Asymptotics of M-estimators in two-phase linear regression models

Hira L. Koul\textsuperscript{a,*,1}, Lianfen Qian\textsuperscript{b,2}, Donatas Surgailis\textsuperscript{c}

\textsuperscript{a}Department of Statistics & Probability, Michigan State University, East Lansing, MI 48824-1027, USA
\textsuperscript{b}Department of Math. Sciences, Florida Atlantic University, Boca Raton, FL 33431, USA
\textsuperscript{c}Institute of Mathematics & Informatics, 2600 Vilnius, Lithuania

Received 16 September 2001; received in revised form 18 April 2002; accepted 26 June 2002

Abstract

This paper discusses the consistency and limiting distributions of a class of M-estimators in two-phase random design linear regression models where the regression function is discontinuous at the change-point with a fixed jump size. The consistency rate of an M-estimator \( \hat{r}_n \) for the change-point parameter \( r \) is shown to be \( n \) while it is \( n^{1/2} \) for the coefficient parameter estimators, where \( n \) denotes the sample size. The normalized M-process is shown to be uniformly locally asymptotically equivalent to the sum of a quadratic form in the coefficient parameter vector and a jump point process in the change-point parameter, in probability. These results are then used to obtain the joint weak convergence of the M-estimators. In particular, \( n(\hat{r}_n - r) \) is shown to converge weakly to a random variable which minimizes a compound Poisson process, a suitably standardized coefficient parameter M-estimator vector is shown to be asymptotically normal, and independent of \( n(\hat{r}_n - r) \).

\textsuperscript{c}© 2002 Elsevier Science B.V. All rights reserved.

MSC: primary 62F12; secondary 62E20; 62J99; 65C05

Keywords: Change-point estimator; Fixed jump size; Compound Poisson process

\textsuperscript{*}Corresponding author. Fax: +1-517-432-1405.
\textsuperscript{E-mail addresses:} koul@stt.msu.edu (H.L. Koul), lqian@pop.fau.edu (L. Qian), sdonatas@ktl.mii.lt (D. Surgailis).
\textsuperscript{1}Research partly supported by the NSF grant DMS-0071619.
\textsuperscript{2}Research partly supported by the Florida Atlantic University Research Initiation Award.
1. Introduction

A regression model with piecewise linear regression function over two different domains of the design variable is called a two-phase linear regression model. There are two types of such models, called restricted and unrestricted. In the restricted case, the regression function is continuous at the change-point but not differentiable while in the unrestricted case, it is discontinuous. Under the unrestricted models there are two types of change-point problems: contiguous and fixed jump size. In the contiguous case, the jump size tends to zero as the sample size tends to infinity, while in the latter case it is constant. The focus of this paper is the development of the asymptotic distributions for a class of M-estimators in unrestricted fixed jump size two-phase linear regression models with random designs.

Examples of important applications of these models in various scientific fields are discussed by numerous researchers. Anderson and Nelson (1975) used a special type of the restricted case of a two-phase regression model (Sprent, 1961), called linear-plateau model, to predict crop yield based on the amount of nitrogen in the soil. Eubank (1984) gave examples of a variety of applications where the regression function is difficult or impossible to specify, but can be approximated by simpler segmented models. Some other important examples are listed in a recent paper of Müller and Stadtmüller (1999) and references therein.

Beginning with the work of Quandt (1958), the literature on the change-point regression problem has become vast. For restricted models with non-random designs, Hudson (1966) gave a concise method for calculating the least squares solution for the change-point while Hinkley (1969, 1971) derived asymptotic results for maximum likelihood estimator (MLE) of the point of intersection for the special case of two line segments under normally distributed errors. Under continuity and suitable identifiability assumptions, Feder (1975a, b) derived the asymptotic distributions of the least square (LS) estimator and the log likelihood ratio statistic for two-phase non-random design regression models with Gaussian errors. In a review paper, Shaban (1980) collected a considerable amount of work on the change-point problem and two-phase regression. Schulze (1987) provided a collection of existing methods mainly focusing on the least squares estimation, testing of hypotheses and testing of model stability for analyzing data using multiphase regression models.

Bhattacharya (1990, 1994) discussed the limiting behavior of MLE of the change-point and of the log-likelihood ratio process for both restricted and unrestricted contiguous two-phase non-random design linear regression models with Gaussian errors. van de Geer (1988) discussed the asymptotics of the least square estimators and tests for some general multiphase regression models. Bai and Perron (1998) investigated the asymptotics of the least squares estimators and the corresponding tests in multi-phase contiguous random design linear regression models when the errors satisfy some martingale type assumptions. Koul and Qian (2002) (KQ) established the consistency and the limiting distribution of the MLE in unrestricted fixed jump size two-phase random design regression models for a class of error densities that excludes the double exponential and such non-smooth densities. See also van de Geer (1988) and Csörgő and Horváth (1997) for some other related results in the fixed design case. In all of
the above papers, except that of KQ, the asymptotic distribution of the standardized change-point estimator is related to a Brownian motion, while in KQ, it is related to a compound Poisson process.


Rukhin and Vajda (1997) considers the change-point estimation problem as a non-linear regression problem. Thus, they deal with an equally spaced non-random design regression problem. They establish the consistency of a class of approximate M-estimators under fairly general conditions. Furthermore, they obtain the joint asymptotic normality of M-estimators of the change-point and the other underlying parameters only for a special class of restricted change-point regression functions and when the dispersion function $\rho$ that appears in the definition of these estimators (see (2.3)) is twice continuously differentiable.

The present paper discusses the consistency and limiting distributions of a class of M-estimators for the unrestricted fixed jump size two-phase random design regression models, under mild conditions on $\rho$, the error and the design variables. The results include the asymptotics of the least absolute deviation and LS estimators for a large class of error distributions.

The minimizer of the underlying M-process with respect to the change-point is non-unique and obtained over an interval of ordered design points. The rates of consistency of any change-point and coefficient parameter vector estimators are shown to be $n$ and $n^{1/2}$, respectively. The normalized M-process is shown to be uniformly locally asymptotically equivalent to the sum of two processes, in probability. One is a quadratic form in the standardized coefficient parameter vector and the other a marked empirical process in the standardized change-point parameter. The latter process is shown to converge weakly to a compound Poisson process whose set of minimizers forms a bounded interval. A suitably standardized smallest change-point M-estimator is shown to converge weakly to the smallest minimizer of this compound Poisson process. The standardized M-estimator of the regression coefficient parameter vector is shown to be asymptotically normal, and independent of the standardized change-point M-estimator. It must be mentioned that because $\rho$ is not assumed to be smooth, many proofs below are necessarily different from those used in KQ.

These findings are thus different from those in the restricted or in the unrestricted contiguous non-random design cases. It is known (see, e.g., van de Geer 1988, Example 6.6), that in the uniform non-random design two-phase unrestricted fixed jump size case the limiting distribution of the standardized change-point estimator is determined by a Brownian motion with a linear drift. It thus follows that the main reason for the difference between these results is the randomness of the design which forces the weak limit of the relevant part of the M-process to be a compound Poisson process.

The paper is organized as follows. Section 2 describes the model and a computational scheme for M-estimators. Section 3 lists all assumptions on $\rho$, the error distribution function, the design variable and contains the proofs of the consistency with and without
a rate, while Section 4 derives the limiting distributions of these estimators. It is perhaps worth mentioning that Theorem 4.2 and Corollary 4.1 give results of some general interest. Section 5 consists of two subsections: The Section 5.1 reports a simulation study while an application of a two-phase regression model to some automobile data is given in Section 5.2.

2. Model and M-estimators

Let
\[ m(x, \vartheta) = (x_0 + x_1 x)I(x \leq s) + (\beta_0 + \beta_1 x)I(x > s), \]
\[ x \in \mathbb{R}, \quad \vartheta = (x_0, x_1, \beta_0, \beta_1, s) \in \mathbb{R}^5, \]
denote the two-phase linear regression function. We consider a set of independent observations \((X_i, Y_i), i = 1, \ldots, n\), such that for some \(\vartheta = (\theta'_1, r)' = (a_0, a_1, b_0, b_1, r)' \in \mathbb{R}^5\),
\[ Y_i - m(X_i, \vartheta) = \epsilon_i, \quad i = 1, \ldots, n \tag{2.1} \]
are independent identically distributed (i.i.d.) random variables. The jump size at the true jump-point \(r\) in the regression function \(m\) is given by \(d = b_0 - a_0 + r(b_1 - a_1)\). We shall make the usual identifiability assumption that the two line segments are different and that \(d\) is fixed and non-zero, i.e.,
\[ d \neq 0. \tag{2.2} \]

It is convenient to write \(m(x, \vartheta) = m_s(x, \vartheta_1)\), for \(\vartheta = (\theta'_1, s)\)' with \(\vartheta_1 \in \mathbb{R}^4, \quad s \in \mathbb{R}\), and refer to \(\vartheta_1\) and \(s\) as the coefficient and the change-point parameters, respectively. Let
\[ \hat{m}_s(x) \equiv (\hat{\hat{\vartheta}}/\hat{\vartheta}\vartheta_1)(m_s(x, \vartheta_1)) \]
\[ = (I(x \leq s), \quad xI(x < s), \quad I(x > s), \quad xI(x > s))', \quad s \in \mathbb{R}, \quad x \in \mathbb{R}, \]
denote the vector of partial derivatives of \(m_s(x, \vartheta_1)\) with respect to \(\vartheta_1\). Observe that \(m(x, \vartheta) \equiv \vartheta'_1 \hat{m}_s(x)\).

To define M-estimators of \(\theta\), we need to compactify the real line. Let \(\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}\). The set \(\overline{\mathbb{R}}\) is compact under the metric \(d(x, y) = |\arctan x - \arctan y|, \quad x, y \in \mathbb{R}\). Throughout we assume that \(\theta\) is an interior point of the parameter space \(\Theta = K \times \overline{\mathbb{R}}\) for a known compact set \(K\) in \(\mathbb{R}^4\). A typical point in \(\Theta\) will be denoted by \(\vartheta = (x_0, x_1, \beta_0, \beta_1, s)' = (\theta'_1, s)'\). Note that \(s = -\infty\) or \(s = \infty\) means that there is no change in regression.

Define the M-process corresponding to a function \(\rho: \mathbb{R} \rightarrow [0, \infty)\) by
\[ M_n(\vartheta) = \sum_{i=1}^{n} \rho(Y_i - \vartheta'_1 \hat{m}_s(X_i)), \quad \vartheta_1 \in \mathbb{R}^4, \quad s \in \overline{\mathbb{R}}. \tag{2.3} \]
A measurable map \(\hat{\theta}_n = \hat{\theta}_n((X_1, Y_1), \ldots, (X_n, Y_n))\) from \(\mathbb{R}^{2n} \rightarrow \Theta\), is an M-estimator, if \(M_n(\hat{\theta}_n) = \inf_{\vartheta \in \Theta} M_n(\vartheta)\), a.s. Often we shall write \(M_n(\vartheta_1, s)\) for \(M_n(\vartheta)\).

Note that the function \(M_n(\vartheta_1, s)\) is not continuous in \(s\), but because of (2.2), for each \(\vartheta_1, \ M_n(\vartheta_1, s)\) as a function of \(s\) is constant on the intervals \((X_{(i-1)}, X_{(i)}], 1 \leq i \leq n+1,\)
where \( \{X_i, 1 \leq i \leq n\} \) are the ordered design variables with \( X_0 = -\infty, X_{n+1} = \infty \). Thus, to compute the M-estimator, proceed as follows: First, for each fixed \( s \), obtain the minimizer \( \vartheta_1(s) \) of \( M_n(\vartheta_1, s) \) with respect to \( \vartheta_1 \) over \( K \). Notice that \( \vartheta_1(s) \) is constant in \( s \) over any interval of two consecutive ordered \( X_i \)'s and that the profile M-process \( M_n(\vartheta_1, s) \) has only finite number of possible values with the change-points located at \( X_i \)'s. At the second stage, compute minimizer \( \hat{\vartheta}_n \) of \( M_n(\vartheta_1, s) \) with respect to \( s \) over \( \{X_i, 1 \leq i \leq n\} \). To make it unique we take this minimizer to be the left end point of the interval over which it is obtained. The associated \( \hat{\vartheta}_1n = \hat{\vartheta}_1n \) becomes the M-estimator of \( \theta_1 \). Then the estimator \( \hat{\theta}_n = (\hat{\vartheta}_1n, \hat{\vartheta}_n) \equiv (\hat{\vartheta}_1n, \hat{\vartheta}n) \) is the M-estimator of the underlying parameter \( \theta \).

3. Consistency

To begin with, we shall state the needed assumptions. Let \( G \) denote the distribution function of the design variable \( X \). Consider the following assumptions:

(a.1) \( \rho \) is convex on \( \mathbb{R} \) with right-continuous non-decreasing almost everywhere derivative \( \psi \) satisfying \( E\psi^2(\varepsilon + y) < \infty, \forall y \in \mathbb{R} \). Moreover, the function \( \lambda(y) := E\psi(\varepsilon + y), y \in \mathbb{R} \), is strictly monotone increasing on \( \mathbb{R} \) and \( \lambda \) is continuous at 0 with \( \hat{\lambda}(0) = 0 \).

(a.2) \( E\psi^2(\varepsilon + c_1 + c_2|X|) < \infty, \forall c_1, c_2 \in \mathbb{R} \).

(a.3) \( G \) has Lebesgue density \( g \) that is continuous and positive at \( r \) and \( EX^2 < \infty \).

(a.4) The function \( y \mapsto E[\psi(\varepsilon + c + y) - \psi(\varepsilon)] \) is continuous at 0, for every \( c \in \mathbb{R} \).

(a.5) The function \( \lambda \) is differentiable in a neighborhood of 0, with its derivative \( \hat{\lambda} \) satisfying \( \hat{\lambda}(0) \neq 0 \), and \( \lim_{\varepsilon \to 0} \varepsilon^{-1} \int_0^\varepsilon |\hat{\lambda}(s) - \hat{\lambda}(0)| \, ds = 0 \).

(a.6) The random variables \( \rho(\varepsilon + d) - \rho(\varepsilon) \) are continuous.

Remark 3.1. On assumptions (a.1)–(a.6). The above assumptions are formulated so as to balance the non-smoothness of \( \psi \) with the smoothness of the error distribution or vice versa.

Also, observe that \( \psi \) non-decreasing and the continuity of \( \lambda \) at 0 implies that the error distribution has zero mass at the jump points of \( \psi \). In particular (a.1) implies the following:

\[
\text{The function } y \mapsto E\{\psi(\varepsilon + y) - \psi(\varepsilon)\}^2 \text{ is continuous at 0. (3.1)}
\]

This will be needed in the sequel.

As will be seen below, the assumptions (a.1)–(a.3) suffice for the consistency of \( \hat{\theta}_n \) while (a.1)–(a.5) are used to obtain the \( n^{-} \)- and \( n^{1/2} \)-consistency of \( \hat{\vartheta}_n \) and \( \hat{\theta}_1n \), respectively, and the limiting distribution of \( n^{1/2}(\hat{\theta}_1n - \theta_1) \). The additional assumption (a.6) is needed only for obtaining the limiting distribution of \( n(\hat{\vartheta}_n - r) \). Of course the continuity of the error distribution and (a.1) with \( \psi \) strictly increasing implies (a.6).

In general, (a.1), (a.2), (a.4) and (a.5) together need not imply (a.6). This may be seen by taking \( \rho(y) = |y|I(|y| > 1) + I(|y| \leq 1) \), and the error d.f. \( F \) to be symmetric around zero and differentiable with positive derivative on \( \mathbb{R} \). Then, \( \psi(y) \equiv -I \)
By the increasing property of arctan function, we have
\[ P(\rho(\varepsilon + d) - \rho(\varepsilon) = 0) \geq P(|\varepsilon + d| \leq 1, |\varepsilon| \leq 1) = P(-1 \leq \varepsilon \leq 1 - d) > 0 \]
so that (a.6) does not hold. \( \square \)

Let \( \| \cdot \| \) denote the Euclidean norm. Note that for all \( x \in \mathbb{R} \), \( \vartheta^* = (\vartheta_1^*, s^*) \), \( \vartheta = (\vartheta_1', s') \in \mathbb{R}^d \times \mathbb{R} \),
\[
\| \hat{m}_s(x) \| = \sqrt{1 + x^2} \leq 1 + |x|,
\| m(x, \vartheta) \| \leq \| \vartheta_1 \| \sqrt{1 + x^2},
\| m_s(x, \vartheta_1) - m_s(x, \vartheta_1^*) \| \leq \| \vartheta_1 - \vartheta_1^* \| \sqrt{1 + x^2},
\| \hat{m}_s(x) - \hat{m}_s^*(x) \| \leq \sqrt{2(1 + x^2)}I(s \wedge s^* < x \leq s \vee s^*),
\leq \sqrt{2(1 + x^2)}I(|x - s| \leq |s^* - s|).
\]
In the sequel, \( C \) denotes a generic positive finite constant that may be different in different consistency context, but will never depend on \( n \). We are now ready to state the strong consistency result.

**Theorem 3.1.** Suppose (2.1), (2.2), and (a.1)–(a.3) hold. Then, \( \hat{\theta}_n \to \theta \), a.s., as \( n \to \infty \).

The following lemma is needed for the proof of the above theorem. Let \( U_\eta(\vartheta) = \{ \vartheta^* = (\vartheta_1^*, s^*) \in \Theta : \| \vartheta_1^* - \vartheta_1 \| < \eta, d(s^*, s) < \eta \} \) denote an \( \eta \)-neighborhood of \( \vartheta \), \( \eta > 0 \).

**Lemma 3.1.** Under the assumptions of Theorem 3.1, for any \( \vartheta \in \Theta \),
\[
E \sup_{\vartheta^* \in U_\eta(\vartheta)} |\rho(Y - m(X, \vartheta^*)) - \rho(Y - m(X, \vartheta))| \to 0, \quad \text{as } \eta \to 0.
\]

**Proof.** Fix a \( \vartheta \in \Theta \) and let
\[
\delta(X, \vartheta) := m(X, \vartheta) - m(X, \vartheta), \quad \varepsilon(\vartheta) := Y - m(X, \vartheta) = \varepsilon + \delta(X, \vartheta).
\]
By the increasing property of arctan function, we have
\[
I(s \wedge s^* < x \leq s \vee s^*) \leq I(d(x, s) \leq d(s^*, s))
\]
For \( \vartheta^* \in U_\eta(\vartheta) \) and for an \( s \in \mathbb{R} \), (3.2) and this inequality imply
\[
|m(X, \vartheta) - m(X, \vartheta^*)| \leq |\vartheta_1| |\hat{m}_s(X) - \hat{m}_s^*(X)| + \| \vartheta_1 - \vartheta_1^* \| |\hat{m}_s^*(X)| \leq \sqrt{2\| \vartheta_1 \| I(d(X, s) \leq d(s^*, s)) \| \vartheta_1 - \vartheta_1^* \| \sqrt{1 + X^2}}
\]
By (a.1) and (3.4), we obtain

\[ \leq \sqrt{2} \| \vartheta_1 \| I(d(X, s) \leq \eta) - \eta(1 + |X|) \]

\[ \equiv A_\eta(X) \quad \text{(say).} \] (3.4)

By (a.1) and (3.4), we obtain

\[ \sup_{\vartheta^* \in U \vartheta \vartheta(\vartheta)} |\rho(Y - m(X, \vartheta^*)) - \rho(Y - m(X, \vartheta))| \leq \int_{-A_\eta(X)}^{A_\eta(X)} |\psi(\varepsilon(\vartheta) + v)| \, dv. \]

Similarly to (3.4), one obtains \( |\delta(X, \vartheta)| \leq C(1 + |X|) \), and hence \( |\varepsilon(\vartheta) + v - \varepsilon| \leq C(1 + |X|) + A_\eta(X) \), for all \( |v| \leq A_\eta(X) \). But note also that for all \( \eta < 1 \), \( A_\eta(X) \leq C(1 + |X|) \).

Hence, by the non-decreasing property of \( \psi \),

\[ |\psi(\varepsilon(\vartheta) + v)| \leq |\psi(\varepsilon + C(1 + |X|))| + |\psi(\varepsilon - C(1 + |X|))|, \quad \forall 0 < \eta < 1. \]

This and the Cauchy–Schwarz inequality imply that the left-hand side of (3.3) is bounded above by \( C\{E^{1/2}|\psi|^2(\varepsilon + C(1 + |X|)) + E^{1/2}|\psi|^2(\varepsilon - C(1 + |X|))\}E^{1/2}A_\eta^2(X) \). Since \( s \in \mathbb{R}, -\pi/2 < \arctan(s) < \pi/2 \), and for a small enough \( \eta (\arctan(s) - \eta, \arctan(s) + \eta) \subseteq (-\pi/2, \pi/2) \), hence, (a.3) and the continuity of \( \tan \) on \( (-\pi/2, \pi/2) \) imply

\[ E A_\eta^2(X) = E(1 + X^2)(\sqrt{2} \| \vartheta_1 \| I(d(X, s) \leq \eta) + \eta^2) \rightarrow 0, \quad \text{as } \eta \rightarrow 0. \]

In view of (a.2), this proves (3.3) for any \( s \in \mathbb{R} \).

In the case \( s = \infty \), by (3.2),

\[ |m(X, \vartheta) - m(X, \vartheta^*)| \leq (\sqrt{2} \| \vartheta_1 \| I(X > s^*) + \eta)(1 + |X|) \equiv A_{1\eta}(X), \]

where \( d(s^*, \infty) < \eta \). Again, the increasing property of \( \arctan \) and \( E|X|^2 < \infty \) imply that \( E A_{1\eta}(X) \rightarrow 0 \), as \( \eta \rightarrow 0 \). The proof is similar in the case \( s = -\infty \), except one replaces \( I(X > s^*) \) by \( I(X < s^*) \) in the above inequality. This completes the proof of Lemma 3.1. \( \square \)

**Proof of Theorem 3.1.** The method of proof is similar to that in Huber (1967). We give details for the sake of completeness.

Let \( h(x) := \sqrt{2} \| \vartheta_1 \| + \| \vartheta_1 \| (1 + |x|) \). By (a.1) and (3.2),

\[ |\rho(Y - m(X, \vartheta)) - \rho(Y - m(X, \vartheta))| \leq \int_{-h(X)}^{h(X)} |\psi(\varepsilon + v)| \, dv. \]

From this it follows that \( E|\rho(Y - m(X, \vartheta)) - \rho(Y - m(X, \vartheta))| < \infty, \forall \vartheta \in \Theta \). Hence the function \( \alpha(\vartheta) = E[\rho(Y - m(X, \vartheta)) - \rho(Y - m(X, \vartheta))] \) for \( \vartheta \in \Theta \) is well defined with \( \alpha(\theta) = 0 \). In view of Lemma 3.1, \( \alpha \) is continuous on \( \Theta \).

Now, by (2.2) and (a.3), for any \( \eta > 0 \), there exists \( x_0 > 0 \), such that

\[ P(|\delta(X, \vartheta)| > x_0) > 0, \quad \forall \vartheta \neq \theta. \] (3.5)

By (a.1), the function \( \ell(x) := E[\rho(x + \varepsilon) - \rho(x)] > 0, \quad x \neq 0 \). Therefore, by (3.5),

\[ \alpha(\vartheta) = E[\rho(\varepsilon + \delta(X, \vartheta)) - \rho(\varepsilon)]|X| = E\ell(\delta(X, \vartheta)) > 0, \quad \forall \vartheta \neq \theta. \]

This and the continuity of \( \alpha \) thus implies that for any neighborhood \( V \) of \( \theta \) there exists \( \vartheta_0 \in V^c \), such that

\[ \inf_{\vartheta \in V^c} \alpha(\vartheta) = \alpha(\vartheta_0) > 0. \]
Let \( \delta_0 = \kappa(\delta_0)/3 > 0 \). For any \( \vartheta \in \mathcal{V}^c \), by Lemma 3.1 again, there exists \( \eta_0 > 0 \), such that

\[
E \inf_{\vartheta^* \in U_{\eta_0}(\vartheta)} [\rho(\varepsilon + \delta(X, \vartheta^*)) - \rho(\varepsilon)] \geq E[\rho(\varepsilon + \delta(X, \vartheta)) - \rho(\varepsilon)] - \delta_0 \\
\geq \kappa(\delta_0) - \delta_0 = 2\delta_0.
\] (3.6)

Recall the definition of \( M_n \) from (2.3). Again, the compactness of \( \mathcal{V}^c \) implies that there exists a finite number \( k \) of \( U_{\eta_0}(\vartheta_j), \vartheta_j \in \mathcal{V}^c, j = 1, 2, \ldots, k \), such that \( U_k = U_{\eta_0}(\vartheta_j) = \mathcal{V}^c \). Then by the strong law of large number and (3.6), we have almost surely for sufficiently large \( n = n(\omega) \), and all \( 1 \leq j \leq k \),

\[
\inf_{\vartheta \in U_{\eta_0}(\vartheta_j)} \frac{1}{n} [M_n(\vartheta) - M_n(\vartheta_j)] \geq E \inf_{\vartheta \in U_{\eta_0}(\vartheta_j)} [\rho(\varepsilon + \delta(X, \vartheta)) - \rho(\varepsilon)] - \delta_0 \geq \delta_0.
\]

Hence, the fact that \( \inf_{\vartheta \in \mathcal{V}} [M_n(\vartheta) - M_n(\vartheta_j)] \leq 0, \forall n \geq 1 \), implies that for any neighborhood \( \mathcal{V} \) of \( \vartheta \) in \( \Theta \), almost surely for sufficiently large \( n \), \( \inf_{\vartheta \in \mathcal{V}} M_n(\vartheta) > \inf_{\vartheta \in \mathcal{V}} M_n(\vartheta_j) \), thereby completing the proof. \( \square \)

The next result gives the \( n \)- and \( n^{1/2} \)-consistency of the estimators \( \hat{\vartheta}_n \) and \( \hat{\theta}_1n \).

**Theorem 3.2.** In addition to (2.1) and (2.2), suppose (a.1)–(a.5) hold. Then,

\[
|n(\hat{\vartheta}_n - \vartheta)| = O_p(1), \quad \|n^{1/2}(\hat{\theta}_1n - \theta_1)\| = O_p(1).
\] (3.7)

**Remark 3.2.** Note that for the least square estimator, i.e., when \( \rho(x) \equiv x^2 \), the conditions (a.1) and (a.2) reduce to requiring only \( E(\varepsilon^2 + X^2) < \infty \), while (a.4)–(a.6) are trivially satisfied. It is perhaps worth comparing this finding with that of Example 6.6 of van de Geer (1988) which shows that in the uniform non-random design case the least square estimator needs at least twelve finite error moments for (3.7) to hold.

The proof of Theorem 3.2 needs some preliminaries. Accordingly, let \( J : \mathbb{R}^2 \to \mathbb{R} \) such that \( EJ^2(x, \varepsilon) < \infty \) for every \( x \in \mathbb{R} \). Define, for \( x \in \mathbb{R} \) and \( u \geq 0 \),

\[
p(x) = EJ(x, \varepsilon), \quad p_1(x) = E|J(x, \varepsilon)|, \quad p_2(x) = EJ^2(x, \varepsilon),
\]

\[
K(u) = EI(r < X \leq r + u), \quad K_n(u) = n^{-1} \sum_{i=1}^n I(r < X_i \leq r + u),
\]

\[
R_n(u) = n^{-1} \sum_{i=1}^n J(X_i, \varepsilon_i)I(r < X_i \leq r + u),
\]

\[
r_n(u) = n^{-1} \sum_{i=1}^n p(X_i)I(r < X_i \leq r + u).
\]
We now state the first result needed as

Lemma 3.2. Suppose that the function \( p_2 \) is bounded on bounded intervals. Then, for each \( \gamma > 0, \eta > 0 \), there is a constant \( 0 < B < \infty \), such that \( \forall \ 0 < \delta < 1 \) and \( \forall n \geq [B/\delta] + 1 \),

\[
P \left( \sup_{B/n < u \leq \delta} \frac{K_n(u) - 1}{K(u)} < \eta \right) > 1 - \gamma, \tag{3.8}
\]

\[
P \left( \sup_{B/n < u \leq \delta} \frac{R_n(u) - r_n(u)}{K(u)} < \eta \right) > 1 - \gamma. \tag{3.9}
\]

A proof of this lemma appears in Koul and Qian (2002); see Lemma 3.2 there.

We need to apply this lemma to \( J(x,z) = \rho(z + d_0 + d_1 x) - \rho(z) \) where \( d_0 = b_0 - a_0, \ d_1 = b_1 - a_1 \). Let

\[
S_n(u) := n^{-1} \sum_{i=1}^{n} \left[ \rho(\varepsilon_i + d_0 + d_1 X_i) - \rho(\varepsilon_i) \right] I(r < X_i \leq r + u), \tag{3.10}
\]

\[
s_n(u) := n^{-1} \sum_{i=1}^{n} E \left[ \left[ \rho(\varepsilon_i + d_0 + d_1 X_i) - \rho(\varepsilon_i) \right] |X_i| \right] I(r < X_i \leq r + u)
\]

\[
:= n^{-1} \sum_{i=1}^{n} p(X_i) I(r < X_i \leq r + u).
\]

Note that under (a.1), for all \( |x| \leq c < \infty, \ |d_0 + d_1 x| \leq |d_0| + |d_1| c = c_1 \), and the corresponding

\[
p_2(x) = E[J^2(x,\varepsilon)] \leq E \left( \int_{-|d_0 + d_1 x|}^{|d_0 + d_1 x|} |\psi(\varepsilon + z)| \, dz \right)^2
\]

\[
\leq 2c_1^2 \ E[\psi^2(\varepsilon + c_1) + \psi^2(\varepsilon - c_1)],
\]

thereby proving that \( p_2 \) is bounded on bounded intervals. This validates the application of the above Lemma 3.2 to this \( J \) and enables us to conclude the following.

Lemma 3.3. Suppose (a.1) holds. Then, for each \( \gamma > 0, \eta > 0 \), there is a constant \( 0 < B < \infty \), such that \( \forall \ 0 < \delta < 1 \) and \( \forall n \geq [B/\delta] + 1 \),

\[
P \left( \sup_{B/n < u \leq \delta} \left| \frac{S_n(u) - s_n(u)}{K(u)} \right| < \eta \right) > 1 - \gamma. \tag{3.11}
\]

We also need the following asymptotic uniform quadraticity result in the coefficient parameter. Let

\[
D_{n1}(w_1) := M_n(\theta_1 + n^{-1/2}w_1, r) - M_n(\theta_1, r)
\]

\[
= \sum_{i=1}^{n} \left[ \rho(\varepsilon_i - n^{-1/2}w_1 m_r(X_i)) - \rho(\varepsilon_i) \right], \quad w_1 \in \mathbb{R}^4. \tag{3.12}
\]
Since \( r \) is fixed, \( D_{n1}(w_1) \) involves only the coefficient parameters. From the results available in linear regression, cf. Heiler and Willers (1988) or Koul (2002, Examples 1 and 2, Chapter 5.4), one obtains the following.

**Lemma 3.4.** Suppose (2.1), (2.2), (a.1), (a.2) with \( c_2 = 0 \), and (a.3)–(a.5) hold. Then, for every \( 0 < b < \infty \),

\[
\sup_{\|w_1\| \leq b} \left| D_{n1}(w_1) + n^{-1/2} w_1' \sum_{i=1}^{n} \hat{m}_r(X_i) \psi(e_i) - \frac{\dot{\gamma}(0)}{2} w_1' E[\hat{m}_r(X) \hat{m}_r(X)'] w_1 \right| = o_p(1).
\]

A proof of this lemma uses (3.1) also.

Next, define

\[
D_{n2}(\vartheta_1, s) := M_n(\vartheta_1, s) - M_n(\vartheta_1, r),
\]

\[
D_n(\vartheta_1, s) := M_n(\vartheta_1, s) - M_n(\vartheta_1, r), \quad \vartheta_1 \in \mathbb{R}^4, \quad s \in \mathbb{R},
\]

\[
\Omega(\delta) := \{ \vartheta \in \Theta : \| \vartheta_1 - \vartheta_1 \| < \delta, |s - r| < \delta \}, \quad \delta > 0.
\]

Note the decomposition

\[
D_n(\vartheta_1, s) \equiv D_{n1}(n^{1/2}(\vartheta_1 - \vartheta_1)) + D_{n2}(\vartheta_1, s). \tag{3.13}
\]

We are now ready to give the

**Proof of Theorem 3.2.** Let, for a \( \delta > 0 \), \( 0 < b < \infty \),

\[
\mathcal{N}_{1b} := \{ \vartheta \in \Omega(\delta) : |s - r| > b/n \}, \quad \mathcal{N}_{2b} := \{ \vartheta \in \Omega(\delta) : \| \vartheta_1 - \vartheta_1 \| > b/n^{1/2} \}.
\]

Because \( \inf_{\vartheta \in \mathcal{N}_{1b} \cup \mathcal{N}_{2b}} D_n(\vartheta_1, s) = \min\{ \inf_{\vartheta \in \mathcal{N}_{1b}} D_n(\vartheta_1, s), \inf_{\vartheta \in \mathcal{N}_{2b}} D_n(\vartheta_1, s) \} \), it suffices to show that \( \forall \gamma > 0, \ 0 < c_1 < \infty, \ 0 < c_2 < \infty, \) there exists a \( 0 < b < \infty \), and an \( N \) such that

\[
P\left( \inf_{\vartheta \in \mathcal{N}_{1b}} D_n(\vartheta_1, s) > c_1 \right) > 1 - \gamma, \tag{3.14}
\]

\[
P\left( \inf_{\vartheta \in \mathcal{N}_{2b}} D_n(\vartheta_1, s) > c_2 \right) > 1 - \gamma, \quad \forall n > N. \tag{3.15}
\]

Clearly, from (3.13) we have

\[
\inf_{\vartheta \in \mathcal{N}_{ab} \cup \mathcal{N}_{ib}} D_n(\vartheta_1, s) \geq \inf_{\vartheta \in \mathcal{N}_{1b}} D_{n1}(n^{1/2}(\vartheta_1 - \vartheta_1)) + \inf_{\vartheta \in \mathcal{N}_{ib}} D_{n2}(\vartheta_1, s), \quad i = 1, 2. \tag{3.16}
\]

We shall show that for \( i = 1 \), the first term in the above lower bound is \( O_p(1) \) and the second term is arbitrarily large and positive with arbitrarily large probability for all sufficiently large \( n \). A similar statement will be proved for \( i = 2 \) with the role of the first and second term reversed.

Consider the second term in the case \( i = 1 \). Since \( \hat{\vartheta}_n \) is strongly consistent by Theorem 3.1, without loss of generality, the parameter space in this part of the proof will be
restricted to a neighborhood $\Omega(\delta)$ of $\theta$, for some $0 < \delta < 1$ to be determined later. Let 
\[ \mathcal{N}_{1\delta} := \Omega(\delta) \cap \mathcal{N}_{1b}. \]

It suffices to show that for all $\gamma > 0$, $0 < c_1 < \infty$, $\exists a \gamma_0 < \infty$, $0 < b_0 < \infty$, $0 < \delta < 1$ and an $n_0$ such that $\gamma_0 b_0 g(r)/2 > c_1$ and that
\[ P \left( \inf_{\tilde{\vartheta} \in \mathcal{A}_{1\delta_0}} \frac{D_{n2}(\tilde{\vartheta}_1, s)}{nK(|s-r|)} > \gamma_0 \right) > 1 - \gamma/2, \quad \forall n > n_0. \quad \tag{3.17} \]

To see this, first, by (a.3), choose a positive $\delta$ sufficiently small such that $\inf_{r \leq x \leq r+\delta} g(x) \geq g(r)/2$. Now let $b_0$ and $\gamma_0$ be as above and note that for $n > b_0/\delta$, $\inf_{r \leq x \leq r+b_0/n} g(x) \geq g(r)/2$. Hence, by (3.17), we obtain
\[ P \left( \inf_{\tilde{\vartheta} \in \mathcal{A}_{1\delta_0}} D_{n2}(\tilde{\vartheta}_1, s) > c_1 \right) \]
\[ \geq P \left( \inf_{\tilde{\vartheta} \in \mathcal{A}_{1\delta_0}} \frac{D_{n2}(\tilde{\vartheta}_1, s)}{nK(|s-r|)} > \frac{2c_1}{b_0 g(r)} \right) \]
\[ \geq P \left( \inf_{\tilde{\vartheta} \in \mathcal{A}_{1\delta_0}} \frac{D_{n2}(\tilde{\vartheta}_1, s)}{nK(|s-r|)} > \gamma_0 \right) > 1 - \gamma/2, \quad \forall n > n_0 \lor (b_0/\delta). \]

We begin to prove (3.17). The details will be given for the case $s > r$ only, they being similar for the case $s < r$. Write $s = r + u$ for some $u > 0$. Let $A_1 := \vartheta_1 - \theta_1$, $\alpha(x) := (-1, -x, 1, x)'$, $b(x) := (0, 0, 1, x)'$, $c(x) := (1, x, 0, 0)'$, $x \in \mathbb{R}$. Note that $a(x) \equiv b(x) - c(x)$ and $\alpha'_1 a(x) = d_0 + d_1 x$, with $d_0 = b_0 - a_0$, $d_1 = b_1 - a_1$. Also, let
\[ S_{n1}(\vartheta_1, u) := n^{-1} \sum_{i=1}^{n} \left[ \rho(\varepsilon_i + \alpha'_1 a(X_i)) - \rho(\varepsilon_i + \alpha'_1 a(X_i)) \right] \]
\[ \times I(r < X_i \leq r + u), \]
\[ S_{n2}(\vartheta_1, u) := n^{-1} \sum_{i=1}^{n} \left[ \rho(\varepsilon_i) - \rho(\varepsilon_i - A'_1 b(X_i)) \right] I(r < X_i \leq r + u), \]
\[ \mathcal{X}_{n1}(u) := S_n(u) - s_n(u), \quad \mathcal{X}_{n2}(u) := n^{-1} \sum_{i=1}^{n} \left[ p(X_i) - p(r) \right] I(r < X_i \leq r + u), \]
where $S_n$, $s_n$ are as in (3.10). Direct calculations show that
\[ n^{-1} D_{n2}(\vartheta_1, s) \equiv p(r) K(u) + p(r) [K_n(u) - K(u)] + \mathcal{X}_{n1}(u) + \mathcal{X}_{n2}(u) \]
\[ + S_{n1}(\vartheta_1, u) + S_{n2}(\vartheta_1, u). \quad \tag{3.18} \]

As will be shown below, the first term of this decomposition makes the main contribution towards (3.17), while the remaining terms are negligible, as $n \to \infty$, and then $\delta \to 0$.

Now, recall that $d = d_0 + d_1 r$. By Fubini and (a.1),
\[ p(r) = E[\rho(\varepsilon + d) - \rho(\varepsilon)] = \begin{cases} \int_{0}^{d} \lambda(z) \, dz, & d > 0, \\
- \int_{d}^{0} \lambda(z) \, dz, & d < 0. \end{cases} \]
Hence, in view of the assumption that $\lambda$ is strictly increasing and $\lambda(0) = 0$, ensured by (a.1),

$$p(r) > 0.$$  \(\text{\(}\)\)  

Also, for all $u \leq \delta$, $|D_n(u)| \leq \sup_{0 \leq v \leq \delta} |p(r + v) - p(r)| K_n(u)$. Thus, by (3.8) and the continuity of $p$ (implied by (a.1)), $\sup_{0 \leq u \leq \delta} |D_n(u)|/K_n = o_p(1)$, as $n \to \infty$ and then $\delta \to 0$. By Lemma 3.3, a similar statement holds for $\mathcal{Z}_{n1}$.

In a similar way, one can handle the two remaining terms $S_{n1}(\theta_1, u)$, $S_{n2}(\theta_1, u)$ on the right-hand side of (3.18). We have, $|\mathcal{A}_i'(x)| \leq C_\delta \leq 1$, for all $\theta_1 \in N_{1,\delta}$, $r < x \leq r + u$, $u \leq \delta$ and for all sufficiently small $0 < \delta < 1$. Hence, by (a.1),

$$\left|S_{n1}(\theta_1, u)\right| \leq n^{-1} \sum_{i=1}^{n} \int_{-C_\delta}^{C_\delta} |\psi(c_i + \theta_1' a(X_i) + t)| \, dt \left( r < X_i \leq r + u \right).$$

Now apply Lemma 3.2 with $J(x, z) := \int_{-C_\delta}^{C_\delta} |\psi(z + \theta_1' a(x) + t)| \, dt$ to obtain that there exists a $B_1$, such that $\sup_{0 \leq v \leq \delta} |S_{n1}(\theta_1, u)|/K(u) = o_p(1)$, as $n \to \infty$, and then $\delta \to 0$. Similarly, there is a $B_2$, such that $\sup_{0 \leq v \leq \delta} |S_{n2}(\theta_1, u)|/K(u) = o_p(1)$, as $n \to \infty$, and then $\delta \to 0$.  

Now, choose an $0 < \eta < p(r)/(4 + p(r))$. From the above facts, (3.8) and (3.18) we readily obtain (3.17) with $\gamma_0 := [(p(r) - \eta(4 + p(r)))/2 > 0$ and $b_0 = \max\{B, B_1, B_2\}$, where $B$ is as in (3.8) and (3.11), in a routine fashion. Now we turn to the first term for the case $i = 2$ in the lower bound (3.16). Again, because $D_{n1}(w_1)$ of (3.12) involves only the coefficient parameters, from the results available in linear regression and using the convexity of $p$, we readily obtain that for every $\gamma > 0$, $0 < \gamma_1 < \infty$, there is a $0 < b_1 < \infty$, and an $N_\gamma$, such that

$$P \left( \inf_{\|w_1\| > b_1} D_{n1}(w_1) > \gamma_1 \right) > 1 - \gamma/2, \quad \forall n > N_\gamma.$$  \(\text{\(}\)\)  

For the same reasons we also have

$$\inf_{\theta \in \mathcal{N}_{1b}} D_{n1}(n^{1/2}(\theta_1 - \theta_1)) = \inf_{\theta_1 \in \mathcal{K}} D_{n1}(n^{1/2}(\theta_1 - \theta_1))$$

$$= \min \left\{ \inf_{\|w_1\| \leq b_1} D_{n1}(w_1), \inf_{\|w_1\| > b_1} D_{n1}(w_1) \right\} = o_p(1),$$

by (3.19) and Lemma 3.4. This proves that the first term in the lower bound (3.16) for the case $i = 1$ is bounded in probability, which in turn together with (3.17) completes the proof of (3.14).

It remains to prove $\inf_{\theta \in \mathcal{N}_{1b}} D_{n2}(\theta_1, s) = o_p(1)$. For this purpose, we further decompose

$$\inf_{\theta \in \mathcal{N}_{1b}} D_{n2}(\theta_1, s) = \min \left\{ \inf_{\theta \in \mathcal{N}_{1b}} D_{n2}(\theta_1, s), \inf_{\theta \in \mathcal{N}_{1b}} D_{n2}(\theta_1, s) \right\},$$

where $\mathcal{N}_{i_1b} := \{ \theta \in \Theta; |s - r| \leq b/n \}$, with the $b$ being as in (3.14). By the convexity of $p$, Lemma 4.1 and Corollary 4.2 below,

$$\inf_{\theta \in \mathcal{N}_{1b}} D_{n2}(\theta_1, s) = \inf_{|r| \leq b, \|w_1\| = b} D_{n2}(\theta_1 + n^{1/2} w_1, r + n^{-1} t) = o_p(1).$$
The inequality $\inf_{\theta \in \mathcal{E}_{a} \cap \mathcal{E}_{b}} D_{n,2}(\hat{\theta}, s) \geq \inf_{\theta \in \mathcal{E}_{1b}} D_{n,2}(\theta, s)$ and (3.17) imply that the first term on the right-hand side of (3.20) is large with arbitrarily large probability for all sufficiently large $n$. Hence, $\inf_{\theta \in \mathcal{E}_{a} \cap \mathcal{E}_{b}} D_{n,2}(\hat{\theta}, s) = O_p(1)$. This and (3.19) completes the proof of (3.15), hence that of the Theorem 3.2. □

4. Limiting distributions

The main focus of this section is to obtain the joint limiting distributions of the M-estimators. In the process we also obtain some auxiliary results of general interest. The main result of this section is given in Theorem 4.1 below. To facilitate its statement we need the following notation.

$$w_{1n} := n^{1/2}(\hat{\theta}_{1n} - \theta_1), \quad t_n = n(\hat{r}_n - r), \quad \mathcal{Z}_n := n^{-1/2} \sum_{i=1}^{n} \tilde{m}_i(X_i)\psi(\epsilon_i),$$

$$\Gamma_r := E[\tilde{m}_i(X)\tilde{m}_i(X)'], \quad \tau^2 := E\psi^2(\epsilon)/\hat{\lambda}^2(0).$$

**Theorem 4.1.** Suppose that (2.1), (2.2) and (a.1)--(a.6) hold. Then

$$w_{1n} = (\hat{\lambda}(0))^{-1} \Gamma_r^{-1} \mathcal{Z}_n + o_p(1). \quad (4.1)$$

Moreover, $(w_{1n}, t_n) \rightarrow_d (\mathcal{Z}, \pi_-)$, where $\mathcal{Z}$ is a $\mathcal{N}_4(0, \tau^2 \Gamma_r^{-1})$ r.v., independent of $\pi_-$, the smallest minimizer of the process $\Pi$. Here,

$$\Pi(t) = \mathcal{P}_1(t)I(t \geq 0) + \mathcal{P}_2(-t)I(t \leq 0), \quad (4.2)$$

$\mathcal{P}_1, \mathcal{P}_2$ are two compound Poisson processes on $[0, \infty)$, with $\mathcal{P}_1(0) = 0 = \mathcal{P}_2(0)$, both having the common rate $g(r)$, and their jumps having the same distribution as that of $\rho(\epsilon + d) - \rho(\epsilon), \rho(\epsilon - d) - \rho(\epsilon)$, respectively. Moreover, the processes $\mathcal{P}_1(t), t > 0$ and $\mathcal{P}_2(-t), t < 0$ are independent.

The proof of this theorem will be a consequence of the following several preliminaries. We need to study the process $D_n$ as a process in the standardized parameters. Recall $\theta = (\theta_1', r)$, $d_0 = b_0 - a_0$, $d_1 = b_1 - a_1$. Consider the standardization $\hat{\theta} = (\hat{\theta}_1', \hat{s}) = (\theta_1' + n^{-1/2}w_1, r + n^{-1}t)'$, $w_1 \in \mathbb{R}^4$, $t \in \mathbb{R}$. It is sometimes convenient to write $w_1' = (u_1, u_2, v_1, v_2)$, $u' = (u_1, u_2) \in \mathbb{R}^2$, $v' = (v_1, v_2) \in \mathbb{R}^2$, $w = (w_1', t)'$. Also let $Z_i = (1, X_i)'$, $Z = (1, X)'$, and $a(X_i) = (-Z_i, Z_i)'$, $1 \leq i \leq n$.

Now, write $D_{n,2}(w_1, t)$ for $D_{n,2}(\theta_1 + n^{-1/2}w_1, r + n^{-1}t)$, for convenience. Then, from (3.12) and (3.13) we readily obtain

$$D_n(\theta_1 + n^{-1/2}w_1, r + n^{-1}t) = D_{n,1}(w_1) + D_{n,2}(w_1, t),$$

where we rewrite

$$D_{n,2}(w_1, t) = \sum_{i=1}^{n} [\rho(\epsilon_i + (d_0 + d_1X_i) - n^{-1/2}(u_1 + u_2X_i))$$

$$- \rho(\epsilon_i - n^{-1/2}(v_1 + v_2X_i))]I(r < X_i \leq r + n^{-1}t), \quad t \geq 0,$$
\[ \sum_{i=1}^{n} \left[ \rho(e_i - (d_0 + d_1 X_i) - n^{-1/2}(v_1 + v_2 X_i)) - \rho(e_i - n^{-1/2}(u_1 + u_2 X_i))]I(r + n^{-1}t < X_i \leq r), \quad t \leq 0. \]

Lemma 3.4 gives an approximation of \( D_{n1}(w_1) \) by a quadratic form in \( w_1 \).

We shall next obtain an approximation for \( D_{n2} \). The details below are given only for \( t \geq 0 \), they being similar for \( t \leq 0 \). The analysis here is relatively intricate because it involves the discontinuity point. It is facilitated by the following preliminaries.

Let \( f_n(X, \varepsilon) \) be a sequence of \( \mathbb{R}^p \)-valued random vectors, and \( h_n(X, \varepsilon) \) be a sequence of real r.v.’s. Put

\[ \mathcal{U}_n := n^{-1/2} \sum_{i=1}^{n} f_n(X_i, \varepsilon_i), \quad \mathcal{T}_n(t) := \sum_{i=1}^{n} h_n(X_i, \varepsilon_i) I(r < X_i \leq r + n^{-1}t), \quad t \geq 0. \]

Let

\[ \chi_n(x, z) := E(\exp\{in^{-1/2}z'f_n(X, \varepsilon)\} \mid X = x) = E \exp\{in^{-1/2}z'f_n(x, \varepsilon)\}, \]
\[ \chi_n(x, v) := E(\exp\{iv h_n(X, \varepsilon)\} \mid X = x) = E \exp\{iv h_n(x, \varepsilon)\}, \quad z \in \mathbb{R}^p, \ x, v \in \mathbb{R}. \]

We have

**Theorem 4.2.** Suppose that \( \{(X_i, \varepsilon_i); \ 1 \leq i \leq n\} \) are i.i.d, with the d.f. \( G \) of \( X \) satisfying (a.3), and that \( \{X_i\} \) are independent of \( \{\varepsilon_i\} \). In addition, suppose the sequences \( f_n, h_n \) satisfy the following:

For every \( t \geq 0 \), and for every \( v \in \mathbb{R} \),
\[ \chi_n(x_n, v) \to \chi(r, v) \quad \text{for any sequence } x_n \in (r, r + t/n], \quad (4.3) \]
where \( \chi(r, v) \) is a characteristic function in \( v \) of some r.v.

\[ \forall z \in \mathbb{R}^p, \quad \text{the sequence } n(1 - \xi_n(\cdot, z)) \text{ is uniformly integrable with respect to } dG(\cdot), \quad \text{and } \forall x \in \mathbb{R}, \ \forall z \in \mathbb{R}^p, \ n(1 - \xi_n(x, z)) \to z' \Lambda(x)z/2, \quad (4.4) \]

where \( \Lambda(x) \) is \( p \times p \) covariance matrix for each \( x \).

Then
\[ (\mathcal{U}_n, \mathcal{T}_n) \Rightarrow (\mathcal{X}, \mathcal{P}) \quad \text{in } \mathbb{R}^p \times D[0, \infty), \quad (4.5) \]

where \( \mathcal{X} \sim \mathcal{N}_p(0, \Lambda) \) is a \( p \)-dimensional mean zero normal random vector with covariance matrix \( \Lambda = EA(X) \), and \( \mathcal{P} \) is a compound Poisson process on \([0, \infty)\), independent of \( \mathcal{X} \). Moreover, the rate of \( \mathcal{P} \) is \( g(r) \), \( \mathcal{P}(0) = 0 \), and its jumps have the same distribution as that of the r.v. with characteristic function \( \chi(r, \cdot) \).

**Proof.** The tightness of the probability measures on \( \mathbb{R}^p \times D[0, \infty) \) corresponding to \((\mathcal{U}_n, \mathcal{T}_n)\) follows from the convergence of finite dimensional distributions and the tightness of the probability measures on \( D[0, \infty) \) corresponding to \( \mathcal{T}_n \). To prove this latter
result we use a result of Whitt (1980) and an argument similar to the one used in the proof of Lemma 3.2 in Ibragimov and Has’minskii (1981, p. 261).

Accordingly, for any function \( \gamma \in D[0, \infty) \), a positive integer \( k \), and a \( \delta > 0 \), let

\[
\omega_\delta^k(\gamma) := \sup_{k \leq u - \delta \leq u' \leq u \leq u' + \delta \leq k + 1} \left[ \min\{|\gamma(u') - \gamma(u)|, |\gamma(u'') - \gamma(u)|\} \right]
\]

\[
+ \sup_{k \leq u \leq k + \delta} |\gamma(u) - \gamma(k)| + \sup_{k + 1 - \delta \leq u \leq k + \delta} |\gamma(u) - \gamma(k + 1)|.
\]

From a result in Whitt (1980), the process \( \mathcal{T}_n \) is tight in \( D[0, \infty) \) if for every positive integer \( k \),

\[
\lim_{\delta \to 0} \sup_n P(\omega_\delta^k(\mathcal{T}_n) > \varepsilon) = 0 \quad \forall \varepsilon > 0.
\] (4.6)

We proceed to verify (4.6). For a \( \delta > 0 \), let \( A_i = A_i(u, u + \delta] \) be the event that a sample path of \( \mathcal{T}_n \) has at least \( i \) discontinuities on the interval \((u, u + \delta]\), \( u \geq 0 \), \( 0 \leq i \leq n \). Then, by (a.3), \( g \) is bounded in a neighborhood of \( r \), and hence

\[
P(A_1) \leq \sum_{i=1}^n P \left( \frac{u}{n} < X_i - r \leq \frac{u + \delta}{n} \right) \leq C\delta,
\]

\[
P(A_2) \leq \sum_{i \neq j=1}^n P \left( \frac{u}{n} < X_i - r, X_j - r \leq \frac{u + \delta}{n} \right) \leq C\delta^2.
\]

Now, fix a positive integer \( k \) and let \( B \) be the event that on the interval \([k, k + 1]\), there are at least two points of discontinuities of \( \mathcal{T}_n \) such that the distance between them is less than \( 2\delta \). Divide the interval \([k, k + 1]\) into \( m := 1/|\delta| \) subintervals \( I_i \), \( i = 1, \ldots, m \), of the length \( 1/m \). Each interval with length less than \( 2\delta \) is totally contained in \( I_i \cup (I_{i+1} \cup I_{i+2}) \), for some \( i \). Therefore,

\[
B \subset \bigcup_{i=1}^m A_2(I_i) \cup \bigcup_{i=1}^{m-2} A_2(I_{i+1} \cup I_{i+2}), \quad P(B) \leq Cm\delta^2 \leq C\delta.
\]

Furthermore, on \( B^c \), the complement of \( B \), any interval of the form \([u - \delta, u + \delta]\) possesses at most one point of discontinuity of \( \mathcal{T}_n \), i.e., that is \( \mathcal{T}_n \) is continuous on either \([u, u + \delta]\) or \([u - \delta, u]\). For example, suppose \( \mathcal{T}_n \) is continuous on \([u, u + \delta]\). Then \( \mathcal{T}_n \) has no jump on \([u, u + \delta]\). Note that \( \mathcal{T}_n \) is a step function, so \( \mathcal{T}_n \) is a constant on \([u, u + \delta]\), i.e.

\[
\sup_{u \leq u' \leq u + \delta} |\mathcal{T}_n(u) - \mathcal{T}_n(u')| = 0.
\]

Finally, on \( B^c \), there is at most one discontinuity point of \( \mathcal{T}_n \) and hence, \( \forall \varepsilon > 0 \),

\[
P \left( \left\{ \sup_{k \leq u \leq k + \delta} |\mathcal{T}_n(u) - \mathcal{T}_n(k)| > \frac{\varepsilon}{2} \right\} \cap B^c \right)
\leq P \left( A_1 \left( k \left( \frac{k + \delta}{n} \right) \right) \right) \leq C\delta,
\]
These results imply that for every \( \varepsilon > 0 \),

\[
P(\beta_0^k(\mathcal{F}_n) > \varepsilon) \\
\leq P(B) + P(B^c \cap \{\beta_0^k(\mathcal{F}_n) > \varepsilon\}) \\
\leq C\delta + P\left(\left\{ \sup_{k \leq u \leq k + \delta} |\mathcal{F}_n(u) - \mathcal{F}_n(k)| > \varepsilon/2 \right\} \cap \mathcal{B}^c \right) \\
+ P\left(\left\{ \sup_{k + 1 - \delta \leq u \leq k + 1} |\mathcal{F}_n(u) - \mathcal{F}_n(k + 1)| > \varepsilon/2 \right\} \cap \mathcal{B}^c \right) \\
\leq C\delta,
\]

thereby verifying (4.6). This proves the tightness of the process \( \{\mathcal{F}_n(z), z \geq 0\} \).

To prove the weak convergence of the finite dimensional distributions, we shall show that the joint characteristic function of \( \mathcal{U}_n \) and of the finite dimensional increments of the process \( \mathcal{F}_n \) converges to that of \( \mathcal{Z} \) and the process \( \mathcal{P} \) of (4.5). Accordingly, let \( z \in \mathbb{R}^p \), and, for a positive integer \( m \), let \( t_0 = 0 < t_1 < t_2 < \cdots < t_m < \infty \) be \( m \) positive numbers, \( (v_1, \ldots, v_m) \in \mathbb{R}^m \). Put

\[
\mathcal{L}_n := z' \mathcal{U}_n + \sum_{k=1}^{m} v_k [\mathcal{F}_n(t_k) - \mathcal{F}_n(t_{k-1})].
\]

It suffices to show

\[
\lim_{n \to \infty} E e^{i \mathcal{L}_n} = \exp \left\{ -\frac{1}{2} z' A z - g(r) \sum_{k=1}^{m} (t_k - t_{k-1})(1 - \chi(r, v_k)) \right\}. \tag{4.7}
\]

Clearly, \( \mathcal{L}_n = \sum_{i=1}^{n} \ell_{ni} \), where \( \{\ell_{ni}; 1 \leq i \leq n\} \) are i.i.d. r.v.’s having the same distribution as the r.v.

\[
\ell_n = n^{-1/2} z' f_n(X, \varepsilon) + h_n(X, \varepsilon) \sum_{k=1}^{m} v_k I(r + t_{k-1}/n < X \leq r + t_k/n).
\]

Therefore \( E e^{i \mathcal{L}_n} = (E e^{i \mathcal{L}_n})^n \), where

\[
E e^{i \mathcal{L}_n} = \sum_{k=1}^{m} \int_{r+t_{k-1}/n}^{r+t_k/n} \mathcal{F}_n(x, v_k) g(x) \, dx + \int_{r+t_{k-1}/n}^{r+t_k/n} \mathcal{F}_n(x, v_k) g(x) \, dx + \delta_n
\]

with

\[
\delta_n := \sum_{k=1}^{m} \int_{r+t_{k-1}/n}^{r+t_k/n} E [e^{i u_h(x, \varepsilon)} (e^{i u_{-1/2} z' f_n(x, \varepsilon)}) - 1)] g(x) \, dx.
\]
We shall show that
\[ |\delta_n| = o(n^{-1}). \] (4.8)

Indeed,
\[ |E[e^{i\xi_h(x,z)}(e^{in^{-1/2}Z_f(x,v)} - 1)]| \leq |E[e^{in^{-1/2}Z_f(x,v)} - 1]|^{1/2} \]
\[ = n^{-1/2}[n(1 - \xi_n(x,z)) + n(1 - \xi_n(x,-z))]^{1/2}. \]

Therefore, by the Cauchy–Schwarz inequality,
\[ |\delta_n| \leq 2n^{-1/2} \int_r^{r+tm/n} |n(1 - \xi_n(x,z))|^{1/2} g(x) \, dx \]
\[ \leq 2n^{-1/2} \left( \int_r^{r+tm/n} |n(1 - \xi_n(x,z))| g(x) \, dx \right)^{1/2}. \]

But, the last integral tends to 0, as \( n \to \infty \), by the uniform integrability of the integrand in assumption (4.4). This proves (4.8).

Next, rewrite
\[ Ee^{i\xi_h} = 1 - n^{-1}(x_{1n} + x_{2n}) + \delta_n, \]
where
\[ x_{1n} := \int_{x \in (r)(r+tm/n]} n(1 - \xi_n(x,z))g(x) \, dx, \]
\[ x_{2n} := n \sum_{k=1}^m \int_{r+tk/n}^{r+tk/n} (1 - \chi_n(x,v_k))g(x) \, dx. \]

With (4.8) in mind, (4.9) will follow if we show
\[ \lim_{n \to \infty} x_{1n} = \frac{1}{2}z'Az, \] (4.9)
\[ \lim_{n \to \infty} x_{2n} = g(r) \sum_{k=1}^m (t_k - t_{k-1})(1 - \chi(r,v_k)). \] (4.10)

But, because \( n(1 - \xi_n(x,z)) + (1/2)z'Az(x,z)|x(\xi(r, r + tm/n)) - 0 \), for almost all \( x(G) \), (4.9) follows from assumption (4.4), and the uniform integrability condition which allows to pass to the limit under the integral sign. The claim (4.10) follows from the fact that \( (1 - \chi_n(r + y/n, v_k))g(r + y/n) \to (1 - \chi(r,v_k))g(r) \), boundedly on each interval \((t_{k-1}, t_k], k = 1, \ldots, m\), implied by assumption (4.3) and the continuity of \( g \) at \( r \). This completes the proof of the theorem.

Further, define
\[ \mathcal{S}_n(t) = \sum_{i=1}^n Z_i \psi(e_i)I(r < X_i \leq r + t/n), \quad t \geq 0. \]
Upon taking \( h_n(X, \epsilon) \equiv a'Z\psi(\epsilon) \), for an \( a \in \mathbb{R}^2 \), in the above theorem, we readily obtain

**Corollary 4.1.** Suppose that (2.1), (2.2), (a.2) with \( c_2 = 0 \), and (a.3) hold. Then, for any \( a \in \mathbb{R}^2 \),

\[
a' \mathcal{S}_n \Rightarrow \mathcal{P}^a \text{ in } D[0, \infty),
\]

where \( \mathcal{P}^a \) is a compound Poisson process on \([0, \infty)\) with the rate \( g(r) \), \( \mathcal{P}^a(0) = 0 \), and the same distribution of the jumps as that of the r.v. \( a'Z\psi(\epsilon) \).

We shall now turn to analyzing the behavior of \( D_{n2} \). Recall \( u' = (u_1, u_2) \), \( v' = (v_1, v_2) \), \( w'_1 = (u', v') \), \( w' = (w'_1, t) \). We are now ready to state and prove

**Lemma 4.1.** Suppose that (2.1), (2.2), (a.1), (a.2) with \( c_2 = 0 \), (a.3) and (a.4) hold.

Then

\[
\sup_{\|w\| \leq b} |D_{n2}(w_1, t) - D_{n2}(0, t)| = o_P(1).
\]

(4.11)

**Proof.** Details will be given only for \( t \geq 0 \), they being similar for \( t \leq 0 \). For a given \( 0 < b < \infty \), let \( \tau_z = z(1 + \|r\| + n^{-1}b) \), \( z > 0 \). Then for any \( u \in \mathbb{R}^2 \), \( \|u\| \leq b \),

\[
|u'Z_iI(r < X_i \leq r + tn^{-1})| \leq \tau_b,
\]

\[
|\theta_i'a(X_i)I(r < X_i \leq r + tn^{-1})| \leq c := \tau_2\theta_i, \quad \forall 1 \leq i \leq n.
\]

Hence, (a.1) implies

\[
\sup_{\|w_1\| \leq b, 0 \leq t \leq b} \left| D_{n2}(w_1, t) - D_{n2}(0, t) + (u - v)'n^{-1/2} \times \sum_{i=1}^{n} Z_i\psi(\epsilon_i)I(r < X_i \leq r + n^{-1}t) \right|
\]

\[
\leq \sum_{i=1}^{n} \int_{0}^{n^{-1/2}t_b} |\psi(\epsilon_i + \theta_i'(a(X_i) + y) - \psi(\epsilon_i))I(r < X_i \leq r + bn^{-1})| \, dy
\]

\[
+ \sum_{i=1}^{n} \int_{0}^{n^{-1/2}t_b} |\psi(\epsilon_i + y) - \psi(\epsilon_i))I(r < X_i \leq r + bn^{-1})| \, dy.
\]

(4.12)

The function \( \psi \) being non-decreasing and the independence of \( X \) and \( \epsilon \) imply that the expected value of the first term in the above bound is bounded above by

\[
n^{1/2} \int_{0}^{n^{-1/2}t_b} \{E|\psi(\epsilon + c + y) - \psi(\epsilon)| + E|\psi(\epsilon - c + y) - \psi(\epsilon)|\} \, dy
\]

\[
\times n^{1/2}[G(r + bn^{-1}) - G(r)].
\]

By (a.2) with \( c_2 = 0 \), and by (a.4), the first factor of this bound is \( O(1) \), while by (a.3), the second factor is \( O(n^{-1/2}) = o(1) \).
Similarly, the expected value of the second term in the upper bound (4.12) is bounded above by

\[ n^{1/2} \int_0^{n^{-1/2} s_n} E|\psi(\varepsilon + y) - \psi(\varepsilon)| \, dy \times n^{1/2} [G(r + bn^{-1}) - G(r)] = o(1). \]

Hence, the claim (4.11) follows from the fact \( \sup_{0 \leq t \leq b} \| S_n(t) \| = O_p(1), \forall b < \infty \), implied by Corollary 4.1, and the observation that

\[ n^{-1/2} \sum_{i=1}^n (u - v)^t Z_i \psi(\varepsilon_i) I(r < X_i \leq r + t n^{-1}) \equiv n^{-1/2} (u - v)^t S_n(t). \]

This completes the proof of the Lemma 4.1. \( \square \)

Upon combining Lemmas 3.4 and 4.1, we obtain the following.

**Theorem 4.3.** Suppose (2.1), (2.2), and (a.1)--(a.5) hold. Then, for every \( 0 < b < \infty \),

\[ M_n(\theta_1 + n^{-1/2} w_1, r + n^{-1} t) \]

\[ = M_n(\theta_1, r) - \hat{w}_1 n^{-1/2} \sum_{i=1}^n \hat{m}_i(X_i) \psi(\varepsilon_i) + \frac{\hat{j}(0)}{2} \hat{w}_1^t \Gamma_1 w_1 + D_n(0, t) + u_p(1), \]

(4.13)

where and \( u_p(1) \) is a sequence of stochastic processes converging to zero uniformly over the set \( \| w_1 \| \leq b, \| t \| \leq b \), in probability.

In view of (3.7) and (4.13), we readily obtain

\[ M_n(\hat{\theta}_{1n}, \hat{r}_{1n}) = Q_n(w_1n) + D_n(0, t_n) + o_p(1), \]

\[ Q_n(w_1) := M_n(\theta_1, r) - \hat{w}_1 n^{-1/2} \sum_{i=1}^n \hat{m}_i(X_i) \psi(\varepsilon_i) + \frac{\hat{j}(0)}{2} \hat{w}_1^t \Gamma_1 w_1, \quad w_1 \in \mathbb{R}^4. \]

(4.14)

Consequently, asymptotically the standardized minimizers \( w_1n \) and \( t_n \) behave in a singular fashion in the sense that a minimizer \( w_1n \) of \( M_n(\theta + n^{-1/2} w_1, r + n^{-1} t_n) \) with respect to \( w_1 \) is asymptotically equivalent to a minimizer of \( Q_n(w_1) \) with respect to \( w_1 \) and does not depend on \( t_n \). This in turn implies (4.1). Similarly, a minimizer \( t_n \) of \( M_n(\theta + n^{-1/2} w_1n, r + n^{-1} t) \) with respect to \( t \) is asymptotically equivalent to a minimizer of \( D_n(0, t) \) with respect to \( t \) and does not depend on \( w_1n \).

In order to obtain the joint weak limit of \( (w_1n, t_n) \), we need to obtain the joint weak limit of \( (Z_n, D_n) \), where

\[ D_n(0, t) \equiv D_{n2}(0, t) \]

\[ = \begin{cases} 
\sum_{i=1}^n \{ \rho(\varepsilon_i + d_0 + d_1 X_i) - \rho(\varepsilon_i) \} I(r < X_i \leq r + t/n), & t \geq 0, \\
\sum_{i=1}^n \{ \rho(\varepsilon_i - d_0 - d_1 X_i) - \rho(\varepsilon_i) \} I(r + t/n < X_i \leq r), & t \leq 0.
\end{cases} \]
Recall $d = d_0 + d_1 r$. Let $\sigma^2_\psi := E\psi^2(\varepsilon)$. We have

**Corollary 4.2.** Under (2.1), (2.2), (a.1) and (a.3),
$$
(\mathcal{L}_n, \{D_n(t), t \geq 0\}) \Rightarrow (\mathcal{L}, \mathcal{P}_1) \quad \text{in} \quad \mathbb{R}^4 \times D[0, \infty),
$$
where $\mathcal{L} \sim N_d(0, \sigma^2_\psi \Gamma_\varepsilon)$, $\mathcal{P}_1$ is a compound Poisson process on $[0, \infty)$, independent of $\mathcal{L}$, with the rate $g(r)$, $\mathcal{P}_1(0) = 0$, and the distribution whose jumps is the same as that of $h(r, \varepsilon) = \rho(\varepsilon + d) - \rho(\varepsilon)$.

**Remark 4.1.** From this corollary and the form of the covariance matrix $\Gamma_\varepsilon$, it readily follows that the first two components of $n^{1/2}(\hat{\theta}_1 - \theta_1)$ are asymptotically independent of the latter two components.

**Proof of Corollary 4.2.** In view of Theorem 4.2, it suffices to verify assumptions (4.3) and (4.4) with $f_n(X, \varepsilon) \equiv \hat{m}_r(X) \psi(\varepsilon)$, and $h_n(X, \varepsilon) \equiv \rho(\varepsilon + d_0 + d_1 X) - \rho(\varepsilon)$.

Here, assumption (4.3) is immediate by the continuity of $\rho$. Next, for a $z \in \mathbb{R}^4$,
$$
|n(1 - \xi_n(x, z))| = n|E[1 - \exp\{in^{-1/2}z'\hat{m}_r(x)\psi(\varepsilon)\}]|.
$$
As $E\psi(\varepsilon) = \lambda(0) = 0$ and $|1 + ix - e^{ix}| \leq x^2/2$, for all $x \in \mathbb{R}$, we obtain
$$
|n(1 - \xi_n(x, z))| \leq \frac{\sigma^2_\psi}{2} (z'\hat{m}_r(x))^2 \leq \frac{\sigma^2_\psi \|z\|^2}{2} (1 + |x|^2) \quad \forall n \geq 1.
$$
Hence, for each $z \in \mathbb{R}^4$, the sequence $\{n(1 - \xi_n(x, z)), n=1, 2, \ldots\}$ is uniformly integrable with respect to $dG(x)$, by (a.3). Furthermore, for any $x \in \mathbb{R}$, $z \in \mathbb{R}^4$, $n(1 - \xi_n(x, z)) \to \sigma^2_\psi z'\hat{m}_r(x)\hat{m}_r(x)'z$. Hence (4.4) is satisfied in the present case with $\Lambda(x) = \hat{m}_r(x)\hat{m}_r(x)'$ and $\Lambda = \Gamma_\varepsilon \equiv E[\hat{m}_r(X)\hat{m}_r(X)']$, thereby completing the proof of the corollary. \hfill \Box

Carrying out a similar argument as above in the case $t \leq 0$, we obtain that there is another compound Poisson process $\mathcal{P}_2$ on $[0, \infty)$, independent of $\mathcal{L}$ of Corollary 4.2, with the rate $g(r)$, $\mathcal{P}_2(0) = 0$, and whose jumps have the same distribution as that of $\rho(\varepsilon - d) - \rho(\varepsilon)$. Moreover, the processes $\mathcal{P}_1(t)$ for $t > 0$ and $\mathcal{P}_2(-t)$ for $t < 0$ are independent because they involve independent sets of random variables. Let $\Pi$ be as in (4.2). We readily obtain the following corollary.

**Corollary 4.3.** Under the assumptions of Corollary 4.2,
$$
(\mathcal{L}_n, D_n) \Rightarrow (\mathcal{L}, \Pi) \quad \text{in} \quad \mathbb{R}^4 \times D(-\infty, \infty),
$$
where $\mathcal{L}$ is as in Corollary 4.2, and independent of $\Pi$.

Note that so far we have not used assumption (a.6). It is used in the next lemma, which is the final step in the proof of Theorem 4.1.

**Lemma 4.2.** Under the assumptions of Theorem 4.1, $n(\hat{r}_n - r)$ converges weakly to the smallest minimizer $\pi_-$ of the process $\Pi$. Moreover, $n(\hat{r}_n - r)$ is asymptotically independent of $n^{1/2}(\hat{\theta}_1 - \theta_1)$. 

**Proof of Lemma 4.2.** The claim about the asymptotic independence follows from the previous corollary.

Next, let $D[-b,b]$ be the Skorokhod space of functions defined on $[-b,b]$, $0 < b < \infty$. Recall that $D \equiv D(-\infty,\infty)$ consists of all functions $l: \mathbb{R} \to \mathbb{R}$ whose restriction $I_b$ on the interval $[-b,b]$ belongs to $D[-b,b]$ for any $b > 0$, where $I_b(t) = I(|t| \leq b)$ is the indicator function. Thus, we need to restrict the discussion to a compact neighborhood $[-b,b]$. Let $r_n^b, \pi_n^b$ be the smallest minimizers of $M_n(\vartheta_1n(s),s)$, $\Pi(s)$, respectively, on the interval $s \in [-b,b]$.

By Theorem 3.1, for any $\gamma > 0$ one can find $b < \infty$ and $n_0$ such that $P(\hat{r}_n = \hat{r}_n^b) > 1 - \gamma$, $\forall n > n_0$. Also, by (a.1), $\Pi(\pm 1) = g(r)E[\rho(\varepsilon \pm d) - \rho(\varepsilon)] > 0$. In view of Theorem 4.3.9 of Gikhman and Skorokhod (1975), this implies that $\Pi(t) \to \infty$ as $|t| \to \infty$, so that for each $\gamma > 0$, there exists $b < \infty$ such that $P(\pi_\gamma = \pi_\gamma^b) > 1 - \gamma$. As a consequence, it suffices to show that for every $0 < b < \infty$,

$$
n(\hat{r}_n^b - r) \Rightarrow \pi_n^b. \tag{4.15}
$$

Now, fix a $0 < b < \infty$. Let

$$
\hat{M}_n^b(t) = [M_n(\vartheta_1n(r + n^{-1}t),r + n^{-1}t) - M_n(\vartheta_1n(r),r)]I_b(t), \quad \Pi^b(t) = \Pi(t)I_b(t).
$$

Theorem 4.3 and Lemma 3.4 imply that

$$
\sup_{|t| \leq b} \left| \vartheta_1n(r + n^{-1}t) - \vartheta_1n(r) \right| = o_P(n^{-1/2}).
$$

This fact and Theorem 4.3 in turn imply that

$$
\sup_{|t| \leq b} \left| \hat{M}_n^b(t) - D_n2(t) \right| = o_P(1). \tag{4.16}
$$

Introduce also

$$
H_n(t) := \begin{cases} 
\sum_{i=1}^{n} \{ \rho(\varepsilon_i + d) - \rho(\varepsilon_i) \}I(r < X_i \leq r + t/n), & t \in [0,b], \\
\sum_{i=1}^{n} \{ \rho(\varepsilon_i - d) - \rho(\varepsilon_i) \}I(r + t/n < X_i \leq r), & t \in [-b,0].
\end{cases}
$$

Then, by (a.4),

$$
E \sup_{|t| \leq b} |D_n2(t) - H_n(t)| \leq n \int_{|x-\varepsilon| \leq b/n} g(x)E|\rho(\varepsilon + d_0 + d_1x) - \rho(\varepsilon + d_0 + d_1x)| dx = o(1).
$$

This fact together with (4.16) in turn yields

$$
\sup_{|t| \leq b} \left| \hat{M}_n^b(t) - H_n(t) \right| = o_P(1). \tag{4.17}
$$
Let \( t^n_b = n(r^n_b - r) \) and \( \pi^n_b \) be the smallest minimizers of \( \hat{M}^b_n \) and \( H_n \), respectively. The claim (4.15) follows from
\[
\begin{align*}
t^n_b - \pi^n_b &\to 0, \\
\pi^n_b &\to \pi^n_-.
\end{align*}
\]
(4.18)
(4.19)

The facts (4.18) and (4.19) are proved below, in Lemmas 4.3, 4.4, respectively. This completes the proof of Theorem 4.1.

**Lemma 4.3.** Under the conditions of Theorem 4.1, (4.18) holds.

This lemma is proved at the end of this section. We use this to prove (4.19). For clarity of the exposition, we shall restrict the following discussion to the interval \([0, b]\) only. The details for the case \([-b, 0]\) are similar. Let \( U_i := \frac{\rho(e_i + d) - \rho(e_i)}{b}, 1 \leq i \leq n \), \( U \) be a copy of \( U_i \),
\[
H_n(t) := \sum_{i=1}^{n} U_i I(r < e_i \leq r + t/n), \quad t \in [0, b],
\]
and let \( \Pi^b(t), t \in [0, b] \), be the compound Poisson process with jump rate \( g(r) \) and jump distribution \( U \):
\[
E e^{ia\Pi^b(t)} = \exp \{tg(r)E(e^{iaU}) - 1)\}, \quad a \in \mathbb{R}.
\]

In the rest of this proof, \( \pi^n_b, \pi^n_- \) will denote the minimizers:
\[
\begin{align*}
\pi^n_b &:= \inf \left\{ 0 \leq t \leq b: H_n(t) = \min_{0 \leq s \leq b} H_n(s) \right\}, \\
\pi^n_- &:= \inf \left\{ 0 \leq t \leq b: \Pi^b(t) = \min_{0 \leq s \leq b} \Pi^b(s) \right\}.
\end{align*}
\]

Write \( 0 = \tau_{n_0} < \tau_{n_1} < \cdots < \tau_{n_k} \leq b \), \( 0 = \tau_0 < \tau_1 < \cdots < \tau_k \leq b \), for consecutive jump points of \( H_n \) and \( \Pi^b \), respectively. By definition, \( \pi^n_b \in \{\tau_{ni}: 0 \leq i \leq k\} \), \( \pi^n_- \in \{\tau_i: 0 \leq i \leq k\} \). Introduce the functionals
\[
\begin{align*}
A_n &= \inf_{i=1, \ldots, k: \tau_{ni} \neq \pi^n_b}(H_n(\tau_{ni}) - H_n(\pi^n_b)), \\
\Delta &= \inf_{i=1, \ldots, k: \tau_i \neq \pi^n_-}(\Pi^b(\tau_i) - \Pi^b(\pi^n_-)).
\end{align*}
\]

By assumption (a.6), \( U \neq 0 \), a.s., and hence, the random variable \( \Delta > 0 \), a.s.

**Lemma 4.4.** Under the assumptions of Theorem 4.1, for every \( 0 < b < \infty \), (4.19) holds and \( A_n \to \Delta \).

**Proof.** We use a coupling of the processes \( H_n, \Pi^b \). We shall show below that one can construct random processes \( \tilde{H}_n, \tilde{\Pi}_n \) indexed by \( t \in [0, b] \), defined on a common probability space such that
\[
\tilde{H}_n \overset{\text{law}}{=} H_n, \quad \tilde{\Pi}_n \overset{\text{law}}{=} \Pi^b
\]
\[
P(\tilde{H}_n = \tilde{\Pi}_n) = 1 - o(1).
\]
(4.20)
(4.21)

Clearly, these two facts imply the statement of the lemma.
We shall first couple Poisson processes on \([0, b]\) with expectations \(\mu_{1t} := g(r)/n\) and \(\mu_{2t} := -\log(1 - P(r < X \leq r + t/n))\). Note \(\mu_{1t} = \mu_{2t} = (1 - P(r < X \leq r + t/n))^{-1}g(r + t/n)/n\) are positive on \([0, b]\). Let \(v'_t := \mu_{1t} - v_t\), \(v''_t := \mu_{2t} - v_t\), where

\[
v_t := \int_0^t (\hat{\mu}_1(s) + \hat{\mu}_2(s)) \, ds = \int_0^t [g(r) \wedge g(r, u)] \, du,
\]

where \(g(r, u) := g(r + u)/(1 - P(r < X \leq r + u))\). Note \(v_t, v'_t, v''_t\) are non-decreasing.

Let \(q(t), q'(t), q''(t)\), be three independent Poisson processes with expectations \(v_t, v'_t, v''_t\), respectively. Put

\[
q_1(t) := q(t) + q'(t), \quad q_2(t) := q(t) + q''(t).
\]

Clearly, \(q_i\) is a Poisson process with \(Eq_i(t) = \mu_{it}\), \(i = 1, 2\). Furthermore, since \(q'\) is non-decreasing, \(q'(b) = 0\) implies \(q'(t) = 0, \forall t \in [0, b]\). A similar fact holds for \(q''\).

Thus,

\[
P \left( \sup_{0 \leq t \leq b} |q_1(t) - q_2(t)| > 0 \right)
\leq P(q'(b) > 0) + P(q''(b) > 0) = 2 - e^{-v'_b} - e^{-v''_b},
\]

where \(v'_b = \int_0^b (g(r) - g(r) \wedge g(r, u)) \, du\), and \(v''_b = \int_0^b (g(r, u) - g(r) \wedge g(r, u)) \, du\). By (a.3), the integrands of \(v'_b\) and \(v''_b\) are continuous at \(u = 0\). This implies \(v'_b = o(1/n) = v''_b\), and we obtain

\[
P \left( \sup_{0 \leq t \leq b} |q_1(t) - q_2(t)| > 0 \right) = o(1/n). \tag{4.22}
\]

To show (4.20), let \(q_{1j}, j = 1, \ldots, n\), be independent copies of the Poisson process \(q_1\) with rate \(\mu_{1j} = g(r)/n\). Let \(0 =: \tau_{0j} < \tau_{1j} < \cdots\), be the times of consecutive jumps of \(q_{1j}\). Next, let \(U_{ij}, i = 1, 2, \ldots\), \(j = 1, 2, \ldots, n\), be i.i.d. copies of \(U\), independent also of \(q_{1j}, j = 1, \ldots, n\). Put \(W_{0j} := 0\), \(W_{ij} := U_{1j} + \cdots + U_{ij}\). Put

\[
z_j(t) := \sum_{i=0}^{\infty} W_{ij} I(\tau_{ij} \leq t < \tau_{i+1, j}).
\]

Then \(z_j\) is a compound Poisson process with rate \(g(r)/n\) and the jump distribution \(U\). As the processes \(z_1, \ldots, z_n\) are independent, so

\[
\hat{H}_n(t) = \sum_{j=1}^n z_j(t), \quad t \in [0, b],
\]

coincides in the distribution with the compound Poisson process \(H^b\), implying the first equality of (4.20).

Next we construct \(\hat{H}_n\). Let \(q_{2j}, j = 1, \ldots, n\), be independent copies of the Poisson process \(q_2\). Put

\[
y_j(t) := q_{2j}(t) \wedge 1, \quad t \in [0, b].
\]

Note \(y_j(t), t \in [0, b]\) are independent and have the same distribution as the process \(I(r < X \leq r + t/n), t \in [0, b]\). Indeed, the random function \(y_j(t), t \in [0, b]\) is
non-decreasing, takes values 0, 1 only, and
\[ P(y_j(t) = 0) = P(q_2^j(t) = 0) = e^{-\mu_2 t} \]
\[ = 1 - P(r < X \leq r + t/n) \]
\[ = P(I(r < X \leq r + t/n) = 0), \quad j = 1, \ldots, n. \]

Therefore
\[ \tilde{H}_n(t) := \sum_{j=1}^{n} U_j y_j(t), \quad t \in [0, b] \]
satisfies (4.20).

It remains to show (4.21). By the above construction,
\[ P \left( \sup_{0 \leq t \leq b} |\tilde{H}_n(t) - \tilde{V}_n(t)| > 0 \right) \leq n P \left( \sup_{0 \leq t \leq b} |U_{1} y_1(t) - z_1(t)| > 0 \right). \]
Note that last supremum is zero unless \(\sup_{0 \leq t \leq b} |q_1^1(t) - q_2^1(t)| > 0\) or \(q_1^1(b) > 1, q_2^1(b) > 1\) hold. Therefore,
\[ P \left( \sup_{0 \leq t \leq b} |\tilde{H}_n(t) - \tilde{V}_n(t)| > 0 \right) \leq n \left( P \left( \sup_{0 \leq t \leq b} |q_1^1(t) - q_2^1(t)| > 0 \right) + P(q_1^1(b) > 1) + P(q_2^1(b) > 1) \right) = o(1), \]
where we used (4.22) and the fact that \(P(q_i^1(b) > 1) = O(n^{-2}), \ i = 1, 2\). This proves (4.21) and Lemma 4.4.

**Proof of Lemma 4.3.** Suppose (4.18) is false. Then, there exist a \(\delta > 0\) and an \(n_0\) such that for all \(n > n_0\),
\[ P(|t_n^b - \pi_n^b| > \delta) > \delta. \quad (4.23) \]
As the sequence \((t_n^b, \pi_n^b), n = 1, 2, \ldots\) is tight, without loss of generality we may assume that it converges in distribution: \((t_n^b, \pi_n^b) \Rightarrow (t^b, \pi^b)\), where \((t^b, \pi^b) \in [0, b]^2\) is a random vector. Note that the sequence \((\hat{M}_n^b(t_n^b), H_n(\pi_n^b)), n \geq 1\), converges in distribution:
\[ (\hat{M}_n^b(t_n^b), H_n(\pi_n^b)) \Rightarrow (\Pi^b(\pi^b_\downarrow), \Pi^b(\pi^b_\uparrow)), \quad (4.24) \]
which follows from the joint convergence \((\hat{M}_n^b, H_n) \Rightarrow (\Pi^b, \Pi^b)\) in \([0, b] \times [0, b]\) and the continuity of the infimum map in \([0, b]\), cf. Billingsley (1968); the latter fact is a consequence of Corollary 4.2, (4.16) and (4.17). Relations (4.24) and (4.17) obviously imply
\[ (H_n(t_n^b), H_n(\pi_n^b)) \Rightarrow (\Pi^b(\pi^b_\downarrow), \Pi^b(\pi^b_\uparrow)), \quad (4.25) \]

Now we use the important fact (mentioned in Section 2) that the M-process \(\hat{M}_n^b\) (similarly as \(H_n\)) is constant on the intervals between consecutive values of \(X_i\); i.e.
on the intervals \([\tau_{n,i−1}, \tau_{ni}], \ i = 1, \ldots, k_n\), between successive jumps of the process \(H_n\). Therefore, the minimizer \(t_n^b \in \{\tau_{ni}: 0 \leq i \leq k_n\}\) a.s. But then

\[
H_n(t_n^b) - H_n(\pi_n^b) \geq \Delta_n I(t_n^b \neq \pi_n^b).
\]

According to (4.24), \(H_n(t_n^b) - H_n(\pi_n^b) \Rightarrow 0\), so that we obtain

\[
\Delta_n I(t_n^b \neq \pi_n^b) \Rightarrow 0. \tag{4.26}
\]

Again, by taking a subsequence, we may assume \((\Delta_n, I(t_n^b \neq \pi_n^b)) \Rightarrow (\Delta, I)\), where \(0 \leq I \leq 1\) is a random variable. From the assumption (4.23) we obtain

\[
P(I > 0) \geq \liminf_{n \to \infty} P(t_n^b \neq \pi_n^b) > \delta.
\]

From Lemma 4.4 and (4.6) we have \(P(\Delta > 0) = 1\). The last two relation clearly contradict (4.26). This completes the proof of Lemma 4.3. \(\square\)

**Remark 4.2.** On MLE and M-estimators. As mentioned earlier, Koul and Qian (2002) (KQ) contain an analog of Theorem 4.1 for the MLE’s. However, the conditions imposed there are much more restrictive than those of the present paper. The first major difference is in the smoothness requirements on the underlying likelihood score, analog of which in the M-estimation method is the function \(\psi\). In the KQ paper the score function was assumed to have smooth first derivative. The other difference is that there \(E|X|^3 < \infty\) was required, compared to now requiring only \(EX^2 < \infty\), cf. (a.3).

For example, Theorem 4.1 is applicable to the case where \(\rho(x) \equiv |x|\), the error distribution is Cauchy, and the design distribution satisfies (a.3), whereas the analogous result in the KQ paper is not applicable to this situation because this procedure is not the MLE at the Cauchy errors. This theorem also covers the case of the MLE at the double exponential errors, i.e., when \(\rho(x) \equiv |x|\) or \(\psi(x) = \text{sgn}(x)\) and the errors are double exponential, because the score \(\psi(x) = \text{sgn}(x)\) is far from being smooth now. A similar remark applies to the Huber score \(\psi(x) \equiv \text{sgn}(x) cI(|x| > c) + xI(|x| \leq c)\), where \(c > 0\) is constant.

Even for the least square estimators, where \(\rho(x) = x^2\), and the Gaussian errors, i.e., even for the MLE at the ideal Gaussian model, Theorem 4.1 is more broadly applicable than its analog in the KQ paper, simply because now we require only finite second moment of the design variable. \(\square\)

5. A simulation study and an application

In this section, we shall report results of a simulation study and an application to an automobile gas mileage-weight data. In both, we shall use \(\rho(x) \equiv |x|\) and \(\rho(x) \equiv x^2\). The corresponding M-estimators are called the least absolute deviation (LAD) and the least square (LS) estimators, respectively.

5.1. A simulation study

The results reported in this section focus on the performance of the LAD and LS estimators for various error densities. The samples were generated from the following
simple model:

\[ Y_i = (0.5 - X_i)I(X_i \leq 0) + (-0.7 + X_i)I(X_i > 0) + \varepsilon_i, \quad i = 1, \ldots, n, \]

where \( \{X_i\} \) is a random sample from the standard normal distribution and \( \{\varepsilon_i\} \) are from various error densities. In other words, we took \( a_0 = 0.5, \quad a_1 = -1, \quad b_0 = -0.7, \quad b_1 = 1, \quad r = 0.0 \) in (2.1). The error densities considered are the double exponential, standard normal, and the student \( t \) with degrees of freedom \( 4 \). The sample sizes used are 100, 200 and 500.

Estimators are computed using the method described in Section 2 above. Our program for computing the LADE (LAD estimator) is based on a specialized linear programming algorithm due to Barrodale and Roberts (1973). The results are computed to single-precision accuracy only.

Table 1 gives the Monte Carlo means (Mean), the standard error (SE) and the mean absolute deviation (MAD) of the LAD and LS estimators, based on 500 repetitions. One observes that under the normal errors, LADE has larger spread (in terms of both SE and MAD) and bias than the LSE, while at the double exponential errors, the LSE has much larger variability and relatively larger bias than the LADE. For example, for \( n = 500 \), the Monte Carlo SE and MAD of the LSE of \( r \) are, respectively, about 1.67 and 2 times larger while the bias is about 3.75 times larger than the LADE of \( r \) at the double exponential errors. We also notice that for other heavy-tail distributions, the efficiency gain by LADE is considerable.

5.2. An application: gasoline mileage data

The data, from Hogg and Ledolter (1992, p. 378), originally reported by Henderson and Velleman (1981), consists of gas mileage and weight of 38 automobiles. The original data set with several other variables was collected by Consumer Reports and by 1974 Motor Trend magazine. It was used by Henderson and Velleman (1981) to investigate the various aspects of automobiles design and performance.

In this section, we only model the relationship between MPG and weights of automobiles. These 38 cars were from the model year 1978–1979. Their weights (in units of 1000 pounds) and fuel efficiencies MPG (miles per gallon) were recorded. Fig. 1 illustrates the scatter plot of MPG against weight which appears a pattern of two-phase linear rather than a simple linear regression. As Henderson and Velleman (1981) and Hogg and Ledolter (1992) pointed out, the linear regression is not an appropriate model. They suggested possible other models such as quadratic regression. In this paper, we suggest the two-phase linear regression model since the scatter plot (Fig. 1) shows the two-phase pattern. For reader’s convenience, we include the data in Table 2.

Thus, we use the two-phase linear regression model (2.1) to fit the data (Weight, MPG). The two estimators used in the two-phase linear regression modeling are LSE and LADE. For the sake of a comparison, we also fit the data by simple linear and quadratic regression models using LSE. Fig. 1 shows the fits for all four estimators for the three types of regression models. Fig. 2 shows the graph of \( M_n(\theta_1, s), s \) as a function of \( s \) using the LAD method.
Table 1
Mean, SE and MAD of LAD and LS estimators for $\theta = (0.5, -1, -0.7, 1, 0)$

<table>
<thead>
<tr>
<th>Error distribution</th>
<th>$N(0,1)$</th>
<th>$D\exp(0,1)$</th>
<th>$t(4)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 100$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}$</td>
<td>0.141</td>
<td>0.036</td>
<td>0.185</td>
</tr>
<tr>
<td>$\hat{\alpha}_0$</td>
<td>0.512</td>
<td>0.037</td>
<td>0.239</td>
</tr>
<tr>
<td>$\hat{\alpha}_1$</td>
<td>-0.990</td>
<td>0.034</td>
<td>0.231</td>
</tr>
<tr>
<td>$\hat{\beta}_0$</td>
<td>-0.569</td>
<td>0.069</td>
<td>-0.593</td>
</tr>
<tr>
<td>$\hat{\beta}_1$</td>
<td>0.925</td>
<td>0.050</td>
<td>0.277</td>
</tr>
<tr>
<td>$\hat{\beta}$</td>
<td>0.102</td>
<td>0.031</td>
<td>0.135</td>
</tr>
<tr>
<td>$\hat{\alpha}_0$</td>
<td>0.511</td>
<td>0.033</td>
<td>0.184</td>
</tr>
<tr>
<td>$\hat{\alpha}_1$</td>
<td>-0.992</td>
<td>0.029</td>
<td>0.190</td>
</tr>
<tr>
<td>$\hat{\beta}_0$</td>
<td>-0.636</td>
<td>0.052</td>
<td>-0.394</td>
</tr>
<tr>
<td>$\hat{\beta}_1$</td>
<td>0.964</td>
<td>0.040</td>
<td>0.231</td>
</tr>
<tr>
<td>$\hat{\beta}$</td>
<td>0.105</td>
<td>0.015</td>
<td>0.065</td>
</tr>
<tr>
<td>$\hat{\alpha}_0$</td>
<td>0.502</td>
<td>0.017</td>
<td>0.154</td>
</tr>
<tr>
<td>$\hat{\alpha}_1$</td>
<td>-1.000</td>
<td>0.016</td>
<td>0.146</td>
</tr>
<tr>
<td>$\hat{\beta}_0$</td>
<td>-0.676</td>
<td>0.025</td>
<td>0.156</td>
</tr>
<tr>
<td>$\hat{\beta}_1$</td>
<td>0.977</td>
<td>0.020</td>
<td>0.164</td>
</tr>
<tr>
<td>$\hat{\beta}$</td>
<td>0.053</td>
<td>0.015</td>
<td>0.065</td>
</tr>
<tr>
<td>$\hat{\alpha}_0$</td>
<td>0.502</td>
<td>0.017</td>
<td>0.154</td>
</tr>
<tr>
<td>$\hat{\alpha}_1$</td>
<td>-1.000</td>
<td>0.016</td>
<td>0.146</td>
</tr>
<tr>
<td>$\hat{\beta}_0$</td>
<td>-0.676</td>
<td>0.025</td>
<td>0.156</td>
</tr>
<tr>
<td>$\hat{\beta}_1$</td>
<td>0.977</td>
<td>0.020</td>
<td>0.164</td>
</tr>
<tr>
<td>$\hat{\beta}$</td>
<td>0.053</td>
<td>0.015</td>
<td>0.065</td>
</tr>
<tr>
<td>$\hat{\alpha}_0$</td>
<td>0.502</td>
<td>0.017</td>
<td>0.154</td>
</tr>
<tr>
<td>$\hat{\alpha}_1$</td>
<td>-1.000</td>
<td>0.016</td>
<td>0.146</td>
</tr>
<tr>
<td>$\hat{\beta}_0$</td>
<td>-0.676</td>
<td>0.025</td>
<td>0.156</td>
</tr>
<tr>
<td>$\hat{\beta}_1$</td>
<td>0.977</td>
<td>0.020</td>
<td>0.164</td>
</tr>
<tr>
<td>$\hat{\beta}$</td>
<td>0.105</td>
<td>0.015</td>
<td>0.065</td>
</tr>
<tr>
<td>$\hat{\alpha}_0$</td>
<td>0.502</td>
<td>0.017</td>
<td>0.154</td>
</tr>
<tr>
<td>$\hat{\alpha}_1$</td>
<td>-1.000</td>
<td>0.016</td>
<td>0.146</td>
</tr>
<tr>
<td>$\hat{\beta}_0$</td>
<td>-0.676</td>
<td>0.025</td>
<td>0.156</td>
</tr>
<tr>
<td>$\hat{\beta}_1$</td>
<td>0.977</td>
<td>0.020</td>
<td>0.164</td>
</tr>
<tr>
<td>$\hat{\beta}$</td>
<td>0.053</td>
<td>0.015</td>
<td>0.065</td>
</tr>
<tr>
<td>$\hat{\alpha}_0$</td>
<td>0.502</td>
<td>0.017</td>
<td>0.154</td>
</tr>
<tr>
<td>$\hat{\alpha}_1$</td>
<td>-1.000</td>
<td>0.016</td>
<td>0.146</td>
</tr>
<tr>
<td>$\hat{\beta}_0$</td>
<td>-0.676</td>
<td>0.025</td>
<td>0.156</td>
</tr>
<tr>
<td>$\hat{\beta}_1$</td>
<td>0.977</td>
<td>0.020</td>
<td>0.164</td>
</tr>
</tbody>
</table>

Table 3 lists the parameter estimators for the two-phase linear regression models for LS and LAD methods. Both methods yield the same value for $\hat{\theta}$ while the estimates of the other parameters are different. It appears there is a change at $\hat{\theta} = 2.7$. Out of the
38 cars, there are 20 having weight at most 2.7 thousand pounds and 18 more than 2.7 thousand pounds.

It is also useful to look at the estimated standard deviation for each piece given by

\[ s_1 = \sqrt{\frac{1}{20} \sum (Y_i - \hat{a}_0 - \hat{a}_1 X_i)^2 I(X_i \leq 2.7)} \]

and

\[ s_2 = \sqrt{\frac{1}{18} \sum (Y_i - \hat{a}_0 - \hat{a}_1 X_i)^2 I(X_i > 2.7)}. \]

Notice that from Table 3, for the LAD and the LS estimators, both \( s_1 \) and \( s_2 \) are smaller using two-phase regression than the single square-root mean square error 2.85 using simple linear regression.
One of the criteria for assessment of the performance of model fitting used in the literature is the mean magnitude of relative error (RE) defined as

\[
\text{MRE} = \frac{1}{n} \sum_{i=1}^{n} \text{RE}_i = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right|
\]

where \( \hat{Y}_i \) is the fitted value of \( Y_i \). The implicit assumption in this summary measure is that the seriousness of the absolute error is proportional to the size of the observations. A companion summary measure related to MRE is the prediction at level \( p \), \( \text{PRED}(p) = \frac{k}{n} \), where \( k \) is the number of observations whose RE is less than or equal to \( p \). These two measures of goodness of fit are recommended by Conte et al. (1986) in software engineering modeling. They also recommended an upper limit of 25% for MRE and a lower limit of 75% for \( \text{PRED}(0.25) \) as the values to be acceptable for model fittings.

We use MRE and \( \text{PRED}(0.25) \) as our model selection criteria. Table 4 reports the two measures MRE and \( \text{PRED}(0.25) \) for gasoline mileage data using all three models: Two-phase, Quadratic and Linear.
Based on this table, one observes that the mean magnitude of relative error using the two-phase regression model is reduced by about $1/7$ to $1/8$ of that using a linear or a quadratic regression model. Within two-phase regression, the two estimators using the LAD and the LS methods are comparable with the LAD estimator being slightly better in terms of the MRE. Thus, based on the given data, we conclude that the two-phase linear regression model with the LAD estimation procedure is the
most desirable among the above models for modeling the MPG against weight of an automobile.

Acknowledgements

Authors thank the two referees for some helpful comments.

References

Hudson, D.J., 1966. Fitting segmented curves whose join points have to be estimated. J. Amer. Statist. Assoc. 61, 1097–1129.