Noname manuscript No.
(will be inserted by the editor)

Fitting Censored Quantile Regression by Variable Neighborhood Search

Rima Sheikh Rajab · Milan Dražić · Nenad Mladenović · Keming Yu

Abstract Censored quantile regression models are very useful for the analysis of censored data when standard linear models are felt to be appropriate. However, fitting censored quantile regression is hard numerically due to the fact that the objective function that has to be minimized is not convex nor concave in regressors. The performance of standard methods is not satisfactory, in particular if a high degree of censoring is present. Usual approach in the literature is to simplify (linearize) estimator function and show theoretically that such approximation tends to good real optimal values. In this paper we suggest different approach, i.e., we solve directly nonlinear non-convex non differentiable optimization problem. Our method is based on variable neighborhood search approach, a recent successful technique for solving global optimization problems. Simulation results presented indicate that our new method can improve the quality of censored quantizing regressors estimator considerably. Moreover, an extramarital affairs example involving censored regression analysis is also used to illustrate the method.

Keywords Censored regression · Powell estimator · Quantile regression · Global optimization · Metaheuristics · Variable neighborhood search

1 Introduction

The censored quantile regression (CQR) model proposed by Powell (1986) is to take a linear model for the response but assume that the observations cannot be

R. Sheikh Rajab
Brunel University, Uxbridge, UB8 3PH, UK

M. Dražić
Faculty of Mathematics, University of Belgrade, Studentski trg 16, 11000 Belgrade, Serbia
Tel.: +381-11-2027-801, Fax: +381-11-2630-151
E-mail: mdrazic@matf.bg.ac.rs

N. Mladenović
Brunel University, Uxbridge UB8 3PH, UK

K. Yu
Brunel University, Uxbridge UB8 3PH, UK

observed above the fixed level y_0. This leads to considering the following censored quantile regression model,

$$y_i = \max\{y_0, \ x_i'\beta_\theta\} + \varepsilon_i \qquad , i = 1, \ldots, n, \qquad (1)$$

where $x_i = (1, x_{1,i}, ..., x_{g-1,i})'$ is the ith observation of g-dimensional covariate vector x, $\beta_\theta = (\beta_{0,\theta}, \beta_{1,\theta}, ..., \beta_{g-1,\theta})$ is the unknown parameter vector which corresponds to the θth $(0 < \theta < 1)$ quantile. ϵ_i is the model error.

There are many applications of CQR, including an alternative direction to link CQR to survival analysis via generalization of Kaplan-Meier and Nelson-Aalen estimators, see the work by Portnoy [1] and Peng and Huang [2] among others. Econometrics and statistics have been interested in CQR models in the recent years, especially due to unknown conditional heteroscedasticity and their robustness to distributional miss-specification of error term. For various applications of CQR, see also [3, 4, 5, 6, 7, 8, 9, 10].

The Powell [11, 12] suggested an intuitive estimator for censored quantile regression model. This estimator solves

$$f(\beta, \theta, y_0) = \min_\beta \{\frac{1}{n} \sum_{i=1}^{n} \rho_\theta(y_i - \max\{y_0, \ x_i'\beta_\theta\}), \qquad (2)$$

where $\rho_\theta(\lambda) = [\theta - I(\lambda < 0)]\lambda$ is the check function and $I(.)$ is the usual indicator function. Since θ and y_0 are given, function (2) depends only on β. Therefore, we denote Powell estimator (2) as $f(\beta)$.

The Powell estimator has several disadvantages. First, the censoring point y_0 must be known. Second, obtaining the global minimum of (2) can be difficult because the objective function $f(\beta)$ is non convex, nor concave and even non differentiable in β. An example of $f(\beta)$ with $q = 2$, $n = 100$, $\theta = 0.95$, $y_0 = 0$ and normally distributed random variable ε is illustrated in Figure 1 (more details of this instance will be given in section & 3.1). Thus, the problem belongs to global (nonlinear) optimization area, and may have many local optima. Consequently, standard optimization tools, that require the objective function to be differentiable and/or convex, may fail to discover the true CQR estimator.

However, several convex optimization algorithms have been adapted for finding CQR, where the Powell estimator (2) has been used (see [13] for a survey of such algorithms). For example, in [14] the problem is linearized by using reduced-gradient algorithm. In that way a local minimizer is found by using linear programming. An interior point algorithm for quantile regression problem is suggested in [15]. A threshold accepting algorithm is given in [16]. The estimators suggest in [4, 7] are asymptotically equivalent to original Powell estimator, they do not allow explicitly for censored observations to be interpolated by the estimated CQR, where the interpolation property suggests finding an exact solution by using a computationally expensive algorithm. An iterative linear programming algorithm is introduced in [17].

Those algorithms have difficulties in solving CQR problems. They exhibit a high degree of complexity in their implementation. Most of them achieve convergence to local optima, whereas finding the global optima for these problems require a heavy computational loads.
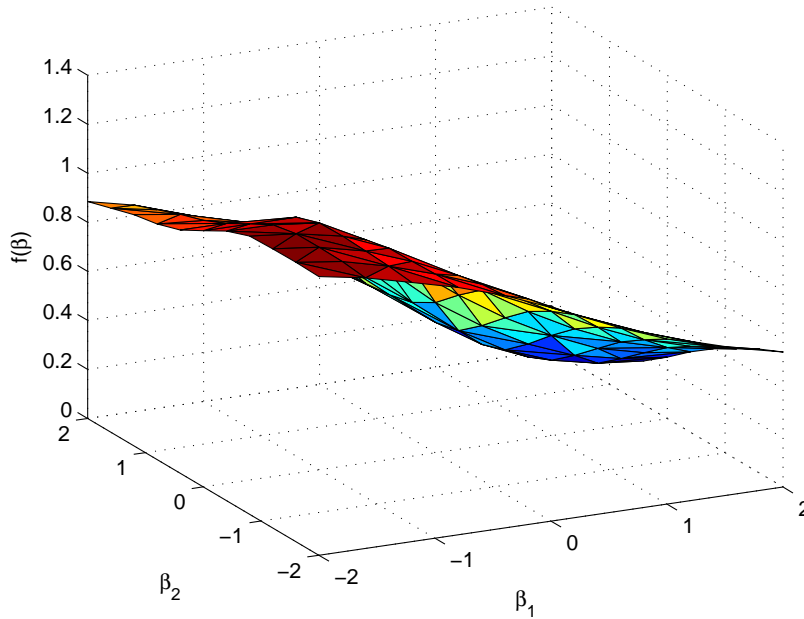
Fig. 1 Powell function $f(\beta_1, \beta_2)$ with $n = 100$, $\theta = 0.95$, $y_0 = 0$ and Gaussian r.v. $\varepsilon \neq 0$
.

Therefore, the basic question we would like to answer in this paper is, what approach is more promising: use original Powell estimator (2) and solve CQR problem approximately, or use simplified approximative model and solving it exactly? Opposite to the most of authors who tried with simplified models, here we suggest, for the first time, use of approximate global optimization method for solving (2). In order to do that, we developed nonlinear programminng (NP) code based on variable neighborhood search (VNS) metaheuristic (or framework for building heuristics). As far as we know, it is the first time that some metaheuristics approach is used for solving CQR. Based on computational results section, it appears that our approach outperforms other methods from the literature.

The paper is organized as follows. In section 2, we briefly give VNS rules, and then explain how it works for solving CQR. In section 3 we first explain how the data instances are generated and then report extensive computational analysis usual in this field. In section 4, we test CQR model and solution methods on the real *Redbook magazine* data of extramarital affairs. The final section contains concluding remarks.

2 Variable neighborhood search for censored quantile regression

As mentioned earlier, CQR problem belongs to continuous global optimization. In this section, we first give rules of VNS for solving global optimization problem (GOP) and then we explain how we use them for solving CQR.

2.1 Variable neighborhood search metahuristics

Variable neighborhood search (VNS) [18] is a metaheuristic based upon systematic changes of neighborhoods in order to enable finding better solution in distant parts of a solution space. VNS is designed for solving both continuous and discrete optimization problems, that may be formulated as

$$\min\{f(\beta)| \ \beta \in \mathcal{B}, \mathcal{B} \subseteq \mathcal{S}\}. \tag{3}$$

$\mathcal{S}, \mathcal{B}, \beta$ and f respectively denote the *solution space*, *feasible set*, a *feasible solution* and a real-valued *objective function*. If $\mathcal{S}$ is a finite but large set, a *combinatorial optimization* problem is defined. If $\mathcal{S} = \mathbb{R}^n$, we refer to *continuous optimization*. An *exact algorithm* for problem (3), if one exists, finds an optimal solution β^*, together with the proof of its optimality, or shows that there is no feasible solution, i.e., $\mathcal{B} = \emptyset$.

Let $\mathcal{N}_k$, $k = 1, \ldots, k_{max}$, denotes a finite set of pre-selected neighborhood structures and let $\mathcal{N}_k(\beta)$ be the set of solutions in k^{th} neighborhood of β. The neighborhood structures $\mathcal{N}_k$ may be induced from one or more metrics introduced into a solution space $\mathcal{S}$, either discrete or continuous. We define $\beta'' \in X$ as a local minimum w.r.t. $\mathcal{N}_k$, if there is no solution $\beta \in \mathcal{N}_k(\beta'') \subseteq \mathcal{B}$ such that $f(\beta) \leq f(\beta'')$. There are three obvious facts that could explain why change of neighborhoods idea works well:

Fact 1. A global minimum is a local minimum with respect to all possible neighborhood structures.

Fact 2. A local minimum with respect to one neighborhood structure is not necessarily a local minimum with respect to another neighborhood structure.

Fact 3. For many problems local minima with respect to one or several neighborhoods are relatively close to each other [19].

Those simple facts are used within VNS in several different ways (see for example recent surveys of VNS in [19,20]). The deterministic change of neighborhoods leads us to a so-called Variable neighborhood descent (VND) heuristic. The basic VNS (BVNS) combines deterministic and random search [18]. Its pseudo-code is given in Algorithm 1.

Function BVNS $(\beta, k_{max}, t_{max})$;
1 **repeat**
2 $k \leftarrow 1$;
3   **repeat**
4 $\beta' \leftarrow$ Shake(β, k);
5 $\beta'' \leftarrow$ LocalSearch(β') ;
6   If $(f(\beta'') < f(\beta))$ $\beta \leftarrow \beta''$; goto 2;
8 $k \leftarrow k + 1$
until $k = k_{max}$;
9 $t \leftarrow$ CpuTime()
until $t > t_{max}$;

Algorithm 1: Steps of the Basic VNS

Let β be the incumbent (the best solution found so far). Within BVNS a point β' from the neighborhood k of β ($\beta' \in \mathcal{N}_k(\beta)$) is taken at random where

$k = 1, \ldots, k_{\max}$. Such a point is initial one for a local search routine that provides local minimum β''. If $f(\beta'')$ is better (smaller in the case of minimization) then the new incumbent is β'' ($\beta \leftarrow \beta''$) also k is set to 1 ($k \leftarrow 1$) and the process is repeated. Otherwise we generate random point from the larger neighborhood ($k \leftarrow k + 1$). The only parameter for the BVNS is the number of neighborhoods used ($k_{\max}$). Once that neighborhood is reached without finding improvement, k is again set to 1. All process is repeated until some stopping criterion, as maximum CPU time $t_{\max}$ used, is satisfied (see Figure 2).
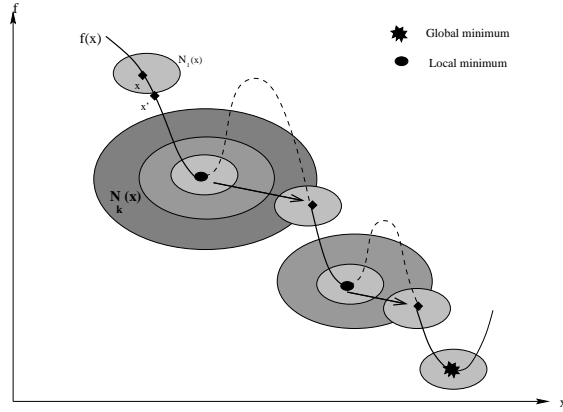

Fig. 2 Illustration of the Basic Variable Neighborhood Search (BVNS)

We may view the VNS as a "shaking" process, where a movement to a neighborhood further from the current solution corresponds to a harder shake. Unlike random restart, the VNS allows a controlled increase in the level of the shake. In this paper we design the GVNS heuristic for solving CQR problem, by minimizing Powell function $f(\beta)$.

2.2 VNS for CQR

In this subsection we explain how we use VNS to solve the CQR problem. $f(\beta)$ defined in 2, is a piece-wise linear objective function with continuous variables $\beta_0, \ldots, \beta_{g-1}$. We approach the CQR problem via an unconstrained nonlinear program. Observe also that any unconstrained nonlinear program may be considered as box constrained, if left and right values of variables that define box are set to the same large negative and positive values a_i and b_i. Therefore, given input data $X = (x_{ij}), i = 1, \ldots, n; \; j = 1, \ldots, g - 1, Y = (y_1, \ldots, y_n), \; \varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ and the value of $\beta = (\beta_0, \ldots, \beta_{g-1})$, the pseudo-code for evaluating Powell estimator is given in Algorithm 2.

Neighborhoods - Shaking. For solving GOP, VNS has already been used in two different ways: with neighborhoods induced by using an ℓ_p norm [21,22,23] and without using an ℓ_p norm [24]. Here we apply VNS that uses ℓ_p norm, i.e., we

```
Function Powell(β, X, Y, y₀, ε)
1  Powell← 0
2  for i = 1, ..., n do
3  |    s ← 0
4  |    for j = 1, ..., g − 1 do
5  |    |    s ← s + x_{ij}β_j
6  |    r ← y_i − max{y₀, s} − ε_i
7  |    Powel ← Powell + θr
8  |    if r < 0 then
9  |    |    Powell ← Powell - r
```

Algorithm 2: Pseudo-code for finding Powell estimator value

define distances between any two solutions β and γ as

$$\delta(\beta, \gamma) = \left(\sum_{i=0}^{g-1} |\beta_i - \gamma_i|^p\right)^{\frac{1}{p}}, \tag{4}$$

or

$$\delta(\beta, \gamma) = \max_{0 \le i \le g-1} |\beta_i - \gamma_i|, \quad p = \infty. \tag{5}$$

The neighborhood $\mathcal{N}_k(\beta)$ denotes the set of solutions in the k–th neighborhood of β, and using the metric δ, it is defined as

$$\mathcal{N}_k(\beta) = \{\gamma \in \mathcal{B} \mid r_{k-1} < \delta(\beta, \gamma) \le r_k\}, \tag{6}$$

where r_k is a given radius of neighborhood $\mathcal{N}_k$ $(k = 1, \ldots, k_{\max})$.

Our `CQR-VNS` procedure for solving CRQ problem contains the following parameters in addition to k_{max} (a maximum number of neighborhoods used in the search) and t_{max} (the maximum time allowed in the search):
(i) _Values of radii_ r_k, $k = 1, \ldots, k_{\max}$. Those values may be defined by the user or calculated automatically during the minimization process. The geometry of the neighborhood structure is induced by the ℓ_1 (4) and ℓ_∞ (5). We use balls as in (6). Radii $r_1 \le r_2 \le \cdots \le r_{k_{\max}}$ are automatically computed as follow: let $\beta = (\beta_0, \ldots, \beta_{g-1})^T \in R^g$ be the current incumbent solution and let

$$a_j \le \beta_j \le b_j, \quad j = 0, \ldots, g-1$$

defines box or hyper-cube

$$H = \prod_{j=0}^{g-1} [a_j, b_j]$$

around the incumbent solution β. In order to find k_{max} neighborhoods automatically and thus make our `CQR VNS` more user-friendly, we divide $\beta_j - a_j$ and $b_j - \beta_j$ into a $k_{\max}$ intervals:

$$\underline{\delta}_j = \frac{\beta_j - a_j}{k_{\max}}; \quad \overline{\delta}_j = \frac{b_j - \beta_j}{k_{\max}}.$$

Then the $k_{\max}$ hyper-cubes (boxes) $H_1, H_2, \ldots, H_{\max}$ around the incumbent (the best solution found so far) β are given

$$a_j + (k-1)\underline{\delta}_j \le \beta_j \le b_j - (k-1)\overline{\delta}_j, k = 1, \ldots, k_{\max}$$

or

$$\underline{d}_{jk} \leq \beta_j \leq \overline{d}_{jk}, k = 1, \ldots, k_{\max}$$

Figure (3) illustrates our construction of continuous neighborhoods as hypercubes for the case of $k_{\max} = 3$ and $g = 2$ (or $\beta = (\beta_0, \beta_1)^T$).
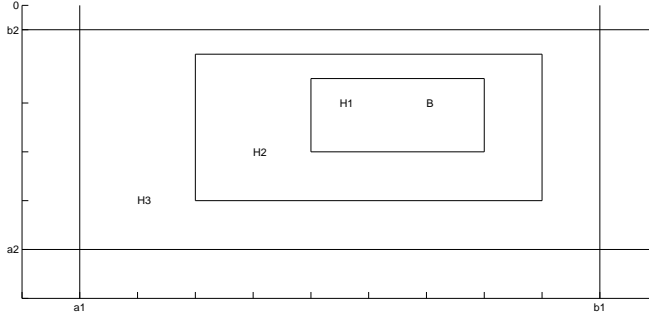

Fig. 3 Automatic construction of neighborhoods with $g = 2$ and $k_{max} = 3$.

(ii) *(Geometry, distribution) pairs*. Geometry of neighborhood structures $\mathcal{N}_k$ is defined by the choice of the metric functions used in the search through the solution space. The usual choices are the ℓ_1, ℓ_2, and ℓ_∞ norms. Their order in the search is also important within VNS. Different *distributions* may be used for obtaining the random point y from the same neighborhood $\mathcal{N}_k$ in the *Shaking* step. Uniform distribution in $\mathcal{N}_k$ is the obvious choice, but other distributions may lead to much better performance on some problems. Beside uniform (u), we also implement hypergeometric distribution (h) [21, 22].

Note that different choices of geometric neighborhood shapes and random point distributions lead to different VNS-based heuristics. We denote them as (α, γ), where α and γ represents geometry (metric) and distribution used, respectively. Therefore, in total we have 6 different variants of VNS defined by (geometry, distribution) pairs: (ℓ_1, u),(ℓ_2, u),(ℓ_∞, u),(ℓ_1, h), (ℓ_2, h), (ℓ_∞, h). Note that "u" denotes uniform distribution, while "h" denotes hypergeometric (special) distribution. For simplicity, we will denote those variants in pseudo-code as (1,1),(2,1),(3,1),(1,2),(2,2) and (3,2): (1,1)= (ℓ_1, u), (2,1)=(ℓ_2, u), etc. For example, pair (3,2) indicates that ℓ_∞ norm (3) and the special distribution (2) are used in the shaking step.

However, after extensive computational analysis, we remain on four (geometry, distribution) pairs in our `CQR-VNS` in the following order: `distribution_type_order` = (1,2) (1,1) (3,1) and (3,2). After that a radius from interval $[0, r_k]$ is taken at random in order to get a point from $\mathcal{N}_k(x)$. Therefore, a random point within Shaking step of `CQR-VNS` is generated in two steps: (i) find random direction; (ii) find random radius along that direction.

Local Search. As a local search for solving CQR we apply the direct search Nelder-Mead nonlinear programming method since it does not use derivatives. The left and right boundaries a_j and b_j for variables are defined as appropriate.

Pseudo-code. The algorithm Glob-VNS for solving CQR is given in Algorithm 3, where $k_{\max}$ and $t_{\max}$ are usual VNS parameters, given by user.

Function CQR-VNS $(\beta^*, k_{max}, t_{max}, X, Y, y_0, \varepsilon)$
1 Select $(geometry, distribution)$ pairs as: (1,2), (1,1), (3,1), (3,2)
2 Choose an initial point $\beta^* \in \mathcal{B}$ at random
3 $f^* \leftarrow$ Powell$(\beta^*, X, Y, y_0, \varepsilon)$
4 $\beta \leftarrow \beta^*$, $t \leftarrow 0$
5 while $t < t_{max}$ **do**
6 $k \leftarrow 1$
7   **repeat**
8   **for** *each (geometry, distribution) pair* **do**
9 $\beta' \leftarrow$ Shake(β, k) // Get $\beta' \in \mathcal{N}_k(\beta)$ at random
10 $\beta'' \leftarrow$ Nelder-Mead(β', f) // Get local minimum β'' by Nelder-Mead
11   **if** $(f < f^*)$ **then**
12 $\beta \leftarrow \beta''$; $f^* \leftarrow f$; go to line 6
13 $k \leftarrow k + 1$
         until $k = k_{max}$
14 $t \leftarrow$ CpuTime$()$

Algorithm 3: VNS for CQR

After choosing $(geometry, distribution)$ pairs and random initial solution $\beta^* \in R^{g-1}$ in steps 1 and 2 respectively, we apply Algorithm 2 to find Powell estimator $f(\beta^*)$. We denote with β the incumbent solution. As explained in Algorithm 1, outer loop of VNS is running until predefined stopping condition is met. The inner loop is repeated k_{max} times, if there is no improvement in regressors β. In each neighborhood a random point from the $\mathcal{N}_k(\beta)$ is taken (line 9) and well known Nelder-Mead unconstrained nonlinear programming code run (line 10). The local minimum value for Powel's estimator is denote with f. If the better solution is obtained, we save it (line 12) and repeat all process with the first neighborhood (i.e., return to step 6).

3 Computational results

We perform extensive computational analysis to investigate how our new CQR-VNS method compares with other approaches. We first give general rules for the computational simulation performed, and then present comparative results on various test instances.

Methods compared. We compare our VNS-CQR with the following approaches from the literature.

1. The first group of methods are the same as in [25]:
   – the direct heteroscedastic method;
   – modified bootstrap and
   – re-sampling methods.
2. The second group of methods are from [4]:
   – CV method which denotes the CQR-LP estimator with log likelihood cross-validated bandwidth;

- CVa method, which denotes the CV estimator with bandwidth adjusted to conform with assumption K;
- PR method, which denotes the CQR-LP estimator with probit estimates for censoring probability;
- HO and HOa are the same as CV and CVa except that the kernel function involves a higher order kernel, and
- CR denotes Powell's estimator.
3. Lastly, we apply our method to an extramarital affairs. Data set is taken from [26].

Computer support. Our code was written in C++ and complied with Microsoft Visual Studio 8.0. The program was run into Intel(R) Core(TM) 2 at 1.73 GHz with 2 GB of RAM. Unfortunately, there is no information about the computers which were used to get results by other methods. Therefore, the efficiency (i.e., the running CPU time of methods) could not be compared in this study. So, we will concentrate the comparison only to their effectiveness or precision.

VNS parameters. Along the space dimension, initial and boundary conditions which are different in each test instance, in the `CQR-VNS` we used the following parameters.

- CPU time was limited to $t_{max} = 5$ seconds;
- The number of neighborhoods structures used is set to 10, $k_{max} = 10$;
- We choose the Nelder-Mead local search method. It stops when one among the following three criteria are met:
  - a diameter of a simplex is less than 0.1e-5 (`ls_eps` = 0.1e-5),
  - the difference between two consecutive objective function values is less than 0.1e-5 (`ls_fun_eps` = 0.1e-5) and
  - the number of iterations reached 500.

3.1 Simulation comparison

In this part, we compared our `CQR-VNS` (whose pseudo-code is given in Algorithm 3) with the three algorithms used in [25]. There the following model is considered:

$$y_i = \max\{\beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + \varepsilon_i, \ y_0\}. \tag{7}$$

The details regarding simulation are listed below:

- x_1 is generated as Bernoulli distribution centered at zero, with the success probability equal to $\frac{1}{2}$;
- x_2 is a standard normal variable $N(0,1)$;
- The censoring point is $y_0 = 0$;
- Three different types of error ε are considered:
  - a standard normal distribution;
  - a heteroscedastic normal $(1 + x_2) \times N(0,1)$, and
  - a normal mixture $0.75 \times N(0,1) + 0.25 \times N(0,4)$, as suggested in [25].
- It is assumed that the best estimator values are known and all equal to 1, $(\beta_0, \beta_1, \beta_2) = (1,1,1)$. Then y_i is calculated by using formula (7).

– For each of the following percentile value $\theta \in \{0.95, 0.90, 0.85\}$ the standard (S) and percentile (P) methods [27] are used to construct confidence intervals. In particular, we compare the 95%, 90% and 85% confidence levels for each type of error.
– A size $n = 100$ of random sample is generated, i.e., $\{(x_{1i}, x_{2i}, y_i), i = 1, 2, \ldots, 100\}$. Those data space points are obtained with the three different types of error ε. Two of them are plot in Figure 4.
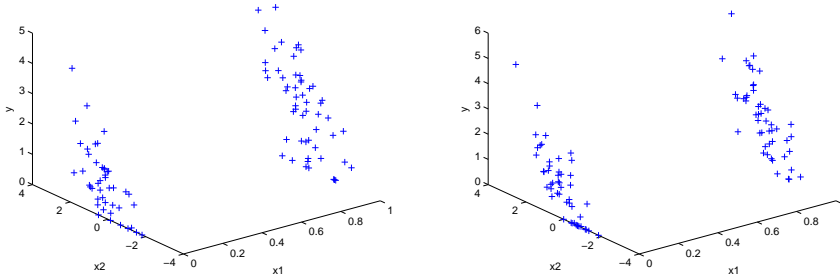– The simulation is repeated 1000 times and the average results reported.

Fig. 4 Points (x_{1i}, x_{2i}, y_i), $i = 1, \ldots, 100$, in data space with the standard normal (left) and normal mixture (right) errors with fixed $\beta_1 = 1$ and $\beta_2 = 1$.

In Figure 1 an instance of this type is plot in the regressor space (β_1, β_2), where the value of β_0 is fixed to 1. Powell's estimator values are obtained by applying Algorithm 2 and taking (β_1, β_2) in each point of the square grid $[0,2] \times [0,2]$ and increment 0.2 for each variable: $\beta_{jk} = 0.2 \cdot k, \forall j = 1, 2; \forall k = 0, \ldots, 10$. The version of the same instance, but with $\varepsilon_i = 0$ in (7), is presented in Figure 5.

Table 1 contains results for estimation of the regression coefficients β_2 only. We compare the empirical coverage probabilities to other three algorithms used in [25]. Therefore, we investigate the finite sample performance of four methods: our VNS for CQR (`CQR-VNS`), the direct heteroscedastics bootstrap method (Bootstrap for short), the resampling method (Resampling for short) and Bilias, Chen and Ying's bootstrap method (M-Bootstrap for short) [25]. The quality of solutions obtained by `CQR-VNS` may be seen in Figure 6 as well

As we can see in Table 1, for standard normal distribution error and normal mixture error, VNS method gives better results when compared to other methods. For heteroscedastic normal error term, our `CQR-VNS` reports better results than others for finding the percentile "P", but it is not the best one in finding standard "S" case. The distribution best solutions obtained by our `CQR-VNS` in 100 runs are presented at Figure 7. Therefore, we can conclude that VNS based heuristic with the Powell's estimator is a new promising method for solving CQR problem. Our results also show that the choice of approximate solution method applied on exact model could be better choice that use of exact methods on approximate model.
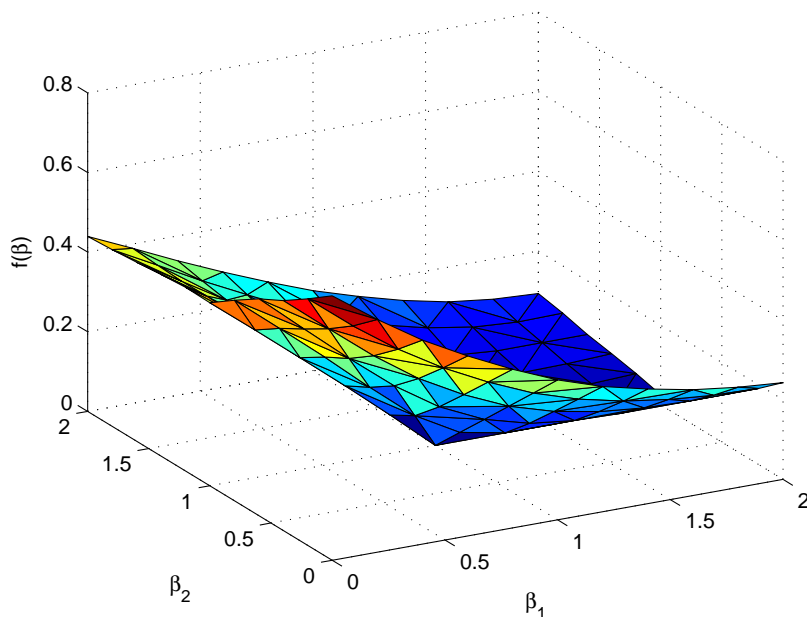
Fig. 5 Censored Quantile Regression function $f(\beta)$ and $\varepsilon = 0$

.
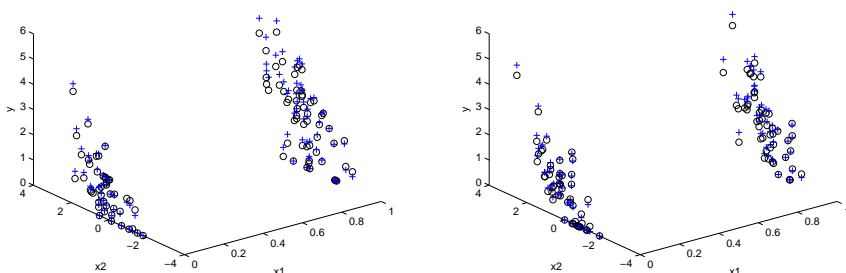
Fig. 6 Points (x_{1i}, x_{2i}, y_i), $i = 1, \ldots, 100$, in data space with the standard normal and normal mixture errors, and their estimated values (denoted as "o"), obtained by `CQR-VNS` (denoted as "+")

3.2 Comparison with Buchinsky's method in terms of root mean square, mean bias, mean absolute deviation and median bias

Buchinsky in [5] provided an alternative method for censored quantile regression fitting. In this subsection, we aim at comparing his method with our algorithm. Reconsider the two regression models in [5]:

Table 1 EMPIRICAL COVERAGE PROBABILITIES FOR CONFIDENCE INTERVALS

		Bootstrap	*Resample*	*M − Bootstrap*	*VNS*
Confidence level		ECP	ECP	ECP	ECP
Standard Normal					
0.95	*S*	0.956	0.929	0.912	0.948
	P	0.974	0.952	0.943	0.951
0.90	*S*	0.909	0.878	0.863	0.897
	P	0.941	0.900	0.886	0.901
0.85	*S*	0.868	0.830	0.812	0.847
	P	0.906	0.846	0.833	0.851
Normal Mixture					
0.95	*S*	0.957	0.936	0.926	0.950
	P	0.975	0.938	0.935	0.951
0.90	*S*	0.923	0.892	0.875	0.899
	P	0.941	0.878	0.872	0.901
0.85	*S*	0.879	0.843	0.824	0.852
	P	0.901	0.829	0.822	0.851
Heteroscedastic Normal					
0.95	*S*	0.963	0.950	0.946	0.937
	P	0.966	0.948	0.943	0.951
0.90	*S*	0.922	0.906	0.896	0.895
	P	0.925	0.898	0.887	0.901
0.85	*S*	0.887	0.859	0.851	0.846
	P	0.868	0.838	0.832	0.851

Note: The model includes three regressors, a constant and two other, the real vector of coefficient is $(1, 1, 1)$, and the censoring point here is $y_0 = 0$. P denotes percentile. S denotes the standard. VNS denotes variable neighborhood search. ECP is the empirical coverage probabilities.

First model is given by

$$y = \max\{\beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + \varepsilon_i, y_0\}, \tag{8}$$

and second one can be written as

$$y = \max\{\beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + x_{i4}\beta_4 + x_{i5}\beta_5 + \varepsilon_i, y_0\}, \tag{9}$$

where $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$ equals $(1, 1, 0.5, -1, -0.5, 0.25)$. x_i are generated as a standard normal distribution, truncated as $\{\| x_i \|_\infty < 2\}$. The error term has the multiplicative herteroscedasticity structure, where it can be formulated as

$$\varepsilon_i = u_i v(x_i),$$

where $v(x_i)$ can be written as

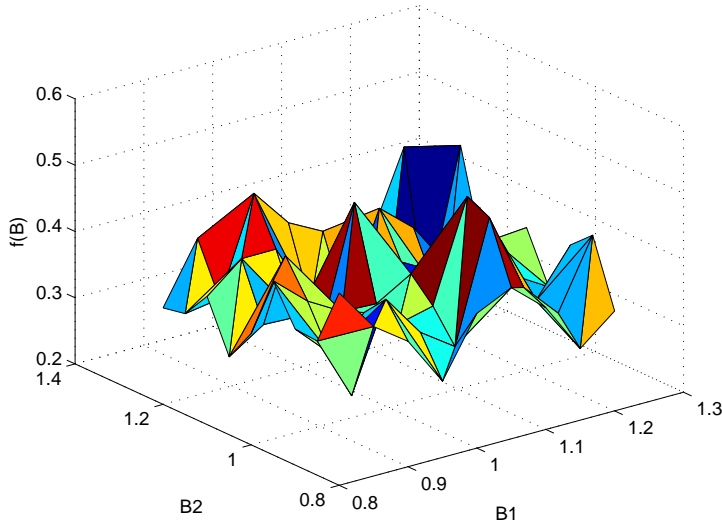$$v(x_i) = a_0 + \sum_{j=1}^{m}(a_{j1}x_{ji} + a_{j2}x_{ji}^2),$$

Fig. 7 Distribution of local minima in (β_1, β_2) space, obtained by 100 restart of `CQR-VNS`

$a_0 = 1$, $a_{j1} = 0.5$ and $a_{j2} = 0.5$. The censoring point is $y_0 = -0.75$.

Two alternative distributions are considered for u_i: a normal distribution $N(0, 25)$, and a χ^2 distribution with four degrees of freedom, re-centered to have zero median. In this part we have done the following:

- We generate the data as above according to [4], i.e. the paper we are comparing with [4].
- We repeat the simulation 10.000 times for each of the three cases of sample size: $\{(y_i, x_i), i = 1, 2, \ldots, 100\}, \{(y_i, x_i), i = 1, 2, \ldots, 400\}$, and $\{(y_i, x_i), i = 1, 2, \ldots, 600\}$.
- We apply the VNS for two cases of regression function. the first one as (8), and the second one as (9).
- We find the root mean square errors (RMSEs), mean bias, mean absolute deviation (MAE), and median bias for all (β_1, β_2).
- Our results compare with the results from the CV method (which denotes the CQR-LP estimator with log likelihood cross-validated bandwidth), the CVa method (which denotes the CV estimator with bandwidth adjusted to conform with assumption K), the PR method (which denotes the CQR-LP estimator with probit estimates for the censoring probability), the HO and HOa methods (which are the same as CV and CVa, except that the kernel function is order kernel), the CR method, which is a Powell's estimator, and the VNS method [4].

We use in Table 2 the regression function as (8). Applying VNS method to original Powell estimator for solving CQR outperforms better than the six other methods for the regression coefficient β_1 for all sample sizes and both kinds of error. On the other hand, the VNS method did not work very well for the regression

Table 2 MONTE CARLO SIMULATION WITH THREE REGRESSORS FOR 0.50 QUANTILE AND 0.75 CENSORING POINT (10,000 REPETITION)

	Intercept - β_0							Slope - β_2						
	CV	CVa	PR	HO	HOa	CR	VNS	CV	CVa	PR	HO	HOa	CR	VNS
							$N(0,25)$							
$n=100$														
$RMSE$	2.88	3.34	1.59	4.39	3.02	4.11	0.42	2.16	2.10	2.18	2.41	1.98	2.85	0.73
$Meanbias$	0.14	0.07	0.60	0.18	0.44	−0.08	−0.23	0.31	0.28	0.70	0.32	0.40	0.33	−0.65
MAE	0.82	0.80	0.94	0.83	0.92	0.74	0.23	0.86	0.83	1.21	0.85	0.90	0.92	0.65
$Meanbias$	0.93	0.32	0.70	0.45	0.60	0.35	−0.00	0.06	0.01	0.49	0.05	0.13	−0.31	−0.69
$n=400$														
$RMSE$	0.58	0.57	0.56	0.60	0.66	0.68	0.28	0.61	0.59	0.90	0.67	0.71	0.66	0.65
$Meanbias$	0.20	0.17	0.17	0.19	0.30	0.19	−0.14	−0.06	−0.10	0.28	−0.05	0.04	−0.45	−0.61
MAE	0.39	0.38	0.36	0.41	0.45	0.41	0.14	0.39	0.39	0.53	0.43	0.44	0.55	0.61
$Meanbias$	0.20	0.16	0.16	0.19	0.29	0.19	−0.00	−0.12	−0.17	0.21	−0.12	−0.05	−0.52	−0.57
$n=600$														
$RMSE$	0.48	0.48	0.46	0.48	0.52	0.49	0.24	0.50	0.48	0.71	0.54	0.56	0.57	0.62
$Meanbias$	0.18	0.16	0.14	0.12	0.20	0.20	−0.12	−0.06	−0.11	0.25	−0.17	−0.10	−0.47	−0.58
MAE	0.33	0.32	0.31	0.33	0.35	0.33	0.12	0.33	0.33	0.44	0.40	0.40	0.50	0.58
$Meanbias$	0.18	0.16	0.15	0.13	0.20	0.19	−0.00	−0.10	−0.15	0.23	−0.22	−0.16	−0.49	−0.54
							$\chi^2(4)$							
$n=100$														
$RMSE$	0.63	0.62	0.62	0.65	0.70	0.67	0.25	0.72	0.70	0.92	0.76	0.79	0.90	0.59
$Meanbias$	0.26	0.23	0.25	0.30	0.37	0.20	−0.12	−0.05	−0.08	0.19	−0.01	0.04	−0.14	−0.52
MAE	0.41	0.40	0.39	0.43	0.45	0.39	0.12	0.46	0.45	0.54	0.48	0.49	0.56	0.52
$Meanbias$	0.24	0.21	0.22	0.27	0.33	0.19	−0.00	−0.11	−0.13	0.12	−0.08	−0.03	−0.37	−0.49
$n=400$														
$RMSE$	0.32	0.31	0.28	0.32	0.35	0.31	0.13	0.35	0.35	0.39	0.37	0.37	0.51	0.50
$Meanbias$	0.17	0.16	0.09	0.17	0.21	0.12	−0.06	−0.11	−0.14	0.00	−0.09	−0.07	−0.45	−0.47
MAE	0.22	0.21	0.18	0.22	0.23	0.20	0.06	0.24	0.24	0.25	0.25	0.25	0.48	0.47
$Meanbias$	0.17	0.16	0.08	0.17	0.20	0.11	−0.00	−0.13	−0.15	−0.02	−0.11	−0.08	−0.47	−0.45
$n=600$														
$RMSE$	0.26	0.265	0.23	0.27	0.29	0.25	0.10	0.30	0.30	0.32	0.30	0.30	0.50	0.48
$Meanbias$	0.14	0.14	0.06	0.15	0.18	0.09	−0.04	−0.10	−0.13	−0.01	−0.09	−0.07	−0.46	−0.46
MAE	0.18	0.18	0.15	0.18	0.19	0.17	0.04	0.21	0.21	0.22	0.21	0.21	0.48	0.46
$Meanbias$	0.14	0.13	0.06	0.14	0.17	0.10	−0.00	−0.12	−0.14	−0.03	−0.10	−0.08	−0.48	−0.44

Note:The model includes three regressors: a constant and two random i.i.d. N(0,1)regressors. The vector of coefficients is (1,1,.5). The censoring point is set at 0.75.

coefficient β_2. The explanation is that the VNS works to give the minimum of the objective function in total, not the minimum of each component of this objective function.

In Table 3 the regression function has five regressors, as in (9). The data are generated in the same way as in Table 2.

Table 3 shows that the VNS method with Powell estimator in general for solving CQR gives very good results in the case of five regressors. We note that VNS is the best for all sample size in both kind of error for slope coefficient. In the case of the x_2 coefficient VNS works well, better than other methods most of the time. We may conclude that, when we increase the dimension of regression function, VNS method with the exact Powell estimator outperforms other methods from the literature.

Table 3 MONTE CARLO SIMULATION WITH SIX REGRESSORS FOR 0.50 QUANTILE AND 0.75CENSORING POINT (10,000 REPETITION)

	Intercept					Slope				
	CV	CVa	PR	CR	VNS	CV	CVa	PR	CR	VNS
$N(0,25)$										
$n = 100$										
$RMSE$	3.29	3.25	4.36	3.93	0.52	2.24	2.24	2.74	3.05	0.61
$Meanbias$	1.66	1.38	3.59	0.70	−0.30	0.26	0.21	0.72	0.43	−0.41
MAE	2.07	1.89	3.56	1.55	0.30	1.12	1.09	1.60	1.11	0.41
$M-bias$	1.77	1.54	3.53	1.22	−0.00	0.05	0.00	0.51	−0.12	−0.13
$n = 400$										
$RMSE$	1.34	1.28	1.46	1.31	0.42	0.78	0.77	1.15	0.89	0.52
$Meanbias$	0.78	0.62	1.15	0.74	−0.23	−0.19	−0.24	0.26	−0.38	−0.32
MAE	0.89	0.80	1.17	0.81	0.23	0.52	0.51	0.69	0.66	0.32
$M-bias$	0.79	0.64	1.15	0.78	−0.00	−0.25	−0.30	0.19	−0.54	−0.01
$n = 600$										
$RMSE$	1.11	1.01	1.07	0.91	0.38	0.63	0.63	0.88	0.71	0.48
$Meanbias$	0.67	0.54	0.81	0.64	−0.20	−0.14	−0.16	0.16	−0.50	−0.29
MAE	0.74	0.66	0.84	0.65	0.20	0.45	0.45	0.54	0.63	0.29
$M-bias$	0.67	0.56	0.81	0.62	−0.00	−0.20	−0.22	0.11	−0.60	−0.00
$\chi^2(4)$										
$n = 100$										
$RMSE$	1.53	1.46	2.01	1.33	0.38	1.02	1.03	1.34	1.23	0.51
$Meanbias$	0.96	0.87	1.63	0.69	−0.20	−0.22	−0.24	0.10	−0.12	−0.32
MAE	0.99	0.93	1.54	0.75	0.20	0.66	0.64	0.80	0.73	0.32
$M-bias$	0.93	0.84	1.53	0.68	−0.00	−0.30	−0.34	−0.02	−0.37	−0.00
$n = 400$										
$RMSE$	0.88	0.80	0.69	0.62	0.19	0.51	0.52	0.57	0.60	0.36
$Meanbias$	0.72	0.62	0.51	0.46	−0.07	−0.24	−0.29	−0.04	−0.47	−0.20
MAE	0.70	0.61	0.52	0.46	0.07	0.37	0.38	0.38	0.54	0.20
$M-bias$	0.70	0.60	0.50	0.45	−0.00	−0.27	−0.31	−0.07	−0.52	−0.00
$n = 600$										
$RMSE$	0.80	0.72	0.54	0.54	0.13	0.42	0.43	0.45	0.57	0.31
$Meanbias$	0.69	0.60	0.40	0.42	−0.04	−0.17	−0.19	−0.04	−0.50	−0.17
MAE	0.69	0.60	0.41	0.41	0.04	0.29	0.31	0.30	0.53	0.17
$M-bias$	0.69	0.60	0.40	−0.00	−0.65	−0.18	−0.20	−0.06	−0.53	−0.00

Note:The model includes six regressors: a constant and five random i.i.d. N(0,1)regressors. The vector of coefficients is (1,1,.5,-1,-.5,.25). The censoring point is set at 0.75.

4 A censored quantile regression model for extramarital affairs

The phenomenon of an extramarital affair is important in many areas of expertise such as psychology, sociology and economic see [28, 26]. In 1974, Redbook magazine published a survey of extramarital affairs. This data set had been used in Fair's study [29].

In this section, we apply CQR-VNS to an extramarital affair data. The same instance is tested by three-step CQR method in [7]. The sample consists of 6366 first-time married women. The 68.5 percent of the data set have no extramarital

affairs. In the CQR model eight independent variables and latent variable were defined as Fair defined them [26].

The dependent variable, which is also called the level of affairs, can be calculated as the number of paramors multiplied by the number of relationships with each partner, divided by the number of years of marriage. If there is no affair, the latent variable $y = 0$. The independent variables can be introduced as follows:

– Marriage Rating: there are five answers here, 5=very good, 4= good, 3= fair, 2= poor and 1= very poor.
– Age: the ages of the women are divided into six groups, 17.5= under 20, 22= from 20 to 24, 27= 25 to 29, 32= 30 to 34, 37= 35 to 39 and 42= 40 and over.
– Years of Marriage: there are divided into seven groups, 0.5= less than one year, 2.5= 1 to 4 years, 6= 5 to 7 years, 9= 8 to 10 years, 13.5= more than 10 years and the with oldest child under 12 years old, 16.5= more than 10 years and with the oldest child between 12 and 17 years old and 23= more than 10 years and with the oldest child 18 years old and over.
– Number of Children per marriage: marriages are divided into six groups, 0= none, 1= 1, 2= 2, 3= 3, 4= 4, 5.5= 5 or more.
– Religiosity: response rating 1 to 4, 1= not religious, 2= mildly, 3= fairly, 4= strongly.
– Education: six groups, 9= if the wife went only to grade school, 12= high school, 14= had some college experience, 16= was a college graduate, 17= had some graduate experience, and 20= an advanced degree.
– Occupation: also it divided to six groups, 6= professional with advanced degree, 5= managerial, administrative, business, 4= teacher, counsellor, social, social worker, nurse, artist, writer, technician, skilled worker, 3= white-collar (sales, clerical, secretarial), semiskilled or unskilled worker, other and 1= student.
– Husband's Occupation: same division as occupation.

as it was mentioned in [26].

4.1 The model

The censored regression model can be written as follow:

$$f(\beta, \theta, y_0) = \min_{\beta} \sum \rho_\theta(y - \max\{y_0, \beta_0 + x_1\beta_1 + x_2\beta_2 + x_3\beta_3$$

$$+x_4\beta_4 + x_5\beta_5 + x_6\beta_6 + x_7\beta_7 + x_8\beta_8\})$$

where the censored point is $y_0 = 0.1$, and the $\theta = \{0.4, ..., 0.9\}$ as in [26]. x_1 denotes marriage rating, x_2 denotes age, x_3 denotes years married, x_4 denotes number of children, x_5 denotes religiosity, x_6 denotes education, x_7 denotes occupation, x_8 denotes husband's occupation and y is a level of the affair. The $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8) = (7.85, -1.53, -0.107, 0.130, -0.0285, -0.944, -0.0853, 0.314, 0.0151)$, which are the final results in Fair's model [26].

4.2 Result analysis

The results from the CQR model are explained in Figure 1. As one can see on Figure 1, that the religiosity effect is strongly negative at all quantiles of affair. The education effect is strongly negative and it is increasing according to quantiles. In Fair's model, the education effect is inexplicable until they appear to have meaning in view of the relational perspectives toward an affair among the educated and intelligent individuals [30].
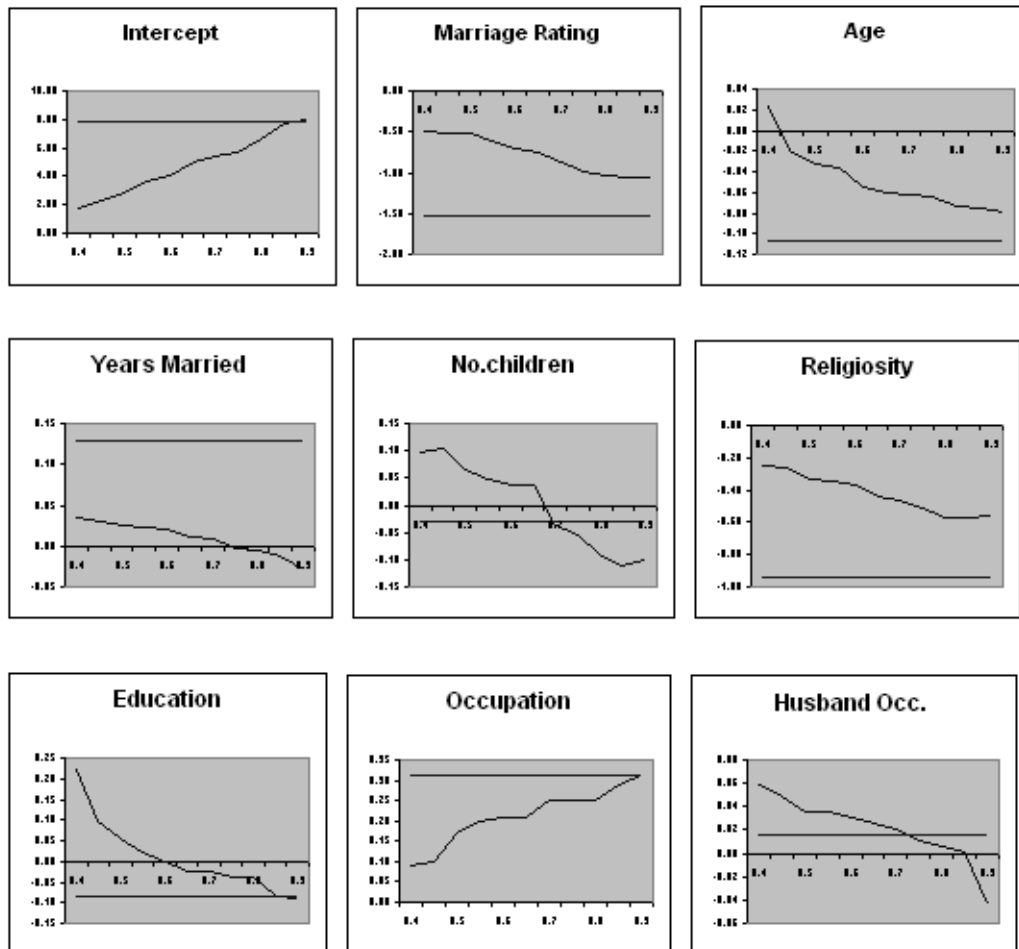

Fig. 8 CQR for extramarital affairs

Also we can note that the age effect is negative at all quantiles. This can be explained by the fact that younger wives are more likely to engage in extramarital affairs than older ones.

The number of children inside the family is strongly negative at all quantiles. This may be related to the relationship between wife and husband is more strong when the family has children.

The women's occupation effect is strongly positive. That mean the wive with high income are more likely to engage in an affair. There are two views of explanation. In [31] explains that such status creates an interactional advantage, increasing the hazard of an affair and subsequent marital dissolution. In [26] there is another explanation, where is related to income status, that means the higher status occupation gives high income. On the other hand, husband's occupation is negative, but [26] predicts the positive effect of husband's occupation, because a higher value of goods consumed in marriage causes wives to substitute labor activities for time spent with family and paramour.

5 Conclusion

In this paper we suggest a new method for solving censored quantile regression (CQR). It is based on variable neighborhood search (VNS) global optimization technique. As objective function it is used Powell estimator, which is known to be non convex-nor concave. Other approaches in the literature try to find linear approximation of Powell function or employ threshold accepting and genetic algorithm. Our method adapts VNS rules in order to solve this global optimization problem. The basic idea of VNS metaheuristic is use of different metric functions in defining neighborhood structure of the current solution.

We perform extensive computational analysis on groups of test instances. Moreover, we use one real world data instance to check how our approach behave in such type of problems.

It is appears that our new approach is competitive with state-of-the-art methods from the literature. Moreover, our results indicate that the better solution are usually obtained by using nonlinear model and effective approximate solution method that use of approximate (linear) model with exact solution procedure.

Future research may include use of other, more sophisticated global optimization techniques for solving CQR. Moreover, it may include extension our approach to semi censored quantile regression ([12]) as well. In addition, new neighborhood structures may be tried out within variable neighborhood approach.

References

1. T. Neocleous, K.V. Branden, and S. Portnoy. Correction to censored regression quantiles by S. Portnoy, 98 (2003), 1001–1012. *Journal of the American Statistical Association*, 101(474):860–861, 2006.
2. L. Peng and Y. Huang. Survival analysis with quantile regression models. *Journal of the American Statistical Association*, 103(482):637–649, 2008.
3. T. Amemiya. Two stage least absolute deviations estimators. *Econometrica*, 50:689–711, 1982.
4. M. Buchinsky and J. Hahn. An alternative estimator for the censored quantile regression model. *Econometrica*, 66:653–671, 1998.

5. P. Chaudhuri, K. Doksum, and A. Samarov. On average derivative quantile regression. *The annals of statistics*, 25:715–744, 1997.
6. S. Chen and S. Khan. Estimating censored regression models in the presence of non-parametric multiplicative heteroskedasticity. *American statistical association*, 98:283–316, 2000.
7. V. Chernozhukov and H. Hong. Three-step censored quantile regression and extramarital affairs. *American statistical association*, 97:872–882, 2002.
8. S. Portnoy. Asymptotic behavior of regression quantiles in non-stationary, dependent cases. *Journal of Multivariate analysis*, 38:100–113, 1991.
9. C.R. Rao and L.C. Zhao. Recent contributions to censored regression models. *American sociological review*, 42:203–213, 1995.
10. K. Yu, Z. Lu, and J. Stander. Quantile regression: application and current. *The statistician*, 52:331–350, 2003.
11. J.L. Powell. Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*, 25:303–325, 1984.
12. J.L. Powell. Censored regression quantiles. *Journal of Econometrics*, 32:143–155, 1986.
13. B. Fitzenberger. *Parctical methods of optimization: constrainrd optimization.* Maddala, G.S. and Rao, C.R., Handbook of statistics, Amsterdam, North-Holland, 1997.
14. R.S. Womersley. Censored discrete linear. *SIAM, Journal of Scientific and Statistical Computing*, 7:105–122, 1986.
15. R. Koenker and B.J. Park. An interior point algorithm for nonlinear quantile regression. *Journal of Econometrics*, 71:1097–1100, 1996.
16. B. Fitzenberger and P. Winker. Improving the computation of censored quantile regressions. *Computional statistics & data analysis*, 52:88–108, 2007.
17. M. Buchinsky. Change in the u.s. wage structure 1963-1987: Application of quantile regression. *Econometrica*, 62:405–458, 1994.
18. N. Mladenović and P. Hansen. Variable neighbourhood search. *Computers and operations research*, 24:1097–1100, 1997.
19. P. Hansen, N. Mladenović, and J.A. Moreno Pérez. Variable neighbourhood search: methods and applications. *4OR*, 6:319–360, 2008.
20. P. Hansen, N. Mladenović, and J.A. Moreno Pérez. Variable neighbourhood search: methods and applications. *Annals of OR*, 175:367–407, 2010.
21. M. Dražić, V. Kovačević-Vujčić, M. Čangalović, and N. Mladenović. Glob- a new vns-based software for global optimization. *Springer US*, 86:135–154, 2006.
22. N. Mladenović, M. Dražić, V. Kovačević-Vujčić, and M. Čangalović. General variable neighbourhood search for the continuous optimization. *European journal of operational research*, 191:753–770, 2008.
23. L. Liberti and M. Dražić. Variable neighbourhood search for the global optimization of constrained nlps. In *In Proceedings of GO Workshop, Almeria, Spain, 2005*, Almeria, Spain, 2005.
24. A.D. Toksari and E. Guner. Solving the unconstrained optimization problem by a variable neighbourhood search. *Mathematical Analysis and Applications*, 328:1178–1187, 2007.
25. Y. Bilias, S. Chen, and Z. Ying. Simple resampling methods for censored regression quantiles. *Journal of Econometrics*, 99:373–386, 2000.
26. R.C. Fair. Cowles foundation for research in economics. *Journal of Political economy*, 86:45–61, 1976.
27. B. Efron and R.J. Tibshirani. *An introduction to bootstrap.* Chapman and Hall, London, UK, 1993.
28. G.S. Becker. A theory of marriage: Part I. *Journal of Political Economy*, 81:813–846, 1973.
29. R.C. Fair. The effect of economic events on votes for president. *The Review of Economics and Statistics*, 60(2):159–173, 1978.
30. I. Reiss. *Family systems in america.* Reinhart and Winston, New York, Holt, 1980.
31. S.J. South and K.M. Lloyd. Spousal alternative and marital dissolution. *American sociological review*, 60:21–35, 1995.