

평균제곱상대오차에 기반한 비모수적 예측[†]

정석오¹⁾, 신기일²⁾

요약

공변량 값이 주어졌을 때 반응변수의 값을 예측하는 데에는 평균제곱오차를 최소화 하는 것을 고려하는 것이 보통이지만, 최근 Park과 Shin (2005), Jones 등 (2007) 등에서 평균제곱오차 대신 평균제곱상대오차에 기반한 예측을 연구한 바 있다. 이 논문에서는 Jones 등 (2007)의 방법을 대체할 새로운 비모수적 예측법을 제안하고, 제안된 방법의 유효성을 뒷받침하는 간단한 모의실험 결과를 제공한다.

주요용어: 평균제곱오차; 평균제곱상대오차; 예측; 커널평활법.

1. 최소 MSRE 예측

이 논문에서는 공변량 혹은 설명변수 X 의 값이 주어졌을 때 “0”보다 큰 반응변수 Y 의 값을 예측(prediction)하는 문제를 다룬다. 이 문제에서 가장 보편적으로 사용하는 방법은 평균제곱오차(mean squared error, 이하 MSE)를 최소화 하는 방법이다. 잘 알려진 결과로

$$\mu(x) := \arg \min_{\tilde{Y}} E \left[(Y - \tilde{Y})^2 \mid X = x \right] = E(Y \mid X = x)$$

가 되므로 $X = x$ 가 주어졌을 때 Y 의 예측을 위해 $E(Y \mid X = x)$ 을 추정하여 사용하게 된다. 그러나 Park과 Stefanski (1998)은 평균제곱상대오차(mean squared relative error, 이하 MSRE)을 고려하는 경우

$$g(x) := \arg \min_{\tilde{Y}} E \left[\left(\frac{Y - \tilde{Y}}{Y} \right)^2 \mid X = x \right] = \frac{E(Y^{-1} \mid X = x)}{E(Y^{-2} \mid X = x)} \quad (1.1)$$

이 되므로 $g(x)$ 의 추정값 $\hat{g}(x)$ 를 Y 의 예측에 사용할 것을 제안한 바 있다. 이 때 Y 와 Y^{-2} 는 음의 상관관계를 가지기 때문에

$$0 \geq \text{Cov}(Y, Y^{-2} \mid X = x) = E(Y \cdot Y^{-2} \mid X = x) - E(Y \mid X = x)E(Y^{-2} \mid X = x)$$

[†] 이 논문은 2008년도 한국외국어대학교 교내학술연구비의 지원에 의하여 이루어진 것임.

1) (449-791) 경기도 용인시 처인구 모현면 왕산리 산 89, 한국외국어대학교, 조교수.

2) (449-791) 경기도 용인시 처인구 모현면 왕산리 산 89, 한국외국어대학교, 교수.

교신저자: keyshin@hufs.ac.kr

가 성립하므로 부등식

$$\frac{E(Y^{-1}|X=x)}{E(Y^{-2}|X=x)} \leq E(Y|X=x) \quad (1.2)$$

을 얻게 되어 최소 MSRE 예측이 축소(shrinkage) 성질을 가짐을 간단히 확인할 수 있다.

이 축소 성질은 다음과 같은 점근근사를 실시하더라도 그대로 유지된다. 우선 기술적 편의를 위해 주어진 $X=x$ 에 대해

$$(i) \mu(x) := E(Y|X=x) > 0,$$

$$(ii) \mu(x) \rightarrow \infty \text{에 따라 } \left\{ \frac{Y - \mu(x)}{\mu(x)} \right\}^k = o_P(1), \quad (k = 2, 3, \dots),$$

임을 가정하자. 위 (ii)는 자료의 평균값이 큰 경우 그 편차(deviation)가 평균값에 비해 상대적으로 작은 값을 가지는 경향이 있음을 가정하는 것이다. 특히 $k=2$ 인 경우 조건부기대값을 따져보면 $cv(x) := \sigma(x)/\mu(x) \approx 0$ (단, $\sigma^2(x) = \text{Var}(Y|X=x)$)가 성립함을 확인할 수 있는데, 이는 우리가 만나는 대부분의 자료가 평균값이 매우 큰 경우 분산은 평균에 비해 상대적으로 작은 값을 가져 변동계수가 매우 작은 분포적 특성을 가지는 경향이 있다는 사실과 일치한다. 이제 $\ell = 0, 1, 2, \dots$ 에 대해

$$Y^{-\ell} = \mu(x)^{-\ell} - \ell\mu(x)^{-\ell-1} \{Y - \mu(x)\} + \frac{1}{2}\ell(\ell+1)\mu(x)^{-\ell-2} \{Y - \mu(x)\}^2 + \dots$$

임을 이용하면

$$E(Y^{-\ell}|X=x) \stackrel{\mu(x) \rightarrow \infty}{\simeq} \mu(x)^{-\ell} \left\{ 1 + \frac{1}{2}\ell(\ell+1)cv(x)^2 \right\}$$

이 된다. 따라서

$$\frac{E(Y^{-1}|X=x)}{E(Y^{-2}|X=x)} \simeq \mu(x) \cdot \frac{1 + cv(x)^2}{1 + 3cv(x)^2} \quad (1.3)$$

와 같은 근사식을 얻을 수 있다. $\{1 + cv(x)^2\}/\{1 + 3cv(x)^2\} < 1$ 이므로 축소 성질을 쉽게 확인할 수 있다. Park과 Shin (2005)에서도 정상 시계열의 최소 MSRE 예측을 위해 비슷한 작업이 이루어진 바 있음을 밝힌다.

2. 비모수적 평활법에 의한 예측

최근 Jones 등 (2007)은 Park과 Stefanski (1998)을 비모수적 모형으로 확장하여 국소다항커널회귀평활법(local polynomial kernel regression smoothing)을 이용한 최소 MSRE 예측법을 제안하였다. 이를 간단히 요약하면 다음과 같다. $\mathcal{X}_n = \{(X_i, Y_i) | i = 1, \dots, n\}$ 를 (X, Y) 의 랜덤샘플, 0을 중심으로 대칭인 확률밀도함수 K 를 커널(kernel), $h > 0$ 를 평활량(smoothing parameter, bandwidth), $p_m(z) = \sum_{j=0}^m \beta_j z^j$ 라 하자. 식 (1.1)에 의거하여 다음의 목적함수(object function)

$$\sum_{i=1}^n \left\{ \frac{Y_i - p_m(X_i - x)}{Y_i} \right\}^2 K \left(\frac{X_i - x}{h} \right)$$

의 값을 최소로 하는 $(\beta_0, \beta_1, \dots, \beta_m)$ 를 찾아 $(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_m)$ 이라 할 때 $\hat{\beta}_0$ 가 g 의 추정치 $\hat{g}(x)$ 가 되는데 이를 $Y|X = x$ 의 예측에 사용하면 된다. 특히 $m = 0$ 인 경우에는 $\hat{\beta}_0$ 가

$$\hat{g}_0(x) = \frac{\sum_{i=1}^n Y_i^{-1} K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^n Y_i^{-2} K\left(\frac{X_i - x}{h}\right)} \quad (2.1)$$

가 되는데 이는 식 (1.1)의 우변에 있는 분자와 분모 항을 Nadaraya-Watson 타입으로 각각 따로 추정하여 대입한 것과 동일한 형태이다. 또한 $m = 1$ 인 경우에는

$$\hat{g}_1(x) = \frac{t_0(x)s_2(x) - t_1(x)s_1(x)}{s_0(x)s_2(x) - s_1^2(x)} \quad (2.2)$$

가 된다. 단, $j = 0, 1, 2$ 에 대해

$$t_j(x) = n^{-1} \sum_{i=1}^n (X_i - x)^j K\left(\frac{X_i - x}{h}\right) Y_i^{-1},$$

$$s_j(x) = n^{-1} \sum_{i=1}^n (X_i - x)^j K\left(\frac{X_i - x}{h}\right) Y_i^{-2}$$

이다. 대개의 비모수적 함수추정 문제에서 그렇듯이 이들도 역시 평활량 h 의 선택에 따라 추정 결과가 매우 민감하게 변하게 된다. Jones 등 (2007)은 대표본 분석을 통해 $\hat{g}_1(x)$ 을 위한 이론적인 최적의 평활량값이

$$h_{\text{opt}} = \left[\frac{R(K) \int r_2(x) V_g(x) dx}{\kappa_2^2 \int r_2(x) \{g''(x)\}^2 f(x) dx n} \right]^{\frac{1}{5}}$$

가 됨을 보였다. 단 $R(K) = \int K(x)^2 dx$, $\kappa_2 = \int x^2 K(x) dx$, $f(x)$ 는 자료의 X 값의 밀도함수, $V_g(x) = r_2^{-4}(x) \{r_2^3(x) - 2r_1(x)r_2(x)r_3(x) + r_1^2(x)r_4(x)\}$, $r_\ell(x) = E(Y^{-\ell}|X = x)$ 이다. Jones 등 (2007)은 위 식에 포함되어 있는 여러 범함수(functional)들을 어림방법(rule-of-thumb, ROT)으로 추정하여 사용하는 것을 제안하였다.

3. 제안된 축소성질을 이용한 비모수적 예측

앞 절에서 소개한 예측방법은 다음과 같은 문제점이 있다. 첫째, 비모수적 함수 추정 문제에서 가장 민감한 문제인 평활량의 선택법이 용이하지 않다. Jones 등 (2007)은 앞 절 마지막에서 언급한 바와 같이 대표본 성질을 분석한 결과에 근거한 어림방법(rule-of-thumb, ROT)을 제시하고는 있으나 개선의 여지가 많고 특히 pilot 추정 단계에서 만나는 문제들을 해결하는 것이 용이하지 않다. 둘째, 경계점(boundary) 문제가 있다. 보통의 국

소선형추정법(local linear estimation)은 경계점 문제가 거의 없지만, $\hat{g}_1(x)$ 의 경우 경계점 영역에서의 함수 추정 시 분산이 커지는 경향이 있음이 Jones 등 (2007)에서 이미 논의된 바 있다. 이 절에서는 Jones 등 (2007)에서 제안한 예측방법과 성능면에서 거의 차이가 없으면서도 위에서 지적된 문제점들을 간단히 피할 수 있는 방법을 제안하고 그 유효성을 논하고자 한다.

본 논문에서는 식 (1.3)에 근거하여 다음

$$\tilde{g}(x) = \hat{\mu}(x) \cdot \frac{1 + \hat{c}\hat{v}(x)^2}{1 + 3\hat{c}\hat{v}(x)^2} \quad (3.1)$$

의 축소 추정량을 사용하여 예측에 이용하는 것을 제안한다. 단 $\hat{\mu}(x)$ 는 국소다항커널회귀 추정량로서

$$\sum_{i=1}^n \{Y_i - p_m(X_i - x)\}^2 K\left(\frac{X_i - x}{h}\right)$$

를 최소로 하는 $(\beta_0, \dots, \beta_m)$ 의 값을 $(\hat{\beta}_0, \dots, \hat{\beta}_m)$ 이라 할 때 $\hat{\mu}(x) = \hat{\beta}_0$ 로 정의된다. 특히 $m = 0$, $m = 1$ 인 경우에 해당하는 $\hat{\mu}(x)$ 를 각각 $\hat{\mu}_0(x)$ (Nadaraya-Watson), $\hat{\mu}_1(x)$ (local linear)이라 하자. 이 추정량들은 이미 많은 통계계산 프로그램에서 제공되고 있으며 평활량의 선택을 위한 표준적인 방법까지 잘 마련되어 있다 (예: R의 stat 패키지의 ksmooth 함수와 loess 함수, KernSmooth 패키지의 locpoly 함수와 dpill 함수 등). 이에 대한 자세한 논의는 Fan와 Gijbels (2000), Wasserman (2006) 등의 해당 장(chapter)을 참조하기 바란다. 한편 $\hat{c}\hat{v}(x)$ 는 적절한 방법으로 $\sigma(x)$ 를 추정한 후 이를 위의 $\hat{\mu}(x)$ 으로 나누어 얻을 수 있다. 이에 대한 자세한 논의는 잠시 후에 계속하기로 한다.

식 (3.1)의 $\tilde{g}(x)$ 을 얻기 위해서는 변동계수 $cv(x) = \sigma(x)/\mu(x)$ 를 미리 추정해야 하고, 이를 위해 분산함수 $\sigma^2(x)$ (혹은 $\sigma(x)$)의 추정 단계가 요구된다. 우선 자료의 Y 값이 x 에 따라 분산이 변화하는 지 아닌 지에 따라 전략을 구별하여 수립해야 한다. 이분산성이 특별히 의심되지 않을 때에는 Rice (1984), Hall 등 (1991), Spokoiny (2002), Lee 등 (2006) 등에서 제안한 방법을 사용해 $\sigma = \sigma(x)$ 를 추정하면 된다. 만약 이분산성이 의심된다면 Box-Cox 변환, 조기종 등 (2006) 과 같은 적절한 분산안정화변환(variance stabilization transformation)을 통해 이를 해결하는 것이 좋다. 즉 자료의 Y 값에 적절한 분산안정화변환을 실시한 후 얻은 자료를 이용해 등분산 가정 하에서 위에서 설명한 방법대로 식 (3.1)에 따라 축소추정한 후 그 추정치에 대해 분산안정화변환의 역변환(back-transform)을 적용한 값을 최종 추정치로 삼으면 된다. 물론 이분산인 경우 분산함수 $\sigma^2(x)$ 를 직접 추정하는 문제를 다룬 Carroll (1982), Hall과 Carroll (1989), Ruppert 등 (1997), Jeong과 Kang (2006) 등과 이 논문들이 참조한 참고문헌들에서 제공하는 방법을 이용한다면 분산안정화변환 단계를 거칠 필요없이 직접 $cv(x)$ 를 추정하여 사용할 수 있기는 하나 계산량이 심하게 많아지게 되는 단점이 있어 실용성이 떨어지므로 본 논문에서는 고려하지 않는다.

이상에서 논의한 축소추정을 이용한 예측법을 정리하면 다음과 같다.

- 이분산성이 의심되지 않는 경우

- Step 1. 자료 $\mathcal{X}_n = \{(X_i, Y_i), i = 1, 2, \dots, n\}$ 에 국소다항커널회귀추정법을 적용해 $\mu(x) = E(Y|X = x)$ 의 추정치 $\hat{\mu}(x)$ 를 얻는다.
- Step 2. 자료 \mathcal{X}_n 를 이용해 $\hat{\sigma}$ 를 구하여 $\hat{c}\hat{v}(x) = \hat{\sigma}/\hat{\mu}(x)$ 를 얻는다.
- Step 3. $\tilde{g}(x) = \hat{\mu}(x) \cdot \{1 + \hat{c}\hat{v}(x)^2\}\{1 + 3\hat{c}\hat{v}(x)^2\}^{-1}$ 을 이용해 $Y|X = x$ 의 예측에 이용한다.

• 이분산성이 의심되는 경우

- Step 1. 자료 $\mathcal{X}_n = \{(X_i, Y_i), i = 1, 2, \dots, n\}$ 의 Y 값에 분산안정화변환을 적용하여 변환된 자료 $\mathcal{X}'_n = \{(X_i, Y'_i), i = 1, 2, \dots, n\}$ 를 얻는다.
- Step 2. 변환된 자료 \mathcal{X}'_n 에 국소다항커널회귀추정법을 적용해 $\mu'(x) = E(Y'|X = x)$ 의 추정치 $\hat{\mu}'(x)$ 를 얻는다.
- Step 3. 변환된 자료 \mathcal{X}'_n 를 이용해 $\sigma' = \sqrt{\text{Var}(Y'|X = x)}$ 의 추정치 $\hat{\sigma}'$ 를 구하여 $\hat{c}\hat{v}'(x) = \hat{\sigma}'/\hat{\mu}'(x)$ 를 얻는다.
- Step 4. $\tilde{g}'(x) = \hat{\mu}'(x) \cdot \{1 + \hat{c}\hat{v}'(x)^2\}\{1 + 3\hat{c}\hat{v}'(x)^2\}^{-1}$ 을 계산한다.
- Step 5. $\tilde{g}'(x)$ 에 위의 Step 1에서 실시한 분산안정화변환의 역변환을 적용해 얻은 값을 $\tilde{g}(x)$ 로 정의하고 이를 $Y|X = x$ 의 예측에 이용한다.

4. 모의 실험

앞 절에서 제안한 축소 예측법과 Jones 등 (2007)에서 제안한 방법의 성능을 비교 조사하기 위해 다음과 같이 모의 실험을 실시하였다. 비교를 간단히 하기 위해 Jones 등 (2007)에서 실시한 모의실험과 같은 모형 하에서 실험을 실시한다.

- 몬테칼로 반복 수: $M = 1000$.
- 표본의 크기: $n = 50, 100, 500$.
- 모형: $X \sim \text{Unif}(0, 1), \log(Y|X) \sim N(\mu(X), \sigma^2), \sigma = 0.2, 0.5$.
 1. $\mu(x) = 0.5 + 2x$.
 2. $\mu(x) = \log\{0.5 + 40(x - 0.5)^2\}$.
 3. $\mu(x) = 5\{\phi(12(x - 0.35))I(x < 0.5) + \phi(12(x - 0.65))I(x \geq 0.5)\}$.

그림 4.1은 이 모형에 따라 생성된 자료를 산점도로 나타낸 것이다.

- 비교 대상: $\hat{g}_1(x)$ vs. $\tilde{g}_1(x)$.
- 커널 K : $K(x) = (2\pi)^{-1/2} \exp(-x^2/2)$ (Gaussian 커널).
- 평활량 h : 각 모형에 해당하는 이론적 최적평활량을 구하여 사용하였다. 이는 평활량 선택의 효과가 비교 과정에 영향을 주지 않도록 하기 위함이다.

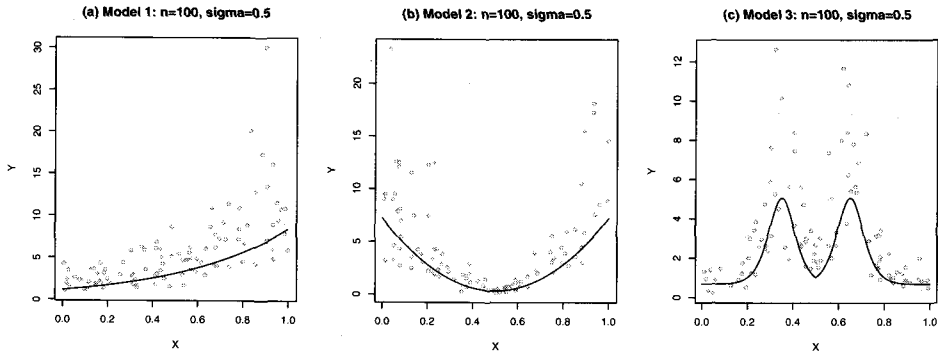


그림 4.1: 모의실험 모형에 따라 생성된 자료의 예 (작은 점들은 생성된 자료를, 실선은 $g(x)$ 을 나타낸 것이다. 단, $n = 100, \sigma = 0.5$.)

- 분산안정화변환: 위의 로그-정규모형에 근거하여 로그 변환을 실시하였다.

\hat{g}_1 를 이용한 예측법과 \tilde{g}_1 을 이용한 예측법의 성능을 비교하기 위해 제곱상대예측오차(squared relative prediction error)의 분포, 즉 $\{(Y - \hat{g}_1(X))/Y\}^2 | X$ 와 $\{(Y - \tilde{g}_1(X))/Y\}^2 | X$ 의 분포를 아래 절차에 따라 비교하였다.

1. 난수발생을 통해 생성된 자료 $(X_1, Y_1), \dots, (X_n, Y_n)$ 를 이용해 주어진 x 점에서의 $\hat{g}_1(x)$ 와 $\tilde{g}_1(x)$ 를 구한다.
2. 주어진 x 에서 새로이 Y 값을 생성하여 제곱상대예측오차 $Y^{-2}\{Y - \hat{g}_1(x)\}^2$ 과 $Y^{-2}\{Y - \tilde{g}_1(x)\}^2$ 의 값을 계산하여 저장한다.
3. 위의 1과 2를 $M = 1000$ 번 반복한다.

그림 4.2-4.4은 위의 절차를 통해 얻은 $M = 1000$ 개의 제곱상대예측오차의 로그값을 이용해 상자그림(boxplot)을 작성한 결과로서, 각각 모형 1, 2, 3에 대해 $n = 100$ 일 때의 결과이다. $n = 50, 500$ 인 경우 $n = 100$ 일 때와 거의 같은 결과를 주기 때문에 본 논문에서는 생략하였다. 각 그림의 위쪽 여섯 개의 패널은 $\sigma = 0.2$ 일 때 각 $x = 0.0, 0.2, 0.4, 0.6, 0.8, 1.0$ 의 여섯 점에서의 로그-제곱상대예측오차(1000개)의 상자그림이고, 아래 쪽의 패널들은 $\sigma = 0.5$ 인 경우에 해당하는 그림이다. 이들 그림을 살펴보면 \hat{g}_1 를 이용하든 \tilde{g}_1 을 이용하든 예측오차 상자의 위치 및 크기가 거의 비슷하여 예측력에는 차이가 없음을 알 수 있다. 다만 앞 절에서 언급한 경계점(boundary) 문제를 비교할 때 본 모의실험의 결과에서는 거의 차이가 발견되지 않았는데 이에 대해 간단히 논의하자면, 경계점 문제가 발생할 가능성이 높음이 이론적으로 증명되어 있는 \hat{g}_1 을 실제 자료 분석에 지속적으로 사용한다면 경계점 문제를 만날 가능성이 늘 잠재되어 있으나 \tilde{g}_1 의 경우 전혀 염려할 필요가 없음을 지적하고자 한다. 지면 상의 이유로 본 논문에 포함시키지 않았지만, \hat{g}_0 과 \tilde{g}_0 을 이용한 예측법의 성능 비교를 위한 모의실험 결과에서도 동일한 결론을 얻었음을 밝힌다.

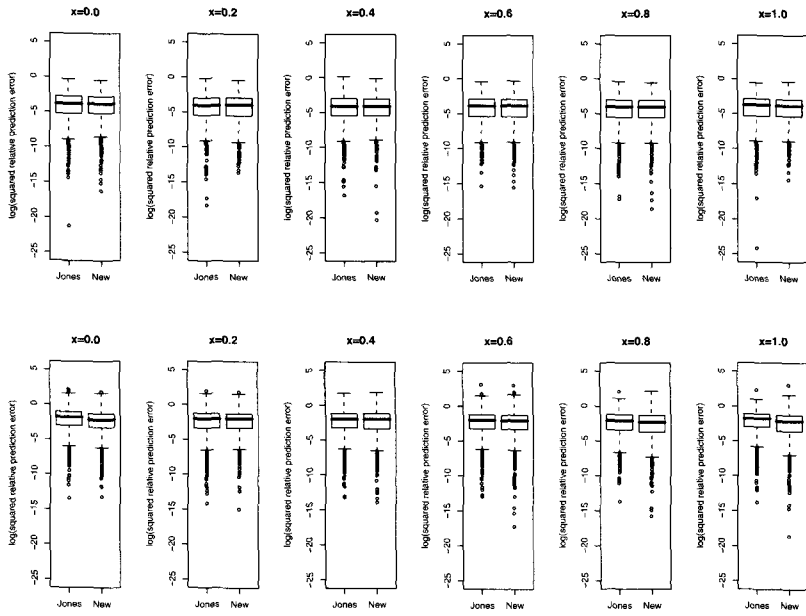


그림 4.2: ($n = 100$) 모형 1의 경우 $x = 0.0, 0.2, 0.4, 0.6, 0.8, 1.0$ 일 때 제공상대오차의 로그값의 상자그림 (위쪽 패널들은 $\sigma = 0.2$, 아래쪽 패널들은 $\sigma = 0.5$ 인 경우에 해당.)

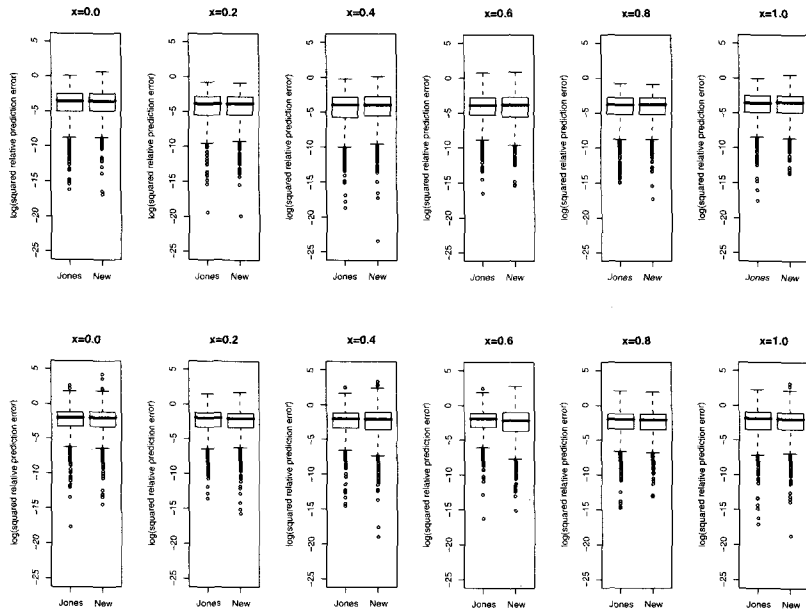


그림 4.3: ($n = 100$) 모형 2의 경우 $x = 0.0, 0.2, 0.4, 0.6, 0.8, 1.0$ 일 때 제공상대오차의 로그값의 상자그림 (위쪽 패널들은 $\sigma = 0.2$, 아래쪽 패널들은 $\sigma = 0.5$ 인 경우에 해당.)

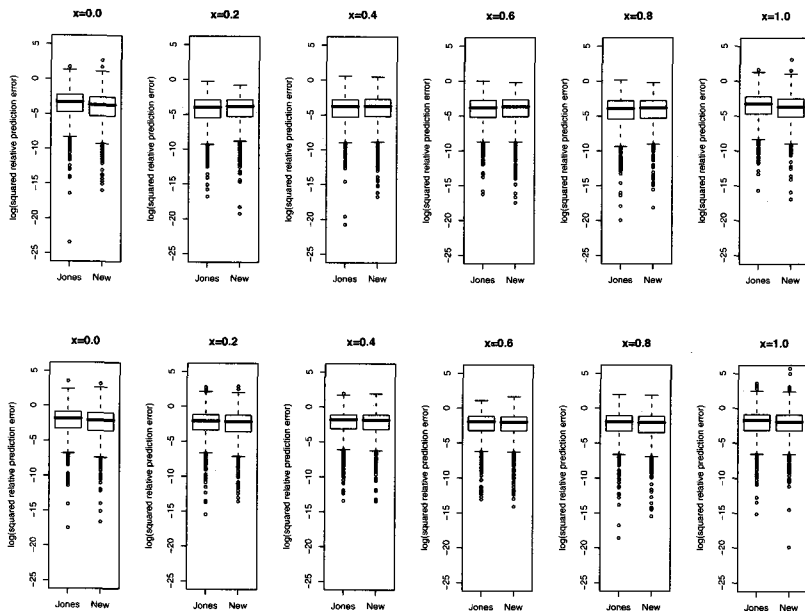


그림 4.4: ($n = 100$) 모형 3의 경우 $x = 0.0, 0.2, 0.4, 0.6, 0.8, 1.0$ 일 때 제공상대오차의 로그값의 상자그림 (위쪽 패널들은 $\sigma = 0.2$, 아래쪽 패널들은 $\sigma = 0.5$ 인 경우에 해당.)

5. 결론

평균제곱오차(MSE)는 모든 통계 자료 분석에서 기본적으로 사용되는 기준이다. 최근 평균제곱오차 뿐 아니라 몇몇 다른 기준, MAE(mean absolute error), MARE(mean absolute relative error), MSRE(mean squared relative error)등이 모형의 우수성 검정을 위하여 사용되고 있다 (Wei, 1990). 이 중에서 MSRE는 MAE, MARE에 비해 분석이 용이하면서도 오차의 상대적 크기를 고려한 기준으로 자체적 의미를 갖고 있다. 본 논문에서는 Jones 등 (2007)이 연구한 국소다항커널회귀 평활량을 이용한 MSRE 예측법의 문제점을 해결하면서도 예측력이 뒤떨어지지 않는 새로운 비모수적 MSRE 예측법을 제안하였으며 모의실험을 통하여 새로운 MSRE 예측법의 우수성을 보였다.

참고문헌

- 조기종, 정석오, 신기일 (2006). 이동-떡변환에 관한 연구. <응용통계연구>, **19**, 283-290.
- Carroll, R. J. (1982). Adapting for heteroscedasticity in linear models. *The Annals of Statistics*, **10**, 1224-1233.

- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall/CRC, New York.
- Hall, P. and Carroll, R. J. (1989). Variance function estimation in regression: The effect of estimating the mean. *Journal of the Royal Statistical Society, Ser. B*, **51**, 3–14.
- Hall, P., Kay, J. W. and Titterton, D. M. (1991). On estimation of noise variance in two-dimensional signal processing. *Advances in Applied Probability*, **23**, 476–495.
- Jeong, S. O. and Kang, K. H. (2006). A difference-based variance function estimator in the multiple regression model, a manuscript.
- Jones, M. C., Park, H., Shin, K. I., Vines, S. K. and Jeong, S. O. (2007). Relative error prediction via kernel regression smoothers. *Journal of Statistical Planning and Inference*, accepted.
- Lee, Y. K., Kim, T. Y. and Park, B. U. (2006). A simple variance estimator in non-parametric regression models with multivariate predictors. *Journal of the Korean Statistical Society*, **35**, 105–114.
- Park, H. and Shin, K. I. (2006). A shrunked forecast in stationary processes favouring percentage error. *Journal of Time Series Analysis*, **27**, 129–139.
- Park, H. and Stefanski, L. A. (1998). Relative-error prediction. *Statistics & Probability Letters*, **40**, 227–236.
- Rice, J. (1984). Bandwidth choice for nonparametric regression. *The Annals of Statistics*, **12**, 1215–1230.
- Ruppert, D., Wand, M. P., Holst, U. and Hössjer, O. (1997). Local polynomial variance function estimation. *Technometrics*, **39**, 262–273.
- Spokoiny, V. (2002). Variance estimation for high-dimensional regression models. *Journal of Multivariate Analysis*, **82**, 111–133.
- Wasserman, L. (2006). *All of Nonparametric Statistics*. Springer, New York.
- Wei, W. W. S. (1990). *Time Series Analysis, Univariate and Multivariate Methods*. Addison Wesley, New York

[2007년 11월 접수, 2008년 1월 채택]

A New Nonparametric Method for Prediction Based on Mean Squared Relative Errors[†]

Seok-Oh Jeong¹⁾, Key-Il Shin²⁾

Abstract

It is common in practice to use mean squared error(MSE) for prediction. Recently, Park and Shin (2005) and Jones *et al.* (2007) studied prediction based on mean squared relative error(MSRE). We proposed a new nonparametric way of prediction based on MSRE substituting Jones *et al.* (2007) and provided a small simulation study which highly supports the proposed method.

Keywords: Mean squared error; mean squared relative error; prediction; kernel smoothing.

[†] This work was supported by Hankuk University of Foreign Studies Research fund of 2008.

1) Assistant professor, Department of Statistics, Hankuk University of Foreign Studies, Kyungki-Do 449-791, Korea.

2) Professor, Department of Statistics, Hankuk University of Foreign Studies, Kyungki-Do 449-791, Korea. Correspondence: keyshin@hufs.ac.kr