W1

8420

UNIVERSITY
OF
WISCONSIN-
MADISON

SSRI
Workshop Series

MONTE-CARLO EVIDENCE ON ADAPTIVE MAXIMUM LIKELIHOOD ESTIMATION OF A REGRESSION

David A. Hsieh[1]
Charles F. Manski[2]

8420

SOCIAL SYSTEMS RESEARCH INSTITUTE

Social Systems Research Institute

University of Wisconsin-Madison

MONTE-CARLO EVIDENCE ON ADAPTIVE MAXIMUM LIKELIHOOD ESTIMATION OF A REGRESSION

David A. Hsieh[1]
Charles F. Manski[2]

8420

---

[1] Assistant Professor, Graduate School of Business, University of Chicago.

[2] Professor, Department of Economics, University of Wisconsin, Madison.

The authors are grateful to Peter Bickel for helpful suggestions. Manski's work was supported under National Science Foundation Grant SES-8319335 to the University of Wisconsin, Madison. Computational facilities have been provided by the Graduate School of Business, University of Chicago.

October 1984

MONTE CARLO EVIDENCE ON ADAPTIVE MAXIMUM LIKELIHOOD ESTIMATION OF A REGRESSION

David A. Hsieh
Graduate School of Business
University of Chicago
Chicago, Illinois 60637

Charles F. Manski
Department of Economics
University of Wisconsin
Madison, Wisconsin 53706

ABSTRACT

This paper reports preliminary monte carlo evidence on the fixed sample size properties of adaptive maximum likelihood(AML) estimates of a simple 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 AML estimators when these parameters are pre-selected 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 identically distributed disturbances is first-order efficient. Bickel's work extends in a straighforward way to non-linear 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(LLE) 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 use of residuals to estimate nonparametrically the score function of the disturbances. To apply the kernel method assumed by 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 with only limited success to determine how the level of smoothing should be set when estimation of the density per se is the objective. In AML estimation, 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 preliminary 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 a simple regression problem, namely 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 $x = 0$ or 1 and where 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 $Pr[x=0] = Pr[x=1] =.5$. The populations are differentiated by their distributions for u. In particular, the shape of u 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)
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 common scale, in some accepted sense. Where necessary, we rescale the above distributions so that they all have 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.

Pseudo-random numbers from each of the six distributions are obtained by taking appropriate functions of standard normal pseudo-random 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 the fixed sample size $N=50$. Given a sample of observations on (y,x), we compute the least squares(LS) and least absolute deviations(LAD) estimates for b. In the problem under study, the LS estimate is simply the difference between the means of y in the subsamples where $x=0$ and $x=1$. The LAD estimate is the analogous difference between subsample medians. Using the LS estimate as the starting point, we compute various AML estimates based on alternative estimated score functions. We also compute the LLE estimate based on the true score function. For each estimator, we report monte carlo estimates of the root mean square error(RMSE) and the interquartile range(IQR).

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

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

$n=1,...N$ 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 $t1,t2,t3$ be positive trimming parameters selected by the statistician. Each triple $(t1,t2,t3)$ defines a score function estimate based on f, namely

$$q(u) = \frac{1}{f(u)} \frac{df(u)}{du} \quad \text{if } |u| < t1,\ f(u) > t2, \text{ and } \frac{1}{f(u)} \frac{df(u)}{du} < t3.$$

$q(u) = 0$ 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 anti-symmetric 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 $(s,t1,t2,t3)$ 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. 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

$$[t1=t,\ t2=\exp(-t^2/2),\ t3=t], \quad t = 3,4,8.$$

Based on the preliminary experiments, we concluded that little would be learned by independently varying the three trimming parameters. The linkeage assumed above is such that when the estimated density f is normal, all three of the parameters t1, t2, and t3 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 is a bootstrap procedure(BAML). The sample distribution of x is taken as the population distribution, the LS estimates of a and b are taken as the true values, and the sample distribution of LS 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 LS 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 fifty independent trials, a rather small number dictated by computational concerns. We define the optimum values of the smoothing and trimming parameters to be those yielding minimum mean square error in the AML estimate for b across the fifty trials. The results reported here fix the trimmming 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 [0.01,3.00] and finds a trial optimum. This is then compared to the MSE 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

Table 1 presents the tentative results of the experiments. We say tentative because the reported statistics are based on a limited number of trials. For all estimators except BAML, the reported RMSEs and IQRs are based on 1000 independent trials. For the BAML estimator, we have thus far only performed 500 trials. The reasons are computational. For a sample of size N=50, computation of a single AML estimate requires a little over one CPU second on the DEC-20 computer on which these experiments were performed. Computation of a single BAML estimate involving fifty AML estimations on the bootstrap population takes about 1.2 minutes. To complete 500 BAML trials on all of the six distributions required about sixty CPU hours.

We have not made a formal assessment of the sampling error of the reported RMSEs and IQRs. Various informal calculations suggest that the numbers in Table 1 may deviate from the actual RMSEs and IQRs by as much as 0.02. For this reason, we report results only to the second decimal place and exercise some caution in drawing conclusions from them.

Keeping the above in mind, examination of Table 1 reveals the following apparent patterns.

TABLE 1: EXPERIMENTAL EVIDENCE ON AML ESTIMATION OF A SHIFT IN LOCATION

A.NORMAL DISTRIBUTION

AML	ROOT MEAN-SQUARE ERROR			INTERQUARTILE RANGE		
	t=3	t=4	t=8	t=3	t=4	t=8
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
BAML	0.29			0.81-1.19		
LS	0.28			0.82-1.20		
LAD	0.37			0.75-1.25		
LLE	0.28			0.82-1.20		

B.VARIANCE CONTAMINATED NORMAL DISTRIBUTION

AML	t=3	t=4	t=8	t=3	t=4	t=8
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	0.26	0.76-1.23	0.80-1.19	0.84-1.16
s=0.10	0.27	0.22	0.18	0.83-1.16	0.87-1.13	0.89-1.11
s=0.25	0.23	0.18	0.17	0.83-1.16	0.88-1.12	0.89-1.12
s=0.50	0.41	0.41	0.40	0.73-1.32	0.74-1.31	0.74-1.31
s=0.75	0.72	0.70	0.69	0.59-1.52	0.60-1.51	0.60-1.50
s=1.00	1.06	0.98	0.94	0.45-1.72	0.53-1.59	0.55-1.59
s=1.50	2.16	1.57	1.16	0.12-2.29	0.62-1.55	0.74-1.38
s=2.00	3.82	2.64	1.17	-0.93-3.15	0.56-1.78	0.83-1.21
BAML	0.19			0.89-1.12		
LS	0.28			0.82-1.17		
LAD	0.13			0.91-1.09		
LLE	0.16			0.90-1.09		

TABLE 1 (continued)

C.t DISTRIBUTION WITH 3 D.O.F.

	ROOT MEAN-SQUARE ERROR			INTERQUARTILE RANGE		
AML	t=3	t=4	t=8	t=3	t=4	t=8
s=0.01	1.39	0.88	0.40	0.45-1.58	0.60-1.43	0.78-1.21
s=0.05	0.40	0.35	0.29	0.74-1.26	0.79-1.21	0.82-1.19
s=0.10	0.32	0.30	0.27	0.79-1.20	0.81-1.18	0.83-1.17
s=0.25	0.25	0.24	0.24	0.83-1.18	0.84-1.16	0.84-1.16
s=0.50	0.24	0.24	0.24	0.85-1.16	0.85-1.16	0.85-1.16
s=0.75	0.28	0.27	0.27	0.83-1.17	0.83-1.16	0.83-1.16
s=1.00	0.35	0.32	0.31	0.80-1.21	0.82-1.19	0.82-1.18
s=1.50	0.63	0.51	0.42	0.73-1.29	0.80-1.20	0.81-1.17
s=2.00	1.06	0.83	0.56	0.68-1.35	0.78-1.21	0.82-1.17
BAML	0.27			0.83-1.17		
LS	0.28			0.82-1.17		
LAD	0.23			0.84-1.16		
LLE	0.21			0.85-1.14		

D.BIMODAL SYMMETRIC MIXTURE OF NORMAL DISTRIBUTIONS

AML	t=3	t=4	t=8	t=3	t=4	t=8
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
BAML	0.61			0.51-1.50		
LS	0.29			0.79-1.19		
LAD	0.85			0.39-1.55		
LLE	0.16			0.90-1.09		

TABLE 1(continued)

E.BETA(2,2) DISTRIBUTION

	ROOT MEAN-SQUARE ERROR			INTERQUARTILE RANGE		
AML	t=3	t=4	t=8	t=3	t=4	t=8
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
BAML	0.30			0.85-1.22		
LS	0.29			0.80-1.20		
LAD	0.42			0.70-1.31		
LLE	0.29			0.82-1.18		

F.LOG-NORMAL DISTRIBUTION

AML	t=3	t=4	t=8	t=3	t=4	t=8
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
BAML	0.17			0.89-1.10		
LS	0.28			0.83-1.17		
LAD	0.18			0.91-1.11		
LLE	0.28			0.82-1.16		

Comparability of RMSE and IQR results

The RMSE and IQR results are very comparable, in the sense that they almost always yield the same rankings of the various estimators. The one exception is in distribution F. There the RMSE and IQR results for AML estimation have somewhat different patterns as one varies s and t.

Effect of trimming on AML estimates

Holding s 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 IQR results for distribution F. There the best performance occurs when t=4. Experiments not reported here show that as one reduces t below t=3, the performance of AML estimates deteriorates further. 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 the RMSEs much more than the IQRs.

Effect of smoothing on AML estimates

Holding t fixed, the performance of AML generally improves as one increases the extent of smoothing by increasing s from s=0.01 to s=0.10. In all cases, the RMSE and IQR reach a local minimum somewhere in the region 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 distributions B,C, and F, the RMSE function is unimodal as a function of s. Under distributions A and E, the local minimum noted above has a small local maximum to its right. As s rises beyond this, the global minimum RMSE is attained and the RMSE function becomes flat as a function of s. Under distribution D, there are two isolated global minima separated by a steep local peak.

Performance of BAML relative to AML with preselected s and t

Under distributions A,B,C, E, and F, the BAML estimate appears to perform as well as the best AML estimate with preselected s and t parameters. The only case in which the BAML has inferior performance is that of distribution D. This exception may have a simple explanation. Inspection of Table 1.D shows that fixing t at t=8, the RMSE of the AML is globally minimized when s is near s=0.25 or s=1.00 and that these minima are separated by a steep peak with local maximum in the vicinity of s=0.75. It is possible that in its attempt to optimize s, the BAML estimator finds the local maximum. Corroborating evidence is that the reported RMSE for the BAML estimate is almost identical to that at the local maximum.

Performance of BAML relative to LS

Under distribution A, the BAML estimator performs about as well as its LS starting point. This is as good as can be expected since LS is optimal under this distribution. In the other cases, LS is suboptimal and we would hope that BAML would outperform it. In fact, BAML substantially outperforms LS under distributions B and F. It performs no better under distributions C and E. As noted above, performance under distribution D is inferior. Note, however, that in this case the best AML with preselected s and t performs much better than LS. This leaves open the possibility that BAML can outperform LS here if the algorithm used to optimize s is refined.

Additional information on the behavior of the BAML estimate relative to LS is given in Table 2. This table reports the percentage of trials in which the step taken by the BAML method from its LS starting point is in the correct direction. The table also reports the percentage of times the BAML step under and over shoots the target value b. Table 2 corroborates the finding that the BAML estimator improves markedly on LS under distributions B and F but not otherwise. Observe the rather strange pattern found for distribution D. The BAML step is generally in the correct direction but is almost always too large.

TABLE 2: DIRECTION AND MAGNITUDE OF THE BAML STEP FROM THE LS STARTING POINT(Percentage of Trials)

	Distribution					
	A	B	C	D	E	F
Correct Direction	50	80	66	89	61	81
Undershoots	41	52	49	4	44	49
Overshoots(1)	4	18	9	11	9	19
Overshoots(2)	5	10	7	74	8	13
Wrong Direction	50	20	34	11	39	19

NOTE: Overshoots(1) is an overshoot that is smaller in magnitude than the distance between b and the LS estimate. Overshoots(2) is an overshoot larger than this distance. Thus, the BAML estimate is closer to b than is the LS estimate if the step either undershoots or overshoots(1).

Performance of BAML relative to LAD

In these experiments, the BAML estimator outperforms LAD more often than not. In particular, the BAML estimator performs substantially better than LAD under distributions A,D, and E and about as well under distribution F. The LAD estimator appears to perform slightly better than BAML under distributions B and C.

Performance of BAML relative to LLE

In adaptive estimation theory, the LLE estimate is the ideal which AML estimates seek to emulate. It is of obvious interest to see how closely they match this ideal. For distributions A,B, C, and D, computation of the LLE estimate is straightforward. Distributions E and F have bounded support. In these cases, the LS residuals sometimes fall outside the permissible range for the disturbances. When this occurs, we set the score equal to zero.

We find that under distributions A,B, and E the BAML estimator performs almost if not as well as the relevant LLE estimator. The gap between the two is more noticeable under distribution C but is still small. Given what has been said earlier, it is not surprising that LLE does substantially better than BAML under distribution D. Note, however, that in this case, the AML estimator with best preselected s and t performs almost as well as LLE. Under distribution F, the BAML method performs much better than does LLE. We can offer no persuasive explanation for this. It may be associated with our treatment of outlying residuals in LLE estimation.

4. CONCLUSION

Should the AML method be used in practice? It would be too much to expect the experiments reported here to provide a clear indication. Our findings do, however, suggest some tentative conclusions. In particular,

i). It is occasionally asserted that AML estimation must perform poorly in small and moderate sized samples because satisfactory nonparametric estimation of score functions requires very large samples. This view, we believe, can be rejected firmly.

ii). We feel that 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.

iii). We feel that bootstrap selection of the smoothing and trimming parameters warrants further study. We are encouraged by the apparent ability of the BAML estimator to mimic the performance of both the LLE estimate and the AML estimate with the best preselected smoothing and trimming parameters. We think it reasonable to conjecture that the version of BAML implemented here can be improved. Recall that our algorithm optimizes only over the smoothing parameter, uses only fifty trials on the bootstrap population, and employs a crude linear search procedure.

REFERENCES

1.Bickel, P.(1982), "On Adaptive Estimation", Annals of Statistics, 10, 647-671.

2.Knuth, D.(1969), The Art of Computer Programming: Seminumerical Algorithms, Vol.2, Addison-Wesley, Reading, Mass.

3.Manski, C.(1984), "Adaptive Estimation of Non-Linear Regression Models", Econometric Reviews, Vol.3, No.2.

4.Stone, C.(1975), "Adaptive Maximum Likelihood Estimators of a Location Parameter", Annals of Statistics, 3, 267-284.