TECHNICAL REPORT 90.01

ONE-STEP M-ESTIMATORS IN THE LINEAR MODEL,
WITH DEPENDENT ERRORS

by

CHRISTOPHER A. FIELD

and

DOUGLAS P. WIENS
2. CONSTRUCTION AND ASYMPTOTIC PROPERTIES OF THE ESTIMATOR

The construction of the estimator of $\theta_0$ in the model defined by (1.1), (1.2) requires a preliminary, $\sqrt{n}$-consistent estimator $\hat{\theta}_*$ of $\theta_0$. In the simulations detailed in Section 3 we have used the minimum $L_1$-norm estimator, whose asymptotic properties are discussed in Basset and Koenker (1978) and Mehra and Rao (1988). See also Dodge (1987).

Given $\theta_*$, we construct an estimate $P_* = \hat{P}(\theta_*)$ of $P$, and then an estimate $U_* = \hat{U}(\theta_*)$ of $U$. As $\hat{U}(\theta)$ one may take any $n \times n$ matrix satisfying $\hat{U}(\theta)\hat{U}^T(\theta) = \hat{P}(\theta)$. Now define

$$A = U^{-1}X, \hat{A}(\theta) = \hat{U}^{-1}(\theta)X, A_* = \hat{A}(\theta_*) = U_*^{-1}X.$$  

Define also transformed observations

$$z = U^{-1}y, z_* = U_*^{-1}y$$

and associated residual vectors

$$\epsilon_+ = z - A\theta_*, \epsilon_* = z_* - A_*\theta_*.$$  

Let the elements of $\epsilon_+, \epsilon_*$ have empirical distribution functions (e.d.f.s) $F_n^+, F_n^*$ respectively. Recall that $F$ is the d.f. of the i.i.d. errors $\epsilon_1, \ldots, \epsilon_n$. Let $\sigma(F)$ be a positive, scale equivariant, shift invariant functional of $F$, and define

$$\sigma_+ = \sigma(F_n^+), \sigma_* = \sigma(F_n^*).$$

For an absolutely continuous function $\psi$ satisfying conditions given below, define

$$\psi_\sigma(x) = \psi(x/\sigma), (\sigma > 0);$$  

$$\psi_\sigma(x) = (\psi_\sigma(x_1), \ldots, \psi_\sigma(x_n))^T, \text{ if } x = (x_1, \ldots, x_n)^T;$$  

$$D(\psi_\sigma, F) = E_F[\psi'_\sigma(\epsilon)]. \quad (2.1)$$

A somewhat more general definition of $D(\psi_\sigma, F)$ is given in assumption S2) below.

Were $P$ known, we would have the exact model

$$z = A\theta_0 + \epsilon,$$  

$$1$$
with independent errors. In this situation Bickel (1975) studied the one-step $M$-estimator

$$\theta_B = \theta_* + (A^T A)^{-1} A^T \psi_{\sigma_*}(\varepsilon_+)/\hat{D}_n(\psi_{\sigma_*}, F),$$

where $\hat{D}_n(\psi_{\sigma}, F)$ estimates $D(\psi_{\sigma}, F)$. Under a set of assumptions contained in ours, Bickel showed that

$$\sqrt{n}(\theta_B - \theta_0) \overset{w}{\Rightarrow} N_p(0, V(\psi_{\sigma}, F)A_0^{-1}) \quad (2.3)$$

where:

$$A_0 = \lim_{n \to \infty} (A^T A)/n,$$

$$V(\psi, F) = E_P[\psi^2(\varepsilon)]/D^2(\psi, F),$$

$$\sigma = \sigma(F).$$

We are proposing the estimator

$$\hat{\theta} = \theta_* + (A_*^T A_*)^{-1} A_*^T \psi_{\sigma_*}(\varepsilon_+)/\hat{D}_n(\psi_{\sigma_*}, F),$$

where $\hat{D}_n(\psi_{\sigma_*}, F)$ is any consistent estimator of $D(\psi_{\sigma}, F)$ for $\sigma = \sigma(F)$. Under the assumptions given below, $\hat{\theta}$ and $\theta_B$ are $\sqrt{n}$-equivalent:

$$\sqrt{n}(\hat{\theta} - \theta_B) \overset{p}{\Rightarrow} 0,$$

and hence by (2.3)

$$\sqrt{n}(\hat{\theta} - \theta_0) \overset{w}{\Rightarrow} N_p(0, V(\psi_{\sigma}, F)A_0^{-1}). \quad (2.4)$$

Revised estimates of $P$ and $\sigma(F)$ are then given by $\hat{P}(\hat{\theta})$ and $\sigma(\hat{F}_n)$, where $\hat{F}_n$ is the e.d.f. of the elements of $\hat{\varepsilon} = \hat{U}^{-1}(\hat{\theta})(y - X\hat{\theta}).$

**Assumptions.** The asymptotic theory will utilize the following assumptions.

A) The distribution function $F$ of $\varepsilon_1$ is symmetric about 0; $\psi$ is odd, bounded and absolutely continuous, with a bounded, piecewise continuous derivative $\psi'$. Write $\psi = \psi^+ - \psi^-$, where $\psi^\pm$ is monotone increasing.

B) $\theta_* - \theta_0 = O_p(n^{-1/2}).$

D) $\sigma(F^n_+), \sigma(F^n^+) = \sigma(F) + O_p(n^{-1/2})$, where $\sigma(F) > 0$;

$$\hat{D}_n(\psi_{\sigma_*}, F) \overset{p}{\Rightarrow} D(\psi_{\sigma}, F) > 0,$$

$$\hat{D}_n(\psi_{\sigma_*}, F) \overset{p}{\Rightarrow} D(\psi_{\sigma}, F),$$

where $\sigma = \sigma(F)$. 

2
G) \( \lim_{n \to \infty} A^T A/n = A_0 \), a positive definite matrix, and \( \lim \max_{i,j} |A_{ij}|/\sqrt{n} = 0 \).

H) \( A_*^T A_*/n \xrightarrow{L} A_0 \), \( A_*^T A/n \xrightarrow{L} A_0 \). By this, if \( \Delta = (A - A_*)/\sqrt{n} \), then \( \Delta^T \Delta \xrightarrow{L} 0 \), hence in particular \( ch_1(\Delta^T \Delta) \xrightarrow{L} 0 \), where \( ch_1 \) denotes the largest characteristic root.

K) The matrix \( S = U_*^{-1}U - I \) satisfies \( tr(STS) \xrightarrow{L} 0 \).

L) \( \max_{i,j} |E[\hat{A}_{ij}(\theta)]| = O(1) \) when \( \|\theta - \theta_0\| \leq M/\sqrt{n} \), where \( M \) is a generic constant and \( \| \cdot \| \) is the Euclidean norm.

S1) For some \( \varepsilon > 0 \) we have

\[
\sup \{ \frac{1}{q^2} \int (\psi^\pm((1 + \lambda + q)(x + h)) - \psi^\pm((1 + \lambda)(x + h)))^2dF(x) : |h| \leq \varepsilon, |\lambda| \leq \varepsilon, |q| \leq \varepsilon \} < \infty,
\]

and for some \( \varepsilon > 0 \) we have

\[
\sup \{ \frac{1}{|h|} \int (\psi^\pm((1 + \lambda)x + h) - \psi^\pm((1 + \lambda)x - h))dF(x) : |h| \leq \varepsilon, |\lambda| \leq \varepsilon \} < \infty.
\]

S2) There exists \( D(\psi^\pm, F) \) such that

\[
\int [\psi^\pm((1 + \lambda)x + h) - \psi^\pm(x)]dF(x) = \frac{D(\psi^\pm, F)}{h} + o(|h|) + O(|\lambda h|) + O(\lambda^2).
\]

Define

\[
D(\psi, F) = D(\psi^+, F) - D(\psi^-, F). \tag{2.5}
\]

Assumptions S1) and S2) are identical to their analogues in Bickel (1975), where it is noted that, in the presence of \( A \), they are satisfied if we may formally differentiate under the integral sign. In that case, (2.1) and (2.5) agree. Assumptions A), B), D), G) also have close analogues in Bickel (1975). Of the remaining assumptions, K) is the most difficult to verify in practice. We believe that our assumptions are far from the weakest possible, and that in particular K) can probably be weakened.

**Asymptotic Theory.** For the remainder of this section we assume, without loss of generality, that \( \theta_0 = 0 \). Define

\[
\epsilon(t) = z - At = \epsilon - At,
\]

3
and put
\[ T_n(t, \lambda) = (1/\sqrt{n})A^T[\psi((1 + \lambda)e(t)) - E[\psi((1 + \lambda)e(t))]]. \]

Lemma 4.1 of Bickel (1975), in our notation, is:

**Lemma 1.** As \( n \to \infty \) and \( \varepsilon_n \downarrow 0 \), and where \( M \) is generic constant,
\[ \sup\{\|T_n(t, \lambda) - T_n(0, 0)\| : \|t\| \leq M/\sqrt{n}, |\lambda| \leq \varepsilon_n\} \overset{p}{\to} 0. \]

Now define
\[ \varepsilon_*(t) = z_* - A_0 t, \]
and put
\[ T_n^*(t, \lambda) = (1/\sqrt{n})A^T[\psi((1 + \lambda)e_*(t)) - E[\psi((1 + \lambda)e_*(t))]]. \]

In the Appendix, we prove

**Lemma 2.** As \( n \to \infty \) and \( \varepsilon_n \downarrow 0 \),
\[ \sup\{\|T_n^*(t, \lambda) - T_n(t, \lambda)\| : \|t\| \leq M/\sqrt{n}, |\lambda| \leq \varepsilon_n\} \overset{p}{\to} 0. \]

As in Bickel (1975), expansions of the expectations in \( T_n \) and \( T_n^* \), using S1), S2), Lemma 1 and Lemma 2, immediately yield:

**Proposition 1.** As \( n \to \infty \),
\[ \sup\{(1/\sqrt{n})\|A^T[\psi((1 + \lambda)e(t)) - \psi(\varepsilon)) + A^T A_{\lambda} \cdot D(\psi, F)]\| : \|t\| \leq M/\sqrt{n}, |\lambda| \leq M/\sqrt{n}\} \overset{p}{\to} 0, \tag{2.6} \]
and
\[ \sup\{(1/\sqrt{n})\|A^T[\psi((1 + \lambda)e_*(t)) - \psi(\varepsilon)) + A^T A_{\lambda} \cdot D(\psi, F)]\| : \|t\| \leq M/\sqrt{n}, |\lambda| \leq M/\sqrt{n}\} \overset{p}{\to} 0. \tag{2.7} \]

We have:

**Theorem 1.** As \( n \to \infty \), \( \sqrt{n}(\hat{\theta} - \theta_B) \overset{p}{\to} 0 \), so that, as well, asymptotic normality as at (2.4) holds.
APPENDIX: PROOFS

Lemma 2: It suffices to show

\[ \sup\{(1/\sqrt{n})\|A^T[\psi((1 + \lambda)e_\ast(t)) - \psi((1 + \lambda)e(t))]\| : \|t\| \leq M/\sqrt{n}, |\lambda| \leq \varepsilon_n \} \xrightarrow{P.T.} 0, \quad (A.1) \]

\[ \sup\{(1/\sqrt{n})\|A^T[E[\psi((1 + \lambda)e_\ast(t)) - \psi((1 + \lambda)e(t))]\| : \|t\| \leq M/\sqrt{n}, |\lambda| \leq \varepsilon_n \} \xrightarrow{P.T.} 0. \quad (A.2) \]

Let \( \delta(t) = e_\ast(t) - e(t) \) and note that then \( \delta(t) = S\varepsilon + \Delta\sqrt{n}t \). We first show that

\[ \max_{1 \leq i \leq n} \sup_{\|t\| \leq M/\sqrt{n}} |\delta_i(t)| \xrightarrow{L^2} 0. \quad (A.3) \]

For this, note that for \( \|t\| \leq M/\sqrt{n}, \|\Delta\sqrt{n}t\|^2 \leq M^2ch_1(\Delta^T\Delta) \xrightarrow{L^1} 0, \) by \( H \). To see that, as well

\[ \|S\varepsilon\| \xrightarrow{L^2} 0, \quad (A.4) \]

first assume that the fourth moment of \( F \) is finite. It follows from the symmetry of \( F \) that the conditional distribution of \( \varepsilon \) given \( P_\ast \) has symmetric univariate and bivariate marginals, and so

\[ E[\|S\varepsilon\|^2] = E[tr\{S^T S \cdot E[\varepsilon\varepsilon^T | P_\ast]\}] = E[E[\varepsilon^2 | P_\ast]trS^T S] \]

\[ \leq \{E[\varepsilon^4] \cdot E[tr^2(S^T S)]\}^{1/2}. \]

Now \( K \) gives \( (A.4) \). To drop the moment restriction, repeat the above argument with the \( \varepsilon \), truncated at \( \pm c \), then let \( c \to \infty \).

Now use \( (A.3) \) and the piecewise continuity of \( \psi' \) to write

\[ (1/\sqrt{n})A^T[\psi((1 + \lambda)e_\ast(t)) - \psi((1 + \lambda)e(t))] = (1 + \lambda)A^T\pi(t)/\sqrt{n}, \]

where \( \pi_i(t) = d_i\delta_i(t) \) and \( d_i = \psi'((1 + \lambda)(\varepsilon_i(t) + \alpha_i\delta_i(t)) \) for some \( \alpha_i \in (0, 1) \). Now

\[ \|(1 + \lambda)A^T\pi(t)/\sqrt{n}\|^2 \leq (1 + \varepsilon_n)^2ch_1(A^T A/n) \cdot \|\pi(t)\|^2 = O(\|\pi(t)\|^2) \quad (A.5) \]

by \( G \), and

\[ \|\pi(t)\|^2 \leq (\sup |\psi'|)^2\|\delta(t)\|^2, \quad (A.6) \]
and so (A.3) gives (A.1). For (A.2), replace \( \pi(t) \) by its expectation in (A.5), then bound 
\[ \| E[\pi(t)] \|^2 \] by 
\[ E[\| \pi(t) \|^2] \] and apply (A.6), (A.3).

\textbf{Theorem 1:} Decompose \( \sqrt{n}(\hat{\theta} - \theta_B) \) as 
\[ 
\left( \frac{A^T A_n}{n} \right)^{-1} \cdot \frac{1}{\sqrt{n}} A^T \psi_{\sigma_*}(\varepsilon_*) [ \hat{D}^{-1}_{n}(\psi_{\sigma_*}, F) - \hat{D}^{-1}_{n}(\psi_{\sigma_*}, F)] \\
+ \hat{D}^{-1}_{n}(\psi_{\sigma_*}, F) \left( \frac{A^T A_n}{n} \right)^{-1} \cdot \frac{1}{\sqrt{n}} [ A^T \psi_{\sigma_*}(\varepsilon_*) - A^T \psi_{\sigma_*}(\varepsilon_+)] \\
+ \hat{D}^{-1}_{n}(\psi_{\sigma_*}, F) \left[ \left( \frac{A^T A_n}{n} \right)^{-1} - \left( \frac{A^T A_n}{n} \right)^{-1} \right] \cdot \frac{1}{\sqrt{n}} A^T \psi_{\sigma_*}(\varepsilon_+). 
\]  
(A.7)

It follows immediately from (2.6) - with \( \lambda = (\sigma/\sigma_*) - 1, \psi = \psi_\sigma \), and \( t = \theta_* \) - and \( G \), as in the proof of Theorem 4.1 of Bickel (1975), that
\[ (1/\sqrt{n}) A^T \psi_{\sigma_*}(\varepsilon_+) = O_p(1) \text{ as } n \to \infty, \] 
(A.8)

hence that the last term in (A.7) tends to zero in probability. We will show that 
\[ (1/\sqrt{n}) A^T [\psi_{\sigma_*}(\varepsilon_*) - \psi_{\sigma_*}(\varepsilon_+)] \overset{p}{\to} 0, \] 
(A.9)

\[ (1/\sqrt{n}) [A^T_* - A^T] \psi_{\sigma_*}(\varepsilon_*) \overset{p}{\to} 0. \]  
(A.10)

From (A.9), (A.10), \( G \) and \( D \) it follows that the second term in (A.7) tends to zero in probability and hence, using (A.8), that the first term does as well.

To establish (A.9), apply (2.6) with \( \psi = \psi_\sigma, \lambda = (\sigma/\sigma_*) - 1, t_* = \theta_* \). Then apply (2.7) with \( \lambda = (\sigma/\sigma_*) - 1 \). By subtraction,
\[ \sup \{ \| A^T [\psi_{\sigma_*}(\varepsilon_*) - \psi_{\sigma_*}(\varepsilon_+)] \| : \| \theta_* \| \leq M/\sqrt{n}, |(\sigma/\sigma_*) - 1| \leq M/\sqrt{n}, |\sigma/\sigma_* - 1| \leq M/\sqrt{n} \} \overset{p}{\to} 0. \]

Now B), D) give (A.9).

By D), it suffices to prove (A.10) with \( \sigma_* \) replaced by \( \sigma(F) \), which we may assume to equal unity. We require
\[ \Delta^T \{ \psi(\varepsilon_*) - E[\psi(\varepsilon_*)] \} \overset{p}{\to} 0, \] 
(A.11)

\[ \Delta^T E[\psi(\varepsilon_*)] \overset{p}{\to} 0. \]  
(A.12)
A conditional argument similar to that used to establish (A.4), together with H), gives (A.11). For (A.12), put \( \hat{Q}(\theta) = \hat{U}^{-1}(\theta)U, \mu = E[\psi(\epsilon)] \). We claim that

\[
\max_{1 \leq i \leq n} \sup_{\|t\| \leq M/\sqrt{n}} \sqrt{n} |E[\psi((\hat{Q}(t)\epsilon(t)),i)| = O(1). \tag{A.13}
\]

Then since \( \hat{Q}(\theta_*)\epsilon(\theta_*) = \epsilon_* \) and \( \psi \) is bounded, we have that \( \max_{1 \leq i \leq n} \mu_i = O(n^{-1/2}) \) and hence

\[
\|\Delta^T E[\psi(\epsilon)]\|^2 \leq (ch_1(\Delta^T\Delta)(\sqrt{n} \max_i \mu_i)^2 \leq 0,
\]

by H). To see (A.13), note that \( E[\hat{Q}(t)\epsilon(t)] = E[\hat{Q}(t)E[\epsilon(t)|\hat{U}(t))] = -E[\hat{A}(t)]t \), so that by L), \( \sqrt{n} \max_{1 \leq i \leq n} \sup_{\|t\| \leq M/\sqrt{n}} |E[\hat{Q}(t)\epsilon(t)],i| = O(1) \). Now for \( \|t\| \leq M/\sqrt{n} \), bound \( |E[\psi((\hat{Q}(t)\epsilon(t)),i)]| \) by \( \sup |\psi'| \cdot |E[\hat{Q}(t)\epsilon(t)],i| \), to get (A.13).