

MONTE CARLO EVIDENCE ON ADAPTIVE MAXIMUM LIKELIHOOD ESTIMATION OF A REGRESSION¹

BY DAVID A. HSIEH AND CHARLES F. MANSKI

University of Chicago and University of Wisconsin-Madison

This paper reports Monte Carlo evidence on the fixed sample size properties of adaptive maximum likelihood estimates of a linear regression. The focus is on the problem of selecting the smoothing and trimming parameters used in estimating the score function. We examine the performance of adaptive maximum likelihood estimators when these parameters are preselected or, alternatively, are determined by a data-based bootstrap method.

1. Introduction. Stone (1975) proposed adaptive maximum likelihood (AML) estimation of a location parameter and proved first-order efficiency. Bickel (1982) modified Stone's idea and proved that AML estimation of the slope parameters of a linear regression with random regressors and identically distributed disturbances is first-order efficient. Bickel's work extends in a straightforward way to nonlinear regressions (Manski, 1984).

These adaptation results are impressive but leave open the question of the fixed sample size behavior of AML estimates. This is always an issue in evaluating asymptotic results but seems particularly critical in the present case. One reason is that AML estimates are approximations to linearized likelihood equations estimates. The latter estimates are themselves based on a first-order approximation.

A second, more specific cause for concern is that an intermediate step in AML estimation is the use of residuals to estimate nonparametrically the score function of the disturbances. To apply the kernel method of Stone and Bickel, the statistician must select smoothing and trimming parameters. Compatibility with first-order theory requires only that the amounts of smoothing and trimming imposed be reduced at appropriate rates as the sample size increases, so that the score function estimate is mean-square consistent. Given a fixed sample size, these restrictions on rates in no way constrain one's choice of magnitudes for the smoothing and trimming parameters. Thus, the first-order theory of AML estimation is inherently incomplete as a guide to practice.

It is not reasonable to expect that a fixed sample size theory will soon be developed. The usual difficulties in parametric problems are compounded by the nonparametric step of score function estimation. The literature on kernel estimation of densities has struggled to determine how the level of smoothing should be set when estimation of the density per se is the objective. In AML estimation,

Received September 1985; revised October 1986.

¹This work was supported under National Science Foundation Grant SES-8319335 to the University of Wisconsin-Madison.

AMS 1980 *subject classifications*. 62E25, 62F11, 62F35, 62J05.

Key words and phrases. Adaptation, efficient estimation, bootstrap.

the situation is much more complex. We need to estimate a density plus its derivative and then to use these to estimate the parameters of interest.

We report Monte Carlo evidence on the fixed sample size properties of AML estimates. Our focus is on the problem of selecting the smoothing and trimming parameters used in estimating the score function. We examine the performance of AML estimates when these parameters are preselected or, alternatively, are determined by a data-based rule. The first-order theory of adaptive estimation assumes preselection. From a first-order perspective, there is no reason to consider data-based rules because preselection yields an efficient estimate. From a fixed sample size perspective, however, data-based rules may have appeal.

2. The experiments. Our experiments concern estimation of the shift parameter separating two distributions, which differ in location but are known to have identical shapes. The common shape of the two distributions is assumed not known. Equivalently, the problem is to estimate the slope parameter b of the linear model $y = a + bx + u$, where x is Bernoulli and u has the same, unknown distribution conditional on both realizations of x . The intercept a is a nuisance parameter, identified by centering the distribution of u at zero.

The data for the experiments are generated by drawing random samples from six populations of (x, u) pairs and converting realizations of (x, u) into observations of (y, x) via the equation $y = a + bx + u$ with $a = -1$ and $b = 1$. In each population, x and u are statistically independent and x is distributed Bernoulli with $\Pr[x = 0] = \Pr[x = 1] = 0.5$. The populations are distinguished by their distributions for u . These may be:

- (a) normal;
- (b) variance-contaminated normal, the mixture $0.9 * N(0, 1/9) + 0.1 * N(0, 9)$;
- (c) t with three degrees of freedom;
- (d) bimodal symmetric mixture of two normals, $0.5 * N(-3, 1) + 0.5 * N(3, 1)$;
- (e) Beta(2, 2); and
- (f) log normal, being the distribution of $\exp(z)$ for $z \sim N(0, 1)$.

For purposes of comparability, it is desirable that the various distributions have a common scale, in some accepted sense. Where necessary, we rescale the above distributions so that they all have a standard deviation equal to one. In order to identify the intercept parameter, we need to center the various distributions at a common point. Where necessary, we shift the rescaled distributions so that they all have mean equal to zero.

Pseudorandom numbers from each of the six distributions are obtained by taking appropriate functions of standard normal pseudorandom numbers. To generate normal numbers, we apply the polar method. We use the multiplicative congruential method and a shuffling table to generate the underlying uniform random numbers. These methods are described in Knuth (1969). Computer programs for generating the random numbers may be obtained from the first author.

In this paper, we report results for samples of 50 and 25 observations on (y, x) . In each sample, we compute the least-squares and least-absolute devia-

tions estimates for b . In the problem under study, the least-squares estimate is simply the difference between the means of y in the subsamples where $x = 0$ and $x = 1$. The least-absolute deviations estimate is the analogous difference between subsample medians. Using the least-squares estimate as the starting point, we compute various AML estimates based on alternative estimated score functions. We also compute the linearized likelihood equations estimate based on the true score function.

For each estimator, we report Monte Carlo estimates of root mean square error and interquartile range. Given that x is Bernoulli, there is an extremely small but positive probability that a sample contains only observations with $x = 0$ or with $x = 1$; hence that b is not estimable. For example, the probability that this occurs in a sample of size 25 is $1/2^{24}$. The treatment of any such samples as may be drawn is immaterial for the reporting of interquartile range statistics. On the other hand, root mean square errors are well defined only if one conditions on the event that the sample contains some observations with $x = 0$ and others with $x = 1$. The root mean square errors reported here should be interpreted as referring to this slight modification of the random regressor model.

The score function estimates used in our AML computations have the form proposed by Bickel (1982). Let N be the sample size. Let (a_N, b_N) be the least-squares estimate for (a, b) and let

$$u_{nN} = y_n - a_N - b_N x_n,$$

$n = 1, \dots, N$, be the associated residuals. Let ϕ be the standard normal density and let s be a positive smoothing parameter selected by the statistician. Each value of s defines a normal kernel estimate of the density function of u , namely,

$$f(u) = \frac{1}{sN} \sum_{n=1}^N \phi\left(\frac{u - u_{nN}}{s}\right).$$

To interpret this, observe that f is the density function of the sum of two independent random variables, one having the empirical distribution of the residuals and the other being $N(0, s^2)$. Thus, s is the standard deviation of the normal noise added to smooth the discrete distribution of the residuals.

Let t_1 , t_2 and t_3 be positive trimming parameters selected by the statistician. Each triple (t_1, t_2, t_3) defines a score function estimate based on f , namely,

$$q(u) = \frac{1}{f(u)} \frac{df(u)}{du}, \quad \text{if } |u| < t_1, f(u) > t_2 \quad \text{and} \quad \frac{1}{f(u)} \frac{df(u)}{du} < t_3,$$

$$q(u) = 0, \quad \text{otherwise.}$$

That is, the trimming parameters set the score to zero if the residual is too large, the density estimate too small or the trial score too large. Note that if the true density function is known a priori to be symmetric, an antisymmetric version of q can be formed. In our experiments, the symmetry assumption is not imposed.

We compute the AML estimates in the manner proposed by Bickel except that we do not split the sample. This technical device was used by Bickel to

simplify certain proofs. Not unexpectedly, experiments reported in Manski (1984) indicated that, in practice, AML estimation works much better when the sample is not split.

Our alternative AML estimates differ in the manner that the parameters (s, t_1, t_2, t_3) are chosen. A set of estimates is obtained by fixing these parameters at predetermined values. In particular, we produce estimates for all cells in a two-dimensional design selected after a round of preliminary experiments with the six distributions. In this design, the smoothing parameter takes the values

$$s = 0.01, 0.05, 0.10, 0.25, 0.50, 0.75, 1.00, 1.50, 2.00.$$

The three trimming parameters are covaried as follows

$$[t_1 = t, t_2 = \exp(-t^2/2), t_3 = t], \quad t = 3, 4, 8.$$

Based on the preliminary experiments, we concluded that little would be learned by independently varying the three trimming parameters. Note that the specified linkage between t_1 , t_2 and t_3 is such that when the estimated density f is normal, all three parameters are effective at the same values of u .

One further score function estimate is obtained using a data-based method for selecting the smoothing parameter. The data-based method evaluated here uses the bootstrap. The sample distribution of x is taken as the population distribution, the least-squares estimates of a and b are taken as the true values, and the sample distribution of least-squares residuals is taken as the population distribution of u . The bootstrap idea is to determine the optimal amount of smoothing and trimming for AML estimation on this artificial population and then to apply the optimum values to the original problem.

One might reasonably question whether a bootstrap approach makes sense here. After all, if the distribution of u really were that of the least-squares residuals, this distribution would not have a Lebesgue density as assumed in adaptive estimation theory. It is therefore not a priori clear that optimal smoothing and trimming for the bootstrap population should be close to optimal for the true population. Nevertheless, we decided to proceed based on the intuitive appeal of the bootstrap and its apparent success in other contexts.

To locate the bootstrap optimum in practice, we perform Monte Carlo AML experiments on the bootstrap population. For this purpose, an experiment consists of 50 independent trials, a rather small number dictated by computational concerns. We define the optimum values of smoothing and trimming parameters to be those yielding minimum mean square error in the AML estimate for b across the 50 trials. The results reported here fix the trimming parameters and use the bootstrap only to optimize the smoothing parameter. This simplification seemed warranted because our experiments with preselected smoothing and trimming parameters indicate that it is always best to trim only rarely. Given this and given that the computation of the bootstrap estimates is very burdensome, we decided to fix the trimming parameters by setting $t = 8$ and to make a line search for the best smoothing parameter.

As programmed, the line search evaluates the mean square errors of the AML bootstrap estimates on a grid of s values in the interval $[0.01, 4.00]$ and finds a

trial optimum. This is then compared to the mean square error at a predicted optimum s value obtained by fitting a quadratic function to the grid. The better of these two points is used.

3. Findings. Tables 1–3 present the results of the experiments. Table 1 shows the precision of the various estimators in samples of 50 observations of (y, x) . Both root mean square error and interquartile range are reported. Table 2 reports the precision of a subset of the estimators in samples of size 25. Table 3 gives quantiles for the sampling distribution of the bootstrap AML estimate.

Each experiment consisted of 2000 independent trials. We have not assessed formally the sampling error of the statistics presented in the tables. We have, however, performed several independent repetitions of some experiments with smaller numbers of trials. We have also, where possible, checked the internal consistency of the experimental findings. In particular, the root mean square errors for least-squares estimation should be the same in all experiments with a given sample size and should differ by a factor of $\sqrt{2}$ across sample sizes. Based on these repetitions and checks, we judge that the precision statistics reported in Tables 1 and 2 are generally within 5–10% of the actual ones. The quantiles in Table 3 may be less accurate in the tails.

The findings reported below refer to Table 1 until stated otherwise.

Comparability of root mean square error and interquartile range results. Measurements of estimator precision by root mean square error and by interquartile range yield very similar results, in the sense that these two statistics almost always rank the various estimators in the same order. The one exception is in the log-normal distribution. There the two sets of statistics for AML estimation have somewhat different patterns as one varies the amounts of smoothing and trimming.

Effect of trimming on AML estimates. Holding the amount of smoothing fixed, the performance of the AML estimate almost always improves as one decreases the extent of trimming by increasing t from $t = 3$ to $t = 8$. The one exception is in the interquartile range results for the log-normal distribution. There $t = 4$ gives the best performance.

Experiments not reported here show that as one reduces t below $t = 3$, the performance of AML estimates deteriorates. As t rises above $t = 8$ to $t = 16$, performance tends to remain unchanged. As t rises beyond $t = 16$ or so, the almost total absence of trimming begins to admit rare, outlying estimated score values and performance begins to deteriorate. Not unexpectedly, these rare events affect root mean square error much more than they do interquartile range.

Effects of smoothing on AML estimates. Holding the amount of trimming fixed, AML performance generally improves as one increases the extent of smoothing by increasing s from $s = 0.01$ to $s = 0.10$. In all cases, the root mean square error and interquartile range reach a local minimum in the region

TABLE 1
Precision of the estimators in samples of size 50

AML	Root mean square error			Interquartile range		
	$t = 3$	$t = 4$	$t = 8$	$t = 3$	$t = 4$	$t = 8$
A. Normal distribution						
$s = 0.01$	5.44	2.88	0.46	0.42-1.55	0.58-1.41	0.75-1.24
$s = 0.05$	0.41	0.33	0.29	0.75-1.27	0.78-1.23	0.82-1.19
$s = 0.10$	0.36	0.32	0.30	0.76-1.25	0.79-1.22	0.81-1.22
$s = 0.25$	0.35	0.35	0.35	0.78-1.24	0.77-1.24	0.77-1.24
$s = 0.50$	0.32	0.32	0.32	0.80-1.21	0.80-1.21	0.80-1.21
$s = 0.75$	0.30	0.30	0.30	0.81-1.21	0.81-1.21	0.81-1.21
$s = 1.00$	0.30	0.29	0.29	0.80-1.20	0.81-1.20	0.81-1.20
$s = 1.50$	0.32	0.29	0.29	0.80-1.20	0.81-1.20	0.81-1.20
$s = 2.00$	0.36	0.28	0.28	0.78-1.20	0.82-1.20	0.82-1.20
Bootstrap AML			0.31		0.79-1.20	
Least squares			0.28		0.82-1.18	
Least-absolute deviations			0.35		0.76-1.22	
Linearized likelihood			0.28		0.82-1.18	
B. Variance-contaminated normal distribution						
$s = 0.01$	> 10	9.51	0.33	0.50-1.55	0.62-1.40	0.77-1.20
$s = 0.05$		0.36	0.31	0.76-1.23	0.80-1.19	0.84-1.16
$s = 0.10$		0.27	0.22	0.83-1.16	0.87-1.13	0.89-1.11
$s = 0.25$		0.23	0.18	0.83-1.16	0.88-1.12	0.89-1.12
$s = 0.50$		0.41	0.40	0.73-1.32	0.74-1.31	0.74-1.31
$s = 0.75$		0.72	0.69	0.59-1.52	0.60-1.51	0.60-1.50
$s = 1.00$		1.06	0.94	0.45-1.72	0.53-1.59	0.55-1.59
$s = 1.50$		2.16	1.16	0.12-2.29	0.62-1.55	0.74-1.38
$s = 2.00$		3.82	1.17	-0.93-3.15	0.56-1.78	0.83-1.21
Bootstrap AML			0.17		0.90-1.11	
Least squares			0.28		0.82-1.17	
Least-absolute deviations			0.13		0.92-1.09	
Linearized likelihood			0.16		0.91-1.09	
C. t distribution with 3 d.o.f.						
$s = 0.01$		1.39	0.88	0.45-1.58	0.60-1.43	0.78-1.21
$s = 0.05$		0.40	0.35	0.74-1.26	0.79-1.21	0.82-1.19
$s = 0.10$		0.32	0.30	0.79-1.20	0.81-1.18	0.83-1.17
$s = 0.25$		0.25	0.24	0.83-1.18	0.84-1.16	0.84-1.16
$s = 0.50$		0.24	0.24	0.85-1.16	0.85-1.16	0.85-1.16
$s = 0.75$		0.28	0.27	0.83-1.17	0.83-1.16	0.83-1.16
$s = 1.00$		0.35	0.32	0.80-1.21	0.82-1.19	0.82-1.18
$s = 1.50$		0.63	0.51	0.73-1.29	0.80-1.20	0.81-1.17
$s = 2.00$		1.06	0.83	0.68-1.35	0.78-1.21	0.82-1.17
Bootstrap AML			0.23		0.85-1.14	
Least squares			0.28		0.82-1.17	
Least-absolute deviations			0.23		0.84-1.14	
Linearized likelihood			0.22		0.85-1.13	

TABLE 1
Continued

AML	Root mean square error			Interquartile range		
	$t = 3$	$t = 4$	$t = 8$	$t = 3$	$t = 4$	$t = 8$
D. Bimodal symmetric mixture of normal distributions						
$s = 0.01$	5.91	5.78	0.35	0.48–1.50	0.64–1.42	0.77–1.22
$s = 0.05$	0.39	0.32	0.27	0.75–1.25	0.80–1.22	0.82–1.18
$s = 0.10$	0.27	0.22	0.19	0.84–1.15	0.86–1.13	0.88–1.10
$s = 0.25$	0.23	0.19	0.18	0.83–1.16	0.87–1.12	0.87–1.12
$s = 0.50$	0.57	0.57	0.57	0.53–1.46	0.53–1.45	0.53–1.45
$s = 0.75$	0.60	0.60	0.60	0.57–1.42	0.57–1.42	0.57–1.42
$s = 1.00$	0.18	0.18	0.18	0.89–1.10	0.89–1.10	0.89–1.10
$s = 1.50$	0.26	0.26	0.26	0.83–1.16	0.83–1.16	0.83–1.16
$s = 2.00$	0.28	0.28	0.28	0.80–1.19	0.80–1.19	0.80–1.19
Bootstrap AML			0.16		0.90–1.11	
Least squares			0.29		0.81–1.19	
Least-absolute deviations			0.84		0.40–1.59	
Linearized likelihood			0.16		0.91–1.09	
E. Beta(2, 2) distribution						
$s = 0.01$	2.33	1.67	0.39	0.38–1.53	0.60–1.41	0.77–1.24
$s = 0.05$	0.41	0.34	0.30	0.72–1.26	0.77–1.24	0.79–1.19
$s = 0.10$	0.36	0.33	0.31	0.74–1.24	0.78–1.21	0.79–1.20
$s = 0.25$	0.36	0.33	0.33	0.76–1.23	0.78–1.23	0.78–1.24
$s = 0.50$	0.29	0.29	0.29	0.81–1.19	0.81–1.19	0.81–1.19
$s = 0.75$	0.28	0.28	0.28	0.82–1.18	0.82–1.18	0.82–1.18
$s = 1.00$	0.28	0.28	0.28	0.81–1.19	0.82–1.19	0.81–1.19
$s = 1.50$	0.28	0.28	0.28	0.80–1.19	0.80–1.19	0.80–1.19
$s = 2.00$	0.29	0.29	0.29	0.80–1.20	0.80–1.20	0.80–1.20
Bootstrap AML			0.29		0.80–1.20	
Least squares			0.29		0.79–1.20	
Least-absolute deviations			0.42		0.70–1.28	
Linearized likelihood			0.29		0.81–1.18	
F. Log-normal distribution						
$s = 0.01$	> 10	> 10	0.34	0.44–1.54	0.59–1.39	0.64–1.38
$s = 0.05$	0.38	0.31	0.27	0.77–1.24	0.80–1.17	0.78–1.23
$s = 0.10$	0.31	0.26	0.21	0.83–1.19	0.87–1.14	0.85–1.15
$s = 0.25$	0.24	0.22	0.22	0.84–1.16	0.86–1.16	0.82–1.20
$s = 0.50$	0.36	0.35	0.35	0.77–1.25	0.78–1.25	0.72–1.29
$s = 0.75$	0.50	0.48	0.47	0.71–1.32	0.72–1.28	0.70–1.30
$s = 1.00$	0.66	0.61	0.59	0.66–1.37	0.74–1.28	0.75–1.25
$s = 1.50$	1.21	1.02	0.86	0.52–1.69	0.82–1.22	0.79–1.21
$s = 2.00$	2.08	1.70	1.18	0.19–2.18	0.83–1.22	0.80–1.19
Bootstrap AML			0.17		0.90–1.10	
Least squares			0.27		0.83–1.18	
Least-absolute deviations			0.17		0.89–1.11	
Linearized likelihood			0.25		0.84–1.16	

TABLE 2
Precision of the estimators in samples of size 25

Estimator	Root mean square error	Interquartile range
A. Normal distribution		
Bootstrap adaptive	0.42	0.72–1.29
Least squares	0.40	0.73–1.28
Least-absolute deviations	0.51	0.67–1.34
Linearized likelihood	0.40	0.73–1.28
B. Variance-contaminated normal distribution		
Bootstrap adaptive	0.25	0.84–1.14
Least squares	0.40	0.76–1.26
Least-absolute deviations	0.19	0.87–1.12
Linearized likelihood	0.27	0.85–1.14
C. <i>t</i> distribution with 3 d.o.f.		
Bootstrap adaptive	0.34	0.77–1.22
Least squares	0.39	0.76–1.24
Least-absolute deviations	0.32	0.79–1.21
Linearized likelihood	0.32	0.81–1.20
D. Bimodal symmetric mixture of normal distributions		
Bootstrap adaptive	0.31	0.84–1.16
Least squares	0.41	0.72–1.30
Least-absolute deviations	0.95	0.31–1.73
Linearized likelihood	0.27	0.84–1.16
E. Beta(2, 2) distribution		
Bootstrap adaptive	0.42	0.72–1.28
Least squares	0.41	0.72–1.27
Least-absolute deviations	0.58	0.63–1.40
Linearized likelihood	0.45	0.72–1.29
F. Log-normal distribution		
Bootstrap adaptive	0.27	0.86–1.14
Least squares	0.40	0.76–1.23
Least-absolute deviations	0.27	0.84–1.15
Linearized likelihood	0.43	0.81–1.18

between $s = 0.05$ and $s = 0.50$. The location and depth of this minimum, however, vary substantially across distributions.

In other respects, the performance of AML as a function of s depends critically on the distribution generating the disturbances. Under the variance-contaminated normal, $t(3)$, and log-normal distributions, the RMSE function is unimodal as a function of s . Under the normal and Beta(2, 2) distributions, the local minimum noted above has a small local maximum to its right. As s rises beyond this, the global minimum root mean square error is attained and the root mean square error function becomes flat as a function of s . Under the bimodal distribution, there are two isolated global minima separated by a steep local peak.

TABLE 3
Quantiles of the standardized bootstrap AML estimate

	Quantile						
	1	5	25	50	75	95	99
$N(0, 1)$ quantiles	-2.33	-1.65	-0.67	0	0.67	1.65	2.33
Sample size							
A. Normal distribution							
50	-2.41	-1.65	-0.67	0.03	0.65	1.61	2.34
25	-2.36	-1.64	-0.66	0.02	0.69	1.63	2.43
B. Variance-contaminated normal distribution							
50	-2.41	-1.56	-0.59	0.01	0.64	1.66	2.61
25	-2.45	-1.60	-0.63	-0.04	0.54	1.59	2.77
C. t distribution with 3 d.o.f.							
50	-2.45	-1.73	-0.65	-0.04	0.61	1.60	2.27
25	-2.50	-1.65	-0.67	-0.01	0.65	1.61	2.27
D. Bimodal symmetric mixture of normal distributions							
50	-2.24	-1.55	-0.62	0.01	0.65	1.64	2.37
25	-2.93	-1.54	-0.53	0.01	0.52	1.56	3.03
E. Beta(2,2) distribution							
50	-2.26	-1.66	-0.70	-0.01	0.68	1.62	2.46
25	-2.25	-1.64	-0.67	0.01	0.67	1.73	2.27
F. Log-normal distribution							
50	-2.69	-1.62	-0.62	-0.02	0.59	1.53	2.49
25	-3.01	-1.55	-0.52	-0.01	0.52	1.43	2.55

Performance of bootstrap AML relative to AML with preselected s and t . Under all distributions, except the normal, the bootstrap AML estimate performs as well as or better than the best AML estimate with preselected s and t parameters. Under the normal distribution, the bootstrap estimate performs slightly less well than the best AML estimate.

Performance of bootstrap AML relative to least squares. Under the normal distribution, the bootstrap estimator performs slightly less well than its least-squares starting point. Some deterioration should be expected since least squares is optimal under this distribution. In the other cases, least squares is suboptimal and we would hope that the bootstrap estimator would outperform it. In fact, bootstrap AML outperforms least squares under the variance-contaminated normal, $t(3)$, bimodal mixture and log-normal distributions. It performs as well under the Beta(2, 2) distribution.

Performance of bootstrap AML relative to least-absolute deviations. In these experiments, the bootstrap AML estimator outperforms least-absolute deviations more often than not. In particular, the bootstrap estimator performs

substantially better than least-absolute deviations under the normal, bimodal and Beta(2, 2) distributions and about as well under the $t(3)$ and log-normal distributions. Least-absolute deviations appears to perform slightly better under the variance-contaminated normal distribution.

Performance of bootstrap AML relative to linearized likelihood equations. In adaptive estimation theory, the linearized likelihood equations estimate is the ideal that AML estimates seek to emulate. It is of obvious interest to see how closely AML estimates can match this ideal. For all distributions but the Beta(2, 2) and log-normal, computation of the linearized likelihood equations estimate is straightforward. The latter two distributions have bounded support. In these cases, the least-squares residuals sometimes fall outside the permissible range for the disturbances. When this occurs, we set the score equal to zero.

We find that under the variance-contaminated normal, $t(3)$, bimodal and Beta(2, 2) distributions, the bootstrap estimator performs about as well as the relevant likelihood equations estimator. Only under the normal distribution does the latter perform noticeably better and even here the difference is small. Under the log-normal distribution, the bootstrap method performs much better. We can offer no persuasive explanation for this. It may be associated with our treatment of outlying residuals in linearized likelihood equations estimation.

Performance of bootstrap AML in samples of size 25. Table 2 reports the precision of the bootstrap AML, least-squares, least-absolute deviations and linearized likelihood equations estimators in samples of size 25. Examination of the table reveals that the qualitative patterns observed in samples of size 50 continue to hold. In particular, the performance of bootstrap AML relative to the other estimators does not deteriorate.

Under all distributions except the bimodal, the root mean square error of bootstrap AML estimation in samples of size 25 is close to $\sqrt{2}$ times the corresponding root mean square error in samples of size 50. Under the bimodal distribution, the root mean square error doubles.

Shape of the distribution of the standardized bootstrap AML estimate. The interquartile range entries in Tables 1 and 2 suggest that the sampling distribution of the bootstrap AML estimate of b is symmetric around b . Table 3 provides more detailed evidence. Let \hat{b} denote a bootstrap estimate and let $r(\hat{b})$ be its root mean square error, as reported in Tables 1 and 2. Table 3 reports selected quantiles of the distribution of the standardized estimate $(\hat{b} - b)/r(\hat{b})$.

Inspection of the table indicates that in all cases, the standardized estimate is at least roughly standard normal. The distributions all appear to be median unbiased. Skewness is minor, if present at all. There is some evidence of moderate long-tailedness, particularly for the bimodal and log-normal cases in samples of size 25.

4. Conclusions. Should some version of the adaptive maximum likelihood method be used in practice? It would be too much to expect the experiments

reported here to provide a definitive answer to the statistical aspect of this question. Moreover, use in practice must depend on the cost of computations. Our findings do, however, suggest two statistical conclusions.

(i) Application of AML estimation with a preselected smoothing parameter should be avoided. The performance of AML estimates has been shown to be rather sensitive to one's choice of smoothing parameter. Moreover, behavior of these estimates as a function of the smoothing parameter is not uniform with respect to the distribution generating the disturbances.

To preselect the smoothing parameter with any confidence, one apparently needs good prior information about the distribution generating the disturbances and possibly about other characteristics of the estimation problem. Such knowledge, of course, assumes away the problem to which adaptive estimation has been proposed as a solution.

(ii) Application of AML estimation with bootstrap selection of the smoothing and trimming parameters warrants serious consideration. In our experiments, the bootstrap estimator performs as well as the AML estimate with the best preselected smoothing and trimming parameters. It compares favorably with both least-squares and least-absolute deviations. It even appears to perform as well as linearized likelihood equations estimation. These findings are very encouraging.

It is sometimes asserted that satisfactory nonparametric estimation of score functions requires very large samples; hence, adaptive estimation should perform poorly in moderate size samples. Our experiments indicate otherwise. The bootstrap AML method appears to work well in such samples, at least for the estimation problems we have examined.

Acknowledgments. Computations were performed on the Cray-1 computer of the University of Minnesota. Additional computational facilities were provided by the Graduate School of Business, University of Chicago. A preliminary version of this paper was presented at the invited session on semiparametric estimation at the 1984 Annual Meeting of the IMS. The authors are grateful to Peter Bickel and to the referees for helpful suggestions.

REFERENCES

- BICKEL, P. (1982). On adaptive estimation. *Ann. Statist.* **10** 647–671.
KNUTH, D. (1969). *The Art of Computer Programming: Seminumerical Algorithms 2*. Addison-Wesley, Reading, Mass.
MANSKI, C. (1984). Adaptive estimation of non-linear regression models. *Econom. Rev.* **3** 145–194.
STONE, C. (1975). Adaptive maximum likelihood estimators of a location parameter. *Ann. Statist.* **3** 267–284.

GRADUATE SCHOOL OF BUSINESS
UNIVERSITY OF CHICAGO
CHICAGO, ILLINOIS 60637

DEPARTMENT OF ECONOMICS
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN 53706