The Annals of Statistics
1999, Vol. 27, No. 3, 1080–1104

TARGET ESTIMATION FOR BIAS AND MEAN SQUARE ERROR REDUCTION

BY JAVIER CABRERA AND LUISA TURRIN FERNHOLZ

Rutgers University and Temple University

Given a statistical functional T and a parametric family of distributions, a bias reduced functional $\widetilde{T}$ is defined by setting the expected value of the statistic equal to the observed value. Under certain regularity conditions this new statistic, called the target estimator, will have smaller bias and mean square error than the original estimator. The theoretical aspects are analyzed by using higher order von Mises expansions. Several examples are given, including M-estimates of location and scale. The procedure is applied to an autoregressive model, the errors-in-variables model and the logistic regression model. A comparison with the jackknife and the bootstrap estimators is also included.

1. Introduction. Statistical procedures for reducing the bias and the variance of estimators have always been of interest for their broad applicability. The well-known methods for bias reduction based on the jackknife, bootstrap and other resampling plans, presented by Efron (1982) and Efron and Tibshirani (1993) have proved to be quite useful, especially in nonparametric settings. However, these methods may not be effective in complex situations when the sampling distribution of the statistic changes too abruptly with the parameter, or when this distribution is very skewed and has heavy tails. In these situations it seems appropriate to consider the median bias of the estimator as well as the mean bias. Results in this direction were studied in Cabrera and Watson (1997) where certain mean and median bias procedures for bias reduction were introduced. The works of Debiche and Watson (1996) and Cabrera and Meer (1996) show some practical applications of these bias reduction procedures. However, most of the results by these authors consisted of practical applications of the median bias procedure, leaving all the theoretical questions and applications for mean bias reduction unanswered.

In this paper we address these unanswered questions. We introduce the definition of target estimators; given a statistic with finite expectation, the corresponding target estimator is defined as a function of the original estimator. We show that, under some regularity conditions, the target estimator will have a smaller bias than the original estimator. Moreover, we show that the

Received May 1996; revised March 1999.
AMS 1991 *subject classifications*. Primary 62G99; secondary 62E20.
Key words and phrases. Bias, mean square error, parametric family, statistical functionals, target estimates, von Mises expansions.

mean square error (MSE) of the target estimator will also be reduced under less restrictive conditions.

Target estimation is a computer intensive procedure that is aimed at improving and optimizing the finite sample properties of a statistic. We present some examples showing this improvement and give some theoretical results using the second-order von Mises expansion of the corresponding functional. Expansions of this kind, with second-order kernels for the functional derivatives, have been recently used by Gatto and Ronchetti (1996) to obtain saddlepoint approximations.

The paper is organized as follows: in Section 2 we define the target estimator and present some of its basic properties and also consider von Mises expansions to identify and study the bias of a statistic and give sufficient conditions for the MSE reduction based on the von Mises remainder. Targeting M-estimates is treated in Section 3. In Section 4 we present some examples and applications as well as a comparison study of target estimators with the jackknife and the bootstrap estimators.

Throughout this paper we shall assume that T is a statistical functional and the statistic $T(F_n)$ estimates the parameter $T(F_\theta)$, where F_n is the empirical d.f. corresponding to the sample $X_1, \dots, X_n$ of i.i.d. random variables with common d.f. F_θ, with $\theta \in \Theta$, for an open subset of real numbers Θ. When a statistical functional T satisfies $T(F_\theta) = \theta$, the functional is said to be "Fisher consistent."

We shall also assume that the expectation of $T(F_n)$, $g(\theta) = E_\theta(T(F_n))$, exists for all $\theta \in \Theta$, where E_θ indicates the expectation with respect to F_θ. Moreover, the function g will be assumed to be one-to-one and differentiable.

2. Target estimators.

DEFINITION. Let $g(\theta) = E_\theta(T(F_n))$ be a one-to-one function. The functional $\widetilde{T}$ induced by T from the relation

$$g^{-1}(T) = \widetilde{T} \tag{1}$$

will be called the *target functional* of T. The statistic $\widetilde{T}(F_n)$ will be the *target estimator*.

REMARK 1. The target estimate of θ corresponds to choosing the value $\theta = \widetilde{T}(\hat{F}_n)$, which solves the equation

$$E_\theta(T(F_n)) = T\big(\hat{F}_n\big),$$

where $\hat{F}_n$ is the observed value of F_n. That is, we set the expectation of a statistic equal to its observed value and we solve for θ.

REMARK 2. Since g depends on n, the basic defining condition $g(\widetilde{T}) = T$ or equivalently $g^{-1}(T) = \widetilde{T}$ depends on the sample size n. The target estimator of some T will be different for different sample sizes. Although $\widetilde{T} = \widetilde{T}_n$

would be a more precise notation, we shall omit the subscript n for simplicity since the sample size will always be assumed to be fixed.

It should be noted that Rousseeuw and Ronchetti (1981) used the function g in an entirely different context to generate functionals $g^{-1}(T)$ for studying the influence curve of statistics in testing hypotheses.

Let the variance and the bias of a functional T at θ be denoted by V_T and $B_T(\theta)$, respectively. The following lemma is a direct consequence of the above definition.

LEMMA 1. *If T is a statistical functional with $g(\theta) = a\theta + b$ for $a \neq 0$, then the corresponding target estimator will be unbiased. The variance of $\widetilde{T}$ will satisfy*

$$V_{\widetilde{T}} = (1/a^2)V_T \tag{2}$$

and the variance of the target estimator will be reduced if and only if $a^2 > 1$, and it will remain unchanged when $a = 1$.

PROOF. By solving (1) when $g(\theta) = a\theta + b$, we obtain the target functional $\widetilde{T} = (T - b)/a$ which directly implies (2).

For the bias of $\widetilde{T}$, note that $E_\theta(\widetilde{T}(F_n)) = (g(\theta) - b)/a = \theta$ and the corresponding target estimator is unbiased. □

For general estimators, the next two theorems give conditions under which the bias and the mean square error (MSE) are reduced after targeting. In particular, Theorem 2 gives several upper bounds for $|B_{\widetilde{T}}(\theta)|$, the absolute value of the bias of the target estimator. For the study of this bias, the ratio $H^+(\theta) = E_\theta[(T - \theta)I_{\{T \geq \theta\}}]/E_\theta[(\theta - T)I_{\{T < \theta\}}]$ plays an important role and will be needed. Note that when $g(\theta) > \theta$ then $H^+(\theta) > 1$. We shall denote the mean absolute deviation of T by $M = E_\theta|T - g(\theta)| = 2E_\theta(T - g(\theta))^+$.

THEOREM 2. *Suppose T is a statistical functional and the function $g(\theta) = E_\theta(T(F_n))$ is increasing for all θ in Θ. Then the following will all hold*:

(a) *If g is convex then $E_\theta(\widetilde{T}) \leq \theta$; if $-g$ is convex then $\theta \leq E_\theta(\widetilde{T})$.*
(b) *If $0 < a < g'(\theta)$ then $|(B_{\widetilde{T}}(\theta))| < (1/a)E_\theta|T - g(\theta)|$.*
(c) *If $0 < a < g'(\theta) \leq b$ then $|(B_{\widetilde{T}}(\theta))| < (b - a)M/2ab$.*
(d) *If g is convex, $g(\theta) > \theta$, with $1 < g'(\theta) \leq b$ and $H^+(\theta) > b$ then*

$$\left|(B_{\widetilde{T}}(\theta))\right| < \left|(B_T(\theta))\right|.$$

(e) *If $-g$ is convex, $g(\theta) < \theta$, with $1/2 < a \leq g'(\theta) \leq 1$ and $H^+(\theta) \leq 2 - 1/a$, then*

$$\left|(B_{\widetilde{T}}(\theta))\right| < \left|(B_T(\theta))\right|.$$

PROOF. Part (a) follows immediately from Jensen's inequality since

$$\theta = g^{-1}(g(\theta)) = g^{-1}(E_\theta(T)) = g^{-1}\Big(E_\theta\big(g(\widetilde{T})\big)\Big)$$

and this last expression is greater than $E_\theta(\widetilde{T})$ when g is convex, but is less than $E_\theta(\widetilde{T})$ when $-g$ is convex.

For part (b) use the mean value theorem for g^{-1} and note that the bias of the target estimator satisfies

$$\begin{aligned}|B_{\widetilde{T}}(\theta)| &= \left|E_\theta(\theta - \widetilde{T})\right| \\ &= \left|E_\theta\left(g^{-1}(g(\theta)) - g^{-1}(T)\right)\right| \\ &= \left|E_\theta([1/g'(\xi)](g(\theta) - T))\right| \\ &< (1/a)E_\theta|(g(\theta) - T)|\end{aligned}$$

since $1/g'(\xi) < 1/a$.

For part (c), consider again the absolute value of the bias of $\widetilde{T}$,

$$\begin{aligned}|B_{\widetilde{T}}(\theta)| &= \left|E_\theta(\theta - \widetilde{T})\right| \\ &= \left|E_\theta\left(g^{-1}(g(\theta)) - g^{-1}(T)\right)\right| \\ &= \left|E_\theta([1/g'(\xi)](g(\theta) - T))\right| \\ &= E_\theta([1/g'(\xi)](g(\theta) - T))^{+} + E_\theta([1/g'(\xi)](g(\theta) - T))^{-} \\ &< (1/a)E_\theta((g(\theta) - T))^{+} + (1/b)E_\theta((g(\theta) - T))^{-} \\ &= (1/a - 1/b)E_\theta|T - g(\theta)|/2 \\ &= (b - a)M/2ab\end{aligned}$$

since g' is bounded below and above.

For part (d) we have g convex so the absolute value of the bias of $\widetilde{T}$ is

$$\begin{aligned}|B_{\widetilde{T}}(\theta)| &= E_\theta(\theta - \widetilde{T}) \\ &= E_\theta\left(g^{-1}(g(\theta)) - g^{-1}(T)\right) \\ &= E_\theta([1/g'(\xi)](g(\theta) - T)) \\ &= E_\theta([1/g'(\xi)](g(\theta) - \theta)) + E_\theta([1/g'(\xi)](\theta - T)) \\ &\le g(\theta) - \theta + E_\theta\left([1/g'(\xi)](\theta - T)I_{\{T \ge \theta\}}\right) \\ &\qquad + E_\theta\left([1/g'(\xi)](\theta - T)I_{\{T < \theta\}}\right) \\ &\le g(\theta) - \theta + (1/b)E_\theta\left((\theta - T)I_{\{T \ge \theta\}}\right) + E_\theta\left((\theta - T)I_{\{T < \theta\}}\right) \\ &< g(\theta) - \theta \\ &= |B_T(\theta)|\end{aligned}$$

since $1/H^{+} < 1/b$.

Finally, for part (e) we have that the absolute value of the bias of T is $\theta - g(\theta)$, whereas the convexity of $-g$ implies that the absolute value for the bias of $\widetilde{T}$ is

$$\begin{aligned}
|B_{\widetilde{T}}(\theta)| &= E_\theta(\widetilde{T} - \theta) \\
&= E_\theta\big(g^{-1}(T) - g^{-1}(g(\theta))\big) \\
&= E_\theta\big([1/g'(\xi)](T - g(\theta))\big) \\
&= E_\theta\big([1/g'(\xi)](\theta - g(\theta))\big) + E_\theta\big([1/g'(\xi)](T - \theta)\big) \\
&\le \frac{1}{a}(\theta - g(\theta)) + \frac{1}{a}E_\theta(T-\theta)I_{\{T \ge \theta\}} + E_\theta(T-\theta)I_{\{T<\theta\}} \\
&= \theta - g(\theta) + \left(\frac{1}{a} - 1\right)(\theta - g(\theta)) \\
&\quad + \frac{1}{a}E_\theta(T-\theta)I_{\{T \ge \theta\}} + E_\theta(T-\theta)I_{\{T<\theta\}} \\
&= \theta - g(\theta) + \left(1 - \frac{1}{a}\right)E_\theta(T-\theta)I_{\{T \ge \theta\}} + \left(1 - \frac{1}{a}\right)E_\theta(T-\theta)I_{\{T<\theta\}} \\
&\quad + \frac{1}{a}E_\theta(T-\theta)I_{\{T \ge \theta\}} + E_\theta(T-\theta)I_{\{T<\theta\}} \\
&= \theta - g(\theta) + E_\theta(T-\theta)I_{\{T \ge \theta\}} + \left(2 - \frac{1}{a}\right)E_\theta(T-\theta)I_{\{T<\theta\}} \\
&\le \theta - g(\theta) + \left(2 - \frac{1}{a}\right)E_\theta(\theta - T)I_{\{T<\theta\}} + \left(2 - \frac{1}{a}\right)E_\theta(T-\theta)I_{\{T<\theta\}} \\
&= |B_T(\theta)|
\end{aligned}$$

since $H^+ \le 2 - 1/a$. □

Note that the bound in (c) is sharp in a sense that it is reached in the limiting case when $g(\theta) = \theta_0 + a(\theta - \theta_0)$ for $\theta < \theta_0$ and $g(\theta) = \theta_0 + b(\theta - \theta_0)$ for $\theta > \theta_0$.

We should also mention that a crucial hypothesis in this last theorem is the lower bound for g'. In most of the applications that we considered, the fact that either $g'(\theta) > 1$ or $g'(\theta) \le 1$ when $g(\theta) > \theta$ seemed to be a decisive factor in determining whether or not the bias was reduced after targeting. Intuitively, this is not too surprising since $\widetilde{T} = g^{-1}(T)$ and g^{-1} "shrinks" T when $g'(\theta) > 1$. The other regularity conditions of Theorem 2(d) insure that this "shrinking" is not done too abruptly. The example in 4.2 will show that these conditions are sufficient but not necessary.

It is important at this point to weight the gain in reducing the bias with the possible increase in the variance. The next theorem will show that a condition much less restrictive than condition (d) in Theorem 2 will insure smaller mean square error. This suggests that for target estimation the requirements for bias and variance reduction do not necessarily work against each other. However, note that there might be bias reduction without the assumptions of Theorem 2 being satisfied, but there might not be a reduction in the MSE (see Section 4.2 below).

Let the mean square error of the target estimator be $\widetilde{\mathrm{MSE}} = \mathrm{V}_{\widetilde{T}} + (\mathrm{B}_{\widetilde{T}}(\theta))^2$; the following result compares $\widetilde{\mathrm{MSE}}$ with the variance of the original estimator.

THEOREM 3. *If T is a statistical functional and $g(\theta) = E_\theta(T(F_n))$ is differentiable then*:

(a) $|g'(\theta)| \geq 1$ *for all* $\theta \in \Theta$, *implies*

$$\widetilde{MSE} \leq V_T. \tag{3}$$

(b) $|g'(\theta)| \leq 1$ *for all* $\theta \in \Theta$, *implies* $\widetilde{MSE} \geq V_T$.

PROOF. Using the mean value theorem for g we have

$$T - g(\theta) = g(\widetilde{T}) - g(\theta) = (\widetilde{T} - \theta) g'(\xi)$$

with ξ between θ and $\widetilde{T}$. Hence

$$\begin{aligned}\mathrm{Var}(T) &= E_\theta(T - g(\theta))^2 \\ &= E_\theta\Big[(\widetilde{T} - \theta)^2 (g'(\xi))^2\Big]\end{aligned}$$

and so (3) holds when $|g'(\theta)| \geq 1$, and when $|g'(\theta)| \leq 1$ the MSE of the target estimator will be larger than or equal to the variance of T. $\square$

This result asserts that the mean square error of an estimator can always be reduced by targeting if the corresponding function g satisfies $|g'(\theta)| \geq 1$ for all θ in Θ. Note that when $g'(\theta) = 1$ for all $\theta \in \Theta$ then $g(\theta) = \theta + b$, and so the bias can always be removed without changing the variance.

2.1. *Von Mises expansions*. Consider a parametric family F_θ and a Fisher consistent statistical functional T. Let T have influence function ϕ_1 [see Hampel (1974) or Hampel, Ronchetti, Rousseeuw and Stahel (1986)]. For a sample $X_1, \dots, X_n$ from F_θ, the first-order von Mises expansion of $T(F_n)$ is

$$T(F_n) = \theta + \frac{1}{n} \sum_i \phi_1(X_i) + \mathrm{Rem}_1 \tag{4}$$

With Hadamard or Frechet differentiability, under certain regularity conditions, we have that $\mathrm{Rem}_1 = o_P(n^{-1/2})$ [see Reeds (1976), Fernholz (1983)].

Taking expected values in (4) we obtain

$$g(\theta) = E_\theta(T(F_n)) = \theta + E_\theta(\mathrm{Rem}_1), \tag{5}$$

so the target functional $\widetilde{T}$ is defined by

$$T = \widetilde{T} + E_{\widetilde{T}}(\mathrm{Rem}_1).$$

Note that the bias of T is

$$B_T(\theta) = E_\theta(\mathrm{Rem}_1),$$

whereas the bias of the target estimator is

$$B_{\widetilde{T}}(\theta) = E_\theta(\mathrm{Rem}_1 - E_{\widetilde{T}}(\mathrm{Rem}_1)),$$

hence Theorem 4 implies that under certain conditions,

$$|E_\theta(\mathrm{Rem}_1 - E_{\widetilde{T}}(\mathrm{Rem}_1))| < |E_\theta(\mathrm{Rem}_1)|.$$

In the following lemma, the bias of the target estimator is expressed in terms of the influence function.

LEMMA 4. *For a statistical functional T with von Mises expansion as in* (4), *the bias of the target estimator is*

$$B_{\widetilde{T}}(\theta) = E_\theta E_{\widetilde{T}}(\phi_1(X_1)).$$

PROOF. We have

$$\begin{aligned}\widetilde{T} &= T - E_{\widetilde{T}}(\mathrm{Rem}_1)\\ &= T - E_{\widetilde{T}}\left(T(F_n) - \theta - \frac{1}{N}\sum_{i=1}^{n}\phi_1(X_i)\right)\\ &= \theta + E_{\widetilde{T}}(\phi_1(X_1))\end{aligned}$$

since $E_{\widetilde{T}}(T(F_n)) = T$. Hence the bias of $\widetilde{T}$ is

$$B_{\widetilde{T}}(\theta) = E_\theta E_{\widetilde{T}}(\phi_1(X_1))$$

and the lemma follows. □

An immediate consequence of this lemma is that

$$|E_\theta E_{\widetilde{T}}(\phi_1(X_1))| < |E_\theta(\mathrm{Rem}_1)|$$

under the conditions of Theorem 2.

Suppose again that T has a first-order von Mises expansion at F_θ. Then taking derivatives in (5) we obtain

$$g'(\theta) = 1 + \frac{\partial}{\partial\theta}E_\theta(\mathrm{Rem}_1),$$

so Theorem 3 implies that when

$$\frac{\partial}{\partial\theta}E_\theta(\mathrm{Rem}_1) > 0, \tag{6}$$

the MSE of the target estimator will be reduced.

In order to study this last inequality, we look at the remainder term and expand it further. For $k \geq 2$, the kth-order von Mises expansion of $T(F_n)$ at

F_θ is

$$T(F_n) - T(F) = \frac{1}{n}\sum_i \phi_1(X_i) + \frac{1}{2n^2}\sum_{i,j}\phi_2(X_i, X_j) + \cdots \frac{1}{k!n^k}\sum_{i_1,\dots,i_k}\phi_k(X_{i_1},\dots,X_{i_k}) + \mathrm{Rem}_k, \tag{7}$$

where the kernels $\phi_1, \phi_2, \dots, \phi_k$, have been properly normalized so that $\int\phi_1(x)\,dF(x) = 0$, $\int\phi_2(x, y)\,dF(x) = \int\phi_2(y, x)\,dF(x) = 0$ and so on. Note that the remainders Rem_1 and Rem_k satisfy

$$\mathrm{Rem}_1 = \frac{1}{2n^2}\sum_{i,j}\phi_2(X_i, X_j) + \cdots + \frac{1}{k!n^k}\sum\phi_k(X_{i_1},\dots,X_{i_k}) + \mathrm{Rem}_k$$

and under certain conditions the remainder of order k satisfies $\mathrm{Rem}_k = o_P(n^{-k/2})$ [see von Mises (1947), Fillipova (1962), Reeds (1976)].

The next two theorems give conditions on the von Mises kernels for bias and variance reduction. When $k = 2$, these conditions can be quite useful in practice since the second-order von Mises expansion gives a reasonable approximation of the statistic and the computations of the first and second kernels are not as intractable as those of higher order. See Fernholz (1996), Fankhauser (1996).

THEOREM 5. *When T has a von Mises expansion as in* (7) *for $k = 2$, and $E_\theta(\mathrm{Rem}_2) = o(n^{-1})$, then for a large sample, the variance of the target estimator will be reduced if*

$$\frac{\partial}{\partial\theta}\left(\frac{1}{2n}\right)E_\theta(\phi_2(X, X)) > 0.$$

PROOF. Suppose that T has a second-order von Mises expansion and consider the first remainder

$$\mathrm{Rem}_1 = \frac{1}{2n^2}\sum_{i,j}\phi_2(X_i, X_j) + \mathrm{Rem}_2,$$

whose expectation is the bias of $T(F_n)$.

Since

$$E\left(\frac{1}{2n^2}\sum_{i,j}\phi_2(X_i, X_j)\right) = \frac{1}{2n}E_\theta(\phi_2(X, X)),$$

the bias of $T(F_n)$ is

$$E_\theta(\mathrm{Rem}_1) = \frac{1}{2n}E_\theta(\phi_2(X, X)) + E_\theta(\mathrm{Rem}_2).$$

Taking derivatives with respect to θ, we obtain

$$\frac{\partial}{\partial\theta}E_\theta(\mathrm{Rem}_1) = \frac{1}{2n}\frac{\partial}{\partial\theta}E_\theta(\phi_2(X,X)) + \frac{\partial}{\partial\theta}E_\theta(\mathrm{Rem}_2).$$

Since by hypothesis $E_\theta(\mathrm{Rem}_2) = o(1/n)$, and $(\partial/\partial\theta)E_\theta(\phi_2(X,X)) > 0$, there is a positive constant C such that

$$\frac{\partial}{\partial\theta}E_\theta(\mathrm{Rem}_1) > C/n + o(1/n).$$

So, for large n, condition (6) holds and the theorem follows. □

THEOREM 6. *Suppose that T has a von Mises expansion of order k for $k \ge 2$ with $E_\theta(\mathrm{Rem}_k) = o(n^{-k/2})$. If*

$$\frac{\partial}{\partial\theta}E_\theta(\phi_2(X,X)) = 0,\dots,\frac{\partial}{\partial\theta}E_\theta(\phi_k(X,\dots,X)) = 0, \tag{8}$$

then

$$B_{\widetilde{T}}(\theta) = o(n^{-k/2}) \quad \text{and} \quad V_T = V_{\widetilde{T}} + o(n^{-k/2}). \tag{9}$$

PROOF. Taking expectations on the von Mises expansion (7) we have

$$E_\theta\left(\frac{1}{2n^2}\sum_{i,j}\phi_2(X_i,X_j)\right) = \frac{1}{2n}E_\theta(\phi_2(X,X)),$$

$$E_\theta\left(\frac{1}{k!n^k}\sum_{i_1,\dots,i_k}\phi_k(X_{i_1},\dots,X_{i_k})\right) = \frac{1}{k!n^{k-1}}E_\theta(\phi_k(X,\dots,X))$$

and differentiating with respect to θ we obtain

$$g(\theta) = \theta + A + o(n^{-k/2}),$$

where A does not depend on θ due to condition (8).

By solving $g(\widetilde{T}) = T$ we have

$$T = \widetilde{T} + A + o(n^{-k/2}),$$

and the corresponding variances satisfy (9), whereas taking expectations on this last equality, we obtain

$$E_\theta(\widetilde{T}) = E_\theta(T) - A + o(n^{-k/2}) = \theta + o(n^{-k/2}),$$

and the theorem is proved. □

3. Targeting *M*-estimates.

3.1. *M-estimates of location.* For a parametric family F_θ, let the functional $T(F_\theta) = \theta$ be defined implicitly by a solution of

$$\int\psi(x-\theta)\,dF_\theta(x) = 0.$$

The corresponding statistic $T(F_n)$ is the well-known M-estimate of location [see Huber (1981) or Hampel, Ronchetti, Rousseeuw and Stahel (1986)]. When the parametric family satisfies $F_\theta(x) = F(x-\theta)$ for all θ and some d.f. F, the corresponding M-estimate is location equivariant, that is $T(F_\theta) = \theta + T(F)$.

The bias of an M-estimate of location will be constant, and an approximation of this bias can be obtained by using second-order von Mises expansions as is done in the following.

LEMMA 7. *Let $T(F_n)$ be an M-estimate of location for the location family F_θ. Assume that T has a von Mises expansion of order two as in* (7) *with $E_\theta(\mathrm{Rem}_2) = o(1/n)$, then the bias of T is constant and given by*

$$E_\theta(\mathrm{Rem}_1) = \frac{K}{2nM^3}\int \psi^2(x)\,dF(x) - \frac{1}{nM^2}\int \psi(x)\psi'(x)\,dF(x) + o(1/n),$$

with $M = \int\psi'(x)\,dF(x)$ and $K = \int\psi''(x)\,dF(x)$. The corresponding target estimate will be unbiased with the same variance as T.

PROOF. If F_n is the empirical d.f. of the sample $X_1,\dots,X_n$, let G_n be the empirical d.f. of the sample $Y_i = X_i - \theta$ for $i = 1,\dots,n$. Since this M-estimate is location equivariant, it satisfies $T(F_n) = \theta + T(G_n)$, and taking expectations we obtain

$$\begin{aligned} g(\theta) &= E_\theta(T(F_n)) \\ &= \theta + E_\theta(T(G_n)) \\ &= \theta + B \end{aligned}$$

with $B = E_\theta(T(G_n))$ a constant bias. Hence, g is linear and by Lemma 1, the target estimator will be unbiased with the same variance as T.

An approximation of this constant bias can be obtained using the second-order von Mises expansion. Recall that the first kernel of T, the influence function, is given by

$$\phi_1(x) = \frac{\psi(x-\theta)}{\int\psi'(x-\theta)\,dF_\theta(x)}$$

The second kernel is

$$\begin{aligned} \phi_2(x,y) = \phi_1(x) + \phi_1(y) & \\ + \frac{1}{M}\Big\{ &\phi_1(x)\phi_1(y)\int \psi''(t-\theta)\,dF(t-\theta) \\ & - \phi_1(x)\psi'(y-\theta) - \phi_1(y)\psi'(x-\theta)\Big\} \end{aligned}$$

with M as above [see Fernholz (1996), Gatto and Ronchetti (1996), Fankhauser (1996)].

Now, if a second-order von Mises expansion as in (7) holds, the bias of an M-estimate of location is

$$\begin{aligned}
E_\theta(\mathrm{Rem}_1) &= \frac{1}{2n} E_\theta(\phi_2(X, X)) + o(1/n) \\
&= \frac{K}{2nM} \int \phi_1^2(x)\, dF(x-\theta) \\
&\quad - \frac{1}{nM} \int \phi_1(x)\psi'(x-\theta)\, dF(x-\theta) + E_\theta(\mathrm{Rem}_2) \\
&= \frac{K}{2nM} \int \phi_1^2(x)\, dF(x-\theta) \\
&\quad - \frac{1}{nM^2} \int \psi(x-\theta)\psi'(x-\theta)\, dF(x-\theta) + E_\theta(\mathrm{Rem}_2) \\
&= \frac{K}{2nM^3} \int \psi^2(x)\, dF(x) \\
&\quad - \frac{1}{nM^2} \int \psi(x)\psi'(x)\, dF(x) + E_\theta(\mathrm{Rem}_2) \\
&= \frac{K}{2nM^3} \int \psi^2(x)\, dF(x) - \frac{1}{nM^2} \int \psi(x)\psi'(x)\, dF(x) + o(1/n),
\end{aligned}$$

where K and M are as above, and the lemma is proved. □

Consider now the particular case of the Huber estimator with

$$\psi(x) = xI_{\{|x| \le b\}} + b\frac{x}{|x|} I_{\{|x| > b\}}.$$

For this case we have

$$M = \int \psi'(x-\theta)\, dF(x-\theta) = F(b) - F(-b)$$

and

$$K = \int \psi''(x-\theta)\, dF(x-\theta) = F'(-b) - F'(b).$$

Therefore, the first and second kernels of the Huber estimate are given by

$$\phi_1(x) = \frac{\psi(x-\theta)}{F(b) - F(-b)}$$

and

$$\begin{aligned}
\phi_2(x, y) &= \phi_1(x) + \phi_1(y) + \phi_1(x)\phi_1(y) \frac{[F'(-b) - F'(b)]}{F(b) - F(-b)} \\
&\quad - \frac{1}{F(b) - F(-b)} [\phi_1(x)\psi'(y-\theta) + \phi_1(y)\psi'(x-\theta)].
\end{aligned}$$

Hence, the bias of the Huber estimate is given by

$$\begin{aligned} E_\theta(\mathrm{Rem}_1) &= \frac{1}{2n} E_\theta(\phi_2(X, X)) + E_\theta(\mathrm{Rem}_2) \\ &= \frac{1}{2n} E_\theta(\phi_1(X))^2 \frac{[F'(-b) - F'(b)]}{[F(-b) - F(b)]} \\ &\quad - \frac{1}{nM^2} \int_{|x|<b} x\, dF(x) + E_\theta(\mathrm{Rem}_2). \end{aligned}$$

Note that the bias of the Huber estimator is $E_\theta(\mathrm{Rem}_1) = E_\theta(\mathrm{Rem}_2)$ when F satisfies $F'(-b) = F'(b)$ and $\int_{|x|<b} x\, dF(x) = 0$.

3.2. *Simultaneous M-estimates of location and scale.* Consider a family of d.f.'s F_θ with $\theta = (\mu, \sigma)$ and the two-dimensional functional $T(F_\theta) = (T_1(F_\theta), T_2(F_\theta))$ defined implicitly by

$$\int \psi\left(\frac{x - T_1(F_\theta)}{T_2(F_\theta)}\right) dF_\theta(x) = 0,$$

where $\psi = (\psi_1, \psi_2)$. The corresponding statistic $T(F_n) = (T_1(F_n), T_2(F_n))$ satisfies the system of equations

$$\sum_{i=1}^{n} \psi_1\left(\frac{X_i - T_1(F_n)}{T_2(F_n)}\right) = 0,$$

$$\sum_{i=1}^{n} \psi_2\left(\frac{X_i - T_1(F_n)}{T_2(F_n)}\right) = 0$$

and is called an M-estimate of location and scale. See Huber (1981) and Hampel, Ronchetti, Rousseeuw and Stahel (1986).

When the family of distributions is such that $F_\theta(x) = F((x - \mu)/\sigma)$ for some fixed d.f. F, the functional T satisfies $T(F_{\mu,\sigma}) = (\mu + \sigma T_1(F), \sigma T_2(F))$ and is said to be location-scale equivariant [see Hampel, Ronchetti, Rousseeuw and Stahel (1986)].

LEMMA 8. *Let $T = (T_1, T_2)$ be the simultaneous M-estimates of location and scale for the family $F_{\mu,\sigma}$ as above. Then*

$$E_{\mu,\sigma}(T_1(F_n), T_2(F_n)) = (\mu + \sigma C_1, \sigma C_2)$$

with C_1 and C_2 constants independent of μ and σ. The corresponding target estimators will be unbiased and their variances will satisfy

$$V_{\tilde{T}_1} = V_{T_1} + (C_1^2) V_{\tilde{T}_2} - 2\sigma_{12}(C_1),$$

$$V_{\tilde{T}_2} = (1/C_2^2) V_{T_2},$$

where σ_{12} is the covariance of (T_1, T_2) and C_1, C_2, are constants.

PROOF. Since T is location-scale equivariant,

$$T(F_n) = (\mu + \sigma T_1(G_n), \sigma T_2(G_n))$$

where F_n is the empirical d.f. of $X_1, \dots, X_n$ and G_n is the empirical d.f. of $Y_i = (X_i - \mu)/\sigma$, for $i = 1, \dots, n$. When we take the expectation of $T(F_n)$, we obtain

$$\begin{aligned} E_{\mu,\sigma}(T_1(F_n), T_2(F_n)) &= g(\mu, \sigma) \\ &= (\mu + \sigma E(T_1(G_n)), \sigma E(T_2(G_n))) \\ &= (\mu + \sigma C_1, \sigma C_2) \end{aligned}$$

with $C_i = E(T_i(G_n))$ constant for $i = 1, 2$.

Now, $\tilde{T} = (\tilde{T}_1, \tilde{T}_2)$ is obtained by solving

$$g\big((\tilde{T}_1, \tilde{T}_2)\big) = (T_1, T_2),$$

which gives

$$\begin{aligned} \tilde{T}_1 + \tilde{T}_2 C_1 &= T_1, \\ \tilde{T}_2 C_2 &= T_2. \end{aligned}$$

Hence, the target functionals are

$$\begin{aligned} \tilde{T}_1 &= T_1 - (C_1/C_2)T_2, \\ \tilde{T}_2 &= (1/C_2)T_2 \end{aligned}$$

with variances satisfying

$$\begin{aligned} V_{\tilde{T}_1} &= V_{T_1} + (C_1/C_2)^2 V_{T_2} - 2C_1\sigma_{12}, \\ V_{\tilde{T}_2} &= (1/C_2)^2 V_{T_2}, \end{aligned}$$

where σ_{12} is the covariance of (T_1, T_2), and C_i, $i = 1, 2$, are constant.

Since

$$E_{\mu,\sigma}\big(\tilde{T}_2(F_n)\big) = (1/C_2)\sigma C_2 = \sigma,$$

and

$$\begin{aligned} E_{\mu,\sigma}\big(\tilde{T}_1(F_n)\big) &= E(T_1(F_n)) - C_1 E\big(\tilde{T}_2(F_n)\big) \\ &= \mu + \sigma C_1 - C_1\sigma \\ &= \mu, \end{aligned}$$

the target estimator $\tilde{T}(F_n)$ is unbiased and the lemma is proved. □

4. Some applications and examples.

4.1. *The sample variance.* Let F be a d.f. with finite second moment. Consider the statistical functional corresponding to its variance σ^2, defined by

$$T(F) = \int (x - \tau(F))^2 \, dF(x),$$

where $\tau(F) = \mu$ is the mean of F. The corresponding estimator,

$$T(F_n) = \sum \frac{(X_i - \overline{X})^2}{n},$$

is well known to be biased.

Taking expectations, we obtain $E_\sigma(T(F_n)) = g(\sigma^2) = \sigma^2(1 - 1/n)$, a linear function of σ^2. So the target estimator is

$$\widetilde{T}(F_n) = \frac{\Sigma(X_i - \overline{X})^2}{n - 1},$$

which is the usual unbiased estimator of σ^2. In this case the variance of the target estimator will not be reduced since $g'(\sigma^2) = 1 - 1/n < 1$.

4.2. *The distance between two means.* Suppose that we have two independent populations F and G with means μ_1 and μ_2, respectively. The parameter to be estimated is $\theta = |\mu_1 - \mu_2|$, which corresponds to the statistical functional

$$T(F,G) = |\tau(F) - \tau(G)|,$$

where τ is the linear functional corresponding to the sample mean $\tau(F) = \int x\,dF(x)$. Given two independent samples with corresponding empirical d.f.'s F_n and G_m, we obtain the statistic that estimates θ,

$$|\tau(F_n) - \tau(G_m)| = |\bar{x} - \bar{y}|.$$

Using the derivatives of a functional of two variables [see Fernholz (1996)] we obtain the von Mises expansion of T at (F,G),

$$|\bar{x} - \bar{y}| = |\mu_1 - \mu_2| + \frac{(\mu_1 - \mu_2)}{|\mu_1 - \mu_2|}(\bar{x} - \mu_1 - \bar{y} + \mu_2) + \mathrm{Rem}_1,$$

where the remainder term is

$$\mathrm{Rem}_1 = \begin{cases} 2(\bar{y} - \bar{x}) I_{[\bar{y} \geq \bar{x}]}, & \text{if } \mu_1 > \mu_2, \\ 2(\bar{x} - \bar{y}) I_{[\bar{x} \geq \bar{y}]}, & \text{if } \mu_2 > \mu_1. \end{cases} \tag{10}$$

In this case the higher order derivatives for T are all zero so that $\mathrm{Rem}_1 = \mathrm{Rem}_k$ for any $k \geq 1$. Note however that the remainder is not zero but is of a smaller order than $1/n^k$ for all $k \geq 1$.

Taking expectations in (10), we obtain

$$\begin{aligned} g(|\mu_1 - \mu_2|) &= E_{|\mu_1 - \mu_2|}(\bar{x} - \bar{y}) \\ &= |\mu_1 - \mu_2| + 2E\{(\bar{y} - \bar{x}) I_{[\bar{y} - \bar{x}]}\} \quad \text{if } \mu_1 > \mu_2 \end{aligned}$$

analogously for the case when $\mu_2 > \mu_1$. Note that when $|\mu_1 - \mu_2| \to \infty$ we have $g(|\mu_1 - \mu_2|) \to |\mu_1 - \mu_2|$.

If the two populations are normal with variances σ_1^2 and σ_2^2, respectively, then putting $\theta = |\mu_1 - \mu_2|$ and $\sigma^2 = \sigma_1^2/n + \sigma_2^2/m$, we have

$$g(\theta) = \theta + (\Phi_-(0) - \Phi_+(0)) + K,$$

where K is given by

$$K = c\sigma \exp(-\theta^2/\sigma^2)$$

with c a positive constant, and where Φ_- and Φ_+ are the normal d.f.'s centered at $-|\mu_1 - \mu_2|$ and $|\mu_1 - \mu_2|$, respectively.

When $\theta \to 0$, then $g(\theta) \to c\sigma > 0$, whereas $g(\theta) - \theta \to 0$ as $\theta \to 0$. For this example, $g'(\theta) < 1$ and the variance of the target estimator will not decrease. However, the bias is reduced, and so this example shows that the conditions of Theorem 2 for bias reduction are sufficient but not necessary. This example also shows that there can be bias reduction without MSE reduction. Figure 1 gives the graph of g for normal populations, when 200 simulations of size $n = 20$ were performed. Figure 2 has the boxplots of both estimators when $|\mu_1 - \mu_2| = 5$.

4.3. *M-estimates for the lognormal distribution.* In this example, we simulated data from the lognormal distribution

$$Y = e^X \quad \text{with } X \sim N(\mu, \sigma^2).$$

The statistical functional was T_H, the Huber estimator "proposal-2" [see Huber (1981)]. If $F_{\mu,\sigma}$ is the d.f. of Y, the parameters

$$(\theta_1, \theta_2) = T_H(F_{\mu,\sigma}),$$

are estimated by the corresponding statistic $T_H(F_n) = (T_1(F_n), T_2(F_n))$ We shall call θ_1 the location parameter and θ_2 the scale parameter. Note that $F_{\mu,\sigma}$ is not a location-scale family of distributions and so the Huber estimator is not location-scale equivariant for this family. Consequently, Lemma 8 cannot be applied here.

Tables 1 and 2 show the values of θ_i and $E(T_i(F_n))$, $i = 1, 2$, for location and scale, respectively, for the grid of values $\mu = \pm 2, \pm 1.56, \pm 1.11, \pm 0.67,$

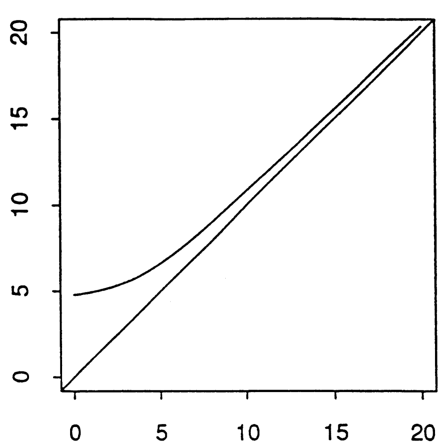

FIG. 1. *Graph of the function $g(\theta)$.*

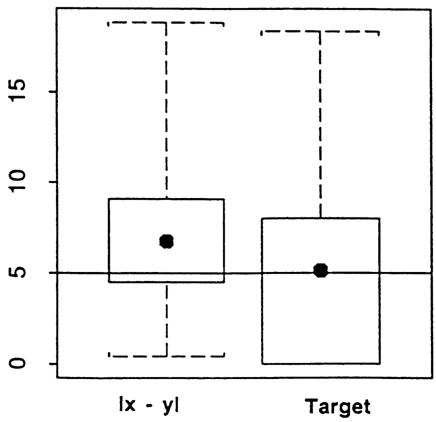

FIG. 2. *Boxplots of the estimator and the corresponding target.*

TABLE 1
The Huber estimator for the lognormal (location). First line: Values of θ_1. Second line: Values of $E(T_1(F_n))$

	Sigma									
Mu	**0.50**	**0.89**	**1.28**	**1.67**	**2.06**	**2.44**	**2.83**	**3.22**	**3.61**	**4.00**
− 2.00	0.14	0.16	0.19	0.23	0.29	0.36	0.45	0.56	0.70	0.90
	0.15	0.18	0.22	0.29	0.41	0.66	1.07	1.23	2.30	4.46
− 1.56	0.23	0.26	0.30	0.37	0.45	0.56	0.69	0.89	1.14	1.43
	0.23	0.26	0.34	0.46	0.63	0.85	1.26	2.02	4.18	7.19
− 1.11	0.35	0.40	0.47	0.57	0.70	0.88	1.10	1.38	1.73	2.22
	0.35	0.42	0.53	0.71	1.01	1.46	2.09	3.61	6.59	11.14
− 0.67	0.55	0.62	0.73	0.89	1.10	1.37	1.72	2.15	2.68	3.37
	0.56	0.66	0.85	1.12	1.60	2.31	3.14	5.41	9.56	11.02
− 0.22	0.86	0.97	1.14	1.39	1.71	2.13	2.64	3.31	4.21	5.36
	0.86	1.02	1.27	1.74	2.36	3.45	5.14	7.76	14.53	22.13
0.22	1.33	1.51	1.78	2.15	2.64	3.27	4.12	5.20	6.68	8.37
	1.40	1.62	1.99	2.83	3.68	5.00	8.82	10.79	24.65	30.40
0.67	2.08	2.36	2.77	3.35	4.13	5.07	6.44	8.09	10.30	13.12
	2.14	2.58	3.17	4.41	5.94	8.67	11.34	20.88	33.06	55.63
1.11	3.24	3.68	4.32	5.24	6.48	8.05	10.09	12.76	16.10	20.31
	3.32	3.85	5.07	6.68	8.51	13.85	22.96	30.13	64.59	64.50
1.56	5.06	5.75	6.75	8.14	10.06	12.44	15.66	19.80	25.00	32.07
	5.13	6.07	7.58	10.57	14.57	21.50	31.80	58.80	83.66	115.29
2.00	7.91	9.01	10.60	12.80	15.74	19.61	24.45	30.75	38.74	49.54
	8.07	9.70	12.15	16.70	23.06	33.44	48.73	61.75	126.91	160.45

TABLE 2
The Huber estimator for the lognormal (scale). First line. Values of θ_2. Second line: Values of $E(T_2(F_n))$

	Sigma									
Mu	**0.50**	**0.89**	**1.28**	**1.67**	**2.06**	**2.44**	**2.83**	**3.22**	**3.61**	**4.00**
− 2.00	0.07	0.12	0.18	0.25	0.35	0.47	0.62	0.80	1.05	1.37
	0.07	0.14	0.22	0.34	0.53	0.93	1.63	1.95	3.77	7.83
-1.56	0.10	0.19	0.28	0.40	0.55	0.73	0.95	1.28	1.69	2.18
	0.11	0.20	0.34	0.54	0.83	1.19	1.92	3.15	7.11	13.27
− 1.11	0.16	0.29	0.44	0.62	0.85	1.15	1.51	1.98	2.57	3.39
	0.17	0.32	0.54	0.83	1.34	2.06	3.23	5.78	11.12	19.64
− 0.67	0.26	0.46	0.69	0.98	1.33	1.79	2.37	3.09	3.97	5.13
	0.26	0.52	0.87	1.33	2.10	3.24	4.70	8.42	15.20	18.12
− 0.22	0.40	0.72	1.08	1.52	2.06	2.77	3.63	4.74	6.25	8.17
	0.40	0.79	1.26	2.07	3.11	4.89	7.69	12.12	23.61	36.42
0.22	0.62	1.11	1.67	2.35	3.19	4.25	5.66	7.46	9.91	12.74
	0.68	1.27	2.00	3.47	4.82	6.91	13.29	16.94	43.15	49.01
0.67	0.97	1.74	2.60	3.65	5.00	6.60	8.85	11.60	15.27	19.96
	1.03	2.02	3.16	5.25	7.73	12.31	16.88	32.83	53.75	95.10
1.11	1.51	2.71	4.05	5.72	7.83	10.47	13.87	18.29	23.85	30.92
	1.59	2.96	5.14	7.86	11.27	19.61	35.03	46.95	114.80	105.68
1.56	2.36	4.23	6.33	8.88	12.18	16.18	21.52	28.42	37.11	48.85
	2.43	4.68	7.50	12.44	19.21	30.52	47.95	93.06	137.37	192.52
2.00	3.69	6.65	9.97	13.99	19.06	25.58	33.62	44.05	57.38	75.32
	3.93	7.66	12.02	19.72	30.34	47.77	73.08	95.96	205.74	257.87

± 0.22 and $\sigma = 0.5, 0.89, 1.28, 1.67, 2.06, 2.44, 2.83, 3.22, 3.61, 4$, used to generate the models. Graphs for each component of the function g are shown in Figure 3 [$E(T_1(F_n))$, for location] and Figure 4 [$E(T_2(F_n))$, for scale]. Figures 5 and 6 show the boxplots of the original Huber estimator and the corresponding target estimator, for location with $\theta_1 = 0.95$, and for scale with $\theta_2 = 1.04$.

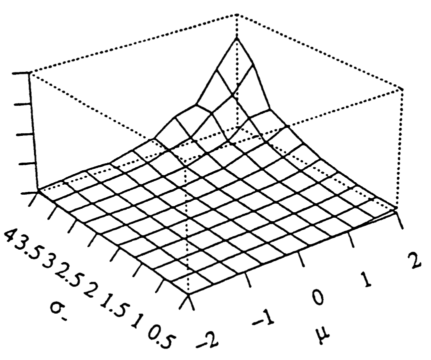

FIG. 3. *Location. Huber estimator for the lognormal. Graph of the expectation $E(T_1(F_n))$, $n = 10$.*

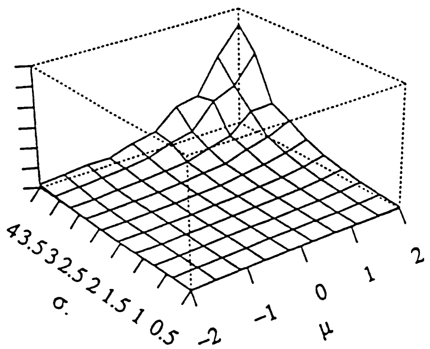

FIG. 4. *Scale. Huber estimator for the lognormal. Graph of the expectation* $E(T_2(F_n))$, $n = 10$.

Huber estimates can be heavily biased when the model is not symmetric as in this case. This example illustrates that targeting Huber estimators is quite effective in reducing both the bias and the variance. Moreover, the robust properties of the Huber estimator seem to be preserved after targeting.

4.4. *An autoregressive model.* In this example we considered an autoregressive model AR(1) of the form

$$X_{t+1} = \theta X_t + \varepsilon_t,$$

where the error term ε_t is Gaussian with $\mu = 0$ and $\sigma = 1$. We want to estimate the parameter θ. Simulations were performed for samples of size 10, 20, 50 and 100 in each one of the five cases when $\theta = 0.5, 0.6, 0.7, 0.8, 0.9$. Table 3 shows the mean square error (MSE) and the bias for each simulation and for the maximum likelihood estimator as compared to the target estimator. These simulations show the dramatic reduction in the bias of the target

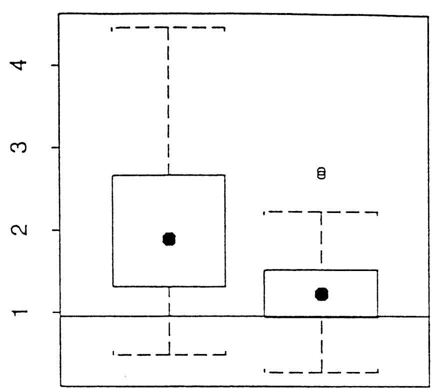

FIG. 5. *Boxplots of Huber and Target Huber for location.*

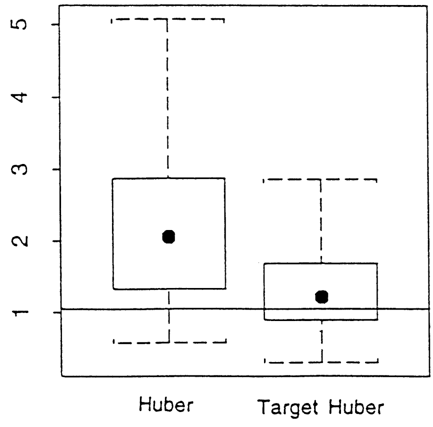

FIG. 6. *Boxplots of Huber and Target Huber for scale.*

estimator as compared to the bias of the MLE in every case. Note also that, except for the two cases when $\theta = 0.5$ and $n = 10, 20$, the MSE is always lower for the target estimator.

4.5. *Applications to semiparametric models.* Consider the usual linear regression model

$$Y = X\theta + \varepsilon$$

when the distribution of the error vector ε is some unknown d.f. G, independent of θ. When the estimated vector $\hat{\theta}$ is not the least squares estimator, $\hat{\theta}$ may have a bias which will depend on the error distribution G. Target estimation to reduce this bias can still be implemented by using the defining condition (1) with the function

$$g(\theta) = E_{\theta, G_n}(\hat{\theta}),$$

with G_n the empirical d.f. of the residuals or any other nonparametric estimator of G.

Median bias reduction using target estimates in the setting of semiparametric models was used for ellipse estimation problems in computer vision [see Cabrera and Meer (1996)] but no theoretical issues were addressed there.

Target estimation can also be extended to nonlinear regression and to the general errors-in-variables model

$$X = U + \varepsilon, \qquad f(U, \theta) = 0$$

by using the expectation function g as above. In these situations, since the d.f. G must be estimated, there will be an additional error on the target

TABLE 3
Estimation of the parameter of an autoregressive model AR(1). *MSE and bias of two estimators*: *MLE and Target*

MSE			Bias		
$n = 10$					
Theta	**MLE**	**Target**	**Theta**	**MLE**	**Target**
0.5	0.1635	0.1789	0.5	−0.288	0.008
0.6	0.1931	0.1703	0.6	−0.329	−0.009
0.7	0.2279	0.1609	0.7	−0.392	−0.050
0.8	0.2673	0.1517	0.8	−0.431	−0.078
0.9	0.3096	0.1431	0.9	−0.506	−0.149
$n = 20$					
Theta	**MLE**	**Target**	**Theta**	**MLE**	**Target**
0.5	0.0643	0.0671	0.5	−0.160	−0.001
0.6	0.0717	0.0641	0.6	−0.175	0.012
0.7	0.0805	0.0595	0.7	−0.203	0.004
0.8	0.0907	0.0536	0.8	−0.213	0.014
0.9	0.1019	0.0470	0.9	−0.267	−0.044
$n = 50$					
Theta	**MLE**	**Target**	**Theta**	**MLE**	**Target**
0.5	0.0197	0.0193	0.5	−0.059	0.000
0.6	0.0196	0.0177	0.6	−0.068	0.001
0.7	0.0197	0.0158	0.7	−0.080	0.001
0.8	0.0200	0.0136	0.8	−0.093	0.001
0.9	0.0203	0.0112	0.9	−0.103	0.001
$n = 100$					
Theta	**MLE**	**Target**	**Theta**	**MLE**	**Target**
0.5	0.0087	0.0085	0.5	−0.026	0.004
0.6	0.0081	0.0075	0.6	−0.039	−0.005
0.7	0.0075	0.0064	0.7	−0.040	0.000
0.8	0.0068	0.0052	0.8	−0.042	0.005
0.9	0.0061	0.0038	0.9	−0.048	0.005

estimator of θ, but our initial results suggest a definite gain in spite of this new error.

In particular, we present here an example of the classical errors-in-variables model where the observable variables are

$$X = U + \delta$$

and the model is

$$Y = a + bU + \varepsilon,$$

where ε and δ are independent Normal errors with mean $\mu = 0$ and standard deviation σ.

Simulations for $a = 0$, $b = 1$ and $b = 10$, were performed for sample sizes 10, 20 and 50. For these simulations, the U's were chosen to be the n

TABLE 4

The errors-in-variables model. The MSE and the bias for two estimators: MLE and Target

		MSE		Bias	
n = 10					
s	***b***	**MLE**	**Target**	**MLE**	**Target**
0.5	1	1.007	0.890	0.044	0.008
0.5	10	4.726	4.238	0.402	−0.104
0.4	1	0.053	0.049	0.024	−0.001
0.4	10	2.596	0.244	0.218	−0.038
0.3	1	0.027	0.025	0.012	−0.000
0.3	10	1.321	1.281	0.119	−0.000
n = 20					
s	***b***	**MLE**	**Target**	**MLE**	**Target**
0.5	1	0.036	0.034	0.018	0.000
0.5	10	1.687	1.607	0.163	−0.021
0.4	1	0.021	0.020	0.012	0.000
0.4	10	1.021	0.996	0.097	−0.009
0.3	1	0.011	0.010	0.006	0.000
0.3	10	0.558	0.547	0.055	0.005
n = 50					
s	***b***	**MLE**	**Target**	**MLE**	**Target**
0.7	1	0.029	0.028	0.013	0.000
0.7	10	1.186	1.143	0.111	−0.024
0.5	1	0.0127	0.0124	0.007	0.000
0.5	10	0.571	0.561	0.051	−0.010
0.4	1	0.0076	0.0757	0.004	0.000
0.4	10	0.359	0.353	0.035	−0.006

quantiles of a standard normal distribution. The errors were added with $\sigma = 0.3, 0.4, 0.5$. The results are given in Table 4, where we can see that in all cases the bias of the target estimator has been substantially reduced when compared to the bias of the MLE. Moreover, note that in all cases the MSE of the target estimator is smaller than that of the MLE.

4.6. *The logistic regression model.* The logistic regression model gives the probability that a binary response Y takes the value one, as a function of X. The formula is

$$\operatorname{logit}(P(Y = 1)) = \alpha + \beta X.$$

The parameter of interest here is the slope β, which is initially estimated by maximum likelihood. Simulations were performed for sample sizes $n = 10, 20, 50, 100$. The values for the predictor X where chosen equally spaced from $X_1 = -1$ to $X_n = 2$.

TABLE 5
The logistic regression model

	Beta = 50		Beta = 1.0		Beta = 2.0	
***n* = 10**	**MLE**	**Target**	**MLE**	**Target**	**MLE**	**Target**
MSE	4.317	1.116	1.367	1.785	1.308	2.822
Bias	0.169	0.067	0.359	0.118	0.237	0.083
***n* = 20**	**Beta = 0.5**		**Beta = 1.0**		**Beta = 2.0**	
	MLE	**Target**	**MLE**	**Target**	**MLE**	**Target**
MSE	0.40	0.231	1.111	0.451	2.678	0.888
Bias	0.09	−0.034	0.294	−0.020	0.648	−0.115
***n* = 50**	**Beta = 0.5**		**Beta = 1.0**		**Beta = 2.0**	
	MLE	**Target**	**MLE**	**Target**	**MLE**	**Target**
MSE	0.134	0.115	0.204	0.158	0.618	0.339
Bias	0.022	−0.014	0.094	0.010	0.251	0.000
***n* = 100**	**Beta = 0.5**		**Beta = 1.0**		**Beta = 2.0**	
	MLE	**Target**	**MLE**	**Target**	**MLE**	**Target**
MSE	0.064	0.058	0.085	0.075	0.218	0.171
Bias	0.031	0.012	0.043	0.002	0.111	−0.001

This is an example where the function g, the expectation of the MLE has slope greater than one and so Theorems 2 and 3 can be applied and the MSE of the target estimator is reduced.

Table 5 shows the excellent performance of the target estimator compared to the MLE.

4.7. *The jackknife*, *the bootstrap and the target estimators*. Target estimation requires certain expectation computations that are not needed by either the jackknife or the bootstrap. However, targeting can provide substantial improvement over both the jackknife and the bootstrap in lowering bias and MSE, at least for larger values of the parameter, as can be seen in the following example.

Figures 7 and 8 present plots for the bias and the MSE of the MLE for the one-parameter logistic regression model compared to three different bias corrected estimators: the jackknife, the bootstrap and the target estimator. The simulations were performed for samples of size 20 where design points were fixed to equally spaced points between -1 and 2. The response values were generated from the logistic model with one slope parameter β and no intercept. A total of 84 values of β were chosen between 0 and 2.2 and the

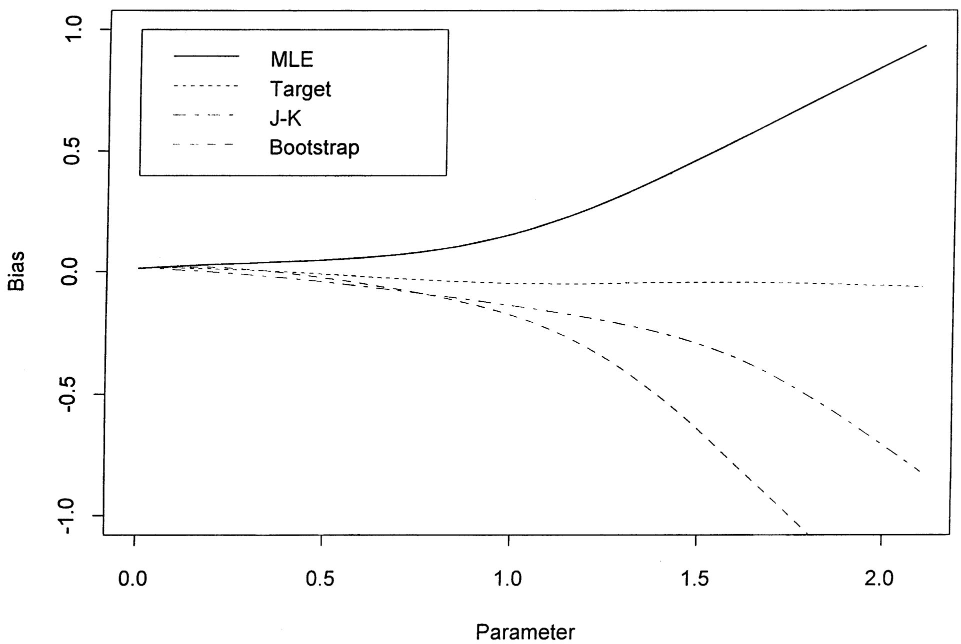

FIG. 7. *The bias of four estimators for the one-parameter logistic regression model.*

average of 30 estimates was obtained for each value of β and for each estimator (MLE, target, bootstrap and jackknife). This resulted in 84 bias estimates for the 84 values of β and for each one of the four estimators considered. The bias curves were obtained by fitting a smoothing spline with six degrees of freedom to the bias points.

In this example we notice a net gain in both bias and MSE reduction for the target estimator, especially for larger values of the parameter when the bias is not negligible. For small values of β, Figure 7 shows that the bootstrap and the jackknife estimators seem to be as effective as the target estimator in correcting the bias. For the bootstrap this can be explained by the fact that for small values of the parameter the bias is small and almost constant; that is, the function g is almost linear with derivative almost equal to one. But for larger β the function g is indeed not linear and the parametric bootstrap bias corrected estimator, $T^*_{B,n} = 2T^* - g(T^*)$, fails to correct the bias effectively. In fact, the more nonlinear g is, the worse the bootstrap estimator will perform.

In comparing targeting to the jackknife, we see that for many common statistics, in particular most maximum likelihood estimators, the biased-corrected jackknife estimator of these statistics will have a bias of order $O(1/n^2)$ (see Schucany, Gray and Owen (1971)). For the target estimators the bias could be reduced to a much smaller order such as $o(n^{-k/2})$ for $k > 3$ and sometimes eliminated as shown by the examples in Section 4.

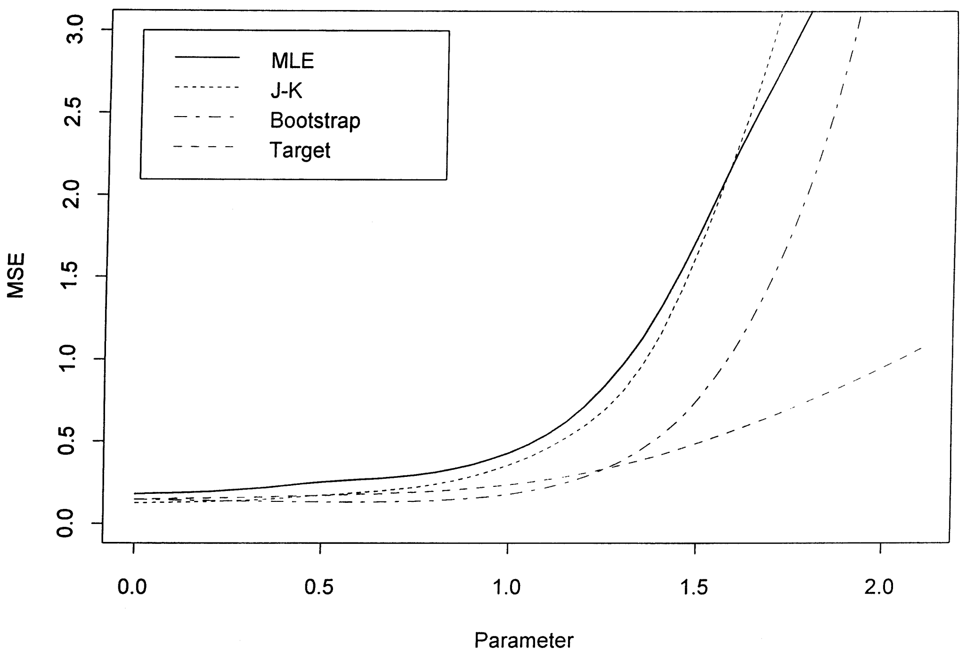

FIG. 8. *The MSE of four estimators for the one-parameter logistic regression model.*

Acknowledgment. The authors thank an anonymous referee for valuable comments that led to an improvement of the paper.

REFERENCES

CABRERA, J., MEER, P. and LEUNG, H. (1994). Simulation methods for bias correction in ellipse estimation. In *Proceedings of the Seattle Symposium on Computer Vision and Performance.*

CABRERA, J. and MEER, P. (1996). Unbiased estimation of ellipses by bootstrapping. *IEEE Trans. Pattern Analysis and Machine Intelligence* **18** 752–756.

CABRERA, J. and WATSON, G. S. (1997). Simulation methods for mean and median bias reduction in parametric estimation. *J. Statist. Plann. Inference* **57** 143–152.

DEBICHE, M. G. and WATSON, G. S. (1996). Confidence limits and bias correction for estimating angles between directions with applications to palaeomagnetism. *J. Geophysical Research* **100** 24, 405–24, 429.

EFRON, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. SIAM, Philadelphia.

EFRON, B and TIBSHIRANI, R. J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, New York.

FANKHAUSER, E. (1996). Etude du developpement de von Mises au deuxieme ordre. Diploma thesis. Ecole Polytechnique Federale, Lausanne, Switzerland.

FERNHOLZ, L. T. (1983). *Von Mises Calculus for Statistical Functionals. Lecture Notes in Statist.* **19**. Springer, New York.

FERNHOLZ, L. T. (1996). On higher order von Mises expansions. Technical Report 96-04, Dept. Statistics, Temple Univ.

FILLIPOVA, A. A. (1962). Mises theorem on the asymptotic behavior of functionals of empirical distribution function and its statistical applications. *Theory Prob. Appl.* **7**, pp. 24–57.

GATTO, R. and RONCHETTI, E. (1996). General saddlepoint approximations of marginal densities and tail probabilities. *J. Amer. Statist. Assoc.* **91** 666–673.

HAMPEL, F. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–393.

HAMPEL, F., RONCHETTI, E., ROUSSEEUW, P. and STAHEL, W. (1986). *Robust Statistics. The Approach Based on the Influence Function.* Wiley, New York.

HUBER, P. (1981). *Robust Statistics.* Wiley, New York.

REEDS, J. A. (1976). On the definition of von Mises functionals. Ph.D. dissertation, Harvard Univ.

ROUSSEEUW, P. and RONCHETTI, E. (1981). Influence curves of general statistics. *J. Comput. Appl. Math.* **7** 161–166.

SCHUCANY, W. R., GRAY, H. R., AND OWEN, D. B. (1971). On Bias Reduction in Estimation. *J. Amer. Statist. Assoc.*, **66**, pp. 524–533.

VON MISES, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Statist.* **18**, pp. 309–348.

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903
E-MAIL: CABRERA@STAT.RUTGERS.EDU

DEPARTMENT OF STATISTICS
TEMPLE UNIVERSITY
SPEAKMAN HALL
PHILADELPHIA, PENNSYLVANIA 19122
E-MAIL: fernholz@sbm.temple.edu