11-2022

Canonical Quantile Regression

Stephen Portnoy
Portland State University

Citation Details
Published as: Portnoy, S. (2022). Canonical quantile regression. Journal of Multivariate Analysis, 105071.

Canonical Quantile Regression

Stephen Portnoy[a]

[a] *University of Illinois at Urbana-Champaign and Portland State University*

Abstract

In using multiple regression methods for prediction, one often considers the linear combination of explanatory variables as an index. Seeking a single such index when here are multiple responses is rather more complicated. One classical approach is to use the coefficients from the leading Canonical Correlation. However, methods based on variances are unable to disaggregate responses by quantile effects, lack robustness, and rely on normal assumptions for inference. An alternative canonical regression quantile (CanRQ) approach seeks to find the linear combination of explanatory variables that best predicts the τth quantile of the best linear combination of response variables. Applying this "regression" approach more generally, subsequent linear combinations are chosen to explain what earlier CanRQ components failed to explain. While numerous technical issues need to be addressed, the major methodological issue concerns directionality: a quantile analysis requires that the notion of a larger or smaller response be well-defined. To address this issue, the sign of at least one response coefficient will be assumed to be non-negative. CanRQ results can be quite different from those of classical canonical correlation, and can offer the kind of improvements offered by regression quantiles in linear models.

Keywords: Canonical Correlation, index prediction, multivariate regression, regression quantile,
2020 MSC: Primary 62H12, Secondary 62J05 62H20

1. Introduction.

Given explanatory X-variables and Y response variables, we seek to find an index based on the X's to predict the best linear combination of Y's in a quantile regression setting. The approach was motivated by a specific financial data set listing various measures of the performance of some large U.S. companies. The aim was to use prior company data to predict future performance; specifically, to define an index in X with better predictive power than CEO compensation. A detailed statistical analysis of the financial data using the Canonical Regression Quantile (CanRQ) methodology is presented in Portnoy and Haimberg [14]. Some additional data analysis is added in Section 5.3, together with a few comments on the earlier analyses. The goal here is to develop the CanRQ method in general, to provide the underlying theoretical properties, and to present some simulations and examples suggesting that the methodology is useful and reliable.

A classical approach might be to use the leading Canonical Correlation, which provides a linear combination of X-variables (the X-index) that is most highly correlated with some linear combination of the Y-variables (the Y-index). However, the inability of correlation methods to disaggregate responses by quantile effects, their inherent lack of robustness, and their need for normal assumptions strongly motivates the search for a novel quantile regression approach. The results here develop such an approach, both for the leading coefficients, and for subsequent coefficients in analogy with classical Canonical Correlations.

The basic idea is to replace the objective of maximizing the correlation of the X- and Y-indices by that of minimizing the regression quantile loss applied to the difference of the indices; see (2). However, this objective function is homogeneous in all the coefficients, and so it is minimized globally by setting all X and Y coefficients to zero. Classical Canonical Correlations solves the corresponding problem by imposing quadratic restraints on both sets of

*sportnoy@illinois.edu

coefficients. Here we replace this by imposing an L_1 constraint on the Y coefficients only: we require the sum of the absolute value of the coefficients of the Y-index to be 1.

Regression quantiles, of course, impose an order on the data and so require that directionality be defined. Here we impose inequality constraints of positivity on at least one response coefficient. This implies that increasing values of the X-index leads to increasing for those Y variables whose coefficients are constrained to be positive. This defines the leading "Canonical Regression Quantiles" (CanRQ) component as a constrained regression quantile solution, which is thus computable using available software (for example, the R quantreg package, Koenker [10])

Subsequent CanRQ components defined to minimize the same regression quantile loss (2) subject to orthogonality to prior components (as well as the L_1 constraint). However, here the sign of the coefficients is not identifiable. Thus, as described below, an (arbitrary) positivity constraint is introduced to provide well-defined coefficients. Asymptotic theory is provided to justify Bootstrap approaches to inference. The fact that the standard bootstrap fails for constrained estimators leads to using an n-choose-m approach (see Andrews [1]). Some simulations and examples are quite promising.

2. Interpretation.

This research does not seek to develop a robust version of Canonical Correlations, nor does it try to define a multivariate notion of quantiles to be used in multivariate regression. The purpose is to develop linear indices based on multiple dependent and independent variables to which standard linear quantile regression can be applied. This allows the use of the well-known computational methods and well-considered means of interpretation of regression quantiles in order to provide a superior regression quantile approach to the linear relationship and dimension reduction goals of classical Canonical Correlation analysis.

Specifically, a fundamental aim of CanRQ is to allow parameters to be interpreted as regression quantile coefficients just as classical canonical correlation allows coefficients to be interpreted in terms of least squares regression. The only difference is that whereas the classical coefficients describe the effect of the explanatory on the conditional mean of the best linear combination of response variables, the CanRQ coefficients describe the effect on the conditional (τth) quantile. For the classical Y-coefficients, the norm constraints restrict their interpretation in that that can only indicate relative influence. Nonetheless, this does allow a qualitative interpretation, and the CanRQ Y-coefficients have an analogous interpretation as indicating the relative response of each variable on the τ-th quantile. That is, the only difference in interpretation between classical canonical correlation coefficients and CanRQ coefficients is that the effect on the conditional τth quantile replaces the effect on the conditional mean.

Furthermore, once the coefficients of the responses are determined, they define a Y-index; and the X coefficients are exactly regression quantile estimates of the Y-index (and so have a direct interpretation as such). While the Y coefficients suggest the influence of the explanatory X-variables on individual Y-responses, the norm and sign constraints require careful consideration. The L_1 norm constraint restricts the interpretation to indicate only a relative effect (just as is the case for classical coefficients). The sign constraints on the Y-coefficients introduced by the CanRQ method require additional consideration. In many cases with multivariate responses that are the main focus of this work, all (or most) of the Y's measure some common component that might be expected to increase with the X-variables. Thus, it might be quite reasonable to constrain all the Y-coefficients to be positive. If this is true, the X-regression quantiles (as a function of the τ probability, defining the quantile) would indicate increasing responses (as τ increases). For example, the $\tau = .75$ CanRQ Y-index would correspond exactly to the 75th conditional quantile of the response Y-index; and would also correspond to generally larger values of each of the Y's. However, if some of the Y-coefficients are negative, then the interpretation is somewhat different, and more akin to Principal Components. Specifically, if only one (or a few) Y-coefficients are constrained to be positive and some CanRQ estimates are negative, the signs of the coefficients separate the Y's into two sets whose response to the X's is in opposite directions. This would suggest that the two sets of variables are related in some way. While this generates a rather more complex interpretation, the dependence on X's would still suggest increasing responses for Y's with positive coefficients and decreasing responses for those with negative coefficients. Again, note that this is exactly the same form of interpretation as for the classical canonical correlations, with meaning of conditional mean replaced by that for conditional quantiles.

Because of the orthogonality constraints for subsequent CanRQ components, the more complex kind of interpretation is even more necessary in going beyond the leading component. The orthogonality constraint essentially ensures that some Y coefficients will be negative; so there is no simple way to impose constraints intended to provide monotonic increasing relationships. In fact, CanRQ coefficients are not unique: reversing the signs of all (X and Y) coefficients does not change the objective function; and so the overall sign is not determined. Thus, to provide specificity, one of the Y coefficients must be constrained to be positive. By default, for subsequent components the first Y-coefficient is constrained to be positive in the current software, but the choice is arbitrary. Nonetheless, once the constraint is chosen, the CanRQ coefficients have a very similar interpretation as for the first component when some coefficients are negative: the signs of the estimates divide the variables into two groups whose relationship is monotonic (increasing and decreasing) in τ. Again, the X coefficients for each component are exactly regression quantiles estimates for the corresponding Y-Index. Therefore, CanRQ can provide useful regression quantile interpretations.

To continue the interpretation of subsequent components, note that classical canonical correlations can be used as a dimension-reduction technique. For this purpose, their application is more concerned with preliminary data analysis and less concerned with confirmatory inference. In this regard, CanRQ can also be used to reduce dimension. However, since different quantile levels may lead to rather different estimates, the reductions may vary with the quantile probability level, τ, and thus may be rather less indicative of the underlying dimension of the data. That is, for each τ, CanRQ may suggest subsets of the data that should be adequate to analyze that quantile, but these subsets may vary with τ. Nonetheless, this is not necessarily a disadvantage, as the indication of heterogeneity can be very useful in avoiding inappropriate interpretations. Furthermore, the dependence of the sign-classification of variables on the quantile level allows a more nuanced interpretation of variable relationships as being dependent on the size of the response.

3. Definition and Methodology

The method of Canonical Quantile Regression seeks first to find a linear combination of the response variables that is best predicted by a linear combination of the explanatory variables in terms of regression quantile objective functions. The basic idea is as follows: let $\rho_\tau(u) = u(\tau - I(u < 0))$ be the quantile objective function so that the τth quantile, Q_τ of a sample $\{x_i\}$ satisfies

$$Q_\tau = \min_\theta \sum_i \rho_\tau(x_i - \theta) \,. \tag{1}$$

Given an $n \times p$ data matrix, X, of explanatory variables (generally including a constant intercept column) and an $n \times q$ data matrix Y of response variables, we would like to define the first Canonical Quantile Regression as the pair of coefficient vectors, (α, β) achieving

$$\min_{\alpha,\beta} \sum_{i=1}^{n} \rho_\tau(x_i'\beta - y_i'\alpha) \,, \tag{2}$$

where x_i and y_i denote the ith rows of X and Y respectively. As noted above, minimizing (2) without constraints would lead to all coefficients being zero. To reflect the quantile setting and to avoid the lack of robustness of the quadratic constraints of classical Canonical Correlations, impose the following L_1 identity:

$$\sum_{j=1}^{q} |\alpha_j| = 1 \;, \tag{3}$$

where α_j are the coordinates of α. However, changing the signs of all coefficients does not change the problem for $\tau = .5$, and reverses τ and $1 - \tau$ otherwise. That is, something further is needed to define directionality for both computation and interpretation.

In many (most) applications, the response variables measure related outcomes, and so we expect at least some of them to be monotonic in the predictors. Since the objective function in (2) is invariant under sign changes (that is, it does not define the sign of the coefficients), either choice of sign can be made. This suggests imposing the side conditions:

$$\sum_j \alpha_j \;=\; 1 \quad ; \quad \alpha_j \geq 0 \quad j \in S \;, \tag{4}$$

where S is a set of Y-indices to be chosen using substantive knowledge of the data. If the Y's are all thought to be positively related to X's, then taking $S = \{1, \ldots, q\}$ might be indicated, but a strict subset $S' \subset S$ might be preferred to account for non-significant variables (see below).

Minimizing (2) subject to this constraint is a constrained regression quantile problem, whose solution is well defined, and it is easily computable since the linear constraints perfectly match the linear programming form of the regression quantile problem (see Koenker [9], p. 213). Thus, letting $(\hat{\alpha}, \hat{\beta})$ solve (2) subject to (3) and (4), consider $\hat{\alpha}$ and $\hat{\beta}$ to define predictive and response indices (respectively), and call $x_i'\beta$ and $y_i'\alpha$ the observed predictive and response indices corresponding to the i-th observation (that is, the i-th row of X and Y).

If $\alpha_j > 0$ (for all j), it is possible to reduce to a standard regression quantile problem. From the side condition (3), we can write $\alpha_1 = 1 - \sum_{j=2}^{q} \alpha_j$ and substitute into the objective function (2) to get

$$\min_{\alpha,\beta} \sum \rho_\tau \left(y_{i,1} - \sum_{j=2}^{q} y_{i,j}\alpha_j + x_i'\beta \right) . \tag{5}$$

The solution to this minimization problem provides $\hat{\beta}$ and $\{\hat{\alpha}_j\}$, $j \in \{2, \ldots, q\}$ (with $\hat{\alpha}_1 = 1 - \sum_{j=2}^{q} \hat{\alpha}_j$) as standard regression quantile coefficients. Furthermore, if all $\alpha_j > 0$, the estimates will all be positive with probability tending to 1 asymptotically. The coefficient estimates will satisfy all known results for regression quantiles (finite sample, asymptotic, and computational; see Koenker [9]).

What if only some $\alpha_j > 0$? For those α_j not assumed to be positive, minimize (1) over all choices for the signs of α_j. For example, suppose $\alpha_1 > 0$ is specified. Define $s_1 = 1$ and let

$$s_j^{(k)} = \pm 1 , \quad j \in \{2, \ldots, q\} \tag{6}$$

denote all choices of signs ($k = 2, \ldots, 2^{q-1}$).

Then, for each k, minimizing (1) subject to the constraints $\alpha_1 > 0$ and $\sum_{j=2}^{q} s_j^{(k)}\alpha_j = 1$. (with $s_1 \equiv 1$) is a linearly constrained regression quantile problem. To complete the computation, simply minimize further over all 2^{q-1} choices of signs. This is essentially just projecting the unconstrained solution onto a union of constraint cones (instead of just one).

Note: given one specification for positive signs (say, $\{\alpha_j \geq 0 \; : \; j \in S\}$), any other specification S' that agrees with the α signs produced under S will give exactly the same coefficients. In this sense, there is a certain uniqueness in the CanRQ methodology: only introducing a positivity constraint on a negative coefficient can provide different estimates.

4. Extension to Subsequent Canonical Regression Quantiles

Subsequent Canonical Correlation components seek to explain what has not been explained by earlier components. In the classical case, the quadratic metric provided by normal assumptions makes orthogonality a reasonable criterion for choosing subsequent components. This criterion is less compelling when considering regression quantiles. Even for classical Canonical Correlation, orthogonality makes more sense for dimension reduction than for choosing good predictors.

Nonetheless, the asymptotic theory for regression quantiles relies on the fact that absolute error loss leads to an approximate quadratic objective function when applied to the data. Furthermore, the author was unable to find any alternative method that seemed more reasonable in general. Thus, subsequent canonical regression quantile components will also be defined by requiring orthogonality to previous components. Again, scale specificity will be obtained by setting the L_1 norm of $\hat{\alpha}$ to one.

Specifically, given leading CanRQ. component, $(\hat{\alpha}^{(1)}, \hat{\beta}^{(1)})$, let $(\hat{\alpha}^{(2)}, \hat{\beta}^{(2)})$ solve (1) as for leading case with $\sum_{j=1}^{q} |\hat{\alpha}_j^{(2)}| = 1$ and, say, $\alpha_1^{(2)} \geq 0$ (to specify signs) plus the orthogonality constraint $\sum_{j=1}^{q} \hat{\alpha}_j^{(2)}\hat{\alpha}_j^{(1)} = 0$. Since subsequent components aim to find differences from earlier components (rather than to estimate responses), no monotonicity (or directionality) constraints are needed. Here, the constraint, $\alpha_1^{(2)} \geq 0$, only serves to fix the sign of the $\hat{\alpha}$ and $\hat{\beta}$ coefficients, which is critical for comparisons and for the use of resampling methods for inference.

Continue inductively to compute subsequent CanRQ components: at stage j take basis elements for R^q orthogonal to $\{\hat{\alpha}^{(1)}, \ldots, \hat{\alpha}^{(j-1)}\}$. Project Y onto this subspace, and find $\hat{\alpha}^{(j)}$ as for the leading canonical regression quantile (including a positivity constraint on at least one $\alpha^{(j)}$). Of course, the last $\hat{\alpha}^{(q)}$ is defined (simply) by orthogonality to the first $(q-1)$ $\hat{\alpha}$'s.

5. Asymptotics and Inference

The presence of constraints can seriously affect theoretical properties if some $\hat{\alpha}$ coefficients are negative or are on the boundary. This can occur with positive probability for finite samples even when all $\alpha_j^{(k)}$ strictly satisfy the inequality constraints. In such cases, the constrained solution will be a projection into the constraint set. While this complicates the asymptotic theory, the basic approach is somewhat straightforward and can be found in Chernoff [4].

To describe the general situation, suppose $\hat{\theta}_n$ is an unconstrained estimator, and satisfies $\sqrt{n}(\hat{\theta}_n - \theta_0) \to \mathcal{N}_m(0, \Sigma)$. Suppose the contraint set, C, can be approximated as an intersection of a finite number (say $k < m$) of half spaces. This is called Chernoff regularity of the constraint set, and is true under appropriate smoothness conditions on the boundary, B, of C. For the estimators here, the constraint set is an intersection of half spaces directly. If θ_0 is interior to C, then $\hat{\theta}_n$ will also lie in C with probability tending to 1; and the constrained estimator will equal $\hat{\theta}_n$ and have the same asymptotically normal distribution. If θ_0 lies outside the interior of C, let θ^* denote the nearest point to θ_0 on the boundary, B, in the metric generated by Σ (that is, the projection onto B under Σ). Then θ^* will lie in a linear manifold in B, which is an intersection of a finite number (less than k) of half-spaces. Thus, the constrained estimator will be a mixture of the projection of the unconstrained estimator onto each of the half-space linear boundaries. Since projection is continuous and linear, $\sqrt{n}(\hat{\theta}_n - \theta_0)$ will converge in distribution to a mixture of multivariate normals.

A general result is given in Geyer [6] while Parker [11] gives a detailed result for constrained regression quantiles. This provides asymptotics for the α and β parameters of the first CanRQ component. Subsequent canonical regression coefficients need to be projected into sets defined by earlier ones, and thus are random projections. However, the random projections are still continuous, and thus convergence in distribution of the unconstrained estimators still implies convergence of the projections (see Chernoff, [5], Theorem 2). Thus, the following result holds, with most details provided in the proof:

Theorem 1. *Consider the vectors $A = Vec(\alpha^{(j)} : j = 1, \ldots, q)$ and $B = Vec(\beta^{(j)} : j = 1, \ldots, p)$ for all CanRQ's defined above (for fixed τ). Suppose that the unconstrained estimators given by* (5) *for any fixed choice of signs of the α's (see* (6)*) satisfy standard regression quantile asymptotics. For example, see Koenker [9] for conditions. Then the vectors $\hat{A}$ and $\hat{B}$ of all canonical regression quantile estimators have an asymptotic distribution in the sense that $\sqrt{n}\{(\hat{A}, \hat{B}) - (A, B)\}$ converges in distribution.*

Proof: Consider the first canonical regression quantile and note that (5) provides standard regression quantile estimators; and so under standard conditions (for example, see Koenker, [9]), $\sqrt{n}\{(\hat{\alpha}, \hat{\beta}) - (\alpha, \beta)\}$ converges in distribution.

Now, continue inductively. At the j-th stage let $\alpha^{(j)}$ denote the coefficients for the j-th component and define $A_j = (\alpha^{(1)}, \ldots, \alpha^{(j-1)})$. Let $\tilde{\alpha}_j$ be the previous (unconstrained) estimates projected orthogonally to the prior A_j parameters, and let B_j denote the corresponding (previous) β's. Note that $\tilde{\alpha}_j$ is obtained from (5) as above and then projected onto a fixed (non-random) space. Thus, from Geyer [6], Theorem 4.4, $\sqrt{n}(\tilde{\alpha}_j - A_j)$ converges to a (joint) distribution in $R^{q(j-1)}$; as does the corresponding B_j.

By the induction hypothesis, $\sqrt{n}(\hat{A}_j - A_j, \hat{B}_j - B_j)$ also converges in distribution. To satisfy the condition that $\hat{\alpha}_j$ be orthogonal to A_j, we must project $\tilde{\alpha}_j$ onto a (random) linear subspace orthogonal to $\hat{A}_j$. This is a continuous linear (random) operation. But since $\sqrt{n}(\hat{A}_j - A_j)$ converges in distribution, the projections converge in probability to the projections orthogonal to the fixed A_j. Thus, from Chernoff [5], Theorem 2, $\sqrt{n}(\hat{\alpha}^{(j)} - \alpha^{(j)})$ converges in distribution; as does $\sqrt{n}(\hat{\beta}^{(j)} - \beta^{(j)})$. □

It now follows that Bootstrap methods can be used to obtain appropriate asymptotic inference. If the parameter inequality constraints hold strictly, standard bootstrap methods will work (asymptotically). However, it seems reasonable to allow some inequality constraints to be equalities. Specifically, if the α-constraints fail, the estimates will lie on the boundary with non-vanishing probability (asymptotically). In such cases, Andrews [1]) presented a specific counterexample showing that the standard bootstrap must fail if the parameter is on the boundary, but referred to Politis, Romano and Wolf [12], Theorem 2.2.1, to show that the n-choose-m subsample bootstrap will work as long as

m tends to infinity more slowly than n. As a consequence, Andrews noted that the subsample bootstrap will provide appropriate inferences for constrained maximum likelihood estimators. Later work extended these results in various directions. Thus, a version of the Andrews bootstrap will be applied here, though it appears that the method has not yet been investigated fully for constrained regression quantiles.

The focus here will use a weighted version of the Andrews bootstrap that is informed by recent research in this area. One complication of subsample bootstrap methods is that the rank of the X-matrix may not be a small fraction of the sample size. Thus, repeated observations in bootstrap samples can lead to singular (or nearly singular) design matrices. To avoid this problem, weighted resampling methods will be used here. Specifically, a subsample size m will be chosen proportional to a power of n a bit less than 1. In minimizing the quantile loss function to obtain the bootstrap estimates, a subsample of size m will be given independent negative exponential weights, and the remaining observations will be given $(1/n)$ times independent uniform weights. This is straightforward computationally since the R-programs in the quantreg package (Koenker [10]) specifically allow such reweighting in the quantile objective function. The weighted subsample process will be repeated R times to provide a "bootstrap" sample. Weighted bootstrap methods are justified in some generality in Chatterjee and Bose [3]. A formal justification of the bootstrap here is given in the following Corollary:

Corollary 1. *Let θ denote the vector of the α and β parameters in the canonical regression quantile model, let $\hat{\theta}$ denote their estimates and let θ^* denote the bootstrap estimates using the weighted n-choose-m bootstrap described above. Specifically, the weighted bootstrap chooses a subsample of size m and uses weights that are either negative exponential or uniform(0, $1/n$), with the weights rescaled so that their mean is 1. Then, under the hypotheses of Theorem 1, the asymptotic distribution of $\{\sqrt{n}(\theta^* - \hat{\theta})\}$ is the same as he asymptotic distribution of $\{\sqrt{n}(\hat{\theta} - \theta)\}$.*

Proof: As noted above, the n-choose-m bootstrap is fully justified by Theorem 2.2.1 of Politis, Romano and Wolf [12] as long as the estimators converge in distribution (which is given by Theorem 1). It remains to show that the weighted version is equivalent. To do this, check the "Basic conditions" (equations (2.1), (2.2), and (2.3)) of Chatterjee and Bose [3]. Since p is assumed to be fixed here, these conditions become: the weights, $\{w_i\}$ are exchangeable, and

$$\begin{aligned} Ew_i &= 1, &(7)\\ 0 < \mathrm{Var}(w_i) &= o(n), &(8)\\ EW_aW_b &= O(n^{-1}), &(9)\end{aligned}$$

where $W_i = (w_i - 1)/\sqrt{Var(w_i)}$. Chatterjee and Bose [3] show that the conditions hold for the multinomial distribution that is equivalent to the standard n-choose-m bootstrap. Now, replace a positive multinomial value by a negative exponential and "0" by a uniform (Unif$(0, n^{-1})$) and rescale to make the marginal expectation equal to 1. The is equivalent to the weighted n-choose-m bootstrap used here, and the conditions can be checked immediately. Specifically, (7) holds by rescaling, (8) holds since the variances are a constant plus $O(n^{-1})$, and (9) holds since the multinomial covariance is $O(n^{-1})$ and the negative exponential and uniform distributions have finite second moments. □

The problem of estimates potentially falling on the boundary might be addressed by an alternative approach. Cavaliere and Georgiev [2] suggest that the main problem generating bootstrap inconsistency in such cases occurs when the probability that a parameter falls on the bounder differs substantially between the bootstrap distribution and the distribution of the estimator under the model. To keep the difference between the bootstrap and empirical probabilities small, a weighted version of a delete-$\sqrt{n}$ jackknife might be used successfully. Portnoy [13] provides evidence that such a procedure is appropriate for censored regression quantiles, where it is also important to keep the resampling distribution close to the empirical distribution. While this method worked in some examples, it did not seem as generally reliable in the simulations and examples, and the weighted version described above was used here.

Based on this weighted n-choose-m bootstrap, several procedures were considered for constructing confidence intervals. The first is the modified percentile method. Let q_{up} and q_{low} be the .975 and .025 quantiles of the bootstrap distribution of an estimator, $\hat{\theta}$. Let $c \equiv \sqrt{m/n}$. Then the upper and lower confidence bounds are: $\theta(1 + c) - c\, q_{low}$ and $\theta(1 + c) - c\, q_{up}$. These come directly from using the bootstrap distribution of $\sqrt{n}(\theta^* - \hat{\theta})$ to approximate the asymptotic distribution of $\sqrt{n}(\hat{\theta} - \theta)$ and from adjusting for the n-choose-m subsample bootstrap.

This method is presented in the simulations results below. Several alternative approaches were tried, with the best being an interval using a standard error estimate based on the interquartile range of the bootstrap sample. Specifically, the interval is $\hat{\theta} \pm IQR\, s/1.349$, where the last factor adjusts for the n-choose-m bootstrap and aims to provide 95% coverage. This method seemed to provide conservative intervals in all cases, and so was not presented here. Future research may suggest ways of combining this approach with the percentile method above to improve coverage.

6. Simulations and examples

Some limited simulations suggest that the above methodology is sound. The CanRQ method was also applied to several real-data sets, two of which are reported here. All results were produced by R-programs (see R Core Team [15]). The tables below were produced from R arrays using the R stargazer package (Hlavac [8]).

6.1. Some simulations

While it is easy to generate multivariate normal samples where classical Canonical Correlations and CanRQ can be applied, it is more difficult to generate such examples where the α coefficients are known but not constant in τ. If some α's vary with τ, then the constraints require others to vary, which eliminates standard approaches to specifying τ-varying coefficients. The β coefficients, however, can be defined by any set of co-monotonic functions of τ (that is, functions satisfying the known minimal conditions needed to define quantile functions; see Koenker [9]). Thus, as noted in Section 2, the β coefficients have exactly the same interpretation in terms of the α responses as standard regression quantiles. The leading α coefficients are subject only to a norm constraint (and positivity constraints), and it is possible, though rather difficult, to define coefficients that vary with τ. As noted above, their interpretation is somewhat compromised by the norm constraint; that is, they can only provide relative indications of influence of the Y-variables. Subsequent α components are also subject to orthogonality and, thus, are more restricted. As noted in Section 2, the subsequent α coefficients provide a more exploratory analysis, more akin to the interpretation of Principal Components. As a consequence, the simulations here take all α coefficients and all β slope coefficients to be constant in τ.

The simulations used additive error regression models (that is, location-scale models) to define auxiliary response vectors Y_i with β coefficients given in Table 1. Note that the intercept in the regression quantile model for the j-th component will be $\beta_{0,j}$ plus the quantile function for the σ_j-scaled error distribution evaluated at τ (where σ_j is defined below).

The specific model is as follows: given an $n \times p$ design matrix, X, define the rows of the auxiliary response matrix, Y_1 by

$$Y_{1i} = (\, 1 \;\, X_i \,)\, B \;+\; (\, S_1 \; S_2 \; S_3 \; S_4 \,)\,, \tag{10}$$

where S_i are independent with S_i distributed as a scale-family with $\sigma_i = .7^{(5-i)}$. Thus, the first regression equation has smallest variance, and the variances increase for subsequent components. Here, B is the 5×4 matrix of β coefficients given as the transpose of the right side of Table 1. The Y matrix of responses is then modeled by taking rows

$$Y_i = Y_{1i}\, A^{-1}\,. \tag{11}$$

The values in the A matrix are given by the transpose of the first 4 columns in Table 1. As noted above, if $Q_j(\tau)$ denotes the quantile function for S_j (in (10)), writing matrix $B(\tau)$ as the matrix B with β_{0j} replaced by $\beta_{0j} + Q_j(\tau)$ gives the CanRQ model (as used in (2)) to be $(1,\, X)B(\tau) = YA$.

Table 1: Specification of the constants defining the regression parameters for the linear regression plus error model and the Y-model used in the simulations. The matrix B in (10) is the transpose of the last 5 columns below and the A in (11) is the transpose of the first 4 columns below.

Component	α_1	α_2	α_3	α_4	β_0	β_1	β_2	β_3	β_4
1	.25	.25	.25	.25	2.5	2	1.5	1	.5
2	.50	-.50	0	0	3.5	3	2.5	2	1.5
3	.25	.25	-.50	0	4.5	4	3.5	3	2.5
4	.1667	.1667	.1667	-.50	5.5	5	4.5	4	3.5

Fig. 1 presents results summarizing coverage probabilities for 8 trials. In all cases $N = 1000$ simulation replications were taken, and the nominal coverage was set at .95. Sample sizes were $n = 200$ and $n = 400$, and all trials used the adjusted bootstrap percentile method described above with $R = 400$ bootstrap replications. Three sets of distributions were used: X and S normal (marked by a single purple dash in the figure), X and S normalized Chi-square (marked by a double navy dash), and X normal with S t-distributed with 3 df (marked by a triple red dash). The quantile level τ was set at .5 and .25 for normal X and S, and kept at .5 for the other two distributional settings. Other values for the various simulation parameters (p, q, n, σ_i, R) were tried, but the results were very similar. Some trials with a data generating process allowing all β coefficients to vary with τ were done, but again results were rather similar.

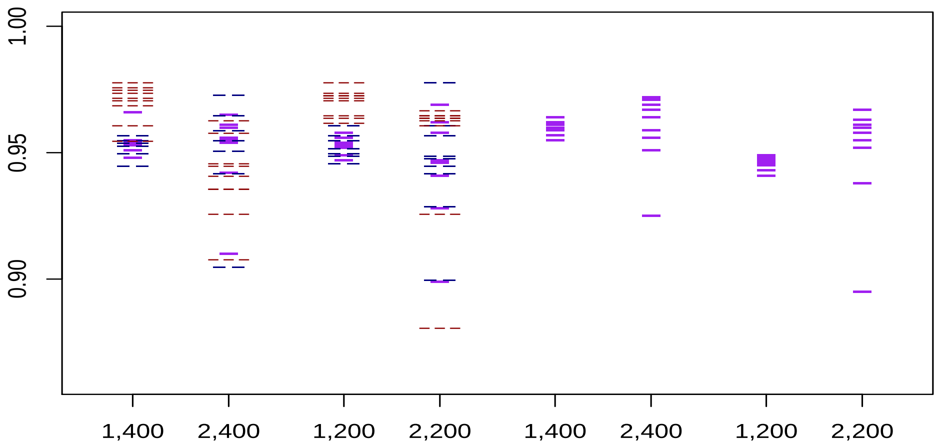

Fig. 1: Simulated Coverage Probabilities for $\hat{\alpha}$ and $\hat{\beta}$ (combined over all simulation trials); x-axis labels = (component, sample size); distributions: - : X and e normal, - - : X and e normalized Chi Square, - - - : X normal and $e \sim t$ with 3 df.

It seems clear that the coverage probabilities for $\tau = .5$ are quite reasonable, especially for the first component. The variation in coverages generally reflect random variation with standard error of order $1/\sqrt{n}$, as expected. The second component gives some low coverages, especially for $n = 200$, but this is not surprising given that the orthogonality constraints promulgate errors from estimating the first component into the second one. Results for $\tau = .25$ used only normal errors. These values seem more variable, but that is to be expected since the asymptotic standard deviation of a quantile is inversely proportional to the density at the quantile. The fact that some coverages are a bit low suggests that some adjustment for subsequent components might be useful, but a formal approach to this problem would require second order asymptotics, which are not available for this model. Overall, the coverage results seem similar to results for other multivariate procedures with relatively modest sample sizes.

The simulations also kept track of biases and the ratio of the average bootstrap standard error to the simulation standard error. Table 2 presents the R-summary results for these values for components 1 and 2. The biases for the first component are quite small, especially considering that the size of the coefficients and fact that the standard errors are generally between .05 and .4. The bias in the second component are again larger, but in fact generally somewhat smaller than the standard errors. The ratio of standard errors show the bootstrap value to be a bit larger than the simulation standard error, which agrees with the coverages generally being slightly conservative. As noted above, confidence intervals based on bootstrap standard errors were overly conservative and are not presented. Their relatively poor behavior is likely because the distribution of the constrained estimates was somewhat far from normal.

Overall, the simulation data indicates that the basic asymptotic theory is useful and suggests that reasonably reliable inference is possible.

Table 2: Summary statistics computed over all simulations for the absolute bias and standard error ratio for CanRQ estimates of the parameters for components 1 and 2.

Statistic	1st Qtile	Median	Mean	3rd Qtile	Max
|bias|,1	0.0004	0.003	0.007	0.007	0.144
|bias|,2	0.077	0.475	0.385	0.655	0.927
bootSE/simSE,1	1.254	1.272	1.279	1.318	1.417
bootSE/simSE,2	1.336	1.406	1.453	1.571	2.012

6.2. A Psychology data set

The data set is available at: `https://stats.idre.ucla.edu/r/dae/canonical-correlation-analysis/` and consists of measurements on a sample of 600 male and female students of three psychological variables (loss-of-control, self-concept, and motivation) and 4 academic variables (reading, writing, math, and science). The question concerned how well the psychological variables explain the academic ones. This clearly suggests a regression setting rather than one of dimension reduction. The following tables present the summary results for canonical regression quantile analysis. The tables give results for the first two components for $\tau \in \{.5, .25, .75\}$. They list the estimates, the bootstrap standard error estimate, and the bootstrap percentile 95% confidence interval (described at the end of the previous section). The tables also list a normalized version of the corresponding coefficient from classical Canonical Correlations. These are normalized to sum to 1 in absolute value in order to provide more direct comparison with the regression quantile values. Canonical Correlation coefficients that are significantly non-zero at a nearly .05 level (or less) are starred.

Without input from the psychologists who undertook the study, it is not clear what constraints to impose on the leading component. The results below used the specification $\alpha_2 \geq 0$, which seemed reasonable since the largest classical Canonical Correlation is associated with the Writing test. In fact several constraint sets were tried, and most others led to less intuitive and interpretable results. As noted in Section 3, specifications that agree with the α signs below under $\alpha_2 \geq 0$ must give the same results.

It should be noted that the first attempt to apply CanRQ specified all α's to be positive (since none of the classical α coefficients were negative). This led to very different solutions that were much closer to the classical Canonical Correlations, especially for the β coefficients. The reason for this is likely the small (non-significant) classical coefficient for Concept, which suggests that this value may be negative (or zero) and, hence, should not be constrained to be positive. With positivity constraints on all Y's, there was still some heterogeneity at $\tau = .75$, but it was much smaller. In all cases, the signs and magnitudes of the coefficients did not seem as intuitive as those given by assuming $\alpha_2 \geq 0$. Finally, note that both objective functions (for both components) were substantially smaller under $\alpha_2 \geq 0$ than under the specification that all α signs are positive.

The CanRQ estimates with 95% confidence intervals are listed in Tables 3 - 8. These tables also give classical Canonical Correlations estimates, with the α coefficients normalized by L_1-norm (rather than by L_2 norm) and β coefficients defined by least-squares regression on the α-index in order to facilitate comparison with CanRQ. Canonical Correlation coefficient that are significant at .05 are starred.

The analyses can be summarized as follows. First consider $\tau = .5$ (Tables 3 and 4). For the leading CanRQ component, indeed, the signs of the α's are not all positive; so the choice of sign constraint is important. The response index weights Writing and Science positive and Reading and Math negatively. All coefficients are significant. Thus, academic performance is not monotonically increasing in the best overall measure of the psychological scores. This is markedly different from the classical Canonical Correlation analysis, where all coefficients are positive, and only the Reading and Writing scores are significant. As will be clear from analysis of other τ values, heterogeneity of the data probably plays a role in this, though non-normality may also contribute. As suggested by differences in the α's, the CanRQ β's differ very substantially from those of the classical Canonical Correlations. Among classical coefficients, Control and Motivation (and the intercept) are significant, while CanRQ has only Control significant.

The second component at $\tau = .5$ also shows substantial differences between the two methods. For the α's, CanRQ has reading and writing positive while math and science are negative, a seemingly intuitive result. All coefficients are significant. Classical Canonical Correlations have none of the α coefficients significant (though science is nearly so). Furthermore, all the β coefficients are significant for CanRQ, while only motivation (and the intercept) are significant for the classical case.

The results for $\tau = .25$ are rather similar to those for $\tau = .5$ (see Tables 5 and 6); and so will not be discussed further here. However, the results for $\tau = .75$ are rather different (see Tables 7 and 8), and so they indicate substantial heterogeneity that would be missed by classical Canonical Correlations. Here the first component α's are positive for reading and writing, but negative for math and science (again, intuitively reasonable). Again, all $\hat{\alpha}$ coefficients are significant. However, only the β intercept is significant, and there is no statistical evidence that the psychological variables explain the third quartile responses. Thus, the CanRQ response index at $\tau = .75$ is minimizing the Y variability about the mean as measured by the .75 quantile loss. This perhaps is like a Principal Components analysis, though the components are now ordered in reverse from smallest variability to largest.

The second CanRQ component at $\tau = .75$ looks rather like the first component at $\tau = .5$, with the coefficient signs lumping reading with math and writing with science. However, here, none of the β's are significant; again perhaps allowing an interpretation akin to reversed Principal Components Analysis.

Clearly, CanRQ appears to correct inadequacies of classical Canonical Correlations most likely caused by non-normality and heterogeneity of the sample. As noted above, the significant differences between the .5 and .75 quantile estimates provide a clear indication of heterogeneity in the sample, most likely reflecting population heterogeneity. The significant differences between the .5 quantile estimates (conditional medians) and their canonical correlation (conditional mean) estimates also suggests inadequacies of the classical analysis, since the median analysis does not require normality and is resistant to outliers. Thus, CanRQ provides a far more nuanced interpretation, especially in terms of predicting the test measurements from the psychological measures. Of course, some caution is required, especially since many of the coefficients are not statistically significant. A larger sample size in subsequent studies would likely be very helpful in this respect.

Table 3: Psych Data, Canonical RQ estimates for Component 1 at $\tau = .5$. The column headings "ci1:low" and "ci1:up" refer to the lower and upper confidence bounds for the bootstrap percentile method described at the end of Section 5. The heading "cancor.coef" refers to the classical canonical correlations renormalized to be directly comparable to the CanRQ estimates, with starred values indicating significance at the .05 level.

Parameter	Value	Std. Error	ci1:low	ci1:up	cancor.coef
α:Read	-0.2408	0.1953	-0.6625	-0.1738	0.3541 *
α:Write	0.2746	0.0348	0.2369	0.3605	0.4206 *
α:Math	-0.2521	0.0927	-0.6329	-0.1986	0.1562
α:Science	0.2325	0.1945	0.1763	0.6664	0.0691
β:int	-0.7795	2.1620	-5.7319	3.1809	-48.5550 *
β:Control	-0.5986	0.4517	-1.8101	-0.1215	-5.2884 *
β:Concept	-0.2777	0.4415	-1.2576	0.2995	0.3554
β:Motivation	0.8128	1.2082	-0.3506	3.9001	-4.4012 *

Table 4: Psych Data, Canonical RQ estimates for Component 2 at $\tau = .5$. The column headings "ci1:low" and "ci1:up" refer to the lower and upper confidence bounds for the bootstrap percentile method described at the end of Section 5. The heading "cancor.coef" refers to the classical canonical correlations renormalized to be directly comparable to the CanRQ estimates, with starred values indicating significance at the .05 level.

Parameter	Value	Std. Error	ci1:low	ci1:up	cancor.coef
α:Read	0.2248	0.0569	0.0789	0.3437	-0.1534
α:Write	0.2510	0.1916	0.1914	0.6935	-0.2711
α:Math	-0.2209	0.1989	-0.6764	-0.1317	-0.0342
α:Science	-0.3032	0.1652	-0.7864	-0.2596	0.5413 *
β:int	4.7393	2.1390	3.0243	11.3630	-6.6817 *
β:Control	0.5009	0.3420	0.0148	1.4170	-0.5167
β:Concept	0.6674	0.3634	0.4222	1.8115	-0.4741
β:Motivation	-2.0594	1.0763	-5.6420	-1.4918	1.9910 *

Table 5: Psych Data, Canonical RQ estimates for Component 1 at $\tau = .25$. The column headings "ci1:low" and "ci1:up" refer to the lower and upper confidence bounds for the bootstrap percentile method described at the end of Section 5. The heading "cancor.coef" refers to the classical canonical correlations renormalized to be directly comparable to the CanRQ estimates, with starred values indicating significance at the .05 level.

Parameter	Value	Std. Error	ci1:low	ci1:up	cancor.coef
α:Read	-0.2685	0.2008	-0.7296	-0.1472	0.3541 *
α:Write	0.2525	0.0389	0.2030	0.3466	0.4206 *
α:Math	-0.2402	0.1286	-0.6331	-0.0741	0.1562
α:Science	0.2388	0.1932	0.1841	0.6827	0.0691
β:int	-1.1681	2.5396	-5.1037	4.6645	-48.5550 *
β:Control	0.0320	0.4608	-0.6871	1.0157	-5.2884 *
β:Concept	0.0954	0.4703	-0.8168	0.9284	0.3554
β:Motivation	1.2788	1.3718	-0.2211	4.9695	-4.4012 *

Table 6: Psych Data, Canonical RQ estimates for Component 2 at $\tau = .5$. The column headings "ci1:low" and "ci1:up" refer to the lower and upper confidence bounds for the bootstrap percentile method described at the end of Section 5. The heading "cancor.coef" refers to the classical canonical correlations renormalized to be directly comparable to the CanRQ estimates, with starred values indicating significance at the .05 level.

Parameter	Value	Std. Error	ci1:low	ci1:up	cancor.coef
α:Read	0.2236	0.0768	0.0057	0.3698	-0.1534
α:Write	0.2728	0.1930	0.2275	0.7572	-0.2711
α:Math	-0.2325	0.1844	-0.6716	-0.0616	-0.0342
α:Science	-0.2711	0.2011	-0.7528	-0.2192	0.5413 *
β:int	0.3470	2.2875	-3.3680	6.3790	-6.6817 *
β:Control	-0.1245	0.4233	-1.0603	0.6283	-0.5167
β:Concept	0.3751	0.4283	-0.3932	1.2665	-0.4741
β:Motivation	-1.3682	1.0208	-4.0058	-0.4218	1.9910 *

Table 7: Psych Data, Canonical RQ estimates for Component 1 at $\tau = .75$. The column headings "ci1:low" and "ci1:up" refer to the lower and upper confidence bounds for the bootstrap percentile method described at the end of Section 5. The heading "cancor.coef" refers to the classical canonical correlations renormalized to be directly comparable to the CanRQ estimates, with starred values indicating significance at the .05 level.

Parameter	Value	Std. Error	ci1:low	ci1:up	cancor.coef
α:Read	0.2738	0.2129	0.2177	0.7483	0.3541 *
α:Write	0.2221	0.0391	0.1423	0.2902	0.4206 *
α:Math	-0.2488	0.1113	-0.6352	-0.1525	0.1562
α:Science	-0.2553	0.1948	-0.6776	-0.1653	0.0691
β:int	3.6275	2.0380	0.5900	8.4320	-48.5550 *
β:Control	0.2145	0.4877	-0.4858	1.3584	-5.2884 *
β:Concept	0.1622	0.5395	-0.6479	1.4561	0.3554
β:Motivation	-1.3087	1.1273	-4.3598	0.2696	-4.4012 *

Table 8: Psych Data, Canonical RQ estimates for Component 2 at $\tau = .75$. The column headings "ci1:low" and "ci1:up" refer to the lower and upper confidence bounds for the bootstrap percentile method described at the end of Section 5. The heading "cancor.coef" refers to the classical canonical correlations renormalized to be directly comparable to the CanRQ estimates, with starred values indicating significance at the .05 level.

Parameter	Value	Std. Error	ci1:low	ci1:up	cancor.coef
α:Read	0.1979	0.0640	0.0120	0.3206	-0.1534
α:Write	-0.2397	0.1908	-0.6968	-0.1044	-0.2711
α:Math	0.2868	0.2119	0.2151	0.8553	-0.0342
α:Science	-0.2756	0.1923	-0.7373	-0.2077	0.5413 *
β:int	3.5058	2.5648	-0.7944	9.3866	-6.6817 *
β:Control	0.0892	0.4039	-0.6060	1.1138	-0.5167
β:Concept	-0.0081	0.3743	-0.5318	0.8141	-0.4741
β:Motivation	-0.7996	1.0561	-3.4221	0.6390	1.9910 *

6.3. A Finance data set

Finally, some results for the motivating example are presented. Annual data was collected from 2009 to 2018, a 10 year period with relatively stable overall economic performance. The response variables were 4 measures of company performance, while the explanatory variables were generally accepted leading indicators of future performance. The analysis focussed on predicting performance two years in advance based on summaries of the 5 previous years explanatory variables, plus an indicator variable for 6 sectors. The canonical regression quantile analysis focussed on using the first component and aimed to construct a linear index based on the previous 5 years of data to predict the company performance variables better than CEO total compensation. Indeed, the coefficients in the first component were able to predict all 4 performance variables (2 years ahead) better than CEO compensation; and the improvement was greater when including past values of the performance variables (as suggested by Granger causality; see Granger [7] and subsequent work). The details of this analysis appear in Haimberg and Portnoy [14].

While the original analysis considered only the first component, it seems reasonable to look at subsequent components. Specifically, consider plotting the first two components to provide a graphical approach for distinguishing the sectors. In particular, let X denote the original design matrix for the explanatory variables, without the sector indicator (and without the previous response variables). Compute the coefficients $\hat{\alpha}^{(1)}$ and $\hat{\alpha}^{(2)}$ for the first two components and plot $Y\hat{\alpha}^{(1)}$ against $Y\hat{\alpha}^{(2)}$. To simplify the picture, consider only 3 groups: the science sectors (health and technology sectors), utilities, and the other 3 sectors combined. The plots are given in Fig. 2. The regression quantile plots (especially for $\tau = .75$) appear to separate the science companies from the others rather more clearly.

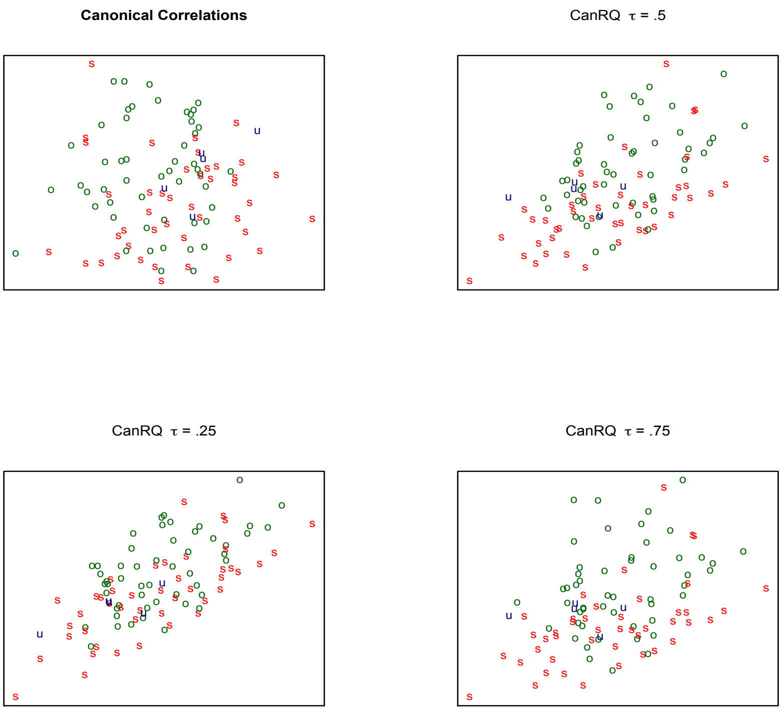

Fig. 2: Scatter plot for fitted Y-indices of first two components. “S” denotes health and technology companies, “U” denotes utilities, and “O” denotes all other sectors. The first component is on the x-axis; the second is on the y-axis.

7. Conclusions.

A Canonical Quantile Regression method has been developed in analogy with Canonical Correlations. Basically, one seeks linear combinations of explanatory and response variables with the L_2 normalization of Canonical Correlations replaced by an L_1 normalization and with mean squared error replaced by the regression quantile objective function. A method for estimating the coefficients for the leading component was developed and extended to subsequent components. An unresolved conceptual problem is the imposition of orthogonality: outside normal models, the resulting indices will not be independent. While it seems clear that specific applications may suggest alternatives to orthogonality, none were pursued here, neither for canonical quantile regression nor for traditional Canonical Correlations.

Inferential methods were also developed based on a weighted n-choose-m bootstrap that followed from work by Andrews [1] on estimators subject to constraints and further developments. Formal asymptotic theory and some preliminary simulations indicate that the methodology appears to be successful. Two real-data examples show that the methodology can achieve the traditional advantages of a conditional quantile analysis. In addition to indicating clear heterogeneity, the quantile analysis provides a more complete, more nuanced, and more robust picture of the data. While much more extensive application and simulation are needed to ensure that such novel methodology is indeed trustworthy in great generality, the initial analyses are very promising.

Acknowledgements

I want to thank Xuming He and Roger Koenker for following earlier versions of this work, and for a number of critically important comments and questions. I want to thank Joseph Haimberg for providing the financial data set in Subsection 6.3 and for providing extensive guidance in its analysis. I also wish to thank two reviewers and the editor whose comments and corrections greatly improved the paper.

References

[1] D. Andrews, Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space, Econometrica 68(2) (2000) 399-405.
[2] G. Cavaliere and I. Georgiev, I, Inference under random limit bootstrap measures, Econometrica, 88 (2020) 2547-2574.
[3] S. Chatterjee and A. Bose, Generalized bootstrap for estimating equations, Annals of Statistics 33 (2005) 414–436.
[4] H. Chernoff, On the distribution of the likelihood ratio, Annals of Mathematical Statistics, 25 (1954) 573-578.
[5] H. Chernoff, Large-sample theory: parametric case, Annals of Mathematical Statistics, 27 (1956) 1-22.
[6] C. Geyer, On the asymptotics of constrained M-estimation, Annals of Statistics, 22 (1994) 1993-2010.
[7] C. Granger, Testing for causality: A personal viewpoint, Journal of Economic Dynamics and Control, 2 (1980) 329-352.
[8] M. Hlavac, `stargazer`: Well-Formatted Regression and Summary Statistics Tables, R package version 5.2.2, `https://CRAN.R-project.org/package=stargazer` (2018)
[9] R. Koenker, Quantile Regression, Cambridge University Press, Cambridge (2005).
[10] R. Koenker, `quantreg`: Quantile Regression. R package version 5.2.2, `https://CRAN.R-project.org/package=quantreg` (2021)
[11] T. Parker, Asymptotic inference for the constrained quantile regression process, Econometrics, 213 (2019) 174-189.
[12] D. Politis, J. Romano, M. Wolf, Subsampling, Springer-Verlag, New York (1999).
[13] S. Portnoy, The jackknife's edge: Inference for censored regression quantiles, Comp. Statist. Data Analysis, 72 (2013) 273-281.
[14] S. Portnoy, J. Haimberg, Canonical quantile regression with application to CEO compensation and predicting company performance, arXiv:2011.08896 (2021).
[15] R Core Team. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, `https://www.r-project.org` (2021).