

# Stepwise Estimation for Multiple Non-Crossing Quantile Regression using Kernel Constraints

Sungwan Bang<sup>a</sup> · Myoungshic Jhun<sup>b</sup> · HyungJun Cho<sup>b,1</sup>

<sup>a</sup>Department of Mathematics, Korea Military Academy

<sup>b</sup>Department of Statistics, Korea University

(Received July 30, 2013; Revised October 25, 2013; Accepted October 25, 2013)

---

## Abstract

Quantile regression can estimate multiple conditional quantile functions of the response, and as a result, it provide comprehensive information of the relationship between the response and the predictors. However, when estimating several conditional quantile functions separately, two or more estimated quantile functions may cross or overlap and consequently violate the basic properties of quantiles. In this paper, we propose a new stepwise method to estimate multiple non-crossing quantile functions using constraints on the kernel coefficients. A simulation study are presented to demonstrate satisfactory performance of the proposed method.

Keywords: Kernel, multiple quantile regression, non-crossing, quadratic programming.

---

## 1. 서론

회귀분석에서 제곱 손실함수를 이용한 최소제곱추정은 계산상의 이점 등에 힘입어 보편적으로 널리 이용되고 있는 추정법이다. 그러나 최소제곱추정은 이상치(outlier)가 존재하거나 꼬리가 두꺼운 오차항의 분포에서는 사용이 제한된다. 이러한 경우에 절대 손실함수를 사용한 최소절대추정은 매우 유용하며 제곱 손실함수에 비해 강건한(robust)한 결과를 제공한다 (Wang 등, 2007; Xu와 Ying, 2010). 또한 Koenker과 Bassett (1978)에 의해 제안된 분위수 회귀는 최소절대추정을 일반화한 것으로 회귀계수 추정의 강건성과 유용성을 바탕으로 의학 (Cole과 Green, 1992; Heagerty와 Pepe, 1999), 경제 (Hendricks와 Koenker, 1992; Koenker과 Hallock, 2001), 생존분석 (Yang, 1999; Koenker과 Geling, 2001), microarray 연구 (Wang과 He, 2007) 등 여러 다양한 분야에 적용되고 있다.

$p$ 차원 설명변수  $\mathbf{x} \in D \subset R^p$ 와 1차원 반응변수  $y \in R$ 로 이루어진 훈련자료  $\{\mathbf{x}_i, y_i\}_{i=1}^n$ 가 주어졌다고 하자. 여기서  $D$ 는 설명변수의 정의역(domain)을 나타낸다. 반응변수  $y$ 에 대한  $100\tau\%$  조건부 분위수 함수  $q_\tau(\mathbf{x})$ 는

$$P(y \leq q_\tau(\mathbf{x})|\mathbf{x}) = \tau, \quad \text{단 } 0 < \tau < 1 \quad (1.1)$$

---

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT and Future Planning(NRF-2013R1A1A1007536) for S. Bang and funded by the Ministry of Education(NRF-2013R1A1A2A10007545) for M. Jhun.

<sup>1</sup>Corresponding author: Associate professor, Department of Statistics, Korea University, Seoul 136-701, Korea. E-mail: [hj4cho@korea.ac.kr](mailto:hj4cho@korea.ac.kr)

와 같이 정의되며, Koenker과 Bassett (1978)은 체크(check) 손실함수  $\rho_\tau(t) = t(\tau - I(t < 0))$ 를 이용한 분위수 회귀의 적합식

$$\min_{q_\tau} \sum_{i=1}^n \rho_\tau(y_i - q_\tau(\mathbf{x}_i))$$

을 통해 조건부 분위수 함수  $q_\tau(\mathbf{x})$ 를 추정하였다. 회귀함수의 추정에서 과대적합의 문제를 해결하고 예측력을 향상시키기 위한 방법으로 벌점함수를 이용한 추정법에 관하여 많은 연구가 진행되었다. Li 등 (2007)는 양정치 커널함수(kernel function)  $K(\cdot, \cdot)$ 에 대응되는 재생 커널 힐버트 공간(reproducing kernel hilbert space; RKHS)  $H_K$ 에서 정의된 벌점함수  $\|q_\tau\|_{H_K}^2$ 를 추가한 커널 분위수 회귀(kernel quantile regression; KQR)의 적합식을

$$\min_{q_\tau \in H_K} \sum_{i=1}^n \rho_\tau(y_i - q_\tau(\mathbf{x}_i)) + \|q_\tau\|_{H_K}^2 \quad (1.2)$$

와 같이 제안하였다. 이때 representer 정리 (Kimeldorf와 Wahba, 1971)를 이용하면 적합식 (1.2)의 해는

$$q_\tau(\mathbf{x}) = \sum_{i=1}^n \beta_{i,\tau} K(\mathbf{x}_i, \mathbf{x}) + b_\tau$$

와 같은 형태로 표현되고, 이를 이용하여 커널 분위수 회귀(KQR)의 적합식 (1.2)를

$$\min_{\beta_\tau, b_\tau} \sum_{i=1}^n \rho_\tau \left( y_i - \sum_{j=1}^n \beta_{j,\tau} K(\mathbf{x}_i, \mathbf{x}_j) - b_\tau \right) + \lambda \sum_{i=1}^n \sum_{j=1}^n \beta_{i,\tau} \beta_{j,\tau} K(\mathbf{x}_i, \mathbf{x}_j)$$

와 같이 다시 표현할 수 있다. 여기서  $\lambda > 0$ 는 조율모수이다. 커널함수는 사용의 편의성, 계산의 효율성 및 유연성 등과 같은 장점을 바탕으로 비선형 함수의 추정에 많이 활용되고 있으며, 본 논문에서는 일반적으로 많이 사용되는 가우스 커널 함수  $K(\mathbf{x}, \mathbf{x}') = \exp(-\|\mathbf{x} - \mathbf{x}'\|^2 / 2\sigma^2)$ 를 주로 사용하였다.

많은 실제자료의 분석에서는 특정한 하나의  $\tau$ 값에 대한 분위수 함수보다는 여러 개의 분위수 함수  $q_{\tau_k}(\mathbf{x})$  ( $k = 1, 2, \dots, K$ )에 대한 추정을 필요로 하는 경우가 빈번히 발생한다. 이때 커널 분위수 회귀의 적합식 (1.3)를 이용하여 각각의 분위수 함수들을 개별적으로 추정하게 되면, 이들이 서로 교차할 가능성이 있다. 이러한 분위수 함수의 교차(quantile crossing) 현상은 식 (1.1)의 조건부 분위수 함수의 정의에 위배되는 것으로, 추정된 분위수 함수의 해석을 어렵게 할 뿐만 아니라 예측력을 감소시킨다. 따라서 본 논문에서는 서로 교차하지 않는(non-crossing) 분위수 함수의 추정법에 대하여 연구하고 새로운 추정법을 제안하고자 한다. 논문의 구성은 다음과 같다. 2절에서는 커널계수에 기반한 비교차 분위수 함수의 순차적 추정법과 효율적인 계산 알고리즘을 제안하였다. 3절에서는 모의실험을 통해 기존의 추정법과 제안한 방법론의 성능을 비교하였으며, 제안된 방법론의 활용가능성을 보였다. 마지막으로 4절에서는 결론과 더불어 차후 연구방향을 제시하였다.

## 2. 비교차 다중 커널 분위수 함수의 순차적 추정법

분위수 회귀는 다양한  $\tau$ 값( $0 < \tau_1 < \tau_2 < \dots < \tau_K < 1$ )에 대하여  $100\tau\%$  조건부 분위수를 추정함으로써 반응변수의 조건부 분포에 대한 포괄적인 정보를 제공하는 이점이 있다. 식 (1.1)에서 정의된 바와 같이 분위수 함수  $q_{\tau_k}(\mathbf{x})$ 는 이론적으로

$$q_{\tau_k}(\mathbf{x}) < q_{\tau_{k+1}}(\mathbf{x}), \quad \text{for } k = 1, 2, \dots, K-1 \text{ and } \forall \mathbf{x} \in D \quad (2.1)$$

을 만족해야 한다. 그러나  $K$ 개의 분위수 함수  $q_{\tau_k}(\mathbf{x})(k = 1, 2, \dots, K)$ 를 KQR의 적합식 (1.3)으로부터 개별적으로 추정하면 이들 분위수 함수들이 서로 교차할 가능성이 있다. 이러한 분위수 교차 현상은 추정의 정확도를 감소시킬 뿐만 아니라 추정된 회귀모형의 해석을 어렵게 한다.

다중 분위수 함수의 추정에서 교차현상을 해결하기 위해 비교차의 제약식을 이용한 방법론이 연구되었다. Takeuchi 등 (2006)은 훈련자료에 기반한 비교차의 제약식을 모형의 적합식에 추가함으로써 서로 다른  $\tau$ 값에 대하여 교차하지 않는 다중 분위수 함수의 추정법(Data based Non-crossing KQR; DN-KQR)을 제안하였으며, DN-KQR의 적합식은

$$\min_{\beta_{\tau_k}, b_{\tau_k}} \sum_{k=1}^K \left\{ \sum_{i=1}^n \rho_{\tau_k} \left( y_i - \sum_{j=1}^n \beta_{j, \tau_k} K(\mathbf{x}_i, \mathbf{x}_j) - b_{\tau_k} \right) \right\} + \lambda \sum_{k=1}^K \left\{ \sum_{i=1}^n \sum_{j=1}^n \beta_{i, \tau_k} \beta_{j, \tau_k} K(\mathbf{x}_i, \mathbf{x}_j) \right\}$$

subject to

$$\sum_{j=1}^n \beta_{j, \tau_k} K(\mathbf{x}_i, \mathbf{x}_j) + b_{\tau_k} \leq \sum_{j=1}^n \beta_{j, \tau_{k+1}} K(\mathbf{x}_i, \mathbf{x}_j) + b_{\tau_{k+1}}, \quad \text{for } k = 1, \dots, K-1 \text{ and } i = 1, \dots, n$$

와 같다. DN-KQR은 비교차 다중 분위수 함수의 추정을 위해  $(n+1) \times K$ 개의 의사결정변수(decision variable),  $2n \times K$ 개의 여유변수(slack variable),  $n \times (K-1)$ 개의 비교차 제약식,  $n \times K$ 개의 등호형 제약식, 그리고  $2n \times K$ 개의 비음 제약식으로 구성된 하나의 QP(quadratic programming) 문제로 공식화된다. 따라서 훈련자료의 크기  $n$ 과 추정하고자 하는 분위수 함수의 개수  $K$ 가 커짐에 따라 계산 알고리즘의 효율이 매우 낮아지게 된다. DN-KQR의 계산 효율을 높이기 위해 Wu와 Liu (2009)는 분위수 함수를 단계적으로 추정하는 방법(Stepwise DN-KQR; SDN-KQR)을 제안하였다. 훈련자료  $\mathbf{x}_i (i = 1, 2, \dots, n)$ 에 대한  $100\tau_k\%$  조건부 분위수의 추정값을  $\hat{q}_{\tau_k}(\mathbf{x}_i) = \sum_{j=1}^n \hat{\beta}_{j, \tau_k} K(\mathbf{x}_i, \mathbf{x}_j) + \hat{b}_{\tau_k}$ 라고 하면, SDN-KQR은  $100\tau_{k+1}\%$  조건부 분위수 함수  $q_{\tau_{k+1}}(\mathbf{x})$ 를

$$\min_{\beta_{\tau_{k+1}}, b_{\tau_{k+1}}} \sum_{i=1}^n \rho_{\tau_{k+1}} \left( y_i - \sum_{j=1}^n \beta_{j, \tau_{k+1}} K(\mathbf{x}_i, \mathbf{x}_j) - b_{\tau_{k+1}} \right) + \lambda \sum_{i=1}^n \sum_{j=1}^n \beta_{i, \tau_{k+1}} \beta_{j, \tau_{k+1}} K(\mathbf{x}_i, \mathbf{x}_j)$$

subject to

$$\sum_{j=1}^n \hat{\beta}_{j, \tau_k} K(\mathbf{x}_i, \mathbf{x}_j) + \hat{b}_{\tau_k} \leq \sum_{j=1}^n \beta_{j, \tau_{k+1}} K(\mathbf{x}_i, \mathbf{x}_j) + b_{\tau_{k+1}}, \quad \text{for } i = 1, 2, \dots, n \tag{2.2}$$

와 같은 적합식으로 추정하고,  $100\tau_{k-1}\%$  조건부 분위수 함수  $q_{\tau_{k-1}}(\mathbf{x})$ 는

$$\min_{\beta_{\tau_{k-1}}, b_{\tau_{k-1}}} \sum_{i=1}^n \rho_{\tau_{k-1}} \left( y_i - \sum_{j=1}^n \beta_{j, \tau_{k-1}} K(\mathbf{x}_i, \mathbf{x}_j) - b_{\tau_{k-1}} \right) + \lambda \sum_{i=1}^n \sum_{j=1}^n \beta_{i, \tau_{k-1}} \beta_{j, \tau_{k-1}} K(\mathbf{x}_i, \mathbf{x}_j)$$

subject to

$$\sum_{j=1}^n \beta_{j, \tau_{k-1}} K(\mathbf{x}_i, \mathbf{x}_j) + b_{\tau_{k-1}} \leq \sum_{j=1}^n \hat{\beta}_{j, \tau_k} K(\mathbf{x}_i, \mathbf{x}_j) + \hat{b}_{\tau_k}, \quad \text{for } i = 1, 2, \dots, n \tag{2.3}$$

을 통해 순차적으로 추정한다.

분위수 회귀의 적합식에 추가된 제약식 (2.2)와 (2.3)은 주어진 훈련자료  $\mathbf{x}_i (i = 1, 2, \dots, n)$ 에서 분위수들의 비교차를 보장하지만, 설명변수의 정의역 전체구간에 대하여 비교차를 보장하지는 못한다. 다시 말해, 주어진 훈련자료로부터 멀리 있는 설명변수의 구간에 대해서는 추정된 분위수 함수들이 서로 교차할 가능성이 여전히 존재한다. 비음 커널함수  $K(\cdot, \cdot) \geq 0$ 에 대하여  $b_{\tau_k} \leq b_{\tau_{k+1}}$  이고  $\beta_{j, \tau_k} \leq$

$\beta_{j,\tau_{k+1}}$  ( $j = 1, 2, \dots, n$ )이던  $K$ 개의 분위수 함수들은 설명변수의 전 구간에 대하여 식 (2.1)을 만족한다. 이러한 커널의 계수(coefficient)에 기반한 비교차의 제약식을 이용하여 Liu와 Wu (2011)는 비교차 다중 분위수 함수의 동시적 추정법을 제안하였다. 그러나 DN-KQR에서와 마찬가지로 Liu와 Wu (2011)의 방법론은 훈련자료의 크기  $n$ 과 추정하고자 하는 분위수 함수의 개수  $K$ 가 커짐에 따라 계산 알고리즘 측면에서 비효율적이다. 따라서 본 논문에서는 커널계수에 기반한 비교차 다중 분위수 함수의 순차적 추정법(Stepwise Coefficients based Non-crossing KQR; SCN-KQR)을 제시하고자 한다.  $\tau_k$ 에 대한 조건부 분위수  $q_{\tau_k}(\mathbf{x}_i)$ 의 커널계수의 추정값을  $\hat{b}_{\tau_k}$ 와  $\hat{\beta}_{j,\tau_k}$  ( $j = 1, 2, \dots, n$ )라고 하면, 제안된 SCN-KQR은  $100\tau_{k+1}\%$  조건부 분위수 함수  $q_{\tau_{k+1}}(\mathbf{x})$ 의 커널계수  $b_{\tau_{k+1}}$ 와  $\beta_{j,\tau_{k+1}}$  ( $j = 1, 2, \dots, n$ )를

$$\min_{\beta_{\tau_{k+1}}, \hat{b}_{\tau_{k+1}}} \sum_{i=1}^n \rho_{\tau_{k+1}} \left( y_i - \sum_{j=1}^n \beta_{j,\tau_{k+1}} K(\mathbf{x}_i, \mathbf{x}_j) - b_{\tau_{k+1}} \right) + \lambda \sum_{i=1}^n \sum_{j=1}^n \beta_{i,\tau_{k+1}} \beta_{j,\tau_{k+1}} K(\mathbf{x}_i, \mathbf{x}_j) \quad (2.4)$$

subject to

$$\hat{b}_{\tau_k} \leq b_{\tau_{k+1}} \quad (2.5)$$

$$\hat{\beta}_{j,\tau_k} \leq \beta_{j,\tau_{k+1}} \quad \text{for } j = 1, 2, \dots, n \quad (2.6)$$

와 같은 상향 적합식으로 추정하고,  $100\tau_{k-1}\%$  조건부 분위수 함수  $q_{\tau_{k-1}}(\mathbf{x})$ 의 커널계수는

$$\min_{\beta_{\tau_{k-1}}, \hat{b}_{\tau_{k-1}}} \sum_{i=1}^n \rho_{\tau_{k-1}} \left( y_i - \sum_{j=1}^n \beta_{j,\tau_{k-1}} K(\mathbf{x}_i, \mathbf{x}_j) - b_{\tau_{k-1}} \right) + \lambda \sum_{i=1}^n \sum_{j=1}^n \beta_{i,\tau_{k-1}} \beta_{j,\tau_{k-1}} K(\mathbf{x}_i, \mathbf{x}_j) \quad (2.7)$$

subject to

$$b_{\tau_{k-1}} \leq \hat{b}_{\tau_k} \quad (2.8)$$

$$\beta_{j,\tau_{k-1}} \leq \hat{\beta}_{j,\tau_k}, \quad \text{for } j = 1, 2, \dots, n \quad (2.9)$$

와 같은 하향 적합식을 통해 순차적으로 추정한다.

SCN-KQR의 커널계수에 대한 초기값으로 본 논문에서는  $\tau = 0.5$ 에서의 분위수 함수  $q_{0.5}(\mathbf{x})$ 의 커널계수  $\hat{b}_{0.5}$ 와  $\hat{\beta}_{j,0.5}$  ( $j = 1, 2, \dots, n$ )를 이용하였으며, 이들은 KQR의 적합식 (1.3)을 이용하여 추정하였다. SCN-KQR의 상향 및 하향 적합식은 각각 QP로 공식화되며 상향 적합식 (2.4)~(2.6)에 대한 계산 알고리즘은 다음과 같다.

$$\min_{\beta_{\tau_{k+1}}, \hat{b}_{\tau_{k+1}}} \sum_{i=1}^n (\tau_{k+1} u_i + (1 - \tau_{k+1}) v_i) + \lambda \sum_{i=1}^n \sum_{j=1}^n \beta_{i,\tau_{k+1}} \beta_{j,\tau_{k+1}} K(\mathbf{x}_i, \mathbf{x}_j)$$

subject to

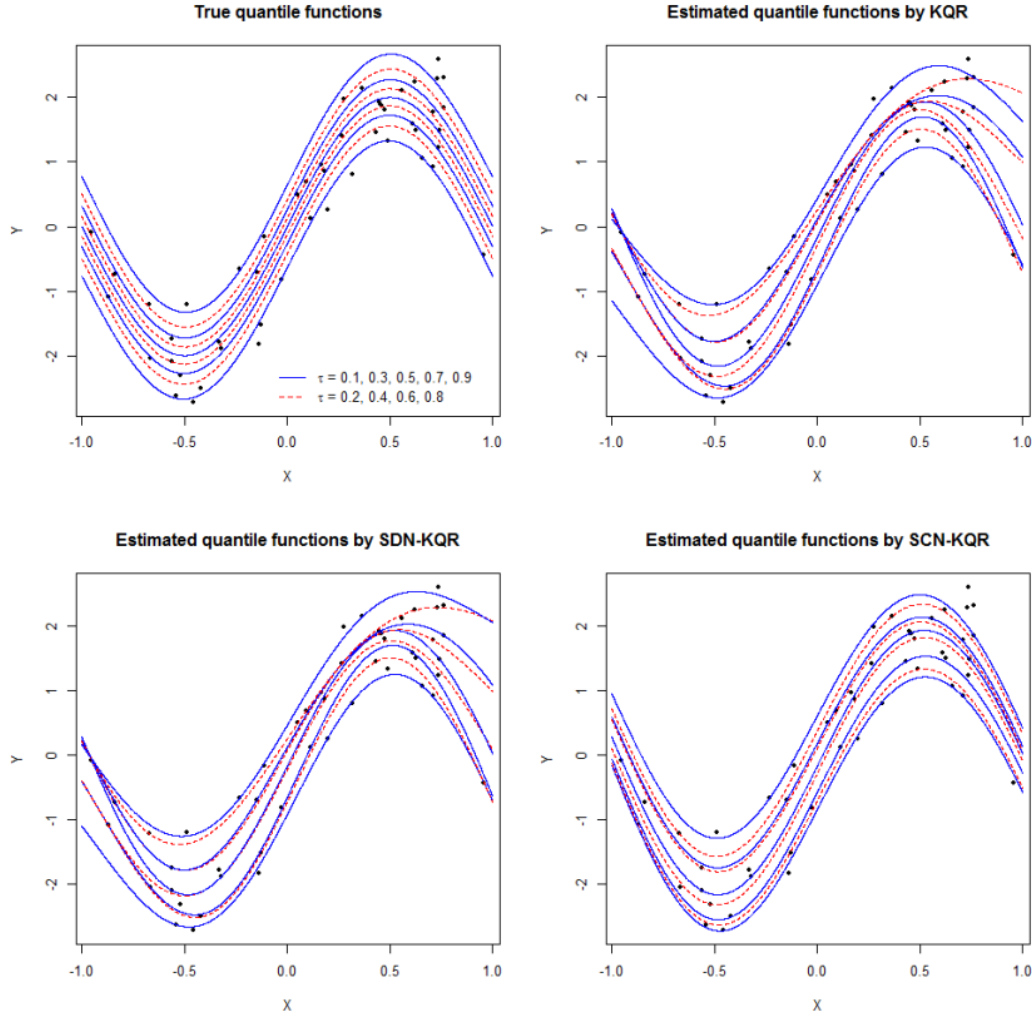
$$\hat{b}_{\tau_k} \leq b_{\tau_{k+1}}$$

$$\hat{\beta}_{j,\tau_k} \leq \beta_{j,\tau_{k+1}}, \quad \text{for } j = 1, 2, \dots, n,$$

$$\sum_{j=1}^n \beta_{j,\tau_{k+1}} K(\mathbf{x}_i, \mathbf{x}_j) - b_{\tau_{k+1}} + u_i - v_i = y_i, \quad \text{for } i = 1, 2, \dots, n,$$

$$u_i \geq 0, v_i \geq 0, \quad \text{for } i = 1, 2, \dots, n.$$

동일한 방법으로 SCN-KQR의 하향 적합식 (2.7)~(2.9)에 대한 계산 알고리즘을 유도할 수 있으며, 본 논문에서는 SCN-KQR의 QP 문제를 실행하기 위해 R 프로그램의 quadprog 패키지에서 제공되는 solve.QP 함수를 사용하였다.



**Figure 3.1.** True quantile functions(top left panel) and estimated quantile functions by KQR(top right panel), SDN-KQR(bottom left panel), and SCN-KQR(bottom right panel)

### 3. 모의실험

비교차 다중 분위수 함수의 추정에서 제안된 SCN-KQR의 성능을 기존의 방법론인 KQR, SDN-KQR과 비교하기 위해 모의실험을 시행하였다. 본 논문에서는 비선형 회귀모형

$$y = 2 \sin(\pi x) + \sigma(x)\epsilon \quad (3.1)$$

을 고려하였으며, 설명변수  $x$ 는 균등분포  $U(-1, 1)$ 를, 오차항  $\epsilon$ 은 표준정규분포  $N(0, 1)$ 를 따르는 것으로 가정하였다. 모형적합을 위해 크기가 50인 훈련자료를 생성하였으며, 조율모수  $\lambda$ 의 선택과 모형 평가를 위해 크기가 1,000인 타당성 자료와 평가 자료를 각각 독립적으로 생성하였다.  $K = 9$ 개의  $\tau_k =$

**Table 3.1.** Mean absolute errors with  $\sigma(x) = 1$ 

Method	Average	$\tau$								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
KQR	0.143 (0.056)	0.174 (0.069)	0.148 (0.064)	0.134 (0.050)	0.126 (0.046)	0.122 (0.045)	0.128 (0.044)	0.131 (0.046)	0.147 (0.050)	0.175 (0.060)
SDN-KQR	0.137 (0.054)	0.169 (0.068)	0.140 (0.053)	0.129 (0.051)	0.122 (0.046)	0.122 (0.045)	0.125 (0.043)	0.126 (0.045)	0.139 (0.048)	0.165 (0.056)
SCN-KQR	0.127 (0.049)	0.146 (0.056)	0.125 (0.048)	0.121 (0.042)	0.116 (0.046)	0.122 (0.045)	0.120 (0.042)	0.122 (0.044)	0.127 (0.046)	0.147 (0.058)

The numbers in parentheses are standard deviations.

**Table 3.2.** Mean absolute errors with  $\sigma(x) = 0.5(1 + 0.2x^2)$ 

Method	Average	$\tau$								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
KQR	0.153 (0.059)	0.187 (0.074)	0.157 (0.065)	0.144 (0.052)	0.136 (0.048)	0.132 (0.045)	0.140 (0.046)	0.140 (0.050)	0.158 (0.053)	0.186 (0.061)
SDN-KQR	0.148 (0.056)	0.181 (0.070)	0.149 (0.055)	0.139 (0.052)	0.132 (0.049)	0.132 (0.045)	0.136 (0.045)	0.136 (0.049)	0.147 (0.051)	0.178 (0.060)
SCN-KQR	0.141 (0.052)	0.166 (0.059)	0.141 (0.049)	0.130 (0.044)	0.126 (0.048)	0.132 (0.045)	0.129 (0.044)	0.134 (0.046)	0.143 (0.051)	0.165 (0.063)

The numbers in parentheses are standard deviations.

$0.1k$  ( $k = 1, 2, \dots, 9$ )에서 조건부 분위수 함수  $q_{\tau_k}(\mathbf{x}_i)$ 를 추정하였으며, 모형평가를 위해 평가 자료를 이용한 평균절대오차(Mean Absolute Error; MAE)

$$\text{MAE}_{\tau_k} = \frac{1}{1,000} \sum_{i=1}^{1,000} |q_{\tau_k}(\mathbf{x}_i) - \hat{q}_{\tau_k}(\mathbf{x}_i)|$$

를 계산하였다. 비선형 함수의 추정을 위해 일반적으로 많이 사용되는 가우스 커널 함수  $K(\mathbf{x}, \mathbf{x}') = \exp(-\|\mathbf{x} - \mathbf{x}'\|^2)$ 를 사용하였으며, 이러한 절차를 100회 독립 반복시행 하였다.

Figure 3.1은  $\sigma(x) = 0.5(1 + 0.2x^2)$ 일 때 비선형 회귀모형 (3.1)의 실제 분위수 함수와 KQR, SDN-KQR, 그리고 SCN-KQR 방법론에 의해 추정된 분위수 함수를 나타내고 있다. 이때, 훈련 자료, 타당성 자료, 평가 자료는 첫 번째 독립 시행에서 생성된 자료를 이용하였다. Figure 3.1에서 KQR에 의해 개별적으로 추정된 분위수 함수들은 서로 교차하는 것을 쉽게 확인할 수 있다. 또한 SDN-KQR의 경우에도 훈련 자료가 희박한 곳에서는 분위수 교차 현상이 발생함을 알 수 있다. 그러나 본 논문에서 제안한 SCN-KQR은 비교한 다중 분위수 함수의 추정에서 강건한 결과를 제공함을 확인할 수 있다. Table 3.1과 Table 3.2는 각각  $\sigma(x) = 1$ 과  $\sigma(x) = 0.5(1 + 0.2x^2)$ 에서 KQR, SDN-KQR, 그리고 SCN-KQR의  $\tau_k$  값에 대한 평균절대오차를 나타내고 있으며, 이로부터 제안하는 SCN-KQR 추정법이 기존의 방법론들에 비해 그 성능이 우수함을 알 수 있다. 여기서 SDN-KQR과 SCN-KQR의 커널계수에 대한 초기값으로  $\tau = 0.5$ 를 사용하였으므로 세 방법론의 평균절대오차가 동일함을 알 수 있다.

#### 4. 결론

분위수 회귀는 반응변수의 조건부 분포에 대한 통계적 분석에서 많은 정보를 제공하는 이점을 바탕으로 의학, 경제, 생존분석 등 여러 다양한 분야에서 널리 이용되고 있다. 그러나 실제 적용분야의 많은 경우

에서 여러 개의 조건부 분위수 함수에 대한 추정을 필요로 하며, 이때 각각의 분위수 함수들을 개별적으로 추정하게 되면 이들이 서로 교차할 가능성이 있다. 이러한 분위수 함수의 교차현상은 추정된 분위수 함수의 해석을 어렵게 할 뿐만 아니라 예측력을 감소시킨다. 따라서 본 논문에서는 다중 비교차 분위수 함수의 추정을 위해 커널 계수에 제약식을 부여하는 순차적 추정법(SCN-KQR)을 제안하였으며, 모의실험을 통해 제안한 SCN-KQR 방법이 기존의 KQR, SDN-KQR 방법에 비해 예측 오차 측면에서 그 성능이 우수함을 확인하였다.

본 논문에서는  $\tau = 0.5$ 에서 커널계수의 추정값을 초기값으로 사용하여 비교차 다중 분위수 함수를 상향 및 하향 적합식을 통해 순차적으로 추정하였다. 차후에는 적절한 초기값  $\tau$ 의 선택 문제와 상·하향식 순차적 추정의 반복을 통한 추정 알고리즘의 연구를 통해 더욱더 강건한 분위수 함수의 추정법이 개발되기를 기대한다.

## References

- Cole, T. and Green, P. (1992). Smoothing reference centile curves: The LMS method and penalized likelihood, *Statistics in Medicine*, **11**, 1305–1319.
- Heagerty, P. and Pepe, M. (1999). Semiparametric estimation of regression quantiles with application to standardizing weight for height and age in U.S. children, *Journal of the Royal Statistical Society: Series C*, **48**, 533–551.
- Hendricks, W. and Koenker, R. (1992). Hierarchical spline models for conditional quantiles and the demand for electricity, *Journal of the American Statistical Association*, **87**, 58–68.
- Kimeldorf, G. and Wahba, G. (1971). Some results on Tchebycheffian spline functions, *Journal of Mathematical Analysis and Applications*, **33**, 82–95.
- Koenker, R. and Bassett, G. (1978). Regression quantiles, *Econometrica*, **4**, 33–50.
- Koenker, R. and Geling, R. (2001). Reappraising Medfly Longevity: A quantile regression survival analysis, *Journal of the American Statistical Association*, **96**, 458–468.
- Koenker, R. and Hallock, K. (2001). Quantile Regression, *Journal of Economic Perspectives*, **15**, 143–156.
- Li, Y., Liu, Y. and Zhu, J. (2007). Quantile regression in reproducing kernel hilbert spaces, *Journal of the American Statistical Association*, **102**, 255–268.
- Liu, Y. and Wu, Y. (2011). Simultaneous multiple non-crossing quantile regression estimation using kernel constraints, *Journal of Nonparametric Statistics*, **23**, 415–437.
- Takeuchi, I., Le, O. V., Sears, T. D. and Smola, A. J. (2006). Nonparametric quantile estimation, *Journal of Machine Research*, **7**, 1231–1264.
- Wang, H. and He, X. (2007). Detecting differential expressions in genechip microarray studies: A quantile approach, *Journal of the American Statistical Association*, **102**, 104–112.
- Wang, H., Li, G. and Jiang, G. (2007). Robust regression shrinkage and consistent variable selection through the LAD-Lasso, *Journal of Business and Economics Statistics*, **25**, 347–355.
- Wu, Y. and Liu, Y. (2009). Stepwise multiple quantile regression estimation using non-crossing constraints, *Statistics and Its Interface*, **2**, 299–310.
- Xu, J. and Ying, Z. (2010). Simultaneous estimation and variable selection in median regression using Lasso-type penalty, *Annals of the Institute of Statistical Mathematics*, **62**, 487–514.
- Yang, S. (1999). Censored median regression using weighted empirical survival and hazard functions, *Journal of the American Statistical Association*, **94**, 137–145.

# 커널 제약식을 이용한 다중 비교차 분위수 함수의 순차적 추정법

방성완<sup>a</sup> · 전명식<sup>b</sup> · 조형준<sup>b,1</sup>

<sup>a</sup>육군사관학교 수학과, <sup>b</sup>고려대학교 통계학과

(2013년 7월 30일 접수, 2013년 10월 25일 수정, 2013년 10월 25일 채택)

---

## 요약

분위수 회귀는 반응변수의 조건부 분위수 함수를 추정함으로써 반응변수와 예측변수의 관계에 대한 포괄적인 정보를 제공한다. 그러나 여러 개의 분위수 함수를 개별적으로 추정하게 되면 이들이 서로 교차할 가능성이 있으며, 이러한 분위수 함수의 교차(quantile crossing) 현상 분위수의 이론적 기본 특성에 위배된다. 본 논문에서는 다중 비교차 분위수 함수의 추정을 위해 커널 계수에 제약식을 부여하는 순차적 추정법을 제안하였으며, 모의실험을 통해 제안한 방법론의 효율적인 성능과 유용성을 확인하였다.

주요어: 커널, 다중 분위수 회귀, 비교차, 이차함수 계획법.

---

---

이 논문은 2013년도 정부(미래창조과학부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것이며(NRF-2013R1A1A1007536)(방성완), 2013년도 정부(교육부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임(NRF-2013R1A1A2A10007545)(전명식)

<sup>1</sup>교신저자: (136-701) 서울시 성북구 안암동 5가, 고려대학교 통계학과, 부교수. E-mail: hj4cho@korea.ac.kr