5-1-2009

Quantile Regression: On Inferences about the Slopes Corresponding to One, Two or Three Quantiles

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

Kathleen Costa
University of Southern California, kcosta@usc.edu

Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

INVITED ARTICLES
Quantile Regression: On Inferences about the Slopes Corresponding to One, Two or Three Quantiles

Rand R. Wilcox   Kathleen Costa
University of Southern California

The problem of testing hypotheses about the slope of a quantile regression line when the sample size is small is considered. A modified bootstrap method is suggested that is found to have certain advantages over the inverse rank method recommended by Koenker (1994). A method is suggested that simultaneously controls the probability of at least one Type I error when performing two or three tests corresponding to two or three specific quantiles. Using data from actual studies, it is illustrated that the new method can yield substantially shorter confidence intervals than the rank inverse method and, even with a large sample size, the choice of method can matter.

Key words: Tests of independence, familywise error, bootstrap methods, Porteus Maze Test, Olympic athletes.

Introduction

Consider the random variables $X_1, \ldots, X_p, Y$ having some unknown (p+1)-variate distribution

Rand R. Wilcox is a Professor of Psychology. He is the author of seven textbooks on statistics, the most recent of which is *Basic Statistics: Understanding Conventional Methods and Modern Insights* (2009, New York, Oxford University Press). Email: rwilcox@usc.edu. Kathleen Costa is in the Department of Kinesiology. Email: kcosta@usc.edu.

and let Y_γ be the conditional γ quantile of Y given $X_1, \ldots, X_p$. When using the Koenker and Bassett (1978) quantile regression method the goal is to estimate Y_γ assuming that

$$Y_\gamma = \alpha_\gamma + \beta_{1\gamma} X_1 + \cdots + \beta_{p\gamma} X_p, \quad (1)$$

where the unknown parameters $\beta_{1\gamma}, \ldots, \beta_{p\gamma}$ and α_γ are estimated based on the random sample $(X_{i1}, \ldots, X_{ip}, Y_i)$, $i = 1, \ldots, n$. The special case $\gamma = .5$ corresponds to what is called the least absolute value regression estimator, meaning

that the parameters are chosen so as to minimize the sum of the absolute values of the residuals. This special case predates ordinary least squares. For a summary of results relevant to $\gamma = .5$, see Birkes and Dodge (1993). A generalization of this method to other quantiles was first considered by Koenker and Bassett (1978). Since then, many new theoretical results have been published plus methods for computing confidence intervals for the parameters (e.g., Koenker, 1994; Koenker & Xiao, 2002). S-PLUS and R provide functions for estimating the parameters, which includes confidence intervals. Although some small-sample size results on the accuracy of these confidence intervals (plus the accuracy of several other methods) were reported by Koenker (1994), the results were limited to p = 1, n=50 and a Type I error probability of $\alpha = 0.1$. Moreover, his results were limited to symmetric distributions. Thus, a goal in this article is to comment on some situations not considered by Koenker (1994). The focus in this article is on testing

$$H_0 : \beta_{1\gamma} = 0.$$

Among the situations considered here, preliminary simulations indicated that the rank inversion method, recommended by Koenker (1994), continues to give fairly accurate confidence intervals for n = 20 (and p = 1) when testing at the .05 level with $\gamma = .5$, .8 and .9. However, for $\alpha = .01$, problems begin to emerge. For $\gamma = .5$, and when both X and Y have standard normal distributions, simulations indicate that now the actual Type I error probability is .002. For $\gamma = .8$ the estimate is .001. This is a concern when making inferences about two or more quantiles because if the goal is to control the probability of at least one Type I error using for example the Bonferroni inequality, having Type I error probabilities well below the nominal level could result in relatively poor power. Accordingly, one goal is to suggest an alternative approach that gives more satisfactory results for this special case.

Practical Reasons for Considering Quantile Regression

Well-known reasons exist for considering quantile regression, but two illustrations are provided that helped motivate this article. The first illustration stems from Costa (2004) where the goal was to study factors that influence increases in horizontal velocity of the body among Olympic athletes who compete in sprints. One issue of specific interest is the rate of horizontal force development (RHFD). Past studies (Henry, 1952; Payne & Blader, 1971; Mero, et al., 1983; Hafez, et al., 1985) indicate horizontal velocity at block departure is dependent on the horizontal impulse generated within the starting blocks. Faster horizontal velocities at the end of the first step out of the blocks are generated with larger net horizontal reaction forces during ground contact (Mero, 1988). These, and related results summarized in Costa (1994), led to the hypothesis that there is an association between horizontal impulse (HI) and RHFD during the first step out of the blocks.

The sample size is n = 39. Initial examination of the data, based on various smooths, hinted at a slightly non-linear association between HI and RHFD. Using, for example, the robust version of the smooth in Cleveland (1979), it appears that as HI increases, RHFD increases somewhat, up to a point, and then decreases. However, a test of the hypothesis that the association is linear, using the method in Stute, et al. (1998) in combination with a least squares fit, failed to reject at the .05 level. Replacing least squares with the more robust Theil (1950) and Sen (1968) regression estimator, again the hypothesis of a linear association is not rejected at the .05 level, but with only 39 pairs of points, this might be due to low power. Testing the hypothesis that the regression line is both straight and horizontal, using the wild bootstrap method in Wilcox (2005, section 9.5), the hypothesis of independence between HI and RHFD was detected at the .1 level. Simple analyses, such as Pearson's correlation and least squares regression, provided no indication of an association (Pearson's correlation is r = -0.04.).

However, consider Figure 1, which shows a scatterplot of the data and three smooths indicated by the three solid lines. The top

Figure 1: A plot of HI versus HFRD Plus the .8 and .2 Quantile Regression Lines
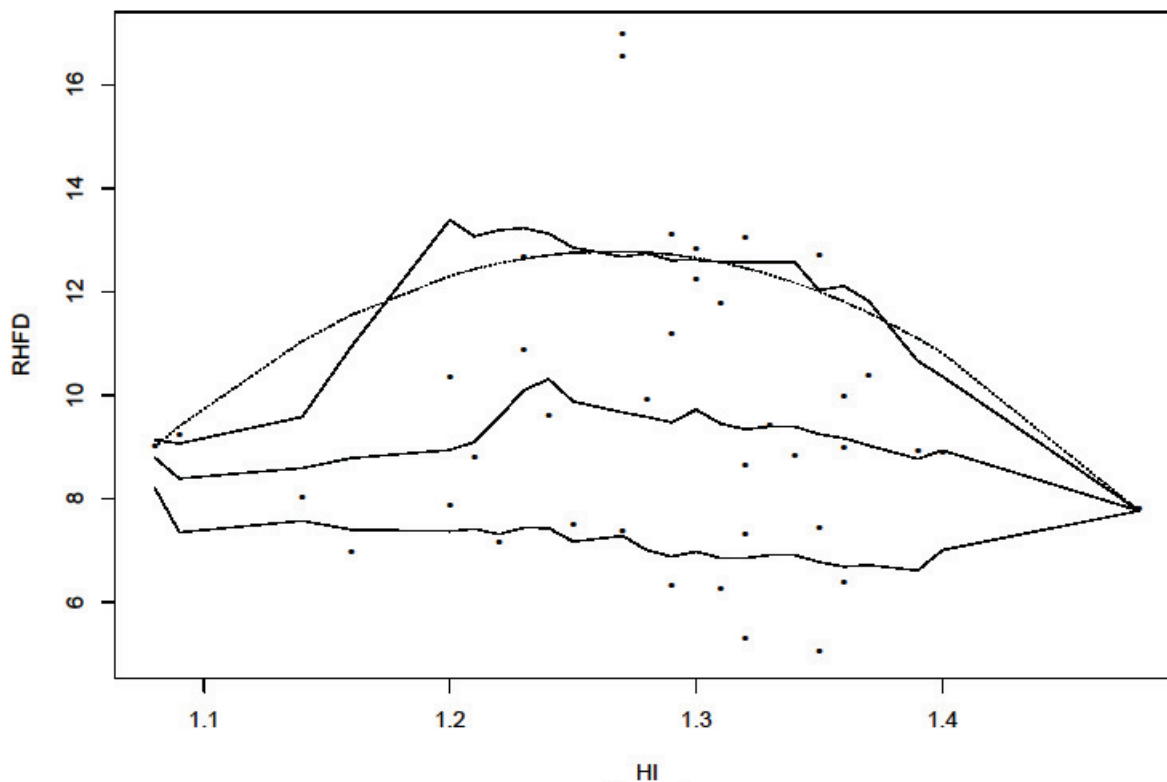
smooth is aimed at estimating the .8 quantile of RHFD given HI. The middle smooth estimates the median of RHFD and the bottom smooth is for the .2 quantile. The so-called running-interval smooth was used, as described for example in Wilcox (2005, section 11.4.4), in conjunction with the Harrell and Davis (1982) quantile estimator. (The S-PLUS function runmq, which comes with the library of functions in Wilcox, 2005, was used.) This suggests that as we move from the lower to the upper quantiles, a non-linear association begins to emerge. As is evident, for the .8 quantile, the association appears to be quadratic. If

$$Y_{.8} = \beta_{0,.8} + \beta_{1,.8}X + \beta_{2,.8}X^2,$$

then the estimates of $\beta_{0,.8}$, $\beta_{1,.8}$ and $\beta_{2,.8}$ are -162.63, 277.13 and -109.47, respectively. The dashed line in Figure 1 shows this fitted model, which appears to be in reasonable agreement with the corresponding smooth.

Thus, an issue is testing both $H_0 : \beta_{1,\gamma} = 0$ and $H_0 : \beta_{2,\gamma} = 0$ in a manner that controls the probability of at least one Type I error in a reasonably accurate fashion.

The second illustration demonstrates that even with n large, quantile regression can help provide a deeper understanding about any association that might exist. Williams, et al. (2005) conducted a study dealing generally with the Porteus Maze Test (PMT), which is used to evaluate intelligence and executive functioning and screen for intellectual deficiency. A portion of the study dealt with the association between the so-called Q score resulting from the PMT test and a measure of maladjustment for the participants in this study. The sample size is n = 1063. Pearson's correlation is r = 0.109, and using the usual Student's T test, the corresponding p-value is less than .001. The .5 quantile regression estimate of the slope is 0 indicating no association.

4

Figure 2 shows a plot of the data. The three straight lines starting from the bottom, are the .5, .8 and .9 quantile regression lines. So it appears that as we move from the median value of Y toward the higher quantiles, an association appears. Using either the inverse rank method or the method considered here, $H_0 : \beta_{1,.9} = 0$ is rejected at the .05 level. The least squares slope is estimated to be .0099, which is close to the estimated .8 quantile regression slope, which is .0098. The estimate of the slope for the .9 quantile is .029. So, although Pearson's correlation rejects at the .001 level, the quantile regression lines provide an interesting perspective on the nature of the association.

Methodology

The Koenker and Bassett (1978) quantile regression method arises as follows. For some γ, $0 < \gamma < 1$, let

$$\rho_\gamma(u) = u(\gamma - I_{u<0}),$$

where the indicator function $I_{u<0} = 1$ if u < 0; otherwise $I_{u<0} = 0$. Assuming that the γ quantile of Y, given X, is given by (1), the Koenker-Bassett quantile regression method estimates the unknown parameters $\beta_{1\gamma},...,\beta_{p\gamma}$ and α_γ with the values $b_{1\gamma},...b_{p\gamma}$ and a_γ, respectively, that minimize

$$\sum \rho_\gamma(r_i), \qquad (2)$$

where $r_i = Y_i - b_{1\gamma}X_{i1} - \cdots - b_{p\gamma}X_{ip} - a_\gamma$ are the residuals. Here, the values that minimize (2) were determined with the function rq that is included in the robust library included with the software S-PLUS.

The proposed method for dealing with very small sample sizes is based in part on a bootstrap estimate of the standard error. The idea of using a bootstrap estimate of the standard error is not new, but the $i = 1,...,n$ more obvious approximation of the null distribution of the test statistic, labeled U below, is already known to be unsatisfactory. (For general results

on bootstrap estimates of the standard error, see Buchinsky, 1994; Hahn, 1994.) More precisely, Koenker (1994) found that the actual probability coverage tended to be higher than the nominal level. Referring to his Table 2, when both X and Y have a Student's T distribution with degrees of freedom 1, 3 or 8, the actual probability coverage, when computing a .9 confidence interval, was estimated to be .920, .948 and .945, respectively. So, in terms of Type I error probabilities, the actual probability of a Type I error can be too low versus the nominal level. Very similar results were obtained here, as indicated in Section 3. One minor goal here is to expand upon Koenker's simulation study by considering sample sizes ranging between 20 and 200, a wider range of α values, and some alternative situations that include skewed distributions. A more major goal is to suggest an adjustment that helps correct the problem just described. And as previously indicated, another goal is to control the probability of at least one Type I error when two or three specific quantiles are of interest.

Let $(X_{i1}^*,...,X_{ip}^*,Y_i^*)$, $i = 1,...,n$, be a bootstrap sample obtained by randomly sampling, with replacement, n vectors of observations $(X_{i1},...,X_{ip},Y_i)$, $i = 1,...,n$. Given, γ label the resulting estimate of the slopes $b_k^*, k = 1,...,p$. Repeat this process B times yielding $b_{1k}^*,...b_{Bk}^*$. Then from basic principles, an estimate of the squared standard error of $b_{k\gamma}$ is

$$S_k^2 = \frac{1}{B-1}\sum_{b=1}^{B}(b_{bk}^* - \bar{b}_k)^2,$$

where $\bar{b}_k = \sum b_{bk}^* / B.$. So an approximate $1 - \alpha$ confidence interval for β_γ is $b_{k\gamma} \pm z_{1-\alpha/2}S_k$, where $z_{1-\alpha/2}$ is the $1 - \alpha / 2$ quantile of a standard normal distribution.

As previously indicated, preliminary simulations indicated that the actual probability coverage is larger than the nominal level. Here, a slight variation of Gosset's original strategy for

Figure 2: Q Scores versus a Maladjustment Score
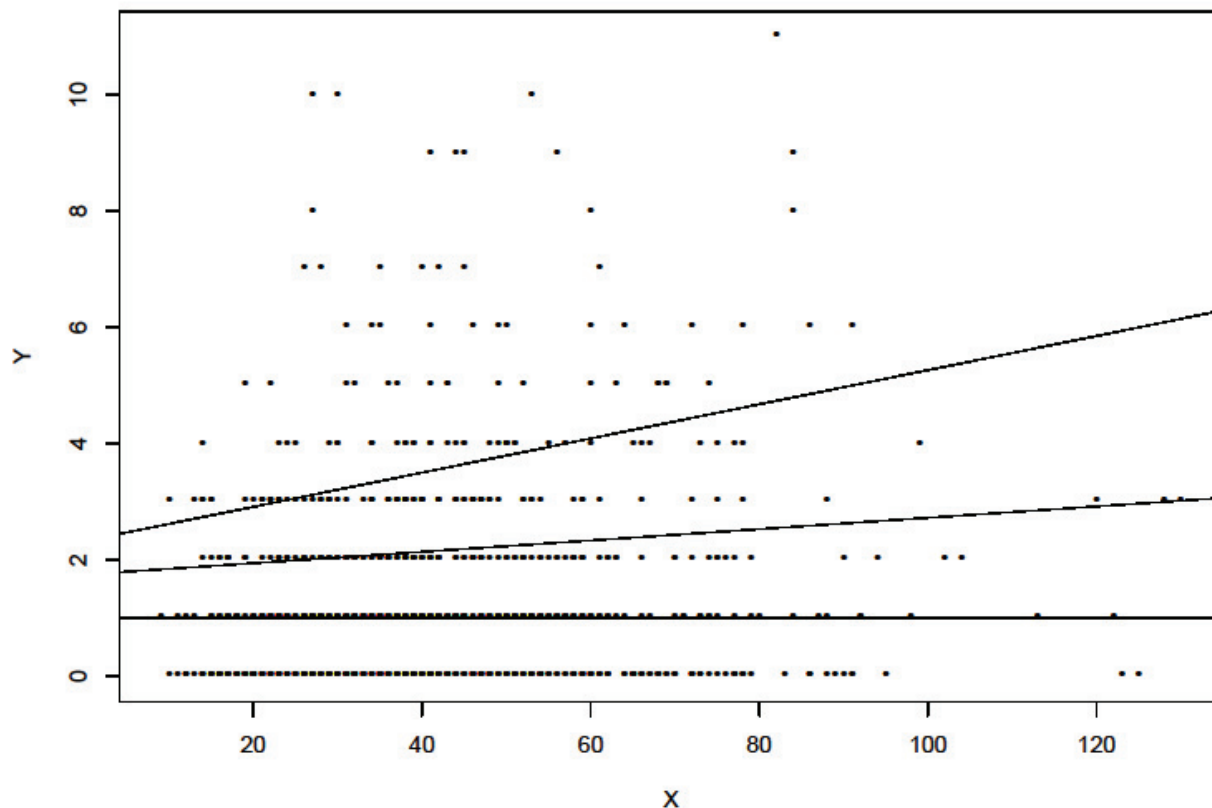
deriving Student's T test is used in an attempt to reduce this problem. That is, assume X and Y are independent, standard normal random variables and use simulations to approximate the distribution of $U_{k\gamma} = \dfrac{|b_{k\gamma}|}{S_k}$. Letting $\hat{u}_{1-\alpha}$ be the resulting estimate of the $1-\alpha$ quantile of distribution of U, the $1-\alpha$ confidence interval for $\beta_{k\gamma}$ is taken to be $b_\gamma \pm \hat{u}_{1-\alpha}S_k$.

Consider p = 1; the $1-\alpha$ quantile of the distribution of $U_{1,.5}$ was estimated for n=10, 20, 30, 40, 60, 100 and 200, and $\alpha = .1$, .05, .025 and .01 using simulations with 1,000 replications. Then a least squares estimate was fitted having the form

$$\hat{u}_{1-\alpha} = d_0 + \frac{d_1}{\sqrt{n}}.$$

The resulting values for d_0 and d_1 are shown in

the top portion of Table 1. Results on how this approximation performs under non-normality are given in the next section of this paper. Still assuming p = 1, next consider the goal of making inferences about the slope corresponding to two different choices for γ: .2 and .8.

Furthermore, the goal is to control the probability of at least one Type I error (cf. Koenker & Machado, 1999). The strategy now is to approximate the null distribution of max($U_{1,.2}, U_{1,.8}$). This was also done for n=10, 20, 30, 40, 60, 100 and 200, and $\alpha =$.1, .05, .025 and .01. The resulting values for d_0 and d_1 are shown in the middle portion of Table 1.

Finally, consideration was given to the goal where three choices for γ are to be used, namely, .2, .5 and .8. The idea is that any one choice for γ might miss an association that would be detected if a different choice were used, and again there is the goal of controlling the probability of at least one Type I error. The

6

resulting values for d_0 and d_1 are shown in the bottom portion of Table 1.

Table 1: Values for d_0 and d_1

α	d_0	d_1
γ =0.5		
.100	1.645	-1.19
.050	1.96	-1.37
.025	2.24	-1.18
.010	2.58	-1.69
γ = (0.2, 0.8)		
.100	1.98	-1.13
.050	2.37	-1.56
.025	2.60	-1.04
.010	3.02	-1.35
γ = (0.2, 0.5, 0.8)		
.100	2.14	-1.31
.050	2.49	-1.49
.025	2.86	-1.52
.010	3.42	-1.85

Simulation Study

Simulations were used to study the small-sample properties of the methods just described, where the critical value is taken to be $\hat{u}_{1-\alpha}$. The distribution for X was taken to be standard normal and the distribution for Y was taken to be one of four g-and-h distributions (Hoaglin, 1985), which contains the standard normal distribution as a special case. If Z has a standard normal distribution, then

$$Y = ((\exp(gZ) - 1) / g)\exp(hZ^2 / 2)$$

if g > 0, and $Y = Z\exp(hZ^2 / 2)$ if g = 0, 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 considered. Additional properties of g-and-h distributions are summarized by Hoaglin (1985).

Table 2: Some Properties of the g-and-h Distribution

g	h	κ_1	κ_2
0.0	0.0	0.00	3.00
0.0	0.2	0.00	21.46
0.2	0.0	0.61	3.68
0.2	0.2	2.81	155.98

Table 3 shows the estimated probability of a Type I error when testing at the .05 or .01 level with n = 20. The estimates are based on 1,000 replications. From Robey and Barcikowski (1992), 1,000 replications is sufficient from a power point of view. More specifically, if we test the hypothesis that the actual Type I error rate is .05, and if we want power to be .9 when testing at the .05 level and the true α value differs from .05 by .025, then 976 replications are required. As is evident, all indications are that reasonable control over the probability of a Type I error is obtained. Similar results were obtained when for a fixed γ, the goal is to test $H_0 : \beta_{1,\gamma} = 0$ and $H_0 : \beta_{2,\gamma} = 0$, or the three hypotheses $H_0 : \beta_{1,\gamma} = 0$, $H_0 : \beta_{2,\gamma} = 0$ and $H_0 : \beta_{3,\gamma} = 0$, provided that when n is small, $.2 \leq \gamma \leq .8$; for brevity, the results are not reported.

Table 3: Probability of at Least One Type I Error, n = 20

g	h	α = 0.05	α = 0.01
0.0	0.0	0.061	0.011
0.0	0.2	0.056	0.008
0.2	0.0	0.064	0.011
0.2	0.2	0.056	0.007

Note that when dealing with the case p=2 or 3, the method used here can be used to control the probability of at least one Type I error when testing simultaneously the two hypotheses $H_0 : \beta_{1,\gamma} = 0$ and $H_0 : \beta_{2,\gamma} = 0$, or the three hypotheses $H_0 : \beta_{1,\gamma} = 0$, $H_0 : \beta_{2,\gamma} = 0$ and $H_0 : \beta_{3,\gamma} = 0$. It is noted that the simulations were repeated when testing these two hypotheses and it was found that the values in Table 3 can be used provided that, when n is small, $.2 \leq \gamma \leq .8$.

Comments and Illustrations Regarding Confidence Intervals

Based purely on simulations, there seems to be little separating the rank inverse method recommended by Koenker (1994) and the bootstrap method used here when α =.05. It is illustrated, however, that when working with real data, the two methods can yield substantially different results.

Consider again the Olympic athlete data and the model $Y_{.8} = \beta_{0,.8} + \beta_{1,.8} X + \beta_{2,.8} X^2$. Using the rank inverse method, the .95 confidence intervals for $\beta_{1,.8}$ and $\beta_{2,.8}$ are $(-1.798(10)^{308}, 320.245)$ and $(-143.70, \ 4903.07)$, respectively. By contrast, using the bootstrap method, the .95 confidence intervals are (25.95, 528.32) and (-208.28, -10.66). Not only do the methods give different results when testing $H_0 : \beta_{1,.8} = 0$ testing, the length of the confidence intervals differ substantially. A similar result is obtained when testing $H_0 : \beta_{2,.8} = 0$, only now the difference between the lengths of the confidence intervals is less dramatic.

Returning to the Porteus maze data, consider the model $Y_{.8} = \beta_{0,.8} + \beta_{1,.8} X$. The estimate of $\beta_{1,.8}$ is zero and using the standard method, the .95 confidence interval is (-0.235, 0.000). Using the bootstrap method studied here, the .95 confidence interval is (-0.126, 0.126). So the standard method is unusual in the sense that the upper end of the confidence interval is equal to the estimated slope. For the .9 quantile, the estimate of the slope is -0.133 and the standard method gives a .95 confidence interval of (-0.247, -0.007). Now the point estimate is near the center of the confidence interval. The bootstrap confidence interval is (-0.253, -0.014).

Conclusion

It is noted that some additional methods and situations were considered beyond those already described. Simulations were run with γ =.1, but now the null distribution of U was found to be rather unstable as a function of the distributions used to generate the data when the sample size is small. A percentile bootstrap method was considered, but it was found to be considerably less satisfactory in terms of probability coverage. The main point is that the adjusted bootstrap method considered here appears to perform reasonably well even under what would seem like extreme departures from normality. Moreover, both methods considered here seem to perform well when sampling from skewed distributions. Generally, when X and Y are independent, the choice between the two methods considered seems to make little difference, but when there is an association, this might no longer be the case, as was illustrated. Finally, R and S-Plus software is available from the author for applying the bootstrap method studied here. Ask for the function qregci.

References

Birkes, D., & Dodge, Y. (1993). *Alternative methods of regression*. NY: Wiley.

Buchinsky, M. (1991). *The theory and practice of quantile regression*. Ph.D. Thesis, Dept. of Economics, Harvard University.

Costa, K. E. (2004). *Control and dynamics during horizontal impulse generation*. Ph.D. Thesis, Dept. of Kinesiology, University of Southern California.

Hahn, J. (1995). Bootstrapping quantile regression estimators. *Econometric Theory*, *11*, 105-121.

Hoaglin, D. C. (1985). Summarizing shape numerically: The g-and-h distributions. In D. Hoaglin, F. Mosteller, & J. Tukey (*Eds.*), *Exploring data tables, trends, and shapes*, 461-515. NY: Wiley.

Koenker, R. (1994). Confidence intervals for regression quantiles. In P. Mandl & M. Huskova (*Eds.*), *Asymptotic statistics*, 349-359. Proceedings of the Fifth Prague Symposium.

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrika*, *46*, 33-50.

Koenker, R., & Machado, J. A. F. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, *94*, 1296-1310.

Koenker, R., & Xiao, Z. J. (2002). Inference on the quantile regression process. *Econometrica*, *70*, 1583-1612.

Robey, R. R., & Barcikowski, R. S. (1992). Type I error and the number of iterations in Monte Carlo studies of robustness. *British Journal of Mathematical and Statistical Psychology*, *45*, 283-288.

Williams, N., et al. (2005). *Porteus' mazes and executive function in children: Standardized administration and scoring, and relationships to childhood aggression and delinquency*. Unpublished manuscript, Dept. of Psychology, University of Southern California.

Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing* (*2nd Ed.*). San Diego, CA: Academic Press.