



Inferences Based on Robust Regression Estimators When There Is Multicollinearity

Rand Wilcox

Dept. of Psychology
University of Southern California

ABSTRACT

The paper deals with the goal of testing hypotheses about the slope parameters of a linear regression model when there is multicollinearity. A heteroscedastic method was recently derived based on a ridge estimator, but it does not guard against the deleterious impact of outliers. Several robust analogs of the ridge estimator have been proposed that might deal with this concern. The goal here is to find a robust method that performs reasonably well in simulations.

Keywords: regression, multicollinearity, heteroscedasticity

INTRODUCTION

For the random variables Y, X_1, \dots, X_p , let $M(Y|X_1, \dots, X_p)$ denote some (conditional) measure of location associated the distribution of Y , given X_1, \dots, X_p . Here it is assumed that $M(Y|X_1, \dots, X_p)$ is given by the linear regression model

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \lambda(X_1, \dots, X_p)\epsilon,$$

where β_0, \dots, β_p are unknown regression coefficients, ϵ is a random random variable having measure of location $M(\epsilon) = 0$, and $\lambda(X_1, \dots, X_p)$ is some unknown function of X_1, \dots, X_p used to model heteroscedasticity. Consider the common goal of testing

$$H_0: \beta_j = 0 \quad (1)$$

for each j ($j = 1, \dots, p$). A well known concern is that when there is multicollinearity, roughly meaning a strong association among the explanatory variables X_1, \dots, X_p , this can result in relatively large standard errors among the estimates of the regression coefficients.

Hoerl and Kennard (1970) suggested dealing with multicollinearity by replacing the ordinary least squares (OLS) estimator with the so-called ridge estimator. Let $(Y_i, X_{i1}, \dots, X_{ip})$ ($i = 1, \dots, n$) denote a random sample. The ridge estimator consists of finding values for β_0, \dots, β_p that minimize

$$\sum (Y_i - \beta_0 - \sum_{j=1}^p \beta_j X_{ij})^2 + k \sum_{j=1}^p \beta_j^2, \quad (2)$$

where the non-negative bias parameter k is to be determined. Let $\hat{\beta}_j(k)$ denote the ridge estimate of β_j . So $\hat{\beta}_j(0)$ corresponds to the least squares estimator. When $k > 0$ and there is multicollinearity, the ridge estimator achieves a smaller standard error than $\hat{\beta}_j(0)$ at the

expense of increased bias when $\beta_j \neq 0$. Generally, when $\beta_j \neq 0$, $|\hat{\beta}_j(k)| \leq |\hat{\beta}_j(0)|$. That is, the ridge estimator down weights the least squares estimator toward zero (e.g., Montgomery et al., 2012). But when $\beta_j = 0$, $\hat{\beta}_j(k)$, is unbiased making it possible to test (2).

Denote the random sample X_{ij} by the $n \times p$ matrix \mathbf{X} and let $\mathbf{C} = \mathbf{X}'\mathbf{X}$. Given k , and when there is a homoscedastic error term, meaning that $\lambda(X_1, \dots, X_p) \equiv 1$,

$$\text{Var}(\hat{\boldsymbol{\beta}}(k)) = \sigma^2(\mathbf{C} + k\mathbf{I}_p)^{-1}\mathbf{X}'\mathbf{X}(\mathbf{C} + k\mathbf{I}_p)^{-1}.$$

Estimating σ^2 with $\hat{\sigma}^2 = \sum r_i^2 / (n - p - 1)$, where r_1, \dots, r_n are the usual residuals, yields an estimate of the squared standard error of $\hat{\beta}_j(k)$, say s_j^2 . A test statistic is simply

$$T_{jk} = \frac{\hat{\beta}_j(k)}{s_j}, \quad (3)$$

where the null distribution is approximated by a Student's t distribution with $n - p - 1$ degrees of freedom. But when there is heteroscedasticity, this estimate of the squared standard error is incorrect.

In the least squares regression literature, there are several heteroscedastic consistent estimates of the standard error (e.g., Wilcox, 2017). Wilcox (2018) found that in the context of ridge regression, an extension of the HC3 estimator studied by Long and Ervin (2000) performed reasonably well in simulations. Let $\mathbf{V} = \text{diag}(r_i^2 / (1 - h_{ii}))$ ($i = 1, \dots, n$), where

$$h_{ii} = \mathbf{x}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'$$

and \mathbf{x}_i is the i th row of \mathbf{X} . Given k , let

$$\mathbf{S}(k) = (\mathbf{C} + k\mathbf{I}_p)^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{C} + k\mathbf{I}_p)^{-1}.$$

Then $s_{jj}(k)$, the j th diagonal element of $\mathbf{S}(k)$, estimates the squared standard error of $\hat{\beta}_j(k)$. Consequently, Wilcox (2018) proposed testing (2) with

$$W_{jk} = \frac{\hat{\beta}_j(k)}{\sqrt{s_{jj}(k)}}, \quad (4)$$

where the null distribution is taken to be a Student's t distribution with $n - p - 1$ degrees of freedom. The bias parameter was estimated using a method based on results in Kibria (2003). This will be called method RHC henceforth.

Simulations reported by Wilcox (2018) indicate that method RHC controls the Type error probability about as well as using OLS, which corresponds to setting $k = 0$. Moreover, when there is no association among the dependent variables, there is very little difference between the power of RHC and simply using the least squares estimator. As the correlation among the independent variables increases, RHC provides better power, sometimes by a substantial amount.

However, it is well-known that the ridge estimator can be overly influenced by a few outliers among the dependent variable Y . Numerous methods have been proposed and compared

regarding how this issue might be addressed (e.g., Adegoke et al. 2016; Arslan Billor, 2000; Ertaa et al., 2017; Kan et al., 2013; Lukman et al., 2014; Samkar & Alpu, 2010; Kan et al., 2013). The focus has been on minimizing mean squared error. Evidently there are no results regarding how these robust estimators perform when testing (1). Another concern is that even a slight departure from normality can result in OLS having poor power relative to using some robust regression estimator (e.g., Wilcox, 2017), which suggests that a similar concern might arise when using RHC.

One approach toward robust analogs of the ridge estimator is as follows. Let $\tilde{\beta}_j$ be the estimate of β_j based on some robust regression estimator, many of which have been derived (e.g., Wilcox, 2017, Chapter 10). Robust analogs of ridge estimators consist of weighting $\tilde{\beta}_j$ based on some function of \mathbf{X} and the residuals, coupled with some estimate of the bias parameter k , yielding say $\tilde{\beta}_j(k)$. Here, two issues are addressed. The first is finding a method that performs reasonably in terms of controlling the Type I error probability via some robust analog of the ridge estimator. The second is gaining some understanding of how the power of a method based on $\tilde{\beta}_j(0)$ compares to the power of a method based on $\tilde{\beta}_j(k)$ when k is estimated via a method to be described. In particular, how do these methods compare, in terms of both Type I errors and power, as a function of the strength of the association among the independent variables, non-normality and heteroscedasticity?

Extant results indicate that a robust regression estimator ($k = 0$), combined with a percentile bootstrap method, performs relatively well in terms of controlling the Type I error probability (Wilcox, 2017). This suggests that for $k > 0$, a percentile bootstrap method might continue to perform well. Preliminary simulations indicate that this is the case for certain estimators but not others. Due to the high execution time associated with the simulations in section 3, extensive results for all of the many estimators that might be used are not practical. Here the goal is to report simulation results when using the Theil (1950) and Sen (1964) estimator that was found to have two positive features compared to simply using $k = 0$. Some results based on the least absolute deviation (LAD) estimator are reported as well. That is, β_j ($j = 0, \dots, p$) are estimated with the values that minimize $\sum |r_i|$, the sum of the absolute residuals. A positive feature of the LAD estimator is that it guards against the deleterious impact of outliers among the dependent variable. A negative feature is the possible impact of leverage points, meaning points $(Y_i, X_{i1}, \dots, X_{ip})$ for which X_{i1}, \dots, X_{ip} is an outlier among the vectors of independent variables. A simple way of dealing with this issue is to remove any leverage points. Simulations in section 3 suggest that for n small and $p = 2$, using the Theil–Sen estimator is a bit better in terms of both Type I errors and power. The main reason for including the LAD estimator is that when n and p are large, execution time using the Theil–Sen estimator can be a practical concern. Using the LAD estimator reduces execution time substantially.

Some alternative robust ridge estimators were found to be reasonably satisfactory in terms of Type I errors when the error term has a normal distribution and there is homoscedasticity. One was based on the least trimmed squares estimator (Rousseeuw Leroy, 1987) and another was based on the MM-estimator derived by Yohai (1987). Using a slight variation of the robust analog of Liu's (1993, 2003) estimator studied by Ertas et al. (2017) also performed well. However, each replication in the simulation study in section 3 required a little over seven seconds using a MacBook pro with a 2.9 GHz processor. So each simulation in section 3 required a little over four hours based on 2000 replications. Consequently, the main focus here is on a just the Theil–Sen and LAD estimators.

It is noted that two approaches were found to be unsatisfactory under normality. The first was based on the Liu (1993, 2003) estimator used in conjunction with the bootstrap method described in section 3. The actual level was estimated to be 0.005 when testing at the 0.05 level using a sample size of $n = 30$ and a common correlation of $\rho = 0.8$ among the explanatory variables. Power was well below the method described in section 3, so this approach was abandoned. The robust analog of Liu's estimator, derived by Kan et al. (2013), suffered from the same problem. This is not to suggest that this estimator be abandoned. In terms of mean squared error it can perform well. But in terms of testing the hypothesis of a zero slope, currently it seems that some alternative estimator is preferable.

The paper is organized as follows. Section 2 describes the estimator that will be used followed by a description of the method used to test the hypothesis of a zero slope. Section 3 reports simulation results and section 4 illustrates the proposed method.

DESCRIPTION OF THE PROPOSED METHOD

Here, the basic strategy was to use a percentile bootstrap method in conjunction with a robust ridge estimator that is based in part on the Theil (1950) and Sen (1964) regression estimator. When using the Theil–Sen estimator ($k = 0$), this approach has been studied extensively and found to perform reasonably well (e.g., Wilcox, 2017). This suggests that this approach might continue to perform well when $k > 0$.

Let $\mathbf{C} = \mathbf{X}'\mathbf{X}$. The basic ridge estimator is

$$\hat{\boldsymbol{\beta}}(k) = (\mathbf{C} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{y},$$

where \mathbf{I} is the identity matrix. When $k = 0$ it reduces to the ordinary least squares estimator. Here, the estimate of k is based on results in Kibria (2003), which was also used by Lukman (2014) when dealing with a robust analog of the ridge estimator. First, determine the orthogonal matrix \mathbf{D} such that $\mathbf{D}'\mathbf{C}\mathbf{D} = \boldsymbol{\Lambda}$, where $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_p)$ contains the eigenvalues of the matrix \mathbf{C} . Let $\boldsymbol{\gamma} = \boldsymbol{\Lambda}^{-1}(\mathbf{X}\mathbf{D})'\mathbf{Y}$ and let

$$\hat{\sigma}^2 = \frac{1}{n - p - 1} \sum r_i^2,$$

where r_1, \dots, r_p are the usual residuals. Then the estimate of k is taken to be

$$\frac{\hat{\sigma}^2}{(\prod_{j=1}^p \lambda_j)^{1/p}}.$$

Adegoke et al. (2016) used another estimate of k when comparing robust ridge estimators in terms of mean squared error. Indeed, several other estimators are listed by Kibria (2003), perhaps in terms of power these alternative estimators offer a practical advantage, but this remains to be determined.

Following, for example, Kan et al. (2013), a robust ridge estimator is taken to be

$$\tilde{\boldsymbol{\beta}} = (\mathbf{C} + k_R\mathbf{I})^{-1}\mathbf{C}\hat{\boldsymbol{\beta}}_R, \quad (6)$$

where $\hat{\beta}_R$ is some robust regression estimator and k_R is estimated as previously described, but with $\hat{\sigma}^2$ based on the residuals associated with $\hat{\beta}_R$.

As previously noted, the first choice for $\hat{\beta}_R$ considered here is the Theil (1950) and Sen (1964) estimator, which has been studied extensively (e.g., Wilcox, 2017). Momentarily focus on $p = 1$. 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 estimated by the median of all the slopes represented by $S_{ii'}$. For $p > 1$ a back-fitting, Gauss–Seidel method is used (e.g., Wilcox, 2017, section 10.2).

Finally, the null hypothesis is tested as follows. Generate a bootstrap sample by randomly sampling with replacement n vectors from the random sample $(Y_1, X_{i1}, \dots, X_{ip})$ ($i = 1, \dots, n$) yielding $(Y_1^*, X_{i1}^*, \dots, X_{ip}^*)$ ($i = 1, \dots, n$). Compute $\tilde{\beta}_j$ ($j = 1, \dots, p$) based on this bootstrap sample yielding $\tilde{\beta}_j^*$. Repeat this process B times yielding $\tilde{\beta}_{1j}^*, \dots, \tilde{\beta}_{Bj}^*$. Put these B values in ascending order yielding $\tilde{\beta}_{j(1)}^* \leq \dots \leq \tilde{\beta}_{j(B)}^*$. Let A denote the number of bootstrap estimates less than zero. From Liu and Singh (1990), a p-value is given by $2\min(p^*, 1 - p^*)$, where $p^* = A/B$. This will be called method RTS henceforth. The percentile bootstrap can be used to compute a confidence interval, but this is not recommended for the situation at hand due to the bias of the estimator. When $k = 0$, meaning that the Theil–Sen estimator is used instead, this is called method TS. When $\tilde{\beta}(0)$ is taken to be the LAD estimator, the percentile bootstrap method for testing the null hypothesis will be called method RQ.

SIMULATION RESULTS

Simulations were used to study the power and Type I error probabilities when testing at the $\alpha = 0.05$ level using methods RTS, TS and RQ. Estimated Type I error probabilities were based on 2000 replications. Four types of distributions were used: normal, symmetric and heavy-tailed, asymmetric and light-tailed, and asymmetric and heavy-tailed. More precisely, values for the error term, ϵ_j were generated from 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 by definition

$$V = \frac{\exp(gZ) - 1}{g} \exp(hZ^2/2)$$

has a g-and-h distribution. For $g = 0$, this last equation is taken to be

$$V = Z \exp\left(h \frac{Z^2}{2}\right).$$

The four distributions used here were the standard normal ($g = h = 0$), 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 ($g = h = 0.2$). The marginal distributions of \mathbf{X} were taken to be the same g-and-h distribution used to generate the values for the error term. More precisely, values were generated from a multivariate normal distribution having correlation ρ , then the marginal distributions were transformed to the

same g-and-h distribution used to generate the error term. Three values for ρ were used: 0.0, 0.5 and 0.8. Table 1 shows the skewness (κ_1) and kurtosis (κ_2) of the g-and-h distributions used. Additional properties of the g-and-h distribution are summarized by Hoaglin (1985).

Table 1: 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

Three choices for $\lambda(\mathbf{X})$ were used: $\lambda(\mathbf{X}) = 1$ (homoscedasticity), $\lambda(\mathbf{X}) = |X_1| + 1$ and $\lambda(\mathbf{X}) = 1/(|X_1| + 1)$. So two types of heteroscedasticity are being considered. The first is where the larger variation among the dependent variable occurs among the more extreme values of the first independent variable, and the second is the reverse situation. For convenience these three choices for λ will be called variance patterns (VP) 1, 2 and 3, respectively.

Estimated Type I error probabilities are shown in Tables 2 and 3 when testing at the 0.05 level and the sample size is $n = 30$. Although the importance of a Type I error can depend on the situation, Bradley (1978) suggests that as a general guide, when testing at the 0.05 level, the actual level should be between 0.025 and 0.075. None of the estimates exceed 0.05. For method RTS, there are five instances where the estimate is less than 0.025, the lowest estimate being 0.018. Note that generally, there is very little difference between methods RTS and TS. That is, switching to a robust ridge estimator appears to have very little impact on the probability of a Type I error. This indicates that any differences in power have little to do with any differences in the Type I error probabilities. The column headed by RQ contains the results when the robust ridge method is based on the LAD estimator. The LAD estimator reduces execution substantially, but as can be seen, the actual level tends to be less than the level using RTS, sometimes by a substantial amount. Using the LAD estimator, the estimated Type I error probability is less than or equal to 0.03 in all of the situations considered. The lowest estimate was 0.011.

Table 2: Estimated Type I error probabilities, symmetric distributions, $\alpha = 0.05, n = 30$

g	h	ρ	VP	RTS	TS	RQ
0.0	0.0	0.0	1	0.034	0.036	0.022
0.0	0.0	0.0	2	0.047	0.045	0.026
0.0	0.0	0.0	3	0.036	0.033	0.018
0.0	0.0	0.5	1	0.034	0.033	0.020
0.0	0.0	0.5	2	0.037	0.036	0.025
0.0	0.0	0.5	3	0.035	0.033	0.017
0.0	0.0	0.8	1	0.027	0.027	0.019
0.0	0.0	0.8	2	0.028	0.027	0.020
0.0	0.0	0.8	3	0.025	0.026	0.015
0.0	0.2	0.0	1	0.040	0.039	0.023
0.0	0.2	0.0	2	0.046	0.044	0.028
0.0	0.2	0.0	3	0.035	0.033	0.016
0.0	0.2	0.5	1	0.032	0.032	0.023
0.0	0.2	0.5	2	0.036	0.037	0.030
0.0	0.2	0.5	3	0.026	0.030	0.016
0.0	0.2	0.8	1	0.023	0.027	0.018
0.0	0.2	0.8	2	0.027	0.026	0.026
0.0	0.2	0.8	3	0.018	0.027	0.010

Table 3: Estimated Type I error probabilities, skewed distributions, $\alpha = 0.05, n = 30$

g	h	ρ	VP	RTS	TS	RQ
0.2	0.0	0.0	1	0.041	0.039	0.020
0.2	0.0	0.0	2	0.048	0.043	0.026
0.2	0.0	0.0	3	0.037	0.035	0.017
0.2	0.0	0.5	1	0.033	0.033	0.019
0.2	0.0	0.5	2	0.037	0.037	0.025
0.2	0.0	0.5	3	0.030	0.029	0.018
0.2	0.0	0.8	1	0.026	0.025	0.017
0.2	0.0	0.8	2	0.030	0.026	0.025
0.2	0.0	0.8	3	0.024	0.027	0.014
0.2	0.2	0.0	1	0.041	0.039	0.023
0.2	0.2	0.0	2	0.048	0.044	0.030
0.2	0.2	0.0	3	0.033	0.033	0.019
0.2	0.2	0.5	1	0.030	0.032	0.010
0.2	0.2	0.5	2	0.037	0.037	0.017
0.2	0.2	0.5	3	0.028	0.028	0.018
0.2	0.2	0.8	1	0.024	0.026	0.019
0.2	0.2	0.8	2	0.028	0.026	0.028
0.2	0.2	0.8	3	0.018	0.026	0.011

Tables 4 and 5 show the estimated power when $\beta_1 = \beta_2 = 0.5$. The column headed by RTS indicates the probability of rejecting $H_0: \beta_1 = 0$, when using a robust ridge estimator, which in this case is also the probability of rejecting $H_0: \beta_2 = 0$. The column headed by TS is the same probability using the Theil–Sen estimator. That is, $k = 0$ is being used. The column headed by RTS.P2 is the probability that both $H_0: \beta_1 = 0$ and $H_0: \beta_2 = 0$ are rejected using the ridge estimator And TS.P2 is this probability using the Theil–Sen Estimator. As expected, when $\rho = 0.0$, there is little separating the two estimators. As ρ increases, situations are encountered where the ridge estimator offers substantially higher power. Notice, however, that the increase in power depends on more than just the correlation; the type of heteroscedasticity plays a role as well.

It is noted that the power associated with method RHC can be substantially higher or lower than method RTS depending on the distribution associated with the error term plus the strength of the association among the dependent variables. Under normality with $\rho = 0.0$, Table 3 indicates that power is 0.569 for VP 1. Using instead RHC, power is 0.637. For $\rho = 0.8$, RTS has power 0.477 versus 0.710 using RHC. However, a slight departure from normality can reverse this last situation. Consider, for example, the mixed normal distribution discussed by Tukey (1960). There is very little visible difference between it and the standard normal distribution. The Kolmogorov distance between the two distributions is small, about 0.04 (Wilcox, 2017). If the error term has this mixed normal distribution, and $\rho = 0.8$, RTS has power 0.400 versus 0.362 using RHC.

Table 4: Estimated power symmetric distributions, $\beta_1 = \beta_2 = 0.5, n = 30$

<i>g</i>	<i>h</i>	ρ	VP	RTS	TS	RTS.P2	TS.P2	RQ
0.0	0.0	0.0	1	0.569	0.572	0.321	0.330	0.462
0.0	0.0	0.0	2	0.174	0.017	0.043	0.044	0.139
0.0	0.0	0.0	3	0.919	0.918	0.850	0.850	0.893
0.0	0.0	0.5	1	0.573	0.444	0.293	0.147	0.476
0.0	0.0	0.5	2	0.223	0.174	0.044	0.024	0.145
0.0	0.0	0.5	3	0.907	0.818	0.820	0.653	0.871
0.0	0.0	0.8	1	0.477	0.240	0.167	0.018	0.362
0.0	0.0	0.8	2	0.193	0.110	0.024	0.004	0.134
0.0	0.0	0.8	3	0.812	0.510	0.636	0.184	0.715
0.0	0.2	0.0	1	0.578	0.580	0.350	0.357	0.565
0.0	0.2	0.0	2	0.164	0.161	0.049	0.045	0.180
0.0	0.2	0.0	3	0.896	0.897	0.812	0.813	0.935
0.0	0.2	0.5	1	0.604	0.445	0.350	0.155	0.644
0.0	0.2	0.5	2	0.236	0.174	0.054	0.017	0.210
0.0	0.2	0.5	3	0.908	0.816	0.824	0.654	0.937
0.0	0.2	0.8	1	0.556	0.261	0.292	0.024	0.561
0.0	0.2	0.8	2	0.208	0.107	0.039	0.002	0.173
0.0	0.2	0.8	3	0.851	0.536	0.716	0.226	0.873

Table 5: Estimated power, skewed distributions, $\beta_1 = \beta_2 = 0.5, n = 30$

<i>g</i>	<i>h</i>	ρ	VP	RTS	TS	RTS.P2	TS.P2	RQ
0.2	0.0	0.0	1	0.588	0.584	0.353	0.343	0.471
0.2	0.0	0.0	2	0.221	0.218	0.044	0.048	0.145
0.2	0.0	0.0	3	0.917	0.912	0.846	0.843	0.900
0.2	0.0	0.5	1	0.593	0.457	0.230	0.170	0.507
0.2	0.0	0.5	2	0.232	0.177	0.046	0.016	0.165
0.2	0.0	0.5	3	0.911	0.817	0.826	0.652	0.885
0.2	0.0	0.8	1	0.494	0.245	0.191	0.017	0.395
0.2	0.0	0.8	2	0.200	0.105	0.026	0.016	0.140
0.2	0.0	0.8	3	0.824	0.509	0.662	0.190	0.717
0.2	0.2	0.0	1	0.591	0.592	0.362	0.363	0.578
0.2	0.2	0.0	2	0.217	0.215	0.043	0.048	0.187
0.2	0.2	0.0	3	0.896	0.901	0.810	0.082	0.916
0.2	0.2	0.5	1	0.637	0.476	0.399	0.191	0.649
0.2	0.2	0.5	2	0.242	0.179	0.060	0.017	0.210
0.2	0.2	0.5	3	0.908	0.806	0.820	0.635	0.935
0.2	0.2	0.8	1	0.569	0.271	0.296	0.027	0.562
0.2	0.2	0.8	2	0.208	0.107	0.039	0.002	0.177
0.2	0.2	0.8	3	0.846	0.540	0.709	0.233	0.868

As previously noted, the non-robust ridge estimator is sensitive to outliers. To provide some indication the extent this issue is addressed using a robust analog, consider again the situation $g = h = 0$ (normality) $\rho = 0$ and VP 1, only now the sample is contaminated by setting both (Y_1, X_{11}, X_{12}) and (Y_2, X_{21}, X_{22}) equal to $(-4, 1, 1)$. For $\beta_1 = \beta_2 = 0.5$, power using method RHC drops from 0.674 to 0.036. Using instead, RTS, power is 0.265.

AN ILLUSTRATION

Data from the Well Elderly 2 study (Clark et al., 2011) are used to illustrate the proposed method and how it compares to using the Theil–Sen estimator versus the LAD estimator. A general goal in the Well Elderly 2 study was to assess the efficacy of an intervention strategy aimed at improving the physical and emotional health of older adults. One specific goal was to understand the association between a measure of life satisfaction (the dependent variable) and

two explanatory variables: a measure of emotional stress and a measure of depressive symptoms. Using the Theil–Sen estimator, the p-values for the two slopes are <0.001 and 0.017 respectively. Using instead the robust ridge estimator, they are <0.001 and 0.010 . So the second independent variable (depressive symptoms) is significant at the 0.01 level when using the ridge estimator, but not when using the Theil–Sen estimator, the only point being that the choice of method might make a practical difference. Using the robust ridge estimator based on the LAD estimator, method RQ, now the p-value is 0.080 , and it is 0.053 using the MM-estimator. As for the method RHC, which reduces to OLS when $k = 0$, the p-value for the second independent variable is 0.024 . Using the OLS estimator it is 0.029 . So the p-value ranges between 0.01 and 0.08 , depending on which method is used again illustrating that the choice of method can be crucial. In this particular instance, method RTS has the lowest p-value, but this is not suggest that it dominates in terms of power.

CONCLUDING REMARKS

In summary, in terms of controlling the Type I error probability, method RTS performs relatively well and generally better than method RQ. A possible practical concern is that RTS can have high execution time when n or p is relatively large. Using method RQ reduces execution time considerably but possibly at the expense of lower power. In terms of power, no method dominates and the choice of method can make a substantial difference. Under normality and homoscedasticity, RHC can have a clear advantage in terms of power, but a few outliers can destroy power.

An open issue is whether a robust method can be found that competes well with method RHC, in terms of power, under normality and homoscedasticity. Several variations of method RTS were considered and found to be unsatisfactory. Using the MM-estimator does not address this issue. The skipped regression estimator in Wilcox (2017, section 10.10) was unsatisfactory. This approach looks for outliers using a projection method, eliminates any that are found, and applies the Theil–Sen estimator using the remaining data. Using instead the OLS estimator after outliers are removed also performed poorly. Another type of skipped estimator (Wilcox, 2017, section 10.13.3) has excellent properties, but when used in the context of a robust ridge estimator, again power was unsatisfactory.

Finally, the R function `rob.ridge.test` performs method RTS and is being added to the R package WRS.

References

- Adegoke, A. S., Adewuyi, E., Ayinde, K. & Lukman, A. F. (2016). A comparative study of some robust ridge and Liu estimators. *Science World Journal*, 11, No. 4. ISSN 1597-6343.
- Arslan, O. & Billor, N. (2000). Robust Liu estimator for regression based on an M-estimator. *Journal of Applied Statistic*, 27, 39–47.
- Bradley, J. V. (1978) Robustness? *British Journal of Mathematical and} Statistical Psychology*, 31, 144–152.
- Clark, F., Jackson, J., Carlson, M., Chou, C.-P., Cherry, B. J., Jordan-Marsh M., Knight, B. G., Mandel, D., Blanchard, J., Granger, D. A., Wilcox, R. R., Lai, M. Y., White, B., Hay, J., Lam, C., Marterella, A. & Azen, S. P. (2011). 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, 782–790. doi:10.1136/jech.2009.099754
- Ertas, H., Kaçiranlar, S. & Güer (2017). Robust Liu-type estimator for regression based on M-estimator. *Communications in Statistics–Simulation and Computation*, 46, 3907–3932.
- Hoaglin, D. C. (1985). Summarizing shape numerically: The g-and-h distribution. In D. Hoaglin, F. Mosteller & J. Tukey (Eds.) *Exploring Data Tables Trends and Shapes*. New York: Wiley, pp. 461--515.

- Hoerl, A. E. & Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, 12, 55–67.
- Kan, B., Alpu, Ö. & Yazici, B. (2013). Robust ridge and robust Liu estimator for regression based on the LTS estimator *Journal of Applied Statistics*, 40, 644–655
- Kibria, B. M. G (2003). Performance of Some New Ridge Regression Estimators. *Communication in Statistics-Simulation and Computation*, 32, 419–435. DOI: 10.1081/SAC-120017499.
- Liu, K. (1993). A new class of biased estimate in linear regression. *Communications in Statistics–Theory and Methods*, 22, 393–402.
- Liu, K. (2003). Using Liu-type estimator to combat collinearity. *Communications in Statistics–Theory and Methods*, 32, 1009–1020.
- Liu, R. G. & Singh, K. (1997). Notions of limiting P values based on data depth and bootstrap. *Journal of the American Statistical Association*, 92, 266–277.
- Long, J. S. & Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *American Statistician*, 54, 217–224.
- Lukman, A., Arowolo, O. & Ayinde, K. (2014). Some robust ridge regression for handling multicollinearity and outlier. *International Journal of Sciences: Basic and Applied Research*, 16, 192–202.
- Montgomery, D. C., Peck, E. A. & Vining, G. G. (2012). *Introduction to Linear Regression} Analysis*, 5th Ed.: New York: Wiley.
- Rousseeuw, P. J. & Leroy, A. M. (1987). *Robust Regression & Outlier Detection*. New York: Wiley.
- Samkar, H. & Alpu, "0}. (2010). Ridge regression based on some robust estimators. *Journal of Modern Applied Statistical Methods*, 9, Issue 2 , Article 17. DOI: 10.22237/jmasm/1288584960
- Sen, P. K. (1968). Estimate of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, 63, 1379–1389.
- Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis. *Indagationes Mathematicae*, 12, 85–91.
- Tukey, J. W. (1960). A survey of sampling from contaminated normal distributions. In I. Olkin, S. Ghurye, W. Hoeffding, W. Madow & H. Mann (Eds.) *Contributions to Probability and Statistics*. Stanford, CA: Stanford University Press (pp. 448-485).
- Wilcox, R. R. (2017). *Introduction to Robust Estimation and Hypothesis Testing} 4th Ed.* San Diego, CA: Academic Press
- Wilcox, R. R. (2018). *Multicollinearity and Ridge Regression: Results on Type I Errors, Power and Heteroscedasticity*. Technical report, Dept. of Psychology, University of Southern California.
- Yohai, V. J. (1987). High breakdown point and high efficiency robust estimates for regression. *The Annals of Statistics*, 15, 642–656.