International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

Asymptotic Properties of Restricted L1-Estimates of Regression

Dupacova, J.

IIASA Working Paper

WP-87-018

February 1987

ASYMPTOTIC PROPERTIES OF RESTRICTED L_1-ESTIMATES OF REGRESSION

J. Dupačová

February 1987
WP-87-18

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

FOREWORD

The recent results on asymptotic behavior of statistical estimates and of optimal solutions of stochastic optimization problems obtained by Dupačová and Wets are used to prove consistency of restricted L_1-estimates under more general assumptions. For the special case of linearly restricted linear L_1-regression, Lagrangian approach is used to achieve asymptotic normality.

Alexander B. Kurzhanski
Chairman
System and Decision Sciences Program

ABSTRACT

Asymptotic properties of L_1-estimates in linear regression have been studied by many authors, see e.g. Bassett and Koenker (1978), Bloomfield and Steiger (1983). It is the lack of smoothness which does not allow to use the known results on asymptotic behavior of M-estimates (Huber (1967)) directly. The additional lack of a convexity in the nonlinear regression case increases the complexity of the problem even under assumption that the true parameter values belong to the interior of the given parameter set; for a consistency result in this case see e.g. Oberhofer (1982).

We shall use the technique developed in Dupačová and Wets (1986), (1987) to get asymptotic properties of the L_1-estimates of regression coefficients which are assumed to belong to an a priori given closed convex set given, e.g., by constraints of general equality and inequality form. The method uses, i.a., tools of nondifferentiable calculus and epi-convergence and it can be applied to other classes of L_1-estimates as well.

CONTENTS

$$\rho(t) = |t| \ ,$$

whose nondifferentiability precludes from direct application of the related asymptotic results. Nevertheless, asymptotic normality of the L_1-estimates of regression coefficient was proved by Bassett and Koenker (1978) for linear regression with nonrandom regressors and by Bloomfield and Steiger (1983) for linear regression with random regressors under ergodicity and stationarity assumption. Consistency result for L_1-estimates of parameters in nonlinear regression model can be found in Oberhofer (1982).

In model (1.1) and, correspondingly, in the optimization problem (1.3), restrictions on the estimated parameter values can be taken into account to respect technical and modeling considerations and, eventually, to guarantee the uniquenesss of the estimate (Barrodale and Roberts (1977)). Inequality constraints on the estimates, however, introduce an essential lack of smoothness. That is why one usually assumes (see e.g. Huber (1967), Oberhofer (1982)) that the true parameter vector is an *interior* point of the given admissible set.

For the case of linearly restricted linear regression the simplex method can be used to get the restricted L_1-estimates, i.e., to get the optimal solution of the mathematical programming problem

$$\min \sum_i |y_i - x^i b| \text{ on the set } S \tag{1.4}$$

where S is a nonempty convex polyhedral set. (For a survey on solution techniques see e.g. Barrodale and Roberts (1977) or Arthanari and Dodge (1982).)

The optimal solution of (1.4) lies very often on the boundary of S what does not conform with the mentioned assumption that the true parameter value is an interior point of S.

We shall use the technique developed in Dupačová and Wets (1986), (1987) to get consistency and asymptotic normality of restricted L_1-estimates. The special form of the objective function together with the use of empirical distribution help to simplify the assumptions used in the mentioned papers. Further simplification is possible in cases where g is linear both in b and x, i.e., for linear regression.

2. CONSISTENCY

Let ξ be an $m+1$ dimensional random vector on $(\Xi, \boldsymbol{a}, P)$ with components $\xi_0, \xi_1, \ldots, \xi_m$ and $f_0 : R^n \times R^{m+1} \to R^1$ the function

$$f_0(b, \xi) = |\xi_0 - g(b, \xi_1, \ldots, \xi_m)| \ . \tag{2.1}$$

Assume that $S \subset R^n$ is a given nonempty closed set of admissible parameter values and define $f : R^n \times R^{m+1} \to \bar{R} = R^1 \cup \{\infty\}$ the function

$$\begin{aligned} f(b, \xi) &= f_0(b, \xi) \quad \text{for} \quad b \in S \quad \text{and} \quad \xi \in R^{m+1} \\ &= +\infty \qquad\quad \text{for} \quad b \notin S \ . \end{aligned} \tag{2.2}$$

Observe that $f(b, \xi) = f_0(b, \xi) + \Psi_s(b)$ where $\Psi_s(b)$ is the indicator function of the set S

$$\Psi_s(b) = 0 \text{ for } b \in S \text{ and } \Psi_s(b) = +\infty \text{ if } b \notin S \ .$$

Let us first discuss the *properties of the functions f and f_0*.

a) If the function $g : R^n \times R^m \to R^1$ in (2.1) is a *continuous* function, then evidently

f is nonnegative and continuous in ξ

f_0 is nonnegative and continuous in b and ξ.

b) If for an arbitrary $\xi \in \Xi$ the function g in (2.1) is *locally Lipschitz* in b then for all $\xi \in \Xi$

$f_0(\cdot, \xi)$ is locally Lipschitz in b and

$f(\cdot, \xi)$ is lower semicontinuous.

c) Taking into account the special form of $f_0(b, \xi)$ we can write (using again the notation $\tilde{\xi}$ for the m-dimensional subvector of components $\xi_1, \ldots, \xi_m$)

$$f_0(b, \xi) = \max \{\Phi_1(b, \xi), \Phi_2(b, \xi)\}$$

with

$$\Phi_1(b, \xi) = \xi_0 - g(b, \tilde{\xi})$$

$$\Phi_2(b, \xi) = -\xi_0 + g(b, \tilde{\xi}).$$

If for an arbitrary $\xi \in \Xi$ the function g is *continuously differentiable* in b then for all $\xi \in \Xi$ (see Rockafellar (1981), prop. 4A and 3H)

$f_0(\cdot, \xi)$ is locally Lipschitz and subdifferentially regular and

$f(\cdot, \xi)$ is lower semicontinuous on R^n.

The *subdifferential* (with respect to b)

$$\partial f_0(b, \xi) = \text{conv}\,\{\nabla_b \Phi_i(b, \xi) \text{ for } i \text{ such that } f_0(b, \xi) = \Phi_i(b, \xi)\}$$

$$= \begin{cases} -\nabla_b g(b, \tilde{\xi}) \text{ if } \xi_0 > g(b, \tilde{\xi}) \\ \nabla_b g(b, \tilde{\xi}) \text{ if } \xi_0 < g(b, \tilde{\xi}) \\ \text{conv}\,\{-\nabla_b g(b, \tilde{\xi}), \nabla_b g(b, \tilde{\xi})\} \text{ otherwise}. \end{cases} \tag{2.3}$$

The estimate of the parameter vector β based on the sample of size ν from the distribution P, i.e., the optimal solution b^ν of the mathematical program

$$\min \sum_{i=1}^{\nu} f_0(b, \xi^i) \text{ on the set } S$$

or, equivalently,

$$\min \sum_{i=1}^{\nu} f(b, \xi^i) \quad \text{on } R^n \quad ,$$

corresponds to the use of the (random) empirical probability measure P^ν which converges to P in distribution almost surely. In our analysis we have to use assumptions concerned jointly with the considered probability measures P, P^ν and the functions f or f_0.

ASSUMPTION 2.1 *To any bounded set $V \subset R^n$ there corresponds a summable function κ such that for any pair $b^0, b^1 \in V$*

$$|f_0(b^0, \xi) - f_0(b^1, \xi)| \leq \kappa(\xi)\|b^0 - b^1\| .$$

COMMENT 2.2 Besides of local Lipschitz property of $f_0(\cdot, \xi)$ which is implied by the same property of $g(\cdot, \tilde{\xi})$, we assume the integrability of the Lipschitz constant κ.

ASSUMPTION 2.3 *The probability measures P, P^ν, $\nu = 1, 2, \ldots$ are f-tight, i.e., $\forall b \in S$ and $\varepsilon > 0$ there is a compact set $K_\varepsilon \subset \Xi$ such that*

$$\int_{\Xi - K_\varepsilon} f(b, \xi) P^\nu(d\xi) < \varepsilon,\ \nu = 1, \ldots$$

$$\int_{\Xi - K_\varepsilon} f(b, \xi) P(d\xi) < \varepsilon \ .$$

Assumption 2.3 is fulfilled automatically for $f(b, \cdot)$ bounded or Ξ compact. For ξ one-dimensional, it is equivalent to uniform integrability of $f_0(b, \cdot)$ in P^ν for $b \in S$ and it is equivalent to the convergence of expectations

$$E^\nu\{f_0(b, \xi)\} = \int_\Xi f_0(b, \xi) P^\nu(d\xi)$$

to a finite expectation

$$E\{f_0(b, \xi)\} = \int_\Xi f_0(b, \xi) P(d\xi)$$

for all $b \in S$ (see Loéve (1955), Section 11.4).

Under Assumption 2.3, similar results hold true in the more-dimensional case as well (see Dupačová and Wets (1986)), namely:

The expectations

$$E^\nu f = \int_\Xi f(b, \xi) P^\nu(d\xi), \ \nu = 1, \ldots$$

$$Ef = \int_\Xi f(b, \xi) P(d\xi)$$

are a.s. finite and lower semicontinuous on S and

$$Ef = \lim_{\nu \to \infty} E^\nu f = \text{epi}-\lim_{\nu \to \infty} E^\nu f \ .$$

In addition, the consistency property follows (see Dupačová and Wets (1986), Theorem. 3.9):

THEOREM 2.4 *Let in the definitions (2.1) and (2.2), ξ be an $(m+1)$-dimensional random vector on $(\Xi, \mathfrak{a}, P)$, $S \subset R^n$ be a closed nonempty set and the function $g : R^n \times R^m \to R^1$ be continuous. Let P^ν, $\nu = 1, 2, \ldots$ be (random) empirical measures based on independent samples of size ν from the distribution P such that Assumptions 2.1 and 2.3 hold true.*
Then:

1) *Any cluster point $\hat{b}$ of any sequence of $\{b^\nu\}_{\nu=1}^\infty$ such that $b^\nu \in \arg\min E^\nu f$, $\nu = 1, 2, \ldots$, almost surely belongs to* $\arg\min Ef$.

2) *If there is a compact set* $D \subset R^n$ *such that for* $\nu = 1, 2, \ldots$

$$(\arg\min E^{\nu}f) \cap D \text{ is nonempty a.s.}$$

and

$$\{\beta\} = (\arg\min Ef) \cap D$$

then there exist a measurable selection $\{b^{\nu}\}_{\nu=1}^{\infty}$ *of* $\{\arg\min E^{\nu}f\}_{\nu=1}^{\infty}$ *such that*

$$\beta = \lim_{\nu \to \infty} b^{\nu} \text{ a.s.} \tag{2.4}$$

and also

$$\inf Ef = \lim_{\nu \to \infty} (\inf E^{\nu}f) \text{ a.s.}$$

For the linear L_1-regression, i.e., for the problem (1.4) with an already given (observed) matrix X of regressors, the existence of the optimal solutions b^{ν} follows via properties of the corresponding linear program, see e.g. Bloomfield and Steiger (1983). For nonlinear L_1-regression this need not be the case. To guarantee the existence of optimal solutions of the programs

$$\min E\{f(b, \xi)\} \text{ and } \min E^{\nu}\{f(b, \xi)\} \tag{2.5}$$

for noncompact S one can use the infcompactness property of the objective functions $E\{f(b, \xi)\}$ and $E^{\nu}\{f(b, \xi)\}$. To this purpose, it is sufficient to assume that for a set $A \in \alpha$ with $P(A) > 0$ (resp. $P^{\nu}(A) > 0$) the set

$$\{(b, \xi) \in R^n \times A : f(b, \xi) \leq \alpha\} \tag{2.6}$$

is bounded for all $\alpha \in R$ (see Dupačová and Wets (1986), Proposition 3.10). For the empirical measure P^{ν}, this property is evidently fulfilled it the function $f(\cdot, \xi)$ is infcompact for a realization of ξ.

Evidently, our assumptions are weaker than those by Oberhofer (1982) and it is possible to proceed in a quite similar way to get the consistency of restricted L_1-estimates for other models without unnatural smoothness assumptions.

3. ASYMPTOTIC NORMALITY

Provided that all the assumptions of Theorem 2.4 needed to get the consistency result (2.4) are fulfilled we can study the rate of convergence for (2.4) in a probabilistic sense. To this purpose, appropriate differentiability properties of our problems (2.5) are needed.

ASSUMPTION 3.1 *For an arbitrary $\xi \in \Xi$, the function $g(\cdot, \xi)$ is continuously differentiable.*

According to results by Clark (1983) (see also the discussion in Dupačová and Wets (1987)) we have with $\partial f_0(b, \xi)$ given by (2.3)

LEMMA 3.2 *Under Assumptions 2.1, 2.3 and 3.1*

$$\partial E\{f_0(b, \xi)\} = E\{\partial f_0(b, \xi)\} \ , \tag{3.1}$$

$$\partial E^{\nu}\{f_0(b, \xi)\} = E^{\nu}\{\partial f_0(b, \xi)\} \tag{3.2}$$

$$a.s. \text{ for } \nu = 1, 2, \ldots$$

and for an arbitrary $b \in S$

$$\partial E\{f(b, \xi)\} \subset \partial E\{f_0(b, \xi)\} + \partial \Psi_s(b) \ , \tag{3.3}$$

$$\partial E^{\nu}\{f(b, \xi)\} \subset \partial E^{\nu}\{f_0(b, \xi)\} + \partial \Psi_s(b) \tag{3.4}$$

$$a.s. \text{ for } \nu = 1, 2, \ldots$$

with equality if Ψ_s is subdifferentially regular at b.

COMMENT 3.3 a) For convex sets or for smooth manifolds, the indicator function Ψ is subdifferentially regular, see Rockafellar (1981).

b) Formula (3.1) together with (2.3) imply that for P absolutely continuous Ef_0 is differentiable.

The properties (3.1), (3.3) imply that for an arbitrary $b \in S$ and $v(b) \in \partial E\{f(b, \xi)\}$ there exist $v_s(b) \in \partial \Psi_s(b)$ and measurable $u_0(b, \cdot)$ such that almost surely

$$u_0(b, \xi) \in \partial f_0(b, \xi)$$

and

$$v(b) = v_0(b) + v_s(b) = E\{u_0(b, \xi)\} + v_s(b) \ .$$

Similarly according to (3.2), (3.4), for an arbitrary $b \in S$ and $v^{\nu}(b) \in \partial E^{\nu}\{f(b, \xi)\}$ we have almost surely

$$v^{\nu}(b) = v_0^{\nu}(b) + v_s(b) = E^{\nu}\{u_0(b, \xi)\} + v_s(b)$$

where

$$Ev_0^{\nu}(b) = E\left\{\frac{1}{\nu}\sum_{i=1}^{\nu} u_0(b, \xi^i)\right\} = E\{u_0(b, \xi)\} = v_0(b)$$

and

$$\text{var } v_0^{\nu}(b) = \frac{1}{\nu} \text{ var } \{u_0(b, \xi)\}$$

due to subdifferential regularity of f_0 and to the definition of P^{ν}.

Application of these results to necessary conditions

$$0 \in \partial E\{f(\beta, \xi)\} \text{ and } 0 \in \partial E^{\nu}\{f(b^{\nu}, \xi)\}, \ \nu = 1, 2, \ldots$$

for the optimal solutions of the problems (2.5), i.e. for

$$\beta \in \arg\min E\{f(b, \xi)\}$$

and

$$b^{\nu} \in \arg\min E^{\nu}\{f(b, \xi)\}, \ \nu = 1, 2, \ldots$$

implies existence of $v_s(\beta) \in \partial\Psi_s(\beta)$, $v_s(b^{\nu}) \in \partial\Psi_s(b^{\nu})$ and random functions $u_0(\beta, \cdot)$, $u_0(b^{\nu}, \cdot)$ such that

$$u_0(\beta, \xi) \in \partial f_0(\beta, \xi) \text{ a.s. },$$

$$u_0(b^{\nu}, \xi) \in \partial f_0(b^{\nu}, \xi) \text{ a.s. for } \nu = 1, 2, \ldots$$

and

$$0 = E\{u_0(\beta, \xi)\} + v_s(\beta) = v(\beta) \tag{3.5}$$

$$0 = E^{\nu}\{u_0(b^{\nu}, \xi)\} + v_s(b^{\nu}) = v^{\nu}(b^{\nu}) \tag{3.6}$$

a.s. for $= 1, 2, \ldots$

For this choice of subgradients $v^{\nu}(b^{\nu})$, the condition

$$\frac{1}{\sqrt{\nu}} v^{\nu}(b^{\nu}) \longrightarrow 0 \text{ in probability as } \nu \longrightarrow \infty$$

is trivially fulfilled.

The basic idea is to apply Huber's approach (see Huber (1967), Section 4) to the subgradients v and v^{ν} of the functions Ef and $E^{\nu}f$ that fulfill (3.5) and (3.6) for to get the asymptotic normality of b^{ν}. The assumptions of Dupačová and Wets (1987) reduce to three basic conditions in our case:

(a) $\sqrt{\nu}\,[v^{\nu}(\beta) + v(b^{\nu})] \longrightarrow 0$ in probability as $\nu \longrightarrow \infty$.

(b) Ef_0 is twice continuously differentiable at the point β with nonsingular Hessian H.

(c) $\sqrt{\nu}\,[v_S(b^{\nu}) - v_S(\beta)] \longrightarrow 0$ in probability as $\nu \longrightarrow \infty$.

The first two properties resemble results of Huber (1967) and their validity can be proved under various sets of sufficient conditions. The property (c) is of a different nature. It is trivially satisfied if β and b^{ν} for ν large enough are interior points of S. For to indicate briefly that all mentioned conditions can be fulfilled we shall concentrate to the case of linearly restricted linear L_1-regression; the nonlinear case is substantially more complicated due to the fact that the function f_0 is neither convex nor differentiable.

We assume that the true parameter vector β is the optimal solution of the mathematical program

$$\text{minimize } E\{f_0(b, \xi)\} \text{ subject to } Ab \geq c \tag{3.7}$$

and it is estimated by optimal solutions b^{ν} of the programs

$$\text{minimize } E^{\nu}\{f_0(b, \xi)\} \text{ subject to } Ab \geq c \ ; \tag{3.8}$$

$A(m, n)$ and $c(m, 1)$ are given matrices of constant elements and $f_0(b, \xi) = |\xi_0 - b^T\tilde{\xi}|$. The corresponding Lagrangian functions have the form

$$L(b, y) = \begin{cases} \int_{\Xi} f_0(b, \xi)P(d\xi) - y^T(Ab - c) & \text{for } y \geq 0 \\ -\infty & \text{otherwise} \end{cases} \tag{3.9}$$

and

$$L^{\nu}(b, y) = \begin{cases} \int_{\Xi} f_0(b, \xi)P^{\nu}(d\xi) - y^T(Ab - c) & \text{for } y \geq 0 \\ -\infty & \text{otherwise} \end{cases} \tag{3.10}$$

Under Assumptions 2.1 and 2.3 (applied to the considered function f_0 instead of f)

an assertion about consistency of saddle points (b^ν, y^ν) parallel to that of Theorem 2.4 can be proved (see Dupačová and Wets (1987), Theorem 5.2). The existence of saddle points in the case of linearly restricted linear L_1-regression is guaranteed thanks to the special type of constraints and of the function f_0. Also in this case,

$$0 \in \partial_b L(\beta, \eta) = \partial_b Ef_0(\beta) - A^T \eta$$

$$0 \in \partial_y \Psi_s(\beta, \eta) + A\beta - c$$

and

$$0 \in \partial_b L^\nu(b^\nu, y^\nu) = \partial_b E^\nu f_0(b^\nu) - A^T y^\nu$$

$$0 \in \partial_y \Psi_s(b^\nu, y^\nu) + Ab^\nu - c$$

are necessary and ***sufficient*** conditions for (β, η) and (b^ν, y^ν) to be saddle points of the Lagrangian functions L and L^ν with respect to the set $S = R^n \times R^m_+$.

The special form of the set S together with consistency of (b^ν, y^ν) help to eliminate the constraints in (3.9) and (3.10) provided that the *strict complementarity* conditions hold true for (β, η), i.e., for $\forall i$

$$\eta_i = 0 \iff \sum_{j=1}^n a_{ij}\beta_j > c_i \ . \tag{3.11}$$

Denote by $I \subset \{1, \ldots, m\}$ the set of indices for which $\eta_i > 0$, i.e., for which the i-th constraint is active for the true parameter vector β. Evidently,

$$y_i^\nu = 0 \text{ for } i \notin I \text{ and } \sum_{j=1}^n a_{ij} b_j^\nu = c_i \text{ for } i \in I \text{ a.s.}$$

for ν large enough. Denote $A_I = (a_{ij})_{\substack{i \in I \\ j=1,\ldots,n}}$. In this situation, we are in fact interested to study asymptotic behavior of the *unconstrained* saddle points (b^ν, y_I^ν) of the reduced Lagrangian function

$$L_I^\nu(b, y_I) = \int_\Xi f_0(b, \xi) P^\nu(d\xi) - \sum_{i \in I} y_i \left[\sum_{j=1}^n a_{ij} b_j - c_i \right]$$

for $\nu \to \infty$. All we need for asymptotic normality of the estimates b^ν are the corresponding versions of conditions (a), (b) with $v^\nu(\beta)$ and $v(b^\nu)$ replaced by $v_0^\nu(\beta) - A_I^T \eta_I$ and $v_0(b^\nu) - A_I^T y_I^\nu$ and with Ef_0 replaced by the reduced Lagrangian function L_I.

THEOREM 3.4 *Let the true parameter vector β be the point of minima of the function $f_0(b, \xi) = |\xi_0 - b^T\tilde{\xi}|$ on the set $S = \{b : Ab \geq c\}$.*

Assume further:

(i) *For the true parameter vector β, the random vector $\tilde{\xi}$ and residual ε in*

$$\xi_0 = \beta^T\tilde{\xi} + \varepsilon$$

are independent with densities h_1 and h_2 such that $h_2(0) > 0$.

(ii) *The absolute values $|\xi_i|$, $i = 0, 1, \ldots, n$, of the components of the random vector ξ are uniformly integrable with respect to P^ν, $\nu = 1, 2, \ldots$.*

(iii) *The absolute moments $E\|\tilde{\xi}\|^k$, $k = 1, 2, 3$ exist and $E\tilde{\xi}\tilde{\xi}^T$ is finite and non-singular.*

(iv) *For the true parameter vector β and for the corresponding saddle point (β, η) of (3.9), the strict complementarity conditions (3.11) hold true. The matrix A_I is of full row rank.*

Then: $\sqrt{\nu}(b^\nu - \beta)$ *is asymptotically normal* $N(0, C\Sigma C^T)$ *with* $\Sigma = \mathrm{var}\,\tilde{\xi}$, $C = H^{-1}(I - A_I^T(A_I H^{-1} A_I^T)^{-1} A_I H^{-1})$ *and* $H = 2h_2(0)E\tilde{\xi}\tilde{\xi}^T$.

4. PROOF OF THEOREM 3.4

The assumed existence of $E\|\tilde{\xi}\|$ and the uniform integrability of $|\xi_i|$, $i = 0, 1, \ldots, n$, imply that Assumptions 2.1 and 2.3 (needed for consistency) are fulfilled for $f_0(b, \xi) = |\xi_0 - b^T\tilde{\xi}|$.

Denote by

$$l(b, y_I, \xi) = \begin{bmatrix} u_0(b, \xi) - A_I^T y_I \\ -A_I b + c_I \end{bmatrix}$$

with $u_0(b, \xi) \in \partial f_0(b, \xi)$ a subgradient of the reduced Lagrangian function $L_I(b, y_I)$. Following our discussion from Section 3, we can choose $u_0(b, \xi)$ in such a way that

$$E^\nu\{l(b^\nu, y_I^\nu, \xi)\} = 0 \;,$$

so that the condition

$$\frac{1}{\sqrt{\nu}} E^\nu\{l(b^\nu, y_I^\nu, \xi)\} \to 0 \text{ in probability as } \nu \to \infty$$

is evidently fulfilled.

Let us study the properties of the subgradients $u_0(b, \xi)$.

LEMMA 4.1 *Denote*

$$\bar{u}_d(b, \xi) = \sup_{\{b': \|b' - b\| < d\}} \|u_0(b, \xi) - u_0(b', \xi)\| .$$

Then under assumption (i) of Theorem 3.4 there is a positive constant k such that

$$E\{\bar{u}_d(b, \xi)\} \leq kdE\|\tilde{\xi}\|^2$$

and

$$E\{\bar{u}_d^2(b, \xi)\} \leq 2kdE\|\tilde{\xi}\|^3 .$$

PROOF According to (2.3), we have

$$u_0(b, \xi) = \begin{cases} -\tilde{\xi} & \text{if } \xi_0 > \tilde{\xi}^T b \\ \tilde{\xi} & \text{if } \xi_0 < \tilde{\xi}^T b \\ \operatorname{conv}\{\tilde{\xi}, -\tilde{\xi}\} & \text{if } \xi_0 = \tilde{\xi}^T b \ , \end{cases} \tag{4.1}$$

so that

$$\bar{u}_d(b, \xi) = \begin{cases} 0 & \text{if } O_d(b) \cap \{b': \tilde{\xi}^T b' = \xi_0\} = \phi \\ \leq 2\|\tilde{\xi}\| & \text{otherwise} . \end{cases} \tag{4.2}$$

For a given d, b and ξ, the condition

$$O_d(b) \cap \{b': \tilde{\xi}^T b = \xi_0\} = \phi$$

can be equivalently expressed as

$$\rho(b, p(\xi)) \geq d$$

where $\rho(b, p(\xi))$ denotes the distance of b from the hyperplane

$$p(\xi) = \{b': \tilde{\xi}^T b' = \xi_0\} ,$$

i.e.,

$$\rho(b, p(\xi)) = \frac{|\tilde{\xi}^T b - \xi_0|}{\|\tilde{\xi}\|} . \tag{4.3}$$

Using (4.2), (4.3), we get

$$E\{\bar{u}_d(b,\xi)\} = \int_{\Xi} \bar{u}_d(b,\xi)P(d\xi) \le 2 \int_{M_d(b)} \|\tilde{\xi}\| P(d\xi)$$

where $M_d(b) = \{\xi : \rho(b, p(\xi)) < d\} = \{\xi : |\tilde{\xi}^T b - \xi_0| < d\|\tilde{\xi}\|\}$. Substituting $\tilde{\xi}^T\beta + \varepsilon$ for ξ_0 we have

$$E\{\bar{u}_d(b,\xi)\} \le 2 \int_{\{\tilde{\xi},\varepsilon : -d\|\tilde{\xi}\| + \tilde{\xi}^T(b-\beta) < \varepsilon < d\|\tilde{\xi}\| + \tilde{\xi}^T(b-\beta)\}} \|\tilde{\xi}\| h_1(\tilde{\xi}) h_2(\varepsilon) d\tilde{\xi} d\varepsilon$$

$$= 2\int \|\tilde{\xi}\| h_1(\tilde{\xi}) \left[\int_{-d\|\tilde{\xi}\| + \tilde{\xi}^T(b-\beta)}^{d\|\tilde{\xi}\| + \tilde{\xi}(b-\beta)} h_2(\varepsilon) d\varepsilon \right] d\tilde{\xi}$$

$$\le 4d \sup h_2(\varepsilon) \cdot E\|\tilde{\xi}\|^2 = kdE\|\tilde{\xi}\|^2 .$$

In a similar way,

$$E\{\bar{u}_d^2(b,\xi)\} = \int_{\Xi} \bar{u}_d^2(b,\xi)P(d\xi) \le 4 \int_{M_d(b)} \|\tilde{\xi}\|^2 P(d\xi) \le 2k\, d\, E\|\tilde{\xi}\|^3 \ . \ \square$$

LEMMA 4.2 Existence of Hessian *Under Assumption (i) of Theorem 3.4, Ef_0 is twice continuously differentiable at the point β with Hessian*

$$H = 2h_2(0)E\tilde{\xi}\tilde{\xi}^T$$

provided that the expectation $E\tilde{\xi}\tilde{\xi}^T$ exists.

PROOF The function Ef_0 can be written as

$$Ef_0(b) = \int_{\Xi} |\xi_0 - \tilde{\xi}^T b| P(d\xi) = \int\int |\tilde{\xi}^T(\beta - b) + \varepsilon| h_1(\tilde{\xi}) h_2(\varepsilon) d\tilde{\xi} d\varepsilon$$

and its gradient (see (4.1), (3.1) and comment 3.3b)

$$\nabla Ef_0(b) = - \iint_{\{\tilde{\xi},\varepsilon : \tilde{\xi}^T(\beta - b) + \varepsilon > 0\}} \tilde{\xi} h_1(\tilde{\xi}) h_2(\varepsilon) d\tilde{\xi} d\varepsilon$$

$$+ \iint_{\{\tilde{\xi},\varepsilon : \tilde{\xi}^T(\beta - b) + \varepsilon < 0\}} \tilde{\xi} h_1(\tilde{\xi}) h_2(\varepsilon) d\tilde{\xi} d\varepsilon$$

$$= -E\tilde{\xi} + 2\int \tilde{\xi} h_1(\tilde{\xi}) \left(\int_{-\infty}^{\tilde{\xi}^T(b-\beta)} h_2(\varepsilon) d\varepsilon \right) d\tilde{\xi} \ .$$

Accordingly, the matrix of the 2-nd order derivatives

$$H(b) = 2\int \tilde{\xi} h_1(\tilde{\xi}) h_2(\tilde{\xi}^T(b - \beta))\tilde{\xi}^T \mathrm{d}\tilde{\xi} ,$$

so that

$$H = H(\beta) = 2h_2(0)E\,\widetilde{\tilde{\xi}\tilde{\xi}}^T . \ \square$$

LEMMA 4.3 *Sufficient conditions that β be an isolated global minimum of*

$$Ef_0(b) = E|\xi_0 - b^T\tilde{\xi}| \ \textit{on}\ S = \{b : Ab \geq c\}$$

and the associated Lagrangian multiplier η be unique are:

(i) $A\beta \geq c,\ \eta^T(A\beta - c) = 0,\ \eta \geq 0$

$$\nabla Ef_0(\beta) - A^T\eta = 0 .$$

(ii) *For $I = \{i : \sum_{j=1}^{n} a_{ij}\beta_j = c_i\}$, the matrix*

$$A_I = (a_{ij})_{\substack{i \in I \\ j = 1, \ldots, n}}$$

is of full row rank and

$$\eta_i > 0,\ i \in I .$$

(iii) *Assumption (i) of Theorem 3.4 comes true and $E\widetilde{\tilde{\xi}\tilde{\xi}}^T$ is nonsingular.*

PROOF Condition (i) is the first-order necessary condition, condition (ii) contains the linear independence condition and strict complementary conditions and condition (iii) together with Lemma 4.2 implies that the second order sufficient condition is fulfilled. The result follows e.g. from Theorem 3.2.2 of Fiacco (1983). □

If condition (ii) is fulfilled, we can rewrite the first order conditions (i) in the form

$$A_I\beta = c,\ \nabla Ef_0(\beta) - A_I^T\eta_I = 0 ,$$

i.e.,

$$\nabla L_I(\beta, \eta_I) = \begin{pmatrix} \nabla_\beta L_I(\beta, \eta_I) \\ \nabla_y L_I(\beta, \eta_I) \end{pmatrix} = 0 .$$

Conditions (ii), (iii) of Lemma 4.3 together with assumption (i) of Theorem 3.4 imply that the matrix L of the second order derivatives of the reduced Lagrange function $L_I(b, y_I)$ at the point β, η_I,

$$L = \begin{bmatrix} H & -A_I^T \\ -A_I & 0 \end{bmatrix},$$

is nonsingular. Accordingly, we have

LEMMA 4.4 *Under assumptions (i), (iv) of Theorem 3.4 complemented by assumption (iii) of Lemma 4.3, condition (b) is fulfilled for $L_I(b, y_I)$.*

Condition (a) can be written as

$$\sqrt{\nu}\,[E^{\nu} l(\beta, \eta_I) + El(b^{\nu}, y_I^{\nu})] = \sqrt{\nu}\begin{bmatrix} v_0^{\nu}(\beta) - A_I^T\eta_I + v_0(b^{\nu}) - A_I^T y_I^{\nu} \\ 0 \end{bmatrix} \longrightarrow 0$$

in probability a.s. as $\nu \longrightarrow \infty$. To get the desired convergence property of

$$\sqrt{\nu}\,[v_0^{\nu}(\beta) - A_I^T\eta_I + v_0(b^{\nu}) - A_I^T y_I^{\nu}]$$

we shall check under which circumstances the conditions (N-1)–(N-4) of Huber (1967) are fulfilled: Measurability and separability of $l(b, y_I, \xi)$, cf. (N-1), is evidently fulfilled, existence and uniqueness of the true β, η_I, cf. (N-2) and (N-3i) follow from assumptions of Lemma 4.3 and properties of subgradients $l(b, y_I, \xi)$, cf. (N-3ii), (N-3iii) and (N-4), can be obtained using Lemma 4.1 and assumption (iii) of Theorem 3.4.

Denote

$$\bar{l}_d(b, y_I, \xi) = \sup_{(b', y_I') \in O_d(b, y_I)} \| l(b, y_I, \xi) - l(b', y_I', \xi) \| .$$

LEMMA 4.5 *Let assumption (i) of Theorem 3.4 be fulfilled and let the absolute moments $E\|\tilde{\xi}\|^2$, $E\|\tilde{\xi}\|^3$ exist. Then there are positive constants K, K' such that*

$$E_d\{\bar{l}(b, y_I, \xi)\} \le Kd$$

$$E_d\{\bar{l}(b, y_I, \xi)\}^2 \le K'd \quad .$$

The expected value $E\{\|l(\beta, \eta_I, \xi)\|^2\}$ is finite.

PROOF We have

$$\bar{l}_d(b, y_I, \xi) \leq \sup_{\|b' - b\| < d} \|u_0(b, \xi) - u_0(b', \xi)\| + \sup_{\|b' - b\| < d} \|A_I(b - b')\|$$

$$+ \sup_{\|y_I' - y_I\| < d} \|A_I^T(y_I - y_I')\| ,$$

so that

$$E\{\bar{l}_d(b, y_I, \xi)\} \leq kd\, E\|\tilde{\xi}\|^2 + a.d \leq Kd$$

according to Lemma 2.

Similarly,

$$\bar{l}_d^2(b, y_I, \xi) = \sup_{(b', y_I') \in O_d(b, y_I)} \|l(b, y_I, \xi) - l(b', y_I', \xi)\|^2$$

$$\leq \sup_{\|b' - b\| < d} \|u_0(b, \xi) - u_0(b', \xi)\|^2 + 2 \sup_{\|b' - b\| < d} \|u_0(b, \xi) - u_0(b, \xi)\|$$

$$\cdot \sup_{\|y_I' - y_I\| < d} \|A_I^T(y_I - y_I')\|$$

$$+ \sup_{\|y_I' - y_I\| < d} \|A_I^T(y_I - y_I)\|^2 + \sup_{\|b' - b\| < d} \|A_I(b - b')\|^2$$

and

$$E\{\bar{l}_d^2(b, y_I, \xi)\} \leq 2kd\, E\|\tilde{\xi}\|^3 + 2kd^2 E\|\tilde{\xi}\|^2 a + 2a^2 d^2 \leq d \cdot K'$$

The last condition is evidently fulfilled as

$$E\{\|u_0(\beta, \xi)\|^2\} = E\|\tilde{\xi}\|^2 < \infty .$$

REFERENCES

Arthanari, T.S. and Y. Dodge (1982): *Mathemtical Programming in Statistics*. Wiley, New York.

Bassett, G.Jr. and R. Koenker (1978): Asymptotic theory of lest absolute error regression, *J. ASA* 73, 618–622.

Barrodale, I. and F.D.K. Roberts (1977): Algorithms for restricted least absolute value estimation. *Commun. Statist.-Simula. Compta*, B6(4), 353–363.

Bloomfield, P. and W.L. Steiger (1983): *Least Absolute Deviations, Theory, Applications and Algorithms*. Birkhäuser, Boston.

Clarke, R. (1983): *Optimization and Nonsmooth Analysis*. Wiley Interscience, New York.

Dupačová, J. (1987): Asymptotic properties of restricted L_1-estimates of regression. To appear in Proc. of the 1st International Conference on Statistical Data Analysis based on the L_1 norm and Related Methods, Neuchatel, 31 August – 4 September 1987. North-Holland, Amsterdam.

Dupačová, J. and R. Wets (1986): Asymptotic behavior of statistical estimators and optimal solutions for stochastic programming problems. WP-86-41 IIASA, Laxenburg.

Dupačová, J. and R. Wets (1987): Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems, II. WP-87-9 IIASA, Laxenburg.

Fiacco, A.V. (1983): *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*. Academic Press, New York.

Huber, P. (1967): The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. 5-th Berkeley Symp. Math. Stat. Prob.* 1, 221–233.

Huber, P. (1973): Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. of Stat.* 1, 799–821.

Loève, M. (1955): *Probability Theory*. van Nostrand, New York.

Oberhofer, W. (1982): The consistency of nonlinear regression minimizing the L_1-norm. *Ann. of Stat.* 10, 316–319.

Rockafellar, R.T. (1981): *The Theory of Subgradients and its Applications to Problems of Optimization. Convex and Nonconvex Functions*. Research and education in mathematics 1, Heldermann Verlag, Berlin.