

A Permutation Test for Quantile Regression

Brian S. CADE and Jon D. RICHARDS

A drop in dispersion, F -ratio like, permutation test (D) for linear quantile regression estimates ($0 \leq \tau \leq 1$) had relative power ≥ 1 compared to quantile rank score tests (T) for hypotheses on parameters other than the intercept. Power was compared for combinations of sample sizes ($n = 20 - 300$) and quantiles ($\tau = 0.50 - 0.99$) where both tests maintained valid Type I error rates in simulations with $p = 2$ and 6 parameters in homogeneous and heterogeneous error models. The D test required two modifications of permuting residuals from null, reduced parameter models to maintain correct Type I error rates when null models were constrained through the origin or included multiple parameters. A double permutation scheme was used when null models were constrained through the origin and all but 1 of the zero residuals were deleted for null models with multiple parameters. Although there was considerable overlap in sample size, quantiles, and hypotheses where both the D and rank score tests maintained correct Type I error rates, we identified regions at smaller n and more extreme quantiles where one or the other maintained better error rates. Confidence intervals on parameters for an ecological application relating Lahontan cutthroat trout densities to stream channel width:depth were estimated by test inversion, demonstrating a smoother pattern of slightly narrower intervals across quantiles than those provided by the rank score test.

Key Words: Ecology; Habitat relationships; Limiting factors; Linear models; Permutation tests; Quantile regression; Rank score tests.

1. INTRODUCTION

Estimating the quantiles ($0 \leq \tau \leq 1$) of a response variable conditional on some set of covariates in a linear model has many applications in the biological and ecological sciences (Cade, Terrell, and Schroeder 1999; Cade and Noon 2003). Quantile regression models allow the entire conditional distribution of a response variable y to be related to some covariates X , providing a richer description of functional changes than is possible by focusing on just the mean (or other central statistics), yet requiring minimal distributional assumptions (Koenker and Bassett 1978, 1982; Koenker and Machado 1999). Quantile regression estimates are especially enlightening for relationships involving heterogeneous responses where by definition rates of change are not the same across all parts of the response

Brian S. Cade is a Biological Statistician and Jon D. Richards is an Operations Research Analyst, Fort Collins Science Center, U. S. Geological Survey, 2150 Centre Avenue, Building C, Fort Collins, CO 80526-8118 (E-mail: brian.cade@usgs.gov).

In the Public Domain

Journal of Agricultural, Biological, and Environmental Statistics, Volume 11, Number 1, Pages 106–126

DOI: 10.1198/108571106X96835

distribution. Many ecological applications of quantile regression have focused on estimating some upper quantiles to characterize effects of limiting factors (Terrell, Cade, Carpenter, and Thompson 1996; Scharf, Juanes, and Sutherland 1998; Cade et al. 1999; Cade and Guo 2000; Haire, Bock, Cade, and Bennett 2000; Eastwood, Meaden, and Grische 2001; Huston 2002). Other applications (Allen, Cade, and Vandever 2001; Dunham, Cade, and Terrell 2002; Cade, Noon, and Flather 2005) have used estimates across the entire $[0, 1]$ interval of quantiles as a flexible method of characterizing effects associated with heterogeneous distributions.

The Dunham et al. (2002) analysis of Lahontan cutthroat trout (*Oncorhynchus clarki henshawi*) standing stock as a function of channel morphology for small streams sampled in the state of Nevada is used as an example of ecological relationships estimated with quantile regression (Figure 1). Trout were captured by electrofishing in seven or more 25-meter (m) sections within a two-kilometer (km) reach in each of 13 streams from 1993 to 1999, and abundance in a stream was estimated with maximum likelihood estimates for removal methods (Dunham et al. 2002). Channel widths and depth were measured in each section and averaged for a stream. Here we considered the model relating density (trout m^{-1}) of age ≥ 1 year trout to stream channel width:depth ratio ($n = 71$). Width:depth ratio is a measure that integrates channel characteristics thought to be related to small stream integrity—and, thus, fish populations—and is easily measured for assessing fish habitat conditions and land use impacts over large regions. Lahontan cutthroat trout are a threatened species of special interest to federal land management agencies. Overgrazing by livestock in riparian areas and road construction impacts on the trout populations are reflected in increasing width:depth ratios of stream channels.

The scatterplot in Figure 1(a) clearly demonstrates a pattern of increasing variation with decreasing width:depth ratios, where highest trout densities were observed only at smaller width:depth ratios. There appears to be some functional nonlinear relationship between trout density and width:depth ratio, but it is not adequately captured by using mean regression. A weighted least squares estimate of the mean function on natural logarithm of densities ($\ln y$) yielded estimates of $b_0 = -1.475$ (90%CI = $-1.94, -1.01$) and $b_1 = 0.001$ (90%CI = $-0.014, 0.012$), clearly supporting the impression that there is little change in mean densities [$P(H_0 : \beta_1 = 0) = 0.90$]. Yet higher quantiles for the same model indicated a decline in densities with width:depth ratio as shown here for the 0.90 quantile regression (Figure 1(a)). Our motivation for developing this permutation test was to provide improved inferences for quantile regression applications with small to moderate sample sizes where methods based on asymptotic arguments are likely to perform poorly.

Inference methods with asymptotic validity for linear quantile regression models based on estimates of the covariance matrices (Koenker and Bassett 1978, 1982; Koenker and Machado 1999) require estimates of the reciprocal of the error density function at the quantile of interest, $f(F^{-1}(0))$. These methods often perform poorly at smaller sample sizes (Koenker 1987; Buchinsky 1991) and the asymptotic theory becomes suspect at more extreme ($\tau > 0.7$ and $\tau < 0.3$) quantiles (Chernozhukov and Umantsev 2001). Koenker (1994) introduced the idea of constructing confidence intervals by inverting a quantile rank

score test (Gutenbrunner, Jurečková, Koenker, and Portnoy 1993) as an alternative inference procedure which does not require estimating the sparsity function and that performed well under linear heteroscedastic regression models and smaller sample sizes. Here we consider a drop in dispersion, F -ratio like, test that is evaluated with permutation arguments based on modifications of the least absolute deviation regression test of Cade and Richards (1996). This test also avoids the sparsity estimation issue, but unlike the quantile rank score tests

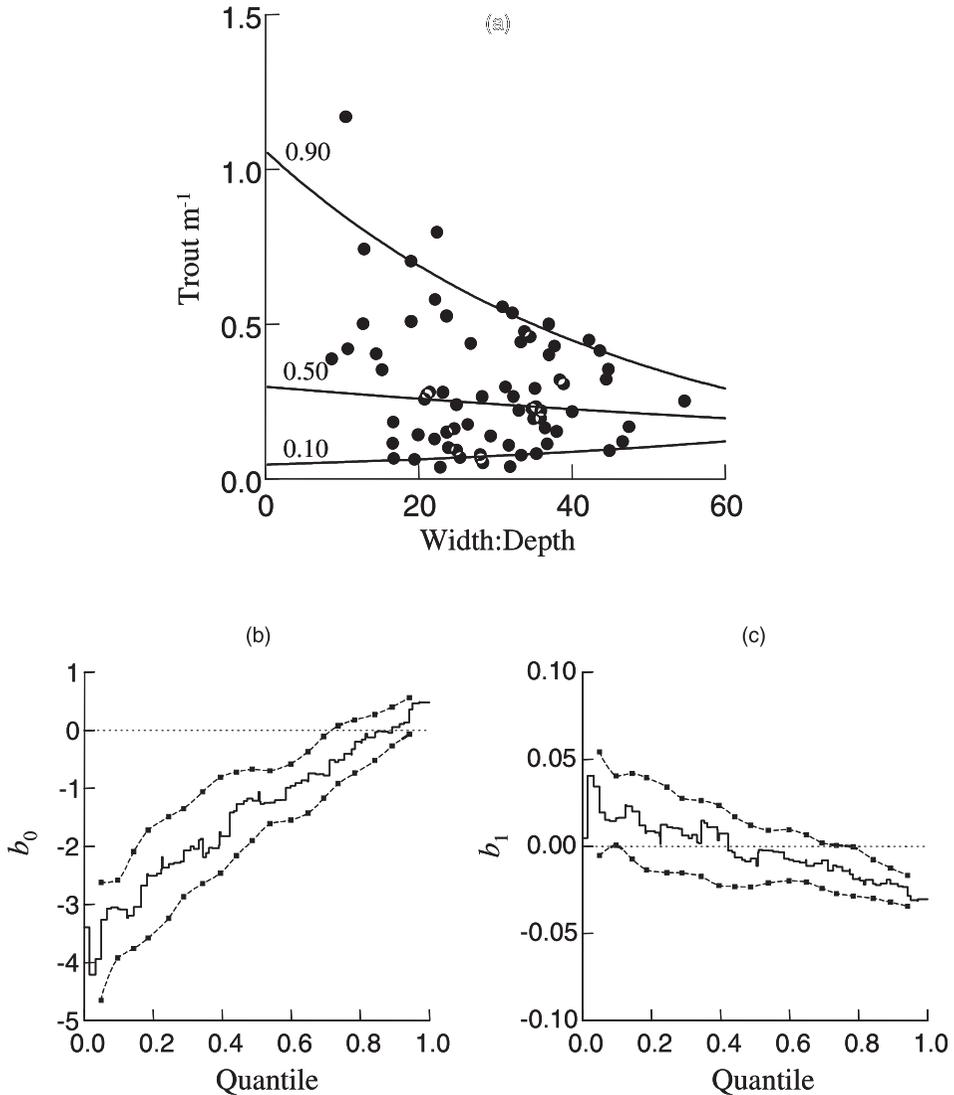


Figure 1. (a) Lahontan cutthroat trout m^{-1} and width:depth ratios for small streams sampled 1993 to 1999 ($n = 71$); exponentiated estimates for 0.90, 0.50, and 0.10 regression quantiles for the weighted model $w(\ln y) = w(\beta_0 + \beta_1 X_1 + (\gamma_0 - \gamma_1 X_1)\varepsilon)$, $w = (1.310 - 0.017 X_1)^{-1}$. Solid lines in (b) and (c) are step functions for estimates of β_0 and β_1 by $\tau = [0, 1]$ and dashed lines connect pointwise 90% confidence intervals for $\tau \in \{0.05, 0.10, 0.15, \dots, 0.95\}$ based on inverting the double permutation D test.

(Koenker 1994) it uses the magnitude of the residuals in its construction, potentially providing greater power and shorter confidence intervals. Two additional modifications of the basic permutation structure used by Cade and Richards (1996) were developed to provide improved Type I error rates when null hypothesis models were constrained through the origin or included multiple parameters.

We evaluated performance of the drop in dispersion permutation test for central to extreme quantiles, a range of error structures, small to moderate sample sizes, and model forms likely to be encountered in ecological applications. Weighted forms of the test based on weighted quantile regression estimates were evaluated for heterogeneous error distributions. The drop in dispersion permutation test was applied to the quantile regression analysis of Lahontan cutthroat trout response to variations in their stream habitat, expanding on the previous analyses of Dunham et al. (2002) made using the quantile rank score test. The simulations and example application parallel those used by Cade, Richards, and Mielke (2006) for evaluating the chi-square distribution and permutation approximations of probabilities for the quantile rank score test.

2. QUANTILE REGRESSION MODEL

The τ th regression quantile ($0 \leq \tau \leq 1$) for the heteroscedastic linear location-scale model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Gamma}\boldsymbol{\varepsilon}$ is defined as $Q_y(\tau|\mathbf{X}) = \mathbf{X}\boldsymbol{\beta}(\tau)$ and $\boldsymbol{\beta}(\tau) = \boldsymbol{\beta} + F_{\boldsymbol{\varepsilon}}^{-1}(\tau)\boldsymbol{\gamma}$; where \mathbf{y} is an $n \times 1$ vector of dependent responses, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression parameters, \mathbf{X} is an $n \times p$ matrix of predictors (first column consists of 1's for an intercept term), $\boldsymbol{\gamma}$ is a $p \times 1$ vector of unknown scale parameters, $\boldsymbol{\Gamma}$ is a diagonal $n \times n$ matrix in which the n diagonal elements are the n corresponding ordered elements of the $n \times 1$ vector $\mathbf{X}\boldsymbol{\gamma}$ ($\text{diag}(\mathbf{X}\boldsymbol{\gamma})$), $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of random errors that are independent and identically distributed (iid), and $F_{\boldsymbol{\varepsilon}}^{-1}$ is the inverse of the cumulative distribution of the errors (Koenker and Bassett 1982; Buchinsky 1991; Gutenbrunner and Jurečková 1992; Koenker and Machado 1999). Homoscedastic regression models are a special case of the linear-location scale model when $\boldsymbol{\gamma} = (1, 0, \dots, 0)'$ and $Q_y(\tau|\mathbf{X}) = \mathbf{X}\boldsymbol{\beta}(\tau)$, $\boldsymbol{\beta}(\tau) = \boldsymbol{\beta} + (F_{\boldsymbol{\varepsilon}}^{-1}(\tau), 0, \dots, 0)'$, where all parameters other than the intercept (β_0) in $\boldsymbol{\beta}(\tau)$ are the same for all τ . More general forms of heteroscedastic errors can be accommodated with regression quantiles (Koenker 1997; Koenker and Machado 1999) but were not considered here.

The restriction imposed on $F_{\boldsymbol{\varepsilon}}$ to estimate regression quantiles is that a τ th quantile of $\mathbf{y} - \mathbf{X}\boldsymbol{\beta}(\tau)$ conditional on \mathbf{X} equals 0, $F_{\boldsymbol{\varepsilon}}^{-1}(\tau|\mathbf{X}) = 0$. Estimates, $\mathbf{b}(\tau)$, of $\boldsymbol{\beta}(\tau)$ are solutions to the following minimization problem:

$$\min \left[\sum_{i=1}^n \rho_{\tau} \left(y_i - \sum_{j=0}^p b_j x_{ij} \right) \right], \quad (2.1)$$

where $\rho_{\tau}(e) = e(\tau - I(e < 0))$ and $I(\cdot)$ is the indicator function assigning 1's for negative residuals and 0's for non-negative residuals. The estimating equations in (2.1) yield primal solutions in a modification of the Barrodale and Roberts (1974) simplex linear program for

any specified value of τ (Koenker and d'Orey 1987). With little additional computation the entire regression quantile process for all distinct values of τ can be estimated (Koenker and d'Orey 1987, 1994).

Consistent estimates with reduced sampling variation for heteroscedastic linear models can be obtained by implementing weighted versions of the regression quantile estimators, where weights are based on the sparsity function at a given quantile and covariate value (Koenker and Portnoy 1996; Koenker and Machado 1999). In the linear location-scale model this simplified to using an $n \times n$ weights matrix, $\mathbf{W} = \Gamma^{-1}$, where the $p \times 1$ vector of scale parameters γ would usually have to be estimated in applications (Gutenbrunner and Jurečková 1992; Koenker and Zhao 1994; Koenker and Machado 1999). The weighted regression quantile estimates then are given by:

$$\min \left[\sum_{i=1}^n \rho_{\tau} \left(y_i - \sum_{j=0}^p b_j x_{ij} \right) w_i \right], \quad (2.2)$$

where w_i is a weight, $\rho_{\tau}(e) = e(\tau - I(e < 0))$, and $I(\cdot)$ is the indicator function assigning 1's for negative residuals and 0's for nonnegative residuals. This is easily implemented by multiplying \mathbf{y} and \mathbf{X} by \mathbf{W} and then using the unweighted estimator (2.1).

3. TEST STATISTIC

The drop in dispersion D test was based on a modification of the permutation test developed for least absolute deviation (LAD) regression (Cade and Richards 1996). The reduced parameter model $\mathbf{y} - \mathbf{X}_2 \boldsymbol{\xi}(\tau) = \mathbf{X}_1 \boldsymbol{\beta}_1(\tau) + \Gamma \boldsymbol{\varepsilon}$ is constructed by partitioning $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$, where \mathbf{X}_1 is $n \times (p - q)$ and \mathbf{X}_2 is $n \times q$; and by partitioning $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$, where $\boldsymbol{\beta}_1(\tau)$ is a $(p - q) \times 1$ vector of unknown nuisance parameters under the null and $\boldsymbol{\beta}_2(\tau)$ is a $q \times 1$ vector of parameters specified by the null hypothesis $H_0 : \boldsymbol{\beta}_2(\tau) = \boldsymbol{\xi}(\tau)$ (frequently $\boldsymbol{\xi}(\tau) = \mathbf{0}$) in the full parameter model $\mathbf{y} = \mathbf{X}_1 \boldsymbol{\beta}_1(\tau) + \mathbf{X}_2 \boldsymbol{\beta}_2(\tau) + \Gamma \boldsymbol{\varepsilon}$; and \mathbf{y} , Γ , and $\boldsymbol{\varepsilon}$ are as above. The sum of weighted absolute deviations minimized in (2.1) for estimating the weighted version of the full parameter model $\mathbf{W}\mathbf{y} = \mathbf{W}\mathbf{X}_1 \boldsymbol{\beta}_1(\tau) + \mathbf{W}\mathbf{X}_2 \boldsymbol{\beta}_2(\tau) + \mathbf{W}\Gamma \boldsymbol{\varepsilon}$, where \mathbf{W} is a weights matrix as in (2.2), are denoted SAFw(τ) and for the reduced parameter model $\mathbf{W}\mathbf{y} - \mathbf{W}\mathbf{X}_2 \boldsymbol{\xi}(\tau) = \mathbf{W}\mathbf{X}_1 \boldsymbol{\beta}_1(\tau) + \mathbf{W}\Gamma \boldsymbol{\varepsilon}$ are denoted SARw(τ). For homogeneous error models $\mathbf{W} = \mathbf{I}$, where \mathbf{I} is the $n \times n$ identity matrix. The test statistic

$$D_o = (\text{SARw}(\tau) - \text{SAFw}(\tau)) / \text{SAFw}(\tau), \quad (3.1)$$

was evaluated by permuting the weighted residuals under the null model to the weighted full model matrix $\mathbf{W}\mathbf{X}$. By taking a large random sample m of the $n!$ possible permutations, probability under the null hypothesis was approximated by the proportion of permuted test statistic values greater than or equal to the observed test statistic, where the observed test statistic counts as one of the permuted values [$P = ((\text{the number of } D \geq D_o) + 1) / (m + 1)$]. When the error distributions are assumed homogeneous so that $\mathbf{W} = \mathbf{I}$ and $\tau = 0.50$, this test statistic is identical to the statistic of Cade and Richards (1996) for LAD regression.

The weights, \mathbf{W} , serve to eliminate the effects of heterogeneous errors so that permuting residuals from the weighted estimates provide an approximation of the sampling distribution of D .

Permuting residuals ($\mathbf{e} = \mathbf{W}\mathbf{y} - \mathbf{W}\mathbf{X}_1\mathbf{b}_1$) under the null, reduced parameter model does not yield exact permutation probabilities except when the null parameter is just an intercept (β_0) and $\mathbf{W} = \mathbf{I}$. This permutation approach due to Freedman and Lane (1983) was found to have perfect correlation asymptotically with the exact test for least squares regression, which is only possible when β_1 is known (Anderson and Robinson 2001), and has performed well in simulations for least squares (Kennedy and Cade 1996; Anderson and Legendre 1999; Legendre 2000) and least absolute deviation regression (Cade and Richards 1996). There is some correlation ($-(n-1)^{-1}$) among the residuals for least squares regression estimates and they do not have constant variance ($E[\mathbf{e}\mathbf{e}'] = \sigma^2(\mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1')$), implying that they are not exchangeable (Good 2002; Commenges 2003). Dependency among the residuals decreases with increasing sample size providing some asymptotic justification for treating them as exchangeable random variables (Randles 1984; Good 2002).

For a linear estimator like least squares regression, Commenges (2003) found that linear transformations that preserve approximate exchangeability of the residuals from a model with $p - q$ parameters must reduce the dimensions of the matrix to $n - (p - q)$ or $n - (p - q) + 1$. A similar proof for quantile regression—which is not a linear estimator—is not available, but the concept remains relevant. It is known that an estimate for a quantile regression model with $p - q$ parameters must have at least $p - q$ residuals equal to 0 (Koenker and Machado 1999). Whenever $p - q > 1$ there will be a mass of residuals at zero ($F_e^{-1}(\tau|\mathbf{X}_1) = 0$), whereas the error distribution that is being approximated under the permutation argument does not have this mass density at zero. Therefore, to provide approximately exchangeable residuals under the null model, we deleted all but 1 of the $p - q$ zero residuals whenever $p - q > 1$ and reduced the dimension of \mathbf{e} to $n - (p - q) + 1$. To permute \mathbf{e} with dimension $n - (p - q) + 1$ against the $n \times p$ matrix \mathbf{X} , we randomly deleted $(p - q) - 1$ rows of \mathbf{X} at each permutation. This allowed all rows of the design matrix to affect the permutation distribution of the test statistic D while using only $n - (p - q) + 1$ rows of \mathbf{X} for the quantile regression estimate at each permutation.

An additional problem with the standard permutation structure occurs whenever the null model is forced explicitly ($H_0 : \beta_0(\tau) = 0$) or implicitly (weighted estimates that do not include all X associated with the weights function) through the origin. There is additional sampling variation not accounted for by the usual permutation distribution of the test statistic because a null model that is constrained through the origin no longer has residuals with τ -quantile = 0, that is, $F_e^{-1}(\tau|\mathbf{X}_1) \neq 0$. This is similar to the property that the mean of the residuals $\neq 0$ when least squares regression is constrained through the origin. If the number of positive, negative, and zero residuals are denoted by N^+ , N^- , N^0 , respectively, and if $N^0 = p - q$ under a null model that includes an intercept, then there are at most $n\tau$ negative residuals where $N^- \leq n\tau \leq N^- + N^0$ and at most $n(1 - \tau)$ positive residuals where $N^+ \leq n[1 - \tau] \leq N^+ + N^0$ (Koenker and Bassett 1978; Koenker and Portnoy 1996).

When the null model does not include an intercept, the limits on the number of positive (negative) residuals can differ from these values consistent with random binomial variation such that $F_e^{-1}(\tau^*|\mathbf{X}_1) = 0$, where τ^* is the proportion from a random binomial variable with probability τ given n . Consequently, we modified a recently proposed double permutation scheme for least squares regression through the origin (Legendre and Desdevises 2002) for quantile regression. At each permutation, the proportion τ^* was generated as a random binomial variable given n and τ . The null model residuals were centered on their τ^* -quantile by $\mathbf{e}^* = \mathbf{e} - F_e^{-1}(\tau^*|\mathbf{X}_1)$ and the centered residuals \mathbf{e}^* were permuted to \mathbf{X} and the test statistic D computed by (3.1).

4. SIMULATION EXPERIMENT

We first conducted a set of Monte Carlo simulations with homogeneous errors to establish the performance for models having a simple error structure. A greater range of simulations were then conducted for models with error heterogeneity requiring weighted estimates because these are the conditions where quantile regression estimates are most useful. Normal ($\mu = 0, \sigma = 1$), uniform (min = -2 , max = 2), and lognormal (median = $0, \sigma = 0.75$) error distributions were used to provide responses with symmetric, unimodal variation with greatest density at the center, symmetric variation with constant density, and asymmetric variation with lower density in the upper tail. Error distributions were centered on their 0.50, 0.75, 0.90, 0.95, or 0.99 quantiles so that $F_e^{-1}(\tau|\mathbf{X}) = 0$, providing a range of central to extreme regression quantiles. Note that similar simulation results for quantiles in the lower tail (0.25, 0.10, 0.05, and 0.01) would be obtained for the symmetric normal and uniform error distributions.

Simple two-parameter and six-parameter multiple regression models were simulated for $n = 20, 30, 60, 90, 150$, and 300 . Independent variables were structured to have a range of values and correlation structure similar to those expected in many ecological investigations. Independent variables were structured so that X_0 was a column of 1's for the intercept; X_1 was uniformly distributed (0, 100); X_2 was negatively correlated ($r = -0.89$) with X_1 specified by the function $X_2 = 4,000 - 20X_1 + N(\mu = 0, \sigma = 300)$; X_3 was positively correlated ($r = 0.94$) with X_1 specified by the function $X_3 = 10 + 0.4X_1 + N(\mu = 0, \sigma = 16)$; X_4 was a 0, 1 indicator variable randomly assigning half the sample to each of two groups; and X_5 was the multiplicative interaction of X_3 and X_4 . Thus, X_1 ranged from 0 to 100, X_2 had most values in the range 0 to 5,000 and was inversely related to X_1 , and X_3 provides values in the range 0 to 60 and was positively related to X_1 . Variables X_2 and X_3 were negatively correlated ($r = -0.85$) with each other through their indirect functional relation with X_1 . The indicator variable (X_4) and its interaction with X_3 (X_5) allowed the effect of X_3 for the regression quantile function to differ in slopes, intercepts, or both terms for the two groups.

The $p = 6$ parameter model was used because it allowed us to evaluate the permutation test for subsets of predictors where multicollinearity existed among continuous variables and for testing interactions with categorical variables. These are conditions common in linear

model applications where improperly constructed permutation tests often fail to maintain correct Type I errors (Cade and Richards 1996; Kennedy and Cade 1996; Anderson and Legendre 1999; Anderson and Robinson 2001; Good 2002). Permutation procedures that perform properly for multicollinear predictors also perform properly for uncorrelated predictors. We limited our simulations for weighted estimates to using the known weights to minimize the extra computing required to estimate weights for the large set of conditions we simulated and to eliminate confounding interpretations of test performance with a particular method of estimating weights. Nonnull parameter values for power simulations were selected to provide a range of small to larger effects that had power < 1 for most effects for most quantiles so that differences across quantiles, sample sizes, and between the D test and quantile rank score test would be evident.

Each combination of conditions (quantile, error distribution, sample size, and model structure) was sampled 1,000 times and the test statistic D_o was computed for each sample. Probabilities were evaluated with separate $m + 1 = 10,000$ random samples of the permutation distribution for D . Cumulative distribution function (cdf) plots of the Type I error probabilities under the null hypothesis were graphed and compared with the expected uniform cdf. Point estimates for $\alpha = 0.05$ and 0.10 , corresponding to coverage for 95% and 90% confidence intervals, were graphed across the combination of model conditions. The 99% binomial confidence intervals for 1,000 simulations (0.076 to 0.124 for $\alpha = 0.10$ and 0.032 to 0.068 for $\alpha = 0.05$) were used as a guide to judge when estimated Type I error rates exceeded variation expected from the sampling simulations. Power under the alternative hypotheses was graphed only for $\alpha = 0.05$ across all combinations of conditions, although cdf plots were initially examined.

All data for the simulation studies were generated with functions in S-Plus 2000 (Insightful, Inc., Seattle, WA). Regression quantile estimates and test statistics were computed by Fortran 95 routines implemented in the Blossom software available from the U. S. Geological Survey (www.fort.usgs.gov/products/software/blossom/blossom.asp).

4.1 HOMOGENEOUS ERROR STRUCTURE: SIMPLE REGRESSION

The simple two-parameter regression model $y = \beta_0 + \beta_1 X_1 + \varepsilon$ with $\beta_0 = 6.0$ and $\beta_1 = 0.0$ was evaluated for $H_0 : \beta_1 = 0$. Type I error rates were well maintained at all sample sizes, error distributions, and quantiles, consistent with exact exchangeability for this hypothesis (Figure 2). Type I errors for the 0.75 quantile were nearly identical to those for the 0.50 quantile for this and subsequent simulations and, therefore, were not graphed. We also evaluated Type I error rates for $H_0 : \beta_0 = 0.0$ with $\beta_1 = 0.10$ and $\beta_0 = 0.0$ in this regression model. The double permutation procedure maintained correct Type I error rates for testing the intercept except for $n < 60$ for $\tau = 0.95$ and $n < 300$ for $\tau = 0.99$ where they were liberal (Figure 2). Using the standard rather than the double permutation scheme for $H_0 : \beta_0 = 0$ resulted in slightly liberal Type I error rates as shown for the lognormal error distribution with $n = 90$ (Figure 3).

Power to detect nonzero slopes for $\beta_1 = 0.01, 0.05, 0.10,$ and 0.20 was progressively

lower moving from the 0.50 to 0.99 quantile and this reduction was greatest for the lognormal error distribution, less for the normal, and least for the uniform error distribution, consistent with changes in densities of observations around the quantiles of these distributions. Power increased with increasing effect size and sample size as desired. Power to detect nonzero intercepts for $\beta_0 = 0.5, 1.0, 2.0,$ and 3.0 with the double permutation scheme decreased with increasing quantile for the lognormal error distribution, with no effective power for

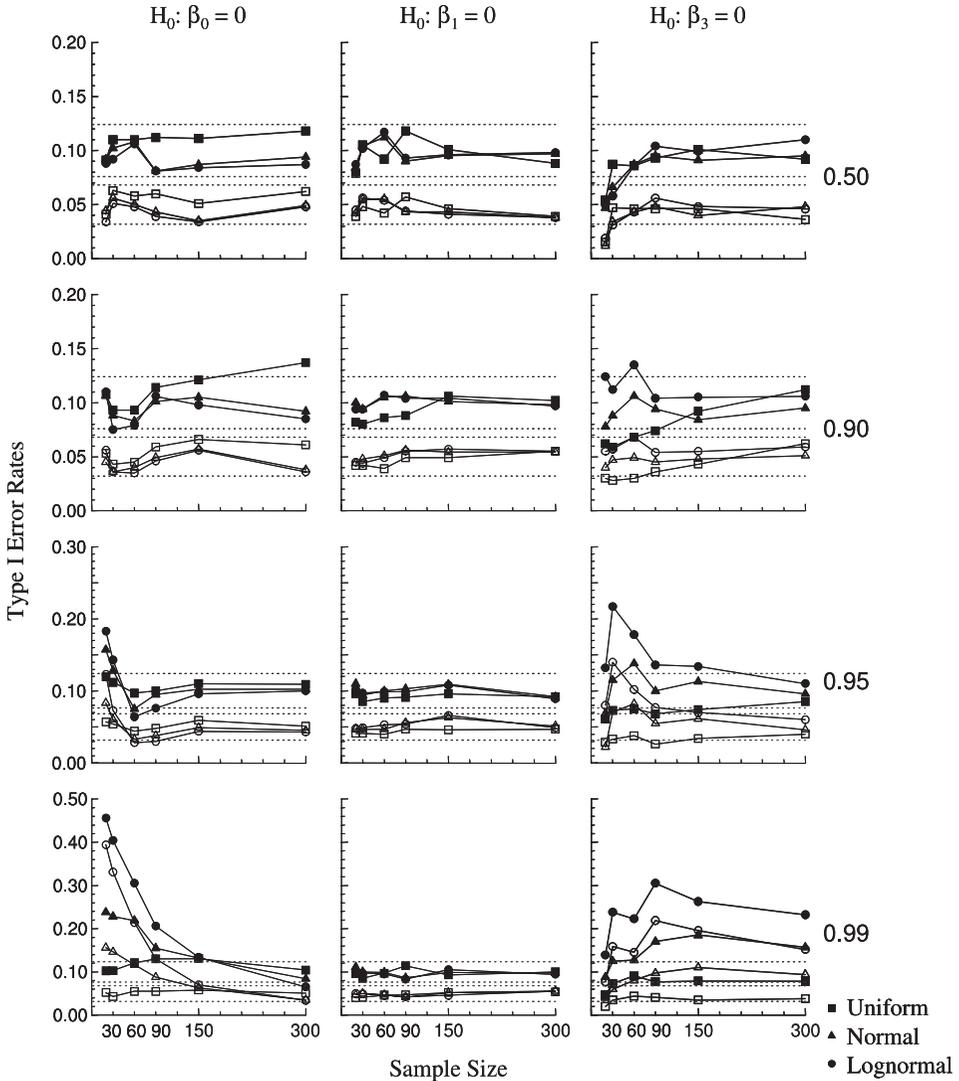


Figure 2. Estimated Type I error rates for $\alpha = 0.05$ (open) and 0.10 (solid); for the permutation D test for homogeneous lognormal (circles), normal (triangles), and uniform (squares) error distributions; for $H_0: \beta_0 = 0$ (double permutation) and $H_0: \beta_1 = 0$ in the model $y = \beta_0 + \beta_1 X_1 + \varepsilon$, and $H_0: \beta_3 = 0$ (drop zero residuals, reduced rank regression) in the model $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon$ for $\tau = \{0.50, 0.90, 0.95, \text{ and } 0.99\}$; and for $n = 20, 30, 60, 90, 150,$ and 300 . Fine dotted lines are 99% binomial confidence intervals around $\alpha = 0.05$ and 0.10 for 1,000 random samples used at each combination of $H_0: n,$ and quantile.

the 0.99 quantile and $n < 150$. Normal and uniform error distributions had less reduction in power across quantiles for this hypothesis, maintaining effective power for all sample sizes up to and including the 0.99 quantile. Power increased with increasing effect size and sample size as desired. Graphs of all power simulations are in the online Appendix.

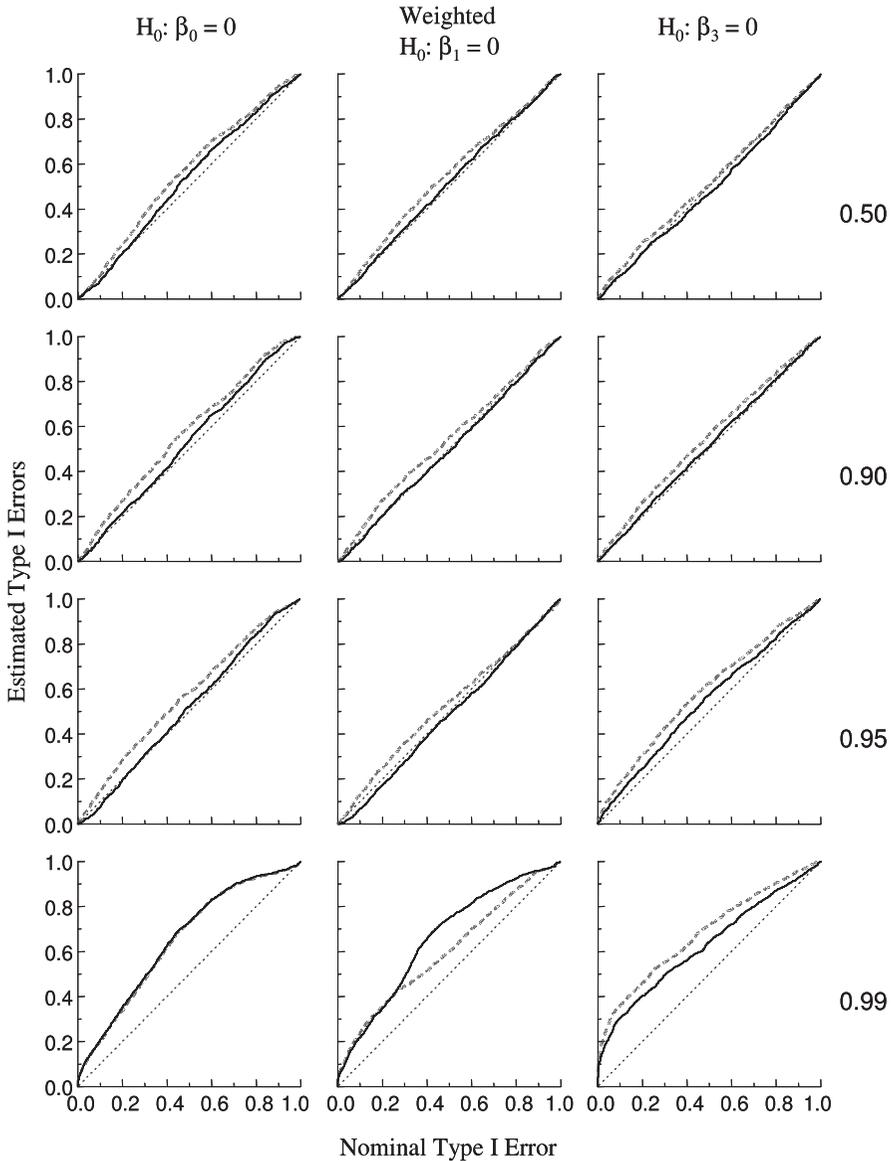


Figure 3. Cumulative distributions of 1,000 estimated Type I errors for permutation approximation of $H_0: \beta_0 = 0$ in the model $y = \beta_0 + \beta_1 X_1 + \varepsilon$; $H_0: \beta_1 = 0$ in the weighted model $wy = w\beta_0 + w\beta_1 X_1 + w(1 + \gamma X_1)\varepsilon$ with $\gamma = 0.05$ and $w = (1 + \gamma X_1)^{-1}$; and for $H_0: \beta_3 = 0$ in the model $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon$; for 0.50, 0.90, 0.95, and 0.99 quantiles; for $n = 90$; and the lognormal error distribution. Cdf's compare the standard approach of permuting residuals from the null model (dashed lines) with double permutation schemes for $H_0: \beta_0 = 0$ and $H_0: \beta_1 = 0$ and deletion of zero residuals and reduced dimension regression for $H_0: \beta_3 = 0$ (solid lines). Fine dotted line is expected cdf of uniform distribution.

4.2 HOMOGENEOUS ERROR STRUCTURE: MULTIPLE REGRESSION

The six-parameter model $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon$ with $\beta_0 = 36.0$, $\beta_1 = 0.10$, $\beta_2 = -0.005$, $\beta_4 = 2.0$, and $\beta_3 = \beta_5 = 0.0$ was evaluated for Type I error rates for $H_0 : \beta_3 = 0$. Here the dimension of \mathbf{e} was reduced to $n - 4$ as there were $p - q = 5$ parameters under the null model. Type I error rates were well maintained for $n > 20$ for 0.50, $n > 60$ for 0.90, and $n > 90$ for 0.95 quantiles (Figure 2). Only the uniform error distribution had Type I error rates near nominal values for the 0.99 quantile (Figure 2). The slightly liberal nature of the standard compared to the reduced zero residuals permutation approach is illustrated for the lognormal error distribution and $n = 90$ (Figure 3). We also evaluated Type I error rates for $H_0 : \beta_4 = 0$ in the six-parameter model with $\beta_0 = 36.0$, $\beta_1 = 0.10$, $\beta_2 = -0.005$, $\beta_3 = 0.05$, and $\beta_4 = \beta_5 = 0.0$ and found a similar pattern. Power was not investigated for multiple regression models with homogeneous errors.

4.3 HETEROGENEOUS ERROR STRUCTURE: SIMPLE REGRESSION

The two-parameter weighted regression model $wy = w\beta_0 + w\beta_1 X_1 + w(1 + \gamma X_1)\varepsilon$ with $\gamma = 0.05$, $\beta_0 = 6.0$, and $\beta_1 = 0.0$ was evaluated for Type I error rates for $H_0 : \beta_1 = 0$ using the known weights $w = (1 + 0.05X_1)^{-1}$. Type I error rates using the double permutation scheme were well maintained for 0.50 to 0.90 quantiles, becoming increasingly liberal for 0.95 to 0.99 quantiles with decreasing sample size for normal and lognormal error distributions (Figure 4). Type I error rates were well maintained for the uniform error distribution at higher quantiles and smaller sample sizes. Examining the cdf's of Type I errors for this hypothesis for the lognormal error distribution and $n = 90$ demonstrated that the double permutation scheme provided better Type I error rates than the standard permutation approach (Figure 3). The null weighted model is implicitly forced through the origin because the column vector of 1's for the intercept have been multiplied by w and, therefore, the residuals no longer have their expected properties when the objective function (2.1) is minimized. Type I error rates for $H_0 : \beta_0 = 0$ with the double permutation scheme also were evaluated for the two-parameter weighted regression model $wy = w\beta_0 + w\beta_1 X_1 + w(1 + \gamma X_1)\varepsilon$ with $\gamma = 0.05$, $\beta_1 = 0.10$, and $\beta_0 = 0.0$ using the known weights $w = (1 + 0.05X_1)^{-1}$ and followed a pattern similar to those for $H_0 : \beta_1 = 0$ (Figure 4).

Power to detect $\beta_1 = 0.01, 0.05, 0.10$, and 0.20 for the weighted regression model with heterogeneous errors decreased with increasing quantile and decreasing sample size more for the lognormal than the normal and uniform error distributions. Power for the lognormal error distribution and the 0.99 quantile was unreliable for $n < 300$ because of excessively liberal Type I error rates associated with the double permutation scheme. Power to detect $\beta_0 = 0.5, 1.0, 2.0$, and 3.0 followed a similar decline with increasing quantile and decreasing samples size as for homogeneous error distributions, becoming almost nonexistent for the 0.99 quantile of the lognormal error distribution with $n = 300$. Uniform and normal error distributions had effective power for the 0.99 quantile at smaller n . Again, power increased with increasing effect size and sample size as desired. Power graphs are in the online Appendix.

4.4 HETEROGENEOUS ERROR STRUCTURE: MULTIPLE REGRESSION

The six-parameter model $wy = w(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + (1 + \gamma X_1)\varepsilon)$ with $\gamma = 0.05$, $\beta_0 = 36.0$, $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$, and known weights $w = (1 + 0.05X_1)^{-1}$ was evaluated for Type I error rates for the full model

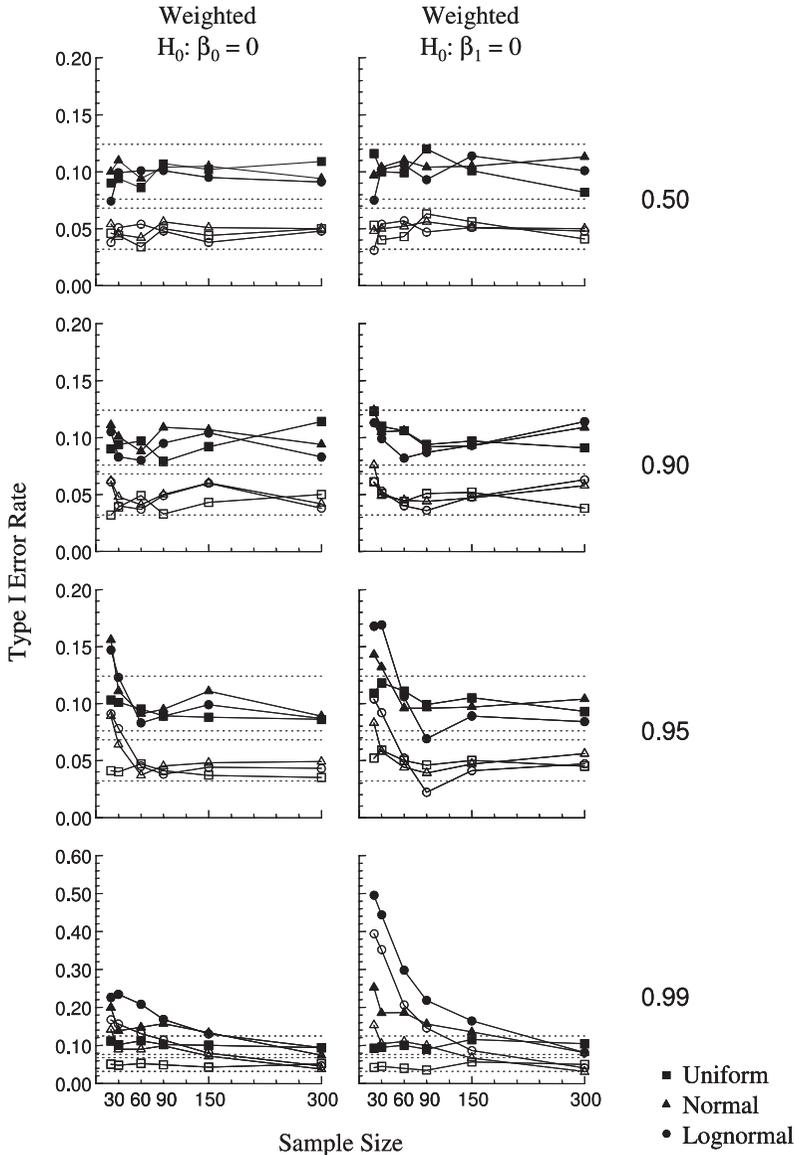


Figure 4. Estimated Type I error rates for $\alpha = 0.05$ (open) and 0.10 (solid); for the double permutation D test for $H_0 : \beta_0 = 0$ and $H_0 : \beta_1 = 0$; for the weighted model $wy = w(\beta_0 + \beta_1 X_1 + (1 + \gamma X_1)\varepsilon)$, $\gamma = 0.05$, and $w = (1 + \gamma X_1)^{-1}$; with lognormal (circles), normal (triangles), and uniform (squares) error distributions; for $\tau = \{0.50, 0.90, 0.95, \text{ and } 0.99\}$; and for $n = 20, 30, 60, 90, 150, \text{ and } 300$. Fine dotted lines are 99% binomial confidence intervals around $\alpha = 0.05$ and 0.10 for 1,000 random samples used at each combination of error distribution, H_0 , n and quantile.

hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$. Type I error rates using the double permutation procedure were well maintained until $n < 60$ for 0.95 and $n < 300$ for 0.99 quantiles, where liberal error rates occurred for the lognormal distribution (Figure 5). Type I error rates were better maintained at smaller n and more extreme quantiles for the

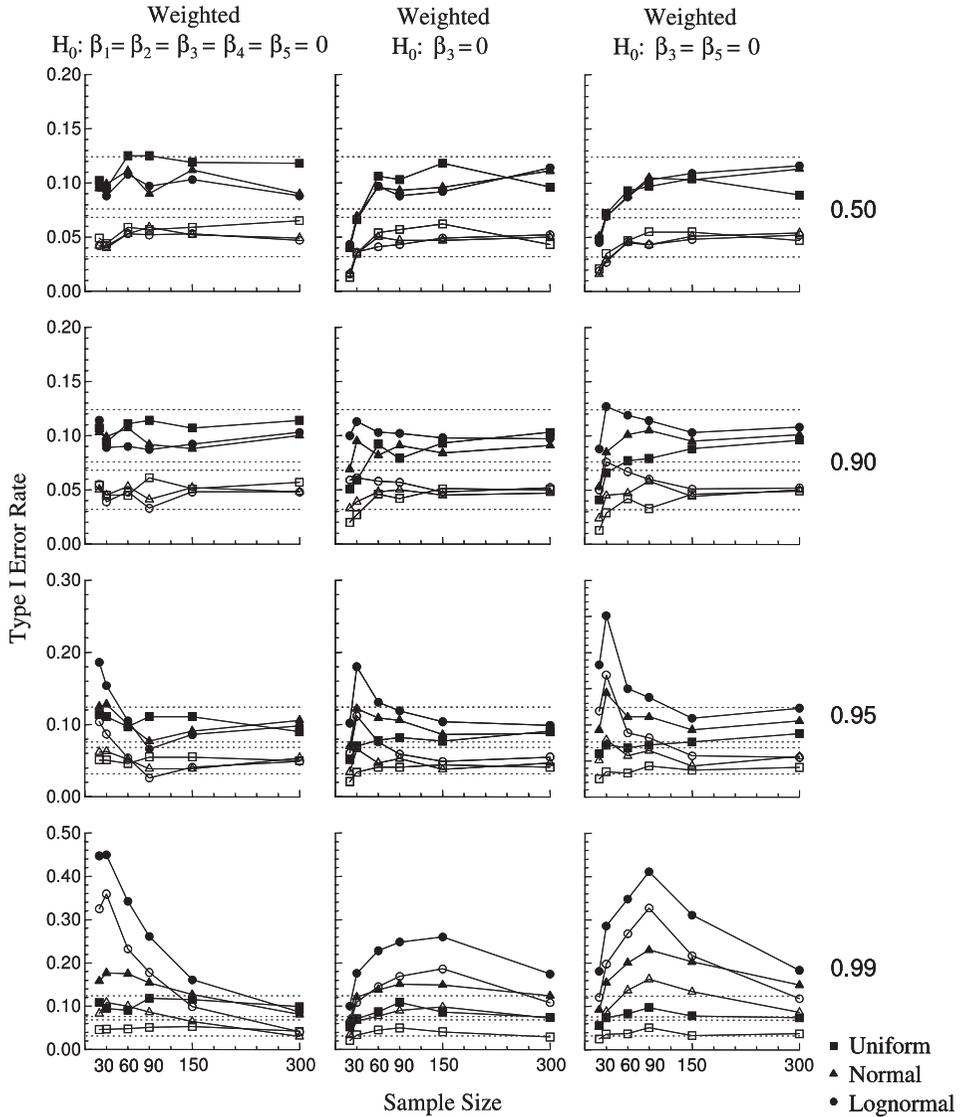


Figure 5. Estimate Type I error rates for $\alpha = 0.05$ (open) and 0.10 (solid); for the permutation D test for $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$ (double permutation), $H_0 : \beta_3 = 0$, and $H_0 : \beta_3 = \beta_5 = 0$ (delete zero residuals, reduced dimension regression); for heterogeneous lognormal (circles), normal (triangles), and uniform (squares) error distributions in the weighted model $wy = w(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + (1 + \gamma X_1)\epsilon)$, $\gamma = 0.05$, $w = (1 + \gamma X_1)^{-1}$; for $\tau = \{0.50, 0.90, 0.95, \text{ and } 0.99\}$; and for $n = 20, 30, 60, 90, 150, \text{ and } 300$. Fine dotted lines are 99% binomial confidence intervals around $\alpha = 0.05$ and 0.10 for 1,000 random samples used at each combination of error distribution, H_0 , n , and quantile.

normal and uniform error distributions. Power estimated with 1 of the 5 slope parameters ($\beta_3 = 0.10, 0.15, 0.20,$ and 0.25) allowed to be nonzero was low to nonexistent for the 0.99 quantile, and increased with increasing effect size and sample size for other quantiles. Power for this and other hypotheses for the multiple regression models was evaluated only for the lognormal error distribution to reduce the amount of computing and reporting (Appendix). Power for normal and uniform error distributions would be greater than or equal to that for the lognormal error distribution for the higher quantiles.

Type I error rates for subhypotheses involving continuous variables in the six-parameter weighted model with $\beta_0 = 36.0, \beta_1 = 0.10, \beta_2 = -0.005, \beta_4 = 2.0,$ and $\beta_3 = \beta_5 = 0.0$ were evaluated for $H_0 : \beta_3 = 0$ and $H_0 : \beta_3 = \beta_5 = 0$. Type I error rates were well maintained by the reduced zero residuals permutation procedure for 0.50 to 0.95 quantiles and $n > 30$, becoming liberal at smaller n and more extreme quantiles for the lognormal and normal error distributions, and more so for $H_0 : \beta_3 = \beta_5 = 0$ than for $H_0 : \beta_3 = 0$ (Figure 5). Only the uniform error distribution had reasonable Type I error rates for the 0.99 quantile. Power for $H_0 : \beta_3 = 0$ was estimated with $\beta_3 = 0.10, 0.15, 0.20,$ and 0.25 for the lognormal error distribution. Power was low for the 0.90 and 0.95 to nonexistent for the 0.99 quantiles but otherwise increased with increasing effect size as desired (Appendix). Subhypotheses involving categorical predictors in the six-parameter weighted model with $\beta_0 = 36.0, \beta_1 = 0.10, \beta_2 = -0.005, \beta_3 = 0.05,$ and $\beta_4 = \beta_5 = 0.0$ were evaluated for Type I error rates for $H_0 : \beta_4 = 0$ and $H_0 : \beta_4 = \beta_5 = 0$ (Appendix) and had similar patterns as for subhypotheses for continuous predictors evaluated above. Power was evaluated for the subhypothesis $H_0 : \beta_4 = 0$ for $\beta_4 = 1.5, 3.0, 6.0,$ and 12.0 and the lognormal error distribution. Power declined with increasing quantiles and decreasing sample size, but increased with increasing effect size (Appendix).

4.5 PERFORMANCE RELATIVE TO RANK SCORE TESTS

We compared the ability of the drop in dispersion D test to maintain valid Type I error rates relative to the conventional rank score T test (Cade et al. 2006) across the same set of simulation conditions. The conventional rank score T test has its probability evaluated under the null hypothesis by referencing the observed test statistic to a chi-square distribution with q df. In Figure 6, we bound the sample size (n) and quantile (τ) space where each test maintained correct Type I error rates by hypothesis tested and number of model parameters (p) for all three error distributions. The rank score T test provided valid Type I error rates for tests of the intercept ($H_0 : \beta_0 = 0$) at more extreme quantiles and smaller samples ($n < 300$) for both homogeneous and heterogeneous error models (Figure 6). The rank score T test also provided valid Type I error rates at smaller samples ($n < 60$) for all quantiles when testing subhypotheses in multiple regression models (Figure 6). However, in the weighted heterogeneous error models, the D test provided valid Type I error rates at more extreme quantiles for moderate sample sizes ($60 < n < 300$). The D test also provided valid Type I error rates at more extreme quantiles and smaller samples ($n < 300$) for testing slopes ($H_0 : \beta_1 = 0$) in simple regression models with homogeneous errors and

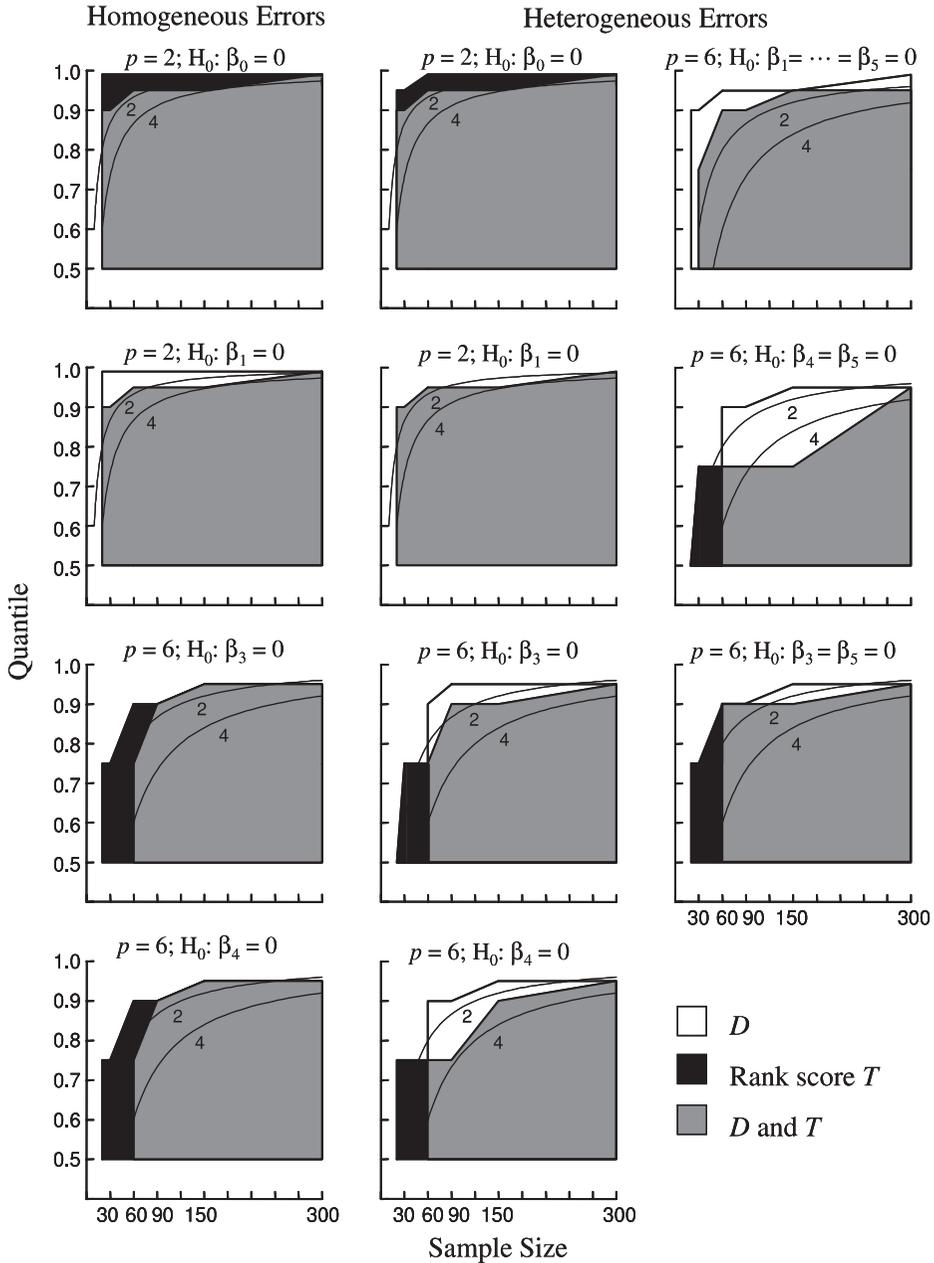


Figure 6. Sample sizes and quantiles providing valid Type I errors and power $> \alpha = 0.05$ for at least one nonzero effect for the permutation D test and rank score T test based on simulations for $n = 20, 30, 60, 90, 150,$ and $300; \tau = \{0.50, 0.75, 0.90, 0.95, 0.99\}$; for lognormal, normal, and uniform homogeneous and heterogeneous error distributions; for $p = 2$ and 6 parameter models; and hypotheses in Figures 2–5. Black shaded regions are where only T rank score test was valid, white regions are where only permutation D test was valid, and gray shaded regions are where both were valid. Thin contour lines are effective ranks of 2 and 4, where effective rank $= ((1 - \tau) \times n) \div p$ for $\tau \geq 0.50$.

when testing all slopes ($H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$) in multiple regression models with heterogeneous errors.

We also graphed functions for the effective rank of the quantile regression models, where effective rank = $((1-\tau) \times n) \div p$ for $\tau \geq 0.50$ summarizes the parameter dimensions as a ratio of order statistic to number of parameters for comparing inference procedures for quantile regression (Chernozhukov and Umantsev 2001). Both the D test and the T rank score test always provided valid Type I error rates for effective ranks ≥ 4 , commonly for effective ranks 2–4, and occasionally for effective ranks < 2 (Figure 6). Thus, both tests provided valid inferences for smaller effective ranks than the ≥ 25 recommended for asymptotic methods based on the variance/covariance matrix (Chernozhukov and Umantsev 2001).

We compared relative power of the D test to the T rank score test by effective rank for those parameter regions in Figure 6 where both tests maintained valid Type I error rates (Figure 7). As effective ranks reached and exceeded 10, relative power approached 1 for all hypotheses. For hypotheses on the intercept, relative power often was < 1 and decreased with decreasing effective ranks especially for lognormal error distributions and smaller n . For hypotheses on other parameters, relative power was ≥ 1 and increased as effective ranks decreased. For effect sizes not included in Figure 7, relative power approached 1 as effect sizes increased away from the null hypothesized value of zero and power of each test approached 1. Relative power increased as effect sizes were closer to the null hypothesized value of zero and power of each test was further from 1. Because power of the permutation version of the F rank score test was similar to the chi-square T rank score test (Cade et al. 2006), the relative power of D to the permutation F rank score test was similar to relative power of D to T . However, like the D test, the permutation F rank score test provides better Type I error rates at smaller sample sizes for more extreme quantiles for parameters other than the intercept (Cade et al. 2006).

5. EXAMPLE APPLICATION

The initial analyses of Lahontan cutthroat trout standing crop as related to stream channel morphology suggested the nonlinear model $y = \exp(\beta_0 + \beta_1 X_1 + \varepsilon)$, where y is trout m^{-1} and X_1 is width:depth ratio, was a reasonable approximation (Dunham et al. 2002). We estimated the weighted linear form $w(\ln y) = w\beta_0 + w\beta_1 X_1 + w(\gamma_0 - \gamma_1 X_1)\varepsilon$, and estimates for selected regression quantiles were plotted by exponentiating to back transform to the nonlinear form (Figure 1, p. 108). The vector of weights w we used were identical to those used with quantile rank score tests (Cade et al. 2006) and the weighted least squares estimate. Weights $w = (1.310 - 0.017X_1)^{-1}$ were estimated by computing the average pairwise differences between the 76 unweighted regression quantile $\tau = [0, 1]$ estimates, $b_0(\tau)$ to estimate γ_0 and $b_1(\tau)$ to estimate γ_1 in the standard deviation function $\gamma_0 - \gamma_1 X_1$. Estimates of parameters for all quantiles were plotted as a step function with 90% confidence intervals for 19 quantiles between 0.05 and 0.95 by increments of 0.05 (Figure 1(b) and 1(c)). Starting values for the manual iteration of the test inversion were based on interval endpoints estimated by the rank score tests. These values were then used

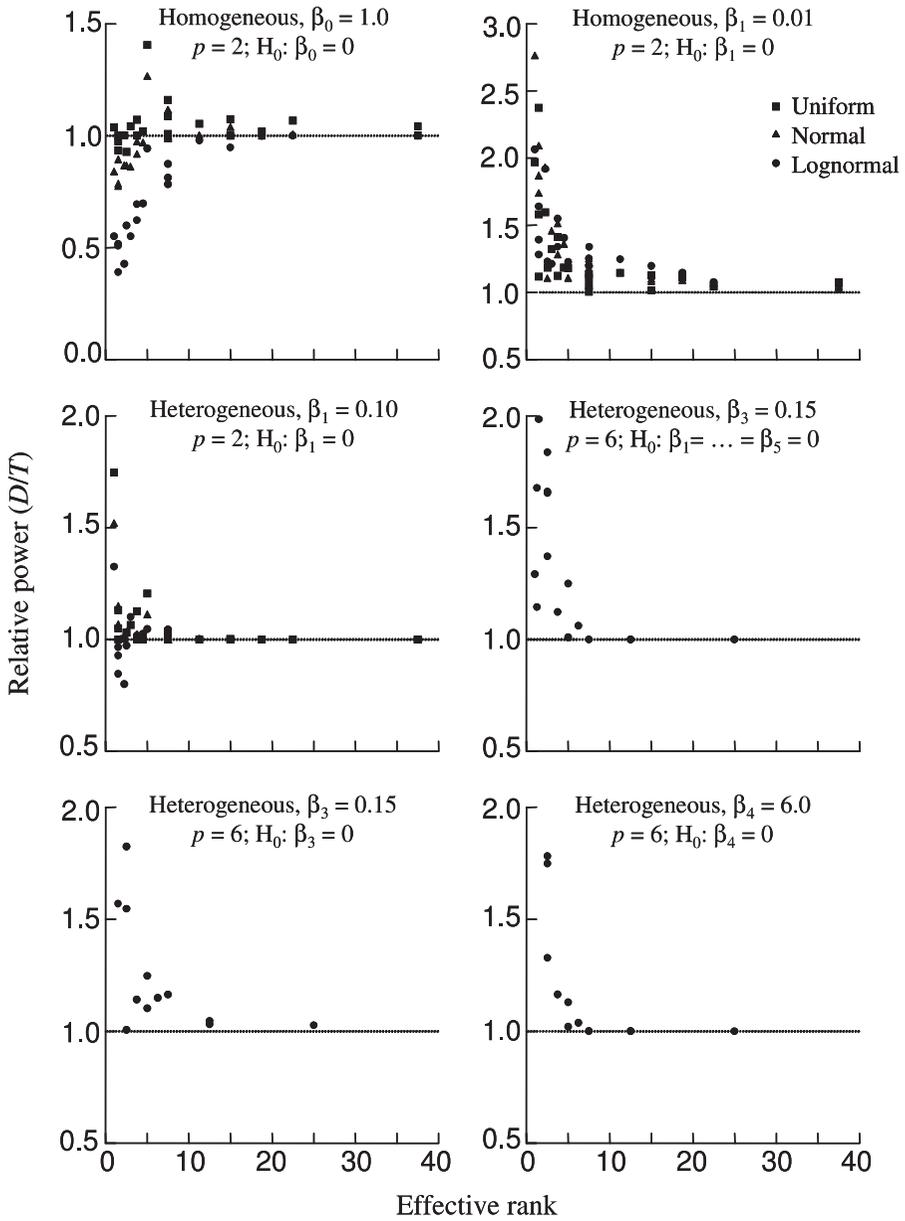


Figure 7. Relative power of the permutation D test compared to T rank score test by effective rank, where effective rank = $((1 - \tau) \times n) \div p$ for $\tau \geq 0.50$, for sample sizes, quantiles, and hypotheses where both tests maintained valid Type I errors in Figure 6. Only one effect size is graphed per hypothesis. Power was estimated for lognormal (circles), normal (triangles), and uniform (squares) error distributions for simple regression models but only for the lognormal error distribution for multiple regression models. Relative power was always near 1 for effective ranks > 40, which were not graphed.

as hypothesized parameter values of $\xi(\tau)$ in the transformation $\mathbf{y} - \mathbf{X}_2\xi(\tau)$ to test the $H_0 : \beta_2(\tau) = \xi(\tau)$ with (3), where β_2 was either β_0 or β_1 . We used $m + 1 = 100,000$ permutations to compute probabilities for the D tests associated with confidence interval endpoints.

The 90% confidence intervals estimated by inverting the D test (Figure 1) were smoother across quantiles than those based on the rank score tests (Cade et al. 2006) because the permutation distribution for the D test was much more continuous than the distribution of the rank score test. Linear interpolation between hypothesized parameter values was not required to achieve $P = \alpha = 0.10$ with the D test as it was for the rank score tests (Koenker 1994; Cade et al. 2006). The D test based 90% confidence intervals for β_1 were 47 to 143% (median = 93%) of the length of those based on the T rank score test (Cade et al. 2006), shorter for 13 of 19 quantiles with greatest reduction in length for lower quantiles. Similarly, the D test based 90% confidence intervals for β_0 were 64 to 118% (median = 96%) of the length of those based on the T rank score tests, shorter for 11 of 19 quantiles with greatest reduction in length also for lower quantiles. Differences between the D test and the quantile rank score based intervals were not sufficient to alter any conclusions about the effects of width:depth ratios on cutthroat trout populations. Confidence bands estimated by the D test supported an interpretation that increasing stream width:depth ratios from 15 to 45 decreased the highest 20% of trout densities ($\tau \geq 0.80$) by 1 to 64% [$\exp(-0.0003 \times 30) = 0.991$ and $\exp(-0.0344 \times 30) = 0.356$], similar to conclusions based on the quantile rank score confidence intervals (Dunham et al. 2002; Cade et al. 2006). There was weak evidence that lowest quantiles ($\tau \leq 0.10$) of trout density increased slightly with stream width:depth ratio, but most quantiles ($0.10 < \tau < 0.80$) of trout density did not change as indicated by confidence intervals including zero (Figure 1).

The interpretation of this pattern was that physical habitat as measured by width:depth ratio was a constraint limiting trout abundance only for some of the observations and other factors were limiting at most locations and times (Dunham et al. 2002). To explore additional factors that might be limiting trout abundance, Dunham et al. (2002) also considered a model that included an indicator variable (X_2) assigning a 0 for streams with nonnative brook trout (*Salvelinus fontinalis*), a possible competitor, or connected to migratory habitats and assigning a 1 for isolated streams without brook trout. We considered a weighted model, $w(\ln y) = w\beta_0 + w\beta_1X_1 + w\beta_2X_2 + w\beta_3X_1X_2 + w(\gamma_0 - \gamma_1X_1)\epsilon$, similar to the unweighted model used by Dunham et al. (2002) to examine whether the data supported separate slopes [$H_0 : \beta_3(\tau) = 0$] for the two groups of streams. Our weighted estimates and the Dunham et al. (2002) unweighted estimates of $\beta_3(\tau)$ were -0.029 to -0.017 for lower quantiles ($\tau < 0.55$), decreasing to -0.044 to -0.035 for higher quantiles ($\tau \geq 0.55$), indicating a stronger negative relationship between Lahontan cutthroat trout abundance and width:depth ratio in isolated streams without brook trout. However, because the 90% confidence intervals for $\beta_3(\tau)$ formed by the rank score test inversion overlapped zero slightly for all but the highest quantiles ($\tau \geq 0.90$), Dunham et al. (2002) concluded there was little evidence supporting the separate slopes model. We estimated 90% confidence intervals for $\beta_3(\tau)$ by inverting the D test and found intervals for $\tau \geq 0.55$ excluded zero, whereas intervals

for $\tau < 0.55$ included zero. The greater power of the D test provided more evidence of a stronger negative relationship between higher quantiles of Lahontan cutthroat trout abundance and width:depth ratio for isolated streams without brook trout.

6. DISCUSSION

Our modifications of the standard approach of permuting residuals from the null model were successful at providing valid Type I error rates when quantile regression models were forced through the origin or had a mass of zero residuals due to multiple parameters. However, using the double permutation scheme for estimates forced through the origin resulted in some loss of power at smaller n and more extreme quantiles compared to the chi-square quantile rank score test (Cade et al. 2006). A similar problem was noted when the double permutation scheme was used to evaluate the rank score test (Cade et al. 2006). Computing the random binomial proportion associated with the double permutation procedure results in many identical values of 1.0 for extreme quantiles with small n . The double permutation procedure will require n much greater than 300 to work reasonably with quantiles ≥ 0.99 .

The deletion of zero residuals and reduced dimension quantile regression used when null models had multiple parameters provided slightly more conservative Type I error rates for the D test than either permutation or Chi-square evaluations of the rank score statistic at smaller n (< 60). This was attributed to the reduction in sample size of the permuted residuals required by the D test when testing subhypotheses. However, power often was substantially greater for the D test compared to rank score test of subhypotheses when $n \geq 60$. The only hypothesis where rank score tests consistently had better power than the D test was when testing the intercept.

Our approach of randomly deleting rows of the design matrix \mathbf{X} to make it conform to the reduced dimension of the residuals \mathbf{e} is only one of several possible approaches. It also would be possible to just delete the rows of \mathbf{X} corresponding to those with the deleted zero residuals. The dimension of \mathbf{X} also could be reduced by Gram-Schmidt orthogonalization (Commenges 2003). It remains to be determined if either of these or other alternatives provide any advantage.

The drop in dispersion D test appears to be a useful addition to the growing arsenal of inference procedures for linear quantile regression analyses with small to moderate sample sizes. Despite the benefits of greater power and reduced interval lengths, inverting the D test to estimate confidence intervals is a computationally burdensome procedure compared to estimating confidence intervals by inverting the chi-square T rank score test (Koenker 1994). Routines that iterate towards selected confidence intervals based on inverting the D test need to be developed with some form of user guided starting values and increments to step towards a solution.

Our simulation experiment avoided the issue of how to estimate weights for heteroscedastic models by using the known standard deviation function. In applications, this function is not known and the weights must be estimated. We used a simple pairwise difference approach based on the initial unweighted estimates for estimating weights in our example application. Other approaches for estimating weights include regressing absolute

values of residuals from an unweighted fit of the 0.50 quantile on the independent variables for linear location-scale models (Zhou and Portnoy 1998) and the sparsity estimation approach for more general heteroscedastic models (Koenker and Machado 1999).

APPENDIX

An online Appendix available at www.fort.usgs.gov/staff/staffprofile.asp?StaffID=115 has graphs for additional Type I error and power simulations.

ACKNOWLEDGMENTS

M. J. Anderson, P. Good, R. Koenker, P. W. Mielke, Jr., and J. E. Roelle reviewed earlier drafts of the article. We thank J. B. Dunham for providing the Lahontan cutthroat trout data.

[Received November 2004. Revised August 2005.]

REFERENCES

- Allen, A. W., Cade, B. S., and Vandever, M. W. (2001), "Effects of Emergency Haying on Vegetative Characteristics Within Selected Conservation Reserve Program Fields in the Northern Great Plains," *Journal of Soil and Water Conservation*, 56, 120–125.
- Anderson, M. J., and Legendre, P. (1999), "An Empirical Comparison of Permutation Methods for Tests of Partial Regression Coefficients in a Linear Model," *Journal of Statistical Computation and Simulation*, 62, 271–303.
- Anderson, M. J., and Robinson, J. (2001), "Permutation Tests for Linear Models," *Australian and New Zealand Journal of Statistics*, 43, 75–88.
- Barrodale, I., and Roberts, F. D. K. (1974), "Algorithm 478: Solution of an Overdetermined System of Equations in the l_1 Norm," *Communications of the Association for Computing Machinery*, 17, 319–320.
- Buchinsky, M (1991), "The Theory and Practice of Quantile Regression," unpublished Ph.D dissertation, Harvard University, Cambridge MA.
- Cade, B. S., and Guo, Q. (2000), "Estimating Effects of Constraints on Plant Performance with Regression Quantiles," *Oikos*, 91, 245–254.
- Cade, B. S., and Noon, B. R. (2003), "A Gentle Introduction to Quantile Regression for Ecologists," *Frontiers in Ecology and the Environment*, 1, 412–420.
- Cade, B. S., and Richards, J. D. (1996), "Permutation Tests for Least Absolute Deviation Regression," *Biometrics*, 52, 886–902.
- Cade, B. S., Terrell, J. W., and Schroeder, R. L. (1999), "Estimating Effects of Limiting Factors with Regression Quantiles," *Ecology*, 80, 311–323.
- Cade, B. S., Noon, B. R., and Flather, C. H. (2005), "Quantile Regression Reveals Hidden Bias and Uncertainty in Habitat Models," *Ecology*, 86, 786–800.
- Cade, B. S., Richards, J. D., and Mielke, P. W., Jr. (2006), "Rank Score and Permutation Testing Alternatives for Regression Quantile Estimates," *Journal of Statistical Computation and Simulation*, 76, 331–355.
- Chernozhukov, V., and Umantsev, L. (2001), "Conditional Value-at-Risk: Aspects of Modeling and Estimation," *Empirical Economics*, 26, 271–292.
- Commenges, D. (2003), "Transformations Which Preserve Exchangeability and Application to Permutation Tests," *Journal of Nonparametric Statistics*, 15, 171–185.
- Dunham, J. B., Cade, B. S., and Terrell, J. W. (2002), "Influences of Spatial and Temporal Variation on Fish-Habitat Relationships Defined by Regression Quantiles," *Transactions of the American Fisheries Society*, 131, 86–98.

- Eastwood, P. D., Meaden, G. J., and Grieco, A. (2001), "Modeling Spatial Variations in Spawning Habitat Suitability for the Sole *Solea solea* using Regression Quantiles and GIS Procedures," *Marine Ecology Progress Series*, 224, 251–266.
- Freedman, D., and Lane, D. (1983), "A Nonstochastic Interpretation of Reported Significance Levels," *Journal of Business and Economic Statistics*, 1, 292–298.
- Good, P. (2002), "Extensions of the Concept of Exchangeability and Their Applications," *Journal of Modern Applied Statistical Methods*, 1, 243–247.
- Gutenbrunner, C., and Jurečková, J. (1992), "Regression Rank Scores and Regression Quantiles," *Annals of Statistics*, 20, 305–330.
- Gutenbrunner, C., Jurečková, J., Koenker, R., and Portnoy, S. (1993), "Tests of Linear Hypotheses Based on Regression Rank Scores," *Nonparametric Statistics*, 2, 307–331.
- Haire, S. L., Bock, C. E., Cade, B. S., and Bennett, B. S. (2000), "The Role of Landscape and Habitat Characteristics in Limiting Abundance of Grassland Nesting Songbirds in an Urban Open Space," *Landscape and Urban Planning*, 48, 65–82.
- Huston, M. A. (2002), "Introductory Essay: Critical Issues for Improving Predictions," in *Predicting Species Occurrences: Issues of Accuracy and Scale*, eds. J. M. Scott, P. J. Heglund, and M. L. Morrison, Covelo, CA: Island Press, pp. 7–21.
- Kennedy, P. E., and Cade, B. S. (1996), "Randomization Tests for Multiple Regression," *Communications in Statistics—Simulation and Computation*, 25, 923–936.
- Koenker, R. (1987), "A Comparison of Asymptotic Testing Methods for L_1 -Regression," in *Statistical Data Analysis Based on the L_1 -norm and Related Methods*, ed. Y. Dodge, Amsterdam: Elsevier Science Publishers, pp. 287–295.
- (1994), "Confidence Intervals for Regression Quantiles," in *Asymptotic Statistics: Proceedings of the 5th Prague Symposium*, eds. P. Mandl and M. Hušková, Berlin: Physica-Verlag, pp. 349–359.
- (1997), "Rank Tests for Linear Models," in *Handbook of Statistics* (vol. 15), eds. G. S. Maddala and C. R. Rao, Amsterdam: North Holland, pp. 175–199.
- Koenker, R., and Bassett, G. (1978), "Regression Quantiles," *Econometrica*, 46, 33–50.
- (1982), "Robust Tests for Heteroscedasticity Based on Regression Quantiles," *Econometrica*, 50, 43–61.
- Koenker, R., and d'Orey, V. (1987), "Computing Regression Quantiles," *Applied Statistics*, 36, 383–393.
- (1994), "A Remark on Algorithm AS229: Computing Dual Regression Quantiles and Regression Rank Scores," *Applied Statistics*, 43, 410–414.
- Koenker, R., and 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., and Portnoy, S. (1996), "Quantile Regression," University of Illinois at Urbana-Champaign, College of Commerce and Business Administration, Office of Research Working Paper 97-0100.
- Koenker, R., and Zhao, Q. (1994), "L-Estimation for Linear Heteroscedastic Models," *Journal of Nonparametric Statistics*, 3, 223–235.
- Legendre, P. (2000), "Comparison of Permutation Methods for the Partial Correlation and Partial Mantel Tests," *Journal of Statistical Computation and Simulation*, 67, 37–73.
- Legendre, P., and Desdevises, Y. (2002), "Independent Contrasts and Regression Through the Origin," available online at www.fas.umontreal.ca/BIOL/legendre/reprints/index.html.
- Randles, R. H. (1984), "On Tests Applied to Residuals," *Journal of the American Statistical Association*, 79, 349–354.
- Scharf, F. S., Juanes, F., and Sutherland, M. (1998), "Inferring Ecological Relationships from the Edges of Scatter Diagrams: Comparison of Regression Techniques," *Ecology*, 79, 448–460.
- Terrell, J. W., Cade, B. S., Carpenter, J., and Thompson, J. M. (1996), "Modeling Stream Fish Habitat Limitations from Wedged-Shaped Patterns of Variation in Standing Stock," *Transactions of the American Fisheries Society*, 125, 104–117.
- Zhou, K. G., and Portnoy, S. L. (1998), "Statistical Inference on Heteroscedastic Models Based on Regression Quantiles," *Journal of Nonparametric Statistics*, 9, 239–260.