The estimation of extreme conditional quantiles is an important issue in different scientific disciplines. Up to now, the extreme value literature focused mainly on estimation procedures based on independent and identically distributed samples. Our contribution is a two-step procedure for estimating extreme conditional quantiles. In a first step nonextreme conditional quantiles are estimated nonparametrically using a local version of [Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46, 33–50.] regression quantile methodology. Next, these nonparametric quantile estimates are used as analogues of univariate order statistics in procedures for extreme quantile estimation. The performance of the method is evaluated for both heavy tailed distributions and distributions with a finite right endpoint using a small sample simulation study. A bootstrap procedure is developed to guide in the selection of an optimal local bandwidth. Finally the procedure is illustrated in two case studies.

**Keywords:** Local polynomial estimation; Quantile regression; Extreme value index; Extrapolation

**1 INTRODUCTION**

In this paper we study the estimation of conditional quantiles with emphasis on the range beyond the data. This problem manifests itself in many areas e.g. an insurance company facing a fire insurance portfolio is typically interested in the claim size that will be exceeded once in, say, 10,000 cases given additional factors such as sum insured and type of building. The extreme value literature on the estimation of extreme quantiles focuses mainly on the independent and identically distributed (i.i.d.) case, see for instance Weissman (1978), Pickands (1975), Dekkers *et al.* (1989), Dekkers and de Haan (1989), Beirlant and Matthys (2001, 2003). On the other hand, the literature in the case of covariates is very sparse. Recent contributions are the work of Chernozhukov (1998, 2001), Gijbels and Peng (2000), and Charnes *et al.* (1995). We apply the regression quantile methodology of Koenker and Bassett (1978) in a nonparametric fashion using a two-step procedure. In a first step we locally approximate the conditional quantile function using a polynomial yielding the regression analogues of univariate order statistics. Next, these nonparametric regression quantiles are used in a fashion analogous to the use of...
order statistics in the i.i.d. case for extrapolating beyond the data. This extrapolation is based on recent extreme value techniques see e.g. Beirlant et al. (1999) and Beirlant and Matthys (2001, 2003). This two-stage procedure generalizes the method based on local maxima proposed in Gijbels and Peng (2000). Chernozhukov (1998, 2001) gave a theoretical study of modern extreme value methods based on extreme regression quantiles in order to extrapolate outside the sample range. Further, these papers by Chernozhukov assume an additive error structure. Recently, Hall and Tajvidi (2000) and Davison and Ramesh (2000) proposed nonparametric estimates of some tail characteristics by means of extreme value modelling based on the generalized extreme value distribution and the generalized Pareto distribution, combined with local curve fitting for the corresponding parameter functions. In the present paper a fully nonparametric approach is pursued, using recent extreme value methods on local polynomial quantile regression estimates.

In Section 2 the nonparametric estimation of regression quantiles on the basis of local polynomial approximations to the true conditional quantile function is discussed. Further it is shown how univariate extreme value methods can be used in the extrapolation step i.e. the estimation of extreme conditional quantiles. The performance of this two step method is evaluated in Section 3 with a small sample simulation and this for both heavy tailed and right bounded distributions. Next, in Section 4, we discuss a bootstrap procedure to guide in the selection of an optimal local bandwidth. In a final section the procedure is illustrated with two practical examples.

2 METHODOLOGY

Consider a random variable $Y$ whose distribution depends on the covariate $x$. Without loss of generality we restrict ourselves to the single covariate case. We denote the conditional distribution of $Y$ given $x$ by $F_{Y|x}$ and the associated quantile function by $Q(\theta; x)$. More precisely
\[
Q(\theta; x) = \inf \{ y : F_{Y|x}(y) \geq \theta \} \quad 0 < \theta < 1.
\]

Interest lies in estimating $Q(\theta; x)$ with $\theta$ close to 1, or sometimes $Q(x) := \lim_{\theta \uparrow 1} Q(\theta; x)$ in the case of a finite right endpoint. We assume no knowledge about $F_{Y|x}$ or $Q$ and follow a nonparametric approach. The estimation problem is considered in its full generality without any particular structure on the model (e.g. linear additive/multiplicative models).

Suppose a dataset $(Y_1, x_1), \ldots, (Y_n, x_n)$ of independent observations according to $F_{Y|x}$ is given and we want to use these to estimate $Q(\theta; x^*)$ for a fixed $x^*$. Since $Q$ is unknown we approximate it locally by a polynomial of degree $p$ centered around $x^*$ in the following way
\[
Q(\theta; x) \approx \sum_{j=0}^{p} \beta_j (x - x^*)^j
\]
for $x$ sufficiently close to $x^*$, where
\[
\beta_j = \beta_j(\theta) = \frac{1}{j!} \left. \frac{d^{j} Q(\theta; x)}{dx^j} \right|_{x=x^*}, \quad j = 0, \ldots, p.
\]

In particular note that $\beta_0 = Q(\theta; x^*)$. Consider a window of size $2h$ centered at $x^*$ in which we apply the above polynomial approximation to $Q(\theta; x)$. Within this window we estimate the coefficients of this approximation as follows:
\[
\min_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} I_{[x^*-h, x^*+h]}(x_i) f_0 \left( Y_i - \sum_{j=0}^{p} \beta_j (x_i - x^*)^j \right)
\]  

(1)
with \( f_0(x) = \theta x^+ + (1 - \theta)x^- \), \( x^+ = \max(0, x) \), \( x^- = \max(0, -x) \) and \( I_A(x) = 1 \) if \( x \in A \), 0 otherwise. This optimization problem is a local version of the Koenker–Bassett approach to estimating regression quantiles (Koenker and Bassett, 1978). The maximum estimator for \( Q(x^*) \) proposed by Gijbels and Peng (2000) can be seen as a special case of (1) by setting \( \theta = 1 \) and \( p = 0 \). Noting that the \( \beta_0 \)-component of the solution to Eq. (1) estimates \( Q(0; x^*) \), we denote this component by \( \hat{Q}(0; x^*) \).

Before considering further the regression case let us first recall some facts from the i.i.d. case in order to motivate our approach.

Consider \( Y_1, \ldots, Y_n \) i.i.d. random variables according to distribution function \( F_Y \) and associated quantile function \( Q(\theta) \). Denote the ascending order statistics corresponding to \( Y_1, \ldots, Y_n \) by \( Y_{1,n} \leq \cdots \leq Y_{n,n} \). Further assume that \( F_Y \) is in the domain of attraction of the Generalized Extreme Value distribution i.e. there exist sequences of constants \( (a_n > 0) \) and \( (b_n) \) such that for some \( \gamma \in \mathbb{R} \)

\[
\lim_{n \to \infty} P\left( \frac{Y_{n,n} - b_n}{a_n} \leq y \right) = \exp\left( - (1 + \gamma y)^{-1/\gamma} \right) \quad 1 + \gamma y > 0. \tag{2}
\]

The parameter \( \gamma \), called the extreme value index, gives important information about the tail of the underlying distribution function \( F_Y \), where tails become heavier with increasing \( \gamma \). Consequently, \( \gamma \) will play an important role when estimating \( Q(\theta) \) for \( \theta \) close to 1. The literature on estimating \( \gamma \) on the basis of an i.i.d. sample is very elaborate, see for instance Hill (1975), Pickands (1975), Dekkers et al. (1989), Beirlant et al. (1999), Feuerverger and Hall (1999).

In case \( \gamma > 0 \), the following approximate representation holds for log-spacings of successive order statistics (Beirlant et al., 1999)

\[
j(\log Y_{n-j+1,n} - \log Y_{n-j,n}) \overset{D}{\approx} \left( \gamma + b_{n,k}\left( \frac{j}{k+1} \right)^{-\rho} \right) F_j \quad j = 1, \ldots, k, \tag{3}
\]

with \( b_{n,k} \in \mathbb{R} \), \( \rho < 0 \) and \( F_j, j = 1, \ldots, k, \) independent standard exponential random variables, from which \( \gamma \) can be estimated jointly with \( b_{n,k} \) and \( \rho \) using the maximum likelihood method. Based on this, Beirlant and Matthys (2001) proposed to use

\[
\hat{Q}_k^{(1)}(\theta) = Y_{n-k,n}\left( \frac{k+1}{(n+1)(1-\theta)} \right) \gamma_k^{(1)} \exp\left( \hat{b}_{n,k} \frac{1 - ((n+1)\theta/k+1)^{-\hat{\rho}_k}}{-\hat{\rho}_k} \right) \tag{4}
\]

as an estimator for extreme quantiles, with \( \gamma_k^{(1)}, \hat{b}_{n,k} \) and \( \hat{\rho}_k \) denoting the maximum likelihood estimators for respectively \( \gamma \), \( b_{n,k} \) and \( \rho \) under Eq. (3) using \( k \) log-spacings of order statistics, \( k \in \{3, \ldots, n-1\} \).

For the more general case where \( \gamma \) can be positive or negative, Beirlant and Matthys (2003) derived the following approximate model for log-ratios of spacings

\[
j \log \frac{Y_{n-j+1,n} - Y_{n-k,n}}{Y_{n-j,n} - Y_{n-k,n}} \overset{D}{\approx} \frac{\gamma}{1 - \left( j/(k+1) \right)^{\rho}} F_j \quad j = 1, \ldots, k-1, \tag{5}
\]
with \( \gamma \in \mathbb{R}, F_j, j = 1, \ldots, k, \) independent standard exponential random variables and \( \gamma \) is estimated using the maximum likelihood method. Model (5) can now be used as a basis for estimating extreme quantiles, see Beirlant and Matthys (2003):

\[
\hat{\mathcal{Q}}^{(2)}_{k+1}(\theta) = Y_{n-k,n} + \hat{a}_{n,k+1} \frac{((k + 1)/(n + 1)(1 - \theta))^{(2)}}{\hat{y}^{(2)}_{k+1}} - 1
\]  

(6)

where

\[
\hat{a}_{n,k+1} = \frac{1}{k} \sum_{j=1}^{k} j(Y_{n-j+1,n} - Y_{n-j,n}) \left( \frac{j}{k + 1} \right)^{y^{(2)}_{k+1}}
\]

(7)

and \( \hat{y}^{(2)}_{k+1} \) denotes the maximum likelihood estimator for \( \gamma \) under Eq. (5) based on the \( k + 1 \) upper order statistics.

**Remarks**

- By imposing a second order condition on the tail of \( F_Y \) a model more complicated than Eq. (5) can be derived for log-ratios of spacings (Beirlant and Matthys, 2003). Even though at a particular fixed \( k \)-value the more complicated model leads to a smaller bias for the estimators, it gives a substantially larger variance and thus a larger mean squared error (MSE). Of course, compared to the reduced model (5) the MSE of the maximum likelihood estimator of \( \gamma \) based on the more complicated model reaches its minimal value far deeper in the sample. Nevertheless, the simulation study described in Beirlant and Matthys (2003) indicates that the simpler model generally outperforms the more complicated model in terms of minimal MSE. Therefore, in our work we only consider the reduced model (5).

- All the above estimators involve the choice of \( k \), the number of upper order statistics used in the estimation. However, in general estimates based on Eqs. (3–6) are quite stable when plotted as a function of \( k \), alleviating the problem of selecting an optimal \( k \)-value to some extent. We will return to this issue when discussing the estimation of extreme conditional quantiles.

- In the special case where \( \gamma < 0 \) the endpoint \( \lim_{\theta \to 1} \mathcal{Q}(\theta) \) is finite and can be estimated by setting \( \theta \) equal to 1 in Eq. (6).

We now return to the case of estimating the conditional quantiles \( \mathcal{Q}(\theta; x^*) \) for \( \theta \) close to 1. The following two-step procedure is proposed for this estimation problem:

1. Compute \( \hat{\mathcal{Q}}(\theta; x^*) \) for \( \theta = 1/(n^* + 1), \ldots, n^*/(n^* + 1) \) with \( n^* = \sum_{i=1}^{n} I_{[x^*-h,x^*+h]}(x_i) \). To ensure monotonicity, the constraints \( \hat{\mathcal{Q}}(i/(n^* + 1); x^*) \geq \hat{\mathcal{Q}}((i - 1)/(n^* + 1); x^*) \), \( i = 2, \ldots, n^* \), were imposed.

2. Estimate \( \gamma \) and extreme conditional quantiles based on Eqs. (3) and (4), or Eqs. (5) and (6), with \( \hat{\mathcal{Q}}(i/(n^* + 1); x^*), i = 1, \ldots, n^* \), taking over the role of univariate order statistics. These estimates will be denoted by \( \hat{\gamma}_{k}^{(1, R)}(\theta) \) and \( \hat{\mathcal{Q}}_{k}^{(1, R)}(\theta) \), respectively \( \gamma_{k+1}^{(2, R)}(\theta) \) and \( \hat{\mathcal{Q}}_{k+1}^{(2, R)}(\theta) \).

**Remarks**

- Note that in Eq. (1) the estimator \( \hat{\mathcal{Q}}(\theta; x^*) \) remains at \( \hat{\mathcal{Q}}(n^*/(n^* + 1); x^*) \) for \( n^*/(n^* + 1) < \theta < 1 \). This does not yield sensible estimators beyond the response data range. This is
illustrated in Figure 1 using a small sample simulation from the Burr($\beta$, $\tau$, $\lambda$) distribution with distribution function given by

$$F_\gamma(y) = 1 - \left( \frac{\beta}{\beta + y^\gamma} \right)^{\frac{1}{\gamma}} y > 0; \quad \beta, \lambda, \tau > 0,$$

for which $\gamma = 1/(\lambda \tau)$ and $\rho = -1/\lambda$. The dependence on the covariate $x$ was obtained by setting $\tau(x) = \exp(1 - x)$; further we took $\beta = \lambda = 1$ so $\gamma(x) = \exp(-1 + x)$. The values for the covariate $x$ were drawn from the $U(-2, 2)$ distribution. In Figure 1 we show $Q(0.9999; x)$ (broken line) and the quartiles of $\hat{Q}(0.9999; x)$ (solid lines) computed over 500 simulated datasets of size $n = 500$. Clearly, a one-step procedure based on local quantile regression underestimates these extreme conditional quantiles. Step 2 in the above proposed procedure is a way of overcoming this shortcoming.

- Monotonicity of $\hat{Q}(\theta; x^*)$ can also be obtained using local versions of restricted regression quantiles; see He (1997) and Zhao (2000). We experimented with this, using as common slope estimator both the 20% and 50% regression trimmed means, but these were not found to perform well.

3 SIMULATION RESULTS

The methodology described in Section 2 was studied extensively through simulation. The distributions from which we simulated are:

- the Burr(1, $\exp(1 - x)$, 1) distribution introduced above,
- the Generalized Pareto Distribution, GPD($\sigma, \gamma$) for which the distribution function is given by

$$F_\gamma(y) = 1 - \left( 1 + \frac{\gamma y}{\sigma} \right)^{-1/\gamma} 1 + \frac{\gamma y}{\sigma} > 0; \quad \gamma \in \mathbb{R}, \sigma > 0,$$

with $\gamma$ denoting the extreme value index. In this study we take $\sigma = 1$ and $\gamma(x) = -\exp(x)$, so that the conditional endpoint is given by $\exp(-x)$. 

FIGURE 1 Burr(1, $\exp(1 - x)$, 1) simulation: $Q(0.9999; x)$ (---) and the quartiles of $\hat{Q}(0.9999; x)$ (——) computed over 500 simulated datasets of size $n = 500$. 

EXTREME CONDITIONAL QUANTILES 571
For both distributions 500 datasets of size $n = 500$ were generated. The values for the covariate $x$ were drawn from the $U(-2, 2)$ distribution. We considered the estimation problem at $x^* = -1$, $x^* = 0$ and $x^* = 1$. In case $\gamma \geq 0$ the main interest is in estimating conditional quantiles such as $Q(1 - (1/10n^*)^{}; x^*)^{}$, while for $\gamma < 0$ we attempt to estimate the conditional endpoint $\lim_{\theta \to 1} Q(\theta; x^*)$. Results are reported for local quadratic approximations to $Q(\theta; x)$, i.e. $p = 2$, using bandwidths $h = 0.25$, $h = 0.5$ and $h = 0.75$. We restrict the discussion to the results obtained with Step 2 of the proposed procedure i.e. the conditional extreme value index estimates and conditional quantile estimates.

Figures 2 and 3 show the results for the Burr$(1, \exp(1 - x), 1)$ simulation. In Figure 2 we show the quartiles (computed over 500 simulated datasets) of the maximum likelihood estimates for $\gamma$ based on Eq. (3), $\hat{\gamma}_{1,k}^{(1,RQ)}$, as a function of $k$ obtained at $x^* = 0$ using (a) $h = 0.25$, (b) $h = 0.5$ and (c) $h = 0.75$. Note that the interquartile range decreases as $h$ increases, since when the bandwidth increases the estimation is based on a larger number of observations. To get a better idea of how the variability of the $\gamma$ estimates is affected by the choice of $k$ and $h$ we also computed the empirical MSE. Figure 3(a) shows the empirical MSE of the estimates $\hat{\gamma}_{1,k}^{(1,RQ)}$ as a function of $k$ and this for the three bandwidths considered. Given the stability of the $\gamma$ estimates over the whole $k$-range it is not unexpected that the empirical MSEs reach their minimal value around $k = n^* - 1$. Note that the smallest MSE is attained for a $k$ value of approximately 150 and a bandwidth $h = 0.75$. Figure 2 shows also the quartiles of the $1 - (1/10n^*)$ conditional quantile estimates, $\hat{Q}_{1,k}^{(1,RQ)}$, as a function of $k$, obtained at $x^* = 0$, using (d) $h = 0.25$, (e) $h = 0.5$ and (f) $h = 0.75$. Again estimates are
very stable over the whole $k$ range. Figure 3(b) contains the corresponding empirical MSE values. In case of extreme quantile (or endpoint) estimators we use a slightly adapted version of the MSE to make the bias-variance trade-off. Instead of the usual squared differences we use squared log-ratios, i.e. $E(( \log \left( \frac{Q_k(y)}{Q(y)} \right))^2)$ where $Q_k(y)$ denotes some extreme quantile estimator. Again for all $h$ values considered the MSE decreases as a function of $k$.

The corresponding results for the GPD$(1, \exp(x))$ simulation are shown in Figures 4, 5 and 6. Figure 4 shows the quartiles of the maximum likelihood estimates for $\gamma$ obtained from Eq. (5), $\gamma^{(LRO)}$, as a function of $k$, at $x^* = 0$ using (a) $h = 0.25$, (b) $h = 0.5$ and (c) $h = 0.75$ and at $x^* = 1$ using (d) $h = 0.25$, (e) $h = 0.5$ and (f) $h = 0.75$. Note that the variability of the estimates (as given by the interquartile range) decreases as $h$ increases. On the other hand the bias increases with increasing $h$. In Figure 6(a) and (b) we show the empirical MSE values of the $\gamma$ estimates at respectively $x^* = 0$ and $x^* = 1$ as a function of $k$ for the three bandwidths considered. At $x^* = 0$ minimal MSE is attained using $h = 0.75$ and a $k$ value of approximately 175, while at $x^* = 1$ using $p = 2$ yields only sensible estimates with $h = 0.25$. Figure 5 shows the corresponding endpoint estimates, as a function of $k$, at $x^* = 0$ using (a) $h = 0.25$, (b) $h = 0.5$ and (c) $h = 0.75$ and at $x^* = 1$ using (d) $h = 0.25$, (e) $h = 0.5$ and (f) $h = 0.75$. At $x^* = 0$ the variability of the endpoint estimates decreases with increasing $h$ and at all bandwidths considered estimates are quite stable as a function of $k$. Again at $x^* = 1$ using $p = 2$ yields only sensible estimates with $h = 0.25$, see also the empirical MSEs of the endpoint estimates at $x^* = 0$ and $x^* = 1$ given in respectively Figure 6(c) and (d).

In case of the GPD simulation, when using $p = 2$ only for $h = 0.25$ sensible estimates are obtained for $\gamma(1)$ and $\lim_{y \to 1} Q(\theta; y)$. The poor performance at bandwidths $h = 0.5$ and $h = 0.75$ is caused by the bias in the upper regression quantiles, see Figure 7. Applying a local cubic approximation gives good estimates for all 3 bandwidths considered. From this it can be concluded that the framework of local polynomial quantile regression offers extra flexibility, which is needed in the given regression context.

Overall we conclude that the choice of the bandwidth is quite important. A bootstrap algorithm to choose this parameter adaptively is given in the next section.

![Figure 3](image-url)
FIGURE 4  GPD(1, exp(x)) simulation: quartiles of $\hat{\pi}_{k+1}^{(2,0)}$ as a function of $k$, $k = 15, \ldots, n^* - 1$, at $x^* = 0$ using (a) $h = 0.25$, (b) $h = 0.5$, (c) $h = 0.75$ and at $x^* = 1$ using (d) $h = 0.25$, (e) $h = 0.5$, (f) $h = 0.75$.

FIGURE 5  GPD(1, exp(x)) simulation: quartiles of $\hat{\pi}_{k+1}^{(2,0)}(1)$ as a function of $k$, $k = 15, \ldots, n^* - 1$, at $x^* = 0$ using (a) $h = 0.25$, (b) $h = 0.5$, (c) $h = 0.75$ and at $x^* = 1$ using (d) $h = 0.25$, (e) $h = 0.5$ and (f) $h = 0.75$. 
FIGURE 6  GPD(1, $-\exp(x)$) simulation: empirical MSE of $\hat{f}_{k+1}^{(\ell_2, RQ)}$ as a function of $k$, $k = 15, \ldots, n^* - 1$, at
(a) $x^* = 0$ and (b) $x^* = 1$ using $h = 0.25$ (——), $h = 0.5$ (———) and $h = 0.75$ (———) and empirical MSE of $\hat{Q}_{k+1}^{(\ell_2, RQ)}(1)$ as a function of $k$, $k = 15, \ldots, n^* - 1$, at (c) $x^* = 0$ and (d) $x^* = 1$ using $h = 0.25$ (——), $h = 0.5$ (———) and $h = 0.75$ (———).

FIGURE 7  GPD(1, $-\exp(x)$) simulation: $Q(\theta; 1)$ (black line) and quartiles of $\hat{Q}(\theta; 1)$ (grey lines) as a function of $\theta$ using (a) $h = 0.25$, (b) $h = 0.5$ and (c) $h = 0.75$. 
4 SELECTION OF THE BANDWIDTH \( h \)

When applying nonparametric techniques, a common issue is the selection of the bandwidth parameter \( h \). On the one hand, \( h \) should be taken sufficiently large in order to have a substantial number of observations in the interval \([x^* - h, x^* + h]\). On the other hand, \( h \) should not be too large since a large \( h \) value will increase the bias due to the fact that the local polynomial approximation to \( \hat{Q}(\theta; x) \) is getting worse. An asymptotic mean squared error (AMSE) criterion could be constructed to guide in the selection of a local optimal \( h \) value. Of course, this requires knowledge of the expressions for the asymptotic variance and bias of \( \hat{Q}(\theta; x^*) \). Further, these expressions will depend on the unknown model parameters, hence requiring a plug-in of estimates.

Here we use the bootstrap estimate of MSE(\( \hat{Q}(\theta; x^*) \)) defined as

\[
\text{MSE}^*(\hat{Q}(\theta; x^*)) = E[(\hat{Q}_b^*(\theta; x^*) - \hat{Q}(\theta; x^*))^2 | \ldots \hat{F}],
\]

with \( \hat{Q}_b^*(\theta; x^*) \) the bootstrap estimate of \( Q(\theta; x^*) \) and \( \hat{F} \) the empirical distribution function, to guide the selection of the bandwidth parameter \( h \), see also Hall (1990) and Efron and Tibshirani (1993). Given bootstrap samples \((Y_1, x_1), \ldots, (Y_n, x_n)\), Eq. (8) can be estimated by

\[
\text{MSE}^*(\hat{Q}(\theta; x^*)) = \frac{1}{B} \sum_{b=1}^{B} (\hat{Q}_b^*(\theta; x^*) - \hat{Q}(\theta; x^*))^2,
\]

where \( \hat{Q}_b^*(\theta; x^*) \) denotes the estimate of \( Q(\theta; x^*) \) from bootstrap sample \( b \). The MSE optimal \( h \) value obtained by the bootstrap is then defined as

\[
\hat{h}_{\text{opt}}^* = \arg \min \text{MSE}^*(\hat{Q}(\theta; x^*)�).
\]

In Figure 8 the bootstrap procedure is illustrated on the basis of a small sample simulation from the GPD(1, -exp(\( x \))) distribution introduced above. The values for the covariate \( x \)

![Figure 8](image-url)
were drawn from the $U(-2, 2)$ distribution. We considered the selection of an optimal local $h$ for the estimation of $Q(0.999; 1)$ using $p = 2$ (i.e. for Step 1 of the proposed procedure). The results are based on 100 samples of size $n = 500$. The bootstrap estimates of $\text{MSE}(\hat{Q}(0.999; 1))$ at a particular $h$ value are based on $B = 500$ bootstrap samples. In Figure 8(a) the histogram of the 100 values of $\hat{h}_{\text{opt}}^*$ for estimating $Q(0.999; 1)$ are shown. Figure 8(b) shows the boxplots of the 100 realisations of $\hat{Q}(0.999; 1)$ obtained with the bootstrap procedure and for some fixed $h$-values. From the GPD simulation described in the previous section we know that at $x^* = 1$ Step 1 of the proposed procedure yields good high quantile estimates for small $h$ values. This is confirmed by the boxplots of the $\hat{Q}(0.999; 1)$ estimates obtained with a fixed prespecified $h$-value given in Figure 8(b). From this picture it is clear that the proposed bootstrap procedure performs quite well in selecting a good $h$-value.

In addition to the above described bootstrap procedure, used to guide in the selection of an optimal local $h$ value, we also analyzed a cross validation procedure developed to obtain a data driven global $h$ value, i.e. a $h$ that performs well over the whole covariate range. Similar to the cross-validation procedure proposed by Aerts and Claeskens (1997) in the context of local polynomial maximum likelihood estimation, the optimal $h$ obtained by cross-validation is here defined as

$$\hat{h}_{\text{CV}} = \arg \min \sum_{i=1}^{n} f_0(Y_i - \hat{b}_{0,i})$$

(9)

where $\hat{b}_{0,i}$ is the local estimator for $Q(\theta; x_i)$ based on the sample without the $i$th observation. Based on the results obtained from a small sample simulation study, this procedure did not prove to be useful for practical purposes and hence will not be considered further in this paper.

5 CASE STUDIES

In this section we illustrate the proposed two-step procedure using two case studies.

5.1 Condroz Data

Our first example comes from pedochemistry. The database contains measurements on the variables Calcium (Ca) content and pH level for soil samples taken in different cities in the Condroz (a geographical region in the southern part of Belgium). These data were already analyzed in Goegebeur et al. (2002) with emphasis on the development of an automatic procedure for highlighting suspicious points, i.e. points that are very unlikely to occur even under a heavy tailed or Pareto-type model. Here we concentrate on the related problem of estimating extreme quantiles of the conditional distribution of the variable Ca given pH level. In fact observations can be considered as unreliable if they exceed a particular extreme conditional quantile. In Figure 9(a) we show the scatterplot of Ca content versus pH level for one of the cities. Next to the global upwards trend of the point cloud, extreme Ca measurements tend to occur more often for the larger pH levels, indicating the need for a tail analysis conditional on the covariate pH. Figure 9(b) shows the bootstrap estimate of $\text{MSE}(\hat{Q}(0.9; 6))$ for $p = 1$ as a function of $h$. This estimated bootstrap MSE attains its minimum at $h = 0.44$. Next we applied the exponential regression model (3) to $\hat{Q}(\theta; \text{pH})$ for $\theta = 1/(n^* + 1), \ldots, n^*/(n^* + 1)$ and this on a grid of pH values. In general tail index
estimates obtained from maximum likelihood fits of Eq. (3) and extreme quantile estimates based on Eq. (4) are very stable when considered as a function of $k$, see for instance Figure 9(c) and (e) where we show $\hat{\gamma}_k^{(1.RQ)}$ at pH = 6.5 as a function of $k$. We therefore propose to summarize the estimates obtained at the different $k$-values using for instance the median. The median of $\hat{\gamma}_k^{(1.RQ)}$, $k = 5, \ldots, n^* - 1$, is given as a function of pH in Figure 9(d). Finally, in Figure 9(f) we show the median of $\hat{Q}_k(0.9995)$, $k = 5, \ldots, n^* - 1$, as a function of pH on the Ca versus pH scatterplot.

5.2 Electric Utility Data

Endpoint or boundary estimation is often encountered when studying the relation between input (e.g. labor, capital) and output (e.g. goods produced) of firms in a productivity analysis. Clearly, given a certain amount of input, the possible output is bounded above. This upper bound is the so-called efficiency curve. We use a data set on 123 American electric utility
companies to illustrate this aspect of the methodology given above. This data set was also used in Gijbels and Peng (2000). Figure 10(a) shows the output versus log(cost) scatterplot. Because of the sparseness of the observations with log(cost) < 1 we restricted the analysis to the observations with 1 ≤ log(cost) ≤ 6. In Figure 10(b) we show $\hat{\text{MSE}}^* (\hat{Q}(0.9; 3))$ as a function of $h$. Application of the bootstrap procedure at different log(Cost)-values indicated the need for a covariate dependent bandwidth parameter. Because of the small number of observations in each window we report the $\gamma$ and endpoint estimates obtained at $k = n^* - 1$. The $\gamma$ estimates at $k = n^* - 1$ shown in Figure 10(c) were obtained by using $h = 0.5$ if log(Cost) < 3.5 and $h = 1$ otherwise. This choice was clearly indicated by the bootstrap algorithm. To avoid difficulties in the estimation of the conditional endpoints, the $\gamma$ estimates were bounded away from 0. However, as can be seen in the plot, this constraint is active only at a small number of log(Cost)-values. Finally, in Figure 10(d) we superimposed the estimated conditional endpoint (solid line) as well as the local maximum estimator proposed by Gijbels and Peng (2000) with $h = 0.5$ (broken line) on the output versus log(Cost) scatterplot (solid line).

![Figure 10](image)

FIGURE 10 Electric utility data: (a) output versus log(Cost) scatterplot, (b) $\hat{\text{MSE}}^* (\hat{Q}(0.9; 3))$ as a function of $h$, (c) $\gamma_{k+1}$ as a function of log(Cost) and (d) $\hat{\text{Q}}^*(k+1; 1)$ as a function of log(Cost).
6 CONCLUSION

We have proposed a flexible nonparametric method for estimating extreme quantiles in a regression setting. In this two-stage procedure we combine the merits of local polynomial quantile regression (first step) and recent extreme value methods (second step) in order to obtain a general and effective technique in comparison with earlier attempts from the literature. In future work we intend to complete this study with inferential tools based on the given quantile estimates. To this end asymptotic results based on the work of Chernozhukov (1998, 2001) could be pursued.

Acknowledgements

Research of Tertius de Wet was partially supported by National Research Foundation grant 2046922 and a research grant from the University of Stellenbosch. Yuri Goegebeur was supported by the Fund for Scientific Research – Flanders (F.W.O. – Vlaanderen).

References