Asymptotic Theory for some High Breakdown Point Estimators

Victoria Zinde-Walsh[*]
Department of Economics,
McGill University,
855 Sherbrooke Street West, Montreal, Quebec H3A 2T7, Canada
Tel. (514) 398 4834

June 25, 2002

ASYMPTOTIC THEORY FOR HBP ESTIMATORS

Victoria Zinde-Walsh
Department of Economics
McGill University,
855 Sherbrooke Street West,
Montreal, Quebec H3A 2T7
Canada

Abstract

High breakdown point estimators in regression are robust against gross contamination in the regressors as well as the errors; the least median of squares (LMS) estimator has the additional property of packing the majority of the sample most tightly around the estimated regression hyperplane in terms of absolute deviations of the residuals and thus is helpful in identifying outliers. Asymptotics for a class of high breakdown point smoothed LMS estimators are derived here under a variety of conditions that allow for time series applications; joint limit processes for several smoothed estimators are examined. The limit process for the LMS estimator is represented via a generalized Gaussian process that defines the generalized derivative of the Wiener process.

1 Introduction

Increasingly, in econometrics attention focuses on statistical methodology that can deal successfully with problems of data contamination, resulting from either miscoding or from heterogeneity when the majority of observations are generated by a process that appears to differ from the one generating some of the observations; various financial data series give examples of observations vastly removed from the majority of data. The concept of a breakdown point of an estimator provides a framework within which it is possible to examine robustness against gross contamination in part of the data; for example, Sakata and White (1998) studied some high breakdown point estimation procedures in application to S&P returns data. Here we examine a classical high breakdown point estimator, the least median of squares, and a class of related estimators and develop asymptotic results that extend to models involving time series data.

The standard definition (see Rousseeuw and Leroy, 1987) of a finite sample breakdown point in a regression estimator is as follows. Consider a regression estimator $T(Z)$, where Z represents a sample of n observations, $Z'(m, n)$ coincides with Z for all but m of the observations that have been replaced with arbitrary unbounded values and define

$$bias(m, T, Z) = \sup \|T(Z'(m,n)) - T(Z)\|,$$

where supremum is taken over all possible $Z'(m, n)$. If the bias $= \infty$, it is said that the estimator "breaks down". The breakdown point is then defined as the minimum amount of contamination to cause the breakdown of an estimator: $\varepsilon(T, Z) = \min\{\frac{m}{n}; \text{bias}(m, T, Z) = \infty\}$.

Clearly the OLS estimator has a finite sample breakdown point $\frac{1}{n}$; the LAD estimator by construction is resilient to high contamination in regression errors but not in the regressors and in the event of such contamination breaks down at $\frac{1}{n}$. A number of regression estimators are resistant to contamination of the regressors as well as errors. An example of such an estimator is the least median of squares (LMS) estimator introduced by Rousseeuw (1984) which has a finite sample breakdown point close to $\frac{1}{2}$, specifically, Theorem 2, p.118 of Rousseeuw and Leroy (1987) demonstrates that generally the breakdown point of LMS is $\frac{([n/2]-k+2)}{n}$ where k is the number of regressors. The breakdown point of the LMS estimator cannot be improved on by any estimator that depends on the sample in a desired fashion, i.e. is regression and/or affine equivariant.

4

There are however many estimators with the same breakdown point and a number of studies in the statistical literature compare their relative merits. Among the various estimators the LMS estimator while robust to gross contamination is characterized by lack of local robustness that is related to its sensitivity to "inliers" due to the unboundedness of the influence function at small values (see Davies, 1993). Finite sample comparisons, e.g. by Ferretti et al.,1999 show that as expected LMS does relatively well when there is substantial gross contamination (demonstrating its global robustness) but not as well against local contamination (lack of local robustness). Since many financial data series are very accurately recorded, but exhibit occasional extreme volatility LMS may provide valuable information.

The LMS estimator additionally possesses the "shorth" property, that is it provides a regression hyperplane around which 50% of the observations is most tightly packed in terms of absolute deviation of the residuals (see Rousseeuw and Leroy(1987) for this property). This property makes the LMS estimator a convenient tool for identifying outliers and potential non-homogeneity (see, e.g. Rousseeuw and van Zomeren (1990) and Atkinson (1994) for examples of use of the LMS to identify outliers).

These properties indicate that the LMS estimator could be a useful tool in applications to data that are likely to lead to a breakdown of other estimators. Closeness of results for standard estimators and the LMS could be used as evidence that contamination/outliers do not present a significant problem for the standard methodology (see Campbell and Galbraith (1993) for a rare example of use of the LMS estimator in econometrics).

It is known that the LMS estimator has a few drawbacks in terms of its asymptotics. It has a slower convergence rate than usual: $n^{-\frac{1}{3}}$ rather than $n^{-\frac{1}{2}}$ in the standard (i.i.d.) case. The description of the asymptotic process of LMS incorporates a random element, a minimizer of a process related to the Wiener process (see Rousseeuw and Leroy(1987), Kim and Pollard (1990)). The current results on the asymptotics of the LMS estimator do not cover cases that allow for models typically used with stochastic time series data, e.g. dynamic models; this paper is the first to provide such results.

The difficulty in deriving the limit process for the LMS estimator is that the absolute value function does not possess a continuous derivative and one cannot rely on straightforward Taylor expansions. Kim and Pollard, 1990 derive the asymptotics for the LMS estimator for the i.i.d. case utilizing empirical distributions approach. A way of circumventing non-differentiability via use of generalized functions was developed by P.C.B. Phillips (1991, 1995)

for the LAD estimator. The approach here is similar in spirit.

Unlike the case of the LAD estimator where generalized functions were used by Phillips to obtain intermediate results, but did not enter into the characterization of the asymptotic process, here the limit process of LMS itself is represented via a generalized Gaussian process in the sense of the definition of Gel'fand and Vilenkin (1964). A generalized Gaussian process is given by a generalized random function g which defines a continuous linear functional $g(\psi)$ on the space K of smooth functions with finite support, such that for any linearly independent set of functions $\psi_1, ..., \psi_l \in K$ the random vector $(g(\psi_1), ..., g(\psi_l))'$ has a multivariate normal distribution. Gel'fand and Vilenkin show that a generalized Gaussian process is uniquely determined by its mean functional and (bilinear) covariance functional. The gain from using the generalized functions approach is that g itself may not be a random function in the ordinary sense (e.g. is a derivative of the sign function) but provides an ordinary random variable when coupled with a function $\psi \in K$, which can be said to smooth g.

In this paper we introducing a class of estimators that are smoothed versions of the LMS estimator and derive their asymptotic properties. The smoothed estimators have the same finite sample breakdown point as the LMS estimator, but can achieve faster convergence rates (though still below the standard rate) and thus are of independent interest. Smoothing may also improve sensitivity to inliers. The smoothed LMS estimators do not, however, possess the "shorth" property.

We derive the joint (Gaussian) asymptotic distribution for any number of smoothed estimators; this result could prove to be useful in itself since unlike joint distributions of standard estimators here we can obtain asymptotic independence and thus can improve the efficiency of estimation by using, say, an average of the asymptotically independent estimators.

We characterize the limit process for the LMS estimator expressing it via a generalized Gaussian random process, specifically one that represents a generalized derivative of a Wiener process described by Gel'fand and Vilenkin (1964). Thus for the LMS estimator we have an asymptotic characterization that differs from the form derived by Kim and Pollard (1990) for the i.i.d. case; they describe the limit process as a minimizer of a functional related to a Wiener process, thus a random element enters into the asymptotic process. Here instead of indexing the limit process by a minimizer of a random process we index it by the smoothing functions from the space K.

We consider a fairly wide class of models that allow for trending regres-

6

sors, lagged dependent variable; we do not require existence of moments for regressors and errors, although certain regularity conditions are imposed. Many of the conditions are sufficient rather than necessary, but in order to streamline exposition we limit the discussion of possible extensions.

The main aim of this paper is to develop the asymptotic results for the LMS estimator and the class of high breakdown point smoothed LMS estimators. Practical application of these estimators will require their computation. While for the LMS estimator itself algorithms are available, the smoothed LMS estimators will have to be computed using some techniques that can deal successfully with cases of many local extrema. In the case of the smoothed maximum score estimator Horowitz (1992) refers to tunneling and generalized simulated annealing as methods of global optimization suitable for computation in this case (for his Monte Carlo experiments he used search over a discrete set of values).

Section 2 of the paper describes the class of estimators that solve smoothed conditions similar to those for LMS and derives asymptotic results for the smoothed LMS estimators. Section 3 examines the joint limit process for several of the smoothed estimators combined and describes the limit process for the LMS estimator via a generalized Gaussian process. Section 4 contains the conclusions. Appendix A provides the derivatives of the smoothed criterion functions; the proofs of the asymptotic results are collected in Appendix B; Appendix C is devoted to proving the finite sample high breakdown point property for the smoothed estimators.

2 The estimators that solve smoothed conditions.

We start by defining and discussing the LMS estimator.

Consider a sample

$$\left\{ (X_i, y_i)' \in R^{k+1}; i = 1, .., n \right\}$$

from a linear model

$$y_i = X_i \beta_0 + u_i. \tag{1}$$

The least median of squares (LMS) estimator β_{LMS} is

$$\beta_{LMS} = \arg \min(median \left\{ (y_i - X_i \beta)^2 \right\}) \tag{2}$$

7

The LMS estimator can also be introduced in a sequential manner. Define a positive function $a(\beta)$ that solves

$$a(\beta) = \operatorname*{arg\,min}_{a \in R} \sum \left| (y_i - X_i\beta)^2 - a^2 \right| \tag{3}$$

then $a(\beta)^2$ is the median of squared residuals, $\{(y_i - X_i\beta)^2\}$. Its minimizer is the LMS estimator

$$\beta_{LMS} = \operatorname*{arg\,min}_{\beta \in R^k} a(\beta); \tag{4}$$

Consider (1) and the sequential conditions (3,4) that the LMS estimator solves. The following Assumption ensures that there is a value for the parameter vector that corresponds to the median of squared errors in the population, which in a symmetric distribution is the square of the 75th (or 25th) percentile.

Assumption 1. *The distribution of the error vector $u = (u_1, ... u_n)'$; $u_i = y_i - X_i\beta_0$ for every n is such that for any i the conditional distribution of its component u_i is symmetric with median 0 and the same 75th percentile $a_0 > 0$.*

It follows that the median of the distribution of squared errors is a_0^2. If $a_0 = 0$ there is a mass of more than 50% at the median, in such a case higher rates of convergence would obtain; this would be excluded by strict monotonicity of the distribution function and it is not investigated here. Assumption 1 holds if the errors are identically and independently distributed with a symmetric continuous density; many other error distributions could satisfy the assumption.

Consider the problem in (3,4). Our approach is to replace the non-differentiable absolute value function in (3) by a smooth function. Smoothing is introduced via a function with finite support that shrinks to zero as sample size n increases, at a rate that is determined by a bandwidth parameter b_n. The assumption of finite support simplifies derivations, but the results can similarly be derived for e.g. smoothing with a Gaussian density function. We can assume without loss of generality that the support of the smoothing function is contained in [-1,1].

Assumption 2. *(a) The smoothing function ψ is a symmetric non-negative continuously differentiable function with finite support;*

(a') The support of ψ is in $[-1, 1]$.

(b) The bandwidth parameter $b_n \to 0$;

8

or alternatively

(b') The bandwidth parameter $b_n = n^{-\alpha}$, $0 < \alpha < \frac{1}{2}$.

Consider a function ψ which satisfies Assumption 2 and introduce the smooth function

$$\Phi_n(\beta, a) = \sum \int \left| (y_i - X_i \beta - v)^2 - a^2 \right| \psi(\frac{v}{b_n}) dv. \tag{5}$$

Here and throughout the paper all integrals are assumed to be from $-\infty$ to ∞; they always converge due to the finite support for the functions. It is easy to verify that $\Phi_n(\beta, a)$ is a stochastic function twice continuously differentiable with respect to β, a. We replace (3) by

$$a(\beta) = \arg\min_{a \in R} \Phi_n(\beta, a).$$

The derivative $\frac{\partial \Phi_n}{\partial a}(\beta, a)$ of $\Phi_n(\beta, a)$ with respect to a can be written as

$$F_n(\beta, a) = \sum_{i=1}^{n} \int sgn((y_i - X_i \beta - v)^2 - a^2) \psi(\frac{v}{b_n}) dv. \tag{6}$$

Then $a(\beta) = \arg\min_{a \in R} \Phi_n(\beta, a)$ satisfies the first order condition

$$F_n(\beta, a) = 0. \tag{7}$$

For any (β, a) that satisfy this equation and such that $\frac{\partial F_n}{\partial a}(\beta, a) \neq 0$ by the Implicit Function Theorem a function $a(\beta)$ is defined in some neighbourhood of the (β, a); the function $a(\beta)$ is continuously differentiable and its derivative is given by

$$\frac{\partial a(\beta)}{\partial \beta} = -\frac{\frac{\partial F_n}{\partial \beta}}{\frac{\partial F_n}{\partial a}}.$$

Then the first order condition for minimization of $a(\beta)$ is

$$\frac{\partial F_n}{\partial \beta}(\beta, a) = 0. \tag{8}$$

Assume that a solution to (7,8) exists; denote it $(\tilde{\beta}', \tilde{a})'$. This is the smoothed version of the LMS estimator (β'_{LMS}, a_{LMS}) that solves (3,4). For this estimator we shall consider the limit process. Of course, the solution

depends on the function ψ; to simplify notation we suppress this dependence and will discuss it later in section 3.

Suppose that Assumption 1 is satisfied. We can write Taylor expansions of (7,8) at $\beta = \beta_0$ and $a = a_0$. For example for (7) one could have

$$F_n(\beta_0, a_0) + \frac{\partial F_n}{\partial \beta}(\beta^*, a^*)(\tilde{\beta} - \beta_0) + \frac{\partial F_n}{\partial a}(\beta^*, a^*)(\tilde{a} - a_0) = 0 \qquad (9)$$

where $\beta^* = \alpha\tilde{\beta} + (1 - \alpha)\beta_0$, for some $0 \le \alpha \le 1$ and a^* is similarly between a_0 and $\tilde{a}$.

For (8) similarly

$$\frac{\partial F_n}{\partial \beta}(\beta_0, a_0) + \frac{\partial^2 F_n}{\partial \beta \partial \beta'}(\beta^*, a^*)(\tilde{\beta} - \beta_0) + \frac{\partial^2 F_n}{\partial \beta \partial a}(\beta^*, a^*)(\tilde{a} - a_0) = 0 \qquad (10)$$

where β^* and a^* are defined similarly to β^*, a^* for (9) but may take different values; we keep the same notation for these "intermediate points"; this does not affect any of our results.

Define C_1, C_2 such that $C_1 > 0$ and $a_0 > C_2 > 0$. Consider a sequence of $k \times k$ matrices V_n and domains U_n defined by

$$U_n = \left\{ (\beta', a)' \in R^{k+1} \left| b_n^{\frac{1}{2}} \| V_n(\beta - \beta_0) \| < C_1, |a - a_0| < C_2 \right. \right\}.$$

We make the following general assumption.

Assumption 3. *As $n \to \infty$ there exists a (possibly random) sequence of $k \times k$ positive definite matrices V_n such that $b_n^{-\frac{1}{2}} V_n^{-1} = o_p(1)$ and*

(a) for $C_1 > 0$ if β satisfies $b_n^{\frac{1}{2}} \| V_n(\beta - \beta_0) \| < C_1$ and $(\beta, a(\beta))$ solve (7), then $|a(\beta) - a_0| = o_p(1)$;

for some constants $d_1(\psi), d_2(\psi) > 0$

(b) if $|a - a_0| < C_2$ then $b_n^{-\frac{1}{2}} V_n^{-1} \frac{\partial F_n}{\partial \beta}(\beta_0, a) \Rightarrow N(0, (d_1(\psi))^2 I_k)$;

(c) $\sup\limits_{U_n} \left(b_n^{-1} V_n^{-1} \frac{\partial^2 F_n}{\partial \beta \partial \beta'}(\beta, a) V_n^{-1} + d_2(\psi) I_k \right) = o_p(1), d_2(\psi) < 0$;

(d) $\sup\limits_{U_n} \left(b_n^{-\frac{1}{2}} V_n^{-1} \frac{\partial^2 F_n}{\partial \beta \partial a}(\beta, a) \right) = o_p(1)$;

(e) $\Pr\left(\sup\limits_{U_n} \frac{\partial F_n}{\partial a}(\beta, a) < \gamma \right) \to 1$ for some $\gamma < 0$.

10

Assumption 3(a) ensures that with probability going to 1 the solution to (7) is in the domain U_n; the conditions 3(b-d) applied to (9) provide the limit distribution for the β that solves (7,8). Condition 3(e) ensures that the solution to (8) actually solves the first-order condition for minimization of $a(\beta)$; note that

$$\frac{\partial^2 a}{\partial\beta\partial\beta'} = \frac{\frac{\partial F_n}{\partial\beta}\frac{\partial^2 F_n}{\partial a\partial\beta'}}{\left(\frac{\partial F_n}{\partial a}\right)^2} - \frac{\frac{\partial^2 F_n}{\partial\beta\partial\beta'}}{\frac{\partial F_n}{\partial a}}$$

and is positive definite in U_n with probability going to 1 by the conditions in Assumption 3. Appendix A provides the form of the derivatives of the function $F_n(\beta, a)$.

Theorem 1 *Under Assumptions 1, 2 (a,b) and 3 as $n \to \infty$ the solution $(\tilde{\beta}', \tilde{a})'$ to (7,8) exists in probability in U_n and is such that*

$$b_n^{\frac{1}{2}} V_n(\tilde{\beta} - \beta_0) \Rightarrow N(0, \frac{d_1^2(\psi)}{d_2(\psi)} I_k).$$

Proof. See Appendix B.

Theorem 1 can accommodate a variety of models with various types of distributional assumptions on the regressors and errors. The following theorem derives the asymptotic distribution of the smoothed LMS estimator for a set of specific assumptions on the regressor and error processes that are the same as in Pollard's Theorem 2 (1991) for the LAD estimator; they cover many cases of interest in models involving time series data.

Theorem 2 *Suppose that Assumption 2(a,a',b') holds and additionally*
(a) the variables $\{u_i = y_i - \beta_0 X_i\}$ are identically and independently distributed; the distribution of u has a symmetric bounded p.d.f. $f(u)$ that strictly monotonically declines from its median at zero and is continuously differentiable in a neighbourhood of the 75th percentile a_0;
(b) for $\{\mathcal{F}_i\}$ an increasing sequence of $\sigma-$fields and a sequence V_n of positive definite matrices (which could be random) $V_n^{-1}X_i'$ is $\mathcal{F}_{i-1}$ measurable for every i;
(c) $\max_{i\leq n}\|V_n^{-1}X_i'\| = o_p(1)$;
(d) $\sum_{i\leq n} V_n^{-1}X_i' X_i V_n^{-1} \to_p I_k$ where I_k is a $k \times k$ identity matrix;

(e) u_i is independent of $\mathcal{F}_{i-1}$ for every i.
Then the conditions of Theorem 1 are satisfied for

$$(d_1(\psi))^2 = f(a_0) \int \psi(w)^2 dw; \ d_2(\psi) = 2(f'(a_0) \int \psi(w)dw)^2.$$

Proof. See Appendix B.

Thus we get that under the conditions of Theorem 2

$$b_n^{\frac{1}{2}} V_n(\hat{\beta} - \beta_0) \Rightarrow N\left(0, \frac{f(a_0) \int \psi(w)^2 dw}{2(f'(a_0) \int \psi(w)dw)^2} I_k\right).$$

Similarly to the case of LAD in pollard (1991, Theorem 2) here also the independence conditions could be replaced by e.g. requiring stationarity and ergodicity of $\{(X_i, u_i)\}$, a martingale difference condition on u_i would suffice.

Models satisfying the theorem include the standard i.i.d. regression case and stationary AR(1) model as demonstrated by Pollard (1991); also, trending regressors are allowed; other cases of interest such as AR(p) dynamics in y can be similarly treated. Thus the theorem can apply to various models involving time series data.

Note that the slower the rate $b_n \to 0$ the closer one gets to standard rate asymptotics for the smoothed estimator; in Appendix C it is demonstrated that the smoothed estimator has the high breakdown point property in the finite sample.

The next section examines the issue of combining several smoothing functions and makes use of generalized Gaussian processes (in the sense of Gel'-fand and Vilenkin, 1964) to characterize the distribution of the LMS estimator.

3 Asymptotics for smoothed estimators combined and for the LMS estimator.

In most standard cases, such as when estimators solve differentiable conditions, or even as in the case of LAD where there is some non-differentiability, but the first derivatives are still ordinary functions (e.g., the sign function), the asymptotic distribution of a smoothed estimator does not depend on the smoothing function; here, however, as Theorems 1 and 2 show the smoothed

LMS estimator has an asymptotic distribution that does depend on the function ψ. It therefore makes sense to consider the asymptotic joint distribution of smoothed LMS estimators corresponding to different smoothing functions. The following Theorem 3 establishes the joint distribution under the assumptions of Theorem 2; a similar result could be established under the Assumptions of Theorem 1 and an additional assumption that would ensure (similarly to Assumption 3(b,c) which provide the variance) that the asymptotic covariances for different smoothing functions exist.

We now subscript the smoothed estimator of Section 2 by the smoothing function for which it was obtained, it is now denoted $\tilde{\beta}_\psi$. We define a transformation $\tilde{\eta}(\psi)$ of $\tilde{\beta}_\psi$, which proves to be a convenient representation, by

$$\tilde{\eta}(\psi) = V_n b_n^{1/2}(\tilde{\beta}_\psi - \beta_0) \int \psi(w)dw \tag{11}$$

In the following theorem we derive the asymptotic distribution of the $(lk) \times 1$ vector $(\tilde{\eta}(\psi_1), .., (\tilde{\eta}(\psi_l))'$ for a linearly independent set of functions $\{\psi_1, .., \psi_l\}$.

Theorem 3 *Under the conditions of Theorem 2 for each of the set of linearly independent functions $\{\psi_1, ..., \psi_l\}$ the joint distribution*

$$(\tilde{\eta}(\psi_1), .., (\tilde{\eta}(\psi_l))' \Rightarrow N(0, \Sigma)$$

where the $lk \times lk$ covariance matrix Σ can be represented as

$$\Sigma = \frac{f(a_0)}{2(f'(a_0))^2} \Psi \searrow I_k,$$

where the $l \times l$ matrix Ψ has elements

$$\{\Psi\}_{i,j} = \int \psi_i(w)\psi_j(w)dw.$$

Proof. See Appendix B.

Corollary Define $D = diag\left(\int \psi_1(w)dw I_k, ..., \int \psi_l(w)dw I_k\right)$. Then

$$\left(b_n^{1/2}\left[V_n(\tilde{\beta}_{\psi_1} - \beta_0)\right]', ..., b_n^{1/2}\left[V_n(\tilde{\beta}_{\psi_l} - \beta_0)\right]'\right)' \Rightarrow N(0, D^{-1}\Sigma D^{-1}).$$

Proof. Follows immediately from Theorem 3 and (11).

Thus we have the joint Gaussian distribution of the estimators smoothed with a variety of smoothing schemes. Note that this is not a standard case of smoothing in that by appropriately selecting smoothing functions e.g. such that $\int \psi_i(w)\psi_j(w)dw = 0$ we can obtain estimators of β_0 that are asymptotically independent (the limit process is joint Gaussian with zero covariances); combining several smoothed estimators that are asymptotically independent, e.g. taking their mean, will reduce the asymptotic variance of the smoothed estimator.

Next, consider the LMS estimator. Under the conditions of Theorem 2 the following theorem establishes the convergence rate for the LMS estimator; for example, in the case $V_n = O(n^{\frac{1}{2}})$ we get the well known convergence rate of $n^{-\frac{1}{3}}$.

Theorem 4 *Under the conditions of Theorem 2 with $n^{\frac{1}{6}}V_n^{-1} = o_p(1)$ the LMS estimator β_{LMS} is such that*

$$n^{-\frac{1}{6}}V_n(\beta_{LMS} - \beta_0) = O_p(1).$$

Proof. See Appendix B.

It follows that β_{LMS} consistently estimates β_0 with the same convergence rate as $\tilde{\beta}_\psi$ for $b_n = n^{-\frac{1}{3}}$. Thus for β_{LMS} with any ψ we can associate the smoothed version $\tilde{\beta}_\psi$ of the LMS estimator: $(\beta_{LMS}, \psi) = \tilde{\beta}_\psi$; the limit process of $\tilde{\beta}_\psi$ represents the smoothed limit process for the LMS estimator if $b_n = n^{-\frac{1}{3}}$ (or any $b_n = O(n^{-\frac{1}{3}})$).

Next we use the apparatus of generalized random processes of Gel'fand and Vilenkin (1964). Consider the linear space K_1 of continuously differentiable functions with finite support; the topology is given by convergence defined for a sequence of functions restricted to a common finite support by uniform convergence of the functions as well as their derivatives.

Recall the definition of $\tilde{\eta}(\psi)$ in (11); we start by representing the limit process for $\tilde{\eta}(\psi)$ via a generalized random process. We introduce a generalized random process η_G by defining for $\psi \in K_1$ the value of the random functional $\eta_G(\psi)$

$$\eta_G(\psi) = -\frac{1}{\sqrt{2}f'(a_0)}b_n^{-\frac{1}{2}}\sum\left[\psi(\frac{u_i - a_0}{b_n}) - \psi(\frac{u_i + a_0}{b_n})\right]V^{-1}X_i. \qquad (12)$$

14

The following theorem shows that the limit process for $\tilde{\eta}(\psi)$ coincides with that of $\eta_G(\psi)$. We show that the limit process for the random functional η_G defined by (12) represents a generalized Gaussian process and produce its mean and covariance functionals which by the results of Gel'fand and Vilenkin (1964) closely link it to the generalized derivative of the Wiener process.

Theorem 5 *Under conditions of Theorem 2 as $n \rightarrow \infty, b \rightarrow 0$ the limit process for $\tilde{\eta}(\psi)$ coincides with that of $\eta_G(\psi)$. The process η_G converges to a generalized Gaussian k-variate process $\eta_{lim} \sim GN(0, WI_k)$ with the expectation functional equal to 0 and covariance functional W defined by its action on a pair (ψ_1, ψ_2) of linearly independent functions from K_1 by*

$$W(\psi_1, \psi_2) = \frac{f(a_0)}{2(f'(a_0))^2} \int \psi_1(w)\psi_2(w)dw. \tag{13}$$

Proof. See Appendix B.

In Gel'fand, Vilenkin (1964, pp. 257-260) the generalized Gaussian process that represents the derivative of the Wiener process is described via its covariance functional $B(\psi_1, \psi_2) = \int \psi_1(w)\psi_2(w)dw$; thus from (13) we see that η_{lim} is proportionate to the generalized derivative of the Wiener process.

From the relation (11) it follows that $\eta_{lim}(\psi)$ represents the limit process of $V_n b_n^{1/2}(\tilde{\beta}_\psi - \beta_0) \int \psi(w)dw$. For $b_n = n^{-\frac{1}{3}}$ it follows from Theorem 4 that $V_n b_n^{1/2}(\tilde{\beta}_\psi - \beta_0)$ represents $V_n b_n^{1/2}(\beta_{LMS} - \beta_0)$ smoothed via the function ψ; the corresponding limit process for $V_n b_n^{1/2}(\tilde{\beta}_\psi - \beta_0)$ provides the limit process for the LMS estimator smoothed via ψ. Thus the limit process for the LMS estimator here is indexed by functions ψ and the limit process for the LMS estimator is related to the generalized derivative of the Wiener process.

4   Conclusions

The conclusions of this paper relate to two issues. One set of conclusions can be stated for the high breakdown point estimators examined here; the second for the methodology utilized here to deal with non-differentiability.

For the high breakdown point estimator that solves smoothed LMS conditions for a particular function ψ we provide the asymptotic Gaussian distribution under conditions applicable in time series contexts. The rate of convergence of this estimator depends on the speed with which the support of the function shrinks to zero; the slower the rate of this bandwidth parameter, the faster the convergence to the limit process.

Next, combining several smoothed estimators for, e.g. some set of basis functions in the space K_1 may lead to a more efficient estimation procedure with Gaussian asymptotics.

Thirdly, the smoothing of the LMS estimator by the methods here (that require that asymptotically the appropriate rate be maintained) provides a description of the limit process of the LMS estimator via a generalized Gaussian process related to the derivative of the Wiener process.

The approach here can be extended for use in other situations where the asymptotic processes are difficult as a result of non-differentiability of the functions defining the estimators. Note that the same method can be used under distributional assumptions that differ from the ones employed here. Other estimators such as the least quantiles of squares estimators (other than the median) or Manski's maximum score estimator (see e.g. Horowitz, 1992 for an analysis of a smoothed version of that estimator) can similarly be treated.

The issues of computation and finite sample performance of the smoothed estimators and their combinations will need to be explored. Computational difficulties are similar to those for the smoothed maximum score estimator and require similarly a global optimization method such as e.g. generalized simulated annealing as suggested by Horowitz.

Appendix A. Derivatives of the function $F_n(\beta, a)$.

Without loss of generality we shall assume that a is bounded away from zero by some positive θ. Assume that $b_n < \theta$ and partition the set of integers $\{1, 2, ...n\}$ into three non-intersecting sets $T_n(-) \cup T_n(+) \cup T_n^c$, where

$$
\begin{aligned}
T_n(-) &= \{i : |y_i - X_i\beta - a| \le b_n\}, & \text{(A1)} \\
T_n(+) &= \{i : |y_i - X_i\beta + a| \le b_n\}, \\
T_n^c &= \{1, 2, ...n\} \setminus [T_n(-) \cup T_n(+)].
\end{aligned}
$$

Denote by T_n the union $T_n(-) \cup T_n(+)$.

Then due to the fact that $T_n(+)$ and $T_n(-)$ do not intersect and by changing the variable in integration one can represent the terms in the right-hand side of (6) for $F_n(\beta, a)$ as (recall that support of ψ is in [-1,1])

$$\int sgn((y_i - X_i\beta - v)^2 - a^2)\psi(\tfrac{v}{b_n})dv = b_n \int sgn((\tfrac{y_i - X_i\beta}{b_n} - w)^2 - \tfrac{a^2}{b_n^2})\psi(w)dw =$$
$$\begin{cases} b_n sgn((y_i - X_i\beta)^2 - a^2) \int \psi(w)dw \text{ if } i \in T_n^c; \\ -b_n \int sgn(\tfrac{y_i - X_i\beta + a}{b_n} - w)\psi(w)dw \text{ if } i \in T_n(+); \\ b_n \int sgn(\tfrac{y_i - X_i\beta - a}{b_n} - w)\psi(w)dw \text{ if } i \in T_n(-). \end{cases}$$

$$(A2)$$

Write $F_n(\beta, a) = F(T_n^c) + F(T_n)$, where $F(T_n^c)$ contains all the terms with $i \in T_n^c$ and $F(T_n)$ all the terms with $i \in T_n$. The function $F_n(\beta, a)$ is differentiable. Indeed, in $F(T_n^c)$ the terms are constants equal to $\pm b_n \int \psi(w)dw$ since for those $|(y_i - X_i\beta)^2 - a^2| > b_n^2$ and the sign of $(y_i - X_i\beta - v)^2 - a^2$ does not change; thus their derivative is zero. By differentiating the terms in $F(T_n)$ in (A2) the result can be written as

$$\frac{\partial F_n}{\partial \beta} = -2\sum \left[\psi(\tfrac{y_i - X_i\beta - a}{b_n}) - \psi(\tfrac{y_i - X_i\beta + a}{b_n})\right] X_i';$$
$$\frac{\partial F_n}{\partial a} = -2\sum \left[\psi(\tfrac{y_i - X_i\beta - a}{b_n}) + \psi(\tfrac{y_i - X_i\beta + a}{b_n})\right].$$

$$(A3)$$

Note that summation here can be assumed to be from $i = 1$ to $i = n$ since for $i \notin T_n$ the values of the function in the sum are zero; note also that if $\frac{\partial F_n}{\partial a} \neq 0$, it is negative.

Next we introduce the following notation:

$$A_i = \psi(\frac{u_i - a}{b_n}) - \psi(\frac{u_i + a}{b_n}); \quad A_i^0 = \psi(\frac{u_i - a_0}{b_n}) - \psi(\frac{u_i + a_0}{b_n});$$
$$B_i = \psi'(\frac{u_i - X_i\eta - a}{b_n}) - \psi'(\frac{u_i - X_i\eta + a}{b_n});$$
$$C_i = \psi'(\frac{u_i - X_i\eta - a}{b_n}) + \psi'(\frac{u_i - X_i\eta + a}{b_n}),$$

$$(A4)$$

where $\eta = \beta - \beta_0$. By taking the appropriate derivatives we obtain

$$\frac{\partial F}{\partial \beta}(\beta_0, a_0) = -2\sum A_i^0 X_i';$$
$$\frac{\partial^2 F}{\partial \beta \partial \beta'}(\beta^*, a^*) = 2\frac{1}{b_n}\sum B_i X_i X_i',$$
$$\frac{\partial^2 F}{\partial \beta \partial a}(\beta^*, a^*) = 2\frac{1}{b_n}\sum C_i X_i'.$$

$$(A5)$$

17

Appendix B. Proofs of the theorems.

Proof of Theorem 1.

Consider first Assumption 3(b); it implies that $b_n^{-\frac{1}{2}} V_n^{-1} \frac{\partial F}{\partial \beta}(\beta_0, a_0)$ is bounded in probability. For $\varepsilon > 0$ then find c such that

$$\Pr\left(\left\| b_n^{-\frac{1}{2}} V_n^{-1} \frac{\partial F}{\partial \beta}(\beta_0, a_0) \right\| > c d_2(\psi)\right) \leq \varepsilon$$

as $n \to \infty$. Construct the U_n of Assumption 3 for some $C_1 > c$. Consider the pairs $(\beta, a(\beta)) \in U_n$ that solve (7); we show that as $n \to \infty$ the probability that some $(\tilde{\beta}, a(\tilde{\beta}))$ solves (8) is at least $1 - 2\varepsilon$.

Examine

$$\frac{\partial F}{\partial \beta}(\beta, a(\beta)) = \frac{\partial F}{\partial \beta}(\beta_0, a_0) + \frac{\partial^2 F}{\partial \beta \partial \beta'}(\beta^*, a^*)(\beta - \beta_0) + \frac{\partial^2 F}{\partial \beta \partial a}(\beta^*, a^*)(a(\beta) - a_0)$$

in U_n. If we multiply both sides by $b_n^{-\frac{1}{2}} V_n^{-1}$ we can write

$$b_n^{-\frac{1}{2}} V_n^{-1} \frac{\partial F}{\partial \beta}(\beta, a(\beta)) = \xi_n + d_2(\psi) V_n b_n^{\frac{1}{2}}(\beta - \beta_0)$$

where ξ_n is such that $\Pr\left(sup \|\xi_n\| > d_2(\psi)c\right) \leq 2\varepsilon$ for large enough n by using the limits in Assumption 3(a-d). The continuous function $b_n^{-\frac{1}{2}} V_n^{-1} \frac{\partial F}{\partial \beta}(\beta, a(\beta))$ maps the closed convex set $B_n = \left\{\beta : b_n^{\frac{1}{2}} V_n \|\beta - \beta_0\| \leq c\right\} \subset R^k$ into a closed domain in R^k with the boundary ∂B_n mapped into the boundary $\xi_n + d_2(\psi)c\lambda$, where $\lambda = c^{-1} b_n^{\frac{1}{2}} V_n(\beta - \beta_0)$ so that $\lambda'\lambda = 1$. By construction with probability $1 - 2\varepsilon$ zero is in the interior of the image, thus for some $\tilde{\beta}, a(\tilde{\beta})$ we get $\frac{\partial F}{\partial \beta}(\tilde{\beta}, a(\tilde{\beta})) = 0$ and $a(\tilde{\beta}) = \tilde{a}$.

Next we consider the expansion in (10); by the limits provided in Assumption 3 (a-d) the rest of the statement of Theorem 1 follows.∎

Proof of Theorem 2.

The proof consists of verifying that Assumptions 1 and 3 are satisfied with the appropriate constants; then the result will follow from Theorem 1. Assumption 1 clearly holds in view of condition (a).

We establish the asymptotic limits of Assumption 3 with

$$
(d_1(\psi))^2 = 2f(a_0) \int \psi(w)^2 dw;
$$

$$
d_2(\psi) = 2 \left(f'(a_0) \int \psi(w)dw \right)^2
$$

by proving the following sequence of statements.

(i) $F(\beta_0, a_0) = \begin{cases} O_p(b_n^3 n) & \text{if } b^3 n \to \infty; \\ O_p(b_n n^{\frac{1}{2}}) & \text{if } b^3 n = O(1); \end{cases}$

$(ii) \dfrac{-b_n^{-\frac{1}{2}} \sum A_i V_n^{-1} X_i}{\sqrt{2f(a) \int \psi(w)^2 dw}} \Rightarrow N(0, I)$ uniformly over $\alpha_1 < a < \alpha_2 < \infty$;

(iii) $\sup\limits_{U_n} \left(b_n^{-1} n^{-1} \frac{\partial F}{\partial a}(\beta, a) + 2f(a) \right) = o_p(1);$

(iv) For $(\beta, a(\beta))$ that solves (7), $a(\beta) - a_0 = \begin{cases} O_p(b_n^2) & \text{for } b_n n^{1/3} \to \infty \\ O_p\left(n^{-1/2}\right) & \text{for } b_n n^{1/3} = O(1). \end{cases}$

(v) $\sup\limits_{U_n} \left[b_n^{-1} V_n^{-1} \sum B_i X_i X_i' V_n^{-1} + \left(2f'(a) \int \psi(w)dw\right) I_k \right] = o_p(1);$

(vi) $\sup\limits_{U_n} b_n^{-1} V_n^{-1} \sum C_i X_i = o_p(1).$

To simplify notation we omit the subscript n in b_n.

(i) Consider the terms in $F(\beta_0, a_0)$ in (A2).

For terms with $i \notin T$ in $F(T^c)$ we get

$E(F(T^c)) = b \int \psi(w)dw \Sigma E(sgn(u_i^2 - a_0^2)) = 0$; since a_0^2 is the median of u_i^2 and $var(F(T^c)) = b^2 \left(\int \psi(w)dw \right)^2 \Sigma E(I\{(u_i^2 - a_0^2) > b^2\})$ where $I\{\cdot\}$ is an indicator function. Since

$$
E(I\{(u_i^2 - a_0^2) \le b^2\}) = \int_{a_0-b}^{a_0+b} f(u)du + \int_{-a_0-b}^{-a_0+b} f(u)du
$$

can be bounded from below by $4bf(a_0+b)$ and from above by $4bf(a_0-b)$, the right-hand side of the expression for the variance can be bounded from below by $b^2 \left(\int \psi(w)dw \right)^2 n(1-4bf(a_0-b))$ and from above by $b^2 \left(\int \psi(w)dw \right)^2 n(1- 4bf(a_0+b))$. By Chebyshev's inequality $F(T^c) = O_p(bn^{\frac{1}{2}})$.

Next, for each term in $F(T)$ we can write (utilizing the symmetry of ψ)

$$
\int sgn(\tfrac{u_i-a}{b} - w)\psi(w)dw - \int sgn(\tfrac{u_i\pm a}{b} - w)\psi(w)dw =
$$
$$
2[\Psi(\tfrac{u_i-a}{b}) - \Psi(-\tfrac{u_i\pm a}{b})], \text{ where } \Psi(w) = \begin{cases} \int_0^w \psi(v)dv & \text{if } |w| < 1, \\ 0 & \text{otherwise.} \end{cases}
$$

19

Note that the support of the function Ψ is in $[-1, 1]$. Since for $i \notin T$ the value of Ψ is zero we can assume that summation of such terms in $F(T)$ goes from 1 to n, and

$$E(2b(\Psi(\tfrac{u_i-a}{b}) - \Psi(\tfrac{u_i+a}{b}))) =$$
$$2b \int \Psi(\tfrac{u-a}{b})f(u)du - 2b \int \Psi(\tfrac{u+a}{b})f(u)du$$
$$= 2b^2 \int \Psi(w)[f(bw+a) - f(bw-a)]dw$$
$$= 2b^3 \int \Psi(w)[f'(a^*) - f'(-a^*)]wd$$
$$= 4b^3 f'(a) \int \psi(w)dw + o(b^3);$$

with $|a^* - a_0| < 2b$ thus the difference $E(F(T)) - 4b^3 n f'(a) \int \Psi(w)wdw = o(nb^3)$ and if $nb^3 \to 0$ this difference goes to zero faster, if nb^3 is bounded the difference goes to zero, finally, if nb^3 diverges to infinity, the difference is either bounded or diverges at a slower rate.

Next

$$E(2b(\Psi(\tfrac{u-a}{b}) - \Psi(\tfrac{u+a}{b})))^2 = 4b^2 E(\Psi(\tfrac{u-a}{b})^2 - 2\Psi(\tfrac{u-a}{b})\Psi(\tfrac{u+a}{b}) + \Psi(\tfrac{u+a}{b})^2) =$$
$$4b^3[\int \Psi(w)^2 f(bw+a)dw + \int \Psi(w)^2 f(bw-a)dw]$$
$$\to 8b^3 \int \Psi(w)^2 dw f(a).$$

The second equality follows from noting that $\Psi(\tfrac{u-a}{b})\Psi(\tfrac{u+a}{b}) = 0$ in the domain since $\Psi(\tfrac{u-a}{b})$ and $\Psi(\tfrac{u+a}{b})$ cannot be non-zero simultaneously due to the restriction on support of Ψ (recall that $b \to 0$) and the last from change of variable to $-w$ and symmetry of the p.d.f. $f(u)$; combining the above results we get the variance for each term as

$$4b^3 \left(\int \Psi(w)^2 f(bw+a)dw + \int \Psi(w)^2 f(bw-a)dw \right) -$$
$$\left(2b^2 \int \Psi(w)[f(bw+a) - f(bw-a)]dw \right)^2.$$

By independence of u_i covariance between terms is zero. Thus taking into account $E(F(T))$ we get that $varF(T) = \begin{cases} O(nb^3) \text{ if } nb^3 = O(1); \\ O(n^2 b^6) \text{ if } nb^3 \to \infty \end{cases}.$

The statement (i) follows from combining the rates.

(ii) First consider

$$E(\psi(\tfrac{u_i-a}{b})) = \int \psi(\tfrac{u-a}{b})f(u)du = b \int \psi(w)f(bw+a)dw =$$
$$b \int \psi(-w)f(-bw-a)dw = b \int \psi(w)f(bw-a)dw$$

20

for any a by symmetry of the functions ψ and f; by change of variable we see that it equals $E(\psi(\frac{u_i+a}{b}))$; $E(A_i) = 0$ follows immediately and thus also $E(\sum A_i X_i) = 0$. Similarly,

$$E(A_i^2) = E\left(\psi(\frac{u_i - a}{b})\right)^2 + E\left(\psi(\frac{u_i + a}{b})\right)^2 = 2b \int \psi(w)^2 f(bw + a)dw$$

since in A^2 the product term $\psi(\frac{u_i-a}{b})\psi(\frac{u_i+a}{b}) = 0$ as a result of the constraints on the support of ψ and $b < a$.

Consider next $\xi_{in} = b_n^{-\frac{1}{2}}\frac{A_i\lambda' V_n^{-1} X_i'}{\sqrt{2\int \psi(w)^2 f(bw+a)dw}}$ and

$$\zeta_{in} = \xi_{in} I\{max \left\|V_n^{-1} X_i'\right\| < M_n\}$$

where $M_n = o(1)$ and $\lambda'\lambda = 1$. Note that $sup |\xi_{in} - \zeta_{in}| \to_p 0$ by condition (c), thus to derive the limit distribution it is sufficient to consider ζ_{in}. We verify that ζ_{in} satisfies all the conditions of the martingale difference Central Limit Theorem of McLeish (see, e.g. Bierens (1994), Theorem 6.1.6). Indeed $E(\zeta_{in} | \mathcal{F}_{i-1}) = 0$;

$$supE(max |\zeta_{in}|^2) \leq supE(b_n^{-1}M_n^2 max A_i^2) < \infty,$$

moreover it goes to 0 since $E(A_i^2)$ is bounded by $2bmax\psi^2$ and $M_n^2 \to 0$. Next, similarly

$$max |\zeta_{in}| < M_n\frac{max |\psi|}{\sqrt{2\int \psi(w)^2 f(bw + a)dw}} \to 0.$$

Finally we need to show that $\Sigma\zeta_{in}^2 \to_p 1$. By condition (c)

$$\Pr\left(\left|\Sigma\zeta_{in}^2 - \Sigma\xi_{in}^2\right| > \varepsilon\right) = \Pr(max \left\|V_n^{-1} X_i'\right\| > M_n)$$

goes to zero thus if we show $\Sigma\xi_{in}^2 \to_p 1$ the required convergence will follow.

Consider $\sum \gamma_{in}$, where

$$\gamma_{in} = \xi_{in}^2 - \lambda' V_n^{-1} X_i' X_i V_n^{-1}\lambda = \left(\frac{b_n^{-1} A_i^2}{2\int \psi(w)^2 f(bw + a)dw} - 1\right)\lambda' V_n^{-1} X_i' X_i V_n^{-1}\lambda;$$

$\rho_{in} = \gamma_{in} I\{max \left\|V_n^{-1} X_i'\right\| < M_n\}$. We have that $E(\rho_{in} | \mathcal{F}_{i-1}) = 0$ thus the ρ_{in} are uncorrelated for each n. It is easy to see that $E(\Sigma\rho_{in})^2 = E(\Sigma\rho_{in}^2) =$

$O(M_n^2)$, thus by Chebychev's inequality $\sum \rho_{in} \to_p 0$ and consequently $\sum \gamma_{in} \to_p 0$. Since

$$\lambda' V_n^{-1} X_i' X_i V_n^{-1} \lambda \to_p 1$$

by condition (d), we get $\Sigma \xi_{in}^2 \to_p 1$ and thus all the conditions of the martingale Central Limit Theorem are satisfied for $\sum \zeta_{in}$.

The $\sum \xi_{in}$ has the same limit distribution;

$$2 \int \psi(w)^2 f(bw+a)dw \to 2f(a) \int \psi(w)^2 dw$$

and by the Cramer-Wold device we get (ii).

(*iii*) For $\psi(\frac{u_i - X_i \eta \pm a}{b_n})$ conditional expectation exists for any lth power of this function and can be written by change of variable as

$$
\begin{aligned}
&E(\psi^l(\tfrac{u_i - X_i \eta \pm a}{b_n}) \,|\mathcal{F}_{i-1}) = \\
&b_n \int \psi^l(w - \tfrac{X_i \eta}{b_n}) f(b_n w \mp a)dw
\end{aligned}
\tag{B1}
$$

and is bounded from below by

$$b_n f(b_n + a) \int \psi^l(w)dw$$

and similarly from above by

$$b_n f(-b_n + a) \int \psi^l(w)dw$$

with probability going to 1.

For terms in $\frac{\partial F}{\partial a}(\beta, a) = -2 \sum \left[\psi(\frac{y_i - X_i \beta - a}{b_n}) + \psi(\frac{y_i - X_i \beta + a}{b_n}) \right]$ define

$$\zeta_{in} = \left[\psi(\frac{u_i - X_i \eta - a}{b_n}) + \psi(\frac{u_i - X_i \eta + a}{b_n}) \right].$$

We have

$$2b_n f(b_n + a) \int \psi(w)dw < E(\zeta_{in}) < 2b_n f(-b_n + a) \int \psi(w)dw$$

and since $b_n \to 0$ and f is continuous we get that

$$sup \left| E(\zeta_{in}) - 2b_n f(a) \int \psi(w)dw \right| \to 0.$$

22

Next, using (B1) we evaluate similarly

$$E(\zeta_{in}^2 \,|\mathcal{F}_{i-1}) = b_n \int \psi^2(w - \frac{X_i\eta}{b_n})\,(f(b_nw + a) + f(b_nw - a))\,dw; \qquad (B2)$$

$$E(\zeta_{in}\zeta_{jn}\,|\mathcal{F}_{j-1}) = b_n^2 \int \psi(w - \frac{X_i\eta}{b_n})\,(f(b_nw + a) + f(b_nw - a))\,dw \times$$
$$\int \psi(w - \frac{X_j\eta}{b_n})\,(f(b_nw + a) + f(b_nw - a))\,dw \qquad (B3)$$

for $i < j$. It is easy to see that $var(\frac{1}{n}\sum \zeta_{in}) \to 0$ since $b_n \to 0$. By Chebyshev's inequality thus

$$\left|\frac{1}{n}\sum \zeta_{in} - 2b_n f(a)\int \psi(w)dw\right| \to_p 0$$

and (iii) follows.

(iv) Consider $\beta : b_n\,\|V_n(\beta - \beta_0)\| < C_1$ and the expansion for $F(\beta, a(\beta)) = 0$ in (9). The order of the first term in that expansion is given by (i); from (A5) and the result in (ii) we get that the second term is $O_p(1)$, finally the coefficient on $(a(\beta) - a_0)$ is examined in (iii) from where it follows that it is $\approx O_p(b_n n)$. By substituting these rates into (9) the result follows.

(v) The expression in braces in statement (v) is a $k \times k$ symmetric matrix; it depends on (β, a) in U_n via B_i defined in (A4). We find conditional expectation for B_i. First calculate conditional expectation $E(\psi'(\frac{u_i - X_i\eta \pm a}{b})\,|\mathcal{F}_{i-1})$; it is

$$\int \psi'(\frac{u_i \pm a - X_i\eta}{b})f(u)du = -b^2 \int \psi(w - \frac{X_i\eta}{b})f'(bw \mp a)dw;$$

thus

$$E(B_i\,|\mathcal{F}_{i-1}) = -b^2 \int \psi(w - \frac{X_i\eta}{b_n})\,[f'(bw - a) - f'(bw + a)]\,dw. \qquad (B4)$$

Similarly conditional variance and covariances for B_i, B_j can be obtained.

Define $\xi_{in}(lm) =$

$$\left(b^{-1}B_i + b\int \psi(w - \frac{X_i\eta}{b_n})\,[f'(bw - a) - f'(bw + a)]\,dw\right)$$
$$\lambda'V_n^{-1}X_i'X_iV_n^{-1}\lambda;$$

23

with the vector λ with ones for the lth and mth components and zeros otherwise. Also define

$$\zeta_{in}(lm) = \xi_{in}(lm)I\{max\left\|V_n^{-1}X_i'\right\| < M_n\}, \text{ with } M_n = o(1);$$

we shall show that $\sum \zeta_{in}(lm)$ has the same probability limit as the sum of the lm and ml elements of the symmetric matrix in braces in (v) and also that this sum is $o_p(1)$ uniformly over all β. Note that from (B4) it follows that $E\zeta_{in}(lm) = 0$.

First by continuity and skew-symmetry of the function f' we see that

$$\int \psi(w - \frac{X_i\eta}{b_n})\left[f'(bw - a) - f'(bw + a)\right]dw - 2f'(a)\int \psi(w)dw$$

has a uniform $o_p(1)$ bound; thus

$$\Sigma\zeta_{in}(lm) = \Sigma\left[\left(b^{-1}B_i + b2f'(a)\int \psi(w)dw + o_p(b)\right)\lambda'V_n^{-1}X_i'X_iV_n^{-1}\lambda\right]$$
$$I\{max\left\|V_n^{-1}X_i'\right\| < M_n\}.$$

Recalling condition (d) the expression in braces can be written as

$$\Sigma\left(\left(b^{-1}B_i + o_p(b)\right)\lambda'V_n^{-1}X_i'X_iV_n^{-1}\lambda\right) + b2f'(a)\int \psi(w)dw(\lambda'\lambda + o_p(1))$$

thus one can see that the difference between $\Sigma\zeta_{in}(lm)$ and the sum of the lm and ml elements of the symmetric matrix in braces in (v) is $o_p(1)$.

Next similarly to the proof in (iii) we can verify that $var(\sum \zeta_{in}(lm)) \to 0$. By Chebyshev's inequality $\sum \zeta_{in}(lm) \to_p 0$ and (v) obtains.

(vi) Next consider

$$\xi_{in} = b_n^{-1}C_iV_n^{-1}X_i$$

and $\zeta_{in} = \xi_{in}I\{max\left\|V_n^{-1}X_i'\right\| < M_n\}$,with $M_n = o(1)$. We show that $\Sigma\zeta_{in} = o_p(1)$ uniformly; this will provide the result in statement (vi).

Similarly to $E(B_i\,|\mathcal{F}_{i-1})$ in (B4) we get $E(C_i\,|\mathcal{F}_{i-1}) =$

$$-b^2\int \psi(w - \frac{X_i\eta}{b})\left[f'(bw - a) + f'(bw + a)\right]dw$$

and corresponding expressions for expectations of squares and products of C_i; note that

$$-b^2\int \psi(w - \frac{X_i\eta}{b})\left[f'(bw - a) + f'(bw + a)\right]dw = o(b^2)$$

24

uniformly since $f'(a) + f'(-a) = 0$. Consider $E(\Sigma \zeta_{in})^2$, similarly to previous proofs by evaluating the terms we can show that this expression goes to zero; the result follows from Chebyshev's inequality.

To conclude the proof of Theorem 2 comparing (i-vi) with Assumption 3 we get (a) from (iv), (b) from (ii), (c) from (v); (d) from (vi) and (e) from (iii). Note that the limit results in (iii), (v) and (vi) hold over all β, not just over U_n.■

Proof of Theorem 3.

Consider without loss of generality two linearly independent functions, ψ_s and ψ_t and $\tilde{\eta}(\psi_s), \tilde{\eta}(\psi_t)$. If we consider jointly the system of equations for the estimators corresponding to the two functions and their expansions in view of the results in Theorem 2 it follows that the joint limit distribution for $\tilde{\eta}(\psi_s), \tilde{\eta}(\psi_t)$ is Gaussian with mean zero; to find its covariance matrix all that we need to establish are the covariances between the A_i and A_j for the different functions ψ_s and ψ_t. If $i \neq j$ such covariance is zero since u_i, u_j are independent. Consider $E(A_i(\psi_s)A_i(\psi_t))$. It is

$$
E\left(\psi_s(\frac{u_i - a}{b_n}) - \psi_s(\frac{u_i + a}{b_n})\right)\left(\psi_t(\frac{u_i - a}{b_n}) - \psi_t(\frac{u_i + a}{b_n})\right)
$$
$$
= \; b_n \int \psi_s(w)\psi_t(w)f(b_n w + a)dw + b_n \int \psi_s(w)\psi_t(w)f(b_n w - a)dw,
$$

where the result is obtained by a change of variable and by noting that terms vanish due to restrictions on the support of ψ. By symmetry of the p.d.f. as $b_n \to 0$

$$
b_n \int \psi_s(w)\psi_t(w)f(b_n w + a)dw + b_n \int \psi_s(w)\psi_t(w)f(b_n w - a)dw
$$
$$
= \; 2b_n f(a) \int \psi_s(w)\psi_t(w)dw + O(b_n^2).
$$

By substituting the the limits obtained in the proof of Theorem 2 the covariances between the different components, ith of the vector $\tilde{\eta}(\psi_s)$ and jth of $\tilde{\eta}(\psi_t)$ are zero if $i \neq j$ and $\frac{f(a)\int \psi_s(w)\psi_t(w)dw}{2f'(a)^2}$ for $i = j$. The result follows.■

Proof of Theorem 4.

Consider for some ψ of Assumption 2 the corresponding $F_n(\beta, a)$. Recall that for any β the vector $(\beta', a(\beta))'$ solves $F_n(\beta, a) = 0$ and $(\tilde{\beta}', a(\tilde{\beta}))'$ additionally solves $\frac{\partial F}{\partial \beta} = 0$; $\tilde{\beta}$ is the minimizer of $a(\beta)$. For β denote by a_β^2 the

median of squared residuals, $\{(y_i - X_i\beta)^2\}$; then β_{LMS} is the minimizer of a_β^2, the minimized value of a_β is a_{LMS}.

Consider now a (random) convex domain

$$U_{n,LMS} = \left\{ (\beta, a) : n^{-\frac{1}{6}} \|V_n(\beta - \beta_{LMS})\| < C_1; |a - a_0| < C_2 \right\}.$$

We establish first that
(1) $|a_\beta - a(\beta)| = O_p(n^{-\frac{1}{2}})$ in $U_{n,LMS}$;
(2) $\left| a_{LMS} - a(\tilde{\beta}) \right| = O_p(n^{-\frac{1}{2}})$.
To prove (1) consider in $U_{n,LMS}$

$$F_n(\beta, a_\beta) = \frac{\partial F}{\partial a}(\beta, a^*)(a_\beta - a(\beta)) \tag{B5}$$

noting that $F_n(\beta, a(\beta)) = 0$ by definition of $a(\beta)$.

For $F_n(\beta, a_\beta)$ when $b_n^3 n = O(1)$ similarly to the proof of (i) for Theorem 2 we establish that $F_n(\beta, a_\beta) = O_p(n^{\frac{1}{6}})$. Consider the terms in (A2). Since a_β^2 is the median of squared residuals we have $\Sigma sgn\left((u_i - X_i\eta)^2 - a_\beta^2\right) = 0$ and we can write $F_n(\beta, a_\beta) =$

$$b \sum_{i \in T_n} \int \left(sgn(\frac{u_i - X_i\eta - a}{b} - w) - sgn(\frac{u_i - X_i\eta + a}{b} - w) \right) \psi(w)dw -$$

$$b \sum_{i \in T_n} sgn\left((u_i - X_i\eta)^2 - a_\beta^2\right) \int \psi(w)dw.$$

We get that $E(F_n(\beta, a_\beta)) = O(b_n^3 n)$ as in the proof of (i). However the squared terms $E\left(bsgn\left((u_i - X_i\eta)^2 - a_\beta^2\right)\int \psi(w)dw\right)^2$ dominate $E(F_n(\beta, a_\beta))^2$ which as a result is $O(b^2 n)$. This leads to $F_n(\beta, a_\beta) = O_p(bn^{\frac{1}{2}})$, here $O_p(n^{\frac{1}{6}})$. We can also similarly to (iii) of Theorem 2 show that $\frac{\partial F}{\partial a}(\beta, a)$ is bounded from above by a negative constant in $U_{n,LMS}$. By (B5) we get that (1) holds.

For any β we have $a(\beta) - a(\tilde{\beta}) \geq 0$ since $\tilde{\beta}$ is the minimizer of $a(\beta)$; for any β by the LMS property $a_\beta - a_{LMS} \geq 0$. We use proof by contradiction to show that (2) holds.

Suppose that (the always non-negative) $n^{\frac{1}{2}}\left(a(\beta_{LMS}) - a(\tilde{\beta})\right)$ is unbounded in probability, in other words, $\Pr(n^{\frac{1}{2}}\left(a(\beta_{LMS}) - a(\tilde{\beta})\right) > B_n) > \varepsilon$ for some $B_n \to \infty$ and some $\varepsilon > 0$. It follows from $|a_{LMS} - a(\beta_{LMS})| = O_p(n^{-\frac{1}{2}})$ that

26

$\Pr(n^{\frac{1}{2}}\left(a_{LMS} - a(\tilde{\beta})\right) > B_n) > \varepsilon$ and that $\Pr(n^{\frac{1}{2}}\left(a_{LMS} - a_{\tilde{\beta}}\right) > B_n) > \varepsilon$, implying that there are cases when $a_{LMS} > a_{\tilde{\beta}}$, which contradicts the definition of LMS, thus (2) holds.

Suppose that $\lim_{n\to\infty} \Pr\left(U_{n,LMS} \cap U_n = \emptyset\right) > 0$ for any C_1. It follows from the proof of Theorem 1 that outside of U_n the value $b_n^{-\frac{1}{2}} V_n^{-1} \frac{\partial F}{\partial \beta}(\beta, a(\beta))$ can be bounded away from zero uniformly in probability; additionally we have that $\frac{\partial F}{\partial a}(\beta, a)$ is bounded from above in $U_{n,LMS}$ by a negative constant. Thus for $\frac{\partial a}{\partial \beta} = -\frac{\frac{\partial F}{\partial \beta}}{\frac{\partial F}{\partial a}}$ the event

$$\inf_{U_{n,LMS}} \left\| n^{\frac{1}{6}} V_n^{-1} \frac{\partial a}{\partial \beta}(\beta) \right\| > p > 0$$

similarly has non-zero probability as $n \to \infty$. For some $\hat{\beta}$ on the boundary of the domain we get $a(\beta_{LMS}) - a(\hat{\beta}) > pC_1$ with nonzero probability; since $\left| a(\beta_{LMS}) - a(\tilde{\beta}) \right| = o_p(n^{-\frac{1}{2}})$ this implies that there are cases when $a(\hat{\beta}) < a(\tilde{\beta})$ which contradicts the fact that $\tilde{\beta}$ minimizes $a(\beta)$. Thus $U_{n,LMS} \cap U_n \neq \emptyset$ with probability that goes to 1; and β_{LMS} is close enough in probability asymptotically to $\tilde{\beta}$ and to β_0 and the result of Theorem 4 follows.∎

Proof of Theorem 5.

Consider the random vector $\sqrt{2} f'(a_0) \tilde{\eta}(\psi)$ which by definition of $\tilde{\eta}(\psi)$ (11) equals $\sqrt{2} f'(a_0) V b^{\frac{1}{2}}(\tilde{\beta}_\psi - \beta_0) \int \psi(w) dw$. Under the conditions of Theorem 2 in its proof in Appendix B the limits of the different terms in the expansion (10) were obtained. It can be seen that $\sqrt{2} f'(a_0) \tilde{\eta}(\psi)$ has the same limit process as $-b_n^{-\frac{1}{2}} \sum A_i V_n^{-1} X_i$ which by the definition of A_i in (A4) equals

$$-b^{-\frac{1}{2}} \sum \left[\psi(\frac{u_i - a_0}{b_n}) - \psi(\frac{u_i + a_0}{b_n}) \right] V^{-1} X_i.$$

Consider (12). First we note that this is a random continuous linear functional on the space K of infinitely differentiable functions (of course, it is also a linear continuous functional on the space $K_1 \supset K$). From Theorem 3 if follows that η_G is a generalized Gaussian random process in the sense of Gel'fand, Vilenkin (1964). Thus by their Theorem 1 (pp. 250-251) this process is determined by the bilinear covariance functional which here from

Theorem 3 is W given by (13). This result extends to functions defined on the space $K_1 \supset K$ since the functionals that need to be considered here in the derivations of results require only taking the first derivative of ψ and are thus well defined for K_1. $\blacksquare$

Appendix C. The breakdown point of the smoothed LMS estimator.

The LMS estimator has the finite sample breakdown point of $\frac{[\frac{n}{2}]-k+2}{n}$; we show here that the same holds for the smoothed LMS estimator; essentially smoothing does not alter the main properties of the objective function that are responsible for the high breakdown point.

The result for LMS is conditional on there being no hyperplane V that passes through zero in R^k and contains more than $k-1$ of $X_i' \in R^k$, where X_i are sample points. We assume that the joint distribution of $Z = \{(X_i, y_i)' \in R^{k+1}, i = 1, ...n\}$ is such that it does not put any mass on any such hyperplane; a sufficient condition is that the joint distribution be continuous (or have a continuous component in the X_i' vector). The following theorem is an adaptation of Theorem 2 in Rousseeuw and Leroy, 1987, pp.118-120, to the case of the smoothed estimator; here we similarly demonstrate that the estimator remains bounded if any number less than $[\frac{n}{2}] - k + 2$ of sample points get replaced by arbitrary values.

Theorem C. *The breakdown point of the smoothed LMS estimator is* $\frac{[\frac{n}{2}]-k+2}{n}$.

Proof.
The reader would need to consult the proof of Theorem 2 (Rousseeuw, Leroy, 1987, pp.118-120) for the LMS estimator (LMS proof) which is closely followed here. We concentrate our attention on part 1 of the LMS proof that establishes that the breakdown point is no lower than $\frac{[\frac{n}{2}]-k+2}{n}$. Part 2 of the proof shows that this value cannot be improved on: the breakdown point cannot be higher; this also holds for the smoothed estimator, but this part is of less interest and the adaptation of the corresponding proof is not provided.

Following the LMS proof denote by Z' the "corrupted" sample, obtained from the original sample Z by replacing no more than $[\frac{n}{2}] - k + 1$ sample points with arbitrary values. Our definition of ρ is the same as in the LMS proof; the property $\rho > 0$ similarly holds.

Note that θ in their notation corresponds to our $\tilde{\beta}$; for the sample Z' for their θ' we write $\tilde{\beta}'$. We use the same notation r_i for residuals from the original sample, $y_i - X_i\tilde{\beta}$, and similarly denote by M the maximum, $max\,|r_i|$. The definition of the set A remains the same; it similarly contains at least $n - [\frac{n}{2}]$ "good" observations (from the original sample). Similarly if $(X_a, y_a) \in A$ we denote the residuals $r_a = y_a - X_a\tilde{\beta}$ and $r'_a = y_a - X_a\tilde{\beta}'$. Note that as in the original LMS proof we use the notation $'$ ("prime") here for the values associated with the corrupted sample Z'; we avoid confusion by not referring to transposes in this proof.

We identify three steps in the original LMS proof:

Step 1 established

$$|r'_a - r_a| > \rho(\left\|\tilde{\beta}' - \tilde{\beta}\right\| - 2\left\|\tilde{\beta}\right\|).$$

Step 2 derived

$$med(r'_i)^2 \leq M^2.$$

Step 3 showed that if $\left\|\tilde{\beta}' - \tilde{\beta}\right\| \geq 2(\left\|\tilde{\beta}\right\| + \frac{M}{\rho})$ then

$$med(r'_i)^2 > M^2,$$

contradicting the result in step 2. It follows that $\left\|\tilde{\beta}'\right\|$ is bounded for any corrupted sample Z'.

Step 1 is obtained by a geometric construction and the result holds regardless of the nature of $\tilde{\beta}, \tilde{\beta}'$.

Denote by $a'(\tilde{\beta})$ for the sample Z' and the original $\tilde{\beta}$ the solution to equation (7). For $\tilde{\beta}'$ a similarly defined $\tilde{a}'(\tilde{\beta}')$ is the minimizer that solves the system of first-order conditions (7,8) for Z'; thus $a'(\tilde{\beta}) \geq \tilde{a}'(\tilde{\beta}')$. We modify steps 2 and 3 to obtain a similar contradiction involving $\left(\tilde{a}'(\tilde{\beta}')\right)^2$, instead of $med(r'_i)^2$.

Note that equation (7) cannot be satisfied for $\beta, a = a'(\beta)$ if for some i $\left|r_i'^2 - (a'(\beta))^2\right| < b$ and all the $n - [\frac{n}{2}]$ differences $r_a^2 - (a'(\beta))^2$ have the same sign. If $\left|r_i'^2 - (a'(\beta))^2\right| > b$ for all i then $a'(\beta)^2 = med(r'_i)^2$; by step 2 of the original LMS proof this value is bounded from above by M^2.

New step 2.

If some of the r_a are such that $\left| r_a^2 - \left(a'(\tilde{\beta}) \right)^2 \right| < b$, then $\left(a'(\tilde{\beta}) \right)^2 < M^2 + b$. If none of the r_a satisfy this, then by the argument above $\left(a'(\tilde{\beta}) \right)^2 < M^2$. Since $a'(\tilde{\beta}) \geq \tilde{a}'(\tilde{\beta}')$ we have

$$\left(a'(\tilde{\beta}') \right)^2 < M^2 + b.$$

New step 3.

Suppose that $\left\| \tilde{\beta}' - \tilde{\beta} \right\| > 2 \left\| \tilde{\beta} \right\| + \frac{M(1+\sqrt{1+\frac{2b}{M^2}})}{\rho})$. Then

$$|r_a' - r_a| \ > \ \rho(\left\| \tilde{\beta}' - \tilde{\beta} \right\| - 2 \left\| \tilde{\beta} \right\|) > \rho \left(2 \left\| \tilde{\beta} \right\| + \frac{M(1 + \sqrt{1 + \frac{2b}{M^2}})}{\rho}) - 2 \left\| \tilde{\beta} \right\| \right)$$

$$= \ M(1 + \sqrt{1 + \frac{2b}{M^2}})$$

where the first inequality follows from the result of step 1 of the original LMS proof. We get

$$|r_a'| \geq |r_a' - r_a| - |r_a| > M(1 + \sqrt{1 + \frac{2b}{M^2}}) - M = M\sqrt{1 + \frac{2b}{M^2}}.$$

Here $\left| r_i'^2 - \left(a'(\tilde{\beta}') \right)^2 \right| < b$ has to be satisfied for some i in order that $a'(\tilde{\beta}')$ be a minimizer satisfying (7,8). From the remark preceding the new step 2 it follows that either $\left| r_a'^2 - \left(a'(\tilde{\beta}') \right)^2 \right| < b$ for some $a \in A$ or some of the differences $r_a'^2 - \left(a'(\tilde{\beta}') \right)^2$ have different signs, in particular, at least one is negative. If some $\left| r_a'^2 - \left(a'(\tilde{\beta}') \right)^2 \right| < b$ then

$$\left(a'(\tilde{\beta}') \right)^2 > \left(M^2 + 2b \right) - b = M^2 + b.$$

If $\left| r_a'^2 - \left(a'(\tilde{\beta}') \right)^2 \right| > b$ always in A at least for one residual $r_a'^2 - \left(a'(\tilde{\beta}') \right)^2 < 0$ in which case the above inequality is also satisfied.

We thus get a contradiction and conclude that $\left\| \tilde{\beta}' \right\|$ is bounded by the same constant for any corrupted sample Z'. ∎

References

[1] Atkinson A.C.(1994) Fast Very Robust Methods for the Detection of Multiple Outliers, Journal of the American Statistical Association, 89, 1329-1339.

[2] Bierens, H.J. (1994) Topics in Advanced Econometrics, Cambridge University Press, N.Y.

[3] Campbell B. and J.W.Galbraith (1993) Inference in Expectations Models of the Term Structure: a Non-parametric Approach, Empirical Economics, 18, 623-638.

[4] Davies P.L. (1993) Aspects of Robust Linear regression, The Annals of Statistics, 21, 1843-1899.

[5] Ellis S.P. (1998), Instability of Least Squares, Least Absolute Deviation and Least Median of Squares Linear Regression, Statistical Sciences, 13, 337-350.

[6] Ferretti N., D. Kelmansky, V.Yohai and R.Zamar (1999) A Class of Locally and Globally Robust Regression Estimates, Journal of the American Statistical Association, 94, 174-188.

[7] Gel'fand I.M. and G.E.Shilov (1964) Generalized Functions, Vol.1, Properties and Operations, Academic Press, San Diego.

[8] Gel'fand I.M. and N.Ya.Vilenkin (1964) Generalized Functions, Vol.4, Applications of Harmonic Analysis, Academic Press, SanDiego.

[9] Horowitz J.L. (1992) A Smoothed Maximum Score Estimator for the Binary Response Model, Econometrica, 60, 505-531.

[10] Kim J. and D.Pollard (1990) Cube Root Asymptotics, Annals of Statistics, 18,191-219.

[11] Phillips P.C.B. (1991) A Shortcut to LAD Estimator Asymptotics, Econometric Theory, 7,450-463.

[12] Phillips, P.C.B. (1995) Robust Non-Stationary Regression, Econometric Theory, 11, 912-951.

[13] Pollard D.(1991) Asymptotics for Least Absolute Deviation Regression Estimators, Econometric Theory, 7, 186-199.

[14] Rousseeuw P.J.(1984) Least Median of Squares Regression , Journal of the American Statistical Association, 79, 871-880.

[15] Rousseeuw P.J. and A.M. Leroy (1987) Robust Regression and Outlier Detection, John Wiley and Sons, New York.

[16] Rousseeuw P.J. and B.C. van Zomeren (1990) Unmasking Multivariate Outliers and Leverage Points, Journal of the American Statistical Association, 85, 633-639.

[17] Sakata Sh. and H.White (1998) High Breakdown Point Conditional Dispersion Estimation with Application to S&P 500 Daily Returns Volatility, Econometrica, 66, 529-568.