Efficient quantile regression with auxiliary information

Ursula U. Müller and Ingrid Van Keilegom

Abstract We discuss efficient estimation in quantile regression models where the quantile regression function is modeled parametrically. Additionally we assume that auxiliary information is available in the form of a conditional constraint. This is, for example, the case if the mean regression function or the variance function can be modeled parametrically, e.g. by a line or a polynomial. In this paper we describe efficient estimators of parameters of the quantile regression function for general conditional constraints and for examples of more specific constraints. We do this more generally for a model with responses missing at random, for which an efficient estimator is provided by a complete case statistic. This covers the usual model as a special case. We discuss several examples and illustrate the results with simulations.

1 Introduction

Completely observed data. The quantile regression model (Koenker and Bassett [2], Koenker [1]) for a random sample (X_i, Y_i), $i = 1, \dots, n$, assumes that the conditional quantile of a response variable Y given a covariate vector X can be modeled parametrically, i.e. it can be written as a parametric quantile regression function $q_\theta(X)$, $\theta \in \mathbb{R}^d$. In this article we consider, more generally, a class of regression models that can be written in the form

$$E\{a_\vartheta(X,Y)|X\} = 0, \quad a_\theta = (a_{1\theta}, \tilde{a}_\theta^\top)^\top, \tag{1}$$

Ursula U. Müller
Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA, e-mail: uschi@stat.tamu.edu

Ingrid Van Keilegom
Institut de statistique, Université catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium, e-mail: ingrid.vankeilegom@uclouvain.be

where the true parameter ϑ belongs to the interior of some compact parameter space $\Theta \subset \mathbb{R}^d$. The first component of the k-dimensional vector a_θ is

$$a_{1\theta}(X,Y) = p - 1\{Y - q_\theta(X) < 0\}, \quad p \in (0,1).$$

This specifies the familiar quantile regression model since

$$\begin{aligned} &0 = E\{a_{1\vartheta}(X,Y)|X\} = E[p - 1\{Y - q_\vartheta(X) < 0\}|X] \\ \iff\ &P\{Y < q_\vartheta(X)|X\} = p. \end{aligned} \tag{2}$$

The vector $\tilde{a}_\theta$ represents auxiliary information in the form of $k-1$ conditional parametric constraints. This is the case, for example, if there are reliable parametric models for certain moments of the conditional distribution of Y given X, including the conditional mean and the conditional variance.

Note that the number of parameters, d, and the number of equations, k, are unrelated. Later on we will transform the equations so as to obtain as many equations as parameters. Also note that the vector θ contains the parameters that determine the model for the p-th quantile, as well as the parameters on which the auxiliary information depends. Usually when we have auxiliary information the latter set is part of (or equal to) the former, but this is not necessarily the case.

We are interested in finding an efficient estimator for ϑ. Efficient estimation of ϑ in model (1), with an arbitrary vector a_θ of constraints, has been addressed by Müller and Van Keilegom [5]. There we also briefly discuss the quantile regression model (2) as an example of a one-dimensional constraint, without assuming the presence of the additional vector $\tilde{a}_\theta$. Let us discuss this model first. The usual estimator under model (2) is based on the check function approach, and solves the quantile regression estimating equation

$$\sum_{i=1}^n \dot{q}_\theta(X_i)[p - 1\{Y_i - q_\theta(X_i) < 0\}] = 0$$

with respect to θ (see e.g. [1]) or more precisely, it minimizes

$$\Big\| \sum_{i=1}^n \dot{q}_\theta(X_i)[p - 1\{Y_i - q_\theta(X_i) < 0\}] \Big\| \tag{3}$$

with respect to θ, where $\|\cdot\|$ denotes the Euclidean norm, since an exact solution of the above equation might not exist. Here, $\dot{q}_\theta(X)$ denotes the $(d \times 1)$-vector of partial derivatives $\partial/(\partial\theta_j)\, q_\theta(X)$, $j = 1, \ldots, d$. This is indeed an unbiased estimating equation since

$$E\big(\dot{q}_\vartheta(X)[p - 1\{Y - q_\vartheta(X) < 0\}]\big) = E\big(\dot{q}_\vartheta(X) E[p - 1\{Y - q_\vartheta(X) < 0\}|X]\big) = 0.$$

This calculation shows that one could, more generally, obtain a consistent estimator $\hat{\vartheta}$ by minimizing the norm of a weighted sum

$$\left\| \sum_{i=1}^{n} W_\theta(X_i)[p - 1\{Y_i - q_\theta(X_i) < 0\}] \right\|,$$

where W_θ is a d-dimensional vector of weights. Müller and Van Keilegom [5] proved that an asymptotically efficient estimator of ϑ is obtained for the weight vector

$$W_\theta(X) = -\frac{f_{Y|X}\{q_\theta(X)\}\dot{q}_\theta(X)}{p^2 + (1-2p)F_{Y|X}\{q_\theta(X)\}}, \tag{4}$$

with $f_{Y|X}(y) = \frac{d}{dy}F_{Y|X}(y)$ (provided it exists) and $F_{Y|X}(y) = P\{Y \leq y|X\}$. A simpler (but asymptotically equivalent) version of this estimator is based on weights

$$f_{Y|X}\{q_\theta(X)\}\dot{q}_\theta(X), \tag{5}$$

since the denominator in (4) equals $p - p^2$ for $\theta = \vartheta$ and hence it does not need to be estimated. The weight vector is undetermined in both cases: it involves the unknown conditional density $f_{Y|X}\{q_\theta(X)\}$ (and $F_{Y|X}$ in the first case) and must therefore be replaced by a suitable consistent estimator. Using these estimated weight vectors in the estimating equation above will yield two asymptotically efficient estimators of ϑ.

Note that if we use the simpler weights (5), then an asymptotically efficient estimator of ϑ is obtained by minimizing

$$\left\| \sum_{i=1}^{n} f_{Y|X}\{q_\theta(X_i)\}\dot{q}_\theta(X_i)[p - 1\{Y_i - q_\theta(X_i) < 0\}] \right\|, \tag{6}$$

which is different from the widely used and commonly accepted estimator given in (3), that corresponds to the check function approach, and that is in fact not efficient.

In this paper we consider model (1) which, aside from (2), assumes that auxiliary information in the form of a constraint $E\{\tilde{a}_\vartheta(X,Y)|X\} = 0$ is available. This is related to Tang and Leng [9] who consider the *linear* quantile regression model with $q_\beta(X) = X^\top\beta$, where β is a parameter vector. They assume additional information in the form of an *unconditional* constraint, $E\{\tilde{a}_\vartheta(X,Y)\} = 0$, and suggest an empirical likelihood approach. Such a constraint applies if, for example, there is knowledge about unconditional moments of the joint distribution of (X,Y). This is conceptually different from our model, since models for moments of the conditional distribution are not included, e.g. models for the conditional mean $E(Y|X)$ or the variance function mentioned above. Another related paper that does consider a conditional constraint is [7] by Qin and Wu, who estimate conditional quantiles. However, neither the quantiles nor the constraint are modeled parametrically. There is more literature dating back several years on estimating unconditional quantiles when auxiliary information is available; see e.g. [4, 8, 10].

Missing data. As in [5] we now assume further that some responses Y are allowed to be *missing at random* (MAR). This means that one has i.i.d. observations $(X_i, \delta_i Y_i, \delta_i)$, $i = 1, \ldots, n$, having the same distribution as $(X, \delta Y, \delta)$, with indicator

$\delta = 1$ if Y is observed and $\delta = 0$ if Y is missing. In particular one assumes that the missingness mechanism depends only on X,

$$P(\delta = 1|X,Y) = P(\delta = 1|X) = \pi(X),$$

where $\pi(\cdot)$ is the propensity score. This implies that Y and δ are conditionally independent given X. One reason for considering the MAR model is that it contains the "full model", where no data are missing, as a special case with $\pi(\cdot) = 1$ (and all indicators $\delta = 1$), so both models can be treated together. Of course, this does not always apply since the construction of reasonable estimators can be quite different, depending on the model. Here we are specifically interested in estimating the *parameter* ϑ. In this case we can work with a simple complete case estimator (an estimator for the full model that ignores observations that are only partially observed). One possibility is to estimate ϑ by a minimizer of

$$\Big\|\sum_{i=1}^{n} \delta_i W_\theta(X_i) a_\theta(X_i,Y_i)\Big\|$$

with respect to θ, where W_θ is a $d \times k$ weight matrix. In this way we obtain a system of d equations with d unknown parameters, although we started off with k constraints. That the weighted sum above leads to an unbiased estimator is easy to see using the fact that δ and Y are conditionally independent given X under the MAR assumption. Beyond that, one can show that a complete case version of *any* consistent estimator of ϑ is again consistent. This can be seen by applying the *transfer principle for complete case statistics*, introduced by Koul et al. [3], which makes it possible to adapt results for the full model to the MAR model. The transfer principle provides the limiting distribution of a complete case version of a statistic as the limiting distribution of that statistic conditional on $\delta = 1$. To verify consistency, one only has to show that the functional of interest, i.e. in our case the parameter vector ϑ, is the same in the unconditional and in the conditional model. This is indeed true, since ϑ is in both models defined as a solution of the same conditional constraint:

$$\begin{aligned} 0 = E\{a_\vartheta(X,Y)|X\} &= \frac{E(\delta|X)E\{a_\vartheta(X,Y)|X\}}{E(\delta|X)} = \frac{E\{\delta a_\vartheta(X,Y)|X\}}{E(\delta|X)} \\ &= E\{a_\vartheta(X,Y)|X,\delta = 1\}. \end{aligned}$$

So far we know that an efficient estimator of ϑ in the (unextended) quantile regression model with MAR responses is given as a minimizer of

$$\Big\|\sum_{i=1}^{n} \delta_i \widehat{W}_\theta(X_i)[p - 1\{Y_i - q_\theta(X_i) < 0\}]\Big\|$$

with respect to θ, where $\widehat{W}_\theta$ is a suitable estimator of the weight vector W_θ given in (4) or (5) (see Section 3.4 in [5].)

In the next section we will provide the formulas for an efficient estimator of ϑ in the general quantile regression model (1) with auxiliary information in the form of a general conditional constraint. In Section 3 we discuss three examples of auxiliary information, namely when we have a parametric model for the mean and for the median, respectively, and when we have two responses that share the same p-th quantile. Section 4 shows the results of a small simulation study, and we end this paper in Section 5 with some general conclusions.

2 The estimator

As in [5] we write

$$\begin{aligned} &\ell_\theta(X,Y) = -W_\theta(X)a_\theta(X,Y), \qquad I = E\{\delta \ell_\vartheta(X,Y)\ell_\vartheta(X,Y)^\top\}, \\ &W_\theta(X) = \big[\frac{\partial}{\partial\theta}E\{a_\theta(X,Y)|X\}\big]^\top A_\theta(X)^{-1}, \end{aligned} \tag{7}$$

now with

$$a_\theta(X,Y) = \begin{pmatrix} a_{1\theta}(X,Y) \\ \tilde{a}_\theta(X,Y) \end{pmatrix} = \begin{pmatrix} p - 1\{Y - q_\theta(X) < 0\} \\ \tilde{a}_\theta(X,Y) \end{pmatrix}, \tag{8}$$

where for $\theta = \vartheta$ the $k \times k$ matrix $A_\vartheta(X)$ is given by

$$A_\vartheta(X) = E\big\{a_\vartheta(X,Y)a_\vartheta(X,Y)^\top|X\big\},$$

and where for $\theta \neq \vartheta$, the matrix $A_\theta(X)$ is obtained by replacing ϑ by θ in the expression of $A_\vartheta(X)$. Note that in general $A_\theta(X)$ and $E\big\{a_\theta(X,Y)a_\theta(X,Y)^\top|X\big\}$ are different, since in certain entries of the matrix $A_\vartheta(X)$ the parameter ϑ will disappear when using the underlying model assumptions. For example, the first entry is $E([p-1\{Y-q_\vartheta(X)<0\}]^2|X) = p^2 + (1-2p)F_{Y|X}\{q_\vartheta(X)\} = p - p^2$, which is independent of ϑ.

The estimator in model (1) can then be written $\hat{\vartheta} = \mathrm{argmin}_\theta \|\sum_{i=1}^n \delta_i \ell_\theta(X_i,Y_i)\|$. In the full model we simply set $\delta_i = 1$ for $i = 1,\dots,n$, i.e. the indicators δ can be ignored. Under the assumptions stated in [5], $\hat{\vartheta}$ is asymptotically linear,

$$n^{1/2}(\hat{\vartheta} - \vartheta) = I^{-1} n^{-1/2} \sum_{i=1}^n \delta_i \ell_\vartheta(X_i,Y_i) + o_p(1),$$

and efficient in the sense of Hájek and Le Cam.

Let us take a closer look at the formula of the weight matrix. The estimating equation for model (1) involves W_θ and a_θ given in equations (7) and (8). Using the specific form of a_θ, the matrix $W_\theta(X)$ computes to

$$W_\theta(X) = \Big(-f_{Y|X}\{q_\theta(X)\}\dot{q}_\theta(X) \ \ \frac{\partial}{\partial\theta}E\{\tilde{a}_\theta(X,Y)^\top|X\} \Big) A_\theta(X)^{-1}, \tag{9}$$

where $A_\vartheta(X)$ is the matrix

$$\begin{pmatrix} p-p^2 & E([p-1\{Y<q_\vartheta(X)\}]\tilde{a}_\vartheta(X,Y)^\top|X) \\ E([p-1\{Y<q_\vartheta(X)\}]\tilde{a}_\vartheta(X,Y)|X) & E\{\tilde{a}_\vartheta(X,Y)\tilde{a}_\vartheta(X,Y)^\top|X\} \end{pmatrix},$$

and where the matrix $A_\theta(X)$ is obtained by replacing in the formula of $A_\vartheta(X)$ every ϑ that does not disappear after using the model assumptions, by θ.

In Section 3.1 of [5] it is shown that if we replace the weight matrix $W_\theta(X)$ given in (9) by an estimator $\widehat{W}_\theta(X)$ that is uniformly consistent in θ and x, i.e. $\sup_{\theta\in\Theta}\sup_x \|\widehat{W}_\theta(x)-W_\theta(x)\| = o_p(1)$, then the resulting estimator (that depends now on $\widehat{W}_\theta$ instead of W_θ) remains asymptotically efficient.

Note that the weight matrix $W_\theta(X)$ involves, among others, the conditional density $f_{Y|X}$. The density can be estimated by using e.g. a kernel smoother of the form

$$\hat{f}_{Y|X=x}(y) = \sum_{i=1}^n \frac{\delta_i k_b(x-X_i)k_h(y-Y_i)}{\sum_{i=1}^n \delta_i k_b(x-X_i)},$$

with kernel k and smoothing parameters b and h, and where $k_b(\cdot) = k(\cdot/b)/b$ for any bandwidth b. The estimation of the other components of the weight matrix $W_\theta(X)$ depends on the specific form of the auxiliary information. We will consider three examples in the next section.

3 Examples

Example 1. We start with a situation in which we have some auxiliary information concerning the conditional mean $r(X) = E(Y|X)$. Suppose that $r(X)$ can be modeled parametrically $r(X) = r_\vartheta(X)$. The function $\tilde{a}_\theta(X,Y)$ is given by

$$\tilde{a}_\theta(X,Y) = Y - r_\theta(X),$$

i.e. $k=2$ and, for example, $r_\theta(X) = \theta^\top X$. Some straightforward algebra shows that the optimal weight matrix is then given by

$$W_\theta(X) = \Big(-f_{Y|X}\{q_\theta(X)\}\dot{q}_\theta(X) \quad -\dot{r}_\theta(X)\Big)A_\theta(X)^{-1},$$

where $A_\theta(X)$ is the 2×2 matrix

$$\begin{pmatrix} p-p^2 & p\,r_\theta(X) - E(1\{Y<q_\theta(X)\}]Y|X) \\ p\,r_\theta(X) - E(1\{Y<q_\theta(X)\}]Y|X) & \mathrm{Var}(Y|X) \end{pmatrix}.$$

The conditional variance $\mathrm{Var}(Y|X)$ can be estimated by standard kernel smoothers, whereas a consistent estimator of the term $E(1\{Y<q_\theta(X)\}]Y|X)$ in the off-diagonal element of the matrix $A_\theta(X)$ is given by

$$\sum_{i=1}^n \frac{\delta_i k_b(x-X_i)1\{Y_i<q_\theta(X_i)\}Y_i}{\sum_{i=1}^n \delta_i k_b(x-X_i)},$$

with kernel k and smoothing parameter b.

Example 2. Let us now consider the case when $p \neq 1/2$, i.e. we want to estimate quantiles other than the median, and we have some auxiliary information regarding the median. For instance we know that the p-th quantile and the median are parallel functions (of X). Let us denote the parametric model for the median by v_θ, and so

$$\tilde{a}_\theta(X,Y) = 1/2 - 1\{Y - v_\theta(X) < 0\}.$$

In this case, it is easily seen that

$$W_\theta(X) = \Big(-f_{Y|X}\{q_\theta(X)\}\dot{q}_\theta(X) \;\; -f_{Y|X}\{v_\theta(X)\}\dot{v}_\theta(X) \Big) A_\theta(X)^{-1},$$

where

$$A_\theta(X) = \begin{pmatrix} p - p^2 & p \wedge (1/2) - p/2 \\ p \wedge (1/2) - p/2 & 1/4 \end{pmatrix},$$

since for $\theta = \vartheta$ the off-diagonal element is given by

$$\begin{aligned} & p/2 - pF_{Y|X}\{v_\vartheta(X)\} - (1/2)F_{Y|X}\{q_\vartheta(X)\} + F_{Y|X}\{q_\vartheta(X) \wedge v_\vartheta(X)\} \\ & \quad = -p/2 + \{p \wedge (1/2)\}. \end{aligned}$$

The estimation of this weight matrix only involves the estimation of the conditional density $f_{Y|X}$, which was discussed in the previous section.

Example 3. The model considered in this paper can be extended to the case where we have a multivariate response $Y = (Y_1, \ldots, Y_{d_Y})^\top$. For simplicity we consider the bivariate case. Let $(X_i, \delta_i Y_i, Y_i)$, $i = 1, \ldots, n$, be an i.i.d. sample, where $Y_i = (Y_{1i}, Y_{2i})^\top$, $X_i = (X_{1i}, \ldots, X_{d_X i})^\top$ and $\delta_i = (\delta_{1i}, \delta_{2i})^\top$. We could then consider the case where the two responses have the same conditional p-th quantile, which means that we should take

$$a_{1\theta}(X,Y) = p - I\{Y_1 - q_\theta(X) < 0\},$$

and

$$\tilde{a}_\theta(X,Y) = p - I\{Y_2 - q_\theta(X) < 0\},$$

and the true parameter vector ϑ satisfies $F_{Y_1|X}\{q_\vartheta(X)\} = p = F_{Y_2|X}\{q_\vartheta(X)\}$. As in the previous two examples, it can be seen that the weight matrix that leads to an asymptotically efficient estimator of ϑ is given by

$$W_\theta(X) = \Big(-f_{Y_1|X}\{q_\theta(X)\} \;\; -f_{Y_2|X}\{q_\theta(X)\} \Big) \dot{q}_\theta(X) A_\theta(X)^{-1}$$

with

$$A_\theta(X) = \begin{pmatrix} p - p^2 & F_{Y_1,Y_2|X}\{q_\theta(X), q_\theta(X)\} - p^2 \\ F_{Y_1,Y_2|X}\{q_\theta(X), q_\theta(X)\} - p^2) & p - p^2 \end{pmatrix},$$

where $F_{Y_1,Y_2|X}(y_1,y_2) = P(Y_1 \leq y_1, Y_2 \leq y_2|X)$. To verify the formula for the off-diagonal element of $A_\theta(X)$ consider the corresponding entry in $A_\vartheta(X)$, which computes to

$$\begin{aligned} &p^2 - p[F_{Y_1|X}\{q_\vartheta(X)\} + F_{Y_2|X}\{q_\vartheta(X)\}] + F_{Y_1,Y_2|X}\{q_\vartheta(X), q_\vartheta(X)\} \\ &\quad = -p^2 + F_{Y_1,Y_2|X}\{q_\vartheta(X), q_\vartheta(X)\}. \end{aligned}$$

4 Simulations

In order to gain some insight into the performance of our proposed method if n is finite, we conducted a small simulation study based on 50,000 simulated samples of size $n = 100$. In this study we consider only the case where all responses are observed. Since our estimator for missing data is a complete case statistic, this essentially means that we use all $n = 100$ data pairs (X,Y), and not just a proportion. The comparisons are equally meaningful.

We considered the scenario from Example 1 in the previous section, with a linear quantile regression function $q_\vartheta(X) = \vartheta_1 + \vartheta_2 X$, and with the auxiliary information that the mean regression function is linear as well, $E(Y|X) = r_\vartheta(X) = \vartheta_3 X$. The parameters of interest are ϑ_1 and ϑ_2, whereas ϑ_3 can be regarded a nuisance parameter. In order to create this scenario we generated responses Y given $X = x$ from a normal distribution with mean $r_\vartheta(x) = \vartheta_3 x$ (with $\vartheta_3 = 1$) and standard deviation $\sigma(x) = a + bx$. Modeling r_ϑ and σ linearly suffices to ensure that the quantile function is also linear: we have $p = \Phi[\{q_\vartheta(x) - r_\vartheta(x)\}/\sigma(x)]$ (see (2)), with Φ the distribution function of the standard normal distribution. Solving this with respect to $q_\vartheta(x)$ gives $q_\vartheta(x) = \vartheta_1 + \vartheta_2 x$ with $\vartheta_1 = \vartheta_1(p) = a\Phi^{-1}(p)$, and $\vartheta_2 = \vartheta_2(p) = \vartheta_3 + b\Phi^{-1}(p)$. The covariates X were generated from a uniform$(-1,1)$ distribution.

The results are in Table 1. For simplicity we used the true (3×2) weight matrix W_θ to implement our efficient estimator. We compared it with the two estimators discussed in the introduction that use only the quantile regression structure, namely the check function approach (3) and the estimator that minimizes (6) (based on weights (5)). To compute the latter estimator we also used the true weights. Comparing the estimators that employ the true weights with the check function approach (which does not require estimation of weights) may not be quite fair, we nevertheless find it interesting since the results make us feel confident that our estimator will outperform the usual approach even if an estimated weight matrix $\widehat{W}_\theta$ is used.

Let us briefly discuss these results. We considered two different slopes for $\sigma(x) = 0.6 + bx$, namely $b = -0.5$ and $b = 0.5$. The first case yields a variance reduction and the second case a variance gain as x increases from -1 to 1. In most cases the efficient estimator (EFF) is clearly better than the two approaches that do not exploit the auxiliary information. This is remarkable, in particular when one considers that the optimization algorithm involves an additional parameter. The efficient estimator performs best in the case of a conditional median ($p = 0.5$), which is not surprising

Table 1 Simulated MSEs of parameter estimators of $q_\vartheta(X) = \vartheta_1 + \vartheta_2 X$

		Estimators of ϑ_1			Estimators of ϑ_2		
$\sigma(x)$	p	QR1	QR2	EFF	QR1	QR2	EFF
$0.6-0.5x$	0.25	0.00632	0.00637	0.00143	0.01083	0.00805	0.00271
	0.5	0.00546	0.00513	0.00095	0.00902	0.00601	0.00190
	0.75	0.00523	0.00502	0.00177	0.00610	0.00438	0.00358
$0.6+0.5x$	0.25	0.00248	0.00151	0.00111	0.00294	0.00158	0.00169
	0.5	0.00287	0.00177	0.00088	0.00495	0.00250	0.00110
	0.75	0.00311	0.00209	0.00169	0.00579	0.00365	0.00149

The table entries give the simulated mean squared errors (MSE) for three estimators of ϑ_1 and ϑ_2. The estimator "QR1" is based on the check function approach (3), the estimator "QR2" is the minimizer of (6), and "EFF" is the efficient estimator that uses the auxiliary information that r_ϑ is linear, $r_\vartheta(X) = \vartheta_3 X$.

since we use a normal density $f_{Y|X}$ in our simulations. The conditional median q_ϑ and the conditional mean r_ϑ are the same in this setting.

The proposed estimator lacks performance only in the case $p = 0.25$ with an increasing variance ($b = 0.5$). Here the estimator QR2 is slightly better for ϑ_2. Simulations with larger sample sizes confirm, however, that our estimator indeed outperforms QR2 *asymptotically*. (For example, for $n = 500$ our simulated MSEs for QR2 and EFF were 0.00097 and 0.00048, respectively.)

Comparing the two estimators QR1 and QR2 that use only the quantile regression structure, we notice that both estimate the intercept similarly well in the case of a decreasing variance function. In all other cases the weighted estimator QR2 is better than the check function approach QR1. Since QR2 is efficient in the original quantile regression model that does not assume auxiliary information, this corresponds to the theoretical findings.

5 Concluding remarks

In this paper we studied a parametric quantile regression model in which the responses are allowed to be missing at random (but do not have to be), and in which the covariates are always observed. We were interested in the estimation of a particular conditional quantile when auxiliary information regarding that quantile is available. We constructed an asymptotically efficient estimator of the model parameters based on weighted estimating equations, and studied three examples in more detail. One of these examples was further examined via a small simulation study, which confirmed the effectiveness of the proposed estimation procedure.

There are numerous other situations where auxiliary information is available. We could, for example, have information regarding the variance, the interquartile range, or the quantile of order $1-p$. It would also be interesting to study a model where responses are subject to censoring, or the case with missing covariates. Such extensions definitely seem feasible, but will be somewhat more challenging from a technical point of view. Finally, an interesting project for future work would be to develop an *efficient* empirical likelihood based method to estimate conditional quantiles, in a similar spirit as [6] or [9]. This would provide an alternative (asymptotically equivalent) approach to exploit information in the form of (conditional) constraints. Although our estimator cannot be outperformed *asymptotically*, it is nevertheless possible that there are situations where the empirical likelihood approach performs better for moderate sample sizes, or where it has computational advantages.

Acknowledgements I. Van Keilegom acknowledges financial support from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement No. 203650, from IAP research network P7/06 of the Belgian Government (Belgian Science Policy), and from the contract 'Projet d'Actions de Recherche Concertées' (ARC) 11/16-039 of the 'Communauté française de Belgique', granted by the 'Académie universitaire Louvain'. The authors would like to thank two referees for their helpful comments.

References

1. Koenker, R.: Quantile Regression. Cambridge (2005)
2. Koenker, R., Bassett, G.: Regression quantiles. Econometrica **46**, 33-50 (1978)
3. Koul, H.L., Müller, U.U., Schick, A.: The transfer principle: a tool for complete case analysis. Ann. Statist. **40**, 3031-3049 (2012)
4. Kuk, A.Y.C., Mak, T.K.: Median estimation in the presence of auxiliary information. J. Roy. Statist. Soc. Ser. B **51**, 261-269 (1989)
5. Müller, U.U., Van Keilegom, I.: Efficient parameter estimation in regression with missing responses. Electron. J. Stat., 1200-1219 (2012)
6. Qin, J., Lawless, J.: Empirical likelihood and general estimating equations. Ann. Statist. **22**, 300-325 (1994)
7. Qin, Y.S., Wu, Y.: An estimator of a conditional quantile in the presence of auxiliary information. J. Statist. Plann. Inference **99**, 59-70 (2001)
8. Rao, J.N.K., Kovar, J.G., Mantel, H.J.: On estimating distribution functions and quantiles from survey data using auxiliary information. Biometrika **77**, 365-375 (1990)
9. Tang, C.Y., Leng, C.: An empirical likelihood approach to quantile regression with auxiliary information. Statist. Probab. Lett. **82**, 29-36 (2012)
10. Zhang, B.: Quantile processes in the presence of auxiliary information. Ann. Inst. Statist. Math. **49**, 35-55 (1997)