sqrt{p} d_n \nu_n) \leq C d_n \nu_n. \tag{2.19}
$$

Let $\{a_{ni}, 1 \leq i \leq n, n = 1,2,\ldots\}$ be a triangular array of real numbers, satisfying

$$
\sum_{i=1}^{n} a_{ni}^2 = 1, \quad |a_{ni}| \leq \mu_n^{-m/2} \quad \text{for} \quad i \geq r_m + 1 \tag{2.20}
$$

$$
n = 1,2,\ldots, \quad m = 1,2,\ldots, M+1.\]
Define the following quantities:

\[
Q_m = \sum_{\mathcal{B} \in G_m} \left\{ \sup_{\mathcal{B} \in \mathcal{B}} \left\{ \sum_{i=r_m+1}^{n} a_{ni}(t_{ni}(e) - t_{ni}(b)) \right\} \geq \varepsilon 3^{-m} \right\},
\]

\[1 \leq m \leq M + 1,\]  \hspace{1cm} (2.21)

\[
U_m = \sum_{\mathcal{B} \in G_m} \left\{ \sup_{\mathcal{B} \in \mathcal{B}} \left\{ \sum_{i=r_m+1}^{r_{m+1}} a_{ni}(t_{ni}(e) - t_{ni}(b)) \right\} \geq \varepsilon 3^{-(m+1)} \right\} \equiv \sum_{\mathcal{B} \in G_m} U(m,\mathcal{B}),
\]

\[1 \leq m \leq M,\]  \hspace{1cm} (2.22)

\[
V_m = \sum_{j_1, \ldots, j_{m+1}} \left\{ \left\| \sum_{i=r_m+1}^{n} a_{ni}(t_{ni}(b_{j_1 \ldots j_m}) - t_{ni}(b_{j_1 \ldots j_m j_{m+1}})) \right\| \geq \varepsilon 3^{-(m+1)} \right\}
\]

\[\equiv \sum_{j_1, \ldots, j_{m+1}} V(m,j_1,\ldots,j_{m+1}), \quad 1 \leq m \leq M, \quad V_M = 0.\]  \hspace{1cm} (2.23)

In the above expressions \( \varepsilon \in (0,1), r_m, m = 1, \ldots, M, M+1, \) has been defined in (2.8). Further, in the definition of \( V_m, \) the summation runs over all such \((j_1, \ldots, j_{m+1})\) that \( D_{j_1 \ldots j_{m+1}} \) is a member of \( G_{m+1}, \) and \( b_{j_1 \ldots j_m} \) is understood as the point chosen in \( D_{j_1 \ldots j_m} \in G_t, \) while in the definition of \( Q_m \) and \( U_m, \) \( b \) is the point chosen in a member \( B \) of \( G_m, \) as stated earlier.

By these definitions it is easily seen that

\[
Q_m \leq U_m + V_m + Q_{m+1}, \quad m = 1, \ldots, M.\]  \hspace{1cm} (2.24)

Therefore, on noticing that \( Q_{M+1} = 0, \) we have

\[
Q_1 \leq \sum_{m=1}^{M} (U_m + V_m).\]  \hspace{1cm} (2.25)
LEMMA 3. Suppose that the conditions of Theorem 1' are met, and \( \{a_{ni}\} \) satisfies (2.20), \( \epsilon \in (0,1) \). Then when \( n \) is large we have

\[
U_m + V_m \leq 4 \nu_n^{2p_m} e^{(-48-\epsilon)(9^{-1} \nu_n)^{m/2}}, \quad m = 1, \ldots, M
\]  

where \( C \) is the constant appearing in (2.17)-(2.19).

Proof. Define

\[
\eta_{ni} \equiv \eta_{ni}(B) = \begin{cases} 
2, \text{ when } |e_{ni} - x_{ni}^i b| < \nu_n^{-2m} \\
0, \text{ otherwise}
\end{cases}
\]  

where \( B \in G_m \) and \( b \) is the point chosen in \( B \). By (2.14), (2.16) and (2.17), we have

\[
\sup_{\beta \in B} \left| \sum_{i=r_m+1}^{r_{m+1}} a_{ni} (t_{ni}(\beta) - t_{ni}(b)) \right| 
\leq \sum_{i=r_m+1}^{r_{m+1}} |a_{ni}| \eta_{ni} - E_{ni} + 2 \sum_{i=r_m+1}^{r_{m+1}} |a_{ni}| E_{ni}.
\]  

From (2.16), (2.17) and (2.27), we have

\[
E_{ni} \leq C \nu_n^{-2m}, \quad i = r_{m+1}, \ldots, r_{m+1}.
\]  

By (2.20), we have for \( n \) large

\[
2 \sum_{i=r_m+1}^{r_{m+1}} |a_{ni}| E_{ni} \leq 2 \left( \sum_{i=r_m+1}^{r_{m+1}} (E_{ni})^2 \right)^{1/2} \leq 2 \left( \nu_n^{-m} C \nu_n^{-4m} \right)^{1/2} \leq 2 \nu_n^{-m} e^{-1/2} (m+1).
\]  

From (2.1), (2.28), (2.30) and by Lemma 1, we obtain
\[ U(m,B) \leq P \left\{ \sum_{i=r_m+1}^{r_{m+1}} |a_{ni}| \leq \varepsilon 2^{-1} 3^{-(m+1)} \right\} \]

\[ \leq 2 \exp \left\{ -\varepsilon 2^{-2} 3^{-(2m+2)} \left[ \left( \max_{r_m < i \leq r_{m+1}} E_n n_i \right)^2 + \varepsilon 3^{-(m+1)} \max_{1 > r_m} |a_{ni}| \right] \right\} \]  

(2.31)

U(m,B) is defined in (2.22). From (2.17), (2.20) and (2.29), we have, for large n,

\[ \max_{1 > r_m} |a_{ni}| \leq \nu_n^{-m/2} \]

\[ \max_{r_m < i \leq r_{m+1}} E_n n_i \leq 2C \nu_n^{-2m} \leq \nu_n^{-m/2} 3^{-(m+1)} \]

Therefore

\[ U(m,B) \leq 2 \exp(-\varepsilon 3^{-(m+1)} \nu_n^{m/2}/16) \]  

(2.32)

which implies

\[ U_m \leq 2 \nu_n^{2m} \exp(-\varepsilon 3^{-(m+1)} \nu_n^{m/2}/16). \]  

(2.33)

By (2.20),

\[ 4 \max_{i > r_m+1} |a_{ni}| \leq 4 \varepsilon 3^{-(m+1)} \nu_n^{-m+(m+1)/2}. \]  

(2.34)

By (2.16) and (2.17), we have, on denoting \( g_{ni} = t_{ni}(b_{j_1}\ldots j_m) - t_{ni}(b_{j_1}\ldots j_{m+1}) \), that

\[ \varepsilon g_{ni}^2 \leq 4 \left| F(x_{ni}^b b_{j_1}\ldots j_m) - F(x_{ni}^b b_{j_1}\ldots j_{m+1}) \right| \]

\[ \leq 2C \nu_n^{-2m} \leq 4 \varepsilon 3^{-(m+1)} \nu_n^{-(m+1)/2}. \]  

(2.35)
Here $b_{j_1 \ldots j_m}$, $b_{j_1 \ldots j_{m+1}}$ were explained following the definition of $V_m$.

From (2.23), (2.34), (2.35), and Bennett inequality (2.1), we have

$$V(m, j_1, \ldots, j_m) \leq 2\exp\left\{-\varepsilon^2 - (2m + 2)\sqrt{2 \left( \max_{1 \leq i \leq m+1} |a_{ni}| + \max_{i > r_{m+1}} |c_{ni}| \right)} \right\}$$

$$\leq 2\exp\left\{-16^{-1}(\mu_n/9)(m+1)/2 \right\}$$

which in turn implies

$$V_m \leq 2^2p^{(m+1)} \exp\{-16^{-1}(\mu_n/9)(m+1)/2\} \leq 2^{2p} \exp\{-16^{-1}(\mu_n/9)\}

Finally, for $n$ large, (2.26) follows from (2.33) and (2.36). Lemma 3 is proved.

**Lemma 4.** Under the assumptions of Theorem 1', we have

$$\lim_{n \to \infty} P(\|s_n\| > v_n) = 0. \quad (2.37)$$

for any constant sequence $\{v_n\}$ such that $\lim_{n \to \infty} v_n = \infty$. (Here we assume (2.2).)

**Proof.** Without losing generality we may assume $d_n v_n^2 \leq 1$. Define

$$\bar{D} \equiv \bar{D}_n \equiv \{\bar{e} = (\bar{e}_1, \ldots, \bar{e}_p)\} : -v_n \leq \bar{e}_i \leq v_n, \quad i = 1, \ldots, p$$

$$\phi_{ni}(\bar{e}) = |e_{ni} - |e_{ni} - x_{ni}^{\prime} \bar{e}|$$

$$\Lambda_{ni}(\bar{e}) = E(\phi_{ni}(\bar{e}))$$

$$R_{ni}(\bar{e}) = \phi_{ni}(\bar{e}) - \Lambda_{ni}(\bar{e}), \quad i = 1, \ldots, n; \quad n = 1, 2, \ldots .$$

The first step is to verify that

$$\lim_{n \to \infty} v_n^2 \sup_{\bar{e} \in \bar{D}} \left| \sum_{i=1}^{n} R_{ni}(\bar{e}) \right| = 0, \quad \text{in probability} \quad (2.38)$$
In order to do this, we partition the interval $\tilde{D}$ in exactly the same manner as we have previously done for $D$ defined by (2.9), with $u_n$ replaced by $v_n$, and that instead of (2.8), we now define $M$ as a positive integer satisfying

$$\sqrt{np} v_n^{-2M} < 2^{-1} \varepsilon^{-3(M+1)} v_n^2$$  \quad (2.39)$$

where $\varepsilon > 0$ is an arbitrarily given constant. The existence of such $M$ follows from the fact that $v_n \to \infty$. Also, the partitioning of $\tilde{D}$ after the $m$-th round will be denoted by $\tilde{G}_m$. A typical interval belonging to $\tilde{G}_m$ will be denoted by $B$ or $B_{j_1, \ldots, j_m}$, and a point selected from it by $b$ or $b_{j_1, \ldots, j_m}$. Define

$$\tilde{V}_0 = \sum_{j_1} P(\sum_{i=1}^n |R_{ni}(b_{j_1})| \geq \varepsilon v_n^2/3)$$

$$\tilde{V}_m = \sum_{j_1, \ldots, j_{m+1}} P\left(\sum_{i=1}^n |R_{ni}(b_{j_1} \ldots j_m) - R_{ni}(b_{j_1} \ldots j_{m+1})| \geq \varepsilon v_n^2/3^{m+1}\right)$$

$$m = 1, 2, \ldots, M;$$

$$\tilde{Q}_m = \sum_{j_1 \ldots j_m} P\left(\sup_{b \in B_{j_1 \ldots j_m}} |\sum_{i=1}^n (R_{ni}(b) - R_{ni}(b_{j_1} \ldots j_m))| \geq \varepsilon v_n^2/3^{m}\right)$$

$$m = 1, 2, \ldots, M+1.$$
which implies that

\[ \tilde{Q}_{M+1} = 0. \]  (2.40)

It is easy to see that

\[ \tilde{Q}_m \leq \tilde{V}_m + \tilde{Q}_{m+1}, \quad m = 1, \ldots, M, \]  (2.41)

\[ P\left( v_n^{-2} \sup_{\beta \in D} \left| \sum_{i=1}^{\infty} R_{n_i}(\varepsilon) \right| \geq \varepsilon \right) \leq \tilde{V}_0 + \tilde{Q}_1. \]  (2.42)

From (2.40)-(2.42), it follows that

\[ P\left( v_n^{-2} \sup_{\beta \in D} \left| \sum_{i=1}^{\infty} R_{n_i}(\varepsilon) \right| \geq \varepsilon \right) \leq \sum_{m=0}^{M} \tilde{V}_m. \]  (2.43)

Since

\[ |\phi_n(b_{j_1})| \leq |x'_{ni} b_{j_1}| \leq d_n v_n \leq v_n^{-1}, \quad \text{since} \quad d_n v_n \leq 1. \]

\[ \sum_{i=1}^{\infty} (\text{var}(\phi_n(b_{j_1}) \leq \sum_{i=1}^{\infty} (x'_{ni} b_{j_1})^2 = \|b_{j_1}\|^2 \leq p v_n^2, \]

\[ |\phi_n(b_{j_1 \ldots j_m} - \phi_n(b_{j_1 \ldots j_{m+1}}) \leq |x'_{ni} (b_{j_1 \ldots j_m} - b_{j_1 \ldots j_{m+1}})| \]

\[ \leq \|x'_{ni}\| \|b_{j_1 \ldots j_m} - b_{j_1 \ldots j_{m+1}}\| \]

\[ \leq p |x'_{ni}| \|b_{j_1 \ldots j_m} - b_{j_1 \ldots j_{m+1}}\| \]

\[ \leq p d_n v_n^{-2m+2} \leq p v_n^{-2m}, \]

and
\[
\sum_{i=1}^{n} \text{Var}(\phi_{ni}(b_{j_1} \ldots j_m) - \phi_{ni}(b_{j_1} \ldots j_{m+1})) \leq \sum_{i=1}^{n} (x'_{ni}(b_{j_1} \ldots j_m - b_{j_1} \ldots j_{m+1}))^2
\]
\[
= \|b_{j_1} \ldots j_m - b_{j_1} \ldots j_{m+1}\|^2
\]
\[
\leq p\|b_{j_1} \ldots j_m - b_{j_1} \ldots j_{m+1}\|^2
\]
\[
\leq pv_n^{-4m+4}.
\]

Applying Lemma 1, we get
\[
\tilde{V}_0 \leq 2v_n^{2p} \exp(-3^{-2} \epsilon 2v_n^4/(2v_n^2 + \epsilon v_n)) \leq 2v_n^{2p} \exp(-cv_n^2). \tag{2.44}
\]
\[
\tilde{V}_m \leq 2v_n^{2p(m+1)} \exp(-\epsilon 3^{-m-1} v_n^2)/(2v_n^{-4m+4} + \epsilon 3^{-m-1} v_n^{-2m+2}) \leq 2v_n^{2p(m+1)} \exp(-c3^{-(m+1)} v_n^{2m+2}), \quad m \geq 1. \tag{2.45}
\]

Here \(c > 0\) is a constant independent of \(n, m\). From (2.43)-(2.45), we obtain
\[
P(\sum_{i=1}^{n} R_{ni}(\tilde{e}) > \epsilon) \leq 2 \sum_{m=0}^{\infty} v_n^{2p(m+1)} \exp(-c3^{-(m+1)} v_n^{2m+2}). \tag{2.46}
\]

Since the right-hand side of (2.46) tends to zero as \(n \to \infty\), we obtain (2.38).

Now we note that
\[
\sup_{\tilde{e} \in D} \max_{1 \leq i \leq n} |x'_{ni} \tilde{e}| \leq d_n v_n \leq v_n^{-1} \to 0, \quad \text{as} \quad n \to \infty. \tag{2.47}
\]

Hence, considering condition 1 of Theorem 1, we have
\[
E\Phi_{ni}(\tilde{e}) = \begin{cases} 
\int_{x'_{ni} \tilde{e} - 2u}^{x'_{ni} \tilde{e}} (2u - 2x'_{ni} \tilde{e}) f(u) du, & \text{if } x'_{ni} \tilde{e} > 0 \\
0 & \text{if } x'_{ni} \tilde{e} = 0 \\
\int_{x'_{ni} \tilde{e} - 2u}^{x'_{ni} \tilde{e}} (2x'_{ni} \tilde{e} - 2u) f(u) du, & \text{if } x'_{ni} \tilde{e} < 0 
\end{cases}
\]
\[
= -f(0)(x'_{ni} \tilde{e})^2(1 + o(1)) \tag{2.48}
\]
for \( \theta \in \theta \), \( i = 1, \ldots, m, n \) sufficiently large, where \( o(1) \to 0 \) as \( n \to \infty \) uniformly for \( \theta \in \tilde{\theta} \) and \( 1 \leq i \leq n \), in view of (2.47). From (2.38) and (2.48), we get

\[
\lim_{n \to \infty} v_n^{-2} \sup_{\theta \in \tilde{\theta}} \left| \frac{1}{n} \sum_{i=1}^{n} \left( |e_{ni}| - |e_{ni} - x_i' \theta| \right) + f(0) \| \theta \|^2 \right| = 0, \text{ in probability}
\]

which implies that

\[
v_n^{-2} \left( \frac{1}{n} \sum_{i=1}^{n} |e_{ni}| - \inf_{|\theta| = v_n} \frac{1}{n} \sum_{i=1}^{n} |e_{ni} - x_i' \theta| \right) \leq -f(0)(1 + o_p(1)) \tag{2.49}
\]

where \( o_p(1) \) tends to zero in probability as \( n \to \infty \). Since \( \sum_{i=1}^{n} |e_{ni} - x_i' \theta| \) as a function of \( \theta \) is convex over \( \mathbb{R}^p \), (2.49) implies

\[
v_n^{-2} \left( \frac{1}{n} \sum_{i=1}^{n} |e_{ni}| - \inf_{|\theta| \geq v_n} \frac{1}{n} \sum_{i=1}^{n} |e_{ni} - x_i' \theta| \right) \leq -f(0)(1 + o_p(1)). \tag{2.50}
\]

Since \( f(0) > 0 \), (2.50) implies (2.37), and Lemma 4 is proved.

**Remark.** Let us return temporarily to model (1.1) and consider the ML\(_1\)N estimate \( \hat{\theta}_n \) defined by (1.2). It follows easily from Lemma 4 and (1.8) that \( \hat{\theta}_n \) is a weakly consistent estimate of \( \theta_0 \). For convenience of presentation we formulated this fact as a corollary of Theorem 1. Now we see that the verification of this fact is, in fact, an important step in the proof of Theorem 1.
3. PROOF OF THEOREMS

Proof of Theorem 1. As mentioned earlier, we need only to prove
Theorem 1'. We begin with verifying that as $n \to \infty$,

$$\sup\{\| \sum_{i=1}^{n} t_n(\xi) x_n \| : \xi \in \tilde{D}\} \to 0, \quad \text{in probability.} \quad (3.1)$$

$D$ and $t_n(\xi)$ were defined by (2.9) and (2.12)-(2.14), respectively. In
view of (1.18), we need only to prove that as $n \to \infty$,

$$Q_0 = P\left( \sup\{ \sum_{i=1}^{n} a_{ni} t_n(\xi) : \xi \in D \} \geq \epsilon \right) \to 0 \quad (3.2)$$

where $(a_{ni})$ is an arbitrarily given constant array satisfying $\sum_{i=1}^{n} a_{ni}^2 = 1$, and $\epsilon \in (0,1)$, also arbitrarily given.

Without loss of generality, assume $|a_{ni}| \geq \ldots \geq |a_{nn}|$. Choose $\nu_n$ and $r_n$ according to (2.7) and (2.8), then (2.20) holds obviously. Define $U_m$, $V_m$, $Q_m$ by (2.21)-(2.23) for $1 \leq m \leq M$, and

$$U_0 = P\left( \sup_{\xi \in \tilde{D}} \left| \sum_{i=1}^{r_1} a_{ni} t_n(\xi) \right| \geq \epsilon/3 \right) \quad (3.3)$$

$$V_0 = \left\{ \sum_{B \in G_1} P \left| \sum_{i=r_{r+1}}^{n} a_{ni} t_n(b) \right| \geq \epsilon/3 \right\} \quad (3.4)$$

with $b \in B$. It is easily verified that with $Q_0$, $U_0$, $V_0$ defined by (3.2)-(3.4), (2.24) also holds true for $m = 0$. Hence

$$Q_0 \leq U_0 + V_0 + Q_1 \leq \sum_{m=0}^{M} (U_m + V_m). \quad (3.5)$$

Here $M$ is defined by (2.8). Now we show that
\[ U_0 + V_0 \to 0, \text{ as } n \to \infty. \] (3.6)

By (2.18), for \( \xi \in D \), we have

\[
\left| \sum_{i=1}^{n} a_{ni} \chi_{ni}(\xi) \right| \leq \left( \sum_{i=1}^{n} \chi_{ni}(\xi) \right)^{1/2} \leq \left( \mu_n (\epsilon / \mu_n) \right)^{1/2} \to 0, \text{ as } n \to \infty. \] (3.7)

Further, \( \psi_{ni}(\xi) = 0 \) when \( \xi \in D \) and \( |e_{ni}| > \sqrt{\delta} d_n \mu_n (|x_{ni}|) \), here \( \psi_{ni}(\xi) \) is defined by (2.12). From (2.7), (2.19) and (3.7), we obtain

\[
U_0 \leq \mathbb{P} \left( \sup_{\xi \in D} \left| \sum_{i=1}^{r_1} a_{ni} \psi_{ni}(\xi) \right| \neq 0 \right) \leq \sum_{i=1}^{r_1} \mathbb{P} \left( |e_{ni}| \leq \sqrt{\delta} d_n \mu_n \right) \leq C d_n \mu_n^2 \to 0. \] (3.8)

Using (2.17)-(2.20) and employing the argument for proving (2.36), it can be shown that there exists constant \( C_1 > 0 \) such that

\[ V_0 \leq 2 \mu_n^{2p} \exp \left( -C_1 \mu_n^{1/2} \right) \to 0, \text{ as } n \to \infty. \] (3.9)

Now (3.6) follows from (3.8) and (3.9). Further, by Lemma 3, we have, for \( n \) large,

\[
\sum_{m=1}^{M} (U_m + V_m) \leq 4 \sum_{m=1}^{\infty} \mu_n^{2p} \exp \left( -48^{-1} (\mu_n / 9)^{m/2} \right). \] (3.10)

Since the function \( x^{3p} \exp \left( -a \left( \frac{x}{b} \right)^{m/2} \right), x > 0 \), attains its maximum at \( x = b (6p/a)^{2/m} \), we have

\[
\mu_n^{2p} \exp \left( -16^{-1} (\mu_n / 9)^{m/2} \right) = \mu_n^{-pm} \mu_n^{3p} \exp \left( -16^{-1} (\mu_n / 9)^{m/2} \right) \leq \mu_n^{-pm} 3p (96p/e)^{6p} \exp \left( -6p \right) \leq (\mu_n / 729)^{pm} (96p/e)^{6p}. \]

Hence, noting that \( \mu_n \to \infty \), we obtain

\[
\sum_{m=1}^{M} (U_m + V_m) \leq \left( \frac{96p/e}{6p} \right)^{6p} \sum_{m=1}^{\infty} (\mu_n / 729)^{-pm} \to 0. \] (3.12)
From (3.5), (3.6) and (3.12), (3.2) follows. This concludes the proof of (3.1).

By (3.1) we have

$$\left| \sum_{i=1}^{n} t_{ni}(\hat{s}_{ni}) \cdot x_{ni} \cdot I(\hat{s}_{ni} \in D) \right| \to 0 \text{ in probability, as } n \to \infty \quad (3.13)$$

and (2.7) gives, for $n$ large, $\Delta/(\sqrt{n} \cdot d_n) > \mu_n^2$ ($\Delta$ is the number appearing in condition 1 of Theorem 1). So by Lemma 2 we have

$$I\left( \sum_{i=1}^{n} \text{sgn}(e_{ni} - x_{ni}^{*} \hat{s}_{ni}) \cdot x_{ni} \right) \geq (p+1)d_n \cdot I(|\hat{s}_{ni}| < \mu_n) = 0, \text{ a.s.} \quad (3.14)$$

Hence by (1.19)

$$\left| \sum_{i=1}^{n} \text{sgn}(e_{ni} - x_{ni}^{*} \hat{s}_{ni}) \cdot x_{ni} \cdot I(|\hat{s}_{ni}| < \mu_n) \right| \leq (p+1)d_n + 0, \text{ a.s.} \quad (3.14)$$

From (2.12)-(2.14), (3.13) and (3.14), we have

$$\left| \sum_{i=1}^{n} \lambda_{ni}(\hat{s}_{ni}) \cdot x_{ni} - \sum_{i=1}^{n} \text{sgn}(e_{ni} \cdot x_{ni}) \cdot I(|\hat{s}_{ni}| < \mu_n) \right| \to 0, \text{ in probability.} \quad (3.15)$$

Since $\sup\{|x_{ni}^{*} \hat{s}_{ni}|: |\hat{s}_{ni}| < \mu_n, 1 \leq i \leq n\} \leq \mu_n d_n \leq \mu_n^2 d_n + 0$, we have

$$\lambda_{ni}(\hat{s}_{ni}) = 2f(0) \left( 1 + o_p(1) \right) x_{ni}^{*} \hat{s}_{ni} \quad \text{on account of condition 1 of Theorem 1,}$$

where $o_p(1) \to 0$ in probability as $n \to \infty$ uniformly for $1 \leq i \leq n$. From this, (1.18) and (3.15), we have

$$\left| 2f(0) \hat{s}_{ni} - \sum_{i=1}^{n} \text{sgn}(e_{ni} \cdot x_{ni}) \cdot I(|\hat{s}_{ni}| < \mu_n) \right| \to 0, \text{ in probability.} \quad (3.16)$$

In view of (1.18), (1.19) and the assumptions on $\{e_{ni}\}$, it follows by Lindeberg's theorem that
\[ \sum_{i=1}^{n} \text{sgn}(e_{ni}) x_{ni} \xrightarrow{P} N(0, \mathbf{I}_p). \]  

(3.17)

From (3.16), (3.17) and Lemma 4, we obtain (1.20) (notice (2.2)). This concludes the proof of Theorem 1', hence Theorem 1.

**Proof of Theorem 3.** The proof differs from the above argument only in some minor details, therefore omitted.

**Proof of Theorem 2.** Define \( T_n \) by (1.12), and

\[
\begin{align*}
\varepsilon_{n0} &= T_n^{1/2} \varepsilon_0, \\
\alpha_{n0} &= \sqrt{n}(\alpha_0 + \varepsilon_{n0}' \varepsilon_0), \\
\gamma_{n0} &= (\alpha_{n0}, \varepsilon_{n0}'). \\
x_{ni} &= T_n^{-1/2} (x_i - \overline{x_n}), \quad i = 1, \ldots, n, \\
z_{ni} &= (1/\sqrt{n}, x_{ni}'), \quad i = 1, \ldots, n.
\end{align*}
\]

We transform the model (1.11) into the following form:

\[ Y_i = z_{ni} \gamma_{n0} + e_i, \quad i = 1, \ldots, n. \]

Since \( \sum_{i=1}^{n} z_{ni} z_{ni}' = I_{p+1} \), and (1.13) guarantees that \( \max_{1 \leq i \leq n} |z_{ni}| \to 0 \) as \( n \to \infty \), Theorem 1' can be applied, and we obtain

\[ 2f(0)(\hat{\gamma}_n - \gamma_{n0}) \xrightarrow{P} N(0, \mathbf{I}_{p+1}), \quad \text{as } n \to \infty \]  

(3.18)

where \( \hat{\gamma}_n = (\hat{\alpha}_{n0}, \hat{\varepsilon}_{n0}') \), and \( \hat{\alpha}_{n0}, \hat{\varepsilon}_{n0} \) are the Minimum \( L_1 \)-Norm estimates of \( \alpha_{n0}, \varepsilon_{n0} \), respectively. Now (3.18) implies the assertion (1.14) and also the asymptotical independence of \( \hat{\alpha}_{n0} \) and \( \hat{\varepsilon}_{n0} \). Finally, (1.15) follows easily from what has already been proved. Theorem 2 is proved.

**Remark.** Wu (1987) proved that in model (1.1) the ML\(_1\)N estimate \( \hat{\varepsilon}_n \) is strong consistent if the following conditions are met:
1°. \( \{e_i\} \) satisfies the condition stated in Theorem 3.

2°. Define \( d_n = \max(1, \|x_1\|, \ldots, \|x_n\|) \) and \( \rho_n \) the smallest eigenvalue of \( S_n \); then \( \rho_n/(d_n^2 \log n) \to \infty \), \( d_n/n^c \to 0 \) for some \( c > 0 \).

At one time it was expected that if condition 1° is replaced by

1': \( e_1, e_2, \ldots \) are i.i.d. and \( e_1 \) has a unique median \( 0 \), the conclusion of Theorem 1 is still true. The motivation behind this conjecture is the simple case of estimating a population median by the sample median, in which the uniqueness of the population median is enough for consistency. Yet the following example shows that this is not true:

**Example.** In model (1.1) take \( p = 1 \) (\( \beta \) is one-dimensional), \( x_n = \log n/\sqrt{n} \), \( n = 1, 2, 3, \ldots \); \( e_1, e_2, \ldots \) are i.i.d., \( e_1 \) has a density function \( f(u) = |u|I(|u| < 1) \). Here \( d_n = 1 \), \( \rho_n = S_n - \frac{1}{3}(\log n)^3 \). Hence condition 2° is fulfilled.

In this example all conditions of Theorem 1, except that \( f(0) > 0 \), are met. In the course of proving Theorem 1' we have already shown this (see (3.15)). (Note that in proving (3.15) we made no use of \( f(0) > 0 \).

\[
\rho_n^{-1/2} \left| \sum_{i=1}^{n} \text{sgn}(e_i) x_i - 2 \sum_{i=1}^{n} x_i \int_{0}^{x_i} f(u)du \right| I(|\hat{\beta}_n| < 1) \to 0, \text{ in probability.} \quad (3.19)
\]

Now if

\[
\hat{\beta}_n \to 0, \text{ in probability,} \quad (3.20)
\]

then since \( \{x_i\} \) is bounded, from (3.19) and \( f(u) = |u|I(|u| < 1) \) we have

\[
\rho_n^{-1/2} \sum_{i=1}^{n} \text{sgn}(e_i) x_i - \rho_n^{-1/2} \hat{\beta}_n^{2} \sum_{i=1}^{n} x_i^{3} \to 0, \text{ in probability.} \quad (3.21)
\]

But by Lindeberg's theorem we have \( \sqrt{\rho_n^{-1}} \sum_{i=1}^{n} \text{sgn}(e_i) x_i \overset{p}{\to} N(0,1) \), while \( \sum_{i=1}^{n} x_i^{3} \) is bounded in \( n \), \( \rho_n \to \infty \) and \( \hat{\beta}_n \to 0 \) in probability. Thus (3.21) is impossible, which in turn implies that (3.20) is impossible.
REFERENCES


APPENDIX 1

In this appendix, all notations and numbering of formula are according to Bassett and Koenker (1978), if not defined here.

Consider the model

\[ y_t = x_t \beta + u_t, \quad t = 1, \ldots, T, \]

where \( \beta \), one-dimensional, is the unknown parameter, \( u_1, u_2, \ldots \) are independent random errors with a common distribution \( N(0,1) \), and

\[ x_1 = 1/\sqrt{2}, \quad x_2 = 1 + \sqrt{2}/10, \quad x_3 = x_4 = \ldots = 1. \]

First we verify that the minimum point of the function

\[ J(\beta) = \sum_{t=1}^{T} |y_t - x_t \beta| \]

is unique. For if this is not true, then owing to the convexity of \( J(\beta) \), there would exist an interval \( [a, b], -\infty < a < b < \infty \), such that \( J(c) = \inf(J(\beta): -\infty < \beta < \infty) \) for each \( c \in (a, b) \). Choose a point \( r \in (a, b), r \neq y_t/x_t, t = 1, \ldots, T \). We should have \( J'(r) = 0 \), i.e.,

\[ -\frac{1}{\sqrt{2}} \text{sgn}(y_1 - r/\sqrt{2}) - (1 + \frac{\sqrt{2}}{10}) \text{sgn}(y_2 - (1 + \frac{\sqrt{2}}{10})r) - \sum_{t=3}^{T} \text{sgn}(y_t - r) = 0. \]

But this is impossible, since the sum of the first two terms is an irrational number while the third term is an integer. This proves the uniqueness stated above.

Now in this model \( H = (1, 2, 3, \ldots, T) \). By the choice of \( \{x_t\} \), the distribution of \( Z_T(\delta, 1) \) is nonlattice. So according to (3.9), we should have

\[ \lim_{T \to \infty} T^{1/2} \text{Pr}(Z_T(\delta, 1) \in C[0, 1]) \text{ exists and not zero.} \quad (\star) \]

But
\[ Z_T(s,1) = (\sqrt{2} + \frac{1}{5}) \text{sgn}(u_2 - T^{-1/2}(1 + \sqrt{2})\delta) + \sum_{t=3}^{T} \sqrt{2} \text{sgn}(u_t - T^{-1/2}\delta). \]

Therefore, when \( T \) is even the right-hand side, with probability one, equals to an odd multiple of \( \sqrt{2} \) plus \( \pm 1/5 \), which is always outside \([-1,1]\). Consequently we have

\[ P(Z_T(s,1) \in C[0,1]) = 0, \quad \text{for} \quad T = 2,4,6, \ldots \]

and (\*) breaks down.
APPENDIX 2

In this appendix, all notations and numbering of formula are according to Amemiya (1982) if not defined here.

1. Denote by \( A_1(\theta_0) \) the value of \( A_1 \) in (3.12) taken at the true parameter point \( \theta_0 \). We shall proceed to show that \( A_1(\theta_0) \to \infty \) in probability as \( T \to \infty \).

Define

\[
\xi_t = \begin{cases} 
1, & \text{if } |y_t - x_t'\theta_0| \leq C_T^{-1} \\
0, & \text{otherwise.}
\end{cases}
\]

Since \( C_T = T^d, \frac{1}{3} < d < \frac{1}{2} \), it follows that \( C_T^{-1} < T^{-1/3} \). Hence \( Z_t \geq \xi_t \), and

\[
A_1(\theta_0) = 2C_T^{-1} \sum_{i=1}^{T} Z_t \log(1 + e^{-C_T|y_t - x_t'\theta_0|}) \\
\geq 2C_T^{-1} \sum_{t=1}^{T} \xi_t \log(1 + e^{-C_T|y_t - x_t'\theta_0|}) \\
\geq 2C_T^{-1} \log(1 + e^{-1}) \sum_{t=1}^{T} \xi_t = \bar{A}.
\]

Since \( y_t - x_t'\theta_0 = u_t, t = 1,2, \ldots \) are independent and identically distributed with a common density function \( f \) which is continuous and \( f(0) > 0 \), it follows that there exist two positive constants \( h_1 \) and \( h_2 \) not depending on \( t \), such that

\[
h_1 C_T^{-1} \leq P(\xi_t = 1) \leq h_2 C_T^{-1}, \quad t = 1,2, \ldots.
\]

Therefore we obtain

\[
E(\bar{A}) \geq 2 \log(1 + e^{-1})h_1 TC_T^{-2} \to \infty.
\]

Here we used the fact that \( C_T = T^d \) and \( 1/3 < d < 1/2 \). Further
Var(\tilde{A}) \leq (2C_T^{-1} \log(1+e^{-1}))^2 \sum_{t=1}^{T} E \xi_t^2 \\
\leq (2C_T^{-1} \log(1+e^{-1}))^2 T \theta C_T^{-1} \\
= 4h_2 \log^2(1+e^{-1})T/C_T^3 \to 0

by the definition of C_T given above. From E(\tilde{A}) \to \infty and Var(\tilde{A}) \to 0, we have \tilde{A} \to \infty in probability as T \to \infty. Since A_1(\tilde{a}_0) \geq \tilde{A}, we obtain A_1(\tilde{a}_0) \to \infty in probability as T \to \infty.

2. Denote by B_1(\bar{\theta}_0) the value of B_1 (in (3.22)) taken at the true parameter point \bar{\theta}_0. We shall now show that B_1(\bar{\theta}_0) \to \infty in probability as T \to \infty.

Define \xi_t as before. Since y_t - x_t^\top \bar{\theta}_0 = u_t, we have

\[ B_1(\bar{\theta}_0) = T^{-1/2} \sum_{t=1}^{T} I(|u_t| < T^{1/3}C_T^{-1}) |\bar{W}_t - G_0(u_t)||x_{it}| \]

\[ \geq \frac{1}{M(e+1)} T^{-1/2} \sum_{t=1}^{T} \xi_t |x_{it}|^2 = \tilde{B} \]

where M = sup(|x_{it}|: t = 1,2,...) \to 0 by assumption. We have

\[ E(\tilde{B}) \geq \frac{1}{M(e+1)} T^{-1/2} h_1 C_T^{-1} \sum_{t=1}^{T} x_{it}^2 \]

\[ = \frac{1}{M(e+1)} h_1 T^{1/2} C_T^{-1} \sum_{t=1}^{T} x_{it}^2 \to \infty, \text{ as } T \to \infty. \]

This is because \( \sum_{t=1}^{T} x_{it}^2 / T \) tends to a positive limit as \( T \to \infty \) (by assumption), and that \( C_T = T^d, 1/3 < d < 1/2 \). Further
\[
\text{Var}(\tilde{B}) \leq M^{-2}(e+1)^{-2} T^{-1} \sum_{t=1}^{T} |x_{it}|^4 E\xi_t^2
\]

\[
\leq M^2(e+1)^{-2} T^{-1} \sum_{t=1}^{T} E\xi_t^2
\]

\[
\leq M^2(e+1)^{-2} T^{-1} \text{Th}_2 C_T^{-1} + O, \quad \text{as} \quad T \to \infty.
\]

Since \( E(\tilde{B}) \to \infty \) and \( \text{Var}(\tilde{B}) \to 0 \), we have \( \tilde{B} \to \infty \) in probability as \( T \to \infty \). Since \( B_1(\varepsilon_0) \geq \tilde{B} \), the same is true for \( B_1(\varepsilon_0) \).