Journal of Modern Applied Statistical Methods

Volume 18 | Issue 1 Article 5

2-20-2020

Regression When There Are Two Covariates: Some Practical Reasons for Considering Quantile Grids

Rand Wilcox
University of Southern California, rwilcox@usc.edu

Recommended Citation

Wilcox, R. (2019). Regression when there are two covariates: Some practical reasons for considering quantile grids. Journal of Modern Applied Statistical Methods, 18(1), eP3227. doi: 10.22237/jmasm/1556670120

Regression When There Are Two Covariates: Some Practical Reasons for Considering Quantile Grids

Erratum

A previous version of this article had a character encoding error with epsilon in equation (1). This has now been corrected.

Journal of Modern Applied Statistical Methods
May 2019, Vol. 18, No. 1, eP3227
doi: 10.22237/jmasm/1556670120

בס"ד

ISSN 1538 − 9472

INVITED ARTICLE

Regression When There Are Two Covariates: Some Practical Reasons for Considering Quantile Grids

Rand Wilcox
University of Southern California
Los Angeles, CA

When dealing with the association between some random variable and two covariates, extensive experience with smoothers indicates that often a linear model poorly reflects the nature of the association. A simple approach via quantile grids that reflects the nature of the association is given. The two main goals are to illustrate this approach can make a practical difference, and to describe R functions for applying it. Included are comments on dealing with more than two covariates.

Keywords: Robust methods, smoothers, interactions, trimmed means, binary data, regression trees

Introduction

A fundamental issue is determining whether two variables, (X_1,X_2), are associated with some dependent variable Y. And there is the related goal of gaining some insight into the nature of the association if one exists. Certainly, the best-known and most commonly used strategy is to assume

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \lambda\left(X_1, X_2\right)\epsilon \tag{1}$$

where $\lambda(X_1,X_2)$ is some unknown function used to model heteroscedasticity and ϵ is some random variable with variance σ^2 and $E(\epsilon) = 0$. The null hypothesis to be tested is

doi: 10.22237/jmasm/1556670120 | Accepted: May 13, 2019; Published: February 20, 2020.
Correspondence: Rand Wilcox, rwilcox@usc.edu

$$H_0 : \beta_j = 0 \tag{2}$$

for each j (j = 1,2) using the least squares regression estimator in conjunction with some method that allows heteroscedasticity.

There are, however, concerns. First, the least squares regression estimator is not robust: outliers among the covariate values (leverage points) can result in estimates of the slopes that poorly reflect the nature of the association among the bulk of the participants. Moreover, outliers among the dependent variable Y can result in relatively poor power. Numerous methods have been derived aimed at dealing with these two concerns (e.g., Wilcox, 2017).

Second, there is the issue of multicollinearity: a sufficiently strong association between the two explanatory variables can negatively impact power when testing (2). This remains the case even when using a robust regression estimator. There are inferential methods for dealing with this issue via a ridge estimator (e.g., Wilcox, 2018, 2019) assuming that a linear model is reasonable. However, these methods are limited to making decisions about whether any of the independent variables are associated with the dependent variable. There is no known method, based on a ridge estimator, for making inferences about which slopes differ from zero.

Third, the linear model given by (1) might provide a poor indication of where and how the two covariate variables are related to the dependent variable Y. One way of trying to justify a linear model is to test the hypothesis that a linear model is correct, which can be done using results stemming from Stute et al. (1998). The R function `lintest` in Wilcox (2017) applies this method. However, it is unclear when this approach will have enough power to detect a situation where a linear model is inadequate. Data from the Well Elderly 2 study (Clark et al., 2011) are used to illustrate this point. The basic goal was to assess the impact of an intervention program aimed at improving the physical and emotional wellbeing of older adults. The focus is on the association between a measure of meaningful activities (MAPA) and two covariates: a measure of life satisfaction (LSIZ) and the cortisol awakening response (CAR), which is the change in the cortisol level upon awakening and measured again 30-45 minutes later. Past studies (e.g., Bhattacharyya et al., 2008; Chida & Steptoe, 2009) indicate that measures of stress are associated with the CAR. The sample size is n = 246. Testing the hypothesis the linear model is correct, the p-value is 0.45. Least squares regression yields a significant association for LSIZ (p-value < 0.001) but not for the CAR (p-value = 0.68). The HC4 method was used to deal with heteroscedasticity in conjunction with a projection-type method for dealing with leverage points (e.g., Wilcox, 2017, section 10.1.1). Switching to the Theil (1950) and Sen (1968)

estimator in conjunction with a percentile bootstrap method, again LSIZ is significant (p-value < 0.001) and CAR is not (p-value = 0.51).

However, consider Figure 1, which shows an approximation of the regression surface using the smoother derived by Cleveland and Devlin (1988), which was applied via the R function `lplot` in Wilcox (2017). (Leverage points were removed.) Notice that for relatively high LSIZ scores, it appears that the CAR does indeed have little or no association with MAPA. But for relatively low LSIZ scores and CAR less than zero, CAR appears to play a role. Figure 1 raises the concern that an association might have been missed despite the fact that a test of the hypothesis that a linear model is correct failed to reject. Results reported later in this paper indicate that indeed this is the case.

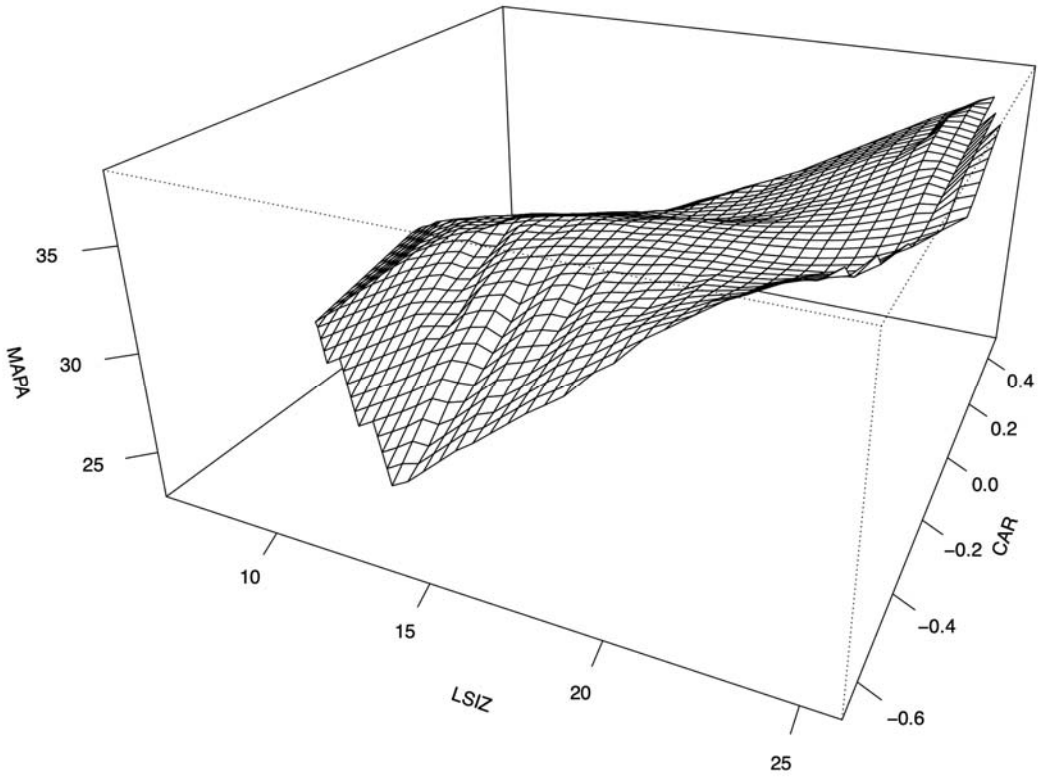

Figure 1. A smooth depicting the association between MAPA (a measure of meaningful activities) and two covariates, namely LSIZ (life satisfaction) and CAR (the cortisol awakening response).

The main goal is to deal with the third issue illustrated by Figure 1, but the approach discussed here is relevant to the first two issues as well. The basic strategy is very simple and has an obvious similarity to regression trees, which are described for example by Hastie et al. (2001). In particular, divide the data into groups based on quantile grids associated with the two explanatory variables and then apply some relevant method for comparing the resulting groups. Here, the choice of grids is based, in some situations, on a smooth as will be illustrated. This in contrast to regression trees where splits are made based on a variation of the least squares method. Another difference is that the goal is not to simply predict the typical value of the dependent variable for points within a given region of the dependent variables. Bagging and random forests (e.g., James et al., 2017) based on regression trees, as well as smoothers (e.g., Wilcox, 2017), are better suited for this purpose. Rather, in addition to establishing an association, the goal is to characterize how regions compare in a simple and readily interpretable manner. For instance, for the data used in Figure 1, imagine that CAR and LSIZ are split at their medians. For low LSIZ scores, let θ_1 and θ_2 denote the median MAPA score for the low and high regions of the CAR. In a similar fashion, for high LSIZ scores, let θ_3 and θ_4 denote the median MAPA score for the low and high regions of the CAR. The question arises whether there is an interaction, meaning whether the data provide reasonably strong evidence that $\theta_1 - \theta_2 \neq \theta_3 - \theta_4$.

The general strategy of using quantile grids will be called method QS henceforth. The simplest version is to split the data based on the median of a single covariate. Another approach is to split the values for the dependent variables into four groups based on the medians of X_1 and X_2 resulting in a 2-by-2 ANOVA design. Of course, other quantiles might be used and each independent variable could be split based on several quantiles as well. For example, the first independent variable might be split into three groups based on the tertiles. An advantage of splitting the data based on two covariates is that interactions can be studied that are missed by splitting the data using a single covariate only. Henceforth, the focus is on splitting the data using both covariates.

Obviously, QS loses the fine detail rendered by a smoother or a linear model when the linear model is correct. Moreover, when the linear model is correct, QS will have less power in terms of establishing an association. But when the linear model is incorrect, QS has the potential of detecting an association that is otherwise missed, and it can provide at least some detail about the nature of the association.

There is an extensive literature on comparing measures of location using robust methods that deal with situations where Y has a skewed or heavy-tailed distribution (e.g., Wilcox, 2017). Roughly, robust methods refer to techniques that

are not substantially altered by a small change in a distribution. For example, a slight departure from normality should not destroy power. Methods based on means are not robust in part because the population variance is not robust. A slight shift away from a normal distribution can inflate the population standard deviation tremendously (e.g., Staudte & Sheather, 1990). Non-normality does not necessarily imply that methods based on means will have poor control over the probability of a Type I error, or have relatively low power. But it is well established that non-normality can be a serious concern when, for example, distributions have different amounts of skewness or when dealing with situations where outliers tend to occur (e.g., Wilcox, 2017).

There are in fact several variations of QS that might be used some of which are summarized here. The primary issue is whether using QS ever makes a practical difference. Several illustrations are provided indicating that the answer is yes. Clearly this is not always the case. But the reality is that a linear model can be highly inadequate in which case switching to QS might yield useful information.

A related issue is dealing with situations where Y is binary. A logistic regression model is one way to proceed but experience with a smoother reveals that this approach can be misleading (e.g., Wilcox, 2017). There is an extensive literature on comparing independent groups when Y is binary (e.g., Storer & Kim 2006; Beal, 1987; Kulinskaya et al., 2010). That is, again method QS can have practical value as illustrated later in this paper.

Using A Trimmed Mean

There are concerns about the robustness of both the population mean and the sample mean (e.g., Wilcox, 2017). The median deals with these issues, there are situations where the sample median provides substantially higher power than the mean, but there are situations where it trims too many values. The focus here is on a compromised amount of trimming: 20%. This is not to suggest that 20% trimming is always optimal. No measure of location is always optimal. In the event there is some reason for choosing some other measure of location, this is easily done for the situation at hand.

Let $Z_1, \ldots, Z_n$ be any n observations. The γ-trimmed mean is $\Sigma_{i=g+1}^{n-g} Z_{(i)} / (n-2g)$, where $Z_{(1)} \leq \cdots \leq Z_{(n)}$ are the Z values written in ascending order and g is the greatest integer less than or equal to γn, $0 \leq \gamma < 0.5$. The 20% trimmed mean corresponds to $\gamma = 0.2$.

One approach to comparing two independent groups is to use a method derived by Yuen (1974). It is based in part on the Winsorized values. The γ Winsorized values $W_1, \ldots, W_n$ corresponding to $Z_1, \ldots, Z_n$ are computed as follows:

$$\begin{aligned} W_i &= Z_{(g+1)} \quad \text{if } Z_i \le Z_{(g+1)} \\ W_i &= Z_i \quad \text{if } Z_{(g+1)} < Z_i < Z_{(n-g)} \\ W_i &= Z_{(n-g)} \quad \text{if } Z_i \ge Z_{(n-g)} \end{aligned}$$

The Winsorized variance is just the usual sample variance based on the Winsorized values.

For two independent groups, let n_j denote the sample size and let g_j indicate the value of g (the number of observations trimmed from each tail) for the jth group ($j = 1,2$). Let $\overline{Y}_j$ denote the trimmed mean and let s_j^2 denote the Winsorized variance. The squared standard error of $\overline{Y}_j$ is estimated with

$$q_j = \frac{\left(n_j - 1\right)s_j^2}{h_j\left(h_j - 1\right)}$$

where $h_j = n_j - 2g_j$ is the number of observations left after trimming. The null distribution of the test statistic

$$\frac{\overline{Y}_1 - \overline{Y}_2}{\sqrt{q_1 + q_2}}$$

is approximated with a Student's t distribution with degrees of freedom $\hat{\nu} = \left(q_1 + q_2\right)^2 / Q$, where

$$Q = \frac{q_1^2}{h_1 - 1} + \frac{q_2^2}{h_2 - 1}$$

Yuen's test has been studied extensively (e.g., Wilcox, 2017, section 5.3). With 20% trimming, power is nearly the same as no trimming (using means) under normality. But of course, no single method dominates in terms of power. Theory and simulations indicate that under general conditions, the ability to control the probability of a Type I error increases as the amount of trimming increases. But

situations can be found where control over the Type I error probability is unsatisfactory when using Yuen's method. This can occur when dealing with skewed distributions, there is heteroscedasticity, the minimum sample size is small, and the difference between the sample sizes is sufficiently large (e.g., Wilcox, 2017, section 5.3.2). This issue can be addressed by switching to a percentile bootstrap method. Even when the sample sizes are equal, but small, a percentile bootstrap method might perform better than Yuen's method.

A percentile bootstrap method is applied as follows. First, generate a bootstrap sample from each group. That is, sample with replacement n_j observations from group j. Based on these bootstrap samples, compute the difference between the trimmed means and label the result D. Repeat this process B times yielding $D_1, \ldots, D_B$. Let A denote the number of D values less than zero and let $p^* = A/B$. From Liu and Singh (1997), a (generalized) p-value is given by $2\min\{p^*, 1 - p^*\}$. Let $D_{(1)} \leq \ldots \leq D_{(B)}$ denote the D values written in ascending order. Then a $1 - \alpha$ confidence interval is $\left(D_{(\ell+1)}, D_{(u)}\right)$, where $\ell = \alpha B/2$ rounded to the nearest integer and $u = B - \ell$. When dealing with the median, a percentile bootstrap method performs very well and is currently the best method for dealing with tied values (Wilcox, 2006). Now $p^* = (A + 0.5C)/B$, where C is the number of times a bootstrap sample from each group yielded the same value for the median.

Consider where the data are split into four groups based on the medians associated with the two independent variables. Focus on whether there is an association between Y and the first covariate. For low values of the second covariate, the typical Y value corresponding to the two regions associated with the first covariate should be the same if there is no association. This can be described more formally as follows. For the random sample $(Y_i X_{i,1}, X_{i,2})$ let $K = \{i{:}X_{i,2} < M_2\}$ $(i = 1, \ldots, n)$, where M_j is the median based on the jth independent variable. Let $K_1 = \{i{:}i \in K \;\&\; X_{i,1} < M_1\}$ and $K_2 = \{i{:}i \in K \;\&\; X_{i,1} \geq M_1\}$. Let $\overline{Y}_{11}$ and $\overline{Y}_{21}$ denote the sample trimmed means based on Y_i values, $i \in K_1$ and $i \in K_2$, respectively, and let μ_{11} and μ_{21} denote the corresponding population trimmed means. Then if the first independent variable has no association with Y, the null hypothesis

$$H_0 : \mu_{11} = \mu_{21} \tag{3}$$

is true. In a similar manner, let $L = \{i{:}X_{i,2} \geq M_2\}$, $L_1 = \{i{:}i \in L \;\&\; X_{i,1} < M_1\}$ and $L_2 = \{i{:}i \in L \;\&\; X_{i,1} \geq M_1\}$. Let $\overline{Y}_{12}$ and $\overline{Y}_{22}$ denote the sample trimmed means based on Y_i values, $i \in L_1$ and $i \in L_2$, respectively. Denote the corresponding

population trimmed means with μ_{12} and μ_{22}. If the first independent variable has no association with Y, then

$$H_0 : \mu_{12} = \mu_{22} \tag{4}$$

is true. So rejecting either (3) or (4) indicates an association with the first independent variable and estimates of the trimmed means provide some sense about the nature of the association. A similar approach can be used to test the hypothesis that the second independent variable has no association. An interaction can be tested as well. That is, one can test

$$H_0 : \mu_{11} - \mu_{21} = \mu_{12} - \mu_{22} \tag{5}$$

using a simple extension of Yuen's method or the percentile bootstrap method (e.g., Wilcox, 2017, sections 7.4.1 and 7.4.9).

When testing (3) and (4), there is the issue of controlling the probability of one or more Type I errors. Here this issue is addressed using the method in Hochberg (1988). When testing at the α level, reject both hypotheses if the maximum of the corresponding p-values is less than or equal to α. If the maximum is greater than α, but the minimum is less than or equal to $\alpha/2$, reject the corresponding hypothesis. When testing more than two hypotheses, p-values can be adjusted via Hochberg's method with the R function `p.adjust`.

This method is readily generalized to situations where data are split into groups based on other quantiles as well as multiple quantiles. Using multiple quantiles provides a more detailed understanding of the nature of any association at the expense of possibly less power due to smaller samples in each group. Controlling the familywise error rate via Hochberg's method reduces power as well. But situations are encountered where this approach has practical value as will be illustrated.

Dealing with Binary Data

When Y is binary, method QS is readily adapted to this situation. Now the goal is to compare the probability of success corresponding to two independent binomial distributions. Many methods have been proposed for dealing with this goal (e.g., Wilcox, 2017, section 5.8). Here the focus is on the Storer and Kim (1990) method, which appears to perform relatively well in terms of both Type I errors and power.

Let r_1 and r_2 denote the number of successes, $\hat{p}_j = r_j / n_j$, and let $\hat{p} = (r_1 + r_2)/(n_1 + n_2)$ be the estimate of the common probability of success assuming the null hypothesis $H_0 : p_1 = p_2$ is true. For any integer u_j, $0 \leq u_j \leq n_j$, let $v_j = u_j / n_j$ and $a_{jk} = 1$ if $|v_j - v_k| \geq |\hat{p}_1 - \hat{p}_2|$; otherwise $a_{jk} = 0$. The p-value is

$$\sum\sum a_{jk} b(u_j, n_1, \hat{p}) b(u_k, n_2, \hat{p})$$

where b is the binomial probability function. A negative feature of the Storer–Kim method is that a confidence interval is not readily computed. A method derived by Kulinskaya et al. (2008) performs relatively well in terms of computing a confidence interval at the possible expense of less power compared to the Storer–Kim method. Interactions and other linear contrasts can be addressed using results in Zou et al. (2009).

Based on prior simulation evidence, if the linear model is correct it performs better than QS in terms of power. As for a binary dependent variable, if the logistic regression model is correct, it performs better than QS as well. Situations exist where QS has more power than these parametric regression models. What is seemingly more important is determining whether QS ever makes a practical difference when dealing with data from a study. Also, there is the issue of developing software that makes method QS easy to use.

Illustrations

The first illustration is based on the data from Well Elderly 2 study shown in Figure 1 and previously described. No association with the CAR was found using OLS with a linear model, p-value = 0.68. And the same was true using a robust regression estimator, p-value = 0.51. However, QS paints a different picture when both independent variables are split at their medians. In particular, testing at the 0.05 level based on 20% trimmed means, the CAR has a signficant association with MAPA when LSIZ is less than its median value. The p-value is 0.014 and adjusting it via Hochberg's method with the goal of controlling FWE for the two hypotheses being tested, now the p-value is 0.028. Moreover, an interaction is indicated, the p-value is 0.012. Using instead the median of the MAPA scores, now the p-value is 0.008. Roughly, if the CAR is negative (cortisol increases after awakening), there appears to be an association when LSIZ is relatively low. For LISZ greater than its median, again no association is found.

There is a method for testing for an interaction based on the generalized additive model (e.g., Wilcox, 2017, section 11.6.3). That is, a parametric model for an interaction is not used. Using the Kolmogorov version of this method, p-value = 0.47. If the interaction is modeled using OLS and a linear model that includes the product of the independent variables, now the p-value is 0.26. One basic concern with this last approach is that in general it does not provide a sufficiently flexible way of modeling an interaction. Using QS, there is a significant interaction.

Next, the same two independent variables are used again, and the dependent variable is taken to be a measure of perceived health. Again, both OLS and the Theil–Sen estimator find no association with the CAR, p-value = 0.46 and 0.25, respectively. However, using QS with median splits, again a significant association with the CAR is found when LSIZ is less than its median.

The next illustration is based on a study dealing with kyphosis, a postoperative spinal deformity. The data are stored in the R variable `rpart::kyphosis`, which reports the presence or absence of kyphosis versus the age of the patient, in months, the number of vertebrae involved in the spinal operation, and a variable called start, which is the beginning of the range of vertebrae involved. The sample size is $n = 81$. Here the focus is on age and start. Using a logistic regression model, start is significant, (p-value < 0.001) but age is not (p-value = 0.17). However, examination of a smooth (using the R function `logSM` in Wilcox, 2017) suggests that the association with age is not monotonic for low start values. Splitting start at its median, the p-value for age using QS is 0.039 for the lower values of start and 0.97 for the higher values. Splitting start based on an estimate of the 0.25 quantile, now for the lower start values the p-value for age is 0.019.

Efron et al. (2004) analyzed data dealing with diabetes, which are available via the R package `lars`. There were ten baseline variables: age, sex, body mass index, average blood pressure, and six blood serum measurements. The sample size is $n = 442$. The dependent variable was a measure of disease progression after one year. The focus is on using two independent variables, where the first is age and the second is any of the remaining independent variables. Using OLS, generally age is found to be significant at the 0.05 level when any other independent variable is included in the model with two exceptions: BMI and the fifth serum measurement. Using QS, for BMI scores above the median, the p-value for age is 0.015, and it is 0.167 for BMI scores below the median. As for the fifth serum measurement it remains non-significant at the 0.05 level. The lowest p-value was 0.078, which occurred for serum measures above the median.

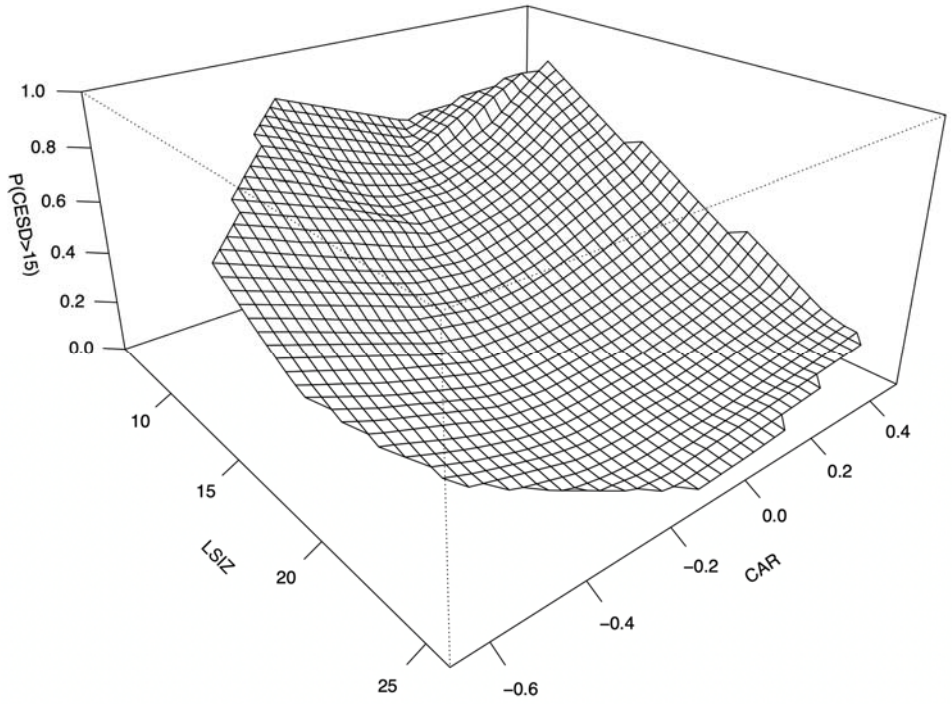

Figure 2. Estimate of the probability of mild or worse depressive symptoms as a function of LSIZ and the CAR.

Consider another example where splitting an independent variable at the median might miss an association. Returning to the Well Elderly study, one of the variables (CESD) measured depressive symptoms. A CESD score greater than 15 is taken to be in indication of mild depression or worse. The goal here is to understand the association between LSIZ and the CAR in terms of P, the probability that CESD is greater than 15. Splitting both independent variables at their median, no association with the CAR is found. However, consider Figure 2, which shows a smooth of the regression surface based on the R function `logSM` previously mentioned. This plot suggests that for low LSIZ scores, P tends to be relatively low when the CAR is near zero, which is close to the median CAR value, –0.03. Moreover, P tends to increase as CAR moves away from its median. Based on this plot, it is not surprising that no association with the CAR is found when splitting the data based on the median CAR value. However, splitting data based on estimates of the 0.6 and 0.8 quantiles of the CAR, two significant results are obtained, both of which correspond to LSIZ scores between 6 and 19. The first

occurs where the first group corresponds to CAR values between –0.675 and 0.001 versus the group where the CAR is between 0.11 and 0.41, p-value = 0.009. The second occurs where for the first group, the CAR values that are between 0.012 and 0.107, and for the second group where CAR is between 0.11 and 0.40, p-value = 0.003.

Software

R functions have seen written that perform method QS. The R function

```
smtest(x,y,IV=1,Qsplit=.5,nboot=1000,est=tmean,tr=.2,PB=FALSE,
     xout=FALSE, outfun=outpro,SEED=TRUE,...)
```

splits the data based on a single covariate and compares the measures of location indicated by the argument est. The argument `x` is a matrix with p columns, where p is the number of independent variables. The argument `IV=1` indicates that a split will be made based on the first independent variable. Setting `xout=TRUE`, leverage points are removed via the function indicated by the argument `outfun`. `PB=TRUE` means that a percentile bootstrap method will be used. The number of bootstrap samples is controlled by the argument `nboot`. With `PB = FALSE`, Yuen's method is used. Using `est=median`, or `hd` (for the Harrell-Davis estimator), a percentile bootstrap method is used automatically. For relatively light-tailed distributions, estimating the median with Harrell and Davis (1982) estimator might provide more power compared to using the usual sample median. The argument `est` indicates the measure of location, which defaults to a 20% trimmed mean. The amount of trimming is controlled by the argument `tr`. For binary data, use the R function

```
smbin.test(x,y,IV=1,Qsplit=.5,method='SK',nboot=1000,xout=FALSE,
     outfun=outpro,SEED=TRUE,...)
```

The argument `method` indicates which method will be used. The default is `SK`, which is the Storer—Kim method. To get confidence intervals, use `method = 'KMS'`.

When splitting based on two independent variables, use the R function

```
smgridRC (x, y, IV = c(1, 2), Qsplit1 = 0.5, Qsplit2 = 0.5, tr = 0.2,
     alpha = 0.05, PB = FALSE, est = tmean, nboot = 1000, pr = TRUE,
     method = "hoch", xout = FALSE, outfun = outpro, SEED = TRUE, ...).
```

The argument `IV` indicates which two covariates will be used. `Qsplit1` (`Qsplit2`) indicates the quantile used to split the first (second) covariate. By default, the median is used. `Qsplit1=c(0.33,0.67)` for example would use the tertiles associated with the first independent variable. Again the amount of trimming is controlled by the argument `tr`.

A portion of the output for the data used in Figure 1 looks like this:

```
$Res.4.IV1
     psihat    ci.lower   ci.upper   p.value      p.adjust     Est.1    Est.2
[1,] -1.281377 -3.220148  0.6573951  1.917055e-01 1.917055e-01 33.02632 34.30769
[2,] -5.101832 -7.199571  -3.0040920 6.971174e-06 1.394235e-05 30.64103 35.74286

$Res.4.IV2
     psihat    ci.lower   ci.upper   p.value      p.adjust     Est.1    Est.2
[1,] 2.385290  0.4991086  4.2714717  0.01393357   0.02786714   33.02632 30.64103
[2,] -1.435165 -3.5796385 0.7093088  0.18637437   0.18637437   34.30769 35.74286
```

Consider the results labeled `$Res.4.IV2`. These are the results for the second covariate, which here corresponds to the CAR. The first line summarizes the results when the first covariate (LSIZ) has a value less than its median and the goal is to compare the two regions associated with the CAR. The estimate of the trimmed mean when the CAR is below its median is 33.03, and it is 30.64 when it is above its median. The difference between these to values is listed under `psihat`. The difference is significant at the 0.05 level, the p-value is 0.014. That is, the data indicate that for low LSIZ scores, typical MAPA scores are higher when the CAR is negative (cortisol increases after awakening). The value under `p.adjust` is the p-value adjusted by Hochberg's method. The values under `ci.low` and `ci.upper` indicate the lower and upper ends of the confidence for the difference between the population trimmed means. The second line reports results when LSIZ has a value greater than its median. Now the estimates of the trimmed means corresponding to low and high values for the CAR are 34.3 and 35.74, respectively; the difference is not significant at the 0.05 level.

Now consider the results under `$Res.4.IV1`. The roles of LSIZ and CAR are reversed. The first line deals with comparing the two LSIZ regions given that CAR is below its median. The corresponding estimates of the trimmed means are 33.02632 and 34.30769. The next line deals with the two LSIZ regions when the CAR values are greater than its median. This second line reports that the p-value is less than 0.001 indicating an association between LSIZ and MAPA when CAR is above its median.

The function also returns a summary of the regions of the covariate values that were used, including the largest and smallest value, the mean and the quartiles. Portions of the results look like this:

```
$Independent.variables.summary[[1]]

V1                  V2
Min.   : 6.0        Min.   : -0.67500
1st Qu.: 14.0       1st Qu.: -0.32941
Median : 16.0       Median : -0.19495
Mean   : 15.5       Mean   : -0.24584
3rd Qu.: 18.0       3rd Qu.: -0.08362
Max.   : 19.0       Max.   : -0.03673
```

The [[1]] at the end of the first line indicates that this is the first region where the values for both LSIZ and CAR are below their respective medians . Results under V1 are for the first covariate, LSIZ. As indicated, the scores range between 6 and 19 and the CAR ranges between –0.675 and –0.03673. Results under `$Independent.variables.summary[[2]]`, not shown here, summarize the LSIZ values below its median and CAR values above its median. If CAR were split based on its tertiles, `$Independent.variables.summary[[2]]`, would summarize the values between of LSIZ less than its median and CAR values between the 0.33 and 0.67 quantiles. Results under `$Independent.variables.summary[[3]]` would be for LSIZ less than its median and CAR values greater than the 0.67quantile.

The plot in Figure 1 was created using default settings for the orientation. Note that based on Figure 1, it is difficult to tell that MAPA scores are higher for low CAR values, versus high CAR values, when the LSIZ scores are low. This is made clearer by rotating the plot resulting in Figure 3. This was done by setting the argument theta in the R function `lplot` to 120. (This function is contained in the R package `WRS` as well as the file `Rallfun-v35` described below.)

The function `sm.inter` can be used to test for an interaction. For binary data, use `smbin.inter`. To perform all pairwise comparisons among the groups, use `smgrid`. The R function `smbinRC` is like the function `smgridRC`, only it is designed for situations where Y is binary. The functions described here are stored in the file `Rallfun-v35` and can be downloaded from https://dornsife.usc.edu/cf/labs/wilcox/wilcox-faculty-display.cfm.

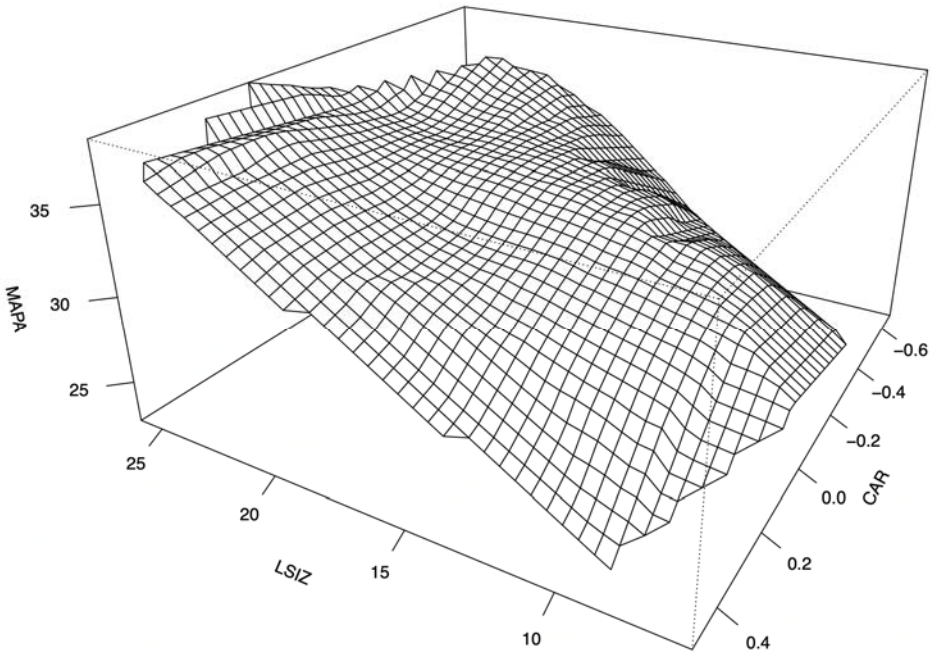

Figure 3. A rotated version of Figure 1 that better reveals how the typical MAPA score changes as the CAR increases.

Conclusion

A fundamental concern is that a linear model can be highly unsatisfactory. This is not always the case, but there is considerable evidence that a linear model can be misleading. One could test the hypothesis that a linear model is correct. But it is unclear when this approach will have enough power to detect a situation where the use of a linear model should be abandoned. Which method best reveals and describes an association depends in part on the nature of the association, which is unknown. It is not being suggested that method QS should be used to the exclusion of all other techniques. Rather, the suggestion is that QS is an option that can play a useful role when the more obvious parametric regression models are inadequate. The most basic version of QS is to split the independent variables at their median. But as previously indicated, this might not suffice; other quantiles might be more effective.

As was illustrated, a smooth can be useful when determining where the independent variables might be split. Here the Cleveland and Devlin method was used in Figures 1 and 3, but another option is the running interval smoother in

Wilcox (2017), which can be used with any robust measure of location. Moreover, it provides a flexible way of capturing any interactions. The running interval smoother can be applied with the R function `rplot`, which is contained in the file `Rallfun-v35` previously mentioned. Here, `rplot` and `lplot` give similar results for the data used in Figures 1 and 3. But there are situations where `lplot` can be misleading due to outliers associated with the dependent variable. Another suggestion is to always check on the impact of removing leverage points, outliers associated with the independent variables. This can be done by setting the argument `xout=TRUE` in the functions `lplot` and `rplot`.

There remains the issue of how QS might be extended to $p > 2$ independent variables. One strategy is to simply use QS with a split on two of the independent variables, which can be done with the R function `smgridRC`. If there is an association among the independent variables, the splits on the two chosen independent variables will impact the values among the remaining independent variables that are included in the four groups. Another possibility is to use regions stemming from regression trees, but typically this results in a complex collection of regions that are difficult to characterize in a simple manner. This is not an argument against regression trees. Random forests, for example, have practical value given the goal of making predictions.

A simpler method is to split the data based on some hyperplane associated with the independent variables. For example, a quantile regression estimator could be used where the first $p - 1$ independent variables are used to predict the typical value for the pth independent variable. The resulting hyperplane could then be used to determine two regions among the points associated with the independent variables, namely points above or below the resulting hyperplane. These two groups could be split again using the same technique and then groups could be compared as previously described. The R function `reg.hyp.split` performs this method. One concern is that it is unknown how to judge the extent a reasonably optimal split has been used. Moreover, a simple characterization of the association might be difficult. In summary, there are crude methods for dealing with $p > 2$ independent variables, but there is considerable room for improvement.

References

Beal, S. L. (1987). Asymptotic confidence intervals for the difference between two binomial parameters for use with small samples. *Biometrics, 43*(4), 941-950. doi: 10.2307/2531547

Bhattacharyya, M. R., Molloy, G.J . & Steptoe A. (2008). Depression is associated with flatter cortisol rhythms in patients with coronary artery disease. *Journal of Psychosomatic Research, 65*(2), 107-113. doi: 10.1016/j.jpsychores.2008.03.012

Chida, Y. & Steptoe, A. (2009). Cortisol awakening response and psychosocial factors: A systematic review and meta-analysis. *Biological Psychology, 80*(3), 265-278. doi: 10.1016/j.biopsycho.2008.10.004

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 Randomised Controlled Trial. *Journal of Epidemiology and Community Health, 66*(9), 782-790. doi: 10.1136/jech.2009.099754.

Cleveland, W. S. & Devlin, S. J. (1988). Locally-weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association, 83*(403), 596-610. doi: 10.1080/01621459.1988.1047863

Efron, B., Hastie, T., Johnstone, I. & Tibshirani, R. (2004). Least Angle Regression (with discussion). *Annals of Statistics, 32*(2), 407-499. doi: 10.1214/009053604000000067

Harrell, F. E. & Davis, C. E. (1982). A new distribution-free quantile estimator. *Biometrika, 69*(3), 635-640. doi: 10.1093/biomet/69.3.635

Hastie, T., Tibshirani, R. & Friedman, J. (2001). *The Elements of Statistical Learning.* New York: Springer. doi: 10.1007/978-0-387-21606-5

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. doi: 10.1002/9781118150702.ch11

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. Biometrika, 75(4), 800-802. doi: 10.1093/biomet/75.4.800

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2017). *An Introduction to Statistical Learning: with Applications in R.* New York: Springer. doi: 10.1007/978-1-4614-7138-7

Kulinskaya, E., Morgenthaler, S. & Staudte, R. (2010). Variance stabilizing the difference of two binomial proportions. *The American Statistician, 64*(4), 350-356. doi: 10.1198/tast.2010.09080

Liu, R. G. & Singh, K. (1997). Notions of limiting P values based on data depth and bootstrap. *Journal of the American Statistical Association, 92*(437), 266-277. doi: 10.2307/2291471

Sen, P. K. (1968). Estimate of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association, 63*(324), 1379-1389. doi: 10.1080/01621459.1968.10480934

Staudte, R. G., & Sheather, S. J. (1990). *Robust estimation & testing*. New York: Wiley. doi: 10.1002/9781118165485

Storer, B. E., & Kim, C. (1990). Exact properties of some exact test statistics for comparing two binomial proportions. *Journal of the American Statistical Association, 85*(409), 146-155. doi: 10.1080/01621459.1990.10475318

Stute, W., Gonzalez Manteiga, W. G. & Presedo Quindimil, M. P. (1998). Bootstrap approximations in model checks for regression. *Journal of the American Statistical Association, 93*(441), 141-149. doi: 10.1080/01621459.1998.10474096

Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis. *Indagationes Mathematicae, 12*, 85–91.

Wilcox, R. R. (2006). Comparing medians. *Computational Statistics & Data Analysis, 51*(3), 1934-1943. doi: 10.1016/j.csda.2005.12.008

Wilcox, R. R. (2017). *Introduction to Robust Estimation and Hypothesis Testing*. (4th ed.). San Diego, CA: Academic Press.

Wilcox, R. R. (2018). Robust regression: Testing global hypotheses about the slopes when there is multicollinearity or heteroscedasticity. *British Journal of Mathematical and Statistical Psychology, 72*(2), 355-369. doi: 10.1111/bmsp.12152

Wilcox, R. R. (2019). Multicollinearity and Ridge Regression: Results on Type I Errors, Power and Heteroscedasticity. *Journal of Applied Statistics, 46*(5), 946-957. doi: 10.1080/02664763.2018.1526891

Yuen, K. K. (1974). The two-sample trimmed *t* for unequal population variances. *Biometrika, 61*(1), 165-170. doi: 10.1093/biomet/61.1.165

Zou, G. Y., Huang, W. & Zhang, X. (2009). A note on confidence interval estimation for a linear function of binomial proportions. *Computational Statistics & Data Analysis, 53*(4), 1080-1085. doi: 10.1016/j.csda.2008.09.033