J. Japan Statist. Soc.
Vol. 31 No. 1 2001 39–51

CALCULATION METHOD FOR NONLINEAR DYNAMIC LEAST-ABSOLUTE DEVIATIONS ESTIMATOR

Kohtaro Hitomi * and Masato Kagihara **

In a nonlinear dynamic model, the consistency and asymptotic normality of the Nonlinear Least-Absolute Deviations (NLAD) estimator were proved by Weiss (1991), even though they are difficult to compute in practice. Overcoming this difficulty will be critical if the NLAD estimator is to become practical. We propose an approximated NLAD estimator with the same asymptotic properties as the original with the exception that ours is easier to use.

Key words and phrases: Calculation method, Least absolute deviations estimator, Nonlinear dynamic model.

1. Introduction

In parametric regression models, the Least-Squares (LS) estimator is usually used for parameter estimates. If the error term is distributed as normal, then the LS estimator is a Maximum Likelihood Estimator (MLE) and attains minimum variance within unbiased estimators. At the same time, however, it is well known that only one outlier may cause a large error in an LS estimator. This occurs in the case of fatter tail distributions of the error term. In such a case, more robust estimators are desirable. One is the Least-Absolute Deviations (LAD) estimator.

In a linear regression model, a linear programming method is available as a calculation method for the LAD estimator. On the other hand, in the nonlinear dynamic model, no computational method is proposed, although it has been shown theoretically by Weiss (1991) that a Nonlinear LAD (NLAD) estimator is consistent and asymptotically normal.

Therefore, we seek an approximate estimator of the NLAD estimator that is practically computable. In a linear case, Hitomi (1997) proposed an estimator of this kind called the Smoothed LAD (SLAD). [1] In order to obtain an SLAD estimator, he approximated the non-differentiable original objective function by the smoothed function which is differentiable. That is, he replaced $|x|$ with $\sqrt{x^2+d^2}$ where $d>0$ is the distance from the origin

Received August 29, 2000. Revised February 9, 2001. Accepted March 17, 2001.

*Department of Architecture and Design, Kyoto Institute of Technology; E-mail: hitomi@hiei.kit.ac.jp.

**Graduate School of Economics, Kyoto University; E-mail: e60x0032@ip.media.kyoto-u.ac.jp.

1Horowitz (1998) also proposed another SLAD estimator, which has different purposes than ours. In his article, the main target is not a calculation method of the LAD estimator. In practice, his SLAD estimator seems to be difficult to compute, as he implied, because it involves an integral of kernel for nonparametric density estimation.

Figure 1.

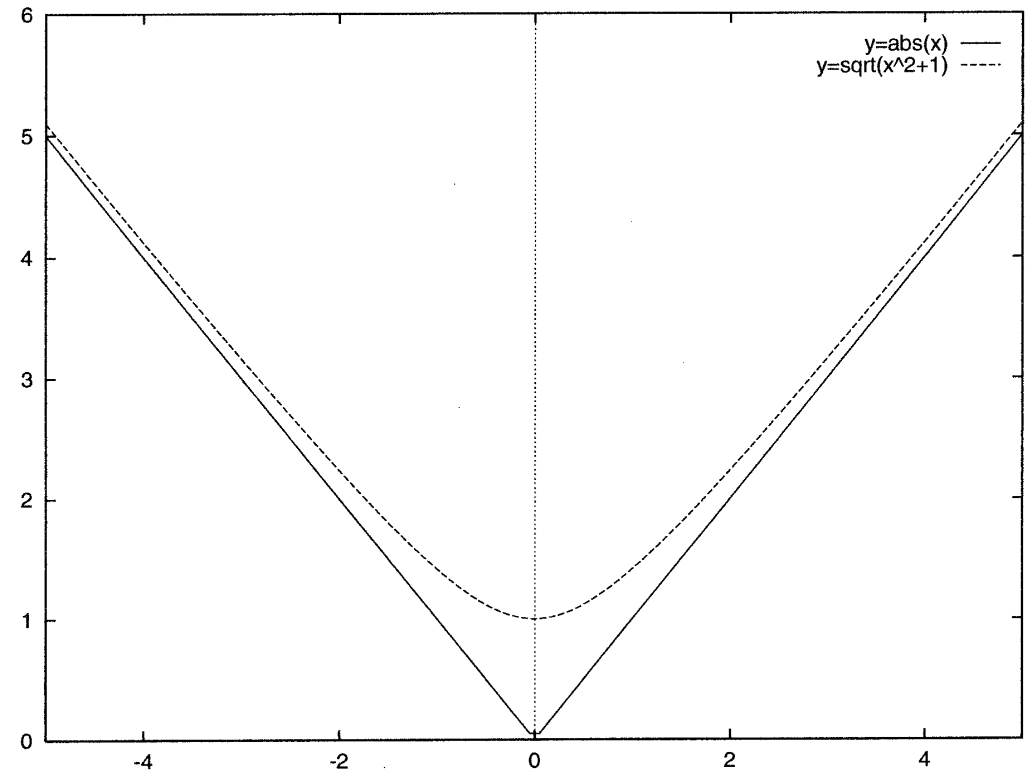
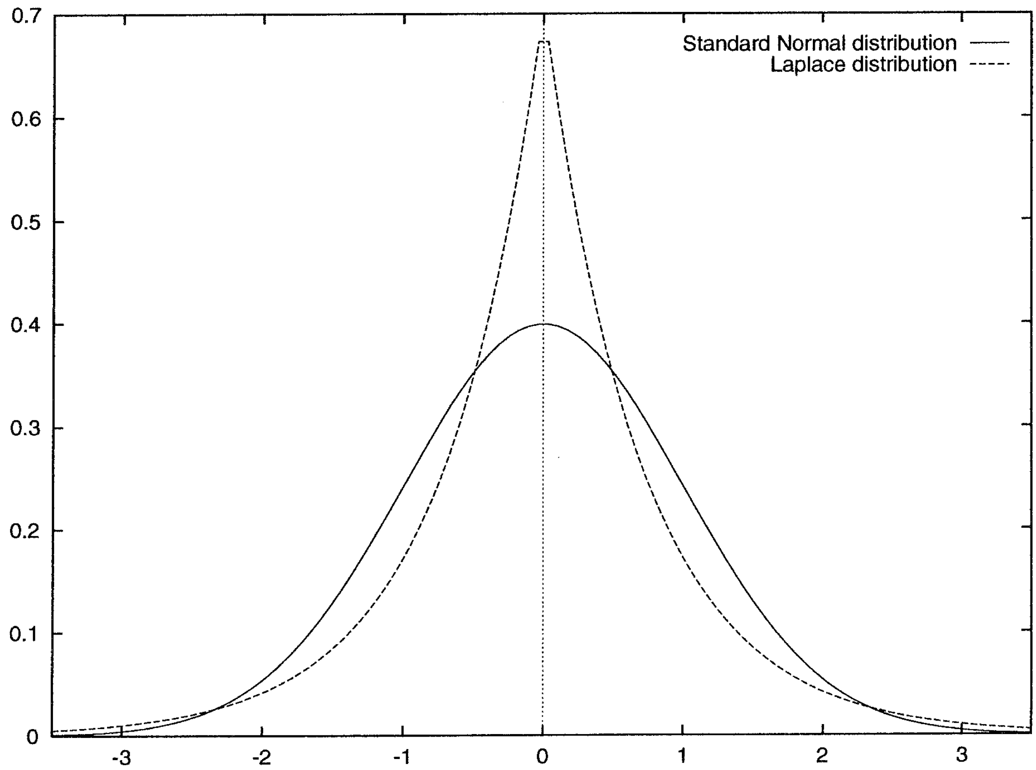


Figure 2.



(see Figure 1). Then the distance between the original and smoothed objective functions is equal to or smaller than d for all x. Hitomi (1997) controlled the parameter d connected with the sample size so that the SLAD estimator has the same asymptotic properties as the original LAD estimator. In this article, we extend this method to the nonlinear dynamic model.

In conclusion, we obtained the general calculation method of the NLAD estimator. We call it the Nonlinear SLAD (NSLAD) estimator.

We introduce a nonlinear dynamic model in Section 2, which was investigated by Weiss (1991). In Section 3, we prove that the NSLAD estimator has the same asymptotic properties as the original NLAD estimator under the nonlinear dynamic model and Section 4 presents the results of the Monte Carlo study. Concluding remarks are given in Section 5. Finally, our assumptions are described in Appendix A and the sketch of proof that our model in Section 4 meets assumptions is given in Appendix B.

2. Model

Weiss (1991) considered the following nonlinear dynamic model.

(2.1) $y_t = g(x_t, \beta_0) + u_t$

g : known function

$x_t = (y_{t-1}, \ldots, y_{t-p}, z_t)$

z_t : vector of exogenous variables

β_0 : $(k \times 1)$ vector of unknown parameter

u_t : unobserved error term which satisfies $\mathrm{Median}(u_t \mid I_t) = 0$

I_t : σ-algebra (information set at period t) generated by $\{x_{t-i}\}$ $(i \geq 0)$ and $\{u_{t-i}\}$ $(i \geq 1)$.

Then the NLAD estimator is defined as the solution of the following problem:

$$\min_{\beta} Q_T(\beta) \equiv \min_{\beta} \frac{1}{T} \sum_{t=1}^{T} |y_t - g(x_t, \beta)|. \tag{2.2}$$

In these basic settings, Weiss (1991) proved that the NLAD estimator $\hat{\beta}$ was consistent and asymptotically normal under the set of assumptions in Appendix A.

THEOREM (WEISS(1991)) *Under the assumptions described in Appendix* A,

$$\hat{\beta} \xrightarrow{p} \beta_0 \quad \textit{and} \quad \sqrt{T}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V), \tag{2.3}$$

where

$$V = D^{-1}AD^{-1},$$

(2.4) $$A = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} E\left[\frac{\partial g_t}{\partial\beta}(\beta_0)\frac{\partial g_t}{\partial\beta'}(\beta_0)\right],$$

(2.5) $$D = 2\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} E\left[f_t(0\mid I_t)\frac{\partial g_t}{\partial\beta}(\beta_0)\frac{\partial g_t}{\partial\beta'}(\beta_0)\right],$$

$f_t(\cdot\mid I_t)$: density function of u_t conditional on I_t.

However, no computational method was proposed by Weiss (1991). This is critical, and we solved this problem.

Following Hitomi (1997), $|y - g(x,\beta)|$ is approximated by $\sqrt{(y-g(x,\beta))^2+d^2}$ $(d > 0)$. The NSLAD estimator is defined as the solution of the following problem which minimizes the smoothed objective function: [2]

(2.6) $$\min_{\beta} Q_T^s(\beta) \equiv \min_{\beta}\frac{1}{T}\sum_{t=1}^{T}\sqrt{(y_t - g(x_t,\beta))^2+d^2}.$$

First and second derivatives are (where $g_t(\beta) := g(x_t,\beta)$)

(2.7) $$\frac{\partial Q_T^s}{\partial\beta} = -\frac{1}{T}\sum_{t=1}^{T}\frac{y_t - g_t(\beta)}{\sqrt{(y_t-g_t(\beta))^2+d^2}}\frac{\partial g_t}{\partial\beta}(\beta)$$

and

(2.8) $$\frac{\partial^2 Q_T^s}{\partial\beta\partial\beta'} = \frac{1}{T}\sum_{t=1}^{T}\left\{\frac{d^2}{((y_t-g_t(\beta))^2+d^2)^{3/2}}\frac{\partial g_t(\beta)}{\partial\beta(\beta)}\frac{\partial g_t(\beta)}{\partial\beta'} - \frac{y_t-g_t(\beta)}{\sqrt{(y_t-g_t(\beta))^2+d^2}}\frac{\partial^2 g_t(\beta)}{\partial\beta\partial\beta'}\right\}.$$

3. Asymptotic properties

3.1. Consistency

The difference between the original and smoothed objective functions defined in (2.2) and (2.6), respectively, is

(3.1) $$0 \le Q_T^s(\beta) - Q_T(\beta) \le d \quad \forall x, y \quad \text{and} \quad \beta.$$

Therefore, if we control the parameter d such that $d \to 0$ as $T \to \infty$,

(3.2) $$\sup_{\beta}\{Q_T^s(\beta) - Q_T(\beta)\} \to 0 \qquad \text{as} \quad T\to\infty.$$

2In appropriate conditions, Q_T converges in probability to a smooth function as $T\to\infty$ to which Q_T^s also converges in probability if $d\to 0$ as $T\to\infty$. Therefore it is understandable that the original NLAD and NSLAD estimator have the same asymptotic properties.

This implies the following theorem.

THEOREM 1. *Assume the original objective function $Q_T(\beta)$ converges in probability uniformly in β to a function that is uniquely minimized at β_0.*[3] *If $d \to 0$ as $T \to \infty$, then the NSLAD estimator is consistent.*

3.2. Asymptotic normality

The asymptotic normality of the NLAD estimator $\hat{\beta}$ is mainly based on the following asymptotic first order condition:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathrm{sgn}(y_t - g_t(\hat{\beta})) \frac{\partial g_t}{\partial \beta}(\hat{\beta}) = o_p(1). \tag{3.3}$$

Thus, if the NSLAD estimator also satisfies (3.3), it is asymptotically normal with the same asymptotic covariance matrix of the original NLAD, according to Weiss'(1991) results. Therefore, we have proved that the NSLAD estimator $\hat{\beta}^s$ satisfies the condition (3.3) in the followings. Now, note the NSLAD estimator $\hat{\beta}^s$ satisfies the first order condition:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{y_t - g_t(\hat{\beta}^s)}{\sqrt{(y_t - g_t(\hat{\beta}^s))^2 + d^2}} \frac{\partial g_t}{\partial \beta}(\hat{\beta}^s) = 0. \tag{3.4}$$

Then, all we have to do is show that the following equation is satisfied:

$$\begin{aligned} A := \frac{1}{\sqrt{T}} &\left\| \sum_{t=1}^{T} \left\{ \mathrm{sgn}(y_t - g_t(\hat{\beta}^s)) - \frac{y_t - g_t(\hat{\beta}^s)}{\sqrt{(y_t - g_t(\hat{\beta}^s))^2 + d^2}} \right\} \frac{\partial g_t}{\partial \beta}(\hat{\beta}^s) \right\| \\ &= o_p(1). \end{aligned} \tag{3.5}$$

Let δ_T be a positive number. Divide the data set into two groups such that $D_1 = \{t : |y_t - g_t(\hat{\beta}^s)| > \delta_T\}$ and $D_2 = \{t : |y_t - g_t(\hat{\beta}^s)| \leq \delta_T\}$. Then the following inequality holds.

$$\begin{aligned} A \leq& \frac{1}{\sqrt{T}} \sum_{t \in D_1} \left| \mathrm{sgn}(y_t - g_t(\hat{\beta}^s)) - \frac{y_t - g_t(\hat{\beta}^s)}{\sqrt{(y_t - g_t(\hat{\beta}^s))^2 + d^2}} \right| \left\| \frac{\partial g_t}{\partial \beta}(\hat{\beta}^s) \right\| \\ &+ \frac{1}{\sqrt{T}} \sum_{t \in D_2} \left| \mathrm{sgn}(y_t - g_t(\hat{\beta}^s)) - \frac{y_t - g_t(\hat{\beta}^s)}{\sqrt{(y_t - g_t(\hat{\beta}^s))^2 + d^2}} \right| \left\| \frac{\partial g_t}{\partial \beta}(\hat{\beta}^s) \right\| \qquad (3.6) \\ <& \left(1 - \sqrt{\frac{1}{1 + d^2/\delta_T^2}}\right) \frac{1}{\sqrt{T}} \sum_{t \in D_1} \left\| \frac{\partial g_t}{\partial \beta}(\hat{\beta}^s) \right\| + \frac{1}{\sqrt{T}} \sum_{t \in D_2} \left\| \frac{\partial g_t}{\partial \beta}(\hat{\beta}^s) \right\| \qquad (3.7) \\ =:& A_1 + A_2. \end{aligned}$$

3Weiss (1991) demonstrated this fact in his nonlinear dynamic model described in Appendix A.

First, we focus on the term A_1. By the Taylor expansion, it follows that

$$(3.8)\qquad 1-\sqrt{\frac{1}{1+d^2/\delta_T^2}}=\frac{1}{2}\left(\frac{d}{\delta_T}\right)^2+o\left(\left(\frac{d}{\delta_T}\right)^2\right).$$

Hence,

$$(3.9)\qquad A_1<\frac{\sqrt{T}}{2}\left(\frac{d}{\delta_T}\right)^2\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{\partial g_t}{\partial\beta}(\hat{\beta}^s)\right\|+o\left(\sqrt{T}\left(\frac{d}{\delta_T}\right)^2\right).$$

Thus, $A_1=o_p(1)$ when $\sqrt{T}\left(\frac{d}{\delta_T}\right)^2\to 0\ (T\to\infty)$ and

$$(3.10)\qquad \frac{1}{T}\sum_{t=1}^{T}\left\|\frac{\partial g_t}{\partial\beta}(\hat{\beta}^s)\right\|\xrightarrow{p} c:\ \text{a finite constant.}$$

Next, consider the term A_2, which we regard as a function of β,

$$(3.11)\qquad \begin{aligned} A_2=A_2(\hat{\beta}^s)&=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}1(|y_t-g_t(\hat{\beta}^s)|\le\delta_T)\left\|\frac{\partial g_t}{\partial\beta}(\hat{\beta}^s)\right\|\\ &=\sqrt{T}\frac{1}{T}\sum_{t=1}^{T}w_t(\hat{\beta}^s)\end{aligned}$$

where

$$w_t(\beta):=1(-\delta_T+\{g_t(\beta)-g_t(\beta_0)\}\le u_t\le\delta_T+\{g_t(\beta)-g_t(\beta_0)\})\left\|\frac{\partial g_t}{\partial\beta}(\beta)\right\|$$

and $1(\cdot)$ represents the indicator function. Here, assume that the conditional density function of u_t on I_t is bounded from above,

$$(3.12)\qquad f_t(u\mid I_t)<M_1\ \forall t,\ I_t$$

and the next condition is satisfied in an open neighborhood B_0 of β_0,

$$(3.13)\qquad \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}E\left\|\frac{\partial g_t}{\partial\beta}(\beta)\right\|<M_2\ \forall\beta\in B_0$$

for some real values $0<M_1,\ M_2<\infty$. Then, for $\forall\beta\in B_0$, we get

$$(3.14)\qquad \begin{aligned} E(w_t(\beta))&=E[E(w_t(\beta)\mid I_t)]\\ &<E\left[\left\|\frac{\partial g_t}{\partial\beta}(\beta)\right\|\int 1(-\delta_T+\{g_t(\beta)-g_t(\beta_0)\}\le u_t\right.\\ &\qquad\left.\le\delta_T+\{g_t(\beta)-g_t(\beta_0)\})f(u_t\mid I_t)du_t\right]\\ &<2M_1\delta_T E\left\|\frac{\partial g_t}{\partial\beta}(\beta)\right\|.\end{aligned}$$

Thus, by the Markov inequality,

$$(3.15) \qquad P(|A_2(\beta)| \geq \epsilon) \leq \frac{E|A_2(\beta)|}{\epsilon} < \frac{2}{\epsilon} M_1 \sqrt{T}\delta_T \frac{1}{T}\sum_{t=1}^{T} E\left\|\frac{\partial g_t}{\partial \beta}(\beta)\right\|.$$

Therefore, when $\sqrt{T}\delta_T \to 0$ as $T \to \infty$, $\forall \beta \in B_0$ $A_2(\beta) \xrightarrow{p} 0$. Hence, invoking the consistency of $\hat{\beta}^s$ (Theorem 1), we can conclude that $A_2(\hat{\beta}^s) \xrightarrow{p} 0$.

From the discussions presented above, $A = o_p(1)$, when the convergence rate of d satisfies

$$(3.16) \qquad \sqrt{T}\left(\frac{d}{\delta_T}\right)^2 \to 0 \qquad \text{and} \qquad \sqrt{T}\delta_T \to 0\ (T \to \infty)$$

and (3.10), (3.12) and (3.13) are satisfied.

THEOREM 2. *Suppose the following conditions are satisfied:*

(i) $y_t = g(x_t, \beta_0) + u_t$, *and the conditional median of* u_t *on* I_t *is zero.*

(ii) *The conditional density of* u_t *on* I_t *is bounded from above for all* t, I_t, *thus,* $\exists M_1$ *such that for all* t, I_t $f_t(u \mid I_t) < M_1$.

(iii) *In an open neighborhood* B_0 *of* β_0 *for a finite constant* M_2,
$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\left\|\frac{\partial g_t}{\partial \beta}(\beta)\right\| < M_2\ \forall \beta \in B_0$.
and $\frac{1}{T}\sum_{t=1}^{T}\left|\left\|\frac{\partial g_t}{\partial \beta}(\hat{\beta}^s)\right\| - E\left\|\frac{\partial g_t}{\partial \beta}(\beta_0)\right\|\right| \xrightarrow{p} 0$

Furthermore, assume the conditions described in Appendix A, *then* $\sqrt{T}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V)$ *for the NLAD estimator* $\hat{\beta}$ (*See Section* 2). *If the smoothing parameter* d *satisfies that* $T^{3/4}d \to 0$, *then the NSLAD estimator* $\hat{\beta}^s$ *is asymptotically normal.*

$$\sqrt{T}(\hat{\beta}^s - \beta_0) \xrightarrow{d} N(0, V).$$

PROOF. By Weiss'(1991) proof method of the asymptotic normality of the NLAD estimator, [4] we can see that all we have to do is show (3.3) for the NSLAD estimator because the other conditions are implied by nonlinear dynamic model described in Appendix A. Hence, we have already completed our proof in the above discussions.

4. Monte Carlo experiments

In this section, we present some examples of the NSLAD estimator and compare the performance of the NSLAD and nonlinear least-squares (NLS)

4See Weiss (1991) pp.61-63, proof of Theorem 3, especially p.62.

Table 1.

$T = 500$	$u \sim$ Normal	$\lambda(=0.5)$	$\beta_1(=1)$	$\beta_2(=1)$	$\phi(=0.5)$
Mean	(SLAD)	0.5053	0.9939	0.9969	0.5012
	(NLS)	0.5020	1.0069	0.9969	0.4985
Standard Deviation	(SLAD)	0.1059	0.1540	0.0622	0.0340
	(NLS)	0.0860	0.1251	0.0496	0.0278
Median	(SLAD)	0.5012	0.9952	0.9989	0.5029
	(NLS)	0.5029	1.0028	0.9990	0.4985
1st Quartile	(SLAD)	0.4307	0.8829	0.9587	0.4788
3rd Quartile		0.5712	1.0878	1.0367	0.5242
	(NLS)	0.4417	0.9198	0.9631	0.4783
		0.5563	1.0898	1.0292	0.5171
	$u \sim$ Laplace				
Mean	(SLAD)	0.5018	0.9930	1.0009	0.5006
	(NLS)	0.5005	1.0052	0.9995	0.4983
Standard Deviation	(SLAD)	0.0691	0.0965	0.0394	0.0213
	(NLS)	0.0884	0.1253	0.0511	0.0271
Median	(SLAD)	0.5002	0.9923	1.0017	0.5008
	(NLS)	0.4977	1.0032	1.0022	0.4991
1st Quartile	(SLAD)	0.4543	0.9311	0.9776	0.4874
3rd Quartile		0.5457	1.0536	1.0262	0.5154
	(NLS)	0.4376	0.9181	0.9652	0.4808
		0.5610	1.0865	1.0345	0.5175

estimators. These experiments are conducted in the following nonlinear dynamic model, which satisfies the assumptions given in Appendix A. [5] We set $\beta_1 = \beta_2 = 1, \lambda = \phi = 0.5$ and $y_0 = 0$.

$$y_t = \phi y_{t-1} + \beta_1 + \beta_2 \frac{z_t^\lambda - 1}{\lambda} + u_t. \tag{4.1}$$

We generated u_t from two distributions. One is the standard normal distribution, where the NLS estimator becomes MLE, and the other is the Laplace distribution, where the NLAD estimator is MLE. The density of the Laplace distribution, whose variance is adjusted to 1, is $f_u(u) = \exp(-\sqrt{2}|u|)/\sqrt{2}$. Median$(u) = 0$, $V(u) = 1$, where $V(\cdot)$ indicates a variance. Next, let z be distributed as a uniform distribution: $z \sim U(0, 9/2)$. Note: this makes $V((z_t^\lambda - 1)/\lambda) = 1 = V(u)$.

The experiments are examined under the following conditions: (1) The number of replications is 1,000. (2) The sample sizes (T) are 50, 100, 200,

5See Appendix B.

Table 2.

$T = 200$	$u \sim$ Normal	λ	β_1	β_2	ϕ
Mean	(SLAD)	0.5219	0.9849	0.9905	0.5015
	(NLS)	0.5133	1.0141	0.9908	0.4955
Standard Deviation	(SLAD)	0.1834	0.2281	0.1053	0.0515
	(NLS)	0.1429	0.1884	0.0824	0.0430
Median	(SLAD)	0.5060	0.9756	0.9995	0.5010
	(NLS)	0.5063	1.0159	0.9952	0.4945
1st Quartile	(SLAD)	0.4012	0.8226	0.9231	0.4683
3rd Quartile		0.6397	1.1380	1.0613	0.5369
	(NLS)	0.4126	0.8913	0.9364	0.4655
		0.6127	1.1408	1.0466	0.5262
	$u \sim$ Laplace				
Mean	(SLAD)	0.5079	0.9845	1.0008	0.5019
	(NLS)	0.5077	1.0049	0.9972	0.4977
Standard Deviation	(SLAD)	0.1144	0.1581	0.0653	0.0347
	(NLS)	0.1459	0.1895	0.0813	0.0418
Median	(SLAD)	0.5017	0.9768	1.0033	0.5029
	(NLS)	0.4994	1.0002	0.9990	0.4973
1st Quartile	(SLAD)	0.4323	0.8748	0.9589	0.4783
3rd Quartile		0.5837	1.0808	1.0434	0.5248
	(NLS)	0.4122	0.8792	0.9429	0.4704
		0.5983	1.1238	1.0561	0.5270

and 500. (3) The smoothing parameter is, $d = T^{-1}$, which satisfies the conditions of Theorems 1 and 2.

The results are reported in the following tables. Each table is reported in the same format. The upper block of this table shows a case with an error term distribution that is standard normal, and the lower block shows a case with the Laplace distribution. In each block, for the NLS and NSLAD estimators, the sample mean, standard deviation, median, and 1st and 3rd quartiles are reported.

The bias becomes negligible as the sample size increases, and there are no differences between the NSLAD and NLS estimators if attention is focused exclusively on this point. In terms of standard deviations, the NLS estimator has a smaller standard deviation when the error term's distribution is standard normal, and the NSLAD estimator has a smaller standard deviation when the error term's distribution is the Laplace distribution. When the error term is distributed as standard normal, standard deviations of the NSLAD estimators are about 20% larger than those of the NLS estimators and at most 28.3% larger in the estimate of λ as $T = 200$. When

Table 3.

$T=100$	$u \sim$ Normal	λ	β_1	β_2	ϕ
Mean	(SLAD)	0.5401	0.9847	0.9834	0.4984
	(NLS)	0.5237	1.0274	0.9839	0.4899
Standard Deviation	(SLAD)	0.2565	0.3252	0.1427	0.0742
	(NLS)	0.2069	0.2640	0.1158	0.0596
Median	(SLAD)	0.5152	0.9417	0.9941	0.5040
	(NLS)	0.5098	1.0269	0.9882	0.4907
1st Quartile	(SLAD)	0.3756	0.7428	0.8982	0.4557
3rd Quartile		0.6817	1.1952	1.0854	0.5484
	(NLS)	0.3835	0.8395	0.9100	0.4496
		0.6371	1.2011	1.0622	0.5300
	$u \sim$ Laplace				
Mean	(SLAD)	0.5235	0.9760	0.9930	0.5028
	(NLS)	0.5189	1.0187	0.9865	0.4938
Standard Deviation	(SLAD)	0.1730	0.2197	0.1017	0.0501
	(NLS)	0.2012	0.2612	0.1148	0.0591
Median	(SLAD)	0.5160	0.9640	0.9981	0.5041
	(NLS)	0.5055	1.0135	0.9923	0.4944
1st Quartile	(SLAD)	0.4206	0.8257	0.9369	0.4699
3rd Quartile		0.6190	1.1127	1.0572	0.5387
	(NLS)	0.3899	0.8478	0.9156	0.4520
		0.6302	1.1828	1.0673	0.5329

the error term is distributed as the Laplace distribution, standard deviations of the NLS estimators are $10\% - 30\%$ larger than those of the NSLAD estimators, and the largest difference is 65.7% larger in the estimate of λ as $T = 50$. The second largest is 29.8% larger in the estimate of β_1 as $T = 500$. With respect to median and quartiles, there are no differences between the NLS and NSLAD estimators regarding their performances. Finally, we compare the computing time of the NLS and NSLAD estimators. The average computing time of the NLS and NSLAD estimators is 0.35 and 0.47 seconds, respectively, when $T = 500$ with the standard normal error, and also 0.35 and 0.42 seconds when $T = 500$ with the Laplace error. In the same manner, the average computing time of the NLS and NSLAD estimators is 0.28 and 0.37 seconds, respectively, when $T = 200$ with standard normal, and also 0.28 and 0.33 seconds when $T = 200$ with the Laplace distribution. Similarly, when $T = 100$ and 50, the NSLAD estimate takes about 30% as much time as the NLS estimate with standard normal, and takes about 20% as much time as the NLS estimate with the Laplace distribution. We use GAUSS for Windows (32 bit), Version 3.2.38., and GAUSS

Table 4.

$T = 50$	$u \sim$ Normal	λ	β_1	β_2	ϕ
Mean	(SLAD)	0.5520	0.9905	0.9705	0.5002
	(NLS)	0.5281	1.0380	0.9808	0.4887
Standard Deviation	(SLAD)	0.4773	0.4344	0.2496	0.1064
	(NLS)	0.3931	0.3717	0.2094	0.0874
Median	(SLAD)	0.4895	0.9134	0.9956	0.5102
	(NLS)	0.4759	1.0267	0.9895	0.4939
1st Quartile	(SLAD)	0.2550	0.6938	0.8074	0.4308
3rd Quartile		0.7687	1.2481	1.1416	0.5777
	(NLS)	0.2828	0.7783	0.8613	0.4317
		0.7274	1.2903	1.1226	0.5501
	$u \sim$ Laplace				
Mean	(SLAD)	0.5133	0.9624	1.0026	0.5053
	(NLS)	0.5455	1.0264	0.9814	0.4919
Standard Deviation	(SLAD)	0.3300	0.2952	0.1915	0.0731
	(NLS)	0.5468	0.3445	0.2029	0.0826
Median	(SLAD)	0.4993	0.9159	1.0154	0.5104
	(NLS)	0.5000	1.0113	0.9880	0.4943
1st Quartile	(SLAD)	0.2964	0.7483	0.8967	0.4608
3rd Quartile		0.6853	1.1311	1.1273	0.5574
	(NLS)	0.2747	0.7812	0.8560	0.4369
		0.7255	1.2624	1.1193	0.5485

Applications Optimization to conduct these experiments in an environment with a CPU that has a PentiumII, 400MHz, and 64MB memory.

From these results, we can conclude that the NSLAD estimator performed very well and equal to the NLS estimator.

5. Concluding remarks

The aim of this article is on the actual usage of the NLAD estimator, which has attractive properties such as robustness. The problem lies in computational difficulty. Therefore, we proposed an NSLAD estimator, a generalization of Hitomi's(1997) SLAD estimator, which is practically computable and has the same asymptotic properties as the NLAD estimator in Weiss'(1991) nonlinear dynamic model. The Monte Carlo experiment implies a good performance of the NSLAD estimator, at least equal to the NLS estimator.

Consequently we obtained a computable approximate to the NLAD estimator, which we called the NSLAD estimator, and the NLAD and NSLAD estimators have the same asymptotic properties.

Appendix A: Set of Assumptions

In this section, we described the set of assumptions under which Weiss (1991) demonstrated the consistency and asymptotic normality of the NLAD estimator. The basic model has already been introduced in Section 2.

ASSUMPTION. In the model (2.1),

(i) (Probability space) $(\Upsilon, \mathcal{F}, P)$ is a complete probability space and $\{u_t, x_t\}_{t=1}^T$ are random vectors on this space.

(ii) (Parameter space) Let B be a compact subset of Euclidean space $\boldsymbol{R}^k$, and β_0 is interior to B.

(iii) (Regression function) $g(x, \beta)$ is measurable in x for each $\beta \in B$ and is A-smooth in β with variables A_{0t} and function ρ. [6] $\nabla_i g_t(\beta)$ is A-smooth with variables A_{it} and functions $\rho_i, i = 1, \dots, k$. In addition, $\max_i \rho_i(s) \leq s$ for $s > 0$ small enough. [7]

(iv) (Conditional distribution of u_t) $f_t(\cdot \mid I_t)$ is Lipschitz continuous uniformly in t, and for each t, u, $f_t(u \mid I_t)$ is continuous in its parameter. Furthermore there exist $h > 0$ such that for all t, and some constant $p_1 > 0, P(f_t(0 \mid I_t) \geq h) > p_1$. [8]

(v) (Unconditional density) An unconditional density f_t of (u_t, x_t) is continuous in its parameter at every point for each t.

(vi) There exists $\Delta < \infty$ and $r > 2$ such that $\alpha(m) \leq \Delta m^\lambda$ for some $\lambda < -2r/(r-2)$, where $\alpha(\cdot)$ stands the usual measure of dependence for α-mixing random variables on $(\Upsilon, \mathcal{F}, P)$ with respect to the σ-algebras generated by $\{u_t, x_t\}$.

(vii) (Dominance conditions) [9] For all t, i, $E|\sup_\beta A_{it}|^r \leq \Delta_1 < \infty$. There exist measurable functions a_{1t}, a_{2t} such that $|\nabla_i g_t(\beta)| \leq a_{1t}, i = 1, \dots,$ $k, |f_t| \leq a_{2t}$ and for all t, $\int a_{2t} du_t dx_t < \infty$ and $\int a_{1t}^3 a_{2t} du_t dx_t < \infty$.

(viii) (Asymptotic stationarity) There exists a matrix A such that $T^{-1} \sum_{t=a+1}^{a+T} E[\nabla g_t(\beta_0) \nabla' g_t(\beta_0)] \to A$ as $T \to \infty$, uniformly in a.

(ix) (Identification condition) There exists $\delta > 0$ such that for all $\beta \in B$ and T sufficiently large, $det(T^{-1} \sum E[\nabla g_t(\beta) \nabla' g_t(\beta) \mid f_t(0 \mid I_t) > h]) > \delta$.

Appendix B: Confirming Assumptions in Our Model

In this section we briefly sketch a proof that our model in Section 4 meets the assumptions described in Appendix A

---

6 $g(x_t, \beta)$ is A-smooth with variables A_t and function ρ if, for each $\beta \in B$, there is a constant $\tau > 0$ such that $\|\tilde{\beta} - \beta\| \leq \tau$ implies that $|g_t(\tilde{\beta}) - g_t(\beta)| \leq A_t(x_t)\rho(\|\tilde{\beta} - \beta\|)$ for all t a.s., where A_t and ρ are nonrandom functions such that $\limsup_{T \to \infty} T^{-1} \sum E[A_t(x_t)] < \infty, \rho(y) > 0$ for $y > 0, \rho(y) \to 0$ as $y \to 0$.

7 $\nabla_i := \partial / \partial \beta_i$ $(i = 1, \dots, k)$, $\nabla := \partial / \partial \beta$.

8 $f_t(\cdot \mid I_t)$ is a density function of u_t conditional on I_t.

9 For a consistent result, these conditions are replaced by the following: There exist measurable functions a_{0t}, a_{1t} such that $|g_t(\beta)| \leq a_{0t}, |\nabla_i g_t(\beta)| \leq a_{1t}, i = 1, \dots, k$, where $E|a_{jt}|^{r_j} \leq \Delta < \infty, j = 0, 1$ for some $r_0 > 1, r_1 > 2$. The other conditions can be relaxed.

First, we begin with Assumption 2. We set $-1+c \leq \phi \leq 1-c, c \leq \lambda \leq 1-c$ $(0 < c < 1)$ and $-C \leq \beta_1, \beta_2 \leq C$ $(0 < C < \infty)$. Note that the parameter space and the support of z are bounded. For Assumption 3, we use the Taylor expansions, and for Assumption 4, we can show $|f(u_2) - f(u_1)| < |u_2 - u_1|$ and $f(u) > h$ is trivial. As to Assumption 6, we transformed our model as follows: $y_t^* = \frac{1}{2} y_{t-1}^* + \eta_t = \sum_{i=0}^{\infty} \frac{1}{2^i} \eta_{t-i}$ s.t. $y_t^* = y_t - 2(2\sqrt{2} - 1)$, $\quad \eta_t = u_t + 2(\sqrt{z_t} - \sqrt{2}) \sim$ i.i.d. To this transformed model, we apply Theorem 14.9 in Davidson (1994) p.219. For Assumption 7, note an unconditional density $f_t = f$ because y_t is stationary. We can see that $E[\nabla g_t(\beta) \nabla' g_t(\beta)] < \infty$ for Assumption 8, and $E[\nabla g_t(\beta) \nabla' g_t(\beta)]$ is nonsingular in Assumption 9.

References

Buchinsky, M. (1995). Quantile regression, Box-Cox transformation model, and the U.S. wage structure, 1963–1987. *Journal of Econometrics*, **65**, 109–154.

Davidson, J. (1994). *Stochastic Limit Theory* . Oxford University Press.

Hitomi, K. (1997). Smoothed least absolute deviations estimator and smoothed censored least absolute deviations estimator. *mimeo.*

Horowitz, J. L.(1998). Bootstrap methods for median regression models. *Econometrica*, **6**, 1327–1351.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics* (ed. Engle, R. F. and McFadden, D. L.), volume 4. Elsevier.

Pollard, D. (1985). New ways to prove central limit theorems. *Econometric Theory*, **1**, 295–314.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory*, **7**, 186–199.

Portnoy, S. and Koenker, R. (1997). The Gaussian hare and the Laplacian tortoise: Computability of squared-error versus absolute-error estimators. *Statistical Science*, **12**, 279–300.

Sposito, V. A. (1990). Some properties of l_p–estimation. *Robust Regression: Analysis and Applications* (ed. Lawrence, K. D. and Arthur, J. L.). Marcel Dekker, Inc.

Weiss, A. A. (1991). Estimating nonlinear dynamic models using least absolute error estimation. *Econometric Theory*, **7**, 46–68.