Abstract

Quantile crossing is a common phenomenon in shape constrained nonparametric quantile regression. A recent study by Wang et al. (2014) has proposed to address this problem by imposing non-crossing constraints to convex quantile regression. However, the non-crossing constraints may violate an intrinsic quantile property. This paper proposes a penalized convex quantile regression approach that can circumvent quantile crossing while better maintaining the quantile property. A Monte Carlo study demonstrates the superiority of the proposed penalized approach in addressing the quantile crossing problem.

Keywords: Quantile function, Quantile crossing, Convex quantile regression, Simultaneous estimation, Regularization

JEL Codes: C1, C6, C13

*Corresponding author.
E-mail addresses: sheng.dai@aalto.fi (S. Dai), timo.kuosmanen@aalto.fi (T. Kuosmanen), xun.zhou@york.ac.uk (X. Zhou).
1 Introduction

Quantile estimation has been widely applied in various fields of economics and econometrics (see, e.g., Wang et al., 2014; Jradi et al., 2019; Tsionas et al., 2020; Kuosmanen & Zhou, 2021; Zhao, 2021). However, when multiple quantiles are separately estimated to obtain a family of conditional quantile functions, two or more quantile curves may cross on the condition that the distribution functions and their associated inverse functions are not monotone increasing (He, 1997). Such quantile crossing is a longstanding problem in quantile regression.

To our knowledge, there are three commonly seen approaches to avoid quantile crossing: post-processing, stepwise estimation, and simultaneous estimation. In the post-processing procedure, a non-crossing assumption is usually enforced via a sorting or monotonic rearrangement of the original estimated non-monotone functions (e.g., Dette & Volgushev, 2008; Chernozhukov et al., 2010). This indirect approach is effective in estimating the conditional quantile, but lacks the ability to quantify the effects of the predictors (Bondell et al., 2010). In the stepwise procedure, it prevents an estimated quantile function from crossing the previously estimated one by adding an extra set of non-crossing constraints iteratively to the regression model (e.g., Liu & Wu, 2009); but this approach cannot offer simultaneous estimates. In the simultaneous estimation, non-crossing constraints are imposed to ensure that the estimated conditional quantile functions are monotone nondecreasing, with all quantiles being estimated simultaneously (e.g., Takeuchi et al., 2006; Bondell et al., 2010). More recently, Wang et al. (2014) extend this simultaneous estimation technique to convex quantile regression (sCQR). However, the non-crossing constraints may disturb the quantile property (Takeuchi et al., 2006).

This paper develops a new non-crossing approach to nonparametric quantile function estimation. Compared with sCQR (Wang et al., 2014), our approach based on the penalized convex quantile regression (pCQR) independently estimates multiple non-crossing quantiles but can better satisfy the intrinsic quantile property. Furthermore, the proposed pCQR approach can fit the true quantile functions more reliably and robustly.
2 Penalized convex quantile regression

Consider a general nonparametric regression model with observations \(\{x_i, y_i\}_{i=1}^n\) satisfying

\[
y_i = f(x_i) + \varepsilon_i, \quad \text{for } i = 1, \ldots, n,
\]

where \(y_i \in \mathbb{R}\) and \(x_i \in \mathbb{R}^d\) are output and inputs variables, and \(\varepsilon_i\) is a error term with zero mean. Accordingly, for a given quantile \(\tau \in (0, 1)\), the nonparametric quantile function \(Q_y(\tau \mid x)\) is defined as

\[
Q_y(\tau \mid x_i) = f(x_i) + F^{-1}_{\varepsilon_i}(\tau),
\]

where \(F_{\varepsilon_i}\) is the distribution function of the error term \(\varepsilon_i\).

To estimate quantiles empirically, we resort to convex quantile regression (CQR) that does not require any assumptions about the functional form of the regression function \(f\) or its smoothness, but imposes the shape constraints such as monotonicity and concavity. Specifically, CQR estimates the quantile function (2) by solving the following linear programming problem (Wang et al., 2014)

\[
\begin{align*}
\min_{\alpha, \beta, \varepsilon^+, \varepsilon^-} & \quad \tau \sum_{i=1}^n \varepsilon_i^+ + (1 - \tau) \sum_{i=1}^n \varepsilon_i^- \\
\text{s.t. } & \quad y_i = \alpha_i + \beta_i' x_i + \varepsilon_i^+ - \varepsilon_i^- \quad \forall i \\
 & \quad \alpha_i + \beta_i' x_i \leq \alpha_h + \beta_h' x_i \quad \forall i, h \\
 & \quad \beta_i \geq 0 \quad \forall i \\
 & \quad \varepsilon_i^+ \geq 0, \varepsilon_i^- \geq 0 \quad \forall i
\end{align*}
\]

where the first set of constraints can be interpreted as a multivariate regression equation, the second set of constraints imposes concavity on the quantile function, the third set of constraints guarantees monotonicity, and the last refers to sign constraints of the error terms. Note that there exists an intrinsic quantile property in terms of the optimal solutions to problem (3), \(\hat{\varepsilon}_i^+\) and \(\hat{\varepsilon}_i^-\).

**Theorem 1.** For any \(\tau \in (0, 1)\), the number of strict positive residuals \((\hat{\varepsilon}_i^+ > 0)\) by \(n_\tau^+\) and the number of strict negative residuals \((\hat{\varepsilon}_i^- > 0)\) by \(n_\tau^-\) always satisfy the inequalities:

\[
\frac{n_\tau^+}{n} \leq 1 - \tau \quad \text{and} \quad \frac{n_\tau^-}{n} \leq \tau.
\]
Proof. See proofs in Wang et al. (2014) and Kuosmanen & Zhou (2021).

Compared with the conventional full frontier estimation, the quantile function estimation is more robust to random noise, heteroscedasticity, and the choice of direction vectors. However, when separately estimating each conditional quantile function $Q(y | x)$, CQR is likely to violate the assumption that the distribution functions and their associated inverse functions should be monotone nondecreasing; see Fig. A1 for an example of the quantile crossing problem detected in our empirical application of CQR.

We notice that the quantile crossing problem could be addressed by simultaneous estimation, which imposes an extra set of linear non-crossing constraints in the CQR approach (see, e.g., Takeuchi et al., 2006; Wang et al., 2014). Following Wang et al. (2014), the simultaneous convex quantile regression (sCQR) estimator of $j$ conditional quantile functions at $0 < \tau_1 < \tau_2 < \cdots < \tau_j < 1$ is formulated as

$$
\min_{\alpha, \beta, \varepsilon^+, \varepsilon^-} \sum_{j=1}^{J} \left( \tau_j \sum_{i=1}^{n} \varepsilon_{i,j}^+ + (1 - \tau_j) \sum_{i=1}^{n} \varepsilon_{i,j}^- \right)
$$

s.t.

$$
\begin{align*}
    y_{i,j} &= \alpha_{i,j} + \beta_{i,j}' x_i + \varepsilon_{i,j}^+ - \varepsilon_{i,j}^- & \forall i, j \\
    \alpha_{i,j} + \beta_{i,j}' x_i &\leq \alpha_{h,j} + \beta_{h,j}' x_i & \forall i, h, j \\
    \alpha_{i,j} + \beta_{i,j}' x_i + C_{i,j} &\leq \alpha_{i,j+1} + \beta_{i,j+1}' x_i & \forall i, 1 \leq j \leq J - 1 \\
    \beta_{i,j} &\geq 0 & \forall i, j \\
    \varepsilon_{i,j}^+ &\geq 0, \varepsilon_{i,j}^- & \geq 0 & \forall i, j
\end{align*}
$$

where $C_{i,j} \geq 0$ are small nonnegative constants for quantiles, which are introduced in sCQR to ensure that $Q_y(\tau_j | x) \leq Q_y(\tau_{j+1} | x), \forall i$ and $j \in J$. For the purpose of non-crossing, $C$ can simply be given by zero; that is, there may exist touching rather than crossing between two neighboring quantiles (see Fig. 1a for an illustration). In practice, however, after enforcing the non-crossing constraints, sCQR may violate the quantile property (Theorem 1) due to the fact that the approach simultaneously optimizes for both the quantile property and the non-crossing property (Takeuchi et al., 2006).

This paper proposes an alternative to sCQR to address the quantile crossing problem. By using the $L_2$-norm regularization on subgradients $\beta_i$, we formulate penalized convex quantile regression (pCQR) as
\[
\min_{\alpha, \beta, \tau, \epsilon, \tau'} \sum_{i=1}^{n} \epsilon_i^+ + (1 - \tau) \sum_{i=1}^{n} \epsilon_i^- + \gamma \sum_{i=1}^{n} ||\beta_i||_2
\]
\[
s.t. \quad y_i = \alpha_i + \beta_i'x_i + \epsilon_i^+ - \epsilon_i^- \quad \forall i
\]
\[
\alpha_i + \beta_i'x_i \leq \alpha_h + \beta_h'x_i \quad \forall i, h
\]
\[
\beta_i \geq 0 \quad \forall i
\]
\[
\epsilon_i^+ \geq 0, \epsilon_i^- \geq 0 \quad \forall i
\]

where \(\gamma \geq 0\) is the tuning parameter and \(|| \cdot ||_2\) denotes the standard Euclidean norm. As \(\gamma\) approaches to zero, pCQR (5) collapses to the original CQR problem (3). The rationale behind the avoidance of quantile crossing in pCQR lies in that as \(\gamma \to \infty\), the regularization will dominate the minimization and then all estimated subgradients \(\beta_i\) “flatten out” to 0. In this case, the estimated quantile functions will be horizontal lines (for \(d = 1\)) or planes \((d > 1)\).

As \(\gamma\) increases, the quantile property may also be disturbed in pCQR. But for sufficiently small \(\gamma\), the quantile property can be guaranteed in theory. Therefore, we design the following Algorithm 1 to obtain the supremum \(\gamma^*\) so as to avoid quantile crossing and ensure the quantile property as well as possible in estimating multiple quantile functions.

**Algorithm 1:** Searching the optimal tuning parameter \(\gamma^*\).

**Data:** \(\{x_i, y_i\}_{i=1}^{n} \in \mathbb{R}^d \times \mathbb{R}, \tau_1 \text{ and } \tau_2 \text{ (} \tau_1 < \tau_2\)\)

1. out = 0 and \(\gamma = 0\);
2. while out = 0 do
3. \quad Solve problem (5) with quantiles \(\tau_1\) and \(\tau_2\), separately, to calculate \(\hat{Q}_y(\tau_1 | x_i)\) and \(\hat{Q}_y(\tau_2 | x_i)\);
4. \quad if \(\sum_{i=1}^{n} \{Q_y(\tau_1 | x_i) - Q_y(\tau_2 | x_i) \leq 0\} \neq n\) then
5. \quad \quad Re-solve problem (5) using the updated \(\gamma\);
6. \quad else
7. \quad \quad out = 1;
8. \quad \quad \gamma = \gamma + 0.01;
9. Result: \(\gamma^*\)

We proceed to illustrate how non-crossing quantile functions look like with a real dataset used in Kuosmanen & Zhou (2021). It contains plant-level data on 130 U.S. electric power plants operating in 2014; see Kuosmanen & Zhou (2021) for a more detailed description of
the data. For the sake of demonstration, we simply consider a univariate case of one input and one output. The input is the total cost involved in electricity production and the output is the net electricity generation of each power plant. Both variables are in natural logarithm.

An application of CQR to the empirical data finds that the 15\textsuperscript{th} quantile curve crosses the 25\textsuperscript{th} quantile curve twice (see Fig. A1). We then demonstrate how the sCQR and pCQR approaches can address this problem. It is evident from Fig. 1 that both approaches manage to circumvent the quantile crossing problem; that is, we observe that \( \hat{Q}_y(0.25 | x_i) \) is greater than or equal to \( \hat{Q}_y(0.15 | x_i) \) in both approaches. However, the shapes of the estimated quantile functions in Figs. 1a and 1b (see particularly the upper right corner) are slightly different. As mentioned earlier, this difference arises because sCQR tries to simultaneously optimize for the non-crossing property, the quantile property, and the production axioms, whereas the pCQR approach independently estimates the quantile production functions. Further, the difference affects which approach can better retain the quantile property.

3 Monte Carlo study

We perform a Monte Carlo study to examine whether pCQR or sCQR can better satisfy the quantile property while addressing quantile crossing. Consider the following data generating process (Dai, 2021)

\[
y_i = \prod_{d=1}^{D} x_{d,i}^{0.8} + v_i - u_i,
\]

where the input matrix \( x_i \in \mathbb{R}^{n \times d} \) is generated independently from \( U[1,10] \), and noise \( v_i \) and inefficiency \( u_i \) are drawn independently from \( N(0, \sigma_v^2) \) and \( N^+(0, \sigma_u^2) \), respectively. To investigate the robustness of the quantile approaches, following Aigner et al. (1977) we use different combinations of noise (\( \sigma \)) and signal to noise ratio (\( \lambda \)), that is, \( (\sigma^2, \lambda) = (1.88, 1.66), (1.63, 1.24), \) and \( (1.35, 0.83) \).

We consider 54 scenarios with \( n \in \{99, 199, 499\} \), \( d \in \{2, 3, 4\} \), and \( \tau \in \{0.85, 0.90, 0.95\} \). Each scenario is replicated 500 times using the pyStoNED package (Dai et al., 2021) on Python with the standard solver Mosek (9.3). We then compute the ramp loss (RL = \( \frac{1}{n} \sum_{i=1}^{n} 1_{y_i > \hat{Q}_y(\tau | x_i)} - \tau \)) (Takeuchi et al., 2006) to examine the quantile property and the mean squared error (MSE) to evaluate the finite-sample performance. Replications for which no quantile crossing happen (i.e., \( \gamma^* = 0 \)) are excluded from the calculations of RL and MSE.
Note that the smaller the ramp loss, the better the quantile performance.

Tables 1 and A1 present the estimated ramp loss and MSE statistics across different scenarios. The results clearly show that compared with sCQR, pCQR has lower ramp loss in virtually all the scenarios and lower MSE in all the scenarios. This is because the optimal tuning parameter $\gamma^*$ used in the pCQR simulations mainly locates in the interval $[0.01, 0.03]$ (see Fig. 2), suggesting that the quantile property can be guaranteed to a certain extent and thus pCQR can better fit the true quantile functions. Several other findings are summarized as follows:

- The higher the dimension or the noise in data space, the lower the ramp loss. As $d$ or $\sigma$ increases, the data space becomes more sparse, thereby indicating that the probability of crossing between two neighboring quantiles is relatively small.

- The differences in the ramp loss among quantiles in pCQR are smaller than those in sCQR due to the different estimation strategies, i.e., independent and simultaneous estimation, respectively.

- For both approaches, the MSE increases as more inputs are included and decreases as the sample size gets larger.

- The performance of both approaches in terms of MSE becomes worse as the signal to noise ratio increases.

Overall, the pCQR approach can better satisfy the quantile property and fit the true quantile functions while at the same time addressing the quantile crossing problem. Regularizing the quantile function, instead of imposing an extra set of non-crossing constraints, proves a better remedy to quantile crossing according to our simulations.
Fig. 1. Empirical illustration of estimated non-crossing quantile functions using the U.S. power plant data.
Table 1. Ramp loss and MSE of two neighboring quantiles with $n = 499$.

| $d$ | $(\sigma^2, \lambda)$ | $\tau_1$ | $\tau_2$ | RL$_{\tau_1}$ | MSE$_{\tau_1}$ | RL$_{\tau_2}$ | MSE$_{\tau_2}$ |
|-----|------------------------|----------|----------|----------------|----------------|----------------|----------------|
| 2   | (1.35, 0.83)           | 0.85     | 0.90     | 0.928          | 0.993          | 0.065          | 0.093          |
|     |                        | 0.90     | 0.95     | 0.998          | 1.078          | 0.074          | 0.117          |
|     | (1.63, 1.24)           | 0.85     | 0.90     | 0.899          | 0.953          | 0.075          | 0.108          |
|     |                        | 0.90     | 0.95     | 0.972          | 1.061          | 0.081          | 0.134          |
|     | (1.88, 1.66)           | 0.85     | 0.90     | 0.898          | 0.967          | 0.079          | 0.117          |
|     |                        | 0.90     | 0.95     | 0.890          | 0.963          | 0.093          | 0.153          |
| 3   | (1.35, 0.83)           | 0.85     | 0.90     | 0.811          | 0.942          | 0.103          | 0.205          |
|     |                        | 0.90     | 0.95     | 0.775          | 0.902          | 0.122          | 0.253          |
|     | (1.63, 1.24)           | 0.85     | 0.90     | 0.832          | 0.950          | 0.110          | 0.226          |
|     |                        | 0.90     | 0.95     | 0.827          | 0.959          | 0.138          | 0.281          |
|     | (1.88, 1.66)           | 0.85     | 0.90     | 0.792          | 0.913          | 0.127          | 0.250          |
|     |                        | 0.90     | 0.95     | 0.789          | 0.914          | 0.172          | 0.332          |
| 4   | (1.35, 0.83)           | 0.85     | 0.90     | 0.658          | 0.837          | 0.163          | 0.368          |
|     |                        | 0.90     | 0.95     | 0.627          | 0.794          | 0.207          | 0.487          |
|     | (1.63, 1.24)           | 0.85     | 0.90     | 0.587          | 0.801          | 0.168          | 0.397          |
|     |                        | 0.90     | 0.95     | 0.696          | 0.861          | 0.217          | 0.535          |
|     | (1.88, 1.66)           | 0.85     | 0.90     | 0.584          | 0.793          | 0.189          | 0.446          |
|     |                        | 0.90     | 0.95     | 0.646          | 0.795          | 0.285          | 0.581          |

Fig. 2. Empirical distribution of $\gamma^*$ with $n = 499$, $(\sigma^2, \lambda) = (1.35, 0.83)$, $\tau_1 = 0.85$, and $\tau_2 = 0.90$. 
4 Conclusions

In this paper, a penalized convex quantile regression approach has been developed to address the quantile crossing problem. The proposed algorithm can search the optimal tuning parameter such that the occurrences of quantile crossing are avoided and the quantile property is ensured as well as possible. A Monte Carlo study confirms the superiority of the proposed approach compared to previous work in addressing the quantile crossing problem. We believe the proposed approach can be readily extended from convex quantile regression to other nonparametric and parametric quantile regression techniques. We leave such extensions as fascinating avenues for future research.

References

Aigner, D., Lovell, C. A. K. & Schmidt, P. (1977), Formulation and estimation of stochastic frontier production function models, *Journal of Econometrics* 6, 21–37.

Bondell, H. D., Reich, B. J. & Wang, H. (2010), Noncrossing quantile regression curve estimation, *Biometrika* 97, 825–838.

Chernozhukov, V., Fernández-Val, I. & Galichon, A. (2010), Quantile and probability curves without crossing, *Econometrica* 78, 1093–1125.

Dai, S. (2021), Variable selection in convex quantile regression: L1-norm or L0-norm regularization?, *arxiv preprint arxiv:2107.03119*.

Dai, S., Fang, Y. H., Lee, C. Y. & Kuosmanen, T. (2021), pyStoNED: A Python package for convex regression and frontier estimation, *arxiv preprint arxiv:2109.12962*.

Dette, H. & Volgushev, S. (2008), Non-crossing non-parametric estimates of quantile curves, *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 70, 609–627.

He, X. (1997), Quantile curves without crossing, *American Statistician* 51, 186–192.

Jradi, S., Parmeter, C. F. & Ruggiero, J. (2019), Quantile estimation of the stochastic frontier model, *Economics Letters* 182, 15–18.

Kuosmanen, T. & Zhou, X. (2021), Shadow prices and marginal abatement costs: Convex quantile regression approach, *European Journal of Operational Research* 289, 666–675.

Liu, Y. & Wu, Y. (2009), Stepwise multiple quantile regression estimation using non-crossing constraints, *Statistics and Its Interface* 2, 299–310.

Takeuchi, I., Le, Q. V., Sears, T. D. & Smola, A. J. (2006), Nonparametric quantile estimation, *Journal of Machine Learning Research* 7, 1231–1264.

Tsionas, M. G., Assaf, A. G. & Andrikopoulos, A. (2020), Quantile stochastic frontier models with endogeneity, *Economics Letters* 188.
Wang, Y., Wang, S., Dang, C. & Ge, W. (2014), Nonparametric quantile frontier estimation under shape restriction, *European Journal of Operational Research* 232, 671–678.

Zhao, S. (2021), Quantile estimation of stochastic frontier models with the normal-half normal specification: A cumulative distribution function approach, *Economics Letters* 206, 109998.
Appendix

A Supplementary tables and figures

Fig. A1. Illustration of the quantile crossing problem in CQR estimation.
Table A1. Ramp loss and MSE of two neighboring quantiles with \( n = 99 \) and 199.

| \( n \) | \( d \) | \((\sigma^2, \lambda)\) | \( \tau_1 \) | \( \tau_2 \) | RL\(_{\tau_1}\) | MSE\(_{\tau_1}\) | RL\(_{\tau_2}\) | MSE\(_{\tau_2}\) |
|-------|------|-----------------|-------|-------|--------|--------|--------|--------|
| 99    | 2    | \((1.35, 0.83)\) | 0.85  | 0.90  | 0.824  | 0.882  | 0.239  | 0.350  |
|       |      | \((1.63, 1.24)\) | 0.85  | 0.90  | 0.806  | 0.882  | 0.248  | 0.378  |
|       |      | \((1.88, 1.66)\) | 0.85  | 0.90  | 0.831  | 0.898  | 0.272  | 0.435  |
|       |      |                  | 0.90  | 0.95  | 0.857  | 0.918  | 0.336  | 0.534  |
|       | 3    | \((1.35, 0.83)\) | 0.85  | 0.90  | 0.719  | 0.753  | 0.349  | 0.619  |
|       |      | \((1.63, 1.24)\) | 0.85  | 0.90  | 0.671  | 0.721  | 0.400  | 0.688  |
|       |      | \((1.88, 1.66)\) | 0.85  | 0.90  | 0.677  | 0.741  | 0.439  | 0.782  |
|       |      |                  | 0.90  | 0.95  | 0.611  | 0.662  | 0.541  | 1.046  |
|       | 4    | \((1.35, 0.83)\) | 0.85  | 0.90  | 0.412  | 0.491  | 0.591  | 1.041  |
|       |      | \((1.63, 1.24)\) | 0.85  | 0.90  | 0.455  | 0.550  | 0.688  | 1.139  |
|       |      | \((1.88, 1.66)\) | 0.85  | 0.90  | 0.502  | 0.606  | 0.664  | 1.180  |
|       |      |                  | 0.90  | 0.95  | 0.412  | 0.449  | 0.931  | 1.624  |
| 199   | 2    | \((1.35, 0.83)\) | 0.85  | 0.90  | 0.837  | 0.916  | 0.139  | 0.201  |
|       |      | \((1.63, 1.24)\) | 0.85  | 0.90  | 0.869  | 0.957  | 0.140  | 0.210  |
|       |      | \((1.88, 1.66)\) | 0.85  | 0.90  | 0.868  | 0.931  | 0.150  | 0.236  |
|       |      |                  | 0.90  | 0.95  | 0.945  | 1.012  | 0.186  | 0.322  |
|       | 3    | \((1.35, 0.83)\) | 0.85  | 0.90  | 0.656  | 0.777  | 0.203  | 0.374  |
|       |      | \((1.63, 1.24)\) | 0.85  | 0.90  | 0.696  | 0.815  | 0.221  | 0.436  |
|       |      | \((1.88, 1.66)\) | 0.85  | 0.90  | 0.713  | 0.837  | 0.249  | 0.461  |
|       |      |                  | 0.90  | 0.95  | 0.657  | 0.716  | 0.335  | 0.653  |
|       | 4    | \((1.35, 0.83)\) | 0.85  | 0.90  | 0.584  | 0.703  | 0.282  | 0.663  |
|       |      | \((1.63, 1.24)\) | 0.85  | 0.90  | 0.503  | 0.670  | 0.331  | 0.736  |
|       |      | \((1.88, 1.66)\) | 0.85  | 0.90  | 0.428  | 0.622  | 0.354  | 0.806  |
|       |      |                  | 0.90  | 0.95  | 0.472  | 0.630  | 0.515  | 1.084  |