THE INSTITUTE OF STATISTICS

UNIVERSITY OF NORTH CAROLINA

L-MOMENT M-ESTIMATES IN LINEAR AND NON-LINEAR REGRESSION

by C.S. Mathers

July 1954

Minutes Series #2524

DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina
Mimeo Series #2324
C. S. Withers
L-Moment M-Estimates in Linear
and Non Linear Regression
Dept. of Statistics
Univ. of North Carolina
July 1994
L-Moment $M$-estimates in Linear and Non-Linear Regression *

C.S. Withers
Applied Mathematics Group
Institute for Industrial Research and Development
Box 31-310, Lower Hutt
New Zealand †

July 29, 1994

Abstract

We introduce a new class of estimates for linear and nonlinear regression models, that combines the robustness features of both $L$-moments and $M$-estimates. We sketch a proof of their asymptotic normality and so derive confidence regions for the regression parameters.

*Research partially supported by the ONR Grants N00014-91-J-1087 and N00014-93-1-0043.
†Written while visiting Stanford University and University of North Carolina.

1991 AMS Classification: Primary 62F20, Secondary 62F35

Key Words and Phrases: regression, $L$-moments, $M$-estimates, confidence regions.

ONR Grant TR No. 12
1 Introduction and Summary

This note combines the notions of $M$-estimates and $L$-moments to obtain a new class of estimates for regression parameters.

Suppose we observe

$$Y_N = z_N(\varphi) + \epsilon_N \text{ in } \mathbb{R} \text{ for } 1 \leq N \leq n$$  \hspace{1cm} (1.1)

where $\varphi$ in $\mathbb{R}^p$ is an unknown parameter, $z_N(\varphi)$ is a given regression function with finite derivatives, and the residuals \{\epsilon_1, \ldots, \epsilon_n\} are independent and identically distributed (i.i.d.) with some unspecified distribution $F$ on $\mathbb{R}$.

We define the LM-estimate (or L-moment $M$-estimate) as \hat{\varphi} minimising

$$\lambda_n = \lambda_n(\varphi) = n^{-1} \sum_{N=1}^{N} \rho(\epsilon_N(\varphi), F_n(\epsilon_N(\varphi)))$$  \hspace{1cm} (1.2)

where $\epsilon_N(\varphi) = Y_N - z_N(\varphi)$, $F_n$ is the empirical distribution of \{\epsilon_N = \epsilon_N(\varphi), 1 \leq N \leq n\}, and $\rho(\epsilon, y)$ is a given function with sufficiently many finite derivatives

$$\rho_{rs}(\epsilon, y) = (\partial/\partial \epsilon)^r(\partial/\partial y)^s \rho(\epsilon, y).$$

If $\rho(\epsilon, y)$ does not depend on $y$, $\hat{\varphi}$ is called an M-estimate and is known to be asymptotically normal. See, for example, Withers (1994a).

A more useful way of writing (1.2) is

$$\lambda_n = n^{-1} \sum_{N=1}^{N} \rho(\epsilon_{(N)}, N/n)$$  \hspace{1cm} (1.3)

where $\epsilon_{(1)} \leq \ldots \leq \epsilon_{(n)}$ are the ordered values of $\epsilon_1, \ldots, \epsilon_n$. The rth L-moment of $F$ is defined to be

$$\beta_r = E \epsilon_1 F(\epsilon_1)^r = \int \epsilon F(\epsilon)^r dF(\epsilon).$$  \hspace{1cm} (1.4)

Estimates based on $L$-moments are more robust than those based on moments and may even be more efficient than maximum likelihood estimates: see Hosking (1990) and
Hosking, Wallis and Wood (1985). They are a special case of probability weighted moments
\[ E a(\epsilon, F(\epsilon)) = \int a(\epsilon, F(\epsilon))dF(\epsilon). \]

For the asymptotic normality and bias reduction of their estimates, see Withers and Pearson (1994).

Set
\[
\begin{align*}
\rho_{rs} &= \rho_{rs}(F) = \rho_{rs}(\epsilon, F(\epsilon))dF(\epsilon), \\
\rho_{rs,rs} &= \rho_{rs,rs}(F) = \int \rho_{rs}(s, F(\epsilon))^2dF(\epsilon) \\
\text{and } \rho_{10:11} &= \int \int_{x \leq y} \rho_{10}(x, F(x))\rho_{11}(y, F(y))dF(x)dF(y).
\end{align*}
\]

In Section 2 we prove our main result:

**Theorem 1.1** If \( \rho_{10} = 0 \) then
\[
n^{1/2}(\hat{\varphi} - \varphi) \xrightarrow{L} \mathcal{N}(0, V(\varphi)) \quad \text{as } n \to \infty
\]
where
\[
V(\varphi) = \rho_{20}^{-2}(\rho_{10,10}V_z^{-1} + \theta
\nu
\nu'),
\]
\[
\nu = V_z^{-1}z, \\
z = \partial \bar{z}/\partial \varphi, \bar{z} = n^{-1} \sum_{N=1}^{n} z_N(\varphi), \\
V_z = n^{-1} \sum_{N=1}^{n} \left\{ \partial z_N(\varphi)/\partial \varphi \right\}\left\{ \partial z_N(\varphi)/\partial \varphi \right\}', \\
\theta = \rho_{11,11} - \rho_{11}^2 + 2\rho_{10:11}.
\]

In fact using Withers (1994b) one may develop a formal Edgeworth expansion for the distribution of \( n^{1/2}(\hat{\varphi} - \varphi). \)

**Note 1.1** Let \( \tilde{F} \) be the distribution of \{\( \tilde{\epsilon}_N = \bar{z} + \epsilon_N, 1 \leq N \leq n \). So \( Y_N = \tilde{z}_N(\varphi) + \tilde{\epsilon}_N \)
where \( \tilde{z}_N(\varphi) = z_N(\varphi) - \bar{z} \) has mean 0. So with centering condition \( \rho_{10}(\tilde{F}) = 0 \),
the asymptotic variance simplifies to
\[
V(\varphi) = \rho_{20}(\tilde{F})^{-2}\rho_{10,10}(\tilde{F})V_z^{-1}
\]
where \( V_\varepsilon = n^{-1} \sum_{i=1}^{n} (z_N - \varepsilon)(z_N - \varepsilon)' = V_{\varepsilon} \varepsilon, z_N = \partial z_N(\varphi)/\partial \varphi. \)

This formula is the same as for M-estimates! \( \square \)

**Note 1.2** The centering condition \( \rho_{10} = 0 \) can be viewed as determining the location parameter. Suppose \( \varphi = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \) with \( \alpha \) scalar and \( z_N(\varphi) = \alpha + x_N(\beta) \). Let \( F^* \) be the distribution of \( \{ \varepsilon_N = \varepsilon_N + \alpha \} \). Then \( Y_N = x_N(\beta) + \varepsilon_N \) so

\[
0 = \rho_{10} = \int (e - \alpha)a(F^*(e))dF^*(e)
\]

giving \( \alpha = \int ea(F^*(e))dF^*(e)/\int b(t)dt \)

for \( F \) continuous. In this way the dimensionality of the problem is decreased by 1 since generally \( \alpha \) is only a nuisance parameter. \( \square \)

**Note 1.3** \( V(\varphi) \) is determined by \( \varphi \) and \( F \), say \( V(\varphi) = V(\varphi, F) \). Let \( \hat{F} \) be the empirical distribution of \( \{ \varepsilon_N(\varphi) \} \). Then for smooth \( a(\cdot, \cdot) \) under weak conditions, \( \hat{V} = V(\hat{\varphi}, \hat{F}) \) is a consistent estimate of \( V(\varphi) \), so that an asymptotically \( 1 - \alpha \) level confidence region for a smooth function \( g(\varphi) \) in \( R^q \) with \( q \leq p \) is

\[
(g(\varphi) - g(\hat{\varphi}))' G^{-1} (g(\varphi) - g(\hat{\varphi})) < \chi^2_{q, 1-\alpha}
\]

where \( G = T(\varphi) \hat{V} T(\varphi)' \) and \( T(\varphi) = \partial g(\varphi)/\partial \varphi' \). \( \square \)

Examples are given in Section 3. In particular \textbf{L-estimates} are shown to be a special case.

## 2 Proof

Here we sketch the proof of Theorem 1.1.

Write the ordered values of \( \{ \varepsilon_N(\varphi) \} \) as \( \{ \varepsilon_{(N)} = \varepsilon_{(N)}(\varphi) = Y((N) - z_{(N)}(\varphi)) \} \).

Let a subscript \( -i,j \ldots \) denote \( \partial_i \partial_j \ldots \) where \( \partial_i = \partial/\partial \varphi_i \). We generally suppress the argument \( \varphi \). Assume that \( \hat{\varphi} \) minimising \( \lambda_n \) of (1.2) satisfies for \( 1 \leq i \leq p \),

\[
0 = \hat{\lambda}_{n,i} = \partial \lambda_n(\hat{\varphi})/\partial \hat{\varphi}_i = n^{-1} \sum_{N=1}^{N=n} \partial p(\varepsilon_{(N)}(\hat{\varphi}), N/n)/\partial \hat{\varphi}_i
\]
\[ = -n^{-1} \sum_{N=1}^{n} \rho_{10}(\epsilon_{(N)}(\hat{\phi}), N/n) z_{(N),i}(\hat{\phi}) \]
\[ = -n^{-1} \sum_{N=1}^{n} \{ \rho_{10}(\epsilon_{(N)}, N/n) - \rho_{20}(\epsilon_{(N)}, N/n) z_{(N),j}\delta_j + \ldots \} \{ z_{(N),i} + z_{(N),ik}\delta_k + \ldots \} \]

expanding about \( \varphi \), where \( \delta = \hat{\phi} - \varphi \) and summation of repeated pairs of suffixes \( j, k, \ldots \) over \( 1, \ldots, p \) is implicit. For example \( z_{(N),j}\delta_j = \sum_{j=1}^{p} z_{(N),j}\delta_j \).

Assume that \( \delta = O_p(n^{-1/2}) \) as \( n \to \infty \), so
\[ 0 = \hat{\lambda}_{n,i} = -n^{-1} \sum_{N=1}^{n} \rho_{10}(\epsilon_{(N)}, F_n(\epsilon_{(N)})) z_{N,i} + O_p(n^{-1/2}) \]
\[ = -n^{-1} \sum_{N=1}^{n} \rho_{10}(\epsilon_{(N)}, F_n(\epsilon_{(N)})) z_{N,i} + O_p(n^{-1/2}) \]

Since \( \tilde{z} \not\to 0 \) as \( n \to \infty \) in general, taking means gives the centering condition
\[ \rho_{10} = 0. \quad (2.1) \]
Conversely this condition implies that \( \delta = O_p(n^{-1/2}) \). Now
\[ 0 = \hat{\lambda}_{n,i} = A_{ni} + B_{nij}\delta_j + C_{niij}\delta_j + O_p(n^{-1}) \quad (2.2) \]
where
\[ A_{ni} = -n^{-1} \sum_{N=1}^{n} \rho_{10}(\epsilon_{(N)}, N/n) z_{(N),i}, \]
\[ B_{nij} = n^{-1} \sum_{N=1}^{n} \rho_{20}(\epsilon_{(N)}, N/n) z_{(N),i} z_{(N),j} \]
\[ = n^{-1} \sum_{N=1}^{n} \rho_{20}(\sum_{N} F_n(\epsilon_{(N)})) z_{N,i} z_{N,j} \]
\[ = B'_{nij} + O_p(n^{-1/2}) = B''_{nij} + O_p(n^{-1/2}), \]
\[ B'_{nij} = n^{-1} \sum_{N=1}^{n} \rho_{20}(\epsilon_{(N)}, F(\epsilon_{(N)})) z_{N,i} z_{N,j}, \]
\[ B''_{nij} = E B'_{nij} = \rho_{20} V_{ij}, \]
\[ V_{ij} = (V_z)_{ij} = n^{-1} \sum_{N=1}^{n} z_{N,i} z_{N,j}, \]
\[ C_{niij} = -n^{-1} \sum_{N=1}^{n} \rho_{10}(\epsilon_{(N)}, N/n) z_{(N),ij}, \]
\[ = -n^{-1} \sum_{N=1}^{n} \rho_{10}(\epsilon_{(N)}, F(\epsilon_{(N)})) z_{N,ij} + O_p(n^{-1/2}) \]
\[ = O_p(n^{-1/2}) \text{ by (2.1)} \]
So by (2.2),

\[ 0 = A_{n_i} + \rho_{20} V_i \delta_j + O_p(n^{-1}) \text{ giving} \]

\[ \delta_j = -\rho_{20}^{-1} V^{ij} A_{n_i} + O_p(n^{-1}) \]  (2.3)

where \((V^{ij}) = (V_{ij})^{-1}\) is \( p \times p \).

Now

\[- A_{n_i} = n^{-1} \sum_{N=1}^{n} \rho_{10}(\epsilon_N, F_n(\epsilon_n)) z_{N;i} = z_i T^i(F_n, F_{n_i}) \]  (2.4)

where

\[ T^i(F_n, F_{n_i}) = \int \rho_{10}(\epsilon, F_n(\epsilon)) dF_{n_i}(\epsilon) \]

and \(F_{n_i}(\epsilon) = n^{-1} \sum_{N=1}^{n} I(\epsilon_N \leq \epsilon) w_{N;i} \) for \(w_{N;i} = z_i^{-1} z_{N;i}\).

(The assumption \(z_i \neq 0\) can be removed by a limiting argument later). By Section 5 of Withers (1994b), \(T^i(F_n, F_{n_i}), 1 \leq i \leq p\) is asymptotically \(N_p(\mu, v_n^{-1})\)

where \(\mu_i = T^i(F, F_i) = 0\) by (2.1),

\[ F_i(x) = E F_{n_i}(x) = F(x), \]

\[ v_{ij} = \sum_{a, b=0}^{n} (w^{ab}) \left[ \frac{a}{1^i}, \frac{b}{1^j} \right] \] \text{ where } \( w_{N,0} = 1 \),

\[ (w^{ab}) = n^{-1} \sum_{N=1}^{n} w_{N,a} w_{N,b} = z^{-1}_a z^{-1}_b V_{ab}, \]

\[ \left[ \frac{a}{1^i}, \frac{b}{1^j} \right] = \int T^i(x) T^j(x) dF(x), \]

and \(T^i(x)\) is the first derivative of \(S(F_a) = T^i(F_0, F_i)\) where \(F_0 = F\).

So \(\sum_{a, b=0}^{p}\) can be replaced by \(\sum_{a=0, 1} \sum_{b=0, j}\) and

\[ T^i(\frac{x}{2}) = \int \rho_{11}(\epsilon, F(\epsilon)) F(\epsilon) dF(\epsilon) \]

where \(F(\epsilon)_x = I(x \leq \epsilon) - F(\epsilon)\) is the first derivative of \(S(F) = F(\epsilon)\),

and \(T^i(\frac{x}{2}) = \rho_{10}(x, F(x)) - T^i(F, F) = \rho_{10}(x, F(x))\).
Since \( \int F(\epsilon_1)F(\epsilon_2)dF(x) = F(\min(\epsilon_1, \epsilon_2)) - F(\epsilon_1)F(\epsilon_2) = C_{12} \) say,

\[
\begin{bmatrix}
0 & 0 \\
i & 1
\end{bmatrix}
= \int \int \rho_{11}(\epsilon_1, F(\epsilon_1))\rho_{11}(\epsilon_2, \mu(\epsilon_2))C_{12}dF(\epsilon_1)dF(\epsilon_2)
= \rho_{11,11} - \rho_{i1}^2.
\]

Also \( \begin{bmatrix}
0 & 0 \\
i & 1
\end{bmatrix}
= \int \int \rho_{11}(\epsilon, F(\epsilon))F(\epsilon)\rho_{10}(x, F(x))dF(\epsilon)dF(x)
= \rho_{10:11},
\]
and \( \begin{bmatrix}
i & 0 \\
i & 1
\end{bmatrix}
= \rho_{10,10}.
\)

So \( v_{ij} = \theta + (w^{ij})\rho_{10,10}. \)

So by (2.3), (2.4), \( n^{1/2}(\hat{\phi} - \phi) \xrightarrow{\mathcal{L}} \mathcal{N}_p(0, V(\phi)) \) where

\[
V_{\alpha\beta}(\phi) = \rho_{20}^{-2}\rho_{\alpha i}^{-1}\rho_{\beta j}^{-1}V^{ij}\]
\[
= \rho_{20}^{-2}\rho_{\alpha i}^{-1}\rho_{\beta j}^{-1}(\theta z_{i,j} + V_{ij}\rho_{10,10})V^{ij}.\]

So Theorem 1.1 holds.

3 Examples

**Example 3.1.** Suppose \( \rho(\epsilon, F) = \rho(\epsilon)a(F) \). Then integrating w.r.t. \( F \) and denoting a first derivative by a subscript 1,

\[
0 = \rho_{10} = \int \rho_1 a(F)dF,
\]

\[
\rho_{10,10} = \int \rho_1^2 a(F)^2 dF,
\]

\[
\rho_{11} = \int \rho_1 a_1(F)dF,
\]

\[
\rho_{11,11} = \int \rho_1^2 a_1(F)^2 dF,
\]

and \( \rho_{10:11} = \int \int_{x \leq y} \rho_1(x)a(F(x))\rho_1(y)a_1(F(y)dF(x)dF(y).\)

\[
\Box
\]

**Example 3.2.** For an \( M \)-estimate, \( \rho(\epsilon, F) = \rho(\epsilon) \) so \( 0 = \int \rho_1 dF, \rho_{10,10} = \int \rho_1^2 dF, \rho_{11} = \rho_{11,11} = \rho_{10:11} = 0 \) giving \( \theta = 0. \)

\[
\Box
\]
Example 3.3. For the "L - LSE estimate", $\rho(\epsilon, F) = \epsilon^2 a(F)/2$ so

$$0 = \rho_{10} = \int \epsilon a(F(\epsilon))dF(\epsilon),$$

$$\rho_{10,10} = \int \epsilon^2 a(F(\epsilon))dF(\epsilon),$$

$$\rho_{11} = \int \epsilon a_1(F(\epsilon))dF(\epsilon),$$

$$\rho_{11,11} = \int \epsilon^2 a_1(F(\epsilon))dF(\epsilon).$$

(a) For $a(F) = F - F^2$, $\rho_{11,11} = \int \epsilon(1 - 2F(\epsilon))^2dF(\epsilon)$ so essentially as for the LSE one needs $\int \epsilon^2 dF(\epsilon) < \infty$.

(b) For $a(F) = (F - F^2)^2/2$, $\rho_{11,11} = \int \epsilon^2(F(\epsilon) - F(\epsilon)^2)(1 - 2F(\epsilon))^2dF(\epsilon)$ so the requirement of finite variance weakens to $\int \epsilon^2 F(\epsilon)^2(1 - F(\epsilon))^2dF(\epsilon) < \infty$ which is a far weaker requirement. For example if

$$F(x) \approx c(-x)^{-\alpha} \quad \text{as} \quad x \to -\infty \quad \text{and} \quad 1 - F(x) \approx dx^{-\beta} \quad \text{as} \quad x \to \infty, \quad (3.1)$$

then it is sufficient that $\min(\alpha, \beta) > 2/3$. So this covers distributions like the Cauchy for which $\alpha = \beta = 1$ and the mean does not exist.

(c) Similarly if $a(F) = (F - F^2)^\gamma$ where $\gamma > 1$ then $\rho_{11,11} < \infty$ if (3.1) holds and $\alpha > 2/(2\gamma - 1)$. So by choosing $\gamma$ large enough one can deal with distributions with power tails decreasing arbitrarily slowly. \hfill \square

Example 3.3.1. Suppose in Example 3.3 that $p = 1$ and $z_N(\varphi) = \varphi$. Then

$$\varphi = \sum_{N=1}^n Y_{(N)} a(N/n) / \sum_{N=1}^n a(N/n)$$

$$= \int_0^1 F_{nY}^{-1}(t)a(t)dt / \left\{ \int_0^1 a(t)dt + O(n^{-1}) \right\}$$

by the Euler-McLaurin expansion, where $F_{nY}$ is the empirical distribution of $\{Y_1, \ldots, Y_n\}$. So $\varphi$ is an $L$-estimate. So by p. 276 of Serfling (1980), $n^{1/2}(\varphi - \varphi_0) \overset{d}{\to} N(0, v)$ where

$$\varphi_0 = \int_0^1 F_{Y}^{-1}(t)a(t)dt / \int_0^1 a(t)dt.$$
\[
\phi = \frac{1}{\int_0^1 F^{-1}(t) a(t) dt / \int_0^1 a(t) dt}
\]

\[
\phi \quad \text{since} \quad \rho_{10} = 0,
\]

\[
F_Y(y) = P(Y_1 \leq y) = F(y - \phi),
\]

\[
v = \int \int a(F(x)) a(F(y)) \{ F(\min(x, y)) - F(x) F(y) \} dx dy.
\]

This gives an alternative formula for the asymptotic variance. \(\square\)

References


