

소지역 추정을 위한 M-분위수 커널회귀[†]

심주용¹ · 황창하²

¹인제대학교 데이터정보학과 · ²단국대학교 정보통계학과

접수 2012년 6월 17일, 수정 2012년 7월 11일, 게재확정 2012년 7월 16일

요약

소지역 추정을 위해 널리 사용되고 있는 방법 중 하나는 선형혼합효과모형이다. 그러나 종속 변수와 독립변수 사이의 관계가 비선형일 때 이 모형은 소지역 관련 모수에 대해 편의된 추정값을 초래한다. 본 논문에서는 M-분위수 커널회귀를 사용하여 소지역의 평균을 추정하는 방법을 제안한다. 그리고 모의실험을 통하여 서포트벡터분위수회귀와 성능을 비교함으로써 제안된 방법의 우수성을 보인다.

주요용어: 서포트벡터분위수회귀, 소지역 추정, 엠-분위수, 커널회귀.

1. 서론

국내외적으로 지역통계에 관한 관심이 높아지고 있으며 이와 관련하여 소지역 추정에 관한 많은 연구가 활발하게 진행되고 있다. 소지역 추정이란 배정된 표본의 크기가 작은 소지역이나 성별, 연령, 교육수준, 소득수준 등과 같은 변수의 특성으로 분류된 소영역 (small domain)에 대한 통계를 생산하는데 이용되는 추정방법이다. 일반적으로 소지역이란 지리학적 지역, 사회 인구학적 집단 또는 그 외의 다른 부모집단 (subpopulation)으로 정의하며, 소지역의 전형적인 특징은 표본의 크기가 작다는 것이다. 소지역 추정에는 계획에 없던 지역을 추정해야 하는 어려움이 있으며 가장 일반적으로 발생하는 어려움은 해당 소지역에 할당된 표본의 수가 부족하다는 것이다.

미국, 캐나다, 영국 등과 같은 통계 선진국에서는 정부기관과 전문 학자들과의 공동연구를 통해 소지역 추정 관련 통계적 방법론에 대해 상당한 결과들을 축적하여 정부 차원의 통계조사인 인구, 소득, 노동력, 환경, 농업, 금융, 보건통계 등 다양한 분야에 이를 활용하고 있다. 최근 국내에서도 소지역 추정에 관하여 많은 연구가 진행되고 있다. 그러나 대부분의 표본설계는 큰 지역의 특성값을 추정할 목적으로 설계되기 때문에 소지역에 배정되는 표본조사단위수가 적어서 신뢰할 수 있는 통계량의 산출이 어렵다. 즉 기존 표본조사의 자료만으로 소지역에 대한 특성값을 추정하면 할당된 표본의 수가 부족하기 때문에 표집오차가 커지게 된다. 그러나 시간 및 경제적 비용을 고려할 때 모든 소지역을 고려한 표본조사는 현실적으로 불가능하다. 따라서 기존 표본조사의 자료를 가지고 일정수준의 정도 (precision)를 만족하는 소지역의 특성값을 추정할 수 있는 연구가 요구된다.

[†] 이 논문은 2011년도 정부 (교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임 (2011-0021125).

¹ (621-749) 경남 김해시 어방동, 인제대학교 데이터정보학과, 겸임교수.

² 교신저자: (448-701) 경기도 용인시 수지구 죽전동 126번지, 단국대학교 정보통계학과, 교수.

E-mail: chwang@dankook.ac.kr

저명한 학술지에 게재된 많은 논문들이 소지역 추정에 관한 다양한 연구결과를 발표하고 있지만 상당수는 선형모형을 이용하고 있어 비선형 현상을 제대로 파악하지 못하는 경향이 있다. 본 논문에서는 일반적인 분위수 (quantile)보다 더 로버스트 (robust)하다고 알려진 M-분위수 (Breckling과 Chambers, 1988)를 이용하여 커널기법 기반의 M-분위수 커널회귀모형을 개발하여 소지역의 평균 추정에 이용하는 방법을 제안하고자 한다. 소지역 추정에 대한 개념 및 응용은 참고문헌 Opsomer 등 (2008), Salvati 등 (2011), Rao (2003) 등에 설명되어 있다. 또한 커널기법에 대한 개념 및 응용은 참고문헌 Cristianini와 Shawe-Taylor (2000), Hwang (2010, 2011), Hwang과 Shin (2010), Shim과 Hwang (2011), Schölkopf와 Smola (2002), Vapnik (1995) 등에 설명되어 있다.

본 논문은 다음과 같이 구성되어 있다. 2절에서는 소지역 추정에 대해서 설명하고, 3절에서는 M-분위수 커널회귀모형을 이용한 소지역 추정에 대해서 기술한다. 4절에서는 모의실험에 대한 결과를 설명한다. 마지막 5절은 결론을 기술하고 있다.

2. 소지역 추정

소지역 추정에서 자료가 각 소지역내에서 일반적으로 상관되어 (correlated) 있으므로 소지역간의 변동을 설명하기 위해 랜덤효과 (random effect)를 고려하는 혼합효과모형을 많이 사용한다. 혼합효과모형은 고정효과 (fixed effect)와 랜덤효과 모두를 이용하여 종속변수와 입력변수들의 관계를 설명하는 모형이다. 고정효과란 종속변수를 관측할 때 부가적으로 관측되는 입력변수들과 관련된 효과를 의미하고 랜덤효과란 각 소지역 내에서는 고정적이지만 주어진 실험에서는 관측되지 않는 어떤 요인들에 의해 랜덤하게 발생하는 소지역간의 변동효과를 의미한다. 일반적으로 소지역 추정에서 고려하는 기본적인 형태의 모형은 다음과 같은 선형 혼합효과모형 (Harville, 1976; Laird와 Ware, 1982)이다.

$$y_{ij} = b_0 + \beta' x_{ij} + \gamma_i + e_{ij}, \quad j = 1, \dots, N_i, \quad i = 1, \dots, m, \quad (2.1)$$

여기서 m 은 소지역의 개수, N_i 는 i 번째 소지역내의 단위 (unit)의 개수, y_{ij} 는 i 번째 소지역의 j 번째 단위의 특성값 (예컨대 인구수, 실업자수 등), $\beta' x_{ij}$ 는 고정효과항이고 γ_i 는 소지역간의 변동성을 나타내는 랜덤효과항이다. 랜덤효과에 관련된 입력벡터 z_{ij} 가 주어지는 경우 γ_i 는 $\gamma_i = z'_{ij} \gamma_i$ 로 표현될 수도 있다. 그리고 γ_i 와 오차항 e_{ij} 는 서로 독립이고, 평균이 0인 정규분포를 따른다고 가정한다.

실제 소지역 추정에서는 각 소지역에서 N_i 개의 단위가 추출되는 것이 아니고 n_i ($< N_i$)개의 단위가 추출된다. 그러므로 각 소지역에서 추출된 n_i ($< N_i$)개의 단위를 이용하여 나머지 $(N_i - n_i)$ 개의 단위의 특성값을 추정하여야 한다. 이를 이용하여 우리는 다음과 같이 각 소지역의 특성값의 합 또는 평균을 추정할 수 있다.

$$\hat{y}_i = \sum_{j \in S_i} y_{ij} + \sum_{j \in R_i} \hat{y}_{ij}, \quad \hat{y}_i = \frac{1}{N_i} \left(\sum_{j \in S_i} y_{ij} + \sum_{j \in R_i} \hat{y}_{ij} \right), \quad (2.2)$$

여기서 S_i 는 N_i 개의 단위들 중 추출된 단위들의 번호의 집합, R_i 는 N_i 개의 단위들 중 추출되지 않은 단위들의 번호의 집합, \hat{y}_{ij} ($j \in R_i$)은 추출된 n_i 개의 단위들의 자료를 이용하여 구한 추출되지 않은 단위 j ($j \in R_i$)의 특성값의 추정값이다.

혼합효과모형은 특성값의 분포가 소지역 구분과 같은 계층적 구조에 의해 설명된다고 가정하고, 특정 확률분포 (일반적으로 정규분포)를 랜덤효과항에 적용한다. 한편 선형 분위수 회귀모형은 비모수적이고 이상값에 대하여 로버스트하다고 알려져 있다 (Koenker와 Bassett, 1978). Chambers와 Tzavidis (2006)는 혼합효과모형 대신 분위수 회귀모형을 소지역 추정에 적용하는 것이 필요하다고

주장하고, 로버스트 통계학의 M-추정법에 기반을 둔 M-분위수를 이용한 선형 M-분위수 회귀모형을 소지역 추정에 활용하였다. Salvati 등 (2011)은 M-분위수에 평활스플라인회귀를 이용한 비모수 M-분위수 회귀모형을 소지역 추정에 적용하였다. 본 논문에서는 일반적인 분위수보다 더 로버스트하다고 알려진 M-분위수에 커널기법을 적용하여 M-분위수 커널회귀모형을 개발하고, 소지역 추정에 활용하는 방법을 제안한다.

3. 소지역 추정을 위한 M-분위수 커널회귀모형

3.1. M-분위수 커널회귀모형

종속변수 y 의 확률밀도함수 $f(y)$ 가 주어진 경우, y 의 θ 번째 분위수는 다음의 식을 최소화하는 q_θ 의 값으로 정의된다.

$$(1 - \theta) \int_{-\infty}^{q_\theta} |y - q_\theta| f(y) dy + \theta \int_{q_\theta}^{\infty} |y - q_\theta| f(y) dy. \quad (3.1)$$

그리고 y 의 θ 번째 M-분위수는 다음의 식을 최소화하는 m_θ 의 값으로 정의된다.

$$(1 - \theta) \int_{-\infty}^{m_\theta} h\left(\frac{y - m_\theta}{s}\right) f(y) dy + \theta \int_{m_\theta}^{\infty} h\left(\frac{y - m_\theta}{s}\right) f(y) dy \quad (3.2)$$

여기서 $h(r/s)$ 는 Huber Proposal 2 손실함수이며 다음과 같이 정의된다 (Huber, 1981).

$$h(r) = \frac{1}{2} r^2 I(|r| < c) + \left(c|r| - \frac{c^2}{2}\right) I(|r| \geq c), \quad s = \frac{\text{med}\{|r|\}}{0.6745} \quad (3.3)$$

이때 $c > 0$ 는 사전에 고정되는 임계값이다. 그리고 M-분위수 회귀모형은 Huber Proposal 2 손실함수에 기반한 다음의 손실함수의 합을 최소화하는 회귀모형이다.

$$h_\theta(r) = \theta h\left(\frac{r}{s}\right) I(r \geq 0) + (1 - \theta) h\left(\frac{r}{s}\right) I(r < 0). \quad (3.4)$$

Salvati 등 (2011)은 평활스플라인회귀를 활용한 비선형 M-분위수 회귀모형을 제안하였다.

구간 $(0, 1)$ 에 속하는 θ 에 대해 입력벡터 \mathbf{x}_j 가 주어질 때 종속변수 y_j 의 θ 번째 M-분위수 회귀함수를 $m_\theta(\mathbf{x}_j)$ 라고 하면 $m_\theta(\mathbf{x}_j)$ 는 다음과 같이 입력벡터 \mathbf{x}_j 와 비선형으로 연관되어 있다고 가정한다.

$$m_\theta(\mathbf{x}_j) = b_0 + \mathbf{w}' \phi(\mathbf{x}_j), \quad j = 1, \dots, n, \quad (3.5)$$

여기서 b_0 는 절편항, $\phi(\mathbf{x}_j)$ 는 비선형 특징사상함수 (nonlinear feature mapping function), 그리고 \mathbf{w} 는 ϕ 에 대응되는 모수벡터이다. ϕ 에 대해 $\phi(\mathbf{x})' \phi(\mathbf{z}) = K(\mathbf{x}, \mathbf{z})$ 를 만족하는 커널함수 $K(\mathbf{x}, \mathbf{z})$ 가 존재한다 (Mercer, 1909). 일반적으로 커널함수로 다음과 같은 Gaussian 커널을 많이 사용한다.

$$K(\mathbf{x}, \mathbf{z}) = e^{-\frac{\|\mathbf{x} - \mathbf{z}\|^2}{\sigma^2}}, \quad (3.6)$$

여기서 σ 는 커널모수이다.

$$\min \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{j=1}^n h_\theta(y_j - b_0 - \mathbf{w}' \phi(\mathbf{x}_j)) \quad (3.7)$$

식 (3.7)을 서포트벡터기계의 최적화 문제로 표현하면 다음과 같이 된다.

$$\begin{aligned} \min \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{j=1}^n (\theta h(\xi_j/s) + (1-\theta)h(\xi_j^*/s)) \\ \text{subject to} \quad & y_j - b_0 - \mathbf{w}'\phi(\mathbf{x}_j) \leq \xi_j \\ & b_0 + \mathbf{w}'\phi(\mathbf{x}_j) - y_j \leq \xi_j^*, \quad \xi_j, \xi_j^* \geq 0 \end{aligned} \quad (3.8)$$

식 (3.8)의 최적화문제는 라그랑주 배수 (Lagrange multiplier)를 이용하여 다음과 같은 최적화 문제로 표현된다.

$$\begin{aligned} \min \quad & \frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i=1}^n (\alpha_i - \alpha_i^*)y_i + \frac{s^2}{2C\theta} \sum_{i=1}^n \alpha_i^2 + \frac{s^2}{2C(1-\theta)} \sum_{i=1}^n \alpha_i^{*2} \\ \text{subject to} \quad & \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0, 0 \leq \alpha_i \leq \epsilon C\theta/s^2, 0 \leq \alpha_i^* \leq \epsilon C(1-\theta)/s^2 \end{aligned} \quad (3.9)$$

그리고 절편항 b_0 의 추정값은 KKT 조건 (Kuhn과 Tucker, 1951)에 의해 구해진다. 그런데 s 는 $(\alpha_i, \alpha_i^*, b_0)$ 를 추정하기 전까지는 알 수 없기 때문에 우리는 반복적으로 $(\alpha_i, \alpha_i^*, b_0)$ 를 추정하여야 한다. 한편 s 는 각 반복에서 잔차를 이용하여 구해진다.

입력벡터 \mathbf{x}_t 에 대해 θ 번째 M -분위수 커널회귀함수는 다음과 같이 구해진다.

$$m_\theta(\mathbf{x}_t) = \sum_{i=1}^n K(\mathbf{x}_t, \mathbf{x}_i)(\alpha_i - \alpha_i^*) + b_0. \quad (3.10)$$

3.2. M-분위수 커널회귀모형을 이용한 소지역 추정

이제 M-분위수 커널회귀모형을 소지역 추정에 활용하는 방법을 설명한다. i 번째 소지역에서 추출된 표본 $\{y_{ij}, \mathbf{x}_{ij}, j \in S_i\}$ 을 훈련자료로 사용하고, 추출되지 않은 표본 $\{y_{ij}, j \in R_i\}$ 을 검정자료로 사용하여 추출되지 않은 단위의 M-분위수 함수의 추정값을 구할 수 있다. 이때 M-분위수 함수의 추정값이 그 단위의 특성값의 추정값이 된다. M-분위수 커널회귀모형을 이용한 소지역 추정은 다음과 같은 단계로 이루어진다.

- (1) 각 소지역 ($i = 1, \dots, m$)에서, 확률 $\theta \in (0, 1)$ 와 추출된 표본 $\{y_{ij}, \mathbf{x}_{ij}, j \in S_i\}$ 를 이용하여 M -분위수 회귀모형을 실행하여 θ 번째 M -분위수 함수의 추정값 $\hat{y}_{ij}(\theta), j \in S_i$ 을 구하고, y_{ij} 와 가장 가까운 값의 $\hat{y}_{ij}(\theta)$ 가 구해질 때의 확률 θ 를 θ_{ij} 라고 한다.
- (2) $\bar{\theta}_i = \frac{1}{n_i} \sum_{j \in S_i} \theta_{ij}$ 를 구한다.
- (3) 확률 $\bar{\theta}_i$ 와 추출된 표본 $\{y_{ij}, \mathbf{x}_{ij}, j \in S_i\}$ 를 훈련자료로 사용하고, 추출되지 않은 표본 $\{y_{ij}, j \in R_i\}$ 를 검정자료로 사용하여 추출되지 않은 단위의 $\bar{\theta}_i$ 번째 M -분위수 함수의 추정값 $\hat{y}_{ij}, j \in R_i$ 을 구한다.
- (4) 만약 $\bar{y}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} y_{ij}$ 의 추정에 관심이 있다면 추정값은 $\hat{y}_i = \frac{1}{N_i} \left\{ \sum_{i \in S_j} y_{ij} + \sum_{i \in R_j} \hat{y}_{ij} \right\}$ 이 된다.

4. 모의실험

이 절에서는 소지역의 평균 추정에 대한 M-분위수 커널회귀 (MQKR)의 유한 표본 성능을 설명하기 위해 모의실험을 수행한다. Salvati 등 (2011)에서와 비슷한 방법으로 모의실험을 수행한다. 모의실험을 통하여 서포트벡터분위수회귀 (SVQR)와 성능을 비교함으로써 제안된 방법의 우수성을 보이고자 한다. M-분위수 회귀모형 관련 Huber Proposal 2 손실함수의 임계값 c 를 $c = 1.345$ 로 정한다. 이 값은 오차가 정규분포를 따를 때 95% 효율성 (efficiency)을 제공한다 (Huber, 1981).

먼저 소지역의 수를 10으로 가정하고, 각 소지역의 100개 단위의 특성값의 평균은 다음 함수를 이용하여 생성한다.

(1) 선형: $m(x) = 3 + 2(x - 0.5)$, $x \sim U(0, 1)$

(2) 비선형: $m(x) = 2\sin(2\pi x)$, $x \sim U(0, 1)$

그리고 i 번째 소지역의 특성값은 다음과 같은 혼합효과모형을 이용하여 생성한다.

$$y_{ij} = m(x_{ij}) + \gamma_i + e_{ij}, \quad j = 1, \dots, 100, \quad i = 1, \dots, 10,$$

여기서 랜덤효과항 γ_i 와 오차항 e_{ij} 는 각각 정규분포 $N(0, 0.04)$ 와 $N(0, 0.16)$ 을 따르고 서로 독립이라고 가정한다. 각 소지역에서 10개의 단위를 단순임의추출하여 그 특성값들을 훈련자료로 사용하고 나머지 90개의 단위의 특성값을 검정자료로 사용한다. 그리고 총 50번의 모의실험을 수행한다. 소지역의 평균 추정에 대한 MQKR과 SVQR의 성능을 비교하기 위해 각 소집단에 대해서 다음과 같은 네 가지 측도를 계산한다.

$$|Bias| = \frac{1}{50} \sum_{t=1}^{50} |\hat{y}_{it} - \bar{y}_i|, \quad RB\% = \frac{|Bias|}{\bar{y}_i} \times 100,$$

$$SE = \sqrt{\frac{1}{50} \sum_{t=1}^{50} (\hat{y}_{it} - \bar{y}_i)^2}, \quad RMSE\% = \frac{RMSE}{\bar{y}_i} \times 100, \quad j = 1, \dots, 10.$$

Table 4.1과 Table 4.2는 선형 또는 비선형의 경우 각 소지역에 대해서 구한 $|Bias|$, $RB\%$, $RMSE$ 및 $RMSE\%$ 의 평균을 나타낸다. 한편 Figure 4.1과 Figure 4.2는 각 소지역별 $|Bias|$, $RB\%$, $RMSE$ 및 $RMSE\%$ 의 값을 나타낸다. 표와 그림을 통하여 MQKR이 SVQR보다 더 좋은 추정결과를 보여주는 것을 알 수 있다.

Table 4.1 Averages and standard deviations of $|Bias|$, $RB\%$, $RMSE$ and $RMSE\%$ for linear case

	$ Bias $	RB	$RMSE$	$RMSE$
MQKR	0.1400 (0.0612)	4.0647 (2.0873)	0.1724 (0.0746)	5.0052 (2.5542)
SVQR	0.1933 (0.0993)	5.8157 (3.3238)	0.2407 (0.1139)	7.0099 (3.8385)

Table 4.2 Averages and standard deviations of $|Bias|$, $RB\%$, $RMSE$ and $RMSE\%$ for nonlinear case

	$ Bias $	RB	$RMSE$	$RMSE$
MQKR	0.3195 (0.0496)	67.3668 (60.7802)	0.4110 (0.0620)	85.0684 (74.1259)
SVQR	0.4102 (0.0356)	84.0601 (68.7454)	0.4984 (0.0360)	102.2735 (83.9272)

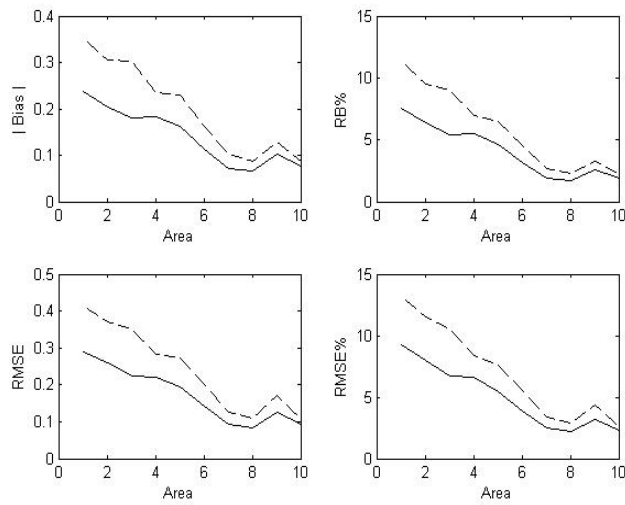


Figure 4.1 Area-specific values of four measures for linear case: MQKR (solid line), SVQR (dashed line)

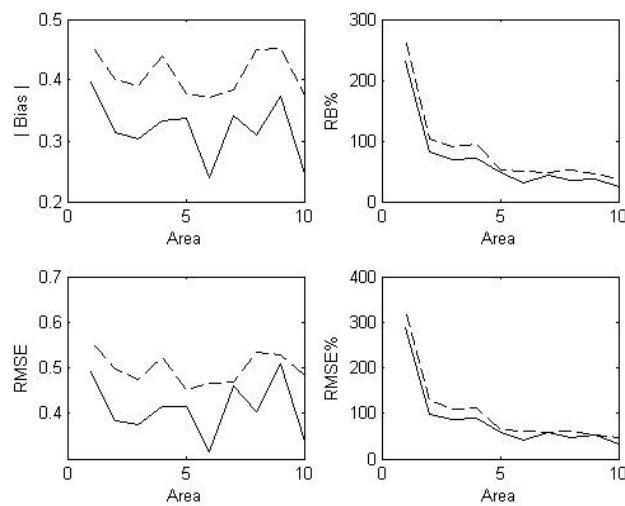


Figure 4.2 Area-specific values of four measures for nonlinear case: MQKR (solid line), SVQR (dashed line)

5. 결론

본 논문에서는 일반적인 분위수보다 더 로버스트하다고 알려진 M-분위수에 커널기법을 적용하여 M-분위수 커널회귀모형을 개발하고, 소지역의 평균 추정에 활용하는 방법을 제안하였다. 기존 연구에서와 비슷한 방법으로 수행된 모의실험을 통하여 제안된 M-분위수 커널회귀모형이 서포트벡터분위수회귀모형 보다 더 좋은 추정결과를 보여주는 것을 알 수 있었다.

참고문헌

- Breckling, J. and Chambers, R. (1988). M-quantiles. *Biometrika*, **75**, 761-771.
- Chambers, R. and Tzavidis, N. (2006) M-quantile models for small area estimation. *Biometrika*, **93**, 255-268.
- Cristianini, N. and Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*, Cambridge University Press.
- Harville, D. A. (1976). Extension of the Gauss-Markov theorem to include the estimation of random effects. *Annals of Statistics*, **4**, 384-395.
- Huber, P. J. (1981). *Robust statistics*, John Wiley & Sons, New York.
- Hwang, C. (2010). Support vector quantile regression for longitudinal data. *Journal of the Korean Data & Information Science Society*, **21**, 309-316.
- Hwang, C. (2011). Asymmetric least squares regression estimation using weighted least squares support vector machine. *Journal of the Korean Data & Information Society*, **22**, 995-1005.
- Hwang, C. and Shin, S. (2010). Estimating GARCH models using kernel machine learning. *Journal of the Korean Data & Information Science Society*, **21**, 419-425.
- Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, **46**, 33-50.
- Kuhn, H. W. and Tucker, A. W. (1951). Nonlinear programming. *Proceedings of 2nd Berkeley Symposium*, University of California Press, Berkeley.
- Laird, N. M. and Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics*, **56**, 89-97.
- Mercer, J. (1909). Functions of positive and negative type and their connection with theory of integral equations. *Philosophical Transactions of Royal Society A*, 415-446.
- Opsomer, J. D., Claesken, G., Ranalli, M. G., Kauermann, G. and Breidt, F. J. (2008). Non-parametric small area estimation using penalized spline regression. *Journal of the Royal Statistical Society B*, **70**, 265-286.
- Rao, J. N. K. (2003). *Small area estimation*, John Wiley & Sons, New York.
- Salvati, N., Ranalli, M. G. and Pratesi, M. (2011). Small area estimation of the mean using non-parametric M-quantile regression: A comparison when a linear mixed model does not hold. *Journal of Statistical Computation and Simulation*, **81**, 945-964.
- Schölkopf, B. and Smola, A. (2002). *Learning with kernels-support vector machines, regularization, optimization and beyond*, MIT Press, Cambridge.
- Shim, J. and Hwang, C. (2011). Forecasting volatility via conditional autoregressive value at risk model based on support vector quantile regression. *Journal of the Korean Data & Information Society*, **22**, 589-596.
- Vapnik, V. N. (1995). *The nature of statistical learning theory*, Springer, New York.

M-quantile kernel regression for small area estimation[†]

Jooyong Shim¹ · Changha Hwang²

¹Department of Data Science, Inje University

²Department of Statistics, Dankook University

Received 17 June 2012, revised 11 July 2012, accepted 16 July 2012

Abstract

An approach widely used for small area estimation is based on linear mixed models. However, when the functional form of the relationship between the response and the input variables is not linear, it may lead to biased estimators of the small area parameters. In this paper we propose M-quantile kernel regression for small area mean estimation allowing nonlinearities in the relationship between the response and the input variables. Numerical studies are presented that show the sample properties of the proposed estimation method.

Keywords: Kernel regression, M-quantile, small area estimation, support vector quantile regression.

[†] This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0021125).

¹ Adjunct professor, Department of Data Science, Institute of Statistical Information, Inje University, Kimhae, Gyeongnam 621-749, Korea.

² Corresponding author: Professor, Department of Statistics, Dankook University, Yongin, Gyeonggido 448-701, Korea. E-mail: chwang@dankook.ac.kr