Improving linear quantile regression for replicated data

Kaushik Jana\textsuperscript{1} and Debasis Sengupta\textsuperscript{2}

\textsuperscript{1} Imperial College London, UK
\textsuperscript{2}Indian Statistical Institute, Kolkata, India

Abstract

When there are few distinct values of the covariates but many replicates, we show that a weighted least squares fit to the sample quantiles of the replicates is asymptotically more efficient than the usual method of linear quantile regression.

Keywords: Asymptotic efficiency; Conditional quantile; Weighted least squares; Löwner order

1 Introduction

Consider a quantile regression problem with a handful of distinct values of covariates, where each covariate profile is replicated many times. A linear regression model for the quantiles are often preferred for such data. If one ignores the fact of replications, the linear quantile regression estimator of Koenker and Bassett (1978) can be used for estimating the parameters and related inference. However, the replicated nature of the data
enables one to fit a linear (mean) regression model to the conditional sample quantiles for each value of covariates. Since these conditional sample quantiles would in general have different variances, a weighted least squares (WLS) estimator with weights inversely proportional to the estimated variances of the respective conditional sample quantiles may be used. Many researchers, apparently oblivious to this common-sense option, have used the method of Koenker and Bassett (1978) for linear quantile regression with replicated data (Redden et al., 2004; Fernández et al., 2004; Elsner et al., 2008; Jagger and Elsner, 2009; Kossin et al., 2013). Before this trend continues further, it would be interesting to study how the two methods compare.

We show in this paper that the WLS estimator is asymptotically more efficient than the estimator of Koenker and Bassett (1978). Small sample simulation are conducted to chart the domain of this dominance relation, and an illustrative data analysis is carried out to demonstrate the gains made.

2 Comparison of asymptotic variances

Suppose the $\tau$-quantile of the conditional distribution of a random variable $Y$ given another random vector $\mathbf{x}$ is $q_Y(\tau|\mathbf{x}) := \inf \{ q : P(Y \leq q|\mathbf{x}) \geq \tau \}$. For a given $\tau \in [0, 1]$, consider the linear regression model (Koenker, 2005)

$$q_Y(\tau|\mathbf{x}) = \mathbf{x}'\beta(\tau),$$

(1)

where $\mathbf{x}$ is the vector of regressors (along with intercept) and $\beta(\tau)$ is the vector of corresponding regression coefficients. Consider independent sets of data of the form $(\mathbf{x}_i, Y_{ij})$ with $j = 1, \ldots, n_i$, $i = 1, \ldots, k$, such that for given $\mathbf{x}_i$, the $Y_{ij}$s are conditionally iid with
common distribution $F_i$. The sample $\tau$-quantile for given $x_i$ is

$$
\hat{q}_i(\tau) = \arg \min_m \sum_{j=1}^{n_i} \rho_\tau(Y_{ij} - m_i), \quad i = 1, \ldots, k,
$$

(2)

where $\rho_\tau(u) = u(\tau - I(u < 0))$. We assume that the distribution $F_i$ has continuous Lebesgue density, $f_i$, with $f_i(u) > 0$ on $\{u : 0 < F_i(u) < 1\}$, for $i = 1, \ldots, k$. The limiting distribution of $\hat{q}_i(\tau)$ has mean $q_Y(\tau|x_i)$ and variance given by (Shorack and Wellner, 2009)

$$
\sigma_i^2(\tau) = \frac{\tau(1 - \tau)}{n_i f_i^2(F_i^{-1}(\tau))}, \quad i = 1, \ldots, k.
$$

(3)

Linear regression of $\hat{q}_i(\tau)$ on $x_i$, with $\sigma_i^{-2}(\tau)$ as weights, produces WLS estimator of $\beta(\tau)$

$$
\tilde{\beta}_{wls}(\tau) = (X' \Omega_i^{-1} X)^{-1} X' \Omega_i^{-1} \tilde{q}(\tau)
$$

(4)

where $X = (x_1 : \ldots : x_k)'$, for $i = 1, \ldots, k$, $\tilde{q}(\tau) = (\hat{q}_1(\tau), \ldots, \hat{q}_k(\tau))'$ and $\Omega_i$ is a diagonal matrix with $\sigma_i^2(\tau), \ldots, \sigma_k^2(\tau)$ as diagonal elements, which have to be replaced by consistent estimates.

The estimator proposed by Koenker and Bassett (1978) is

$$
\tilde{\beta}_{kb}(\tau) = \arg \min_{\beta \in \mathbb{R}^2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \rho_\tau(Y_{ij} - x_i' \beta(\tau)).
$$

(5)

This estimator (the KB estimator) works even if $n_i = 1$ for some or all $i$.

In order to show that (4) is asymptotically more efficient than (5), we need the following regularity conditions.

**Condition A1.** For some vector $(\xi_1, \xi_2, \ldots, \xi_k)^T$ with positive components,

$$
\left( \frac{n_1}{n}, \frac{n_2}{n}, \ldots, \frac{n_k}{n} \right)^T \to (\xi_1, \xi_2, \ldots, \xi_k)^T
$$

(6)
in Euclidean norm, as \( n = \sum_{i=1}^{k} n_i \to \infty \).

**Condition A2.** The distribution functions \( F_i \) are absolutely continuous, with continuous density \( f_i \) uniformly bounded away from 0 and \( \infty \) at \( F_i^{-1}(\tau) \).

**Condition A3.** \( \max_{i=1,\ldots,k} ||X_i||/\sqrt{n} \to 0 \) as \( n \to \infty \). Further, the sample matrices \( D_{0n} = n^{-1} \sum_{i=1}^{k} n_i X_i X_i^T \), \( D_{1n} = n^{-1} \sum_{i=1}^{k} n_i f_i(F_i^{-1}(\tau))X_i X_i^T \) and \( D_{2n} = n^{-1} \sum_{i=1}^{k} n_i f_i^2(F_i^{-1}(\tau))X_i X_i^T \) converge to positive definite matrices \( D_0 \), \( D_1 \) and \( D_2 \), respectively, as \( n \to \infty \).

**Theorem 1:** Under Conditions A1, A2 and A3, and assuming the \( \Omega_{\tau} \) in (4) is replaced by a consistent estimator,

(a) \( \sqrt{n}(\hat{\beta}_{kb}(\tau) - \beta(\tau)) \to \mathcal{N}(0, \tau(1-\tau)D_1^{-1}D_0D_1^{-1}) \),

(b) \( \sqrt{n}(\hat{\beta}_{wls}(\tau) - \beta(\tau)) \to \mathcal{N}(0, \tau(1-\tau)D_2^{-1}) \),

(c) the limiting dispersion matrix of \( \sqrt{n}(\hat{\beta}_{kb}(\tau) - \beta(\tau)) \) is larger than or equals to that of \( \sqrt{n}(\hat{\beta}_{wls}(\tau) - \beta(\tau)) \) in the sense of the Löwner order \(^1\).

**Proof:** The result of part (a) follows from (Koenker (2005), page 121). Part (b) follows from the fact that the WLS estimator is a linear function of the conditional sample quantiles \( \hat{q}_i(\tau), \ i = 1, \ldots, k \), whose limiting distribution under the given conditions are well known (Shorack and Wellner, 2009). The continuous mapping theorem ensures that a consistent estimator of \( \Omega_{\tau} \) would be an adequate substitute for it.

Note that the asymptotic dispersion matrices of \( \sqrt{n}(\hat{\beta}_{kb}(\tau) - \beta(\tau)) \) and \( \sqrt{n}(\hat{\beta}_{wls}(\tau) - \beta(\tau)) \) are the limits of \( \tau(1-\tau)D_{1n}^{-1}D_{0n}D_{1n}^{-1} \) and \( \tau(1-\tau)D_{2n}^{-1} \), respectively, where \( D_{0n}, D_{1n} \) and \( D_{2n} \) are as defined in Condition A3. Thus, part (c) is proved if we can show that for every \( n, D_{2n}^{-1} \leq D_{1n}^{-1}D_{0n}D_{1n}^{-1} \) in the sense of the Löwner order. It suffices to show that \( D_{1n}D_{0n}^{-1}D_{1n} \leq D_{2n} \).

\(^1\)A symmetric matrix \( A \) is said to be greater than or equal to another symmetric matrix \( B \) in the sense of the Löwner order if \( A - B \) is a non-negative definite matrix.
Let \( D_{0n} = n^{-1}B'B, \ D_{1n} = n^{-1}A'B = n^{-1}B'A \) and \( D_{2n} = n^{-1}A'A \), where

\[
B = \begin{bmatrix} \sqrt{n_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{n_k} \end{bmatrix} X, \quad A = \begin{bmatrix} \sqrt{n_1} f_1(F_1^{-1}(\tau)) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{n_k} f_k(F_k^{-1}(\tau)) \end{bmatrix} X. \quad (7)
\]

It follows that

\[
D_{1n} D_{0n}^{-1} D_{1n} = n^{-1} A'B(B'B)^{-1} B'A = n^{-1} A' P_B A \leq n^{-1} A'A = D_{2n},
\]

where \( P_B \) is the orthogonal projection matrix for the column space of \( B \). Part (c) is proved by taking limits of the two sides of the above inequality as \( n \) goes to infinity.

The next theorem provides a necessary and sufficient condition for the Löwner order of part (c) to hold with equality.

**Theorem 2:** Suppose Conditions A1, A2 and A3 hold and assume that \( \Omega_{\tau} \) in (4) is replaced by a consistent estimator.

(a) The asymptotic dispersion matrices of the estimators (4) and (5) coincide if all \( f_i(F_i^{-1}(\tau))'s \) in (3) for \( i = 1, \ldots, k \) are equal.

(b) Suppose \( x_i = (\frac{1}{d_i}) \) for \( i = 1, \ldots, k \), where \( z_1, \ldots, z_k \) are samples from a \( p \)-variate continuous distribution not restricted to any lower dimensional subspace. The asymptotic dispersion matrices of the estimators (4) and (5) coincide only if all \( f_i(F_i^{-1}(\tau))'s \) in (3) for \( i = 1, \ldots, k \) are equal.

**Proof:** For simplicity of notation, we refer to \( f_i(F_i^{-1}(\tau)) \) simply by \( f_i \) in this proof. The point of departure of the proof of this theorem is part (c) of Theorem 1, where a Löwner order between the two dispersion matrices has been established. This order follows from the inequality at the end of the proof of that theorem, which holds with equality if and
only if the column space of $A$ is contained in the column space of $B$. From the definition of $A$ and $B$ given in (7), this condition amounts to the containment of the column space of $FX$ in that of $X$, where $F$ is the diagonal matrix with $f_1, \ldots, f_k$ as its diagonal elements.

Part (a) is proved by using the fact that if all the $f_i$’s are equal, then $FX$ is a constant multiple of $X$, implying the equivalence of the column spaces of these two matrices.

In order to prove part (b), we start from the assumption that the column space of $FX$ is contained in that of $X$, that is, there is a $(p+1) \times (p+1)$ matrix $C$ such that $XC' = FX$. By writing this matrix equation in terms of equality of the corresponding rows of the two sides, we have

$$C x_i = f_i x_i \quad \text{for } i = 1, \ldots, k.$$ 

Therefore, every $f_i$ is an eigen value of the $(p+1) \times (p+1)$ matrix $C$ with eigen vector $x_i$. Lemma 1 proved below implies that all the $f_i$’s have to be the same almost surely over the distribution of the $z_i$’s mentioned in the statement of the theorem.

**Lemma 1:** Suppose $z_1, \ldots, z_k$ are samples from a $p$-variate continuous distribution not restricted to any lower dimensional subspace. If $C$ is a $(p+1) \times (p+1)$ matrix with $(\frac{1}{z_1}), \ldots, (\frac{1}{z_k})$ as eigen vectors, then $C$ is almost surely a multiple of the $(p+1) \times (p+1)$ identity matrix.

**Proof:** Suppose $z_1, \ldots, z_{p+1}$ are samples drawn initially as in the statement of the lemma and $C$ is a $(p+1) \times (p+1)$ matrix having $(\frac{1}{z_1}), \ldots, (\frac{1}{z_{p+1}})$ as eigen vectors. If $C$ is not a multiple of the identity matrix, no eigen value of $C$ has multiplicity $(p+1)$. Therefore, the eigenspace (space of eigenvectors) corresponding to each eigenvalue has dimension $p$ or less. For $(\frac{1}{z_{p+2}}), \ldots, (\frac{1}{z_k})$ to be eigen vectors of $C$, they have to belong to the union of these eigenspaces (each with dimension $< p$). This event has probability zero, according to the hypothesis of the lemma. The result follows.

**Remark 1:** The condition $f_1(F_1^{-1}(\tau)) = \cdots = f_k(F_k^{-1}(\tau))$ mentioned in Theorem 2
may occur when, for instance, the model (1) arises from the more restrictive observation model

$$Y_{ij} = \beta_0 + \beta_1 X_i + e_{ij}, \ j = 1, \ldots, n_i, \ i = 1, \ldots, k,$$

where $e_{ij} \sim F$ for some common distribution $F$ that does not depend on $X_i$. This is a special case of (1) with $\beta_0(\tau) = \beta_0 + F^{-1}(\tau)$ and $\beta_1(\tau) = \beta_1$ for all $\tau$. By denoting $\mu_i = \beta_0 + \beta_1 X_i$, we get $F_i(y) = F(y - \mu_i)$ and $f_i(y) = f(y - \mu_i)$. Thus, the conditional $\tau$-quantile is $F_i^{-1}(\tau) = F^{-1}(\tau) + \mu_i$ and the value of the conditional density at that quantile is $f_i(F_i^{-1}(\tau)) = f(F_i^{-1}(\tau) - \mu_i) = f(F^{-1}(\tau))$, for $i = 1, \ldots, k$. The equality holds for all $\tau$, which is a much stronger condition than the conditions of Theorem 2.

In order to define the estimator (4) completely, one has to choose a consistent estimator of $\Omega_\tau$, which may obtained by plugging any consistent estimator of $1/(f_i(F_i^{-1}(\tau)))$ in (3). Let us denote $s_i(\tau) = 1/(f_i(F_i^{-1}(\tau)))$ and consider some consistent estimators of this parameter under various conditions.

A simple plug-in estimator is obtained by using the sample quantile to estimate $F_i^{-1}$ and the kernel density estimator (Silverman, 1986) of $f_i$, for each $i$. If $h_{ni}$ is the kernel bandwidth, then this estimator would be consistent as long as $h_{ni} \to 0$ and $n_i h_{ni} \to \infty$ as $n_i \to \infty$, and the conditions of Theorem 1 hold.

By noting that $s_i(\tau) = \frac{d}{d\tau} F_i^{-1}(\tau)$, Siddiqui (1960) proposed the finite difference estimator

$$\hat{s}_i(\tau) = \frac{[\hat{q}_i(\tau + h_{ni}) - \hat{q}_i(\tau - h_{ni})]}{2h_{ni}}, \quad (8)$$

which has been quite popular. This estimator is consistent under the conditions of Theorem 1 when the bandwidth parameter $h_{ni}$ tends to 0 as $n_i \to \infty$. A bandwidth rule, suggested by Hall and Sheather (1988) for the purpose of obtaining confidence intervals of
the \( \tau \)-quantile based on Edgeworth expansions is

\[
h_{ni} = n_i^{-1/3} z_\alpha^{2/3} [1.5 s_i(\tau) / s_i''(\tau)]^{1/3},
\]

where \( z_\alpha \) satisfies \( \Phi(z_\alpha) = 1 - \frac{\alpha}{2} \) and \( 1 - \alpha \) is the specified coverage probability of the said confidence interval. In the absence of any information about \( s_i(\cdot) \), one can use the Gaussian model, as in Koenker and Machado (1999), to choose

\[
h_{ni} = n_i^{-1/3} z_\alpha^{2/3} [1.5 \phi^2(\Phi^{-1}(\tau)) / (2(\Phi^{-1}(\tau))^2 + 1)]^{1/3}. \tag{9}
\]

3 Simulations of performance

We now compare the small sample performances of the estimators \( \hat{\beta}_{wls}(\tau) \) and \( \hat{\beta}_{kb}(\tau) \) defined in (4) and (5), in terms of their empirical Mean Squared Error (MSE). The specific version of the WLS estimator we use here is defined by (4) with \( \Omega_{\tau} \) replaced by

\[
\tilde{\Omega}_{\tau} = \begin{pmatrix}
\frac{1}{n_1} \tau (1 - \tau) \hat{s}_1(\tau) & 0 & \cdots & 0 \\
0 & \frac{1}{n_2} \tau (1 - \tau) \hat{s}_2(\tau) & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \frac{1}{n_k} \tau (1 - \tau) \hat{s}_k(\tau)
\end{pmatrix},
\]

where \( \hat{s}_i(\tau) \) is defined as in (8) together with (9) and \( \alpha = 0.05 \).

For \( i = 1, \ldots, k \), we simulate a scalar covariate \( x_i \) from the gamma distribution with shape parameter \( p = 2 \) and scale parameter \( \theta = 0.5 \). Then, for every \( i \) and \( j = 1, \ldots, n_i \), we simulate \( Y_{ij} \) from \( \mathcal{N}(\mu_i, \eta_i^2) \) where \( \mu_i = \beta_1 + \beta_2 x_i - \eta_i \Phi^{-1}(\tau) \), so that the \( \tau \)-quantile of \( Y_{ij} \) is \( \beta_1 + \beta_2 x_i \). As for \( \eta_i^2 \), we choose two different values: \( \eta_i = 1/x_i \) and \( \eta_i = 1 \). Only the second choice ensures asymptotic equivalence of the two estimators as per Theorem 2.
We use $\beta_1 = 1$, $\beta_2 = 0.5$, quantile $\tau = 0.1$, 0.3, 0.5, 0.7 and 0.9 and number of distinct covariate values $k = 5$, 10 and 30. As for the number of replicates $n_i$ for the $i$th distinct value of the covariate, we choose the balanced design $n_1 = \cdots = n_k = n_0$ (say), and use the values 50, 100, 200 and 500 for $n_0$. These choices of $\tau$, $k$ and $n_i$ by and large cover the data analytic problems of Redden et al. (2004), Fernndez et al. (2004), Elsner et al. (2008), Jagger and Elsner (2009) and Kossin et al. (2013).

We compute the KB estimator (5) by using the quantile regression package quantreg (R package version 5.29;//www.r-project.org).

Table 1 shows the empirical MSE of the WLS and KB estimators of the two regression parameters, for $\eta_i = 1/x_i$ and the specified values of the other parameters, based on 10,000 simulation runs. It can be seen that the empirical MSE of the WLS estimator is generally less than that of the KB estimator. The only case where the KB estimator has much smaller MSE than the WLS estimator occurs for the extreme quantiles ($\tau = 0.1$ or 0.9) and small sample size, ($n_i = 50$ and $k = 30$). This may be because $n_i = 50$ is too small for the estimation of variance of extreme quantiles. For $n_i = 200$ or higher, the MSE of the WLS estimator is smaller for all the quantiles considered here. For $\tau = 0.3$, 0.5 and 0.7, the superiority holds for all the sample sizes considered. These small sample findings nicely complement the large sample superiority of the WLS estimator over the KB estimator, as described in Theorem 1.

We now turn to the case $\eta_i = 1$ for all $i$, so that the condition of Theorem 2 holds and the two estimators have asymptotically equivalent performance. Table 2 shows the empirical MSE of the WLS and the KB estimators of the regression of parameters, based on 10,000 simulation runs, for $\eta_i = 1$ and other parameters having specified values as in Table 2. It is found that there is no clear dominance of any one estimator over the other, for any choice of sample size. The WLS estimator of $\beta_0$ generally has smaller MSE than the KB estimator, while the KB estimator appears to work better for $\beta_1$. Overall, the
empirical MSE of two estimators are very close to one another.

4 Data analysis

We now use the WLS and the KB estimator to fit model (1) to the tropical cyclone data considered in Elsner et al. (2008) and available at http://myweb.fsu.edu/jelsner/temp/extspace/globalTCmax4.txt. The satellite based data set consists of lifetime maximum wind speed (metre per second) for each of the 2097 cyclone occurred globally over the years 1981 to 2006. The focus is on the upper quantiles, as these are the storms that may cause major damage.

In Table ??, we report the KB estimator (also used by Elsner et al. (2008)) along with the WLS estimator for the cyclone data at the 0.85, 0.9, 0.95, 0.975 and 0.99 quantiles. We also show the large sample standard errors of the above two estimators of the intercept and the slope parameters. We observe that the WLS estimator has less standard error in all the cases.

Figure 1 shows the observed wind speeds in successive years and the regression lines fitted by the WLS method for the 0.85, 0.9, 0.95, 0.975 and 0.99 quantiles. It may be observed that higher quantiles generally have positive slopes of the fitted regression lines, which point towards extreme cyclone becoming progressively more fierce over the years.

5 Concluding remarks

Thus, the limited simulations and a real data analysis conducted here generally support the wisdom of using the WLS estimator as an alternative to the KB estimator in the case of replicated data, particularly for the middle quantiles.

The key to better performance of the WLS estimator is its utilization of replications
Table 1: Empirical MSE of $\hat{\beta}_{wls}$ and $\hat{\beta}_{kb}$ for $\eta_i = 1/x_i$, $i = 1, \ldots, k$ and for different values of $\tau$, $k$ and $n_0$.

| $\tau$ | $k$ | $n_0=50$ | $n_0=100$ | $n_0=200$ | $n_0=500$ |
|-------|-----|---------|------------|------------|------------|
|       |     | $\hat{\beta}_0$ | $\hat{\beta}_1$ | $\hat{\beta}_0$ | $\hat{\beta}_1$ | $\hat{\beta}_0$ | $\hat{\beta}_1$ |
| 0.1   | 5   | WLS 0.2662 0.2769 | 0.1291 0.1521 | 0.0610 0.0640 | 0.0238 0.0273 |
|       |     | KB 0.3272 0.3521 | 0.1589 0.1875 | 0.0809 0.0872 | 0.0332 0.0389 |
| 0.3   | 10  | WLS 0.0912 0.0445 | 0.0380 0.0203 | 0.0174 0.0094 | 0.0063 0.0035 |
|       |     | KB 0.0877 0.0521 | 0.0426 0.0250 | 0.0208 0.0120 | 0.0087 0.0051 |
| 0.5   | 30  | WLS 0.0299 0.0064 | 0.0104 0.0025 | 0.0042 0.0011 | 0.0014 0.0004 |
|       |     | KB 0.0172 0.0059 | 0.0084 0.0028 | 0.0044 0.0015 | 0.0017 0.0006 |
| 0.7   | 5   | WLS 0.1390 0.1514 | 0.0739 0.0892 | 0.0347 0.0414 | 0.0090 0.0054 |
|       |     | KB 0.1889 0.2027 | 0.0999 0.1241 | 0.0477 0.0550 | 0.0128 0.0076 |
| 0.9   | 10  | WLS 0.0373 0.0212 | 0.0184 0.0102 | 0.0133 0.0165 | 0.0035 0.0021 |
|       |     | KB 0.0494 0.0286 | 0.0247 0.0139 | 0.0189 0.0213 | 0.0050 0.0029 |
| 0.7   | 30  | WLS 0.0079 0.0024 | 0.0036 0.0011 | 0.0090 0.0054 | 0.0017 0.0005 |
|       |     | KB 0.0104 0.0035 | 0.0052 0.0017 | 0.0128 0.0076 | 0.0025 0.0008 |
| 0.9   | 5   | WLS 0.1228 0.1420 | 0.0605 0.0709 | 0.0156 0.0089 | 0.0031 0.0010 |
|       |     | KB 0.1707 0.1870 | 0.0846 0.0978 | 0.0228 0.0132 | 0.0046 0.0016 |
| 0.7   | 10  | WLS 0.0320 0.0190 | 0.0304 0.0350 | 0.0078 0.0044 | 0.0015 0.0005 |
|       |     | KB 0.0459 0.0278 | 0.0425 0.0477 | 0.0114 0.0066 | 0.0023 0.0007 |
| 0.9   | 30  | WLS 0.0061 0.0019 | 0.0122 0.0134 | 0.0031 0.0018 | 0.0006 0.0002 |
|       |     | KB 0.0092 0.0031 | 0.0167 0.0175 | 0.0046 0.0028 | 0.0009 0.0003 |
| 0.9   | 5   | WLS 0.1453 0.1890 | 0.0696 0.0832 | 0.0185 0.0104 | 0.0037 0.0011 |
|       |     | KB 0.1981 0.2505 | 0.1003 0.1202 | 0.0259 0.0148 | 0.0052 0.0017 |
| 0.7   | 10  | WLS 0.0380 0.0221 | 0.0362 0.0469 | 0.0089 0.0051 | 0.0017 0.0005 |
|       |     | KB 0.0512 0.0314 | 0.0481 0.0639 | 0.0125 0.0072 | 0.0025 0.0008 |
| 0.9   | 30  | WLS 0.0079 0.0024 | 0.0141 0.0182 | 0.0036 0.0020 | 0.0006 0.0002 |
|       |     | KB 0.0104 0.0035 | 0.0192 0.0239 | 0.0051 0.0030 | 0.0010 0.0003 |
| 0.9   | 5   | WLS 0.2719 0.2833 | 0.1302 0.1570 | 0.0375 0.0192 | 0.0107 0.0026 |
|       |     | KB 0.3546 0.3697 | 0.1623 0.1863 | 0.0437 0.0254 | 0.0085 0.0029 |
| 0.9   | 10  | WLS 0.0912 0.0432 | 0.0612 0.0717 | 0.0177 0.0099 | 0.0043 0.0011 |
|       |     | KB 0.0870 0.0490 | 0.0802 0.1000 | 0.0213 0.0124 | 0.0044 0.0015 |
| 0.9   | 30  | WLS 0.0304 0.0067 | 0.0253 0.0297 | 0.0063 0.0034 | 0.0013 0.0004 |
|       |     | KB 0.0174 0.0058 | 0.0349 0.0414 | 0.0083 0.0047 | 0.0017 0.0005 |
Table 2: Empirical MSE of $\hat{\beta}_{wls}$ and $\hat{\beta}_{kb}$ for $\eta_i = 1$, $\forall i$ and for different values of $\tau$, $k$ and $n_0$.

| $\tau$ | $k$ | Estimator | $n_0=50$ | $n_0=100$ | $n_0=200$ | $n_0=500$ |
|--------|-----|-----------|---------|----------|---------|---------|
| 0.1    | 5   | WLS       | 0.0631  | 0.0795   | 0.0305  | 0.0487  |
|        |     | KB        | 0.0679  | 0.0797   | 0.0313  | 0.0421  |
| 0.3    | 10  | WLS       | 0.0227  | 0.0256   | 0.0106  | 0.0128  |
|        |     | KB        | 0.0231  | 0.0233   | 0.0112  | 0.0119  |
| 0.5    | 10  | WLS       | 0.0038  | 0.0069   | 0.0012  | 0.0013  |
|        |     | KB        | 0.0059  | 0.0064   | 0.0016  | 0.0013  |

12
Figure 1: Scatter plot of the lifetime maximum wind speeds over the years 1981-2006 along with the regression fit using the WLS estimator at 0.85, 0.90, 0.95, 0.975 and 0.99 quantiles.

Table 3: For a $\tau \in (0, 1)$, consider, quantile regression, $q_y(\tau) = \beta_0 + \beta_1 x$, where, $\beta_0$ and $\beta_1$ are regression parameters.

| $\tau$ | $\beta_0$ | $\beta_1$ | p-value ($\beta_0$) | p-value ($\beta_1$) |
|--------|-----------|-----------|---------------------|---------------------|
| 0.5    | 0.017     | 0.997     | 0                   | 0                   |
| 0.75   | -0.03     | 1.006     | 0                   | 0                   |
| 0.9    | -0.109    | 1.102     | 0.02                | 0                   |
| 0.95   | 0.47      | 1.147     | 0                   | 0                   |
| 0.975  | 1.876     | 1.170     | 0                   | 0                   |
through weights. A weighted version of the KB estimator can also accomplish this. Knight (2001) have shown in an unpublished work that a weighted quantile regression estimator with weights \( \sigma_i(\tau) \) as defined in (3) is first order equivalent to the WLS estimator with those weights and is neither uniformly better nor uniformly worse than it in second order. Our simulations (not reported here) confirmed this finding.

References

Elsner, J. B., Kossin, J. P., and Jagger, T. H. (2008). The increasing intensity of the strongest tropical cyclones. *Nature*, 455:92–95.

Fernández, J. R., Redden, D. T., Pietrobelli, A., and Allison, D. B. (2004). Waist circumference percentiles in nationally representative samples of african-american, european-american, and mexican-american children and adolescents. *The Journal of Pediatrics*, 145(4):439 – 444.

Hall, P. and Sheather, S. J. (1988). On the distribution of a studentized quantile. *Journal of the Royal Statistical Society. Series B (Methodological)*, 50(3):381–391.

Jagger, T. H. and Elsner, J. B. (2009). Modeling tropical cyclone intensity with quantile regression. *International Journal of Climatology*, 29(10):1351–1361.

Knight, K. (2001). Comparing conditional quantile estimators: first and second order considerations. *University of Toronto, Mimeo*.

Koenker, R. (2005). *Quantile regression*. Cambridge University press.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.

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(448):1296–1310.

Kossin, J. P., Olander, T. L., and Knapp, K. R. (2013). Trend analysis with a new global record of tropical cyclone intensity. *Journal of Climate*, 26(24):9960–9976.

Redden, D. T., Fernández, J. R., and Allison, D. B. (2004). A simple significance test for quantile regression. *Statistics in Medicine*, 23(16):2587–2597.

Shorack, G. and Wellner, J. (2009). *Empirical Processes with Applications to Statistics*. Society for Industrial and Applied Mathematics.

Siddiqui, M. M. (1960). Distribution of quantiles in samples from a bivariate population. *Journal of Research of the National Bureau of Standards*, 64B:145–150.

Silverman, Bernard, W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall.
Maximum wind speed (ms$^{-1}$)

| Year | 1980 | 1985 | 1990 | 1995 | 2000 | 2005 |
|------|------|------|------|------|------|------|
| 0.85 | 0.9  | 0.95 | 0.975| 0.99 |      |      |

Year