VGAM Family Functions for Quantile Regression

T. W. Yee

February 25, 2008

Beta Version 0.7-6

Department of Statistics,
University of Auckland,
New Zealand
yee@stat.auckland.ac.nz
http://www.stat.auckland.ac.nz/∼yee

Contents

[Important note: This document and code is not yet finished, but should be completed one day . . .]

1 Introduction

This document describes in detail VGAM family functions for quantile and expectile regression. Many of VGAM's features come from `glm()` and `gam()` so that readers unfamiliar with these functions are referred to Chambers and Hastie (1993). Additionally, the VGAM *User Manual* should be consulted for general instructions about the software.

Most VGAM family functions for quantile regression are based on the LMS idea, which is described in Section 2. Another method, called expectile regression, is very much related to quantile regression and is described in Section 3. The rest of this section is largely background reading.

1.1 Some Notation and Background

In this documentation we use the words *quantile*, *centile*, and *percentile* interchangeably, and note that, for example, a 0.5 quantile is equivalent to 50-percentile, which is the median. VGAM uses the argument `percentile` to input quantiles, so it should be assigned values between 0 and 100. Related terms are *quartiles*, *quintiles* and *deciles*, which divide the distribution into 4, 5 and 10 equal parts respectively. The word *expectile* is not used synonomously with quantile—see later. Table 1 summarizes the notation used in this article.

Suppose a real-valued random variable Y has cumulative distribution function (cdf) $F(y) = P(Y \leq y)$. Then the τth quantile of Y is defined to be

$$Q(\tau) \quad = \quad \inf\{y : \tau \leq F(y)\} \tag{1}$$

where $0 < \tau < 1$. Thus a proportion τ lies below $Q(\tau)$ and $1 - \tau$ lies above $Q(\tau)$. Like the cdf, the quantile function $Q(\tau)$ *completely* describes the distribution of Y; there is no information loss. This contrasts sharply with ordinary regression where usually only the mean of Y is modelled.

Quantiles have equivariance to monotone transformations: if ψ is a nondecreasing function on $\mathbb{R}$ then

$$Q_{\psi(Y)}(\tau) \;\; = \;\; \psi(Q_Y(\tau)). \tag{2}$$

This means the quantiles of the transformed variable $\psi(Y)$ are the transformed quantiles of the original variable Y.

Applications of quantile regression come from many fields. Here are some.

- Popular medical examples include investigating height, weight, body mass index (BMI) as a function of age of the person, e.g., Royston (1991), Gasser et al. (1994), Royston and Wright (1998), Cole et al. (1998), Heagerty and Pepe (1999), Wei et al. (2006) to name a few. Historically, the construction of 'growth charts' was probably the first example of age-related reference intervals. Another example is Campbell and Newman (1971)—the ultrasonographic assessment of fetal growth has become clinically routine. Many researchers have proposed gestational-age-specific centile charts and tables for a variety of relevant measurements, e.g., Chitty et al. (1994).

- Economics, e.g.,it has been used to study determinants of wages, discrimination effects, and trends in income inequality. See Koenker and Hallock (2001) and Koenker (2005) for more references.

- Education, e.g., the performance of students in public schools on standardized exams as a function of socio-economic variables such as parents' income and educational attainment. Other variables are policy variables, e.g., class size, teacher qualifications. It is unrealistic to believe that covariates should shift the distribution of the exam results by a fixed amount. We wish to determine whether policy variables alter the performance of the best students in the same way that weaker students are affected. See Koenker and Hallock (2001) for a substantial list of references in education too.

- Climate data, e.g., Figure 2 plots some Melbourne temperature data which was examined by Hyndman et al. (1996) and shown to exhibit bimodal behaviour.

Koenker (2005) is a very accessible book on both the theory and practice of quantile regression. Another book is Hao and Naiman (2007). Other literature include the reviews of Wright and Royston (1997), Buchinsky (1998), Koenker and Hallock (2001) and Yu et al. (2003); see also Cizek (2000).

1.2 Classical Quantile Regression

There have been quite a number of methods that have been developed for quantile regression. One particular popular method of solution, which shall be referred to in this document as the "classical" method, is now described. It serves as the basis for expectile regression (Sections 1.3 and 3). It may be shown that quantiles can be formulated in terms of a minimization problem: minimize with respect to ξ the expectation of $\rho_\tau(Y - \xi)$ where

$$\rho_\tau(u) \;\; = \;\; u(\tau - I(u < 0)) \tag{3}$$

Table 1: *Notation and jargon used in this article. Data is* $(y_i, \boldsymbol{x}_i)$, $i = 1, \ldots, n$.

Notation	Comments
Y	Response. Has mean μ, cdf $F(y)$, pdf $f(y)$
$Q_Y(\tau) = \tau$-quantile of Y	$0 < \tau < 1$
$\xi(\tau) = \xi_\tau = \tau$th quantile	Koenker and Bassett (1978), $\xi(\frac{1}{2}) =$ median
$\mu(\omega) = \mu_\omega = \omega$-expectile	$0 < \omega < 1$, $\mu(\frac{1}{2}) = \mu$, Newey and Powell (1987)
$\widehat{\xi}(\tau)$, $\widehat{\mu}(\omega)$	Sample quantiles and expectiles
centile	Same as quantile and percentile here
regression quantile	Koenker and Bassett (1978)
regression expectile	Newey and Powell (1987)
regression percentile	All forms of asymmetric fitting (Efron, 1992)
$\rho_\tau(u) = u \cdot (\tau - I(u < 0))$	Check function corresponding to $\xi(\tau)$
$\rho_\omega^{[2]}(u) = u^2 \cdot \lvert\omega - I(u < 0)\rvert$	Check function corresponding to $\mu(\omega)$
$u_+ = \max(u, 0)$	Positive part of u
$u_- = \min(u, 0)$	Negative part of u

is known as a *check* function. That is, we minimize

$$E\left[\rho_\tau(Y - \xi)\right] \;=\; (\tau - 1)\int_{-\infty}^{\xi} (y - \xi)\, dF(y) + \tau \int_{\xi}^{\infty} (y - \xi)\, dF(y). \tag{4}$$

Using the formula

$$\frac{\partial}{\partial x}\int^{x} h(t, x)\, dt \;=\; \int^{x} \frac{\partial\, h(t, x)}{\partial x}\, dt$$

for a function h, we set the derivative with respect to ξ equal to zero to give

$$\begin{aligned} 0 \;&=\; (\tau - 1)\int_{-\infty}^{\xi} (-1)\, dF(y) + \tau \int_{\xi}^{\infty} (-1)\, dF(y) \qquad (5)\\ &=\; (1 - \tau)F(\xi) - \tau[1 - F(\xi)] \\ &=\; F(\xi) - \tau \end{aligned}$$

so that $F(\widehat{\xi}) = \tau$ [cf. (1)]. The solution $\widehat{\xi}(\tau)$ may not necessarily be unique.

The above dealt with population quantities. With random sample data $y_1, \ldots, y_n$ of Ys, one can replace F by its empirical cdf

$$F_n(y) \;=\; \frac{1}{n}\sum_{i=1}^{n} I(y_i \leq y).$$

so that the τth sample quantile can be found by solving

$$\begin{aligned} \min_{\xi \in \mathbb{R}} R(\xi) \;&=\; \min_{\xi \in \mathbb{R}} \sum_{i=1}^{n} \rho_\tau(y_i - \xi) \qquad (6)\\ &=\; \min_{\xi \in \mathbb{R}} \sum_{i=1}^{n} \left[\tau(y_i - \xi)_+ + (1 - \tau)(y_i - \xi)_-\right]. \qquad (7) \end{aligned}$$

Equation (7) is more convenient than defining the τth sample quantiles in terms of the order statistics $y_{(1)}, \ldots, y_{(n)}$ because the optimization problem can be generalized to the situation where there are covariates $\boldsymbol{x}$. Specifically, if the regression model $y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + \varepsilon_i$, $\varepsilon_i \sim F$, is assumed then the τth quantile is defined as any solution to the quantile regression minimization problem

$$\widehat{\boldsymbol{\beta}}(\tau) \;=\; \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \;\sum_{i=1}^{n} \rho_\tau\Big(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}\Big) \tag{8}$$

[cf. (6)]. This gives rise to the *linear conditional quantile function* $Q_Y(\tau|\boldsymbol{X}=\boldsymbol{x}) = \boldsymbol{x}^T\widehat{\boldsymbol{\beta}}(\tau)$.

Computationally, (7) is solved by linear programming techniques (e.g., as described in Koenker (2005)) because it involves minimizing the sum of *asymmetric least absolute deviations* (ALADs).

As a whole, there are some compelling advantages this classical methodology has over LMS-type methods (see Section 2). They include:

1. Inference is based on a well-established bed of theory. In contrast, inference for LMS methods are ad hoc. Koenker and Bassett (1978) gives the fundamental results, and some of these are summarized in Cizek (2000).
2. The idea can be generalized to multivariate responses, albeit, so far it is not satisfactory.
3. The idea can be generalized to other responses, e.g., binary responses.
4. The idea can be generalized to expectile regression—see Sections 1.3 and 3.

1.3 Expectile regression

Rather than minimizing the expectation of $\rho_\tau(Y-\xi)$ in (3), one can instead consider minimizing the expectation of $\rho_\omega^{[2]}(Y-\mu)$ with respect to μ where

$$\rho_\omega^{[2]}(u) = u^2 \cdot |\omega - I(u<0)|, \qquad 0<\omega<1. \tag{9}$$

This is a very natural alternative, and the results will be shown below to be quite interpretable.

Applying the same argument as (4)–(5), we minimize the *asymmetric least squares* (ALS) criterion

$$E\left[\rho_\omega^{[2]}(Y-\mu)\right] \;=\; (1-\omega)\int_{-\infty}^{\mu}(y-\mu)^2\,dF(y) + \omega\int_{\mu}^{\infty}(y-\mu)^2\,dF(y), \tag{10}$$

and setting the derivative with respect to μ to zero gives

$$(1-\omega)\int_{-\infty}^{\mu(\omega)}(y-\mu(\omega))\,dF(y) + \omega\int_{\mu(\omega)}^{\infty}(y-\mu(\omega))\,dF(y) \;=\; 0. \tag{11}$$

Theorem 1 of Newey and Powell (1987) show that a unique solution exists if $E(Y)=\mu(0.5)=\mu$ exists, and they called the quantities *expectiles*. They used this name for regression surfaces obtained by ALS. This was deliberate so as to distinguish them from the original *regression quantiles* of Koenker and Bassett (1978). Efron (1991, 1992) use the general name *regression percentile* to apply to all forms of asymmetric fitting.

In terms of interpretation, it can be seen that, given $X=x$, the quantile $\xi_\tau(x)$ specifies the position below which $100\tau\%$ of the (probability) mass of Y lies while the expectile $\mu_\omega(x)$ determines (again at $X=x$) the point such that $100\omega\%$ of the mean distance between it and Y comes from the mass below it. Thus expectiles are quite interpretable. Note that the 0.5-expectile $\mu(\frac{1}{2})$ is the mean μ while the 0.5-quantile $\xi(\frac{1}{2})$ is the median.

The above corresponds to the population. In terms of the sample, expectiles involve minimizing the asymmetrically weighted least-squares criterion

$$R(\mu) \;=\; \sum_{i=1}^{n} (1-\omega)\,[(y_i-\mu)_-]^2 + \omega\,[(y_i-\mu)_+]^2 \tag{12}$$

with respect to μ. The sample mean corresponds to $\omega = \frac{1}{2}$.

Although quantiles and expectiles are interrelated (see Section 1.3.1), each hold certain advantages and disadvantages so that neither is uniformly superior (it is like choosing between the conditional mean and median in conventional regression). The choice will usually depend on the particular application at hand. Here are some further brief notes.

1. In view of the one-to-one mapping between expectiles and quantiles (Section 1.3.1), Efron (1991) proposes that the τ quantile be estimated by the expectile for which the proportion of in-sample observations lying below the expectile is τ. This provides justification for practitioners who use expectile regression to perform quantile regression.

2. An ALS estimator is easier to compute since least-squares calculations are more natural to statisticians than linear programming techniques!

3. An ALS estimator is reasonably efficient under normality conditions (cf. Efron (1991)).

4. The ALS method is not as robust as the ALAD method against outliers. Quantiles depend only on local features of the distribution whereas expectiles have a more global nature. For example, increasing values in the upper tail of a distribution do not affect the quantiles of the lower tail but it affects the values of *all* the expectiles.

5. Although the interpretation of quantiles is easier than for expectiles, the latter is a valuable alternative to the former.

6. For quite a large class of nonlinear regression models, the conditional expectiles as functions of x are in a one-one correspondence with the conditional percentiles. Therefore, the ALS approach can be adapted to estimate conditional percentiles directly.

7. ALS regression is the least squares analogue of quantile regression.

8. Both quantile and expectile regression coincide with the MLE solutions where the data are assumed to be drawn from some specific distributions. See Exercise 12.

1.3.1 Interrelationship between expectiles and quantiles

Quite generally, Newey and Powell (1987) stated that "expectiles have properties that are similar to quantiles". Jones (1994) explored this statement in detail, and showed that the reason for this is that expectiles of a distribution F are quantiles a distribution G which is related to F. The main details are as follows.

Let $P(s) = \int_{-\infty}^{s} y f(y) dy$ be the partial moment, $\rho_\tau^{[1]}(u) = \tau - I(u \leq 0)$ and $\rho_\omega^{[2]}(u) = |u|(\omega - I(u < 0))$. From (4)–(5) we saw that one way of defining the ordinary τ-quantile of a continuous distribution with density f, $0 < \tau < 1$, is as the value of ξ that equates $\int \rho_\tau^{[1]}(y-\xi) f(y)\, dy$ to zero. In a similar way, for expectiles $\mu(\omega)$, (11) corresponds to the equation

$$\int \rho_\omega^{[2]}(y - \mu(\omega))\, f(y)\, dy \;=\; 0. \tag{13}$$

Then solving this equation shows immediately that $\omega = G(\mu(\omega))$ where

$$G(t) = \frac{P(t) - tF(t)}{2(P(t) - tF(t)) + t - \mu} \tag{14}$$

(a rearrangement of Equation (2.7) in Newey and Powell (1987)). Thus, G is the inverse of the expectile function, and its derivative is

$$g(t) = \frac{\mu F(t) - P(t)}{\{2(P(t) - tF(t)) + t - \mu\}^2}. \tag{15}$$

It can be shown that G is actually a distribution function (so that g is its density function). That is, the expectiles of F are precisely the quantiles of G defined here.

Jones (1994) illustrated his results with the standard normal, standard uniform and standard exponential distributions (reproduced in Figure 1; Exercise 13). For example, for the standard uniform,

$$g(t) = 2t(1-t)/\{2t(1-t) - 1\}^2, \qquad 0 \le t \le 1,$$

which is symmetric.

For other theory relating expectiles with quantiles see Abdous and Rémillard (1995), and Yao and Tong (1996).

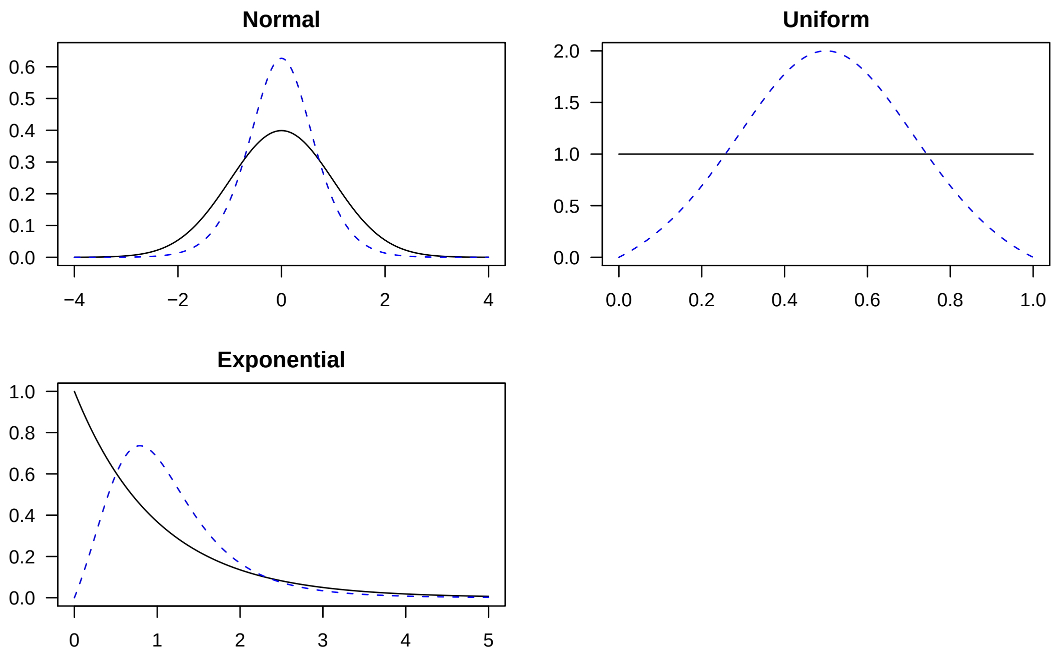

Figure 1: *Density plots of expectile g (blue dashed lines; Equation (15)) for the original f of standard normal, standard uniform and standard exponential distributions (black solid lines).*

1.3.2 Expected shortfall

The *expected shortfall* (ES) is a concept used in financial mathematics to measure portfolio risk. It is also called the *Conditional Value at Risk* (CVaR), *expected tail loss* (ETL) and *worst conditional expectation* (WCE). The ES at the $100\tau\%$ level is the expected return on the portfolio in the worst $\tau\%$ of the cases. It is often defined as

$$\text{ES}(\tau) = E(Y|Y < a) \tag{16}$$

where a is determined by $P(X < a) = \tau$ and τ is the given threshold.

The ES is very much related to expectiles (Taylor, 2008) via (11). That is, the solution $\mu(\omega)$ of this minimization satisfies

$$\left(\frac{1-2\omega}{\omega}\right) E\left[(Y-\mu(\omega)) \cdot I(Y<\mu(\omega))\right] \;=\; \mu(\omega) - E(Y). \tag{17}$$

This is another rearrangement of Newey and Powell's equation (2.7). They explain that (17) indicates that the solution $\mu(\omega)$ is determined by the properties of the expectation of the random variable Y conditional on Y exceeding $\mu(\omega)$. This suggests a link between expectiles and ES. Equation (17) can be rewritten

$$E\left[Y|Y<\mu(\omega)\right] \;=\; \left(1+\frac{\omega}{(1-2\omega)F(\mu(\omega))}\right)\mu(\omega) - \frac{\omega}{(1-2\omega)F(\mu(\omega))}\,E(Y).$$

This provides a formula for the ES of the quantile that coincides with the ω expectile. Referring to this as the τ quantile, we can write $F(\mu(\omega)) = \tau$ and rewrite the expression as

$$\text{ES}(\tau) \;=\; \left(1+\frac{\omega}{(1-2\omega)\,\tau}\right)\mu(\omega) - \frac{\omega}{(1-2\omega)\,\tau}\,E(Y). \tag{18}$$

This equation relates the ES associated with the τ quantile of the distribution of Y and the ω expectile that coincides with that quantile. The equation is for ES in the lower tail of the distribution. The equation for the upper tail of the distribution is produced by replacing ω and τ with $(1-\omega)$ and $(1-\tau)$, respectively.

By the way, another popular measure of financial risk is the *Value at Risk* (VaR). The VaR (ν_p, say) specifies a level of excessive *losses* such that the probability of a loss larger than ν_p is less than p (often $p = 0.01$ or 0.05 is chosen). The ES is defined as the conditional expectation of the loss given that it exceeds the VaR (see Yamai and Yoshiba (2002)). The ES is an increasingly popular risk measure in financial risk management and it possesses the desired sub-additivity[1] property, which is lacking for the VaR. Another major shortcoming of VaR, in addition to not being a coherent risk measure, is that it provides no information on the extent of excessive losses other than specifying a level that defines the excessive losses.

[1]The sub-additivity of a risk measure means that the risk for the sum of two independent risky events is not greater than the sum of the risks of the two events.

1.4 Software for Quantile Regression

There are a substantial number of software implementations for quantile regression. Here is an incomplete enumeration.

Koenker's generalized L_1 approach is available for R (the quantreg package is available at CRAN) and S-PLUS (on STATLIB (http://lib.stat.cmu.edu/S/quantreg). The archive is called "quantreg", the function is "rq"). More recent versions for Windows or Unix platforms are available from R. Koenker's web site (http://www.econ.uiuc.edu/∼roger), and R. Koenker also has a Koenker-D'Orey program that computes quantile linear regression estimates.

In large packages, STATA has a "qreg" command, and SAS has some capabilities for quantile regression. Shazam, and a number of other statistical/econometric packages, also allow for quantile regression. The package XploRe also allows for quantile regression modelling; see Cizek (2000).

For the classical LMS method I know of three other implementations. T. Cole has a stand-alone FORTRAN program, there is the gamlss package for R, and V. Carey has S-PLUS code at http://biosun1.harvard.edu/∼carey/index.ssoft.html; there is a link to "lmsqreg".

Other implementations include the package GROSTAT by M. Healy and co-workers, and a FORTRAN implementation by Wade and Ades (1994). It is known that, over the years, several people have written their own code which they have not made available from a website.

Koenker and Hallock (2001) express the general lack of functionality for inference in most software implementations of quantile regression.

2 LMS-type Methods

The classical LMS method for quantile regression is described in Section 2.1, a less known one in Section 2.2, and a newer one in Section 2.3.

2.1 The Box-Cox-normal Distribution Version

Suppose we have scatterplot data (x_i, y_i). The idea behind the LMS method is that a Box-Cox power transformation of the y_i, given x_i, is standard normal. That is,

$$Z \;=\; \begin{cases} \dfrac{\left(\dfrac{Y}{\mu(x)}\right)^{\lambda(x)} - 1}{\sigma(x)\,\lambda(x)}\,, & \lambda(x) \neq 0; \\[2ex] \dfrac{1}{\sigma(x)}\;\log\!\left(\dfrac{Y}{\mu(x)}\right)\,, & \lambda(x) = 0, \end{cases} \tag{19}$$

is $N(0,1)$. The LMS method gets its name from the first letters of the Greek letters λ, μ and σ, which are the three parameters. There are other statistical methods with the name "LMS" somewhere (e.g., the LMS method (Least Median Squares) introduced by Rousseeuw (1984)), so don't get confused!

The parameter σ must be positive, therefore VGAM chooses $\boldsymbol{\eta}(x) = (\lambda(x), \mu(x), \log(\sigma(x)))^T$, i.e., its default parameter link is the log link. Given $\hat{\boldsymbol{\eta}}$, the α percentile (e.g., $\alpha = 50$ for median) can be estimated by inverting the Box-Cox power transformation to give

$$\hat{\mu}(x)\left[1 + \hat{\lambda}(x)\,\hat{\sigma}(x)\,\Phi^{-1}(\alpha/100)\right]^{1/\hat{\lambda}(x)}. \tag{20}$$

The three parameters may be estimated elegantly by penalized likelihood. This was first described by Cole and Green (1992) who used an iterative smoothing spline solution to fit the LMS method. VGAM gives slightly different results—see Yee (1998) for details—because VGAM smooths all three functions simultaneously using a vector smoothing spline. For further information about the LMS method see also Green and Silverman (1994), Wright and Royston (1997). The LMS method fits neatly into the VGAM framework of Yee and Wild (1996).

Of the three functions, it is usual to allocate more degrees of freedom to $\mu(x)$. In contrast, the two functions $\lambda(\cdot)$ and $\sigma(\cdot)$ usually varies as a function of x more smoothly. In fact, it is sometimes a good idea to set λ and σ to be an intercept term only. These preferences can be easily be chosen in VGAM: e.g., setting, `s(age, df=c(2,4,2))` and/or `lms.bcn(zero=c(1,3))`. An example is given in Section 2.5.

It is worth noting about the residuals. In the LMS method, the conditional distribution of the Z-scores defined in (19) was assumed standard normal. It is natural to define the residuals as the Z-scores themselves. Thus fitting the λ, μ and σ curves with appropriate degrees of freedom, we can obtain the raw residuals as

$$z_i \;=\; \frac{\left(\dfrac{y_i}{\mu(x_i)}\right)^{\lambda(x_i)} - 1}{\sigma(x_i)\,\lambda(x_i)}\,, \tag{21}$$

The availability of a definition of residual can be thought of as an advantage of the LMS method. In parametric or nonparametric quantile regression methods it is not alway possible to define residuals, particularly when quantiles are estimated separately. From a practical point of view, the residuals should be checked for standard normality by a QQ-plot or overlaying the probability density function of a $N(0, 1)$ onto a histogram. These types of residuals have not yet been implemented in VGAM.

The VGAM solution involves maximizing an *approximate* likelihood, and using *approximate* derivatives. In fact, if too many iterations are performed, the solution may diverge and fail! Also, good initial values are usually required.

The LMS method does not have to be regressed on a single covariate X. One can adjust for other predictors such as SEX and RACE. For example, one could use

```
vgam(y ~ s(age) + as.factor(sex) + as.factor(race), lms.yjn, data)
```

For goodness-of-fit tests for quantile data see Royston and Wright (2000).

2.2 The Box-Cox-gamma Distribution Version

From (20), one problem with the above formulation is that

$$1 + \lambda(x)\,\sigma(x)\,\Phi^{-1}(\alpha/100) > 0$$

is required in order for the centiles to be available. Hence a disadvantage is that the range of transformation depends on λ. To overcome this range-restriction problem in its parameters, Lopatatzidis and Green considered replacing the normal distribution by a gamma distribution. The transformed variable

$$W \;=\; (Y/\mu)^{\lambda}$$

is assumed gamma with unit mean and variance $\lambda^2\sigma^2$. Then the α percentile of Y at x is

$$\mu(x)\,W_{\alpha}^{1/\lambda(x)}$$

where W_α is the equivalent deviate of size α for the gamma distribution with mean 1 and variance $\lambda(x)^2\sigma(x)^2$. In the case $\lambda(x) = 0$, the distribution of W at x reduces to

$$\log\left(\frac{y}{\mu(x)}\right) \;\sim\; N(0, \sigma(x)^2), \tag{22}$$

giving 100α% centile of Y at x as

$$\mu(x)\;\exp\{\sigma(x)Z_\alpha\}.$$

Now the (two) parameter gamma distribution has density

$$g_U(u) \;=\; \frac{\beta^{\alpha}}{\Gamma(\alpha)}u^{\beta-1}\,\exp(-\beta u), \qquad \alpha > 0,\quad \beta > 0,\quad u > 0,$$

giving $E(U) = \alpha/\beta$ and $\mathrm{Var}(U) = \alpha/\beta^2$. The distribution of W can thus be achieved from this with $\alpha = \beta = (\sigma\lambda)^{-2}$. Note that it is still easy to compute W_α easily in S-PLUS/R using `qgamma()`, which is for the one parameter gamma distribution, because $T = \theta W \sim \Gamma(\theta)$ where $\theta = (\sigma\mu)^{-2}$.

Although the gamma model is not traditional in quantile regression it avoids a range of problems. It has finite expectations of the required derivatives of the likelihood function, which can prove to be an important problem for the normal version, particularly when σ is small. In such situations it appears that the penalized likelihood can go to infinity and thus the algorithm cannot converge. In such cases the gamma version of the LMS method can give an acceptable solution. Moreover, the parameterization reduces or eliminates correlations since the off-diagonal elements of the $\mathbf{W}_i$ are zero or close to zero (relative to the diagonal elements). Lastly, an especially attractive feature of the gamma model is that, unlike the normal case, the range of transformation does not depend on λ. Thus, in the gamma model Y ranges from $(0, \infty)$ for all λ, μ and σ. Further technical details of the gamma version and a comparison between the normal and gamma distributions is given in Lopatatzidis and Green.

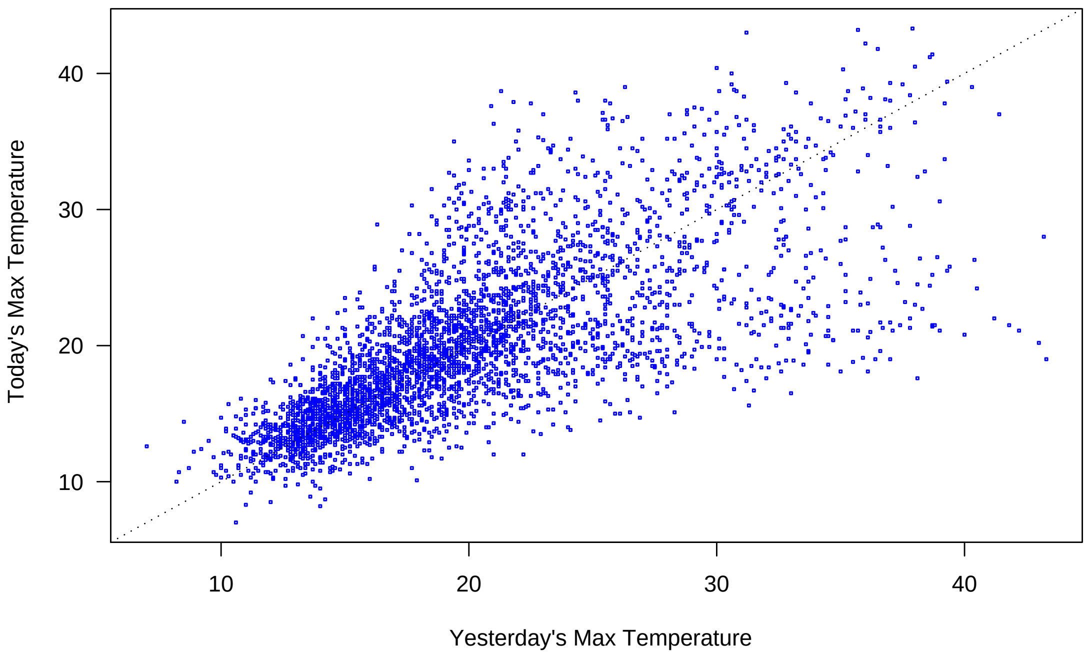

Figure 2: *Melbourne temperature data.*

2.3 The Yeo-Johnson-normal Distribution Version

Yeo and Johnson (2000) introduce a new power transformation which is well defined on the whole real line (the Box-Cox transformation is only valid for positive values) and potentially useful for improving normality. It is given by

$$\psi(\lambda, y) = \begin{cases} \{(y+1)^{\lambda} - 1\}/\lambda & (y \geq 0, \lambda \neq 0), \\ \log(y+1) & (y \geq 0, \lambda = 0), \\ -\{(-y+1)^{2-\lambda} - 1\}/(2-\lambda) & (y < 0, \lambda \neq 2), \\ -\log(-y+1) & (y < 0, \lambda = 2). \end{cases} \tag{23}$$

An LMS method based on this power transformation has been implemented in the VGAM family function `lms.yjn()`, which chooses the three parameters λ, μ, σ to transform to normality. The observed Hessian matrices are almost always not positive-definite, but 3 elements of the expected Hessian are tractable. The other 3 elements can be approximated.

A value of $\lambda = 1$ corresponds to the identity transformation, which is why the `"loge"` link is provided by `lms.yjn()`—when used by `vglm()` the `summary()` can test the hypothesis that $\lambda = 1$. The Yeo-Johnson transformation is equivalent to the generalized Box-Cox transformation for $y > -1$ where the shift constant 1 is included. For the many properties of the transformation see Yeo and Johnson (2000). The LMS-Yeo-Johnson method was proposed by Yee (2002). See also Yee (2004) for details.

From a practical point of view, `lms.yjn()` subtracts the sample median of the response from the response. This is done so that the numerical integration scheme described in Yee (2004) will be accurate. If problems still persist, one idea is to fit a regression through the data (e.g., fit a quadratic or even a regression spline) and then perform the quantile regression on the residuals.

Another VGAM family function, `lms.yjn2()`, estimates the EIM using simulation.

2.4 VGAM Family and Generic Functions

A summary of functions written for age-reference centile data is given in Table 2.

1. It is easy to see why the VGAM family function names are `lms.bcn()`, `lms.bcg()`, `lms.yjn()`: the "bc" stands for Box-Cox, "`yj`" for Yeo-Johnson, "`n`" for normal, and the "g" stands for gamma.

2. Because $\sigma > 0$, VGAM uses the log link for this parameter as default. However, an identity link can easily be chosen instead by `link="identity"`.

3. Output: Suppose `fit` is a `lms.bcn()` VGAM object. Then the fitted values (in `fitted(fit)`) are a n-row matrix of centiles. The default percentiles are 25, 50 and 75. The user can specify alternative values, e.g., `lms.bcn(percentiles=c(50,95))` for 50% and 95%. Thus `fitted(fit)` do not contain the mean, as for GLMs.

4. Constraints: The `lms.bcn()` family function has the `zero` argument which can be assigned a vector taking the values 1, 2, and/or 3. If set to a value j, it forces η_j to be modelled as a intercept only. By default, `zero=NULL`, meaning all the η_j are modelled as functions of the predictors.

5. The `yoffset` argument of `lms.yjn()` adds that value to the response before fitting the model. Thus a good value to input is the negative of some value of center of location, e.g., −mean or −median. The idea is to center the data. The value is stored on the object in `@misc$yoffset`, and used by `qtplot()` and `deplot()`. Note that `offset` can be used in `vglm()`/`vgam()`, and is the offset to the linear/predictors, and not the response.

6. `lmscreg.control()` is called by `vglm()` and `vgam()`. It has some options that can be chosen, for example, setting `cdf=F` suppresses the computing of the cdf values $\hat{P}(Y_i \leq y_i|\boldsymbol{x}_i)$ for observation i.

7. The methods functions `qtplot.lmscreg()`, `deplot.lmscreg()`, `cdf.lmscreg()` have an `Attach` option. If set to `TRUE`, the object is returned with the answer in the slot `@misc`.

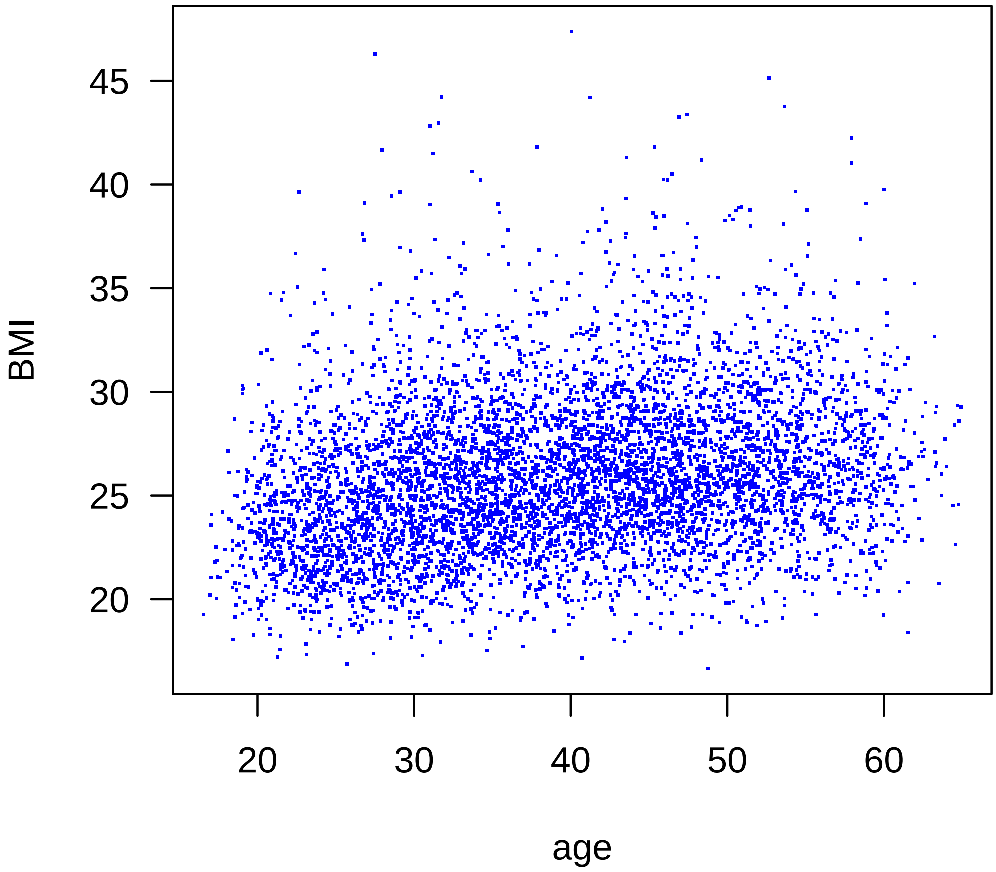

Figure 3: *Obesity of 4786 white male New Zealanders from a workforce study.*

2.5 Example

In an undistributed dataset called workforce.txt are the ages (in years) and body mass indexes (BMI; in kg/m^2) of 6183 European-type New Zealanders aged between 16 and 65. They can be thought of as an approximate random sample from the adult population. Loosely speaking, the larger the BMI the more obese the person; it is defined as a person's weight $\div$ height^2. Because BMIs are affected by age, gender and race, healthy values can range anywhere from about 20 to 30 for most adults.

In this analysis we will illustrate the new LMS-Yeo-Johnson method by restricting ourselves to males only, of which there are 4786 individuals in a data frame called `wfmen`. A scatter plot of the data is given in Figure 3. With such a mass of data it is sometimes difficult to see certain features, however, BMI appears to increase with age and then plateau. A simple fit is

Table 2: *Functions for age-reference centile data. The top are* VGAM *family functions, and the bottom are generic functions.*

`lms.bcn`	Box-Cox transformation to normality
`lms.bcg`	Box-Cox transformation to gamma distribution
`lms.yjn`	Yeo-Johnson transformation to normality
`alsqreg`	Asymmetric least squares—see Section 3.
`amlbinomial`	Asymmetric maximum likelihood—for binomial. See Section 3.
`amlpoisson`	Asymmetric maximum likelihood—for Poisson. See Section 3.
`deplot`	Plots the fitted probability density functions
`qtplot`	Plots the fitted quantiles
`cdf`	Cumulative distribution function of individual observations

```
> fit = vgam(BMI ~ s(age, df = c(2, 4, 2)), family = lms.yjn(yoffset = -25),
+     data = wfmen)

VGAM  s.vam  loop  1 :  loglikelihood = -8007.024
VGAM  s.vam  loop  2 :  loglikelihood = -7998.88
VGAM  s.vam  loop  3 :  loglikelihood = -7998.372
VGAM  s.vam  loop  4 :  loglikelihood = -7998.284
VGAM  s.vam  loop  5 :  loglikelihood = -7998.28
VGAM  s.vam  loop  6 :  loglikelihood = -7998.274
VGAM  s.vam  loop  7 :  loglikelihood = -7998.275
VGAM  s.vam  loop  8 :  loglikelihood = -7998.274
VGAM  s.vam  loop  9 :  loglikelihood = -7998.275

> par(mfrow = c(3, 1), mar = c(6, 5, 2, 1) + 0.1, las = 0)
> plot(fit, se = TRUE, rug = FALSE, cex = 1, lcol = "blue",
+     scol = "green4", lwd = 2, slwd = 2)
```

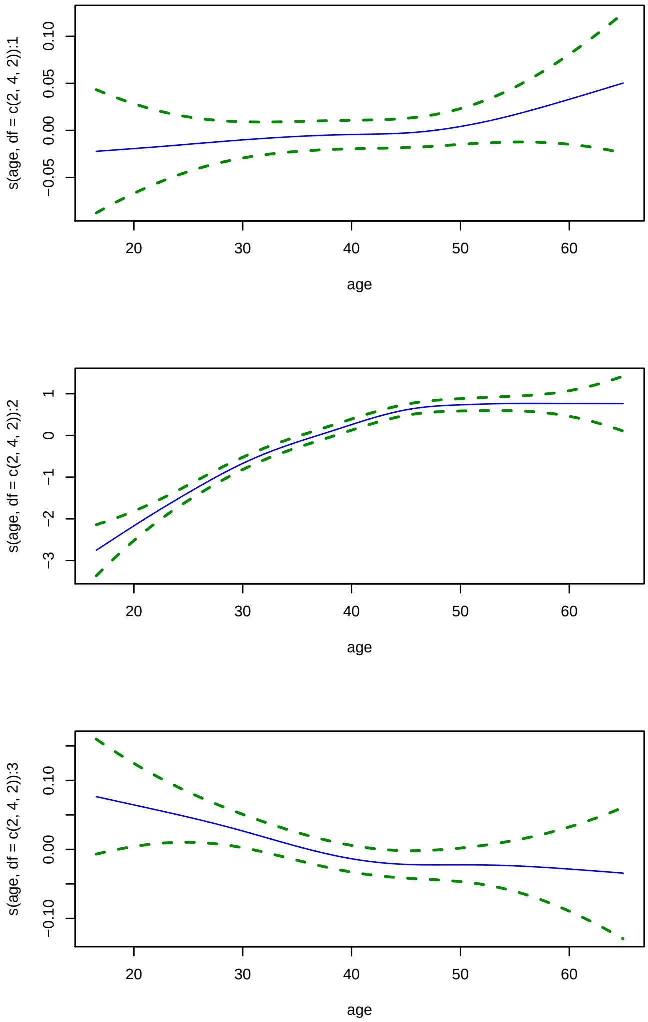

Figure 4: *Fitted functions* $\hat{\eta}_1 = \hat{\lambda}(x)$, $\hat{\eta}_2 = \hat{\mu}(x)$, $\hat{\eta}_3 = \log(\hat{\sigma}(x))$ *of the LMS method fitted to BMI data.*

Also, Figure 5 was produced by

```
> par(mfrow = c(3, 1), mar = c(6, 5, 2, 1) + 0.1, las = 0)
> plot(fit, deriv = 1, scale = 0.2, cex = 1, rug = FALSE,
+      lcol = "blue", lwd = 2)
```

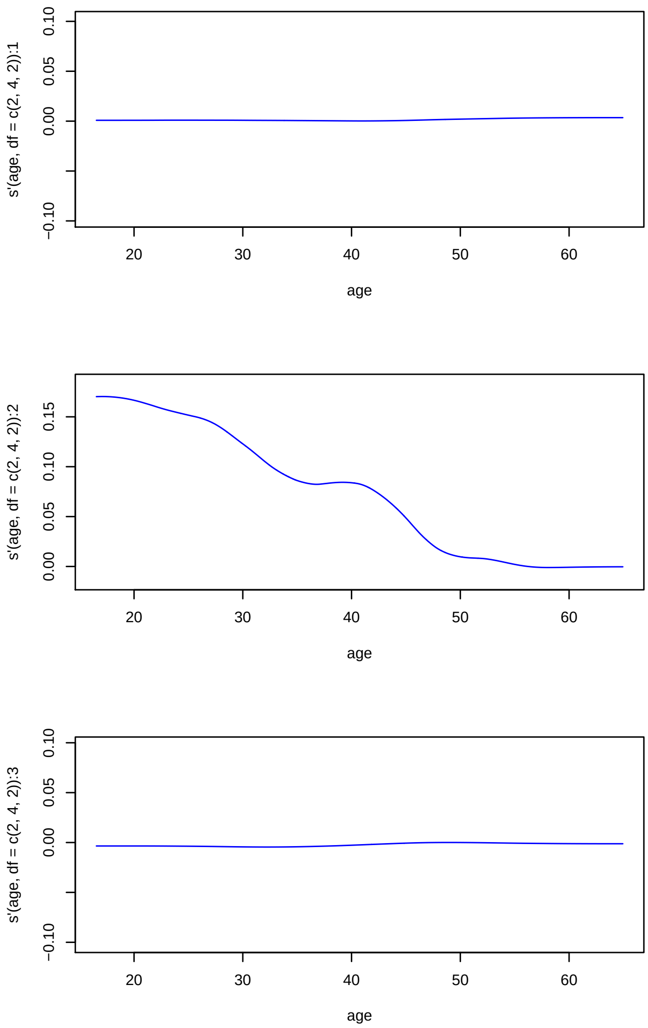

Figure 5: *First derivatives of the fitted functions:* $\hat{\eta}'_j(x)$, $j = 1, 2, 3$.

In Figure 4 the first and third functions appear linear, but the second ($\hat{\mu}(x)$) is clearly nonlinear. The first function looks flat, so we'll replace it by a constant. In Figure 5 the derivatives, which are plotted on the same scale, confirm this. A more formal summary is provided by

```
> summary(fit)

Call:
vgam(formula = BMI ~ s(age, df = c(2, 4, 2)), family = lms.yjn(yoffset = -25),
    data = wfmen)

Number of linear predictors:    3

Names of linear predictors: lambda, mu, log(sigma)

Dispersion Parameter for lms.yjn family:   1

Log-likelihood: -7998.274 on 14347.13 degrees of freedom

Number of Iterations:  9

DF for Terms and Approximate Chi-squares for Nonparametric Effects

                          Df Npar Df Npar Chisq   P(Chi)
(Intercept):1              1
(Intercept):2              1
(Intercept):3              1
s(age, df = c(2, 4, 2)):1  1     0.9      1.987 0.143571
s(age, df = c(2, 4, 2)):2  1     3.0     54.506 0.000000
s(age, df = c(2, 4, 2)):3  1     1.0      3.378 0.061782
```

The P-values for testing linearity of the functions appear to match the visual inspections of the plots. So let's try

```
> fit2 = vgam(BMI ~ s(age, df = c(4, 1)), lms.yjn(yoffset = -25,
+     zero = 1), data = wfmen)

VGAM  s.vam  loop  1 :  loglikelihood = -8008.065
VGAM  s.vam  loop  2 :  loglikelihood = -8001.304
VGAM  s.vam  loop  3 :  loglikelihood = -8001.215
VGAM  s.vam  loop  4 :  loglikelihood = -8001.201
VGAM  s.vam  loop  5 :  loglikelihood = -8001.198
VGAM  s.vam  loop  6 :  loglikelihood = -8001.198
```

which models λ as an intercept only and σ linearly.

Then Figure 6 is obtained from

```
> par(mar = c(5, 5, 1, 3) + 0.1, bty = "l", las = 1, mfrow = c(1,
+     1))
> more = qtplot(fit2, per = c(1, 5, 25, 50, 75, 95, 99),
+     pcol = 1, ylab = "BMI", pcex = 0.25, pch = 15, lwd = 2,
+     lcol = "blue", tcol = "blue", tadj = 1)
```

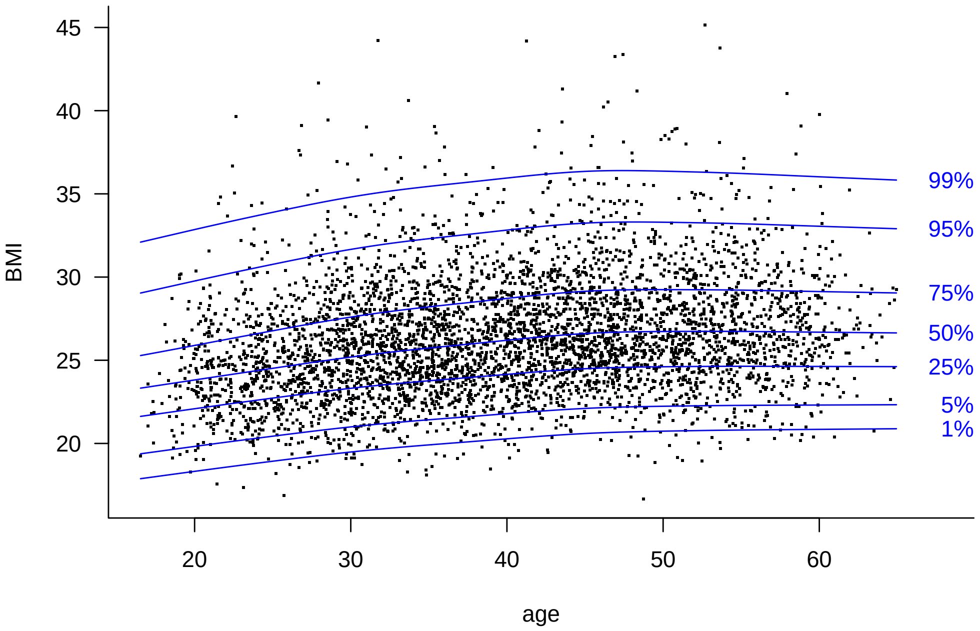

Figure 6: *Fitted centiles from* `fit2`.

Such plots of quantiles are often seen in hospitals, doctors surgeries and patients' waiting rooms. For fixed ages, the distribution of BMI is positively skewed. These results show that BMI generally increases with age until about 50, and then stabilizes. Sometimes data like this decreases past a certain point—this is probably a selection bias because people with high BMIs are more prone to diseases (particular cardiovascular types) and hence die younger.

Density plots can be obtained as in

```
> par(mfrow = c(1, 1), mar = c(5, 5, 0.1, 0.1) + 0.1, bty = "l",
+     las = 1, lwd = 2)
> at = seq(15, 43, by = 0.25)
> deplot(fit2, x0 = 20, y = at, xlab = "BMI", col = "green4")
> deplot(fit2, x0 = 40, y = at, add = TRUE, lty = 2, col = "blue")
> a <- deplot(fit2, x0 = 60, y = at, add = TRUE, lty = 3,
+     col = "red", lwd = 3)
```

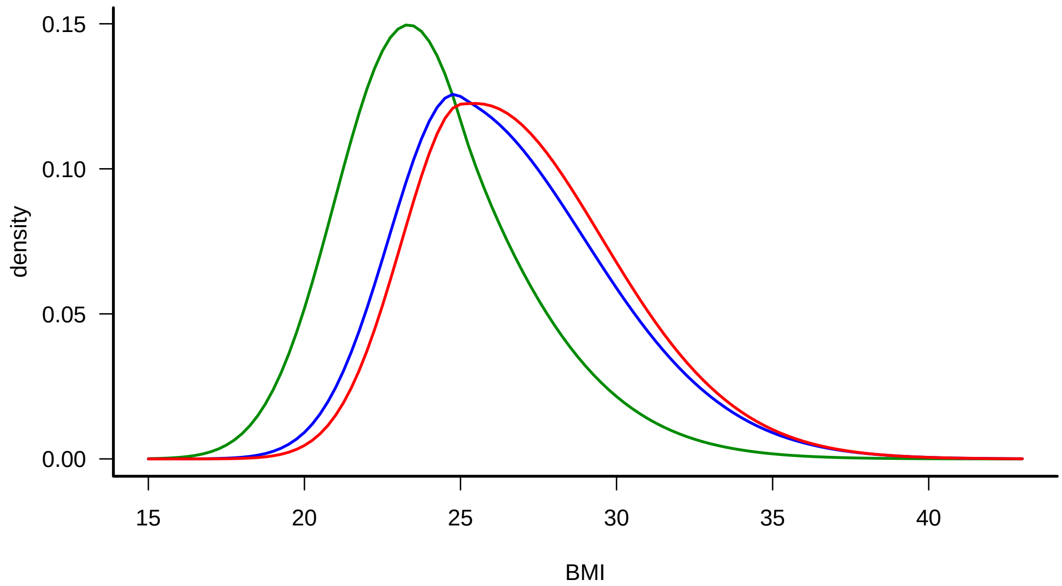

Figure 7: *Density plots of BMI from* `fit2` *at ages 20 (solid line), 40 and 60.*

This shows the density of BMI at certain values of age. There is little difference in the distribution of BMIs between the 40 and 60 year olds, but they are both substantially higher than 20 year olds. (This is called middle-aged spread!) The positive skew is also evident.

Then

```
> wfmen[1:2, c("age", "BMI")]

     age    BMI
1 29.961 24.811
2 55.937 29.047

> fitted(fit2)[1:2, ]

       25%      50%      75%
1 23.31083 25.18645 27.58766
2 24.63723 26.72540 29.19981

> cdf(fit2)[1:2]

        1         2
0.4511025 0.7371614
```

That is, person 1 with a BMI of 24.811 is on the 45.11 percentile of the BMI distribution amongst other 30 year olds.

Here is some other output that is available.

```
> more@post$qtplot$percentiles

[1]  1  5 25 50 75 95 99

> more@post$qtplot$fitted[1:2, ]

        1%       5%      25%      50%      75%      95%      99%
1 19.48175 20.99086 23.31083 25.18645 27.58766 31.65674 34.80149
2 20.81650 22.29291 24.63723 26.72540 29.19981 33.17789 36.19084

> a@post$deplot$newdata

  age
1  60

> a@post$deplot$y[1:5]

[1] 15.00 15.25 15.50 15.75 16.00

> a@post$deplot$density[1:5]

[1] 1.234249e-07 2.549810e-07 5.152472e-07 1.018604e-06 1.970411e-06

> fit2@constraints

$`(Intercept)`
     [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    1    0
[3,]    0    0    1

$`s(age, df = c(4, 1))`
     [,1] [,2]
[1,]    0    0
[2,]    1    0
[3,]    0    1

> args(lms.yjn)

function (percentiles = c(25, 50, 75), zero = NULL, link.lambda = "identity",
    link.sigma = "loge", elambda = list(), esigma = list(), dfmu.init = 4,
    dfsigma.init = 2, init.lambda = 1, init.sigma = NULL, rule = c(10,
        5), yoffset = NULL, diagW = FALSE, iters.diagW = 6)
NULL
```

2.6 Advanced Usage of Some Methods Functions

With VGAM family functions for quantile regression the resulting fitted models can be plotted in more ways than just the generic function `plot()`. Here are some more details.

2.6.1 `qtplot()`

In theory the LMS method can operate with more than one 'primary' covariate (usually age), e.g., adjust for other variables such as gender and race. To plot the results using `qtplot()`, however, is not easy, but possible by using its `newdata` argument. This allows prediction to occur in that data frame. Note that plotting the quantiles against the primary variable only makes sense if the non-primary variables have a common value.

Below is some code to illustrate this. We will use `wf`, which consists of a random sample of 350 men and 350 women from the `workforce` data frame, and there is another variable, `wf$sex`, which one adjusts for. It has value 0 for men and 1 for women. The 'primary' variable is `wf$age`.

Figure 8 was obtained from the following.

```
> fit3 = vgam(BMI ~ s(age, df = 4) + sex, fam = lms.yjn(percentile = 50,
+     yoffset = -25, zero = c(1, 3)), data = wf, trace = FALSE)
> nn = 80
> Age = seq(18, 64, len = nn)
> half = split(wf, wf$sex)
> par(bty = "l", mar = c(5, 4, 1, 3) + 0.1, las = 1)
> plot(wf$age, wf$BMI, xlab = "age", ylab = "BMI", type = "n")
> points(half[["0"]]$age, half[["0"]]$BMI, pch = 15, cex = 0.7,
+     col = "green4")
> newdat = data.frame(age = Age, sex = rep(0, nn))
> men = qtplot(fit3, newdata = newdat, add = TRUE, lcol = "green4",
+     tcol = "green4", lwd = 2, tadj = 0)
> points(half[["1"]]$age, half[["1"]]$BMI, pch = "o", col = "red",
+     cex = 0.9)
> newdat = data.frame(age = Age, sex = rep(1, nn))
> women = qtplot(fit3, newdata = newdat, add = TRUE, llty = 2,
+     lcol = "blue", tcol = "blue", llwd = 2, tadj = 0)
```

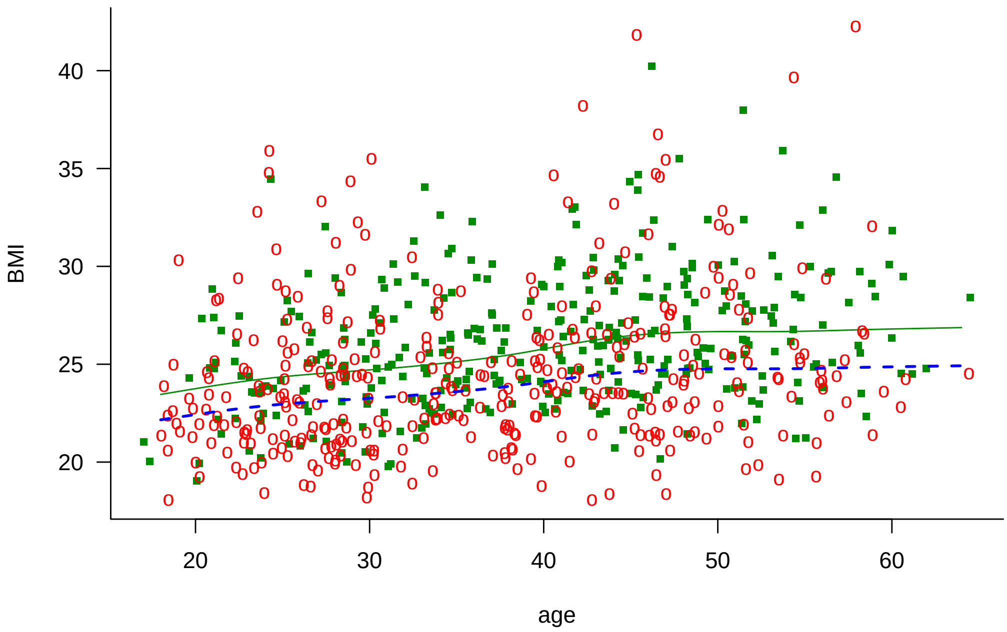

Figure 8: *Output of* `qtplot()` *when there are non-primary variables. The solid line and squares are for men, and the dashed line and circles are for women.*

Not surprising, men are more solid build than women of the same age. In fact,

```
> range(men@post$qtplot$fitted - women@post$qtplot$fitted)

[1] 1.296857 1.949876
```

shows that the median BMI of men is between about 1.3 to 1.95 kg/m^2 higher than women.

Here are some notes about `qtplot()`.

1. When `newdata` is specified in `qtplot()` it is very important to note that `qtplot()` chooses the first term in the formula as the primary variable. If one used

   ```
   fit = vgam(BMI ~ sex + s(age, df=c(2,4,2)), fam=lms.bcn, data=wf)
   ```

   then `qtplot(fit, newdata=newdat)` would try using `newdat$sex` as the primary variable.

2. `qtplot()` uses the percentile values from the original model if the argument `percentiles` isn't specified.

3. Some arguments beginning with "p" such as `pcex`, `pcol` correspond to points. Some arguments beginning with "t" such as `tcol` correspond to the text labelling.

2.6.2 `deplot()`

The generic function `deplot()` stands for 'density plot', and is relevant for quantile regression. The methods function `deplot.vglm()` dispatches a quantile regression VGAM object by looking at the last element of the vector `@vfamily`.

2.6.3 cdf()

When one fits an LMS-type VGAM, the cumulative distribution function values are returned in `@post$cdf`. This can be suppressed by setting the argument `cdf=FALSE`, which is found in the function `lmscreg.control()`.

The generic function `cdf()` extracts or computes the cumulative distribution function of the response for individual observations. The answer is a vector. It has the argument `Attach=FALSE`, which if assigned `TRUE`, returns the object with the vector of cumulative probabilities in `@post$cdf`. The methods function `cdf.lmscreg()` accepts a data frame `newdata` for which the cumulative probabilities are calculated. Here is an example.

```
> data(bminz)
> fit = vgam(BMI ~ s(age, df = c(2, 4, 2)), fam = lms.bcn,
+     data = bminz, trace = FALSE)
> fit@post$cdf[1:4]

        1         2         3         4 
0.2297406 0.6375906 0.6325457 0.4372778 

> fitted(fit)[1:4, ]

       25%      50%      75%
1 22.98046 25.38744 28.33684
2 23.65124 26.19280 29.20880
3 24.09322 26.70356 29.77494
4 23.23800 25.70082 28.67144

> bminz[1:4, ]

       age      BMI
1 31.52966 22.77107
2 39.38045 27.70033
3 43.38940 28.18127
4 34.84894 25.08380

> cdf(fit, data.frame(BMI = fitted(fit)[1:4, "25%"], age = bminz$age[1:4]))

   1    2    3    4 
0.25 0.25 0.25 0.25 
```

That is, a 39.4 year old with a BMI of 27.7 corresponds to a cumulative probability of 0.638.

3 Expectile regression

Section 2 focussed on quantile regression based on the LMS idea. The VGAM package implements another method that is similar to quantile regression which was proposed by Aigner et al. (1976) and Newey and Powell (1987) and further developed by Efron (1991). We call this method "expectile regression". For normally distributed responses, it is based on *asymmetric least squares* (ALS) estimation, a variant of ordinary least squares (OLS) estimation. A short summary is given below and full details can be obtained from the papers. The method proposed by Koenker and Bassett (1978) is similar but is based instead on minimizing *asymmetric least absolute deviations* (ALAD). Then, Efron (1992) generalized ALS estimation to families in the exponential family, and in particular, the Poisson distribution. He called this *asymmetric maximum likelihood* (AML) estimation.

Consider the linear model $y_i = \boldsymbol{x}_i \boldsymbol{\beta} + \varepsilon_i$ for $i = 1, \ldots, n$. Let $r_i(\boldsymbol{\beta}) = y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}$ be a residual. The asymmetric squared error loss function for a residual r is r^2 if $r \leq 0$ and wr^2 if $r > 0$. Here w is a positive constant and is related to ω in Section 1.3 by $w = \omega/(1-\omega)$ (a renormalization of (10)). The solution is the set of regression coefficients that minimize the sum of these over the data set, weighted by the `weights` argument (so that it can contain frequencies). Written mathematically, the asymmetric squared error loss $S_w(\boldsymbol{\beta})$ is

$$S_w(\boldsymbol{\beta}) \;=\; \sum_{i=1}^{n} w_i \, Q_w^*(r_i(\boldsymbol{\beta})) \tag{24}$$

and Q_w^* is the asymmetric squared error loss function

$$Q_w^*(r) \;=\; \begin{cases} r^2, & r \leq 0, \\ w\, r^2, & r > 0. \end{cases} \tag{25}$$

The w_i are known prior weights, inputted using the `weights` argument (of `vglm()`, etc.) and retrievable afterwards as `weights(fit, type="prior")`.

Here are some notes about ALS quantile regression.

1. Usually the user will specify some desired value of the percentile, e.g., 75 or 95. Then the necessary value of w needs to be numerically solved for to obtain this. One useful property is that the percentile is a monotonic function of w, meaning one can solve for the root of a nonlinear equation. See Section 3.1 for a numerical example.

2. A rough relationship between w and the percentile 100α is given on p.102 of Efron (1991). Let $w^{(\alpha)}$ denote the value of w such that $\boldsymbol{\beta}_w$ equals $z^{(\alpha)} = \Phi^{-1}(\alpha)$, the 100α standard normal percentile point. If there are no covariates (intercept-only model) and y_i are standard normal then

$$w^{(\alpha)} \;=\; 1 + \frac{z^{(\alpha)}}{\phi\left(z^{(\alpha)}\right) - (1-\alpha) z^{(\alpha)}} \tag{26}$$

where $\phi(z)$ is the probability density function of a standard normal. Here are some values.

```
> alpha = c(1/2, 2/3, 3/4, 0.84, 9/10, 19/20)
> zalpha = qnorm(p = alpha)
> walpha = 1 + zalpha/(dnorm(zalpha) - (1 - alpha) * zalpha)
> round(cbind(alpha, walpha), dig = 2)
```

```
      alpha walpha
[1,]   0.50   1.00
[2,]   0.67   2.96
[3,]   0.75   5.52
[4,]   0.84  12.81
[5,]   0.90  28.07
[6,]   0.95  79.73
```

3. The ALS loss function (25) leads to an important invariance property: if the y_i are multiplied by some constant c then the solution vector $\widehat{\boldsymbol{\beta}}_w$ is also multipled by c. Also, a shift in location to $y_i + d$ means the estimated intercept (the first element in $\boldsymbol{x}$) increases by d too.

4. ALS quantile regression is consistent for the true regression percentiles $y^{(\alpha)}|\boldsymbol{x}$ in the cases where $y^{(\alpha)}|\boldsymbol{x}$ is linear in $\boldsymbol{x}$. See also Newey and Powell (1987) for a more general proof of this.

5. An iterative solution is required, and the Newton-Raphson algorithm is used. In particular, for Poisson regression with the canonical (log) link, following in from Equation (2.16) of Efron (1992),

$$\begin{aligned}
\boldsymbol{\beta}^{(a+1)} &= \boldsymbol{b}^{(a)} + d\boldsymbol{b}^{(a)} \\
&= \boldsymbol{b} - \ddot{\mathbf{S}}_w^{-1} \ddot{\boldsymbol{S}}_w \\
&= (\mathbf{XWVX})^{-1}\mathbf{XWV}\left[\boldsymbol{\eta} + (\mathbf{WV})^{-1}\,\mathbf{W}\boldsymbol{r}\right]
\end{aligned}$$

   are the Newton-Raphson iterations (iteration number a suppressed for clarity). Here, $\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{b})$, $\mathbf{V} = \mathrm{diag}(v_1(\boldsymbol{b}), \ldots, v_n(\boldsymbol{b})) = \mathrm{diag}(\mu_1, \ldots, \mu_n)$ contains the variances of y_i and $\mathbf{W} = \mathrm{diag}(w_1(\boldsymbol{b}), \ldots, w_n(\boldsymbol{b}))$ with $w_i(\boldsymbol{b}) = 1$ if $r_i(\boldsymbol{b}) \leq 0$ else w.

   Consequently one expects order-2 convergence, meaning that the number of correct decimal places (approximately) doubles after each successive iteration.

3.1 ALS Example

ALS quantile regression is implemented with the VGAM family function `alsqreg()`. The `@deviance` slot computes the weighted asymmetric squared error loss $S_w(\boldsymbol{\beta})$ (see (24)) summed over all `w.als` values.

Here is a simple example involving the BMI of 700 New Zealanders. The value of w here was largely chosen at random.

```
> data(bminz)
> o = with(bminz, order(age))
> bminz = bminz[o, ]
> fit = vglm(BMI ~ bs(age, df = 3), fam = alsqreg(w.als = 3),
+     data = bminz, trace = TRUE, crit = "coef")

VGLM    linear loop  1 :  coefficients =
27.524503547, -0.780304638,  6.482936311, -3.917888792
VGLM    linear loop  2 :  coefficients =
27.544342508, -0.716448787,  7.992961885, -4.498998892
VGLM    linear loop  3 :  coefficients =
27.545049746, -0.719421948,  8.032319764, -4.511034213
VGLM    linear loop  4 :  coefficients =
27.545049746, -0.719421948,  8.032319764, -4.511034213
```

Note the rapid convergence. The fitted model is

```
> fit

Call:
vglm(formula = BMI ~ bs(age, df = 3), family = alsqreg(w.als = 3),
    data = bminz, trace = TRUE, crit = "coef")

Coefficients:
     (Intercept) bs(age, df = 3)1 bs(age, df = 3)2 bs(age, df = 3)3
       27.545050        -0.719422         8.032320        -4.511034

Degrees of Freedom: 700 Total; 696 Residual
Residual Deviance: 27882.83
```

Here, the `"Residual deviance"` is $S_w(\widehat{\boldsymbol{\beta}}_w)$ where $w = 3.0$. Some useful output is

```
> fit@extra

$w.als
[1] 3

$M
[1] 1

$n
[1] 700
```

```
$y.names
[1] "w.als=3"

$percentile
w.als=3
     71

$individual
[1] TRUE

$deviance
 w.als=3
27882.83
```

```
> coef(fit)

     (Intercept) bs(age, df = 3)1 bs(age, df = 3)2 bs(age, df = 3)3
       27.545050        -0.719422         8.032320        -4.511034
```

```
> coef(fit, matrix = TRUE)

                 mu(w.als=3)
(Intercept)        27.545050
bs(age, df = 3)1   -0.719422
bs(age, df = 3)2    8.032320
bs(age, df = 3)3   -4.511034
```

Figure 9 is a quantile plot produced by

```
> with(bminz, plot(age, BMI, main = paste(round(fit@extra$percentile,
+     dig = 1), "percentile curve"), col = "blue", las = 1))
> with(bminz, lines(age, c(fitted(fit)), col = "red"))
```

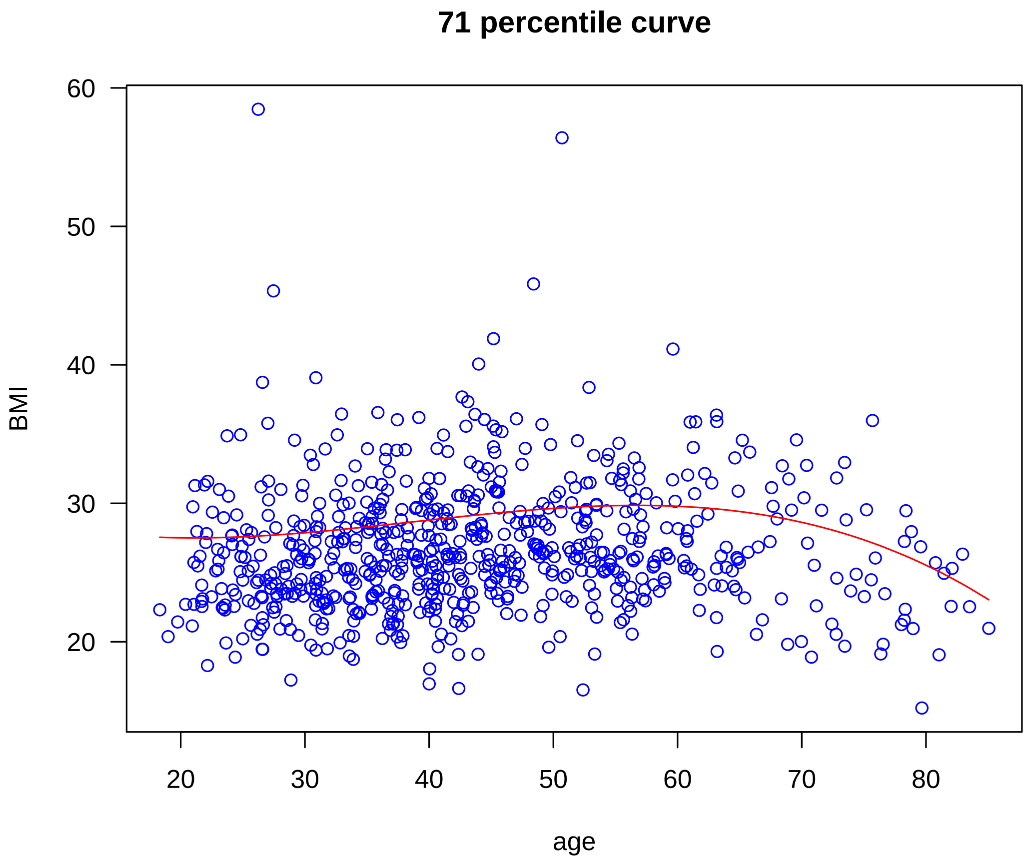

Figure 9: *Quantile plot from* `alsqreg()` *applied to* `bminz`*.*

Suppose we want to obtain the running quartiles, i.e., the 25, 50 and 75 percentiles. Then we need to numerically search for the appropriate w for these and there are several ways of attempting this. One of them is to first define a new function

```
> findw = function(w, percentile = 50) {
+     fit = vglm(BMI ~ bs(age, df = 3), fam = alsqreg(w = w),
+         data = bminz)
+     fit@extra$percentile - percentile
+ }
```

Solving for the root as a function of w gives the desired percentile.

The following code gives Figure 10. It uses `uniroot()` since the percentile is a monotonic function of w. We assume the appropriate w lies between 0.0001 and 10000. For each percentile the optimal w is found, then a model is fitted with this w, and then the curve is added to the plot.

```
> with(bminz, plot(age, BMI, main = "25, 50 and 75 percentile curves",
+     col = "blue", las = 1))
> for (myp in c(25, 50, 75)) {
+     bestw = uniroot(f = findw, interval = c(1/10^4, 10^4),
+         percentile = myp)
+     fit = vglm(BMI ~ bs(age, df = 3), fam = alsqreg(w = bestw$root),
+         data = bminz)
+     with(bminz, lines(age, c(fitted(fit)), col = "red"))
+ }
```

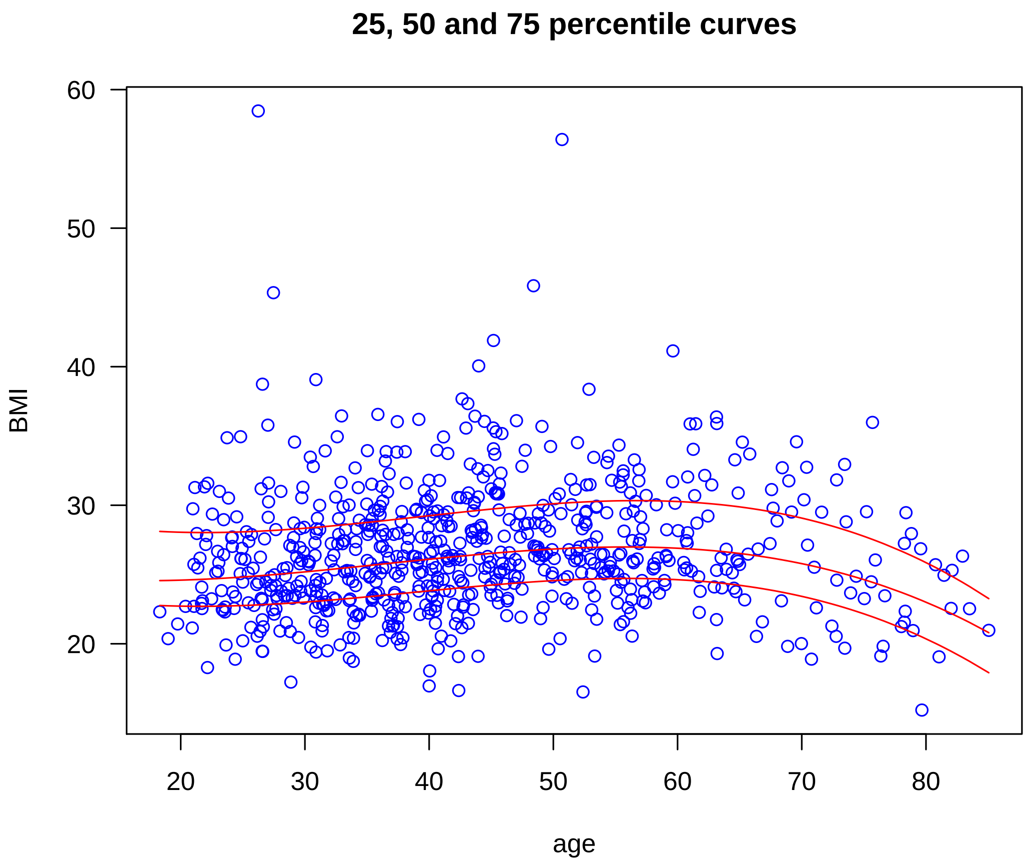

Figure 10: *Quantile plot from* `alsqreg()` *applied to* `bminz`*: 25, 50 and 75 percentile curves. Each regression curve is a regression spline with 3 degrees of freedom (1 would be a linear fit).*

Altogether, this appears to give a reasonable fit to the data. Presently, the generic functions `qtplot()`, `deplot()` etc. do not work on an `alsqreg()` object.

The following code gives Figure 11. It uses a vector of w values.

```
> data(bminz)
> o = with(bminz, order(age))
> bminz = bminz[o, ]
> fit3 = vgam(BMI ~ s(age, df = 4), fam = alsqreg(w.als = c(0.1,
+     1, 10), parallel = TRUE), data = bminz)
> with(bminz, plot(age, BMI, main = paste(paste(round(fit3@extra$percentile,
+     dig = 1), collapse = ", "), "percentile curves")),
+     col = "blue")
> with(bminz, matlines(age, fitted(fit3), col = 1:fit3@extra$M,
+     lwd = 2))
```

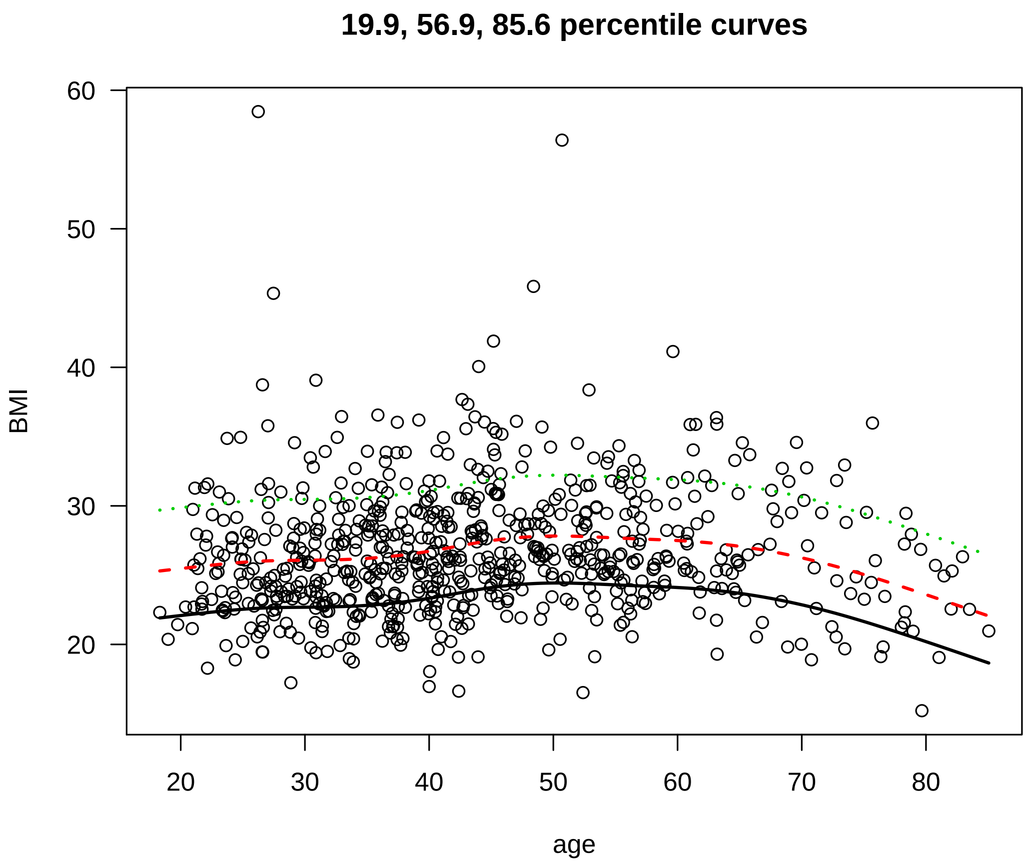

Figure 11: *Quantile plot from* `alsqreg()` *applied to* `bminz`*: several* w *values.*

For each value in the `w` vector we have

```
> fit3@extra

$w.als
[1]  0.1  1.0 10.0

$M
[1] 3
```

```
$n
[1] 700

$y.names
[1] "w.als=0.1" "w.als=1"   "w.als=10"

$percentile
w.als=0.1   w.als=1  w.als=10
 19.85714  56.85714  85.57143

$individual
[1] TRUE

$deviance
w.als=0.1   w.als=1  w.als=10
 2972.257 14285.684 54564.238
```

3.2 Poisson regression

ALS is a special case of AML estimation. The latter was used by Efron (1992) to obtain expectiles from the Poisson distribution. More generally,

$$S_w(\boldsymbol{\beta}) \;=\; \sum_{i=1}^{n} w_i \, D_w(y_i, \mu_i(\boldsymbol{\beta})) \tag{27}$$

is minimized (cf. (24)), where

$$D_w(\mu, \mu') \;=\; \begin{cases} D(\mu, \mu') & \text{if } \mu \leq \mu', \\ w\, D(\mu, \mu') & \text{if } \mu > \mu'. \end{cases} \tag{28}$$

Here, D is the deviance from a model in the exponential family

$$g_\eta(y) \;=\; \exp(\eta y - \psi(\eta)).$$

The user is referred to that paper for more details. The VGAM family function `amlpoisson()` is an implementation of AML estimation to the Poisson model.

Here is a simple example.

```
> set.seed(1234)
> mydat = data.frame(x = sort(runif(n <- 200)))
> mydat = transform(mydat, y = rpois(n, exp(0 - sin(8 * x))))
> (fit = vgam(y ~ s(x), fam = amlpoisson(w.aml = c(0.02,
+     0.2, 1, 5, 50)), data = mydat))

Call:
vgam(formula = y ~ s(x), family = amlpoisson(w.aml = c(0.02,
    0.2, 1, 5, 50)), data = mydat)

Degrees of Freedom: 1000 Total; 975.37 Residual
Residual Deviance: 1332.287
```

Then

```
> with(mydat, plot(x, jitter(y), main = paste(paste(round(fit@extra$percentile,
+      dig = 1), collapse = ", "), "percentile curves")),
+      col = "blue", las = 1)
> with(mydat, matlines(x, fitted(fit), col = c(1, 2, 4),
+      lwd = 2))
```

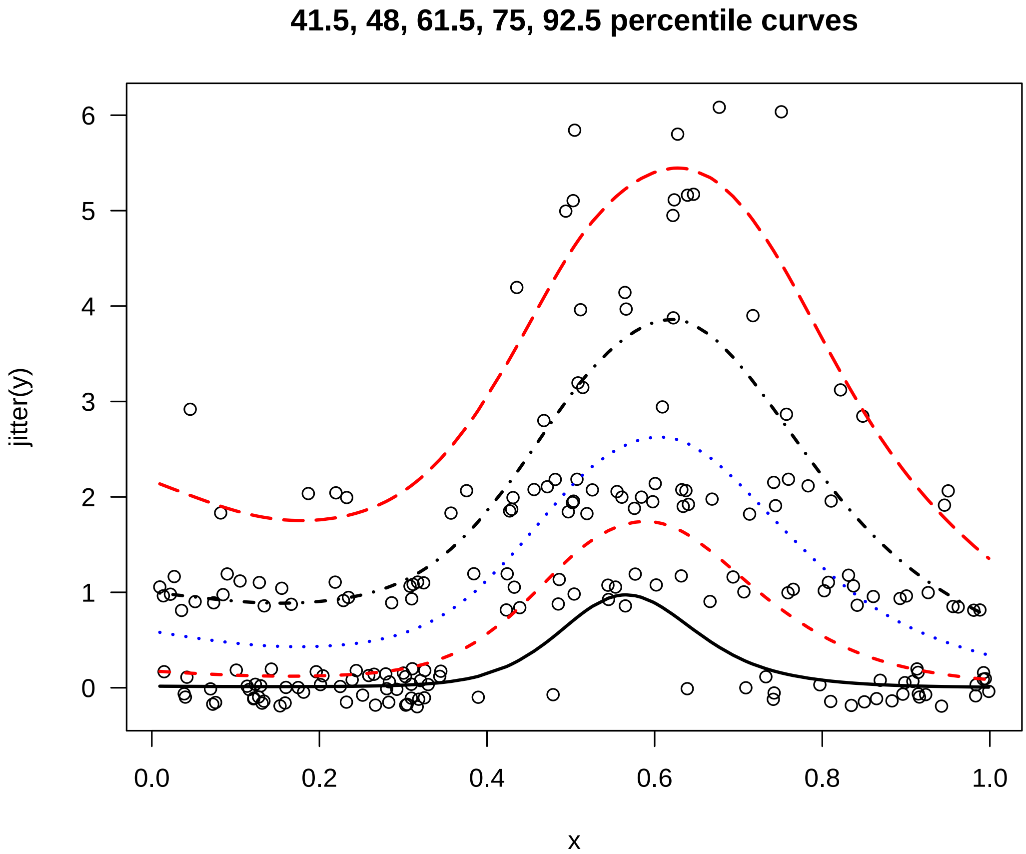

Figure 12: *Quantile plot from `amlpoisson()` from simulated data. The response has been jitted to aid clarity.*

The results look ok. Some more results can be seen with

```
> fit@extra

$w.aml
[1]  0.02  0.20  1.00  5.00 50.00

$M
[1] 5

$n
[1] 200

$y.names
[1] "w.aml=0.02" "w.aml=0.2"  "w.aml=1"    "w.aml=5"    "w.aml=50"

$individual
[1] TRUE

$percentile
w.aml=0.02  w.aml=0.2    w.aml=1    w.aml=5   w.aml=50
      41.5       48.0       61.5       75.0       92.5

$deviance
w.aml=0.02  w.aml=0.2    w.aml=1    w.aml=5   w.aml=50
  22.16201   95.65031  210.03791  374.37732  630.05903
```

The VGAM family function `amlbinomial()` fits a logistic regression using AML estimation. It is based on the methodology of Efron (1992) which is based on the exponential family.

4 Discussion

Quantile regression is an interesting and useful type of regression that allows modelling the entire distribution of a response rather than just through its mean function. It is a large topic and it has a wide range of applications. In this document we have focussed on estimating quantiles given a random sample from a reference population, as well as the related topic of expectiles.

Another type of related data that is often collected is extreme value data where the R_i largest values of some phenomenon is measured at each value of a covariate x_i. For example, the 20 most intelligent children in each age group (integer valued) in a large school are tested with the same IQ test. Fixing the definition of "gifted" as being within the top 1%, the data helps determine the cut-off score for that particular IQ test for each age group in order to screen for gifted children. Compared to the simple random sampling scheme assumed in this paper, this type of data consists of the top portion of the data set only and omitting the vast majority of the data. This censoring often occurs from the literature because tables only include the most extreme values, e.g., worse disasters, largest insurance claims, most exceptional people. Fortunately, the VGAM framework based on penalized likelihood can accommodate such problems too! Such extreme value data can be modelled using the Gumbel distribution, which is a generalized extreme value distribution with $\xi = 0$. Then one can apply a penalized likelihood argument similar to the LMS method to obtain extreme centiles (e.g., $\alpha = 0.99$) based on such data. See Rosen and Cohen (1996), who extended the work of Smith (1986). VGAM family functions have been written to implement their method, with a methods function for `qtplot()` applicable as well. See the VGAM documentation on extreme value data.

The ability to handle extreme value data illustrates some of the capability of the VGAM framework: its sheer breadth enables many models to be encompassed, and therefore can be fitted using a function such as `vglm()` and `vgam()`, often with only the need to write a single VGAM family function.

A second type of data is longitudinal data. Work needs to be done for this common form of data in terms of applying quantile regression.

Exercises

1. Consider the model $y_i = \boldsymbol{\beta}^T \boldsymbol{x}_i + \varepsilon_i, \quad i = 1, \ldots, n$, where the errors are i.i.d. and that we observe $y_i^* = \max(y_i, a)$ where a is some real censoring point. Using the equivariance property to monotone transformations show that

$$Q_{y_i^*}(\tau|\boldsymbol{x}_i) = Q_{f(y_i)}(\tau|\boldsymbol{x}_i) = f\{Q_{y_i}(\tau|\boldsymbol{x}_i)\} = f(\boldsymbol{\beta}^T \boldsymbol{x}_i) = \max(\boldsymbol{\beta}^T \boldsymbol{x}_i, a).$$

   Explain why this means we can estimate $\boldsymbol{\beta}(\tau)$ by solving

$$\hat{\boldsymbol{\beta}}(\tau) = \mathrm{argmin}_{\boldsymbol{\beta} \in \mathbf{R}^p} \sum_{i=1}^{n} \rho_\tau\Big(y_i - \max(\boldsymbol{\beta}^T \boldsymbol{x}_i, a)\Big).$$

2. Repeat the BMI analyses of Section 2.5 using `lms.bcn()` and `lms.bcg()` instead. Is there any significant difference between the results?

3. To check the fit of Figure 6 count the number of observations above the 99% curve. Is it approximately 1% of the sample size? Do the same for below the 1% curve.

4. Add to VGAM the type of residual described in (21). That is, write S functions called `residuals.lms.bcn()`, `residuals.lms.bcg()` and `residuals.lms.yjn()` to output the residuals in defined in (21), as well as the usual residuals available for VGLMs and VGAMs.

5. The function `lms.bcg()` does not handle $\lambda = 0$. Modify the function so that it does. To do this, you will first need to use (22) to derive ℓ_i and the first two derivatives.

6. Check that the derivatives in `lms.bcg()` are correct.

7. John and Draper (1980) propose an alternative to the power family transformation which they called the *modulus* transformation:

$$y^{(\lambda)} = \begin{cases} \operatorname{sgn}\left\{\dfrac{(|y|+1)^{\lambda}-1}{\lambda}\right\}, & \lambda \neq 0, \\ \operatorname{sgn}\left\{\log(|y|+1)\right\}, & \lambda = 0. \end{cases} \tag{29}$$

   It is a one-parameter transformation, and is appropriate for dealing with a fairly summetric but non-normal error distribution. Show that when $\lambda < 0$, observations thus transformed are restricted to the interval $\{1/\lambda, 1/(-\lambda)\}$. Write a VGAM family function to implement this as an LMS-type method transforming to normality. Call it `lms.jdn()`, and make sure it calls `qtplot.lms.jdn()` to compute the quantiles. Once `lms.jdn()` is working, write the functions `cdf.lms.jdn()` and `deplot.lms.jdn()` so that `cdf()` and `deplot()` work for this model.

8. With several of your own data sets (these may be simulated data) compare `lms.bcn()` with `lms.yjn()`. Do you find any method does better in general than the other?

9. Improve the VGAM code so that if `fit` is a LMS-type model with `s()` terms then `qtplot(fit, deriv=1)` will plot the first derivatives of the quantiles. Of course, `deriv=0` is default, and `qtplot(fit, deriv=2)` should be handled too.

10. Hyndman et al. (1996) show that, during the summer in Melbourne, the distribution of temperature for a given day, given the previous day's temperature, is bimodal. See Figure 2. Attempt to write a VGAM family function that applies a Yeo-Johnson transformation to obtain

$$\frac{1}{2}N(-\delta,1) + \frac{1}{2}N(\delta,1)$$

    where $\delta > 0$. Does this method work?

11. Implement the method described by Rigby and Stasinopoulos (2006) as a VGAM family function. Call it `lms.bct()`.

12. Both quantile and expectile regression can be thought of as arising when the data each come from a specific distribution.

    (a) Suppose the observations y_i are a random sample drawn from an asymmetric double exponential (aka Laplace) distribution

$$f(y|\xi) = \tau(1-\tau)\exp(-\rho_\tau(y-\xi)), \qquad -\infty < y, \xi < \infty. \tag{30}$$

    Show that f is a valid pdf and that the MLE of ξ means solving (7).

(b) Suppose the y_i are a random sample drawn from an asymmetric normal distribution

$$f_\omega(y;\mu) \;=\; \frac{2\,\sqrt{\omega(1-\omega)}}{\sqrt{\pi}\left(\sqrt{\omega}+\sqrt{1-\omega}\right)}\;\exp\left(-\rho_\omega^{[2]}(y-\mu)\right), \qquad -\infty < y, \mu < \infty. \quad (31)$$

Show that f is a valid pdf and that the MLE of μ results in an expectile regression (12). Plot (31) for $\mu = 1$ and $\omega = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, \frac{7}{8}$.

13. Obtain expressions for $g(t)$ and $G(t)$ in equations (14)–(15) for the standard normal, standard uniform and standard exponential distributions and plot them as in Figure 1.

Software Changes

3 Jan 2004 `deplot.lmscreg()` has argument `at` replaced by `y`. In the output, the names `at` and `y` replaced by `y` and `density` respectively.

ACKNOWLEDGEMENTS

The author wishes to thank Professor Peter Green for providing the Lopatatzidis and Green manuscript, and T. Cole for helpful comments.

References

Abdous, B., Rémillard, B., 1995. Relating quantiles and expectiles under weighted-symmetry. Ann. Inst. Statist. Math. 47 (2), 371–384.

Aigner, D. J., Amemiya, T., Poirer, D., 1976. On the estimation of production frontiers: Maximum likelihood estimation of the parameters of a discontinuous density function. International Economic Review 17, 377–396.

Buchinsky, M., 1998. Recent advances in quantile regression models: a practical guide for empirical research. The Journal of Human Resources 33 (1), 88–126.

Campbell, S., Newman, G. B., 1971. Growth of the fetal biparietal diameter during normal pregnancy. Journal of Obstetrics and Gynaecology of the British Commonwealth 78 (6), 513–519.

Chambers, J. M., Hastie, T. J. (Eds.), 1993. Statistical Models in S. Chapman & Hall, New York.

Chitty, L. S., Altman, D. G., Henderson, A., Campbell, S., 1994. Charts of fetal size: 2, head measurements. British Journal of Obstetrics 101 (1), 35–43.

Cizek, P., 2000. Quantile regression. In: Härdle, W., Hlávka, Z., Klinke, S. (Eds.), XploRe© Application Guide. Springer-Verlag, Berlin, pp. 19–48.

Cole, T. J., Freeman, J. V., Preece, M. A., 1998. British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. Statistics in Medicine 17 (4), 407–429.

Cole, T. J., Green, P. J., 1992. Smoothing reference centile curves: The LMS method and penalized likelihood. Statistics in Medicine 11 (10), 1305–1319.

Efron, B., 1991. Regression percentiles using asymmetric squared error loss. Statistica Sinica 1 (1), 93–125.

Efron, B., 1992. Poisson overdispersion estimates based on the method of asymmetric maximum likelihood. Journal of the American Statistical Association 87 (417), 98–107.

Gasser, T., Ziegler, P., Seifert, B., Prader, A., Molinari, L., Largo, R., 1994. Measures of body mass and of obesity from infancy to adulthood and their appropriate transformation. Annals of Human Biology 21 (2), 111–125.

Green, P. J., Silverman, B. W., 1994. Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach. Chapman & Hall, London.

Hao, L., Naiman, D. Q., 2007. Quantile Regression (Quantitative Applications in the Social Sciences). Vol. 149. Sage Publications.

Heagerty, P. J., Pepe, M. S., 1999. Semiparametric estimation of regression quantiles with application to standardizing weight for height and age in us children. Applied Statistics 48 (4), 533–551.

Hyndman, R. J., Bashtannyk, D. M., Grunwald, G. K., 1996. Estimating and visualizing conditional densities. Journal of Computational and Graphical Statistics 5 (4), 315–336.

John, J. A., Draper, N. R., 1980. An alternative family of transformations. Applied Statistics 29 (2), 190–197.

Jones, M. C., 1994. Expectiles and M-quantiles are quantiles. Statistics & Probability Letters 20 (2), 149–153.

Koenker, R., 2005. Quantile Regression. Cambridge University Press, Cambridge.

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

Koenker, R., Hallock, K. F., 2001. Quantile regression: An introduction. Journal of Economic Perspectives 15 (4), 143–156.

Newey, W. K., Powell, J. L., 1987. Asymmetric least squares estimation and testing. Econometrica 55 (4), 819–847.

Rigby, R. A., Stasinopoulos, D. M., 2006. Using the Box-Cox t-distribution in GAMLSS to model skewness and kurtosis. Statistical Modelling 6 (3), 209–229.

Rousseeuw, P. J., 1984. Least median of squares regression. Journal of the American Statistical Association 79 (388), 871–880.

Royston, P., 1991. Constructing time-specific reference ranges. Statistics in Medicine 10 (5), 675–690.

Royston, P., Wright, E. M., 1998. A method for estimating age-specific reference intervals ('normal ranges') based on fractional polynomials and exponential transformation. Journal of the Royal Statistical Society, Series A, General 161 (1), 79–101.

Royston, P., Wright, E. M., 2000. Goodness-of-fit statistics for age-specific reference intervals. Statistics in Medicine 19 (21), 2943–2962.

Taylor, J. W., 2008. Estimating value at risk and expected shortfall using expectiles. Journal of Financial Econometrics 6.

Wade, A. M., Ades, A. E., 1994. Age-related reference ranges: significance tests for models and confidence intervals for centiles. Statistics in Medicine 13 (22), 2359–2367.

Wei, Y., Pere, A., Koenker, R., He, X., 2006. Quantile regression methods for reference growth charts. Statistics in Medicine 25 (8), 1369–1382.

Wright, E. M., Royston, P., 1997. A comparison of statistical methods for age-related reference intervals. Journal of the Royal Statistical Society, Series A, General 160 (1), 47–69.

Yamai, Y., Yoshiba, T., 2002. On the validity of value-at-risk: Comparative analyses with expected shortfall. Monetary and Economic Studies 20 (1), 57–85.

Yao, Q., Tong, H., 1996. Asymmetric least squares regression estimation: a nonparametric approach. J. Nonparametr. Statist. 6 (2-3), 273–292.

Yee, T. W., 1998. On an alternative solution to the vector spline problem. Journal of the Royal Statistical Society, Series B, Methodological 60 (1), 183–188.

Yee, T. W., 2002. An implementation for regression quantile estimation. In: Härdle, W., Rönz, B. (Eds.), Proceedings in Computational Statistics COMPSTAT 2002. Physica-Verlag, Heidelberg, pp. 3–14.

Yee, T. W., 2004. Quantile regression via vector generalized additive models. Statistics in Medicine 23 (14), 2295–2315.

Yee, T. W., Wild, C. J., 1996. Vector generalized additive models. Journal of the Royal Statistical Society, Series B, Methodological 58 (3), 481–493.

Yeo, I.-K., Johnson, R. A., 2000. A new family of power transformations to improve normality or symmetry. Biometrika 87 (4), 954–959.

Yu, K., Lu, Z., Stander, J., 2003. Quantile regression: applications and current research areas. Journal of the Royal Statistical Society, Series D (The Statistican) 52 (3), 331–350.