THE INSTITUTE OF STATISTICS

THE CONSOLIDATED UNIVERSITY
OF NORTH CAROLINA

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

by C.S. Withers

July 1994

Mimeo Series #2324

DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina

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

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

where φ in 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 R.

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

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

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 sufficently 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

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

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

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

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** $Ea(\epsilon_1, F(\epsilon_1)) = \int a(\epsilon, F(\epsilon))dF(\epsilon)$.

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

Set

$$\begin{aligned} \rho_{rs} &= \rho_{rs}(F) = \rho_{rs}(\epsilon, F(\epsilon))dF(\epsilon), \\ \rho_{rs,rs} &= \rho_{rs,rs}(F) = \int \rho_{rs}(s, F(\epsilon))^2 dF(\epsilon) \\ \text{and} \quad \rho_{10:11} &= \int\int_{x \leq y} \rho_{10}(x, F(x))\rho_{11}(y, F(y))dF(x)dF(y). \end{aligned}$$

In Section 2 we prove our main result:

Theorem 1.1 If $\rho_{10} = 0$ then

$$\begin{aligned} n^{1/2}(\hat{\varphi} - \varphi) &\xrightarrow{\mathcal{L}} \mathcal{N}(0, V(\varphi)) \quad \text{as} \quad n \to \infty \\ \text{where} \quad V(\varphi) &= \rho_{20}^{-2}(\rho_{10,10}V_z^{-1} + \theta\nu\nu'), \qquad (1.5) \\ \nu &= V_z^{-1}\bar{z}_., \\ \bar{z}_. &= \partial\bar{z}/\partial\varphi, \bar{z} = n^{-1}\sum_{N=1}^{n} z_N(\varphi), \\ V_z &= n^{-1}\sum_{N=1}^{n}\{\partial z_N(\varphi)/\partial\varphi\}\{\partial z_N(\varphi)/\partial\varphi\}', \\ \theta &= \rho_{11,11} - \rho_{11}^2 + 2\rho_{10:11}. \end{aligned}$$

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

$$V(\varphi) = \rho_{20}(\tilde{F})^{-2}\rho_{10,10}(\tilde{F})V_{\tilde{z}}^{-1} \qquad (1.6)$$

where $V_{\bar{z}} = n^{-1}\sum_{N=1}^{n}(z_{N\cdot} - \bar{z})(z_{N\cdot} - \bar{z}.)' = V_z - \bar{z}.\bar{z}'_{\cdot}$, $z_{N\cdot i} = \partial z_N(\varphi)/\partial\varphi$.

This formula is the same as for M-estimates! □

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

$0 = \rho_{10} = \int(e-\alpha)a(F^*(e))dF^*(e)$ giving $\alpha = \int ea(F^*(e))dF^*(e)/\int_0^1 a(t)dt$

for F continuous. In this way the dimensionality of the problem is decreased by 1 since generally α is only a nuisance parameter. □

Note 1.3 $V(\varphi)$ is determined by φ and F, say $V(\varphi) = V(\varphi, F)$. Let $\hat{F}$ be the empirical distribution of $\{e_N(\hat{\varphi})\}$. 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 $G = T(\hat{\varphi})\hat{V}T(\hat{\varphi})'$ and $T(\varphi) = \partial g(\varphi)/\partial\varphi'$. □

Examples are given in Section 3. In particular 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 $\{\epsilon_N(\varphi)\}$ as $\{\epsilon_{(N)} = \epsilon_{(N)}(\varphi) = Y((N) - z_{(N)}(\varphi)\}$. Let a subscript $\cdot ij\ldots$ denote $\partial_i\partial_j\ldots$ where $\partial_i = \partial/\partial\varphi_i$. We generally suppress the argument φ. Assume that $\hat{\varphi}$ minimising λ_n of (1.2) satisfies for $1 \leq i \leq p$,

$$0 \;=\; \hat{\lambda}_{n.i} = \partial\lambda_n(\hat{\varphi})/\partial\hat{\varphi}_i = n^{-1}\sum_{N=1}^{n}\partial\rho(\epsilon_{(N)}(\hat{\varphi}), N/n)/\partial\hat{\varphi}_i$$

$$= \quad -n^{-1} \sum_{N=1}^{n} \rho_{10}(\epsilon_{(N)}(\hat{\varphi}), N/n) z_{(N)\cdot i}(\hat{\varphi}))$$

$$= \quad -n^{-1} \sum_{N=1}^{n} \{\rho_{10}(\epsilon_{(N)}, N/n) - \rho_{20}(\epsilon_{(N)}, N/n) z_{(N)\cdot j}\delta_j + \ldots\} \left\{z_{(N)\cdot i} + z_{(N)\cdot ik}\delta_k + \ldots\right\}$$

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

Assume that δ is $O_p(n^{-1/2})$ as $n \to \infty$, so

$$\begin{aligned} 0 = \hat{\lambda}_{n\cdot i} &= -n^{-1} \sum_{N=1}^{n} \rho_{10}(\epsilon_N, F_n(\epsilon_N)) z_{N\cdot i} + O_p(n^{-1/2}) \\ &= -n^{-1} \sum_{N=1}^{n} \rho_{10}(\epsilon_N, F_n(\epsilon_N)) z_{N\cdot i} + O_p(n^{-1/2}) \end{aligned}$$

Since $\bar{z}. \not\to 0$ as $n \to \infty$ in general, taking means gives the centering condition

$$\rho_{10} = 0. \tag{2.1}$$

Conversely this condition implies that $\delta = O_p(n^{-1/2})$. Now

$$0 = \hat{\lambda}_{n-i} = A_{ni} + B_{nij}\delta_j + C_{nij}\delta_j + O_p(n^{-1}) \tag{2.2}$$

where

$$\begin{aligned} A_{ni} &= -n^{-1} \sum_{N=1}^{n} \rho_{10}(\epsilon_{(N)}, N/n) z_{(N)\cdot i}, \\ B_{nij} &= n^{-1} \sum_{N=1}^{n} \rho_{20}(\epsilon_{(N)}, N/n) z_{(N)\cdot i} z_{(N)\cdot j} \\ &= n^{-1} \sum_{N=1}^{n} \rho_{20}(\sum_N, F_n(\epsilon_N)) z_{N\cdot i} z_{N\cdot j} \\ &= B'_{nij} + O_p(n^{-1/2}) = B''_{nij} + O_p(n^{-1/2}), \\ B'_{nij} &= n^{-1} \sum_{N=1}^{n} \rho_{20}(\epsilon_N, F(\epsilon_N)) z_{N\cdot i} z_{N\cdot j}, \\ B''_{nij} &= EB'_{nij} = \rho_{20} V_{ij}, \\ V_{ij} &= (V_z)_{ij} = n^{-1} \sum_{N=1}^{n} z_{N\cdot i} z_{N\cdot j}, \\ C_{nij} &= -n^{-1} \sum_{N=1}^{n} \rho_{10}(\epsilon_{(N)}, N/n) z_{(N)\cdot ij}, \\ &= -n^{-1} \sum_{N=1}^{n} \rho_{10}(\epsilon_N, F(\epsilon_N)) z_{N\cdot ij} + O_p(n^{-1/2}) \\ &= O_p(n^{-1/2} \quad \text{by (2.1)} \end{aligned}$$

So by (2.2),

$$\begin{aligned} 0 &= A_{ni} + \rho_{20} V_{ij} \delta_j + O_p(n^{-1}) \quad \text{giving} \\ \delta_j &= -\rho_{20}^{-1} V^{ji} A_{ni} + O_p(n^{-1}) \qquad (2.3) \\ \text{where} \quad (V^{ij}) &= (V_{ij})^{-1} \quad \text{is} \quad p \times p. \end{aligned}$$

Now

$$-A_{ni} = n^{-1} \sum_{N=1}^{n} \rho_{10}(\epsilon_N, F_n(\epsilon_n)) z_{N \cdot i} = \bar{z}_{\cdot i} T^i(F_n, F_{ni}) \qquad (2.4)$$

where

$$T^i(F_n, F_{ni}) = \int \rho_{10}(\epsilon, F_n(\epsilon)) dF_{ni}(\epsilon)$$

and $F_{ni}(\epsilon) = n^{-1} \sum_{N=1}^{n} I(\epsilon_N \leq \epsilon) w_{N \cdot i}$ for $w_{N \cdot i} = \bar{z}_{\cdot i}^{-1} z_{N \cdot i}$.

(The assumption $\bar{z}_{\cdot i} \neq 0$ can be removed by a limiting argument later). By Section 5 of Withers (1994b), $\{T^i(F_n, F_{ni}), 1 \leq i \leq p\}$ is asymptotically $\mathcal{N}_p(\mu, vn^{-1})$ where $\mu_i = T^i(F, F_i) = 0$ by (2.1),

$$F_i(x) = EF_{ni}(x) = F(x),$$

$$v_{ij} = \sum_{a,b=0}^{n} (w^{ab}) \left[{}^a_1 i, {}^b_1 j\right]_1 \quad \text{where } w_{N \cdot 0} = 1,$$

$$(w^{ab}) = n^{-1} \sum_{N=1}^{n} w_{N \cdot a} w_{N \cdot b} = \bar{z}_{\cdot a}^{-1} \bar{z}_{\cdot b}^{-1} V_{ab},$$

$$\left[{}^a_1 i, {}^b_1 j\right]_1 = \int T^i({}^a_x) T^j({}^b_x) dF(x),$$

and $T^i({}^a_x)$ 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

$$T^i({}^0_x) = \int \rho_{11}(\epsilon, F(\epsilon)) F(\epsilon)_x dF(\epsilon)$$

where $F(\epsilon)_x = I(x \leq \epsilon) - F(\epsilon)$ is the first derivative of $S(F) = F(\epsilon)$, and $T^i({}^0_x) = \rho_{10}(x, F(x)) - T^i(F, F) = \rho_{10}(x, F(x))$.

Since $\int F(\epsilon_1)_x F(\epsilon_2)_x dF(x) = F(\min(\epsilon_1,\epsilon_2)) - F(\epsilon_1)F(\epsilon_2) = C_{12}$ say,

$$\begin{aligned}
\left[{}^0_1 i, {}^0_1 j\right]_1 &= \int\int \rho_{11}(\epsilon_1, F(\epsilon_1))\rho_{11}(\epsilon_2, \mu(\epsilon_2)) C_{12} dF(\epsilon_1) dF(\epsilon_2) \\
&= \rho_{11,11} - \rho_{11}^2. \\
\text{Also} \quad \left[{}^0_1 i, {}^j_1 j\right]_1 &= \int\int \rho_{11}(\epsilon, F(\epsilon)) F(\epsilon)_x \rho_{10}(x, F(x)) dF(\epsilon) dF(x) \\
&= \rho_{10:11}, \\
\text{and} \quad \left[{}^i_1 i, {}^j_1 j\right]_1 &= \rho_{10,10}. \\
\text{So} \quad v_{ij} &= \theta + (w^{ij})\rho_{10,10}.
\end{aligned}$$

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

$$\begin{aligned}
V_{\alpha\beta}(\varphi) &= \rho_{20}^{-2} V^{\alpha i}(\bar z_{\cdot i} v_{ij} \bar z_{\cdot j}) V^{j\beta} \\
&= \rho_{20}^{-2} V^{\alpha i}(\theta \bar z_{\cdot i} \bar z_{\cdot j} + V_{ij}\rho_{10,10}) V^{j\beta}.
\end{aligned}$$

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,

$$\begin{aligned}
0 = \rho_{10} &= \int \rho_1 a(F) dF, \\
\rho_{10,10} &= \int \rho_1^2 a(F)^2 dF, \\
\rho_{11} &= \int \rho_1 a_1(F) dF, \\
\rho_{11,11} &= \int \rho_1^2 a_1(F)^2 dF, \\
\text{and} \quad \rho_{10:11} &= \int\int_{x\le y} \rho_1(x) a(F(x)) \rho_1(y) a_1(F(y) dF(x) dF(y).
\end{aligned}$$

□

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 $\theta = 0$. □

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

$$\begin{aligned} 0 &= \rho_{10} = \int \epsilon a(F(\epsilon))dF(\epsilon), \\ \rho_{10,10} &= \int \epsilon^2 a(F(\epsilon))^2 dF(\epsilon), \\ \rho_{11} &= \int \epsilon a_1(F(\epsilon))dF(\epsilon), \\ \rho_{11,11} &= \int \epsilon^2 a_1(F(\epsilon))^2 dF(\epsilon). \end{aligned}$$

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

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

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

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 γ large enough one can deal with distributions with power tails decreasing arbitrarily slowly. □

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

$$\begin{aligned} \hat{\varphi} &= \sum_{N=1}^{n} Y_{(N)} a(N/n) / \sum_{N=1}^{n} a(N/n) \\ &= \int_0^1 F_{nY}^{-1}(t)a(t)dt / \left\{ \int_0^1 a(t)dt + O(n^{-1}) \right\} \end{aligned}$$

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

$$\varphi_0 = \int_0^1 F_Y^{-1}(t)a(t)dt / \int_0^1 a(t)dt,$$

$$
\begin{aligned}
&= \varphi + \int_0^1 F^{-1}(t)a(t)dt \Big/ \int_0^1 a(t)dt \\
&= \varphi \quad \text{since} \quad \rho_{10} = 0, \\
F_Y(y) &= P(Y_1 \le y) = F(y-\varphi), \\
v &= \int\int a(F(x))a(F(y))\left\{F(min(x,y)) - F(x)F(y)\right)dxdy.
\end{aligned}
$$

This gives an alternative formula for the asymptotic variance. □

References

[1] Hosking, J.R.M. (1990) L-moments: analysis and estimation of distributions using linear combinations of order statistics. *J. R. Statist. Soc.* **52**, 105-124.

[2] Hosking, J.R. M., Wallis, J.R. and Wood, E.F. (1985) Estimation of the generalised extreme-value distribution by the method of probability-weighted-moments. *Technometrics*, **27**, 251-261.

[3] Serfling, R.J. (1980) Approximation theorems of mathematical statistics. Wiley, New York.

[4] Withers, C.S. (1994a) The bias and skewness of M-estimates in regression. Submitted.

[5] Withers, C.S. (1994b) Edgeworth expansions for functions of weighted empirical distributions. Submitted.

[6] Withers, C.S. and Pearson, C.P. (1994) Bias-reduced estimates of probability weighted moments with application to L-skewness and L-kurtosis. Submitted.