International Journal of Statistics: Advances in Theory and Applications
Vol. 1, Issue 1, 2020, Pages 21-43
Published Online on April 12, 2019

http://jyotiacademicpress.org

THE RUNNING INTERVAL SMOOTHER: A CONFIDENCE BAND HAVING SOME SPECIFIED SIMULTANEOUS PROBABILITY COVERAGE

RAND R. WILCOX

Department of Psychology
University of Southern California
3620 S. McClintock Avenue
Los Angeles, CA 90089-1061
USA
e-mail: rwilcox@usc.edu

Abstract

Let $M(Y|X)$ be some conditional measure of location associated with the random variable Y, given X. Many nonparametric regression estimators of $M(Y|X)$ have been proposed. One that is particularly convenient when dealing with robust measures of location is a running interval smoother. The paper deals with the goal of computing K confidence intervals for $M(Y|X)$ corresponding to K values of the covariate X, where K is relatively large, that have simultaneous probability coverage $1 - \alpha$. When working with a 20% trimmed mean, methods based on the Studentized maximum modulus distribution or the Bonferroni method, for example, can be highly unsatisfactory. The paper describes and compares methods that provide more satisfactory results. When $M(Y|X)$ is taken to be the median of Y, given X, the approach based on 20% trimmed means performs poorly. An alternative approach, based in part on the Bonferroni method, was found that gives

2020 Mathematics Subject Classification: 62G08, 62G10, 62G35.
Keywords and phrases: confidence band, smoothers, robust regression, trimmed mean.
Communicated by Gaorong Li.
Received February 21, 2017; Revised March 7, 2017

reasonably satisfactory results. It is illustrated that achieving reasonably accurate probability coverage depends in part on the choice for the span and that a good choice for the span is a function of the strength of the association.

1. Introduction

Let $M(Y|X)$ be some conditional measure of location associated with the random variable Y, given X. Certainly, the best-known and most frequently used method for estimating $M(Y|X)$ is to assume $M(Y|X) = \beta_0 + \beta_1 X$, where β_0 and β_1 are unknown parameters. However, it is well established that this linear model might not provide an adequate approximation of the true regression line. Something other than a straight regression line might be needed. A simple approach is to include a quadratic term or some other type of parametric model, but even this approach can be unsatisfactory. Another approach is to use some nonparametric regression estimator, commonly called smoothers, many of which have been proposed (e.g., Hastie & Tibshirani [13]; Efromovich [3]; Eubank [4]; Fan & Gijbels [5]; Fox [6]; Green & Silverman [7]; Gyöfri et al. [8]; Härdle [11]; Wilcox [21]). The typical smoother is aimed at estimating $E(Y|X)$, the population mean of Y given X. It is well known, however, that the population mean is not robust (e.g., Huber & Ronchetti [17]; Hampel et al. [10]; Staudte & Sheather [19]). When dealing with robust measures of location, a relatively simple and effective method for estimating $M(Y|X)$ is the running interval smoother. It is readily adapted to any robust estimator of location and it has been studied extensively. For a summary of relevant results, see Wilcox [21].

For a single value of the covariate, x, it is a trivial matter to compute a confidence interval for $M(Y|X = x)$ based on the running interval smoother provided that the span is chosen appropriately. (Details are made clear in Section 2.) But when dealing with K points, say $x_1, \ldots, x_K$, there is the issue of computing confidence intervals that have some specified simultaneous probability coverage, $1 - \alpha$. If K is relatively

small, a Bonferroni adjustment might be made or a method based on the Studentized maximum modulus distribution might be used. However, if K is relatively small, important details about $M(Y|X)$ might be missed. If K is reasonably large, say 10 or 25 or even larger, a method based on the Studentized maximum modulus distribution might still be used, but preliminary simulations clearly demonstrated that this approach can be rather unsatisfactory when using a 20% trimmed mean: in some situations the actual simultaneous probability coverage can be substantially larger than the nominal level. That is, the widths of the confidence intervals are larger than necessary to achieve the desired simultaneous probability coverage. And of course there is the related issue of testing $H_0 : M(Y|X) = \mu_0$, where μ_0 is some specified constant. Power can be relatively low when using a Studentized maximum modulus distribution because the actual probability of one or more Type I errors is substantially smaller than the nominal level.

The paper examines methods for dealing with this issue when using the running interval smoother in conjunction with one of two measures of location: a 20% trimmed mean and the population median. For the 20% trimmed mean, the proposed method is based in part on the Tukey and McLaughlin [20] method for computing a confidence interval. When there is interest in the population median, the Tukey–McLaughlin method breaks down. Instead, the method derived by Hettmansperger and Sheather [15] is used. The 20% trimmed mean was chosen because it has good efficiency compared to the sample mean when sampling from a normal distribution and it has a reasonably high breakdown point, namely, 0.2. (The breakdown point refers to the smallest proportion of observations that must be altered to make the estimator arbitrarily large or small.) Moreover, it has been studied extensively and found to perform relatively well compared to other estimators that might be used (e.g., Wilcox [21]). This is not to suggest that it dominates, clearly this is not the case. The only point is that it is a relatively good choice.

The basic strategy for computing confidence intervals that have simultaneous probability coverage $1 - \alpha$ is to determine an appropriate adjustment when dealing with normal distributions and there is no association, and then use the same adjustment when sampling from a non-normal distribution or when there is an association. Roughly, when using a 20% trimmed mean, the adjustment has a certain similarity to using a Studentized maximum modulus, an important difference being that the method used here takes into account the correlation among the statistics that are used. As will be seen, this approach was found to perform well in simulations provided the span of the running interval smoother is not too large. As for using medians, a Bonferroni method was found to perform relatively well.

Section 2 describes the details of the proposed methods. Section 3 reports simulation results and Section 4 illustrates the methods using data from the Well Elderly 2 study.

2. Description of the Methods

First, the details of the running interval smoother are described followed by a description of the Tukey–McLaughlin method and the Hettmansperger–Sheather method. Then the strategy for adjusting the confidence intervals is described.

2.1. The running interval smoother

The running interval smoother is based on a relatively simple process. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample and consider the goal of estimating $M(Y|X = x)$, where x is some specified value of the independent variable X. Let MAD be the median absolute deviation statistic. That is, MAD is the median of $|X_1 - M|, \ldots, |X_n - M|$, where M is the usual sample median based on $X_1, \ldots, X_n$. Let MADN = MAD/0.6745. Under normality, MADN estimates the population standard deviation. Let f be some constant (called the span) that is chosen in a manner to be describe. Now, the point x is said to be close to X_i if

$$|X_i - x| \leq f \times \text{MADN}.$$

So for normal distributions, x is close to X_i if x is within f standard deviations of X_i. Let

$$N(x) = \{i : |X_i - x| \leq f \times \text{MADN}\}.$$

That is, $N(x)$ indexes the set of all X_i values that are close to x. Let $\hat{\theta}_i$ be an estimate of some parameter of interest, based on the Y_i values such that $i \in N(x)$. That is, use all of the Y_i values for which X_i is close to x. The running interval smoother simply computes $\hat{\theta}_i$ $(i = 1, \ldots, n)$, which provides an estimate of $M(Y|X = x)$.

Typically, taking the span to be 0.8 suffices in terms of providing a relatively accurate estimate of $M(Y|X)$, based on mean squared error and bias, but of course exceptions are encountered (e.g., Wilcox [21]). This assumes that bias is measured with

$$\sum E(\hat{\theta}_i - \theta_i). \tag{1}$$

However, this measure of bias is unsatisfactory for present purposes. To illustrate why, consider data generated via the model

$$Y = \beta_1 X + \epsilon,$$

where both X and ϵ have standard normal distributions. For the case $\beta_1 = 0$, there is little bias when estimating $M(Y|X = 1)$ or $M(Y|X = -1)$. However, when $\beta_1 = 1$ and the span is $f = 0.8$, the bias is rather severe. It is, $1 - 0.81 = 0.19$ for $X = 1$ and -0.19 for $X = -1$ (based on a simulation with 10,000 replications). More generally, for $X = x$, bias is positive or negative depending on whether $x > 0$. So bias is small based on (1), but the bias for a specific choice for x can be severe, which in turn can result in an inaccurate confidence interval for $M(Y|X = x)$.

Note that for the situation at hand, Pearson's correlation is $\rho = \beta_1 / \sqrt{\beta_1^2 + 1}$, which is 0.7 when $\beta_1 = 1$. The degree of bias is a function of ρ. Still assuming $f = 0.8$, if $\beta_1 = 0.5$, so $\rho = 0.45$, the bias at $X = 1$ is 0.12 and for $\beta_1 = 0.25$ ($\rho = 0.24$), the bias is 0.05. So for a relatively weak association, the bias is relatively low and a reasonably accurate confidence interval can be computed, but otherwise this is not the case.

Reducing the span reduces the bias at any specific design point. Using $f = 0.5$, the bias associated with $\rho = 0.24$, 0.45, and 0.7, again at the point $X = 1$, is 0.02, 0.04, and 0.077, respectively. For $f = 0.2$, bias is now 0.001, 0.011, and 0.014, respectively. In the context of computing confidence intervals having some specified simultaneous probability coverage, this helps explain why the methods consider here do not perform well in general when $f = 0.8$, except when the strength of association is relatively weak. For this reason, the focus here is on $f = 0.5$ and 0.2 henceforth.

2.2. The Tukey–McLaughlin method

To describe the Tukey–McLaughlin method, momentarily ignore the covariate X. Let $Y_{(1)} \leq \ldots \leq Y_{(n)}$ be the observations written in ascending order. Suppose the desired amount of trimming has been chosen to be γ, $0 \leq \gamma < 0.5$. Let $g = [\gamma n]$, where $[\gamma n]$ is the value of γn rounded down to the nearest integer. The sample trimmed mean is computed by removing the g largest and g smallest observations and averaging the values that remain. More formally, the sample trimmed mean is

$$\bar{Y}_t = \frac{Y_{(g+1)} + \ldots + Y_{(n-g)}}{n - 2g}. \tag{2}$$

As seems evident, the optimal amount of trimming depends on the situation–no single amount is always optimal based on efficiency and achieving relatively high power when testing hypotheses. As previously

noted, the focus here is on $\gamma = 0.2$ because this results in relatively good efficiency under normality, versus the sample mean. Moreover, empirical studies summarized by Wilcox [21] suggest that often it has good efficiency relative to other amounts of trimming as well as other robust estimators that might be used. (More comments about this issue are relegated to the final section of this paper).

Next, let

$$W_i = \begin{cases} Y_{(g+1)}, & \text{if } Y_i \le Y_{(g+1)}, \\ Y_i, & \text{if } Y_{(g+1)} < Y_i < Y_{(n-g)}, \\ Y_{(n-g)}, & \text{if } Y_i \ge Y_{(n-g)}. \end{cases}$$

The Winsorized sample mean is

$$\overline{W} = \frac{1}{n}\sum W_i,$$

and the Winsorized standard deviation is

$$s_w^2 = \frac{1}{n-1}\sum (W_i - \overline{W})^2.$$

The two-sided Tukey–McLaughlin $1-\alpha$ confidence interval for the population trimmed mean is

$$\overline{Y}_t \pm t_{1-\alpha/2}\frac{s_w}{(1-2\gamma)\sqrt{n}}, \tag{3}$$

where $t_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of Student's t distribution with $n-2g-1$ degrees of freedom. This will be called method TM henceforth.

2.3. The Hettmansperger–Sheather method

Method TM performs poorly when the amount of trimming approaches 0.5. When dealing with the median, the method derived by Hettmansperger and Sheather [15] is used here.

Let U be a binary random variable that has a binomial distribution with probability of success $p = 0.5$. For any integer k greater than 0 and less than $[n/2]$, let $\zeta_k = P(k \le U \le n - k)$. Then a distribution-free ζ_k confidence interval for the median is

$$(Y_{(k)}, Y_{(n-k+1)})$$

(e.g., Hettmansperger & McKean [14]).

Because the binomial distribution is discrete, it is not possible, in general, to choose k so that the probability coverage is exactly equal to $1 - \alpha$. For example, if $n = 10$, a 0.891 and 0.978 confidence interval can be computed, but not a 0.95 confidence interval as is often desired. However, linear interpolation can be used along the lines suggested by Hettmansperger and Sheather [15] so that the probability coverage is approximately $1 - \alpha$. First determine k such that $\zeta_{k+1} < 1 - \alpha < \zeta_k$. Next, compute

$$I = \frac{\zeta_k - (1 - \alpha)}{\zeta_k - \zeta_{k+1}},$$

and

$$\lambda = \frac{(n - k)I}{k + (n - 2k)I}.$$

Then an approximate $1 - \alpha$ confidence interval is

$$(\lambda X_{(k+1)} + (1 - \lambda)X_{(k)},\ \lambda X_{(n-k)} + (1 - \lambda)X_{(n-k+1)}). \tag{4}$$

This will be called method HS henceforth. Results reported by Sheather and McKean [18], as well as Hall and Sheather [9], support the use of this method.

2.4. Computing confidence intervals

As is evident, when using methods TM and HS, a confidence interval can be computed based on the Y_j values for which $j \in N(X_i)$. Let $\#N(X_i)$ denote the cardinality of the set $N(X_i)$. With 20% trimming, the expectation is that generally, a reasonably accurate confidence can be obtained when $\#N(X_i) \geq 12$ (Wilcox [21]). So here, when computing confidence intervals for $M(Y|X = X_i)$, only X_i values satisfying $\#N(X_i) \geq 12$ are used. Note that when using method HS, if both n and α are sufficiently small, a confidence interval cannot be computed. To avoid this issue, only X_i values satisfying $\#N(X_i) \geq 16$ are used.

Next, focus on method TM. Let $x_1, \ldots, x_K$ be K covariate values of interest, where K is relatively large. Momentarily consider the goal of testing

$$H_0 : M(Y|X = x_k) = 0,$$

for each $k = 1, \ldots, K$ and let $p_1, \ldots, p_K$ be the corresponding p-values based on the Tukey–McLaughlin method. Let $p_{\min} = \min(p_1, \ldots, p_K)$. As is evident, if the α quantile of the distribution of $p_{\min}$ could be obtained, say p_α, then the probability of one or more Type I errors is α. The strategy is to determine p_α when $Y = \epsilon$, and both X and ϵ have a standard normal distribution. Then simulations are used to determine the impact on the probability of one or more Type I errors when there is an association and the error term has a non-normal distribution. (So the strategy is similar in spirit to classic ANOVA methods.)

Here, two strategies for choosing $x_1, \ldots, x_K$ are considered, which are labelled methods M1 and M2. The basic idea behind both strategies is to choose points such that the number of nearest neighbours, $\#N(x_k)$, is sufficiently large, say greater than to equal to $n_{\min}$, so as to yield a confidence having reasonably accurate probability coverage. For reasons

previously indicated, $n_{\min} = 12$ is used in conjunction with method TM, given the goal that the simultaneous probability coverage is to be $1 - \alpha = 0.95$. Perhaps $n_{\min}$ needs to be adjusted when say $1 - \alpha = 0.99$ or when using some other robust estimator, but this is not pursued here.

The same method for determining p_α was considered when using method HS, with $n_{\min} = 16$, but this was found to be unsatisfactory in simulations: the estimate of α often exceeded 0.08. A more successful method was to simply rely on the Bonferroni method. That is, probability coverage for each of the K confidence intervals is set at $1 - \alpha / K$.

Method M1

M1 uses a specified number of covariate values. How many points to use depends on how much detail is desired, which presumably depends on the situation. Here, the focus is on $K = 25$ points evenly space between x_1 and x_{25}, inclusive, where x_1 is taken to be smallest X_i value such that $\# N(X_i) \geq n_{\min}$ and let x_{25} is taken to be the largest X_i value such that $\# N(X_i) \geq n_{\min}$. Some consideration is given to $K = 10$ as well.

Method M2

M2 uses all X_i values such that $\# N(X_i) \geq n_{\min}$. This approach can be implemented when using method TM, but it breaks down when using HS and $n_{\min} = 16$, again because if α / K is sufficiently small, a confidence interval cannot be computed.

As previously indicated, p_α is estimated via a simulation. To elaborate on the details, the first step was to generate data and compute p-values for each $H_0 : M(Y|X = x_k) = 0 (k = 1, \ldots, K)$ followed by $\hat{p}$, the minimum of the K p-values. This process was repeated 4000 times yielding $\hat{p}_1, \ldots, \hat{p}_{4000}$, which were then used to estimate the α quantile of $p_{\min}$ via the quantile estimator derived by Harrell and Davis [12].

As previously noted, $f = 0.8$ seems to suffice in most situations in terms of capturing any curvature that might exist. However, preliminary simulations revealed that in some situations, $f = 0.8$ resulted in estimates of the simultaneous probability that were less than 0.90 when using M1. Decreasing the span to $f = 0.5$ substantially improved matters except in situations where there is substantial curvature, or more generally when the strength of the association is relatively high, in which case $f = 0.2$ provides substantially better results. Henceforth, $f = 0.5$ is assumed when using M1 unless stated otherwise. As for M2, $f = 0.5$ can be unsatisfactory even when the regression line is straight. Using $f = 0.2$ resulted in much better control over the Type I error probability.

Table 1 shows some estimates of p_α, when $\alpha = 0.05$, based on 4000 replications and sample sizes n ranging from 50 to 1000. For $n < 50$ with $f = 0.2$, situations are encountered where $\#N(x_k)$ is not sufficiently large for any x_k, $k = 1, \ldots, K$, which explains the missing entries in Table 1. A similar problem occurs when using M1 and $n \leq 40$, which is why the smallest sample size in Table 1 is 50. Note that based on the Bonferroni method, using method M1 with $K = 25$ and $K = 10$, each of the K tests would be performed at the 0.002 and 0.005 level, respectively. So Table 1 indicates that as n increases, the estimates of p_α decrease and are only slightly larger than the Bonferroni values when $n = 1000$ when using M1.

Table 1. Estimates of p_α based on 4000 replications

n	M1 ($K = 25$)	M1 ($K = 10$)	M2
50	0.0048	0.0076	****
60	0.0045	0.0084	****
70	0.0042	0.0070	0.0114
80	0.0041	0.0068	0.0057
100	0.0035	0.0057	0.0024
150	0.0033	0.0057	0.0014
200	0.0030	0.0061	0.0012
300	0.0030	0.0054	0.0008
400	0.0028	0.0055	0.0006
500	0.0025	0.0052	0.0006
600	0.0026	0.0056	0.0006
800	0.0026	0.0054	0.0005
1000	0.0028	0.0055	0.0005

3. Simulation Results

Simulations were used to study the small-sample properties of method TM in conjunction with M1 and M2. Estimates of the simultaneous probability coverage were based on 4000 replications. Data were generated based on the model

$$Y = X^a + \epsilon, \tag{5}$$

for $a = 0$, 1 and 2.

Four types of distributions were used: normal, symmetric and heavy-tailed, asymmetric and light-tailed, and asymmetric and heavy-tailed. More precisely, both the error term and the distribution of the independent variable were taken to be one of four g-and-h distributions (Hoaglin [16]) that contain the standard normal distribution as a special case. If Z has a standard normal distribution, then

$$V = \begin{cases} \dfrac{\exp(gZ) - 1}{g} \exp(hZ^2 / 2), & \text{if } g > 0, \\ Z \exp(hZ^2 / 2), & \text{if } g = 0, \end{cases}$$

has a g-and-h distribution where g and h are parameters that determine the first four moments. The four distributions used here were the standard normal $(g = h = 0.0)$, a symmetric heavy-tailed distribution $(h = 0.2, g = 0.0)$, an asymmetric distribution with relatively light tails $(h = 0.0, g = 0.2)$, and an asymmetric distribution with heavy tails $(g = h = 0.2)$. Table 2 shows the skewness (κ_1) and kurtosis (κ_2) for each distribution. Additional properties of the g-and-h distribution are summarized by Hoaglin [16].

Table 2. Some properties of the g-and-h distribution

g	h	κ_1	κ_2
0.0	0.0	0.00	3.0
0.0	0.2	0.00	21.46
0.2	0.0	0.61	3.68
0.2	0.2	2.81	155.98

Table 3 summarizes the simulation results for method TM, based on M1, for $a = 0$ and 2, where the random variables X and ϵ were generated from identical g-and-h distributions. Shown are estimates of α when the goal is to achieve simultaneous probability coverage $1 - \alpha = 0.95$ when computing confidence intervals for the population 20% trimmed means. Bradley [1] has suggested that as a general guide, when computing a 0.95 confidence interval, at a minimum the actual probability coverage should be between 0.925 and 0.975. Note that for $a = 0$ and $f = 0.2$ as well as 0.5, this criterion is met for all of the situations considered. As previously noted, $f = 0.5$ is generally small enough to achieve a reasonably accurate approximation for any regression line that exhibits curvature. However, if the strength of the association is sufficiently high, $f = 0.2$

can be required given the goal of achieving simultaneous probability coverage equal to 0.95. For $a = 2$ and $f = 0.2$, Bradley's criterion is met for all but one situation, where the estimate is 0.077 for $g = 0.2$, $h = 0.0$, and $n = 1000$. Plots reveal the problem: a slightly smaller choice for the span is required. For $a = 1$, not shown in Table 3, again $f = 0.2$ is required to get reasonably good control over the probability coverage.

Table 3. Estimates of α, when the goal is to achieve simultaneous probability coverage $1 - \alpha = 0.95$, based on method TM using M1 with $K = 25$

g	h	n	$a = 0, f = 0.5$	$a = 0, f = 0.2$	$a = 2, f = 0.2$
0.0	0.0	50	0.050	****	****
0.0	0.0	100	0.050	0.066	0.065
0.0	0.0	200	0.050	0.055	0.057
0.0	0.0	1000	0.050	0.059	0.071
0.0	0.2	50	0.036	****	****
0.0	0.2	100	0.040	0.046	0.046
0.0	0.2	200	0.040	0.041	0.044
0.0	0.2	1000	0.049	0.049	0.063
0.2	0.0	50	0.050	****	****
0.2	0.0	100	0.055	0.054	0.064
0.2	0.0	200	0.048	0.060	0.065
0.2	0.0	1000	0.057	0.065	0.077
0.2	0.2	50	0.034	****	****
0.2	0.2	100	0.037	0.049	0.044
0.2	0.2	200	0.036	0.044	0.046
0.2	0.2	1000	0.052	0.052	0.060

It is noted that additional simulations were run using $\beta_1 = 0.5$, $a = 1$, and $f = 0.5$. For $50 \leq n \leq 500$, the estimates were reasonably close to the nominal level. For example, when $g = h = 0$ and $n = 50$, the estimate was 0.055. For $n = 500$ the estimate was 0.052, but for $n = 800$

the estimate was 0.08. The difficulty is that bias becomes more of a factor when the sample size is large. Using $f = 0.2$ corrects this problem, the estimate being 0.065. This suggests using $f = 0.2$ routinely, but a concern is that this can result in relatively wide confidence intervals.

Table 4 shows the simulation results for method M2. There are two situations where the estimate drops below 0.025, both of which occur when $n = 50$, $a = 0$, $f = 0.5$ and sampling is from a heavy-tailed distribution. No estimate exceeds 0.075, the largest estimate being 0.68.

Table 4. Estimates of α, when the goal is to achieve simultaneous probability coverage $1 - \alpha = 0.95$, based on method TM using M2

g	h	n	$a = 0, f = 0.5$	$a = 0, f = 0.2$	$a = 2, f = 0.2$
0.0	0.0	50	0.050	****	****
0.0	0.0	100	0.050	0.050	0.048
0.0	0.0	200	0.050	0.050	0.052
0.0	0.0	1000	0.050	0.048	0.065
0.0	0.2	50	0.021	****	****
0.0	0.2	100	0.036	0.034	0.021
0.0	0.2	200	0.027	0.032	0.033
0.0	0.0	1000	0.044	0.045	0.059
0.2	0.0	50	0.028	****	****
0.2	0.0	100	0.038	0.050	0.046
0.2	0.0	200	0.038	0.055	0.057
0.2	0.0	1000	0.059	0.052	0.068
0.2	0.2	50	0.018	****	****
0.2	0.2	100	0.037	0.036	0.032
0.2	0.2	200	0.052	0.031	0.037
0.2	0.2	1000	0.057	0.044	0.059

Table 5 shows the results when using method HS. The largest estimate is 0.054. The main difficulty is that estimates drop below 0.025. The two smallest estimates, 0.014 and 0.019, occur when $n = 50$, $f = 0.5$ and sampling is from a skewed distribution ($g = 0.2$).

Table 5. Estimates of α, when the goal is to achieve simultaneous probability coverage $1 - \alpha = 0.95$, based on method HS ($K = 25$)

g	h	n	$a = 0, f = 0.5$	$a = 0, f = 0.2$	$a = 2, f = 0.2$
0.0	0.0	50	0.022	****	****
0.0	0.0	100	0.037	0.023	0.023
0.0	0.0	200	0.034	0.034	0.037
0.0	0.0	1000	0.042	0.049	0.054
0.0	0.2	50	0.022	****	****
0.0	0.2	100	0.033	0.020	0.021
0.0	0.2	200	0.031	0.028	0.028
0.0	0.2	1000	0.045	0.040	0.046
0.2	0.0	50	0.014	****	****
0.2	0.0	100	0.034	0.020	0.022
0.2	0.0	200	0.031	0.033	0.031
0.2	0.0	1000	0.037	0.038	0.052
0.2	0.2	50	0.019	****	****
0.2	0.2	100	0.029	0.022	0.021
0.2	0.2	200	0.036	0.030	0.033
0.2	0.2	1000	0.034	0.038	0.054

There is the issue how the lengths of the confidence intervals using M1 compare to the lengths based on the Studentized maximum modulus (SMM) distribution. To provide at least some indication, consider the case $Y = X + \epsilon$, where both X and ϵ have standard normal distributions. For the k-th value among the K values of the independent variable, let L_{k1} denoted that the length of the confidence interval and let L_{k2} denote the length based on method SMM ($k = 1, \ldots, K$). Let $\overline{L}_m = \sum L_{km} / K (m = 1, 2)$. A simulation estimate of $E(\overline{L}_1) / E(\overline{L}_2)$, when $K = 10$ and $n = 50$, was 0.958. Increasing n to 100, the estimate was 1.0. For $K = 25$ and $n = 50$, the estimate was 0.918, and for $n = 100$ the estimate was 0.953.

4. Some Illustrations

The methods are illustrated using data from the Well Elderly 2 study (Clark et al. [2]) that dealt with an intervention program aimed at improving the physical and emotional wellbeing of older adults. A portion of the study focused on the association between the cortisol awakening response (CAR) and a measure of depressive symptoms based on the Center for Epidemiologic Studies Depressive Scale (CESD). CAR refers to the change in cortisol concentration that occurs 30-60 minutes after waking from sleep. A CESD score greater than 15 is regarded as an indication of mild depression. A score greater than 21 indicates the possibility of major depression.

It seems fairly evident that simply computing a 0.95 confidence interval for each of the K covariate values of interest can result in a substantially different result compared to method TM, where the goal is to compute confidence intervals having simultaneous probability coverage 0.95. Using measures taken after intervention, Figures 1 and 2 illustrate this point. Figure 1 shows the confidence intervals based on the former strategy with $K = 25$ covariate values chosen as done by M1 and when the span is $f = 0.5$. Figure 2 shows the results when using method TM based on M1. As can be seen, the length of the confidence intervals differ substantially from those in Figure 1, as would be expected.

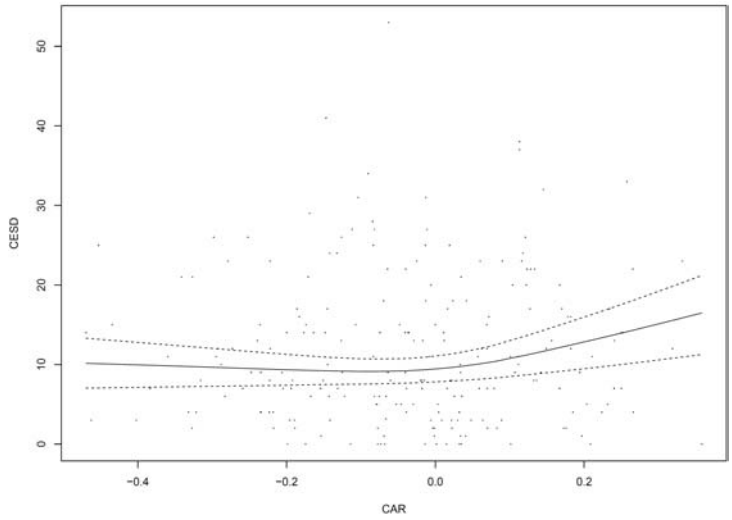

Figure 1. Confidence intervals based on 20% trimmed means and the Well Elderly 2 data where each confidence interval has, approximately, probability coverage 0.95.

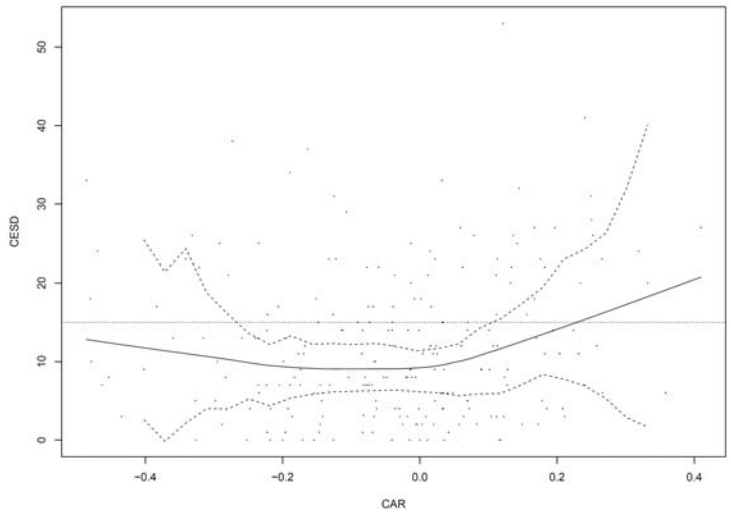

Figure 2. Confidence intervals based on the same data used in Figure 1, only now the simultaneous probability coverage is approximately 0.95. The horizontal dotted line corresponds to CESD = 15. (CESD values greater than 15 are considered an indication of mild depression.)

The horizontal dotted line in Figure 2 corresponds to CESD = 15. So Figure 2 indicates that for CAR values between – 0.2 to 1.5, after intervention, a reasonable decision is that the typical participant does not have any indication of mild depression. Outside this interval, it is unclear the extent to which this is the case.

Figure 3 shows an estimate of the regression line prior to intervention. Note that now, it is less clear whether the typical participant does not show signs of mild depression. Moreover, there is no strong empirical evidence that there is an association. This in contrast to Figure 2 where it appears there is a positive association between CAR and CESD when CAR is positive (cortisol decreases after awakening). For CAR greater than zero, the slope of the regression line, based on the Theil-Sen estimator, is significant at the 0.05 level (using a percentile bootstrap method), $p = 0.038$. (A significant result is also obtained using least squares regression in conjunction with the HC4 estimate of the standard error, $p = 0.012$.) This raises the concern that after intervention, for CAR sufficiently large, the typical participant might exhibit mild depression. In Figure 2, for example, for CAR greater than 2.1, the typical CESD measure is estimated to be greater than 15. However, based on the confidence intervals in Figure 2, there is no compelling evidence that this is the case.

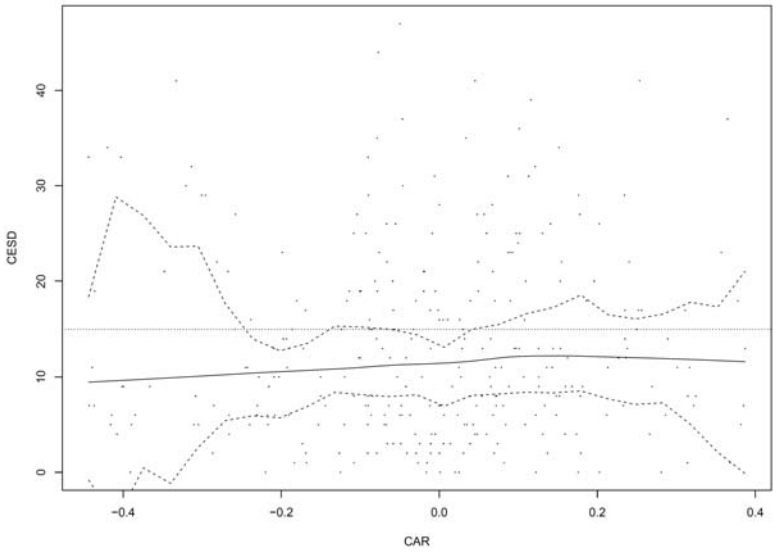

Figure 3. Confidence intervals based on measures taken prior to intervention. The simultaneous probability coverage is approximately 0.95.

5. Concluding Remarks

There are, of course, many variations of the methods considered here and there is the practical issue that no single estimator dominates in terms of efficiency. So, using a robust estimator other than the 20% trimmed mean or median might have practical value. For example, a robust M-estimator might be used, but it is known that non-bootstrap methods, based on some estimate of the standard error, can perform poorly in terms of achieving reasonably accurate probability coverage when sampling from skewed distributions (Wilcox [21]). This issue can be addressed with a percentile bootstrap method. For the situation at hand, perhaps a percentile bootstrap method gives satisfactory results, but this remains to be determined.

One of the main points is that when using a 20% trimmed mean, the strategy of using a Studentized maximum modulus distribution can result in confidence intervals for which the actual simultaneous probability coverage can be substantially greater than the nominal level. Of course, switching to the Bonferroni method only makes matters worse. This is less of an issue when using the median, but for $n = 100$ there are situations where the actual probability coverage is greater than 0.975.

The choice for the span is crucial. If there is a fairly weak association, taking the span to be $f = 0.8$ is satisfactory, but otherwise the span should 0.5 or smaller. For a very strong association, $f = 0.2$ should be used. For sample sizes $n > 1000$, perhaps $f < 0.2$ is required, particularly when the strength of the association is fairly strong. The simulations suggest that to minimize the bias associated with any estimate of $M(Y|X)$, f should be a decreasing function of n, with the complication that any method for choosing the span also depends on the strength of the association. For $n \leq 1000$, an argument for using $f = 0.2$ routinely is that the simultaneous probability coverage is controlled reasonably well, at least among the situations considered here. But a negative feature is that $f = 0.2$ can result in relatively wide confidence intervals compared to using $f = 0.5$.

Finally, the R function rplotCI applies method TM based on M1, and rplotCIv2 uses M2, both of which have been added to the R package WRS. The R function rplotCIM applies method HS and has been added to WRS as well.

References

[1] J. V. Bradley, Robustness? British Journal of Mathematical and Statistical Psychology 31 (1978), 144-152.

[2] F. Clark, J. Jackson, M. Carlson, C.-P. Chou, B. J. Cherry, M. Jordan-Marsh, B. G. Knight, D. Mandel, J. Blanchard, D. A. Granger, R. R. Wilcox, M. Y. Lai, B. White, J. Hay, C. Lam, A. Marterella and S. P. Azen, Effectiveness of a lifestyle intervention in promoting the well-being of independently living older people: Results of the Well Elderly 2 randomise controlled trial, Journal of Epidemiology and Community Health 66 (2012), 782-790.

DOI: http://dx.doi.org/10.1136/jech.2009.099754

[3] S. Efromovich, Nonparametric Curve Estimation: Methods, Theory and Applications, Springer-Verlag, New York, 1999.

[4] R. L. Eubank, Nonparametric Regression and Spline Smoothing, Marcel Dekker, New York, 1999.

[5] J. Fan and I. Gijbels, Local Polynomial Modeling and its Applications, CRC Press, Boca Raton, FL, 1996.

[6] J. Fox, Multiple and Generalized Nonparametric Regression, Sage, Thousands Oaks, CA, 2001.

[7] P. J. Green and B. W. Silverman, Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach, CRC Press, Boca Raton, FL, 1993.

[8] L. Györfi, M. Kohler, A. Krzyzk and H. Walk, A Distribution-Free Theory of Nonparametric Regression, Springer Verlag, New York, 2002.

[9] P. Hall and S. J. Sheather, On the distribution of a studentized quantile, Journal of the Royal Statistical Society B 50 (1988), 380-391.

[10] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel, Robust Statistics, Wiley, New York, 1986.

[11] W. Hardle, Applied Nonparametric Regression, Econometric Society Monographs No. 19, Cambridge University Press, Cambridge, UK, 1990.

[12] F. E. Harrell and C. E. Davis, A new distribution-free quantile estimator, Biometrika 69 (1982), 635-640.

[13] T. J. Hastie and R. J. Tibshirani, Generalized Additive Models, Chapman and Hall, New York, 1990.

[14] T. P. Hettmansperger and J. W. McKean, Robust Nonparametric Statistical Methods, Arnold, London, 1998.

[15] T. P. Hettmansperger and S. J. Sheather, Confidence intervals based on interpolated order statistics, Statistics and Probability Letters 4 (1986), 75-79.

[16] D. C. Hoaglin, Summarizing shape numerically: The *g*-and-*h* distribution, In D. C. Hoaglin, F. Mosteller & J. W. Tukey (Eds.), Exploring Data Tables Trends and Shapes, Wiley, New York, 1985, pp. 461-514.

[17] S. J. Huber and E. Ronchetti, Robust Statistics, 2nd Edition, Wiley, New York, 2009.

[18] S. J. Sheather and J. W. McKean, A comparison of testing and confidence intervals for the median, Statistical Probability Letters 6 (1987), 31-36.

[19] R. G. Staudte and S. J. Sheather, Robust Estimation and Testing, Wiley, New York, 1990.

[20] J. W. Tukey and D. H. McLaughlin, Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/Winsorization 1, Sankhya A 25 (1963), 331-352.

[21] R. R. Wilcox, Introduction to Robust Estimation and Hypothesis Testing, 4th Edition, Academic Press, San Diego, CA, 2017.

■