

Qtools: A Useful Package for Quantiles

Marco Geraci

University of South Carolina

Abstract

Quantiles play a fundamental role in statistics. The quantile function defines the distribution of a random variable and, thus, provides a way to describe the data that is specular but equivalent to that given by the corresponding cumulative distribution function. There are many advantages in working with quantiles, starting from their properties. The renewed interest in their usage seen in the last years is due to the theoretical, methodological and software contributions that have broadened their applicability. This paper presents the R package **Qtools**, a collection of utilities for unconditional and conditional quantiles.

Keywords: discrete random variables, goodness of fit, imputation, location-scale-shape measures, transformations.

1. Applications and theory of quantiles

Quantiles have a long history in applied statistics, especially the median. The analysis of astronomical data by Galileo Galilei in 1632 (Hald 2003, p.149) and geodic measurements by Roger Boscovich in 1757 (Koenker and Bassett 1985; Koenker 2005, p.2) are presumably the earliest examples of application of the least absolute deviation (L_1) estimator in its, respectively, unconditional and conditional forms. However, it was Sir Francis Galton with his remarkable studies on anthropometry to provide a more systematic definition of quantiles and to popularize terms such as ‘median’, ‘quartile’ and ‘percentile’ (Galton 1882, 1885). He was also responsible for introducing the ogive (Galton 1875), which today it is called ‘quantile function’. (However, it is not uncommon to find that the ogive is used as synonym with cumulative distribution function.) Over time, the interest in the applications of quantiles has grown in parallel with advances in the corresponding inference and computing algorithms. The theoretical studies on unconditional and conditional quantiles of continuous random variables started to appear in the statistical literature of the 20th century. According to David (1995), it seems that the term ‘quantile’ itself was first used in print by Kendall (1940). In his paper, Kendall showed that sample quantiles of independent and identically distributed (*iid*) observations have an asymptotic normal distribution (see also Hald, 1998, for historical notes on asymptotic studies on unconditional quantiles). The asymptotic theory of conditional quantile functions is more recent, beginning from the work of Bassett and Koenker (1978) on the median estimator and its generalization to other quantiles by Koenker and Bassett (1978).

In the case of discrete data, studies have somewhat lagged behind, most probably because of the analytical drawbacks surrounding the discontinuities that characterise discrete quantile

functions. Some forms of approximation to continuity have been recently proposed to study the large sample behavior of quantile estimators. For example, [Ma, Genton, and Parzen \(2011\)](#) have demonstrated the asymptotic normality of unconditional sample quantiles based on the definition of the mid-distribution function ([Parzen 2004](#)). [Machado and Santos Silva \(2005\)](#) proposed inferential approaches to the estimation of conditional quantiles for counts based on data jittering.

Finally, it is worth mentioning that asymptotic results are also available for unconditional (see for example [Oberhofer and Haupt 2005](#), and references therein) and conditional ([Parente and Santos Silva 2016](#)) sample quantiles when data are not independent, as well as for nonlinear median functions ([Wang 1995](#)).

Functions implementing quantile methods can be found in common statistical software. A considerable number of R ([R Core Team 2015](#)) packages that provide such functions is available on the Comprehensive R Archive Network (CRAN). The base package **stats** ([R Core Team 2015](#)) contains basic functions to estimate sample quantiles or to compute quantiles of common parametric distributions. Functions for less traditional parametric models are interspersed between packages as auxiliary functions or collected systematically in dedicated packages (e.g., [Hankin and Lee 2006](#); [Nadarajah and Rocha 2016](#)). The **quantreg** package ([Koenker 2013](#)) is arguably a benchmark for distribution-free estimation of linear quantile regression models, as well as the base for other packages which makes use of linear programming (LP) algorithms ([Koenker and D'Orey 1987](#); [Koenker and Park 1996](#)). Other contributions to the modeling of conditional quantile functions include packages for Bayesian regression, e.g., **bayesQR** ([Benoit, Al-Hamzawi, Yu, and den Poel 2014](#)) and **BSquare** ([Smith and Reich 2013](#)), and the **lqmm** package ([Geraci and Bottai 2014](#); [Geraci 2014](#)) for random-effects regression.

The focus of this paper is on the R package **Qtools**, a collection of utilities for unconditional and conditional quantiles which includes functions for the analysis of the location, scale and shape of a distribution, and transformation-based quantile regression. Additional functions in the package provide methods for restricted quantile regression, quantile-based multiple imputation, and quantile regression for discrete responses.

2. Unconditional quantiles

2.1. Definition of quantiles and their properties

Let Y be a random variable with cumulative distribution function (CDF) F_Y and support S_Y . The CDF, calculated at $y \in S_Y$, returns the probability $F_Y(y) \equiv p = \Pr(Y \leq y)$. The quantile function (QF) is defined as $Q(p) = \inf_y \{F_Y(y) \geq p\}$, $0 < p < 1$. (Some authors consider $0 \leq p \leq 1$. For practical purposes, it is simpler to exclude the endpoints 0 and 1.) When F_Y is continuous and strictly monotone (hence, $f_Y(y) \equiv F'_Y(y) > 0$ for all $y \in S_Y$), the quantile function is simply the inverse of F_Y . In other cases, the one-to-one relationship between the values of Y and the probability p is lost where the distribution function is piecewise constant and, by convention, the quantile p is defined as the smallest value y such that $F_Y(y)$ is at least p .

Quantiles enjoy a number of properties. An excellent overview is given by [Gilchrist \(2000\)](#). The advantages of using quantiles rather than probabilities lie in their ‘algebraic’ properties which can be summarized in:

1. $Q_{a+bY}(p) = a + bQ_Y(p)$, $a \in \mathbb{R}$, $b \in \mathbb{R}_+$ (**location-shift rule**);
2. if $Q_1(p)$ and $Q_2(p)$ are two QFs, then $Q_1(p) + Q_2(p)$ is a QF (**addition rule**);
3. if $Q_1(p)$ and $Q_2(p)$ are two QFs, then $\pi Q_1(p) + (1 - \pi)Q_2(p)$, $0 \leq \pi \leq 1$, lies between the two distributions (**intermediate rule**);
4. if $Q_1(p)$ and $Q_2(p)$ are two positive QFs, then $Q_1(p) \cdot Q_2(p)$ is a QF (**multiplication rule**);
5. if $Q_Y(p)$ is the QF of Y , the reflection $-Q_Y(1 - p)$ is the QF of $-Y$ (**reflection rule**);
6. if $Q_Y(p)$ is the QF of Y , the reciprocal $1/Q_Y(1 - p)$ is the QF of $1/Y$ (**reciprocal rule**);
7. if $h(\cdot)$ is a non-decreasing function on \mathbb{R} , then $Q_{h(Y)}(p) = h\{Q_Y(p)\}$. Hence $Q_Y(p) = h^{-1}\{Q_{h(Y)}(p)\}$ (**Q-transformation rule** or equivariance to monotone transformations).

(Note that, in general, the last property does not hold for the expected value, i.e., $E\{h(Y)\} \neq h\{E(Y)\}$.)

Sample quantiles for a random variable Y can be calculated in a number of ways, depending on how they are defined (Hyndman and Fan 1996). For example, the function `quantile` in the base package `stats` provides nine different sample quantile estimators, which are based on the sample order statistics or the inverse of the empirical CDF. These estimators are distribution-free as they do not depend on any parametric assumption about F (or Q). Alternatively, one could consider a model for F (or Q), indexed by some low-dimensional parameter, say θ . Estimation of θ can be carried out efficiently using one of several available methods, such as maximum likelihood estimation (MLE), the method of moments and the method of percentiles (Gilchrist 2000). The focus of the package `Qtools` is on the distribution-free approach.

2.2. Sample quantiles and large- n properties

Let $\mathcal{Y}_n = (Y_1, Y_2, \dots, Y_n)$ be an *iid* sample of size n from the population F_Y . Let ξ_p denote the p th population quantile and $\hat{\xi}_p$ the corresponding sample quantile based on the sample \mathcal{Y}_n . (The subscripts will be dropped occasionally to ease notation, e.g., F will be used in place of F_Y or ξ in place of ξ_p .) In the continuous case, it is well known that $\sqrt{n}(\hat{\xi}_p - \xi_p)$ is approximately normal with mean zero and variance

$$\omega^2 = \frac{p(1-p)}{\{f_Y(\xi_p)\}^2}. \quad (1)$$

A more general result is obtained when the Y_i 's are independent with common ξ_p but different distribution function, i.e., $Y_i \sim F_{Y_i}$. In this case the variance depends on the density f_{Y_i} , i.e.,

$$\omega_i^2 = \frac{p(1-p)}{\{f_{Y_i}(\xi_p)\}^2}. \quad (2)$$

The density evaluated at the p th quantile, $f(\xi_p)$, is called density-quantile function by Parzen (1979). Its reciprocal, $s(p) \equiv 1/f(\xi_p)$, is called sparsity function (Tukey 1965) or quantile-density function (Parzen 1979). It is easy to verify that $s(p) = dQ(p)/dp$.

As mentioned previously, the discontinuities of F_Y when Y is discrete represent a mathematical inconvenience. [Ma et al. \(2011\)](#) derived the asymptotic distribution of the sample mid-quantiles, that is, the sample quantiles based on the mid-distribution function (mid-CDF). The latter is defined as $F_Y^{mid}(y) = F_Y(y) - 0.5p_Y(y)$, where $p_Y(y)$ denotes the probability mass function ([Parzen 2004](#)). In particular, they showed that, as n becomes large, $\sqrt{n}(\hat{\xi}_p - \xi_p)$ is approximately normal with mean 0. Under *iid* assumptions, the expression for the variance ω^2 is similar to that in (1); see [Ma et al. \(2011\)](#) for details.

The package **Qtools** provides the functions `midecdf` and `midquantile`, which return objects of class ‘`midecdf`’ or ‘`midquantile`’, respectively, containing: the values or the probabilities at which mid-cumulative probabilities or mid-quantiles are calculated (‘`x`’), the mid-cumulative probabilities or the mid-quantiles (‘`y`’), and the function that linearly interpolates those coordinates (‘`fn`’). An example is shown below using data simulated from a Poisson distribution.

```
R> library("Qtools")
R> set.seed(467)
R> y <- rpois(1000, 4)
R> pmid <- midecdf(y)
R> xmid <- midquantile(y, probs = pmid$y)
R> pmid
```

```
Empirical mid-ECDF
Call:
midecdf(x = y)
```

```
R> xmid
```

```
Empirical mid-ECDF
Call:
midquantile(x = y, probs = pmid$y)
```

A confidence interval for sample mid-quantiles can be obtained using `confint.midquantile`. This function is applied to the output of `midquantile` and returns an object of class ‘`data.frame`’ containing sample mid-quantiles, lower and upper bounds of the confidence interval of a given level (95% by default), along with standard errors as an attribute named ‘`stderr`’. This is shown below using the sample `y` generated in the previous example.

```
R> xmid <- midquantile(y, probs = 1:3/4)
R> x <- confint(xmid, level = 0.95)
R> x
```

	midquantile	lower	upper
25%	2.540000	2.416462	2.663538
50%	3.822816	3.693724	3.951907
75%	5.254902	5.072858	5.436946

```
R> attr(x, "stderr")
```

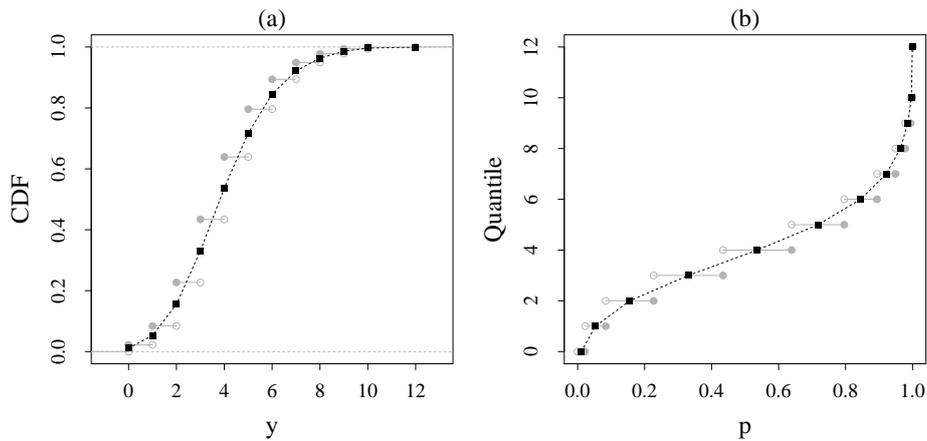


Figure 1: Cumulative distribution (a) and quantile (b) functions for simulated Poisson data. The ordinary cumulative distribution function (CDF) and quantile function (QF) are represented by step-functions (grey lines), with the convention that, at the point of discontinuity or ‘jump’, the function takes its value corresponding to the ordinate of the filled circle as opposed to that of the hollow circle. The mid-CDF and mid-QF are represented by filled squares, while the piecewise linear functions (dashed lines) connecting the squares represent continuous versions of, respectively, the ordinary CDF and QF.

```
[1] 0.06295447 0.06578432 0.09276875
```

Finally, a plot method is available for both `midcdf` and `midquantile` objects. An illustration is given in Figure 1. The mid-distribution and mid-quantile functions are discrete and their values are marked by filled squares. The piecewise linear functions connecting the filled squares represent continuous versions of the CDF and QF which interpolate between the steps of, respectively, the ordinary CDF and quantile functions. Note that the argument `jumps` is a logical value indicating whether values at jumps should be marked.

```
R> par(mfrow = c(1,2))
R> plot(pmid, xlab = "y", ylab = "CDF", jumps = TRUE)
R> points(pmid$x, pmid$y, pch = 15)
R> plot(xmid, xlab = "p", ylab = "Quantile", jumps = TRUE)
R> points(xmid$x, xmid$y, pch = 15)
```

2.3. LSS - Location, scale and shape of a distribution

Since the cumulative distribution and quantile functions are two sides of the same coin, the location, scale and shape (LSS) of a distribution can be examined using one or the other. Well-known quantile-based measures of location and scale are the median and inter-quartile range (IQR), respectively. Similarly, there are also a number of quantile-based measures for skewness and kurtosis (Groeneveld and Meeden 1984; Groeneveld 1998; Jones, Rosco, and Pewsey 2011).

Define the ‘central’ portion of the distribution as that delimited by the quantiles $Q(p)$ and $Q(1-p)$, $0 < p < 0.5$, and define the ‘tail’ portion as that lying outside these quantiles. Let $IPR(p) = Q(1-p) - Q(p)$ denote the inter-quantile range at level p . Building on the results by Horn (1983) and Ruppert (1987), Staudte (2014) considered the following identity:

$$\underbrace{\frac{IPR(p)}{IPR(r)}}_{\text{kurtosis}} = \underbrace{\frac{IPR(p)}{IPR(q)}}_{\text{tail-weight}} \cdot \underbrace{\frac{IPR(q)}{IPR(r)}}_{\text{peakedness}}, \quad (3)$$

where $0 < p < q < r < 0.5$. These quantile-based measures of shape are sign, location and scale invariant. As compared to moment-based indices, they are also more robust to outliers and easier to interpret (Groeneveld 1998; Jones *et al.* 2011).

It is easy to verify that a quantile function can be written as

$$Q(p) = \underbrace{Q(0.5)}_{\text{median}} + \frac{1}{2} \underbrace{IPR(0.25)}_{\text{IQR}} \cdot \underbrace{\frac{IPR(p)}{IPR(0.25)}}_{\text{shape index}} \cdot \left(\underbrace{\frac{Q(p) + Q(1-p) - 2Q(0.5)}{IPR(p)}}_{\text{skewness index}} - 1 \right). \quad (4)$$

This identity establishes a relationship between the location (median), scale (IQR) and shape of a distribution. (This identity appears in Gilchrist (2000, p.74) with an error of sign. See also Benjamini and Krieger (1996, eq.1).) The quantity $IPR(p)/IPR(0.25)$ in (4) is loosely defined as ‘shape index’ (Gilchrist 2000, p.72), although it can be seen as the tail-weight measure given in (3) when $p < 0.25$. For symmetric distributions, the contribution of the skewness index (Gilchrist 2000, p.53) vanishes. Note that the skewness index not only is location and scale invariant, but is also bounded between -1 and 1 (as opposed to the Pearson’s third standardized moment which can be infinite or even undefined).

The function `qlss` provides a quantile-based LSS summary as defined in (4) of either a *theoretical* or an *empirical* distribution. It returns an object of class ‘`qlss`’, which is a list containing measures of location (median), scale (IQR and IPR), and shape (skewness and shape indices) for each of the probabilities specified in the argument `probs` (by default, `probs = 0.1`). The quantile-based LSS summary of the normal distribution is given in the example below for $p = 0.1$. The argument `fun` can take any quantile function whose probability argument is named ‘`p`’ (this is the case for many standard quantile functions in R, e.g., `qt`, `qchisq`, `qf`, etc.).

```
R> qlss(fun = "qnorm", probs = 0.1)
```

```
call:
```

```
qlss.default(fun = "qnorm", probs = 0.1)
```

Unconditional Quantile-Based Location, Scale, and Shape

```

** Location **
Median
[1] 0
** Scale **
Inter-quartile range (IQR)
[1] 1.34898
Inter-quantile range (IPR)
      0.1
2.563103
** Shape **
Skewness index
0.1
      0
Shape index
      0.1
1.900031

```

An empirical example is now illustrated using the `faithful` data set, which contains 272 observations on waiting time (minutes) between eruptions and the duration (minutes) of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA. Summary statistics are given in Table 1.

	Minimum	Q1	Q2	Q3	Maximum
Waiting time	43.0	58.0	76.0	82.0	96.0
Duration	1.6	2.2	4.0	4.5	5.1

Table 1: Minimum, maximum and three quartiles (Q1, Q2, Q3) for waiting time and duration in the Old Faithful Geyser data set.

Suppose the interest is in describing the distribution of waiting times. The density, as estimated using `density.default`, is plotted in Figure 2, along with the mid-quantile function. The distribution is bimodal with peaks at around 54 and 80 minutes. Note that `qlss` takes the argument `type` as well as any other argument for the function `quantile`.

```

R> y <- faithful$waiting
R> qlss(y, probs = c(0.05, 0.1, 0.25), type = 7)

```

```

call:
qlss.numeric(x = y, probs = c(0.05, 0.1, 0.25), type = 7)

```

Unconditional Quantile-Based Location, Scale, and Shape

```

** Location **
Median

```

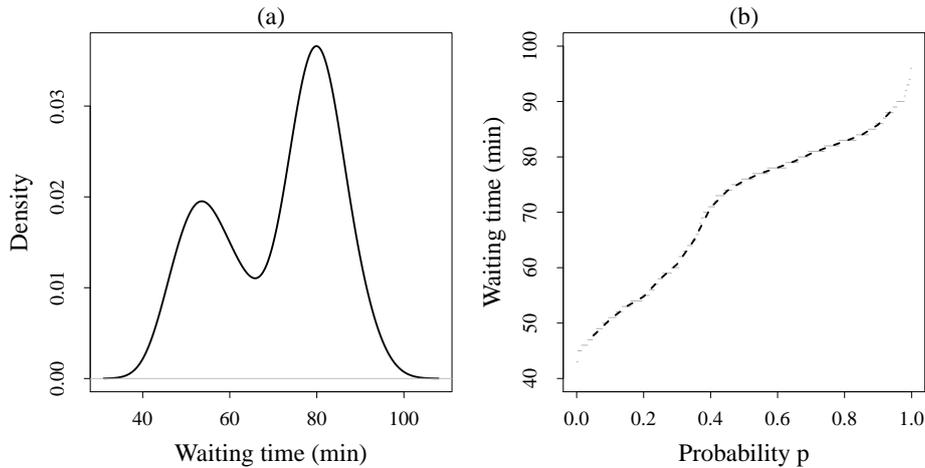


Figure 2: Estimated density (a) and empirical mid-quantile (b) functions of waiting time between eruptions in the Old Faithful Geyser data set.

```
[1] 76
** Scale **
Inter-quantile range (IQR)
[1] 24
Inter-quantile range (IPR)
0.05  0.1  0.25
   41   35   24
** Shape **
Skewness index
      0.05      0.1      0.25
-0.3658537 -0.4285714 -0.5000000
Shape index
      0.05      0.1      0.25
1.708333 1.458333 1.000000
```

At $p = 0.1$, the skewness index is approximately -0.43 , which denotes a rather strong left asymmetry. As for the shape index, which is equal to 1.46 , one could say that the tails of this distribution weigh less than those of a normal distribution (1.90), though of course a comparison between unimodal and bimodal distributions is not meaningful.

3. Conditional quantiles

3.1. Linear models

Conditional modeling of quantiles can be easily illustrated starting from the simple case in which two populations are contrasted in relation to some characteristic Y . Let $F_0(y)$ and $F_1(y)$ denote the cumulative distribution functions (CDFs) of Y in the two populations, and let $\mu_0 = \int_{-\infty}^{\infty} y dF_0(y)$ and $\mu_1 = \int_{-\infty}^{\infty} y dF_1(y)$ be their respective means. The two CDFs are

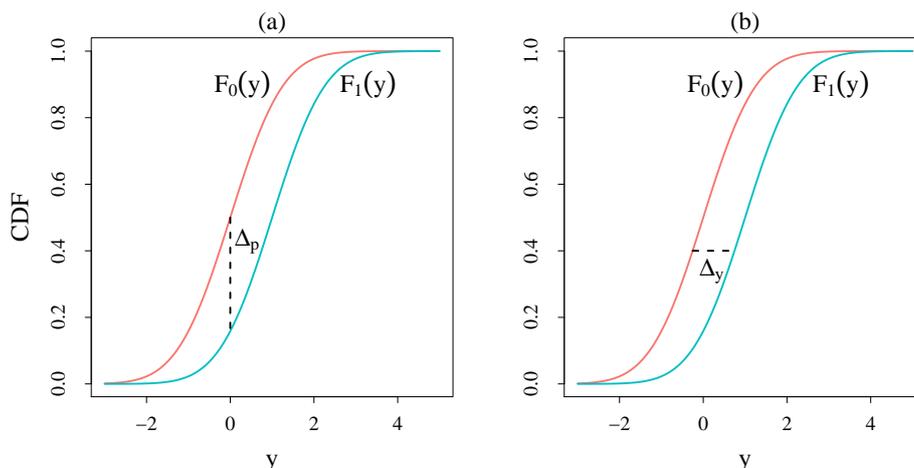


Figure 3: (a) Probability difference (à la Kolmogorov–Smirnov); (b) Quantile difference (à la Lehmann–Doksum).

plotted in Figure 3. The question of whether and how F_1 differs from F_0 can be approached by comparing, for any fixed value y , the distance between the CDFs along the probability (vertical) axis. Therefore

$$\Delta_p(y) = F_1(y) - F_0(y).$$

A discrepancy between the two distributions at some value y would result in a positive or negative $\Delta_p(y)$. The largest (in magnitude) of all the differences $\Delta_p(y)$ is used to calculate the Kolmogorov–Smirnov test statistic. Plot (a) in Figure 3 shows the case in which F_1 is stochastically dominated by F_0 , thus $\Delta_p(y) < 0$ for $-\infty < y < \infty$. A complementary approach consists in comparing, for any fixed value p , the distance of the CDFs along the quantile (horizontal) axis. Therefore

$$\Delta_y(p) = F_1^{-1}(p) - F_0^{-1}(p)$$

gives the contrast between the p th quantiles of the two distributions. This is also called ‘quantile treatment effect’ (Doksum 1974; Lehmann 1975; Koenker and Xiao 2002). An illustration is given in plot (b), Figure 3.

In general, the p th linear QR model is of the form

$$Q_{Y|X}(p) = \mathbf{x}^\top \boldsymbol{\beta}(p) \tag{5}$$

where \mathbf{x} is a k -dimensional vector of covariates (including 1 as first element) and $\boldsymbol{\beta}(p) = [\beta_0(p), \beta_1(p), \dots, \beta_{k-1}(p)]^\top$ is a vector of coefficients. The ‘slopes’ $\beta_j(p)$, $j = 1, \dots, k - 1$, have the usual interpretation of partial derivatives. For example, in case of the simple model $Q_{Y|X}(p) = \beta_0(p) + \beta_1(p)x$, one obtains

$$\frac{\partial Q_{Y|X}(p)}{\partial x} = \beta_1(p).$$

Estimation can be carried out using LP algorithms which, given a sample (\mathbf{x}_i, y_i) , $i = 1, \dots, n$, solve

$$\min_{\mathbf{b} \in \mathbb{R}^k} \sum_{i=1}^n \kappa_p \left(y_i - \mathbf{x}_i^\top \mathbf{b} \right),$$

where $\kappa_p(u) = u(p - I(u < 0))$, $0 < p < 1$, is the check loss function.

Large- n approximation of standard errors can be obtained from the sampling distribution of the linear quantile estimators. The quantity $\sqrt{n}(\hat{\beta}(p) - \beta(p))$ has a limiting centered normal distribution whose scale depends on whether the observations are assumed to be *iid* or *nid*. Let $f_{Y|X}$ denote the density of Y conditional on X and $\xi_p \equiv \mathbf{x}^\top \beta(p)$. In the *nid* case, the asymptotic variance-covariance matrix of p -th regression quantile is defined by the sandwich variance estimator (Koenker 2005, p.74)

$$V = p(1-p)H^{-1}DH^{-1} \quad (6)$$

where $D = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top$ and $H = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top f_{Y_i|X_i}(\xi_p)$. In the *iid* case, the conditional density is constant and the above expression simplifies to

$$V = \omega^2 D^{-1} \quad (7)$$

where $\omega^2 = \frac{p(1-p)}{\{f_{Y_i|X_i}(\xi_p)\}^2}$, $i = 1, \dots, n$.

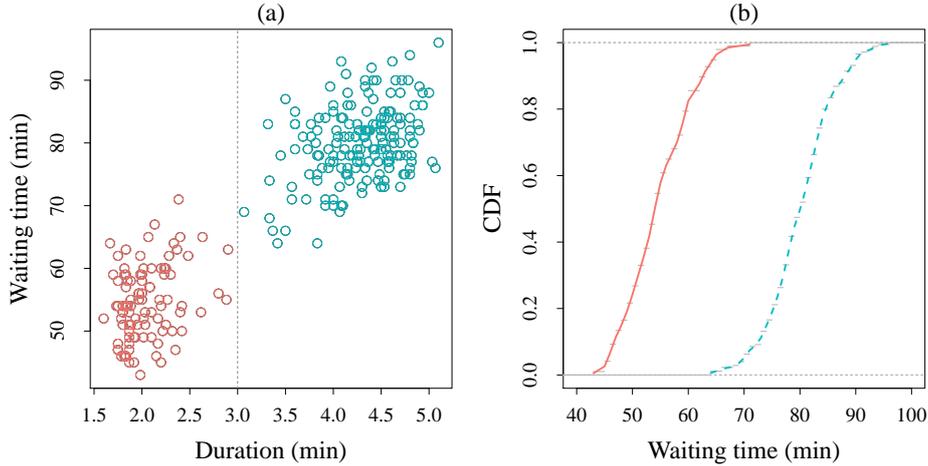


Figure 4: (a) Waiting times between eruptions against durations of eruptions (dashed vertical line drawn at 3 minutes) in the Old Faithful Geyser data set. (b) Mid-CDF of waiting time by duration of eruption (solid line, shorter than 3 minutes; dashed line, longer than 3 minutes).

Waiting times between eruptions are plotted against the durations of the eruptions in Figure 4. Two clusters of observations can be defined for durations below and above 3 minutes (see also Azzalini and Bowman 1990). The distribution shows a strong bimodality as already illustrated in Figure 2, which is also apparent in the sample quantile function. The latter is initially convex up to about $p = 0.35$, then concave until $p = 0.65$, and then convex again. Unsurprisingly, the proportion of observations with duration less than 3 minutes is approximately 35%.

A dummy variable for durations equal to or longer than 3 minutes is created to define the two distributions and included as covariate X in a model as the one specified in (5). The latter is then fitted to the Old Faithful Geyser data using the function `rq` in the R package **quantreg** (Koenker 2013). The coefficient for \mathbf{x} , therefore, is an estimate of the quantile difference $\Delta_y(p)$. The latter is calculated for $p \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$ in the following example:

```
R> require("quantreg")
R> y <- faithful$waiting
R> x <- as.numeric(faithful$eruptions >= 3)
R> fit <- rq(formula = y ~ x, tau = c(0.1, 0.25, 0.5, 0.75, 0.9))
R> fit
```

Call:

```
rq(formula = y ~ x, tau = c(0.1, 0.25, 0.5, 0.75, 0.9))
```

Coefficients:

	tau= 0.10	tau= 0.25	tau= 0.50	tau= 0.75	tau= 0.90
(Intercept)	47	50	54	59	63
x	26	26	26	25	25

Degrees of freedom: 272 total; 270 residual

From the output above, it is quite evident that the distribution of waiting times is shifted by an approximately constant amount at all considered values of p . The location-shift hypothesis can be tested for by using the Khmaladze test. The null hypothesis is that two distributions, say F_0 and F_1 , differ by a pure location shift (Koenker and Xiao 2002), that is

$$H_0 : F_1^{-1}(p) = F_0^{-1}(p) + \delta_0,$$

where δ_0 is the quantile treatment effect, constant over p . The location–scale-shift specification of the test considers

$$H_0 : F_1^{-1}(p) = \delta_1 F_0^{-1}(p) + \delta_0.$$

The alternative hypothesis is that the model is more complex than the one specified in the null hypothesis. The Khmaladze test is implemented in the **quantreg** package (see `?quantreg::KhmaladzeTest` for further details). The (asymptotic) critical values of the test and corresponding significance levels are not readily available in the same package. These are provided by the function `KhmaladzeFormat` in the **Qtools** package.

```
R> kt <- KhmaladzeTest(formula = y ~ x, taus = seq(.05, .95, by = .01),
+   nullH = "location")
R> kt
```

\$nullH

```
[1] "location"
```

\$Tn

```
[1] 1.867598
```

\$THn

```
[1] 1.867598
```

attr(,"class")

```
[1] "KhmaladzeTest"
```

```
R> KhmaladzeFormat(kt, 0.05)
```

```
Khmaladze test for the location-shift hypothesis
Joint test is not significant at 10% level
Test(s) for individual slopes:
not significant at 10% level
```

Distribution-free quantile regression does not require introducing an assumption on the functional form of the error distribution (Koenker and Bassett 1978), but only weaker quantile restrictions (Powell 1994). Comparatively, the linear specification of the conditional quantile function in Equation 5 is a much stronger assumption and thus plays an important role for inferential purposes.

The problem of assessing the goodness of fit (GOF) is rather neglected in applications of QR. Although some approaches to GOF have been proposed (Zheng 1998; Koenker and Machado 1999; He and Zhu 2003; Khmaladze and Koul 2004), there is currently a shortage of software code available to users. The function `GOFTest` implements a test based on the cusum process of the gradient vector (He and Zhu 2003). Briefly, the test statistic is given by the largest eigenvalue of

$$n^{-1} \sum_i^n \mathbf{R}_n(\mathbf{x}_i) \mathbf{R}_n^\top(\mathbf{x}_i)$$

where $\mathbf{R}_n(\mathbf{t}) = n^{-1/2} \sum_{j=1}^n \psi_p(r_j) \mathbf{x}_j I(\mathbf{x}_j \leq \mathbf{t})$ is the residual cusum (RC) process and $\psi_p(r_j)$ is the derivative of the loss function κ_p calculated for residual $r_j = y_j - \mathbf{x}_j^\top \boldsymbol{\beta}(p)$. The sampling distribution of this test statistic is non-normal (He and Zhu 2003) and a resampling approach is used to obtain the p -value under the null hypothesis.

An example is provided further below using the New York Air Quality data set, which contains 111 complete observations on daily mean ozone (parts per billion – ppb) and solar radiation (Langleys – Ly). For simplicity, wind speed and maximum daily temperature, also included in the data set, are not analyzed here.

Suppose that the model of interest is

$$Q_{\text{ozone}}(p) = \beta_0(p) + \beta_1(p) \cdot \text{Solar.R.} \quad (8)$$

Three conditional quantiles ($p \in \{0.1, 0.5, 0.9\}$) are estimated and plotted using the following code:

```
R> dd <- airquality[complete.cases(airquality), ]
R> dd <- dd[order(dd$Solar.R), ]
R> fit.rq <- rq(Ozone ~ Solar.R, tau = c(.1, .5, .9), data = dd)
R> x <- seq(min(dd$Solar.R), max(dd$Solar.R), length = 200)
R> yhat <- predict(fit.rq, newdata = data.frame(Solar.R = x))
R> plot(Ozone ~ Solar.R, data = dd)
R> apply(yhat, 2, function(y,x) lines(x,y), x = x)
```

As a function of solar radiation, the median of the ozone daily averages increases by 0.09 ppb for each Ly increase in solar radiation (Figure 5). The 90th centile of conditional ozone shows

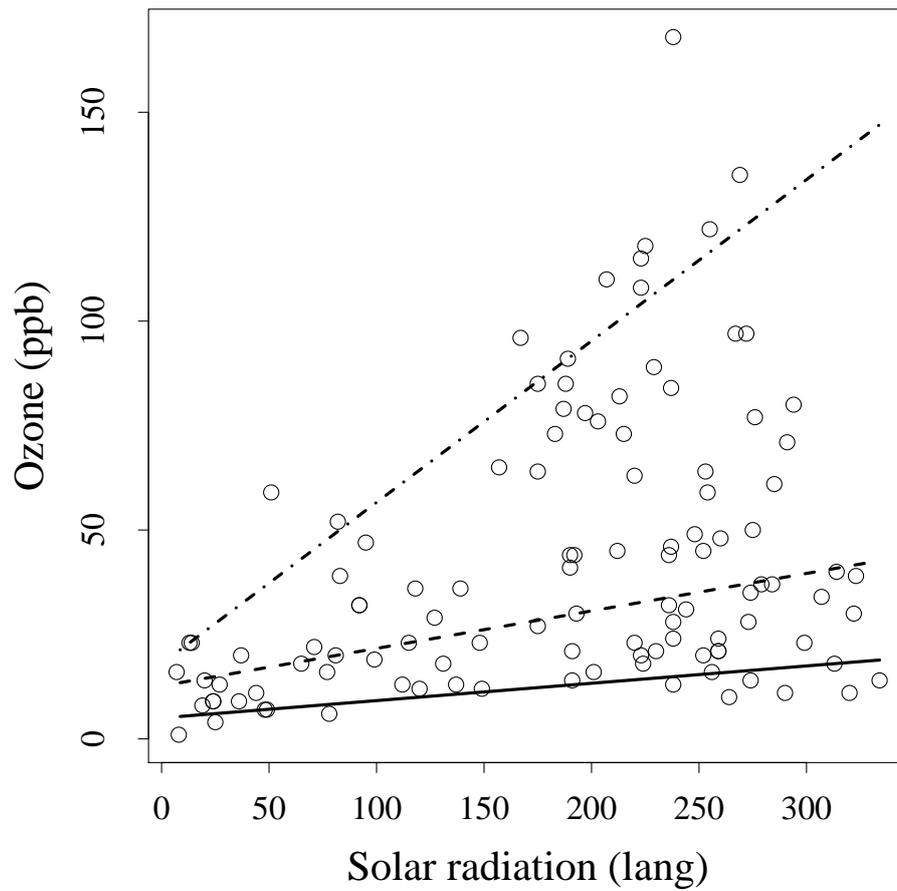


Figure 5: Predicted 10th (solid line), 50th (dashed line), and 90th (dot-dashed line) centiles of ozone conditional on solar radiation in the New York Air Quality data set.

a steeper slope at 0.39 ppb/Ly, about nine times larger than the slope of the conditional 10th centile at 0.04 ppb/Ly.

The RC test applied to the the object `fit.rq` provides evidence of lack of fit for all quantiles considered, particularly for $p = 0.1$ and $p = 0.5$. Therefore the straight-line model in Equation 8 for these three conditional quantiles does not seem to be appropriate. The New York Air Quality data set will be analyzed again in the next section, where a transformation-based approach to nonlinear modeling is discussed.

```
R> gof.rq <- GOFTest(fit.rq, alpha = 0.05, B = 1000, seed = 987)
R> gof.rq
```

```
Goodness-of-fit test for quantile regression based on the cusum process
Quantile 0.1: Test statistic = 0.1057; p-value = 0.005
Quantile 0.5: Test statistic = 0.2191; p-value = 0
Quantile 0.9: Test statistic = 0.0457; p-value = 0.066
```

3.2. Transformation models

Complex dynamics may result in nonlinear effects in the relationship between the covariates and the response variable. For instance, in kinesiology, pharmacokinetics, and enzyme kinetics, the study of the dynamics of an agent in a system involves the estimation of nonlinear models; phenomena like human growth, certain disease mechanisms and the effects of harmful environmental substances such as lead and mercury, may show strong nonlinearities over time. In this section, the linear model is abandoned in favor of a more general model of the type

$$Q_{Y|X}(p) = g \left\{ \mathbf{x}^\top \boldsymbol{\beta}(p) \right\} \quad (9)$$

for some real-valued function g . If g is nonlinear, the alternative approaches to conditional quantile modeling are

1. nonlinear parametric models: this approach may provide a model with substantive interpretability, possibly parsimonious (in general more parsimonious than polynomials), and valid beyond the observed range of the data. A nonlinear model depends on either prior knowledge of the phenomenon or the introduction of new, strong theory to explain the observed relationship with potential predictive power. Estimation may present challenges;
2. polynomial models and smoothing splines: this approach goes under the label of *nonparametric regression*, in which the complexity of the model is approximated by a sequence of locally linear polynomials (a naïve global polynomial trend can be considered to be a special case). A nonparametric model need not introducing strong assumptions about the relationship and is essentially data-driven. Estimation is based on linear approximations and, typically, requires the introduction of a penalty term to control the degree of smoothing;
3. transformation models: a flexible, parsimonious family of parametric transformations is applied to the response seeking to obtain approximate linearity on the transformed

scale. The data provide information about the ‘best’ transformation among a family of transformations. Estimation is facilitated by the application of methods for linear models.

The focus of this section is on the third approach. More specifically the functions available in **Qtools** refer to the methods for transformation-based QR models developed by Powell (1991), Chamberlain (1994), Mu and He (2007), Dehbi, Cortina-Borja, and Geraci (2016) and Geraci and Jones (2015). Examples of approaches to nonlinear QR based on parametric models or splines can be found in Koenker and Park (1996) and Yu and Jones (1998), respectively.

The goal of transformation-based QR is to fit the model

$$Q_{h(Y;\lambda_p)}(p) = \mathbf{x}^\top \boldsymbol{\beta}(p). \quad (10)$$

The assumption is that the transformation h is the inverse of g , $h(Y;\lambda_p) \equiv g^{-1}(Y)$, so that the p th quantile function of the transformed response variable is linear. (In practice, it is satisfactory to achieve approximate linearity.) The parameter λ_p is a low-dimensional parameter that gives some flexibility to the shape of the transformation and is estimated from the data. In general, the interest is on predicting $Q_{Y|X}(p)$ and estimating the effects of the covariates on $Q_{Y|X}(p)$. If h is a non-decreasing function on \mathbb{R} (as is the case for all transformations considered here), predictions can be easily obtained from (10) by virtue of the equivariance property (Q-transformation rule) of quantiles, i.e.,

$$Q_{Y|X}(p) = h^{-1} \left\{ \mathbf{x}^\top \boldsymbol{\beta}(p); \lambda_p \right\}. \quad (11)$$

The marginal effect of the j th covariate x_j can be obtained by differentiating the quantile function $Q_{Y|X}(p)$ with respect to x_j . This can be written as the derivative of the composition $Q \circ \eta$, i.e.,

$$\frac{\partial Q(p)}{\partial x_j} = \frac{\partial Q(p)}{\partial \eta(p)} \cdot \frac{\partial \eta(p)}{\partial x_j}, \quad (12)$$

$\eta(p) = \mathbf{x}^\top \boldsymbol{\beta}(p)$. Once the estimates $\hat{\boldsymbol{\beta}}(p)$ and $\hat{\lambda}_p$ are obtained, these can be plugged in Equations 11 and 12.

The package **Qtools** provides several transformation families, namely the Box–Cox (Box and Cox 1964), Aranda-Ordaz (Aranda-Ordaz 1981) and Jones (Jones 2007; Geraci and Jones 2015) transformations. A distinction between these families is made in terms of the support of the response variable to which the transformation is applied and the number of transformation parameters. The Box–Cox is a one-parameter family of transformations which applies to singly bounded variables, $y > 0$. The Aranda-Ordaz symmetric and asymmetric transformations too have one parameter and are used when responses are bounded on the unit interval, $0 < y < 1$ (doubly bounded). Geraci and Jones (2015) developed two families of transformations which can be applied to either singly or doubly bounded responses:

- Proposal I transformations – this family has one parameter and it comes in both symmetric and asymmetric forms;
- Proposal II transformations – this family has two parameters, with one parameter modeling the symmetry (or lack thereof) of the transformation.

Originally, [Box and Cox \(1964\)](#) proposed using power transformations to address lack of linearity, homoscedasticity and normality of the residuals in mean regression modeling. [Sakia \(1992, p.175\)](#) reported “that seldom does this transformation fulfil the basic assumptions of linearity, normality and homoscedasticity simultaneously as originally suggested by Box & Cox (1964). The Box-Cox transformation has found more practical utility in the empirical determination of functional relationships in a variety of fields, especially in econometrics”. Indeed, the practical utility of power transformations has been long recognized in QR modeling ([Powell 1991](#); [Buchinsky 1995](#); [Chamberlain 1994](#); [Mu and He 2007](#)). Model 10 is the Box–Cox QR model if

$$h_{BC}(Y; \lambda_p) = \begin{cases} \frac{Y^{\lambda_p} - 1}{\lambda_p} & \text{if } \lambda_p \neq 0 \\ \log Y & \text{if } \lambda_p = 0. \end{cases} \quad (13)$$

Note that when $\lambda_p \neq 0$, the range of this transformation is not \mathbb{R} but the singly bounded interval $(-1/\lambda_p, \infty)$. This implies that the inversion in (11) is defined only for $\lambda_p \mathbf{x}^\top \boldsymbol{\beta}(p) + 1 > 0$.

The ‘symmetric’ Aranda-Ordaz transformation is given by

$$h_{AOs}(Y; \lambda_p) = \begin{cases} \frac{2}{\lambda_p} \frac{Y^{\lambda_p} - (1 - Y)^{\lambda_p}}{Y^{\lambda_p} + (1 - Y)^{\lambda_p}} & \text{if } \lambda_p \neq 0, \\ \log \left(\frac{Y}{1 - Y} \right) & \text{if } \lambda_p = 0. \end{cases} \quad (14)$$

(The symmetry here is that $h_{AOs}(\theta; \lambda_p) = -h_{AOs}(1 - \theta; \lambda_p) = h_{AOs}(\theta; -\lambda_p)$.) There is a range problem with this transformation too since, for all $\lambda_p \neq 0$, the range of h_{AOs} is not \mathbb{R} , but $(-2/|\lambda|, 2/|\lambda|)$. The ‘asymmetric’ Aranda-Ordaz transformation is given by

$$h_{AOa}(Y; \lambda_p) = \begin{cases} \log \left\{ \frac{(1 - Y)^{-\lambda_p} - 1}{\lambda_p} \right\} & \text{if } \lambda_p \neq 0, \\ \log \{-\log(1 - Y)\} & \text{if } \lambda_p = 0. \end{cases} \quad (15)$$

For $\lambda_p = 0$, this is equivalent to the complementary log-log. The asymmetric Aranda-Ordaz transformation does have range \mathbb{R} , although one can consider alternatives to it. Note that $h_{AOa}(Y; 1) = \log(Y/(1 - Y))$, i.e., it is symmetric.

To overcome range problems which give rise to computational difficulties, [Geraci and Jones \(2015\)](#) proposed to use instead one-parameter transformations with range \mathbb{R} . Proposal I is written in terms of the variable, say, W

$$h_I(W; \lambda_p) = \begin{cases} \frac{1}{2\lambda_p} \left(W^{\lambda_p} - \frac{1}{W^{\lambda_p}} \right) & \text{if } \lambda_p \neq 0 \\ \log W & \text{if } \lambda_p = 0, \end{cases} \quad (16)$$

which takes on four forms depending on the relationship of W to Y , as described in Table 2. For each of domains $(0, \infty)$ and $(0, 1)$, there are two forms, one symmetric, the other asymmetric.

Support of Y	Symmetric	Asymmetric
$(0, \infty)$	$W = Y$ $h_{Is}(Y; \lambda_p)$	$W = \log(1 + Y)$ $h_{Ia}(Y; \lambda_p)$
$(0, 1)$	$W = Y/(1 - Y)$ $h_{Is}(Y; \lambda_p)$	$W = -\log(1 - Y)$ $h_{Ia}(Y; \lambda_p)$

Table 2: Choices of W and corresponding notation for transformations based on (16).

Since the transformation in (16) has range \mathbb{R} for all λ_p , it admits an explicit inverse transformation. In addition, in the case of a single covariate, every estimated quantile that results will be monotone increasing, decreasing or constant, although different estimated quantiles can have different shapes from this collection. Geraci and Jones (2015) also proposed a transformation that unifies the symmetric and asymmetric versions of in two-parameter transformations, namely

$$h_{II}(W; \lambda_p) = h_I(W_{\delta_p}; \lambda_p) \tag{17}$$

where h_I is given in (16) and

$$W_{\delta_p} = h_{BC}(1 + W; \delta_p) = \begin{cases} \frac{(1 + W)^{\delta_p} - 1}{\delta_p} & \text{if } \delta_p > 0 \\ \log(1 + W) & \text{if } \delta_p = 0, \end{cases}$$

with $W = Y$ if $Y > 0$ and $W = Y/(1 - Y)$ if $Y \in (0, 1)$. The additional parameter δ_p controls the asymmetry: symmetric forms of h_I correspond to $\delta_p = 1$ and asymmetric forms of h_I to $\delta_p = 0$.

All transformation models discussed above can be fitted using a two-stage (TS) estimator (Chamberlain 1994; Buchinsky 1995) whereby $\beta(p)$ is estimated conditionally on a fine grid of values for the transformation parameter(s). Let us consider one-parameter transformations, although the case of two-parameter transformation is analogous. Briefly, the algorithm proceeds as follows:

1. Obtain the p th regression quantile $\hat{\beta}_p(\lambda)$ conditional on λ by minimising

$$\hat{\beta}_p(\lambda) = \underset{b \in \mathbb{R}^k}{\operatorname{argmin}} \sum_{i=1}^n \kappa_p \left(h(y_i; \lambda) - \mathbf{x}_i^\top \mathbf{b} \right),$$

where κ_p is the loss function defined in Section 3.1.

2. Estimate the parameter λ_p by

$$\hat{\lambda}_p = \underset{\lambda \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^n \kappa_p \left(y_i - h^{-1} \left\{ \mathbf{x}_i^\top \hat{\beta}_p(\lambda); \lambda \right\} \right).$$

Alternatively, point estimation can be approached using the RC process (Mu and He 2007), which is akin to the process that leads to the RC test introduced in Section 3.1. Similarly to the TS estimator, the RC estimator is based on a grid search for λ_p . However, the objective function to be minimized is defined as

$$n^{-1} \sum_i^n \{\mathbf{R}_n(\mathbf{x}_i; \lambda)\}^2$$

where $\mathbf{R}_n(\mathbf{t}; \lambda) = n^{-1} \sum_{j=1}^n \psi_p(\tilde{r}_j) I(\mathbf{x}_j \leq \mathbf{t})$ and $\tilde{r} = h(y; \lambda_p) - \mathbf{x}^\top \boldsymbol{\beta}_p$. This estimator avoids the troublesome inversion of the Box-Cox and Aranda-Ordaz transformations, but it is computationally more intensive than the TS estimator.

There are several methods for interval estimation, including those based on large- n approximations and the ubiquitous bootstrap. Both the TS and RC estimators have an asymptotic normal distribution. The large-sample properties of the TS estimator for monotonic quantile regression models have been studied by Powell (1991) (see also Chamberlain 1994; Machado and Mata 2000). Under regularity conditions, it can be shown that the joint estimator $\sqrt{n}(\hat{\boldsymbol{\beta}}_p^\top, \hat{\lambda}_p)$ will converge to a $k + 1$ -variate normal distribution with 0 mean and a sandwich-type limiting covariance matrix $p(1-p)H^{-1}DH^{-1}$, similar to (6), where the matrices H and D both depend on the gradient $\partial h^{-1}\{\mathbf{x}^\top \boldsymbol{\beta}_p; \lambda_p\} / \partial(\boldsymbol{\beta}_p^\top, \lambda_p)$ and the density of the residuals $r = y - h^{-1}\{\mathbf{x}^\top \boldsymbol{\beta}_p; \lambda_p\}$. In contrast, the form of the covariance matrix of the sampling distribution for the RC estimator is rather complicated and its estimation requires resampling (Mu and He 2007).

The estimation procedures discussed so far reduces to those for standard quantile regression problems when the transformation parameter is assumed to be known. In this case, conditional inference is apposite. Note, however, that standard errors will be underestimated if these are computed conditional on the transformation parameter when in fact the latter has been estimated from the data.

In **Qtools**, model fitting for one-parameter transformation models can be carried out using the function `tsrq`. The `formula` argument specifies the model for the linear predictor as in (10), while the argument `tsf` provides the desired transformation h as specified in Equations 13-16: `bc` for the Box-Cox model, `ao` for Aranda-Ordaz families, and `mcjI` for proposal I transformations. Additional arguments in the function `tsrq` include: `symmetry`, a logical flag to specify the symmetric or asymmetric version of `ao` and `mcjI`; `dbounded`, a logical flag to specify whether the response variable is doubly bounded (default is strictly positive, i.e., singly bounded); `lambda`, a numerical vector to define the grid of values for estimating λ_p ; and `conditional`, a logical flag indicating whether λ_p is assumed to be known (in which case the argument `lambda` provides such known value).

There are other functions to fit transformation models. The function `rcrq` fits one-parameter transformation models using the RC estimator. The functions `tsrq2` and `nlrq2` are specific to Geraci and Jones's (2015) Proposal II transformations. The former employs a two-way grid search while the latter is based on Nelder-Mead optimization as implemented in `optim`. Simulation studies in Geraci and Jones (2015) suggest that, although computationally slower, a two grid search is numerically more stable than the derivative-free approach.

A summary of the basic differences between all fitting functions is given in Table 3. The table also shows the available methods in `summary.rqt` to estimate standard errors and confidence intervals for the model's parameters. Unconditional inference is carried out jointly on $\boldsymbol{\beta}(p)$

and the transformation parameter by means of bootstrap using the package **boot** (Canty and Ripley 2014; Davison and Hinkley 1997). Large- n approximations (Powell 1991; Chamberlain 1994; Machado and Mata 2000) are also available for the one-parameter TS estimator under *iid* or *nid* assumptions.

When `summary.rqt` is executed with the argument `conditional = TRUE`, confidence interval estimation for β_p is performed with one of the several methods developed for linear quantile regression estimators (Koenker 2005, p.110) (see options `rank`, `iid`, `nid`, `ker`, and `boot` in `quantreg::summary.rq`).

Function name	Transformation parameters	Estimation	Standard errors or confidence intervals	
			Unconditional	Conditional
<code>tsrq</code>	1	Two-stage	<code>iid, nid, boot</code>	All types
<code>rcrq</code>	1	Residual process	<code>boot</code>	All types
<code>tsrq2</code>	2	Two-stage	<code>boot</code>	All types
<code>nlrq2</code>	2	Nelder–Mead	<code>boot</code>	–

Table 3: Transformation-based quantile regression in package **Qtools**. ‘All types’ consists of options `rank`, `iid`, `nid`, `ker`, and `boot` as provided by function `summary` in package **quantreg**.

In the New York Air Quality data example, a linear model was found unsuitable to describe the relationship between ozone and solar radiation. At closer inspection, Figure 5 reveals that the conditional distribution of ozone may in fact be nonlinearly associated with solar radiation, at least for some of the conditional quantiles. The model

$$Q_{h_{I_s}\{\text{ozone}\}}(p) = \beta_0(p) + \beta_1(p) \cdot \text{Solar.R}, \tag{18}$$

where h_{I_s} denotes the symmetric version of (16) for a singly bounded response variable, is fitted for the quantiles $p \in \{0.1, 0.5, 0.9\}$ using the following code:

```
R> system.time(fit.rqt <- tsrq(Ozone ~ Solar.R, data = dd, tsf = "mcjI",
+   symm = TRUE, dbounded = FALSE, lambda = seq(1, 3, by = 0.005),
+   conditional = FALSE, tau = c(.1, .5, .9)))
  user system elapsed
   0.5   0.0   0.5
R> fit.rqt
```

```
call:
tsrq(formula = Ozone ~ Solar.R, data = dd, tsf = "mcjI", symm = TRUE,
     dbounded = FALSE, lambda = seq(1, 3, by = 0.005), conditional = FALSE,
     tau = c(0.1, 0.5, 0.9))
```

Proposal I symmetric transformation (singly bounded response)

```
Optimal transformation parameter:
tau = 0.1 tau = 0.5 tau = 0.9
  2.210   2.475   1.500
```

```

Coefficients linear model (transformed scale):
              tau = 0.1  tau = 0.5  tau = 0.9
(Intercept) -3.3357578 -48.737341 16.557327
Solar.R      0.4169697   6.092168  1.443407

```

```
Degrees of freedom: 111 total; 109 residual
```

The TS estimator makes a search for λ_p over the grid 1.000, 1.005, ..., 2.995, 3.000. The choice of the search interval usually results from a compromise between accuracy and performance: the coarser the grid, the faster the computation, the less accurate the estimate. A reasonable approach would be to start with a coarse, wide-ranging grid (e.g., `seq(-5, 5, by = 0.5)`), then center the interval about the resulting estimate using a finer grid, and re-fit the model. The output above reports the estimates $\hat{\beta}(p)$ and $\hat{\lambda}_p$ for each quantile level specified in `tau`. Here, the quantities of interest are the predictions on the ozone scale and the marginal effect of solar radiation, which can be obtained using the function `predict.rqt`

```

R> x <- seq(9, 334, length = 200)
R> qhat <- predict(fit.rqt, newdata = data.frame(Solar.R = x),
+   type = "response")
R> dqhat <- predict(fit.rqt, newdata = data.frame(Solar.R = x),
+   type = "maref", namevec = "Solar.R")
The linear component of the marginal effect is calculated as derivative of
Ozone ~ beta1 * Solar.R
with respect to Solar.R

```

The calculations above are based on a 200-long sequence of ozone values in the interval [9, 334] Ly, as provided via the argument `newdata` (if this argument is missing, the function returns the fitted values). There are three types of predictions available:

1. `link`, predictions of conditional quantiles on the transformed scale (10), i.e., $\hat{Q}_{h(Y;\lambda_p)}(p) = \mathbf{x}^T \hat{\beta}(p)$.
2. `response`, predictions of conditional quantiles on the original scale (11), i.e., $\hat{Q}_{Y|X}(p) = h^{-1} \left\{ \mathbf{x}^T \hat{\beta}(p); \hat{\lambda}_p \right\}$.
3. `maref`, predictions of the marginal effect (12).

In the latter case, the user must specify the name of the covariate with respect to which the marginal effect has to be calculated using the argument `namevec`. The function `maref.rqt` computes derivatives symbolically using the `stats` function `deriv` and these are subsequently evaluated numerically. While the nonlinear component of the marginal effect in Equation 12 (i.e., $\partial Q(p)/\partial \eta(p)$) is rather straightforward to derive for any of the transformations (13)-(16), the derivative of the linear predictor (i.e., $\partial \eta(p)/\partial x_j$) requires parsing the `formula` argument in order to obtain an expression suitable for `deriv`. The function `maref.rqt` can handle simple expressions with common functions like `log`, `exp`, etc., interaction terms, and ‘as is’ terms (i.e., `I()`). However, using functions that are not recognized by `deriv` will trigger an error.

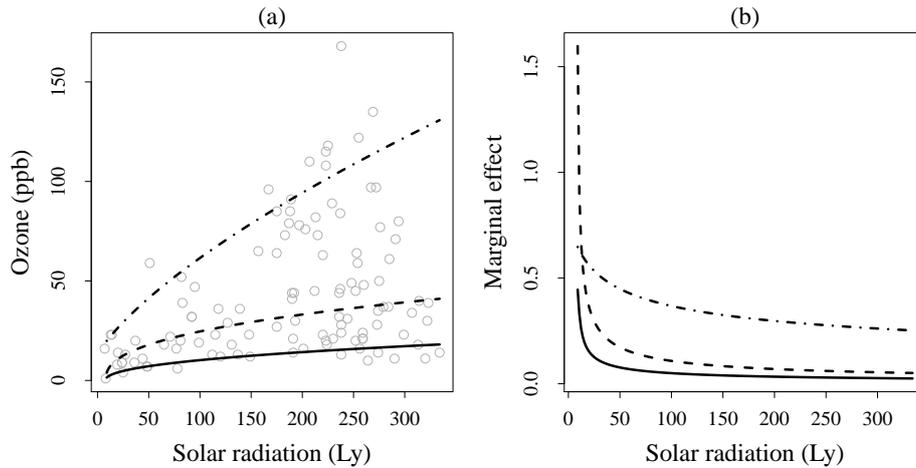


Figure 6: Predicted 10th (solid line), 50th (dashed line), and 90th (dot-dashed line) centiles of ozone conditional on solar radiation (a) and corresponding estimated marginal effects (b) using the symmetric proposal I transformation in the New York Air Quality data set.

The predicted quantiles of ozone and the marginal effects of solar radiation are plotted in Figure 6 using the following code:

```
R> par(mfrow = c(1, 2))
R> plot(Ozone ~ Solar.R, data = dd, xlab = "Solar radiation (lang)",
+      ylab = "Ozone (ppb)")
R> for(i in 1:3){lines(x, qhat[,i], lty = c(1, 2, 4)[i], lwd = 2)}
R> plot(range(x), range(dqhat), type = "n", xlab = "Solar radiation (lang)",
+      ylab = "Marginal effect")
R> for(i in 1:3){lines(x, dqhat[,i], lty = c(1, 2, 4)[i], lwd = 2)}
```

The effect of solar radiation on different quantiles of ozone levels shows a nonlinear behavior, especially at lower ranges of radiation (below 50 Ly) and on the median ozone. It might be worth testing the goodness-of-fit of the model. In Section 3.1, it was found evidence of lack of fit for the linear specification (8). In contrast, the output reported below indicates that, in general, the goodness of fit of the quantile models based on the transformation model (18) has improved since the test statistics are now smaller at all values of p . However, such improvement is still far from being sufficient for the median.

```
R> GOFTest(fit.rqt, alpha = 0.05, B = 1000, seed = 416)
```

Goodness-of-fit test for quantile regression based on the cusum process

Quantile 0.1: Test statistic = 0.0393; p-value = 0.077

Quantile 0.5: Test statistic = 0.1465; p-value = 0

Quantile 0.9: Test statistic = 0.0212; p-value = 0.289

The TS and RC estimators generally provide similar estimates and predictions. However, computation based on the cusum process tends to be somewhat slow, as shown further below. This is also true for the RC test provided by `GOFTest`.

```
R> system.time(fit.rqt <- rcrq(Ozone ~ Solar.R, data = dd, tsf = "mcjI",
+   symm = TRUE, dbounded = FALSE, lambda = seq(1, 3, by = 0.005),
+   tau = c(.1, .5, .9)))
   user system elapsed
36.88   0.03  37.64
```

An example using doubly bounded transformations is demonstrated using the A-level Chemistry Scores data set. The latter is available from **Qtools** and it consists of 31022 observations of A-level scores in Chemistry for England and Wales students, 1997. The data set also includes information of prior academic achievement as assessed with General Certificate of Secondary Education (GCSE) average scores. The goal is to evaluate the ability of GCSE to predict A-level scores. The latter are based on national exams in specific subjects (e.g., chemistry) with grades ranging from A to F. For practical purposes, scores are converted numerically as follows: A = 10, B = 8, C = 6, D = 4, E = 2, and F = 0. The response is therefore doubly bounded between 0 and 10. It should be noted that this variable is discrete, although, for the sake of simplicity, here it is assumed that the underlying process is continuous.

The model considered here is

$$Q_{h_{AOa}\{\text{score}\}}(p) = \beta_0(p) + \beta_1(p) \cdot \text{gcse}, \quad (19)$$

where h_{AOa} denotes the asymmetric Aranda-Ordaz transformation in (15). This model is fitted for $p = 0.9$:

```
R> data(Chemistry)
R> fit.rqt <- tsrq(score ~ gcse, data = Chemistry, tsf = "ao", symm = FALSE,
+   lambda = seq(0,2,by=0.01), tau = 0.9)
```

The predicted 90th centile of A-level scores conditional on GCSE and the marginal effect of GCSE are plotted in Figure 7. There is clearly a positive, nonlinear association between the

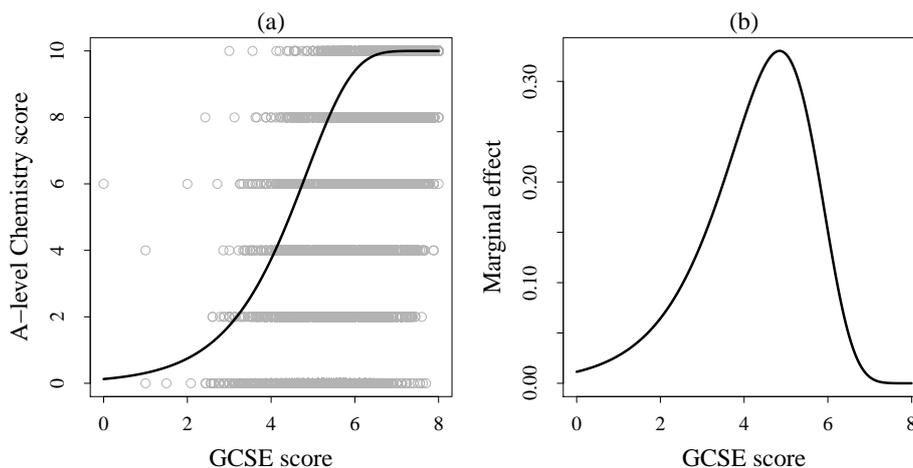


Figure 7: Predicted 90th centile of A-level scores conditional on GCSE scores (a) and corresponding estimated marginal effect (b) using the asymmetric Aranda-Ordaz transformation in the A-level Chemistry Scores data set.

two scores. The nonlinearity is partly explained by the floor and ceiling effects which result from the boundedness of the measurement scale. Note, however, that the S-shaped curve is not symmetric about the inflection point. As a consequence, the marginal effect is skewed to the left. Indeed, the estimate $\hat{\lambda}_{0.9} = 0$ and the narrow confidence interval give support to a complementary log-log transformation:

```
R> summary(fit.rqt, conditional = FALSE, se = "nid")
```

```
call:
```

```
summary.rqt(object = fit.rqt, se = "nid", conditional = FALSE)
```

```
Aranda-Ordaz asymmetric transformation (doubly bounded response)
```

```
Summary for unconditional inference
```

```
tau = 0.9
```

```
Optimal transformation parameter:
```

Value	Std. Error	Lower bound	Upper bound
0.000000000	0.001364422	-0.002674218	0.002674218

```
Coefficients linear model (transformed scale):
```

	Value	Std. Error	Lower bound	Upper bound
(Intercept)	-4.3520060	0.015414540	-4.3822179	-4.3217941
gcse	0.8978072	0.002917142	0.8920898	0.9035247

```
Degrees of freedom: 31022 total; 31020 residual
```

Alternatively, one can estimate the parameter δ_p using a two-parameter transformation:

```
R> coef(tsrq2(score ~ gcse, data = chemsub, dbounded = TRUE,
+ lambda = seq(0, 2, by = 0.1), delta = seq(0, 2, by = 0.1),
+ tau = 0.9), all = TRUE)
```

(Intercept)	gcse	lambda	delta
-4.1442274	0.8681246	0.0000000	0.0000000

These results confirm the asymmetric nature of the relationship since $\hat{\delta}_{0.9} = 0$. Similar results (not shown) were obtained with `nlrq2`.

3.3. Conditional LSS

Quantile-based measures of location, scale, and shape can be assessed *conditionally* on covariates. A simple approach is to fit a linear model as in (5) or a transformation-based model as in (10), and then predict $\hat{Q}_{Y|X}(p)$ to obtain the conditional LSS measures in Equation 4 for specific values of \mathbf{x} .

Estimation of conditional LSS can be carried out by using the function `qlss.formula`. The conditional model is specified in the argument `formula`, while the probability p is given in

`probs`. (As seen in Equation 4, the other probabilities of interest to obtain the decomposition of the conditional quantiles are $1 - p$, 0.25, 0.5, and 0.75.) The argument `type` specifies the required type of regression model, more specifically `rq` for linear models and `rqt` for transformation-based models (see Section 3.2). The function `qlss.formula` will take any additional argument to be passed to `quantreg::rq` or `tsrq` (e.g., `subset`, `weights`, etc.).

Let's consider the New York Air Quality data example discussed in the previous section and assume that the transformation model (18) holds for the quantiles $p \in \{0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95\}$. Then the conditional LSS of the distribution of ozone conditional on solar radiation for $p = 0.05$ and $p = 0.1$ is predicted by

```
R> fit.qlss <- qlss(formula = Ozone ~ Solar.R, data = airquality, type =
+   "rqt", tsf = "mcjI", symm = TRUE, dbounded = FALSE, lambda =
+   seq(1, 3, by = 0.005), probs = c(0.05, 0.1))
R> fit.qlss
```

call:

```
qlss.formula(formula = Ozone ~ Solar.R, probs = c(0.05, 0.1),
  data = airquality, type = "rqt", tsf = "mcjI", symm = TRUE,
  dbounded = FALSE, lambda = seq(1, 3, by = 0.005))
```

Conditional Quantile-Based Location, Scale, and Shape

-- Values are averaged over observations --

**** Location ****

Median

[1] 30.2258

**** Scale ****

Inter-quartile range (IQR)

[1] 43.40648

Inter-quantile range (IPR)

0.05 0.1

88.02909 73.93430

****Shape****

Skewness index

0.05 0.1

0.5497365 0.5180108

Shape index

0.05 0.1

1.960315 1.661648

The output, which is of class 'qlss', is a named list with the same LSS measures seen in the case of unconditional quantiles. However, these are now conditional on solar radiation. By default, the predictions are the fitted values, which are averaged over observations for printing purposes. An optional data frame for predictions can be given via the argument `newdata` in `predict.qlss`. If `interval = TRUE`, the latter computes confidence intervals at the specified `level` using R bootstrap replications (it is, therefore, advisable to set the seed before calling `predict.qlss`). The conditional LSS measures can be conveniently plotted

using the `plot.qlss` function as shown in the code below. The argument `z` is required and specifies the covariate used for plotting. Finally, the argument `which` specifies which of the probabilities given in `probs` should be used (in this example, `which = 2` corresponds to $p = 0.1$).

```
R> set.seed(567)
R> x <- seq(9, 334, length = 200)
R> qhat <- predict(fit.qlss, newdata = data.frame(Solar.R = x),
+   interval = TRUE, level = 0.90, R = 500)
R> plot(qhat, z = x, which = 2, interval = TRUE, type = "l",
+   xlab = "Solar radiation (lang)", lwd = 2)
```

Figure 8 shows that both the median and the IQR of ozone increase nonlinearly with increasing solar radiation. The distribution of ozone is skewed to the right and the degree of asymmetry is highest at low values of solar radiation. This is due to the extreme curvature of the median which takes on values close to the 10th centile (Figure 6). However, the sparsity of observations at the lower end of the observed range of solar radiation determines substantial uncertainty as reflected by the wider confidence interval (Figure 8). At $p = 0.1$, the conditional shape index is on average equal to 1.66 and it increases monotonically from 1.32 to about 1.85, remaining always below the tail-weight threshold of a normal distribution (1.90).

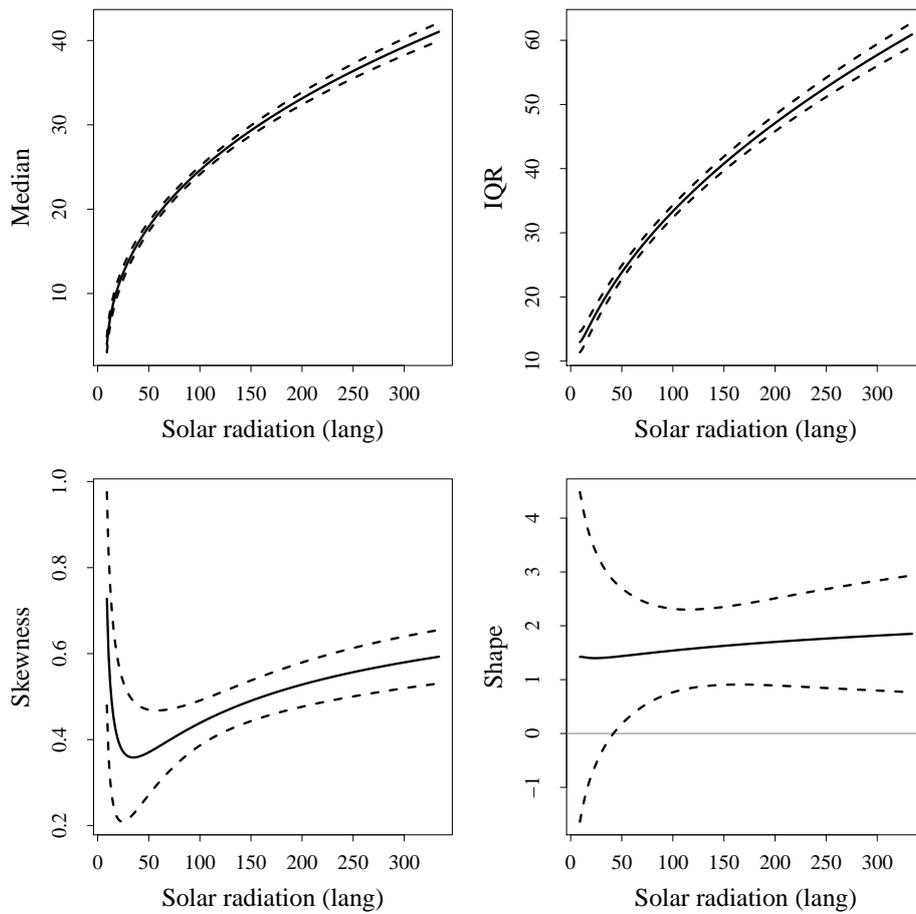


Figure 8: Location, scale and shape of ozone levels conditional on solar radiation in the New York Air Quality data set. Dashed lines denote the bootstrapped 90% point-wise confidence intervals.

4. Other functions in **Qtools**

4.1. Restricted quantile regression

Besides a loss of precision, high sparsity (low density) might also lead to a violation of the basic property of monotonicity of quantile functions. Quantile crossing occurs when $\mathbf{x}_i^\top \hat{\beta}(p) > \mathbf{x}_i^\top \hat{\beta}(p')$ for some \mathbf{x}_i and $p < p'$. This problem typically occurs in the outlying regions of the design space (Koenker 2005) where also sparsity occurs more frequently. Balanced designs with larger sample sizes would then offer some assurance against quantile crossing, provided, of course, that the QR models are correctly specified (see Section 3.1). Model's misspecification, indeed, can still be a cause of crossing of the quantile curves. Restricted regression quantiles (RRQ) (He 1997) might offer a practical solution when little can be done in terms of modeling. This approach applies to a subclass of linear models

$$Y = \mathbf{x}^\top \beta + \epsilon$$

and linear heteroscedastic models

$$Y = \mathbf{x}^\top \beta + (\mathbf{x}^\top \gamma) \epsilon$$

where $\mathbf{x}^\top \gamma > 0$ and $\epsilon \sim F$. Basically, it consists in fitting a reduced regression model passing through the origin. The reader is referred to He (1997) for details. Here, it is worth stressing that when the restriction does not hold, i.e., if the model is more complex than a location–scale–shift model, then RRQ may yield unsatisfactory results He (1997). See also Zhao (2000) for an examination of the asymptotic properties of the restricted QR estimator. In particular, the relative efficiency of RRQ as compared to RQ depends on the error distribution. For some common unimodal distributions, Zhao (2000) showed that RRQ in *iid* models is more efficient than RQ. This property is lost when the error is asymmetric. In contrast, the efficiency of RRQ in heteroscedastic models is comparable to that of RQ even for small samples.

The package **Qtools** provides the functions `rrq`, `rrq.fit` and `rrq.wfit` which are, respectively, the ‘restricted’ analogous of `rq`, `rq.fit`, and `rq.wfit` in **quantreg**. S3 methods `print`, `coef`, `predict`, `fitted`, `residuals`, and `summary` are available for objects of class `rrq`. In particular, confidence intervals are obtained using the functions `boot` and `boot.ci` from package **boot** (Canty and Ripley 2014; Davison and Hinkley 1997). Future versions of the package will develop the function `summary.rrq` to include asymptotic standard errors (Zhao 2000).

An application is shown below using an example discussed by Zhao (2000). The data set, available from **Qtools**, consists of 118 measurements of esterase concentrations and number of bindings counted in binding experiments.

```
R> data("esterase")
R> taus <- c(.1, .25, .5, .75, .9)
R> fit.rq <- rq(Count ~ Esterase, data = esterase, tau = taus)
R> yhat1 <- fitted(fit.rq)
R> fit.rrq <- rrq(Count ~ Esterase, data = esterase, tau = taus)
R> yhat2 <- fitted(fit.rrq)
```

The predicted 90th centile curve crosses the 50th and 75th curves at lower esterase concentrations (Figure 9). The crossing is removed in predictions based on RRQs.

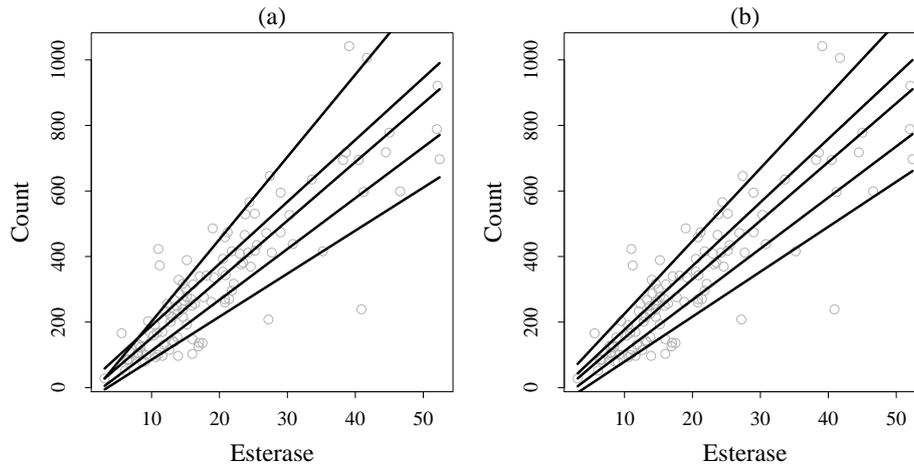


Figure 9: Predicted quantiles of number of bindings conditional on esterase concentration using regression quantiles (a) and restricted regression quantiles (b) in the Esterase data set.

As discussed above, the reliability of the results depends on the validity of the restriction carried by RRQ. A quick check can be performed using the location–scale–shift specification of the Khmaladze test discussed in Section 3.1 as shown below.

```
R> kt <- KhmaladzeTest(formula = Count ~ Esterase, data = esterase,
+ taus = seq(.05,.95,by = .01), nullH = "location-scale")
R> KhmaladzeFormat(kt, 0.05)
```

```
Khmaladze test for the location-shift hypothesis
Joint test is not significant at 10% level
Test(s) for individual slopes:
not significant at 10% level
```

4.2. Multiple imputation

Regression models play an important role in conditional imputation of missing values. QR can be used as an effective approach for multiple imputation (MI) when location-shift models are inadequate (Muñoz and Rueda 2009; Bottai and Zhen 2013; Geraci 2013).

In **Qtools**, `mice.impute.rq` and `mice.impute.rrq` are auxiliary functions written to be used along with the functions of the R package **mice** (van Buuren and Groothuis-Oudshoorn 2011). The former is based on the standard QR estimator (`rq.fit`) while the latter on the restricted counterpart (`rrq.fit`). Both imputation functions allow for the specification of the transformation-based QR models described in Section 3.2. The equivariance property is therefore very useful not only to achieve linearity of the conditional model but also to ensure that imputations lie within some interval when the values being imputed are singly or doubly bounded.

An example available from `?mice.impute.rq` using the `nhanes` data set is reported below. Missing values in continuous variables are imputed using quantile regression, while those in categorical variables are imputed using multinomial and binomial regressions.

```
R> require("mice")
R> data("nhanes")
R> nhanes2 <- nhanes
R> nhanes2$hyp <- as.factor(nhanes2$hyp)
R> set.seed(199)
R> imp <- mice(nhanes2, meth = c("polyreg", "rq", "logreg", "rq"), m = 5)
```

A linear regression model is then fitted for each of the five imputed data sets and results are pooled using Rubin's rules.

```
R> fit <- lm.mids(bmi ~ hyp + chl, data = imp)
R> pool(fit)
```

```
Call: pool(object = fit)
```

```
Pooled coefficients:
(Intercept)      hyp2      chl
23.09613686  0.79895370  0.01652075
```

```
Fraction of information about the coefficients missing due to nonresponse:
(Intercept)      hyp2      chl
  0.3626414  0.6585308  0.4835364
```

4.3. Conditional quantiles of discrete data

Modeling discrete response variables, such as categorical and count responses, has been traditionally approached with distribution-based methods: a parametric model $F_{Y|X}(y; \theta)$ is assumed and then fitted by means of MLE. Binomial, negative binomial, multinomial and Poisson regressions are well-known in many applied sciences. Because of the computational advantages and the asymptotic properties of MLE, these methods have long ruled among competing alternatives.

Modeling conditional functions of discrete data is less common and, on a superficial level, might even appear as an unnecessary complication. However, a deeper look at its rationale will reveal that a distribution-free analysis can provide insightful information in the discrete case as it does in the continuous case. Indeed, methods for conditional quantiles of continuous distributions can be—and have been—adapted to discrete responses.

The package **Qtools** offers some limited functionalities for count and binary data. Further research is needed to develop the theory of QR for discrete data and to improve computational algorithms. Therefore, the user should use these functions with caution.

Let Y be a count variable such as, for example, the number of car accidents during a week or the number of visits of a patient to their doctor during a year. As usual, X denotes a vector of covariates. Poisson regression, which belongs to the family of generalized linear models (GLMs), is a common choice for this kind of data, partly because of its availability in many statistical packages. Symbolically, $Y \sim \text{Pois}(\theta)$ where

$$\theta \equiv E(Y|X = \mathbf{x}) = h^{-1}(\mathbf{x}^\top \boldsymbol{\beta})$$

and h is the logarithmic link function. Note that the variance also is equal to θ . Indeed, moments of order higher than 2 governing the shape of the distribution depend on the same parameter. Every component of the conditional LSS in a Poisson model is therefore controlled by θ . If needed, more flexibility can be achieved using a distribution-free approach.

Machado and Santos Silva (2005) proposed the model

$$Q_{h(Z;p)}(p) = \mathbf{x}^\top \boldsymbol{\beta}(p). \quad (20)$$

where $Z = Y + U$ is obtained by jittering Y with a $[0, 1)$ -uniform noise U , independent of Y and X . In principle, any monotone transformation h can be considered. A natural choice for count data is a log-linear model (Machado and Santos Silva 2005), i.e.,

$$h(Z; p) = \begin{cases} \log(Z - p) & \text{for } Z > p \\ \log \zeta & \text{for } Z \leq p. \end{cases}$$

where $0 < \zeta < p$. It follows that $Q_{Z|X}(p) = p + \exp(\mathbf{x}^\top \boldsymbol{\beta}(p))$. (Note that the p th quantile of the conditional distribution of Z is bounded below by p .) Given the continuity between counts induced by jittering, standard inference for linear quantile functions (Koenker and Bassett 1978) can be applied to fit (18). In practice, a sample of M jittered responses Z is taken to estimate $\hat{\boldsymbol{\beta}}_m(p)$, $m = 1, \dots, M$; the noise is then averaged out, $\hat{\boldsymbol{\beta}}(p) = \frac{1}{M} \sum_m \hat{\boldsymbol{\beta}}_m(p)$.

Machado and Santos Silva's (2005) methods, including large- n approximations for standard errors, are implemented in the function `rq.counts`. The `formula` argument specifies a linear model as in (18), while the argument `tsf` provides the desired transformation h . By default, this is the log transformation (i.e., Box-Cox with parameter $\lambda_p = 0$) but other transformations described in Section 3.2 are allowed. Note that `GOFtest` can be applied to `rq.counts` objects as well.

Qtools provides functions for modeling binary responses as well. First of all, it is useful to note that the classical GLM for a binary response $Y \sim \text{Bin}(1, \pi)$ establishes a relationship between the *probability* $\Pr(Y = 1) = \pi$ and a set of predictors \mathbf{x} , that is

$$\pi \equiv \mathbb{E}(Y|X = \mathbf{x}) = h^{-1}(\mathbf{x}^\top \boldsymbol{\beta}).$$

Common choices for the link function $h : (0, 1) \rightarrow \mathbb{R}$ are logit, probit and c-log-log, which can be considered special cases of the Aranda-Ordaz families of transformations. Another way to formulate the regression problem above is to consider the continuous latent variable Y^* , where

$$Y^* = \mathbf{x}^\top \boldsymbol{\beta} + \epsilon, \quad (21)$$

and assume that the binary observations are the result of the dichotomization $Y = I(Y^* > 0)$. Maximum score estimation, originally developed by Manski (1975, 1985), is equivalent to estimating the conditional quantiles of the latent variable Y^* in (21). Using the same notation introduced in Section 3.1, the problem to be solved is given by

$$\min_{\mathbf{b} \in \mathbb{R}^k} \sum_{i=1}^n \kappa_p \left(y_i - I(\mathbf{x}_i^\top \mathbf{b} > 0) \right). \quad (22)$$

Since the indicator function is a monotone transformation, the Q-transformation rule applies. Indeed, $I(Q_{Y^*|X}(p) > 0) = Q_{I(Y^*) > 0|X}(p) = Q_{Y|X}(p)$. In practice, the goal is to find \mathbf{b} such that the number of matches between observed and fitted 0's and 1's is as large as possible. However, the minimization problem in (22) offers numerical challenges due to the the piecewise linearity of the indicator function and the nonconvexity of the loss function. Smoothed approximations of $I(\cdot)$ and simulated annealing algorithms have been suggested (Horowitz 1992; Kordas 2006).

The function `rq.bin` is the main function to obtain binary regression quantiles. It is a wrapper for the function `rqbin.fit` which calls Fortran code written for simulated annealing estimation (Goffe, Ferrier, and Rogers 1994). **Qtools** offers general methods for objects of class `rq.bin` including `coef` and `predict`. In particular, the commands `predict(object, type = "latent")` and `predict(object, type = "probability")` provide predictions for, respectively, the latent quantiles $\mathbf{x}_i^T \boldsymbol{\beta}(p)$ and the individual probabilities $\Pr(Y_i^* > 0)$, $i = 1, \dots, n$. A `summary` method to calculate standard errors has not yet been implemented.

5. Final remarks

Quantiles have long occupied an important place in statistics. The package **Qtools** builds on recent methodological and computational developments of quantile functions and related methods to promote their application in statistical data modeling.

Acknowledgements

This work is partially supported by an ASPIRE grant from the Office of the Vice President for Research at the University of South Carolina. The maintainers of the UCI Machine Learning Repository (Lichman 2013) are gratefully acknowledged for making the Wine Quality data set available.

References

- Aranda-Ordaz FJ (1981). "On Two Families of Transformations to Additivity for Binary Response Data." *Biometrika*, **68**(2), 357–363.
- Azzalini A, Bowman AW (1990). "A Look at Some Data on the Old Faithful Geyser." *Journal of the Royal Statistical Society C*, **39**(3), 357–365.
- Bassett G, Koenker R (1978). "Asymptotic Theory of Least Absolute Error Regression." *Journal of the American Statistical Association*, **73**(363), 618–622.
- Benjamini Y, Krieger AM (1996). "Concepts and Measures for Skewness with Data-Analytic Implications." *Canadian Journal of Statistics*, **24**(1), 131–140.
- Benoit DF, Al-Hamzawi R, Yu K, den Poel DV (2014). *bayesQR: Bayesian Quantile Regression*. R package version 2.2, URL <http://CRAN.R-project.org/package=bayesQR>.
- Bottai M, Zhen H (2013). "Multiple Imputation Based on Conditional Quantile Estimation." *Epidemiology, Biostatistics, and Public Health*, **10**(1), e8758.

- Box GEP, Cox DR (1964). “An Analysis of Transformations.” *Journal of the Royal Statistical Society B*, **26**(2), 211–252.
- Buchinsky M (1995). “Quantile Regression, Box-Cox Transformation Model, and the US Wage Structure, 1963-1987.” *Journal of Econometrics*, **65**(1), 109–154.
- Canty A, Ripley BD (2014). *boot: Bootstrap R (S-PLUS) Functions*. R package version 1.3-15, URL <http://CRAN.R-project.org/package=boot>.
- Chamberlain G (1994). *Quantile Regression, Censoring, and the Structure of Wages*, volume 1. Cambridge University Press, Cambridge, UK.
- David HA (1995). “First (?) Occurrence of Common Terms in Mathematical Statistics.” *The American Statistician*, **49**(2), 121–133.
- Davison AC, Hinkley DV (1997). *Bootstrap Methods and Their Applications*. Cambridge University Press, Cambridge.
- Dehbi HM, Cortina-Borja M, Geraci M (2016). “Aranda-Ordaz Quantile Regression for Student Performance Assessment.” *Journal of Applied Statistics*, **43**(1), 58–71.
- Doksum K (1974). “Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case.” *The Annals of Statistics*, **2**(2), 267–277.
- Galton F (1875). “IV. Statistics by Intercomparison, With Remarks on the Law of Frequency of Error.” *Philosophical Magazine Series 4*, **49**(322), 33–46.
- Galton F (1882). “Report of the Anthropometric Committee.” In *Report of the 51st Meeting of the British Association for the Advancement of Science, 1881*, pp. 245–260.
- Galton F (1885). “Some Results of the Anthropometric Laboratory.” *The Journal of the Anthropological Institute of Great Britain and Ireland*, **14**, 275–287.
- Geraci M (2013). “Estimation of Regression Quantiles in Complex Surveys With Data Missing at Random: An Application to Birthweight Determinants.” *Statistical Methods in Medical Research*. doi:10.1177/0962280213484401.
- Geraci M (2014). “Linear Quantile Mixed Models: The **lqmm** Package for Laplace Quantile Regression.” *Journal of Statistical Software*, **57**(13), 1–29.
- Geraci M, Bottai M (2014). “Linear Quantile Mixed Models.” *Statistics and Computing*, **24**(3), 461–479.
- Geraci M, Jones MC (2015). “Improved Transformation-Based Quantile Regression.” *Canadian Journal of Statistics*, **43**(1), 118–132.
- Gilchrist W (2000). *Statistical Modelling with Quantile Functions*. Chapman & Hall/CRC, Boca Raton, FL.
- Goffe WL, Ferrier GD, Rogers J (1994). “Global Optimization of Statistical Functions With Simulated Annealing.” *Journal of Econometrics*, **60**(1), 65–99.

- Groeneveld RA (1998). “A Class of Quantile Measures for Kurtosis.” *The American Statistician*, **52**(4), pp. 325–329.
- Groeneveld RA, Meeden G (1984). “Measuring Skewness and Kurtosis.” *Journal of the Royal Statistical Society D*, **33**(4), pp. 391–399.
- Hald A (1998). *A History of Mathematical Statistics from 1750 to 1930*. John Wiley & Sons, New York, NY.
- Hald A (2003). *A History of Probability and Statistics and their Applications before 1750*. John Wiley & Sons, New York, NY.
- Hankin RKS, Lee A (2006). “A New Family of Non-Negative Distributions.” *Australia and New Zealand Journal of Statistics*, **48**, 67–78.
- He X (1997). “Quantile Curves without Crossing.” *The American Statistician*, **51**(2), 186–192.
- He XM, Zhu LX (2003). “A Lack-of-fit Test for Quantile Regression.” *Journal of the American Statistical Association*, **98**(464), 1013–1022.
- Horn PS (1983). “A Measure for Peakedness.” *The American Statistician*, **37**(1), 55–56.
- Horowitz JL (1992). “A Smoothed Maximum Score Estimator for the Binary Response Model.” *Econometrica*, **60**(3), 505–531.
- Hyndman RJ, Fan Y (1996). “Sample Quantiles in Statistical Packages.” *The American Statistician*, **50**(4), 361–365.
- Jones MC (2007). “Connecting Distributions with Power Tails on the Real Line, the Half Line and the Interval.” *International Statistical Review*, **75**(1), 58–69.
- Jones MC, Rosco JF, Pewsey A (2011). “Skewness-invariant Measures of Kurtosis.” *The American Statistician*, **65**(2), 89–95.
- Kendall MG (1940). “Note on the Distribution of Quantiles for Large Samples.” *Supplement to the Journal of the Royal Statistical Society*, **7**(1), 83–85.
- Khmaladze EV, Koul HL (2004). “Martingale Transforms Goodness-of-fit Tests in Regression Models.” pp. 995–1034.
- Koenker R (2005). *Quantile Regression*. Cambridge University Press, New York, NY.
- Koenker R (2013). *quantreg: Quantile Regression*. R package version 5.05, URL <http://CRAN.R-project.org/package=quantreg>.
- Koenker R, Bassett G (1978). “Regression Quantiles.” *Econometrica*, **46**(1), 33–50.
- Koenker R, Bassett G (1985). “On Boscovich’s Estimator.” *The Annals of Statistics*, **13**(4), 1625–1628.
- Koenker R, D’Orey V (1987). “Algorithm AS 229: Computing Regression Quantiles.” *Journal of the Royal Statistical Society C*, **36**(3), 383–393.

- Koenker R, Machado JAF (1999). “Goodness of Fit and Related Inference Processes for Quantile Regression.” *Journal of the American Statistical Association*, **94**(448), 1296–1310.
- Koenker R, Park BJ (1996). “An Interior Point Algorithm for Nonlinear Quantile Regression.” *Journal of Econometrics*, **71**(1-2), 265–283.
- Koenker R, Xiao ZJ (2002). “Inference on the Quantile Regression Process.” *Econometrica*, **70**(4), 1583–1612.
- Kordas G (2006). “Smoothed Binary Regression Quantiles.” *Journal of Applied Econometrics*, **21**(3), 387–407.
- Lehmann EL (1975). *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day, San Francisco, CA.
- Lichman M (2013). “UCI Machine Learning Repository.” URL <http://archive.ics.uci.edu/ml>.
- Ma Y, Genton MG, Parzen E (2011). “Asymptotic Properties of Sample Quantiles of Discrete Distributions.” *Annals of the Institute of Statistical Mathematics*, **63**(2), 227–243.
- Machado JAF, Mata J (2000). “Box-Cox Quantile Regression and the Distribution of Firm Sizes.” *Journal of Applied Econometrics*, **15**(3), 253–274.
- Machado JAF, Santos Silva JMC (2005). “Quantiles for Counts.” *Journal of the American Statistical Association*, **100**(472), 1226–1237.
- Manski CF (1975). “Maximum Score Estimation of the Stochastic Utility Model of Choice.” *Journal of Econometrics*, **3**(3), 205–228.
- Manski CF (1985). “Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator.” *Journal of Econometrics*, **27**(3), 313–333.
- Mu YM, He XM (2007). “Power Transformation Toward a Linear Regression Quantile.” *Journal of the American Statistical Association*, **102**(477), 269–279.
- Muñoz JF, Rueda M (2009). “New Imputation Methods for Missing Data Using Quantiles.” *Journal of Computational and Applied Mathematics*, **232**(2), 305–317.
- Nadarajah S, Rocha R (2016). “**Newdistns**: An R Package for New Families of Distributions.” *Journal of Statistical Software*, **69**(10), 1–32.
- Oberhofer W, Haupt H (2005). “The Asymptotic Distribution of the Unconditional Quantile Estimator Under Dependence.” *Statistics & Probability Letters*, **73**(3), 243–250.
- Parente PMDC, Santos Silva JMC (2016). “Quantile Regression With Clustered Data.” *Journal of Econometric Methods*, **5**(1), 1–16.
- Parzen E (1979). “Nonparametric Statistical Data Modeling.” *Journal of the American Statistical Association*, **74**(365), 105–121.
- Parzen E (2004). “Quantile Probability and Statistical Data Modeling.” *Statistical Science*, **19**(4), 652–662.

- Powell JL (1991). *Estimation of Monotonic Regression Models Under Quantile Restrictions*, pp. 357–384. Cambridge University Press, New York, NY.
- Powell JL (1994). *Estimation of Semiparametric Models*, volume Volume 4, chapter 41, pp. 2443–2521. Elsevier.
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Ruppert D (1987). “What is Kurtosis?: An Influence Function Approach.” *The American Statistician*, **41**(1), 1–5.
- Sakia RM (1992). “The Box-Cox Transformation Technique: A Review.” *Journal of the Royal Statistical Society D*, **41**(2), 169–178.
- Smith L, Reich B (2013). *BSquare: Bayesian Simultaneous Quantile Regression*. R package version 1.1, URL <http://CRAN.R-project.org/package=BSquare>.
- Staudte RG (2014). “Inference for Quantile Measures of Kurtosis, Peakedness and Tail-weight.” *arXiv preprint arXiv:1047.6461v1 [math.ST]*. doi:10.1080/03610926.2015.1056366.
- Tukey JW (1965). “Which Part of the Sample Contains the Information?” *Proceedings of the National Academy of Sciences of the United States of America*, **53**(1), 127–134.
- van Buuren S, Groothuis-Oudshoorn K (2011). “**mice**: Multivariate Imputation by Chained Equations in R.” *Journal of Statistical Software*, **45**(3), 1–67.
- Wang JD (1995). “Asymptotic Normality of L_1 -Estimators in Nonlinear Regression.” *Journal of Multivariate Analysis*, **54**(2), 227–238.
- Yu KM, Jones MC (1998). “Local Linear Quantile Regression.” *Journal of the American Statistical Association*, **93**(441), 228–237.
- Zhao QS (2000). “Restricted Regression Quantiles.” *Journal of Multivariate Analysis*, **72**(1), 78–99.
- Zheng JX (1998). “A Consistent Nonparametric Test of Parametric Regression Models under Conditional Quantile Restrictions.” *Econometric Theory*, **14**(1), 123–138.

Affiliation:

Marco Geraci
Department of Epidemiology and Biostatistics
Arnold School of Public Health
University of South Carolina
915 Greene Street, Columbia SC 29208, USA
E-mail: geraci@mailbox.sc.edu
URL: <http://marcogeraci.wordpress.com/>