LINEAR REGRESSION: ROBUST HETEROSCEDASTIC CONFIDENCE BANDS THAT HAVE SOME SPECIFIED SIMULTANEOUS PROBABILITY COVERAGE

Rand R. Wilcox
Dept of Psychology
University of Southern California

October 23, 2016
ABSTRACT

Let \( M(Y|X) = \beta_0 + \beta_1 X \) be some conditional measure of location associated with the random variable \( Y \), given \( X \), where the unknown parameters \( \beta_0 \) and \( \beta_1 \) are estimated based on the random sample \((X_1, Y_1), \ldots, (X_n, Y_n)\). When using the ordinary least squares (OLS) estimator and \( M(Y|X) = E(Y|X) \), several methods for computing a confidence band have been derived that are aimed at achieving some specified simultaneous probability coverage assuming a homoscedastic error term and normality. There is an extant technique that allows heteroscedasticity, but a remaining concern is that the OLS estimator is not robust. Extant results indicate how a confidence interval can be computed via a robust regression estimator when there is heteroscedasticity and attention is focused on a single value of \( X \). The paper extends this method by describing a heteroscedastic technique for computing a confidence interval for each \( M(Y|X = X_i) \) \((i = 1, \ldots, n)\) such that the simultaneous probability coverage has some specified value. The small-sample properties of the method are studied when using the ordinary least squares estimators as well as three robust regression estimators.

Keywords: prediction intervals, heteroscedasticity, robust regression, analysis of covariance, Well Elderly 2 study.

1 Introduction

Let \( M(Y) \) be some population measure of location associated with the random variable \( Y \) and consider a situation where the conditional measure of location of \( Y \), given \( X \), is

\[
M(Y|X) = \beta_0 + \beta_1 X,
\]

where \( \beta_0 \) and \( \beta_1 \) are unknown parameters that are estimated based on the random sample \((X_1, Y_1), \ldots, (X_n, Y_n)\). Further assume that given \( X \),

\[
Y = \beta_0 + \beta_1 X + \lambda(X)\epsilon,
\]

where \( \epsilon \) has some unknown distribution with \( M(\epsilon) = 0 \) and \( \lambda(X) \) is some unknown function of \( X \) used to model heteroscedasticity. As is evident, a fundamental goal is computing
a confidence interval for $M(Y|X)$. From basic principles, if $\epsilon$ has a normal distribution with $M(\epsilon) = E(\epsilon) = 0$ and $\lambda(X) \equiv 1$ (homoscedasticity), a $1 - \alpha$ confidence interval for $M(Y|X)$ can be computed for a single value of $X$ based on the ordinary least squares (OLS) estimator (e.g., Montgomery and Peck, 1992). Moreover, several methods for computing a confidence band have been derived that are designed to have some specified simultaneous probability coverage, again assuming normality and homoscedasticity (Liu et al., 2008). However, homoscedastic methods use an incorrect estimate of the standard error when there is heteroscedasticity. A method that allows heteroscedasticity was derived and studied by Faraway and Sun (1995). It is well known, however, that OLS is not robust (e.g., Huber & Ronchetti, 2009; Staudte & Sheather, 1990; Rousseeuw & Leroy, 1990; Marrona et al. 2006). In practical terms, OLS can result in a poor fit for the bulk of the points and it can have poor efficiency relative to various robust regression estimators (e.g., Wilcox, 2017). When using a robust regression estimator and when focused on a single $X_i$ ($1 \leq i \leq n$), there is a straightforward method for computing a confidence interval for $M(Y|X = X_i)$ as indicated in Section 2. But what is unclear is how to compute a confidence interval for each $M(Y|X_i)$, $i = 1, \ldots, n$ such that the simultaneous probability coverage is $1 - \alpha$. If there are tied values, let $U_1, \ldots, U_m$ be the $m$ unique values among the $X_1, \ldots, X_n$, in which case the goal is to compute confidence intervals for $M(Y|X = U_i)$, $i = 1, \ldots, m$.

A simple strategy aimed at achieving simultaneous probability coverage greater than or equal $1 - \alpha$ would be to use the Bonferroni inequality and compute a $1 - \alpha/n$ confidence interval for each $X_i$, $i = 1, \ldots, n$. It seems fairly evident, however, that even when $n$ is moderately small, this approach can result in relatively wide confidence intervals.

A related issue is testing the $n$ hypotheses

$$H_0 : M(Y|X = X_i) = \theta_0,$$

for each $i = 1, \ldots, n$, where $\theta_0$ is some specified constant. The goal is to test all $n$ hypotheses such that the family wise error (FWE) rate, meaning the probability of one or more Type I errors, is equal to $\alpha$. Consider, for example, the Well Elderly 2 study (Jackson et al, 2009; Clark et al. 2012) that dealt with an intervention program aimed at improving the physical and emotional wellbeing of older adults. The cortisol awakening response (CAR) refers to the change in cortisol concentration that occurs during the first hour after waking from sleep. (CAR is taken to be the cortisol level upon awakening and measure again 30-60
minutes later.) Extant studies (e.g., Clow et al., 2004; Chida & Steptoe, 2009) indicate that various forms of stress are associated with the CAR. One issue was understanding the association between the CAR and a measure of depressive symptoms based on the Center for Epidemiologic Studies Depressive Scale (CESD). The CESD (Radloff, 1977) is sensitive to changes in depressive status over time and has been successfully used to assess ethnically diverse older people (Lewinsohn et al., 1988; Foley et al., 2002). Higher scores indicate a higher level of depressive symptoms. When the CAR is positive, a positive association was found between the CAR and CESD based on the Theil–Sen regression estimator, which is reviewed in section 2.1. (A percentile bootstrap method was used and leverage points were removed.) More precisely, the hypothesis of a zero slope was rejected at the 0.05 level ($p = 0.021$).

A simple interpretation is that when the CAR is positive (cortisol decreases after awakening), as CAR increases, typical depressive symptoms tend to increase as well. To get a better understanding of the practical implications of this result, it is noted that 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. So an issue is whether anything can be said about whether $M(Y|X)$ is greater than or less than 15 over the range of observed $X$ values, where $X$ is the CAR and $Y$ is CESD.

Note that testing these $n$ hypotheses also arises in the context of an analysis of covariance (ANCOVA) where two outcome variables are of interest, say $Z_1$ and $Z_2$, are possibly dependent. For example, $Z_1$ and $Z_2$ might be measures taken before and after some treatment. If $C_1$ and $C_2$ are corresponding covariates measured at times one and two, of interest is understanding $M(Z_1 − Z_2|C_1 − C_2)$.

As indicated in section 2, testing a single hypothesis is easily done when using a robust regression estimator. When testing $n$ hypotheses, there is the issue of controlling the probability of one or more Type I errors. One strategy is to use some improvement on the Bonferroni method, such as the method derived by Hochberg (1988) or Hommel (1988). A fairly obvious concern, however, is that when $n$ is even moderately large, relatively low power can result.

Consider, for example, Hochberg’s sequentially rejective method, which is applied as
follows. Let $p_1, \ldots, p_n$ be the p-values associated with the $n$ tests, put these p-values in descending order, and label the results $p_{[1]} \geq p_{[2]} \geq \cdots \geq p_{[n]}$. Beginning with $k = 1$ (step 1), reject all hypotheses if

$$p_{[k]} \leq \frac{\alpha}{k}.$$ 

That is, reject all hypotheses if the largest p-value is less than or equal to $\alpha$. If $p_{[1]} > \alpha$, proceed as follows:

1. Increment $k$ by 1. If $p_{[k]} \leq \frac{\alpha}{k}$, stop and reject all hypotheses having a p-value less than or equal $p_{[k]}$
2. If $p_{[k]} > \frac{\alpha}{k}$, repeat step 1.
3. Repeat steps 1 and 2 until all $n$ hypotheses have been tested.

Because at the $k$th step, hypotheses are tested at the $\alpha/k$ level, this suggests that Hochberg’s method can have relatively low power. Simulation results in section 3 illustrate that this is indeed the case and that the same is true when using Hommel’s method.

The remainder of the paper is organized as follows. Section 2 describes the basic strategy for computing a confidence interval for each $M(Y|X_i) (i = 1, \ldots, n)$ such that the simultaneous probability coverage is approximately $1 - \alpha$. And of course there is related goal of testing the $n$ hypotheses indicated by (3) such that the probability of one or more Type I error probabilities is approximately equal to $\alpha$. Section 3 reports simulation results on how well four variations of the method perform in terms of controlling the probability of one or more Type I errors, one of which is based on the OLS estimator. Some results on power are included as well. An illustration is given in section 4.

## 2 Description of the Proposed Method

The basic strategy mimics the approach used by the two-sample version of Student’s T test. First, determine an appropriate critical value, based on an obvious test statistic, assuming
normality and homoscedasticity and in conjunction with a regression estimator of interest. Then study the impact of non-normality and heteroscedasticity via simulations.

Momentarily consider a single value for the covariate, \( X = x \). Let \( \tau^2 \) denote the squared standard error of \( \hat{Y} = b_0 + b_1 x \), where \( b_0 \) and \( b_1 \) are estimates of \( \beta_0 \) and \( \beta_1 \), respectively, based on some regression estimator to be determined. A basic percentile bootstrap method is used to estimate \( \tau^2 \) (e.g., Efron & Tibshirani, 1993). More precisely, generate a bootstrap sample by randomly sampling with replacement \( n \) pairs of points from \( (X_1, Y_1), \ldots, (X_n, Y_n) \) yielding \( (X_1^*, Y_1^*), \ldots, (X_n^*, Y_n^*) \). Based on this bootstrap sample, estimate the intercept and slope and label the results \( b_0^* \) and \( b_1^* \), which yields \( \hat{Y}^* = b_0 + b_1 x \). Repeat this \( B \) times yielding \( \hat{Y}_1^*, \ldots, \hat{Y}_B^* \), in which case an estimate of \( \tau^2 \) is

\[
\hat{\tau}^2 = \frac{1}{B - 1} \sum (\hat{Y}_k^* - \bar{Y}^*)^2,
\]

where \( \bar{Y}^* = \sum \hat{Y}_k^*/B \). (In terms of controlling the probability of a Type I error, \( B = 100 \) appears to suffice, at least for the situations considered in section 4, but a few results are reported when \( B = 1000 \).) Then, of course, the hypothesis given by (3) can be tested with

\[
W = \frac{\hat{Y} - \theta_0}{\hat{\tau}}
\]

once an appropriate critical value has been determined.

Momentarily assume that \( W \) has a standard normal distribution, in which case a p-value can be determined for each \( X = X_i, \ i = 1, \ldots, n \). Denote the resulting p-values by \( p_1, \ldots, p_n \) and let \( p_m = \min(p_1, \ldots, p_n) \). As is evident, if \( p_\alpha \), the \( \alpha \) quantile of \( p_m \), can be determined, the probability of one or more Type I errors can be controlled simply by rejecting \( i \)th hypothesis if and only if \( p_i \leq p_\alpha \). And in addition, confidence intervals for each \( M(Y|X = X_i) \) can be computed that have simultaneous probability coverage \( 1 - \alpha \).

The distribution of \( p_m \) is approximated in the following manner. Assume both the error term \( \epsilon \) and \( X \) have a standard normal distribution and consider the case \( \beta_0 = \beta_1 = 0 \). Then a simulation can be performed yielding an estimate of the \( \alpha \) quantile of the distribution of \( p_m \). In effect, generate \( n \) pairs of observations from a bivariate normal distribution having correlation zero yielding \( (X_1, Y_1), \ldots, (X_n, Y_n) \). Compute \( p_m \) and repeat this process \( A \) times yielding \( p_{m1}, \ldots, p_{mA} \). Put these \( A \) values in ascending order yielding \( p_{m(1)} \leq \ldots \leq p_{m(A)} \) and
let $k = \alpha A$ rounded to the nearest integer. Then the $\alpha$ quantile of $p_m$, $p_\alpha$, is estimated with $p_{m(k)}$. Moreover, the simultaneous probability coverage among the $n$ confidence intervals

$$\hat{Y}_i \pm z_{\tau_i}(i = 1, \ldots, n)$$

is approximately $1 - \alpha$, where $z$ is the $1 - p_{m(k)}/2$ quantile of a standard normal distribution, $\hat{Y}_i = b_0 + b_1 X_i$ and $\tau_i$ is the corresponding estimate of the standard error.

When dealing with least squares regression, the choice for the slope is irrelevant when estimating $p_\alpha$ because from basic principles (e.g., Montgomery & Peck, 1992, p. 27) the standard error does not depend on the slope. However, it is not completely clear whether this remains the case when dealing with robust regression estimators, so some additional simulations were run for determining $p_\alpha$ when $\beta_1 = 1$. All indications are that altering the slope makes at most a trivial difference. For example, with $n = 50$ and when using the Theil–Sen estimator, the estimate of $p_\alpha$ is 0.01038807. Increasing the slope to $\beta_1 = 1$, the estimate is 0.01044386.

Note that the method for determining the critical value $z$ relaxes somewhat the assumption that $W$ has a standard normal distribution. Consider, for example, the OLS estimator. Assuming $W$ has a standard normal distribution is not quite satisfactory when the sample size is relatively small. However, in terms of the probability of a Type I error, it is the value of $p_{m(k)}$ relative to the $p$-values associated with $W$ that matters. In effect, the simulation used to determine $p_{m(k)}$ adjusts the critical value so that the level of the test is $\alpha$.

### 2.1 Choosing a Regression Estimator

As previously indicated, it is known that OLS is not robust. Some practical consequences are that outliers can result in a misleading summary of the association among the bulk of the points. Also, outliers among the dependent variable can result in a relatively high standard error, which in turn can mean relatively low power compared to various robust estimators that have been derived. Numerous robust estimators have been proposed that are aimed at dealing with the concerns with OLS (e.g., Wilcox, 2017, Chapter 10). No single estimator is optimal, but one that performs relatively well is the Theil (1950) and Sen (1964) estimator. Peng et al. (2008) established that it can be super efficient when
the error term is discontinuous. Its efficiency compares well to OLS under normality and
there are situations where the Theil–Sen estimator has a substantially smaller standard error
(Wilcox, 2017). Accordingly, results based on the Theil–Sen estimator are reported plus
results based on three alternative estimators: the OLS estimator, the quantile regression
estimator derived by Koenker and Bassett (1978) and a slight modification of the Theil–
Sen estimator considered by (Wilcox & Clark, 2013). The modification of the Theil–Sen
estimator has been found to have a practical advantage when there are tied values among
the dependent variable.

For a single independent variable, the Theil–Sen estimator is computed as follows. For
any \( i < i' \), for which \( X_i \neq X_{i'} \), let

\[
S_{ii'} = \frac{Y_i - Y_{i'}}{X_i - X_{i'}}.
\]

The Theil–Sen estimate of the slope is

\( b_{1ts} \), the median of all the slopes represented by \( S_{ii'} \).

Two strategies for estimating the intercept have been proposed. The first is

\[
M_y - b_1 M_x,
\]

where \( M_y \) and \( M_x \) are the usual sample medians of the \( Y \) and \( X \) values, respectively.
The second approach, which is used here, estimates the intercept with the median of \( Y_1 - b_1 X_1, \ldots, Y_n - b_1 X_n \). Wilcox (2017) summarizes various strategies for extending the Thiel–
Sen estimator to more than one independent variable.

The Modified Theil–Sen estimator replaces the usual sample median with the quantile
estimator derived by Harrell and Davis (1982). Let \( V \) be a random variable having a beta
distribution with parameters \( a = (n+1)q \) and \( b = (n+1)(1-q) \). Let

\[
W_i = P\left( \frac{i-1}{n} \leq V \leq \frac{i}{n} \right).
\]

Then the Harrell–Davis estimate of the \( q \)th quantile is

\[
\hat{\theta}_q = \sum_{i=1}^{n} W_i X_{(i)}.
\]

As for the Koenker and Bassett (1978) estimator, it estimates the slope and intercept
with the goal of predicting the \( q \)th quantile of \( Y \) given \( X \). Let

\[
\psi_q(u) = u(q - I_{u<0}),
\]
where $I$ is the indicator function. Then the slope and intercept of the regression line are determined by minimizing

$$\sum \psi_q(r_i),$$

where $r_1, \ldots, r_n$ are the usual residuals. Here the focus is on $q = 0.5$.

For completeness, it is noted that classic inferential methods based on the OLS estimator, which assume homoscedasticity, are based on an incorrect estimate of the standard error of $b_0$ and $b_1$ when in fact there is heteroscedasticity (e.g., Godfrey, 2006; Long & Ervin, 2000). Several theoretically sound methods for estimating standard errors, when there is heteroscedasticity, have been derived (e.g., White, 1980; Cribari-Neto, 2004; Cribari-Neto et al., 2007; Cribari-Neto, Souza & da Silva, 2011). Perhaps one of these estimators can be used in conjunction with some modification of the method used here, which has practical value. But this issue is not pursued here.

### 2.2 Some Estimates of $p_\alpha$

Table 1 shows some estimates of $p_\alpha$ based on $A = 4000$ replications and sample sizes ranging from 10 to 600. As can be seen, the estimate when using OLS depends on the sample size, as might be expected. In contrast, when using the robust regression estimators, the estimate of $p_\alpha$ appears to be virtually independent of $n$. Moreover, the results suggest that in general, as $n \to \infty$, $p_{m(k)} \to 0.10$, or some value approximately equal to 0.10, but there is no formal proof that this is the case.

### 3 Simulation Results

Simulations were used to study the small-sample properties of the method in section 2. The sample sizes considered were 20 and 40. Result for $n = 40$ added no new insights, so for brevity they are not reported. Estimated Type I error probabilities, $\hat{\alpha}$, were based on 4000 replications.

Four types of distributions were used: normal, symmetric and heavy-tailed, asymmetric
Table 1: Estimates of $p_α$ based on 4000 replications

<table>
<thead>
<tr>
<th>n</th>
<th>TS</th>
<th>OLS</th>
<th>TSHD</th>
<th>KB</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.011</td>
<td>0.001</td>
<td>0.009</td>
<td>0.011</td>
</tr>
<tr>
<td>20</td>
<td>0.010</td>
<td>0.004</td>
<td>0.008</td>
<td>0.009</td>
</tr>
<tr>
<td>50</td>
<td>0.010</td>
<td>0.008</td>
<td>0.009</td>
<td>0.009</td>
</tr>
<tr>
<td>100</td>
<td>0.010</td>
<td>0.008</td>
<td>0.009</td>
<td>0.008</td>
</tr>
<tr>
<td>400</td>
<td>0.011</td>
<td>0.011</td>
<td>0.012</td>
<td>0.009</td>
</tr>
<tr>
<td>600</td>
<td>0.010</td>
<td>0.011</td>
<td>0.010</td>
<td>0.010</td>
</tr>
</tbody>
</table>

OLS=Ordinary Least Squares
TS=Theil–Sen
TSHD=Modified Theil–Sen
KB=Koenker–Bassett

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, 1985) that contain the standard normal distribution as a special case. If $Z$ has a standard normal distribution, then

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

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 (1985).

Three choices for $λ$ were used: $λ(X) = 1$, $λ(X) = |X| + 1$ and $λ(X) = 1/(|X| + 1)$. For convenience, these three choices are denoted by variance patterns (VP) 1, 2, and 3. As is evident, VP 1 corresponds to the usual homoscedasticity assumption.

Table 3 summarizes the simulation results when testing (3) at the 0.05 level and the sample size is $n = 20$. Although the seriousness of a Type I error depends on the situation, Bradley (1978) has suggested that as a general guide, when testing at the 0.05 level, at a
Table 2: Some properties of the g-and-h distribution

<table>
<thead>
<tr>
<th>g</th>
<th>h</th>
<th>$\kappa_1$</th>
<th>$\kappa_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>0.00</td>
<td>3.0</td>
</tr>
<tr>
<td>0.0</td>
<td>0.2</td>
<td>0.00</td>
<td>21.46</td>
</tr>
<tr>
<td>0.2</td>
<td>0.0</td>
<td>0.61</td>
<td>3.68</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
<td>2.81</td>
<td>155.98</td>
</tr>
</tbody>
</table>

minimum the actual level should be between 0.025 and 0.075. All of the estimates are less than 0.075 but when dealing with heavy-tailed distributions ($h = 0.2$) and VP 3, estimates drop below 0.025, even when using a robust regression estimator.

As previously indicated, another approach to controlling the probability of one or more Type I errors is to use the improvements on the Bonferroni method derived by Hochberg (1988) or Hommel (1988). Next, simulations are used to contrast the power of these methods with the approach based on (4). Consider the situation where both $X$ and $\epsilon$ have standard normal distributions and $Y = 0.25X + \epsilon$. The goal is to test the $n$ hypotheses indicated by equation (3) with $\theta_0 = 0$. Let $Q$ indicate the number of hypotheses rejected. Using the Hommel method, $E(Q/n)$, the expected proportion of rejections, was estimated to be 0.097. Switching to Hochberg’s method gave exactly the same result. Using instead (4), the estimated proportion was 0.475. However, at least in principle, situations can arise where the Hommel and Hochberg methods reject in contrast to the method based on (4). If, for example, the p-values range between 0.02 and 0.049, then the Hommel and Hochberg methods reject all $n$ hypotheses at the 0.05 level. In contrast, the method based on (4) would fail to reject any of the hypotheses. But in practice, perhaps such a situation is highly unlikely.

4 An Illustration

Consider again the Well Elderly 2 study described in the introduction. As previously indicated, when the CAR (the cortisol awakening response) is positive, a positive association was found between the CAR and a measure of depressive symptoms. But to what extent is this a
Table 3: Estimated probability of one or more Type I errors, $\alpha = 0.05$, $n = 20$

<table>
<thead>
<tr>
<th>$g$</th>
<th>$h$</th>
<th>$VP$</th>
<th>OLS</th>
<th>TS</th>
<th>TSHD</th>
<th>KB</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>1</td>
<td>0.050</td>
<td>0.050</td>
<td>0.050</td>
<td>0.050</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>2</td>
<td>0.071</td>
<td>0.061</td>
<td>0.056</td>
<td>0.049</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>3</td>
<td>0.031</td>
<td>0.032</td>
<td>0.034</td>
<td>0.036</td>
</tr>
<tr>
<td>0.0</td>
<td>0.2</td>
<td>1</td>
<td>0.032</td>
<td>0.036</td>
<td>0.033</td>
<td>0.022</td>
</tr>
<tr>
<td>0.0</td>
<td>0.2</td>
<td>2</td>
<td>0.056</td>
<td>0.042</td>
<td>0.037</td>
<td>0.038</td>
</tr>
<tr>
<td>0.0</td>
<td>0.2</td>
<td>3</td>
<td>0.019</td>
<td>0.018</td>
<td>0.020</td>
<td>0.018</td>
</tr>
<tr>
<td>0.2</td>
<td>0.0</td>
<td>1</td>
<td>0.050</td>
<td>0.045</td>
<td>0.049</td>
<td>0.052</td>
</tr>
<tr>
<td>0.2</td>
<td>0.0</td>
<td>2</td>
<td>0.075</td>
<td>0.058</td>
<td>0.060</td>
<td>0.059</td>
</tr>
<tr>
<td>0.2</td>
<td>0.0</td>
<td>3</td>
<td>0.027</td>
<td>0.033</td>
<td>0.034</td>
<td>0.032</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
<td>1</td>
<td>0.035</td>
<td>0.034</td>
<td>0.031</td>
<td>0.025</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
<td>2</td>
<td>0.056</td>
<td>0.041</td>
<td>0.038</td>
<td>0.041</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
<td>3</td>
<td>0.017</td>
<td>0.018</td>
<td>0.023</td>
<td>0.016</td>
</tr>
</tbody>
</table>

OLS=Ordinary Least Squares
TS=Theil–Sen
TSHD=Modified Theil–Sen
KB=Koenker–Bassett
serious health concern? In particular, is it the case that CESD scores greater than 15 occur for the typical participant over the range of CAR values that are available? Here, leverage points were removed, after which the range of CAR values is 0-0.358. Figure 1 shows the regression line (based on the modified Theil–Sen estimator) along with 0.95 confidence band. The individual confidence intervals indicate that for CAR between zero and 0.10, the typical CESD score is significantly less than 15. Using OLS, the range is zero to 0.07. For CAR greater than 0.23, the typical CESD score is estimated to be greater than 15. But over the entire range of available CAR values, the typical CESD scores are not significantly greater than 15. Perhaps for CAR values greater than 0.358 this would no longer be the case, but this remains to be determined.

5 Concluding Remarks

All indications are that the confidence band based on (4), and the hypothesis testing method based on the test statistic $W$, perform reasonably well, in terms of controlling the Type I error probability and providing reasonably accurate confidence bands, with a bit of an advantage when using a robust estimator. The main difficulty is that when dealing with symmetric, heavy-tailed distributions, the actual level can be substantially lower than the nominal level when dealing with variance pattern VP 3. Another limitation is that there is uncertainty about an appropriate critical value when $n > 600$. A critical value could be determined as described in section 2, with the understanding that execution time can be somewhat high based on current technology.

In principle, the method studied here can be extended to situations where there are two or more independent variables. A few simulations were run with two independent variables, which revealed that some modification of the method studied here is needed. Also, some modification and an extension of the approach used here might prove to be useful when comparing two independent groups and there is a covariate, but this remains to be determined.

Finally, the R function regYci is being added to the R package WRS, which applies the proposed method. The R function regYband creates a plot of the regression line as well as the confidence band.
Figure 1: Regression line and 0.95 confidence band for predicting depressive symptoms based on the cortisol awakening response.
REFERENCES


