

Estimation of smooth monotone frontier function under stochastic frontier model

Danbi Yoon^a · Hohsuk Noh^{a,1}

^aDepartment of Statistics, Sookmyung Women's University

(Received July 17, 2017; Revised August 23, 2017; Accepted August 31, 2017)

Abstract

When measuring productive efficiency, often it is necessary to have knowledge of the production frontier function that shows the maximum possible output of production units as a function of inputs. Canonical parametric forms of the frontier function were initially considered under the framework of stochastic frontier model; however, several additional nonparametric methods have been developed over the last decade. Efforts have been recently made to impose shape constraints such as monotonicity and concavity on the nonparametric estimation of the frontier function; however, most existing methods along that direction suffer from unnecessary non-smooth points of the frontier function. In this paper, we propose methods to estimate the smooth frontier function with monotonicity for stochastic frontier models and investigate the effect of imposing a monotonicity constraint into the estimation of the frontier function and the finite dimensional parameters of the model. Simulation studies suggest that imposing the constraint provide better performance to estimate the frontier function, especially when the sample size is small or moderate. However, no apparent gain was observed concerning the estimation of the parameters of the error distribution regardless of sample size.

Keywords: frontier function, monotonicity constraint, production efficiency, stochastic frontier model

1. 서론

생산성 분석이란 각각의 생산단위에서의 투입과 산출변수 측면에서의 생산결과를 토대로 생산단위의 성과를 측정하는 일련의 분석방법이라고 할 수 있다. 이러한 생산성 분석은 이루어진 투입에 대한 최적의 산출에 대한 실제 달성한 산출의 비율로 이루어지는 경우가 많다. 따라서, 주어진 생산 자료를 기반으로 투입 대비 최대산출량을 나타내는 최대산출량 곡선(production frontier curve)를 추정하는 것이 필요하다. 효율성 측정에 유용한 최대산출량 곡선 흔히 프런티어 함수라고 불리는 함수의 다양한 추정 방법들에 대해서는 현재까지 많은 연구가 진행되어 왔다. 현재까지 개발된 프런티어 함수 추정방법들은 얻어진 생산량이 최대산출량에 도달하지 못한 원인에 대한 가정을 기준으로 크게 결정적 프런티어 모형(deterministic frontier models; DFM)과 확률 프런티어 모형(stochastic frontier models; SFM) 두 가지로 구분할 수 있다.

This research was supported by the 'Basic science research program' through the National Research Foundation of Korea funded by the Ministry of Education (2014R1A1A2059875).

¹Corresponding author: Department of Statistics, Sookmyung Women's University, Cheongpa-ro, 47-gil 100, Yongsan-gu, Seoul 04310, Korea. E-mail: word5810@gmail.com

결정적 프런티어 모형은 비효율성의 원인을 통계 가능한 조직 내 비효율성으로만 보는 모형이다. 이 모형 하에서의 대부분의 프런티어 함수 추정방법들은 프런티어 함수에 대한 모수적 가정없이 자료 전체를 덮는 곡선으로 추정하는 자료포락분석(data envelopment analysis)이라는 아이디어에 기반하고 있는 것이 특징이다. DFM 모형 하에서의 대표적인 추정량들은 프런티어 함수에 대한 기본적인 가정을 만족시킨다는 장점으로 인해 생산성 분석에 많이 사용되고 있긴 하지만, 이상치(outlier)에 민감하고 생산단위의 산출량이 모두 측정오차 없이 측정되었다는 가정이 다소 현실성이 떨어진다는 단점을 가지고 있다. 이러한 단점을 보완하는 분석모형으로 확률 프런티어 모형을 고려할 수 있다.

확률적 프런티어 모형에서는 생산단위에서 관측되는 비효율성이 통제 가능한 조직내 비효율성뿐만 아니라 통제 불가능한 외부 환경 즉 확률오차(random error)에도 기인한다고 가정한다. 하지만 확률 프런티어 분석은 생산량 측정에 오차의 존재를 가정하는 장점을 가지고 있음에 비해 프런티어 함수에 대해 콥-더글라스(Cobb-Douglas), 초월대수(translog)와 같은 특정한 모수적 모형을 가정하고 추정이 진행되는 한계가 존재하였다. 지난 10년 동안 이러한 추정의 한계를 극복하기 위해 프런티어 함수에 대한 모수적 모형을 가정하는 것을 완화하는 연구가 활발히 이루어져 왔다. 그와 연관된 대표적인 업적으로 Fan 등 (1996)은 pseudo-likelihood라는 방법을 사용하여 비효율성과 오차에 대해 각각 절단된 정규(truncated normal) 분포와 정규분포를 가정하되 프런티어 함수는 특정한 모수적 모형을 가정하지 않고 추정하는 준모수적 추정방법을 개발하였다. 이후에 Keshvari와 Kuosmanen (2013)는 비효율성과 오차에 대해 동일한 가정하에서 추정된 프런티어 함수가 단조성(monotonicity)을 만족시키도록 하는 비모수적 추정방법을 제안하였고, Kuosmanen와 Kortelainen (2012)는 단조성과 오목성(concavity)을 동시에 만족시키는 추정량을 제안하였다. 제안된 추정량들은 SFM하에서 프런티어 함수가 만족해야하는 기본적인 가정을 만족시키는 장점을 가지는 추정량이지만 추정량이 계단함수(step function)이나 조각적 선형함수(piecewise linear function)로 주어지는 특성 때문에 추정된 프런티어 함수가 해석이 용이하지 않는 불연속점이나 미분불가능한 점을 여러 곳에 가지는 단점을 가지고 있다. 이러한 단점을 극복하고자 본 논문에서는 SFM하에서 프런티어 함수의 기본 가정인 단조성을 만족하면서도 추정된 함수가 매끄러운 곡선으로 주어지게 되는 방법을 개발하고자 했다. 아울러, 제안된 방법이 기존의 SFM하에서 단조성 제약조건을 고려하지 않은 smooth 프런티어 함수 추정방법인 Fan 등 (1996) 방법과 단조성은 만족하지만 계단함수로 주어지는 Keshvari와 Kuosmanen (2013)의 방법과 비교하여 추정성능면에서 어떤 장점이 있는지 비교하는 시뮬레이션을 진행하였다.

이후 본 논문의 진행은 다음과 같다. 2장에서는 확률 프런티어 모형하에서 프런티어 함수를 비모수적으로 추정하는 Fan 등 (1996)의 방법에 대해 간단히 설명하고 3장에서는 그 방법을 확장하여 단조성을 만족하는 프런티어 함수 추정 방법을 소개한다. 4장에서는 본 논문에서 제안된 단조성을 만족하는 프런티어 함수 추정 방법과 기존의 방법들과의 추정성능을 비교하는 시뮬레이션 결과를 제시하고자 한다.

2. 확률프런티어 모형에서 비모수적 프런티어 함수추정

확률 프런티어 모형은 i 번째 생산단위에서 투입량을 x_i , 산출량을 y_i 라고 할 때

$$y_i = g(x_i) - \epsilon_i = g(x_i) + v_i - u_i \quad (2.1)$$

이다. 여기서 v_i 는 측정오차이며 u_i 는 생산단위의 (기술) 비효율성이다. 다수의 확률 프런티어 분석에서 계산상의 용의성을 위해 v_i 는 평균이 0, 분산이 σ_v^2 인 정규분포로, u_i 는 평균이 0, 분산이 σ_u^2 인 half-normal 분포를 따른다고 가정하고, v_i 와 u_i 는 서로 독립이라 가정한다. 이 때, 비효율성 분포에 대한 가정은 생산성 분석에서 매우 중요한 문제이나 본 논문에서는 가정의 적절성 문제나 적절한 비효율성 분포의 선택에 대해서는 논의하지 않고 주어진 가정하에서 프런티어 함수 추정 문제에만 초점을 두고 논의를

진행하려 한다. 하지만, 비효율성 u_i 의 분포에 대해 일반적인 truncated normal 분포나 gamma 분포와 같은 다른 분포를 고려하는 경우에도 본 논문에서 제안된 방법은 사용가능하며, 다만 전체오차 $\epsilon_i = v_i - u_i$ 의 확률밀도함수의 특성에 기반하여 프런티어와 함수와 전체오차와 연관된 모수(parameter) 추정이 필요하다. 비효율성의 분포로 일반적인 truncated normal 분포나 gamma 분포를 고려할 때의 프런티어 함수와 전체오차의 모수의 추정에 대한 내용은 Fan 등 (1996)의 4절에 자세히 나와 있다.

일반적으로 자주 사용되는 정규분포를 따르는 측정오차와 half-normal 분포를 따르는 생산단위 비효율성을 가정할 때 전체오차 $\epsilon_i = v_i - u_i$ 의 확률밀도함수는 식 (2.2)와 같이 유도할 수 있다.

$$f_\epsilon(t; \sigma_u, \sigma_v) = \frac{2}{\sqrt{\sigma_u^2 + \sigma_v^2}} \phi\left(\frac{t}{\sqrt{\sigma_u^2 + \sigma_v^2}}\right) \Phi\left(-t \frac{\sigma_u/\sigma_v}{\sqrt{\sigma_u^2 + \sigma_v^2}}\right) \quad (2.2)$$

이때, $\sigma^2 = \sigma_u^2 + \sigma_v^2$ 와 $\lambda = \sigma_u/\sigma_v$ 라고 정의하면 전체오차 ϵ_i 의 확률밀도함수는 식 (2.3)과 같이 재매개화될 수 있다.

$$f_\epsilon(t; \sigma, \lambda) = \frac{2}{\sigma} \phi\left(\frac{t}{\sigma}\right) \Phi\left(-\frac{\lambda t}{\sigma}\right). \quad (2.3)$$

만약 전체오차 ϵ 의 관측치 $\epsilon_1, \dots, \epsilon_n$ 이 주어져 있다면 최대우도추정(maximum likelihood estimation) 방법으로 σ 와 λ 를 추정할 수 있지만 그러한 관측치가 주어지지 않음 때문에 ϵ_i 가 y_i , $m(x) = E(Y|X = x)$, σ , λ 로 표현되는 것을 이용하여 σ 와 λ 를 추정한다. 이를 설명하기 위해 먼저 프런티어 함수 $g(x)$ 와 조건부 평균함수 $m(x)$ 의 관계를 살펴보면 (2.4)와 같다.

$$g(x) = m(x) + \frac{\sqrt{2}\sigma\lambda}{\sqrt{\pi(1 + \lambda^2)}} \equiv m(x) + \gamma(\theta), \quad \theta = (\sigma, \lambda)^\top. \quad (2.4)$$

식 (2.4)를 이용하면 우리가 조건부 평균함수 $m(x)$ 를 아는 경우에 ϵ_i 가 $\epsilon_i = y_i - g(x_i) = y_i - m(x_i) - \gamma(\theta)$ 로 표현되는 것을 이용하여 주어진 모수 θ 에 대해 우도함수(likelihood function)의 값을

$$\sum_{i=1}^n \log f_\epsilon(y_i - m(x_i) - \gamma(\theta); \theta) \quad (2.5)$$

와 같이 계산하는 것이 가능하다. 현실에서는 $m(x)$ 의 값을 알지 못하지만 그 추정치 $\hat{m}(x)$ 는 스플라인이나 국소다항식과 같은 비모수 함수추정 방법을 통해서 얻는 것이 가능하므로 식 (2.5)에서 $m(\cdot)$ 을 $\hat{m}(\cdot)$ 으로 대체한 pseudo-likelihood (2.6)을 최대화시키는 θ 를 추정치 $\hat{\theta}$ 로 얻어낸다.

$$\sum_{i=1}^n \log f_\epsilon(y_i - \hat{m}(x_i) - \gamma(\theta); \theta) \quad (2.6)$$

θ 의 추정치 $\hat{\theta}$ 가 얻어지면 $\hat{\theta}$ 로부터 프런티어 함수 $g(x)$ 의 추정량은 식 (2.7)과 같이 주어진다.

$$\hat{g}(x) = \hat{m}(x) + \frac{\sqrt{2}\hat{\sigma}\hat{\lambda}}{\sqrt{\pi(1 + \hat{\lambda}^2)}}. \quad (2.7)$$

3. 단조성 제약조건하에서 확률 프런티어 함수추정

생산성 분석의 많은 경우에서 투입량이 증가하면 최대산출량은 증가한다고 볼 수 있으므로 프런티어 함수가 투입량에 대한 증가함수로 주어진다고 가정할 수 있다. 결정적 프런티어 모형에서는 자유가처분

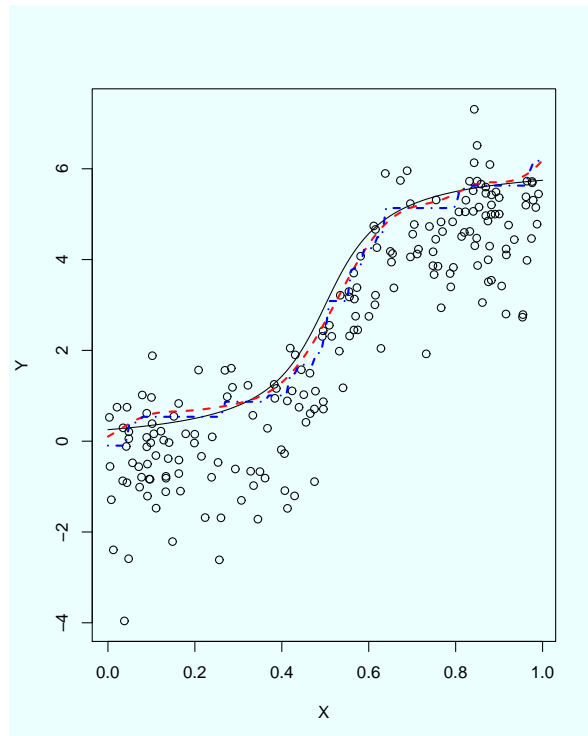


Figure 3.1. An example of the estimated frontier function (the true frontier function - solid line, the estimated function by Keshvari and Kuosmanen (2013) - dash-single-dotted line, and the estimated function by local linear method with monotonicity constraint - dash-double-dotted line).

성(free disposability) 가정에 기반한 FDH 추정량 (Deprins 등, 1984)과 FDH 추정량을 선형화하여 얻어지는 LFDH 추정량 (Jeong과 Simar, 2006) 등이 이러한 단조증가성을 가지는 전통적인 추정량이다. 또한, 최근에 Noh (2014), Daouia 등 (2016)에 의해서 개발된 추정량들 역시 결정적 프런티어 모형 가정하에서 단조증가성을 만족하도록 제안된 추정량이다. 이에 비해 아직 확률적 프런티어 모형에서는 단조증가성을 만족하는 추정량이 활발하게 연구되지 않았다. Keshvari와 Kuosmanen (2013)는 Fan 등 (1996)의 비효율성과 오차 가정하에서 프런티어 함수 $g(x)$ 와 조건부평균 함수 $m(x)$ 가 상수만큼만 차이가 나는 것에 착안하여 Fan 등 (1996)의 추정 방법에서 다른 것은 동일하게 하되 식 (2.6)의 $m(x)$ 의 추정량만 식 (3.1)의 isotonic nonparametric least square 방법에 의해 추정된 단조증가성을 만족하는 추정량($\hat{m}_{\text{inls}}(x_i) = \hat{\alpha}_i$, $\hat{m}_{\text{inls}}(x) = \max_{x_i \leq x} \hat{\alpha}_i$)으로 대체하는 것을 제안하였다.

$$(\hat{\alpha}_1, \dots, \hat{\alpha}_n) = \arg \min_{\alpha_1, \dots, \alpha_n} \sum_{i=1}^n (y_i - \alpha_i)^2, \quad (3.1)$$

subject to $\alpha_i \leq \alpha_j \quad \forall i, j \quad x_i \leq x_j$.

Keshvari와 Kuosmanen (2013)에 의해서 제안된 프런티어 함수 추정량 $\hat{g}_{\text{inls}}(\cdot)$ (Figure 3.1의 1점 쇄선)은 계단함수로 주어지는 조건부 평균함수 추정량 $\hat{m}_{\text{inls}}(\cdot)$ 와 상수차이가 나는 함수로 주어지기 때문에 Figure 3.1에서 알 수 있듯이 x_i 값들의 일부에서 불연속인 함수로 주어진다. 이러한 특성은 추정량의 편이(bias) 문제를 심화시킬 수도 있을 뿐만 아니라 해석상의 불편함이 따르게 된다. 왜냐하면 프런티어 함수의 불연속점은 생산특성의 변화등이 나타나는 점을 의미할 수 있으나 $\hat{g}_{\text{inls}}(\cdot)$ 의 불연속점들은 방법상의 특성으로 인해 나타나는 점이기에 그런 해석을 부여할 수 없기 때문이다. 따라서 추정성능 향

상과 해석상의 편의를 위해 확률프런티어 모형 하에서 단조증가성을 만족하는 매끄러운(smooth) 프런티어 함수 추정을 고려하는 것이 필요하다. 본 논문에서는 조건부 평균함수 $m(x)$ 를 단조성을 만족하도록 하면서 매끄러운 함수로 추정하는 대표적인 두가지 방법을 이용하여 추정하고 추정된 조건부 평균 함수로부터 Fan 등 (1996)의 pseudo-likelihood 방법을 이용하여 프런티어 함수를 추정함으로써(Figure 3.1의 파선) Keshvari와 Kuosmanen (2013)의 추정량이 가지는 단점을 보완하는 추정량을 제안하였다.

3.1. 스플라인 방법을 통한 단조 증가 프런티어 함수추정

스플라인은 미분가능하거나 연속인 함수를 근사하기 위해 사용하는 부드럽게 이어진 여러 구간들에서 정의된 정해진 차수이하의 조각다항식(piecewise polynomial)의 모음이라고 할 수 있다. 조건부 평균 함수 $m(x)$ 가 정의역이 $[a, b]$ 이고, $[a, b]$ 의 내부의 knot가 $a < \xi_1 < \dots < \xi_k < b$ 인 d 차 스플라인 함수로 표현된다고 가정하자. 그러면 스플라인 함수를 표현하는데 흔히 사용하는 B-spline 기저에 대해 식 (3.2)가 성립함을 알 수 있다.

$$m(x) = \sum_{l=1}^L \beta_l B_l(x), \quad L = k + d + 1. \tag{3.2}$$

이 때, $B_l(x)$, $l = 1, \dots, L$ 은 B-spline 기저함수들이다. Schumaker (2007)에 의하면 $m(x)$ 의 스플라인 계수 β_1, \dots, β_L 가

$$\beta_1 \leq \beta_2 \leq \dots \leq \beta_L \tag{3.3}$$

을 만족하면 $m(x)$ 가 단조증가함수가 된다. 따라서, $m(x)$ 를 단조증가성을 가지도록 하면서 추정하려면 식 (3.4)와 같이 조건부 평균함수의 회귀계수를 추정하기 위한 최소제곱방법에서 식 (3.3)을 만족시켜야 할 제약조건으로 고려하면 된다.

$$\begin{aligned} \min_{\beta_1, \dots, \beta_L} \sum_{i=1}^n \left(y_i - \sum_{l=1}^L \beta_l B_l(x_i) \right)^2 \\ \text{subject to } \beta_1 \leq \dots \leq \beta_L. \end{aligned} \tag{3.4}$$

식 (3.4)의 회귀계수 벡터 $\beta = (\beta_1, \dots, \beta_L)^\top$ 를 구하는 문제는 식 (3.5)와 같은 quadratic programming 문제로 보고 해결할 수 있다.

$$\begin{aligned} \min_{\beta} (\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta) \\ \text{subject to } \mathbf{C}\beta \geq \mathbf{0}_{L-1}. \end{aligned} \tag{3.5}$$

이 때, \mathbf{X} 는 B-spline 기저함수를 나타내는 $n \times L$ 행렬이고 \mathbf{C} 는 회귀계수의 제약조건을 나타내기 위한 $(L - 1) \times L$ 행렬로

$$\mathbf{X} = \begin{pmatrix} B_1(x_1) & \dots & B_L(x_1) \\ \vdots & \ddots & \vdots \\ B_1(x_n) & \dots & B_L(x_n) \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} -1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -1 & 1 \end{pmatrix} \tag{3.6}$$

와 같이 주어지며 $\mathbf{0}_{L-1}$ 은 $(L - 1)$ 차원 영벡터이다. 본 논문에서 스플라인 방법에 기반한 단조증가성을

만족하는 조건부 평균함수의 추정량 $\hat{m}_{sp}(x)$ 를 얻기 위해 2차 스플라인을 사용하였으며 식 (3.5)를 풀기 위해 R package ‘quadprog’를 사용하였다 (2차 스플라인의 경우 식 (3.3)은 단조증가하기 위한 필요충분조건이 됨). 최종적으로 스플라인 방법에 기반한 단조증가하는 프런티어 함수의 추정량 $\hat{g}_{sp}(x)$ 는 식 (2.6)에 $\hat{m}(x_i)$ 대신에 $\hat{m}_{sp}(x_i)$ 를 대입하여 얻어지는 $\hat{\theta} = (\hat{\sigma}, \hat{\lambda})^\top$ 로 부터 식 (2.7)을 사용하여 얻어진다. 한편, 프런티어 함수를 스플라인을 이용하여 추정하는 경우 제시된 방법과 같이 두단계에 걸쳐서 프런티어 함수와 전체오차의 모수를 추정하는 대신

$$l(\beta, \theta) = \sum_{i=1}^n \log f_\epsilon \left(y_i - \sum_{l=1}^L \beta_l B_l(x_i); \theta \right)$$

와 같은 로그우도함수를 $\mathbf{C}\beta \geq \mathbf{0}_{L-1}$ 의 단조성 제약조건하에서 최대화함으로써 한 번에 프런티어 함수와 전체오차의 모수를 추정하는 방법을 생각할 수 있다. 하지만, 이러한 방법은 제시된 방법과 달리 추정량을 구하는 과정이 제약조건이 있는 비선형최적화 과정으로 주어져 nloptr과 같은 R package를 이용하여 추정량을 구해야 할 뿐 아니라 추정량이 비선형최적화를 위한 초기치설정 상 상당히 민감하다는 단점을 가지고 있다. 또한, 논문에서 제시하지는 않았지만 Fan 등 (1996)의 스플라인 버전 추정량을 이용하여 적절한 초기치 설정을 하여도 제시된 방법보다 추정성능이 더 나쁘게 나타났다.

3.2. 국소다항식 방법을 통한 단조증가 프런티어 함수추정

국소다항식(local polynomial) 방법은 자료의 각 관측치 (x_i, y_i) 에 똑같은 가중치를 부여하는 것이 아니라 회귀함수를 추정하고자 하는 점에 가까울수록 가중치를 더 많이 부여하고, 멀어질수록 가중치를 적게 부여하여 다항함수를 적합함으로써 주어진 점에서의 회귀 함수를 추정하는 기법이다. Local constant 추정방법은 추정하고자 하는 점 근처에서 추정 하고자 하는 함수를 국소적 상수함수로 보고 추정하는 방법으로 Nadaraya (1965), Watson (1964)에 의해 제안되었고 식 (3.7)과 같이 추정량이 주어진다.

$$\hat{m}_{lc}(x) = \sum_{i=1}^n \frac{K\left(\frac{x_i-x}{h}\right)}{\sum_{j=1}^n K\left(\frac{x_j-x}{h}\right)} y_i \equiv \sum_{i=1}^n W_{lc,i}(x) y_i. \quad (3.7)$$

이 때, $K(\cdot)$ 는 각 관측치에 가중치를 부여하기 위해 사용하는 커널함수이며 h 는 추정하는 점 주위의 자료들에 대해 얼마나 많이 가중치를 부여할 것인지 결정하는 평활량(bandwidth)모수이다. 한편, local linear 추정방법은 추정하고자 하는 점 근처에서 추정하고자 하는 함수를 국소적 선형 함수로 보고 추정하는 방법으로 식 (3.8)과 같이 추정량이 주어진다.

$$\hat{m}_{ll}(x) = \sum_{i=1}^n \frac{(S_2(x) - S_1(x)(x_i - x))K\left(\frac{x_i-x}{h}\right)}{S_2(x)S_0(x) - S_1^2(x)} y_i \equiv \sum_{i=1}^n W_{ll,i}(x) y_i, \quad (3.8)$$

여기서 $S_j(x) = \sum_{i=1}^n (x_i - x)^j K((x_i - x)/h)$ 이다.

Local constant, local linear 추정방법 모두 조건부 평균함수의 추정량이 $\hat{m}(x) = \sum_{i=1}^n W_i(x) y_i$ 형태의 linear smoother로 주어진다. 이것을 Du 등 (2013)에서는 $\hat{m}(x; \mathbf{p}) = \sum_{i=1}^n p_i W_i(x) y_i$ ($\mathbf{p} = (p_1, \dots, p_n)^\top$, $\sum_{i=1}^n p_i = n$)의 특수한 한 형태($p_1 = \dots = p_n = 1$)로 보고 $\mathbf{p} = (p_1, \dots, p_n)^\top$ 를 $\mathbf{p}_{unif} = (1, \dots, 1)^\top$ 로부터 크게 멀어지기 않게 하면서 $\hat{m}(x; \mathbf{p})$ 가 단조증가성을 만족하도록 선택함으로써 단조증가하는 조건부 평균함수를 추정하는 방법을 제안하였다. $\hat{m}(x; \mathbf{p})$ 가 단조증가성을 만족하도록 하게 하기 위해 Du 등 (2013)에서는 관측된 자료의 x 값의 범위(range)에 조밀한 grid인 $\mathcal{M} = \{\min_{1 \leq i \leq n} x_i = x_0^*, x_1^*, \dots, x_{m-1}^*, x_m^* = \max_{1 \leq i \leq n} x_i\}$ 를 고려하여 $\hat{m}(x; \mathbf{p})$ 의 x 에 대한 도함수

가 \mathcal{M} 에서 음이 아닌 값을 가지게 하는 제약조건을 식 (3.9)와 같이 고려하였다.

$$\hat{m}'(x; p_1, \dots, p_n) = \sum_{i=1}^n p_i W_i'(x^*) y_i \geq 0, \quad \forall x^* \in \mathcal{M}. \quad (3.9)$$

이 때, $W_i'(x^*)$ 는 함수 $W_i'(\cdot)$ 의 $x = x^*$ 에서의 도함수 값이며 local constant, local linear 추정량의 $W_i'(x^*)$ 의 자세한 형태는 Racine (2016)에 나와 있는 것을 참조하였다. 최종적으로 국소다항식에 기반한 단조증가하는 조건부회귀함수의 추정량 $\hat{m}_{lp}(x) = \sum_{i=1}^n \hat{p}_i W_i(x) y_i$ 는 quadratic programming 문제 (3.10)의 해로부터 얻어질 수 있다.

$$\hat{\mathbf{p}} = (\hat{p}_1, \dots, \hat{p}_n)^\top = \operatorname{argmin}_{\mathbf{p}} (\mathbf{p} - \mathbf{p}_{unif})^\top (\mathbf{p} - \mathbf{p}_{unif}) \quad (3.10)$$

$$\text{subject to } \begin{pmatrix} W_1'(x_0^*) y_1 & \cdots & W_n'(x_0^*) y_n \\ \vdots & \ddots & \vdots \\ W_1'(x_m^*) y_1 & \cdots & W_n'(x_m^*) y_n \end{pmatrix} \mathbf{p} \geq \mathbf{0}_n \quad (3.11)$$

국소다항식에 기반한 단조증가하는 프런티어 함수의 추정량 $\hat{g}_{lp}(x)$ 는 식 (2.6)에 $\hat{m}(x_i)$ 대신에 $\hat{m}_{lp}(x_i)$ 를 대입하여 얻어지는 $\hat{\theta} = (\hat{\sigma}, \hat{\lambda})^\top$ 로 부터 식 (2.7)을 사용하여 얻어진다.

4. 모의실험

본 논문에서 제안된 단조성을 만족하면서 smooth한 프런티어 함수 추정량이 단조성 제약조건을 고려하지 않은 smooth 프런티어 함수 추정량인 Fan 등 (1996)의 추정량과 단조성은 만족하지만 계단함수로 주어지는 Keshvari와 Kuosmanen (2013)의 추정량과 비교하여 추정의 효율면에서 어떤 장점이 있는지 비교하는 시뮬레이션을 진행하였다. 시뮬레이션 모형은 Martins-Filho와 Yao (2015)에서 고려한 모형 중 두개의 모형을 선택하였다. 두 모형 모두 단조성 제약조건을 모형으로 Model 1은 단조성뿐 아니라 오목성도 만족하는 모형이며 Model 2는 단조성만 만족하는 모형이다.

- Model 1: $y_i = 1 + \log(1 + x_i) + \epsilon_i$
- Model 2: $y_i = 3 + 2 \tan^{-1}(10(x_i - 0.5)) + \epsilon_i$

이때 x_i 의 분포는 0부터 1사이의 값을 갖는 균일분포를 가정한다. $\epsilon_i = u_i - v_i$ 이며 오차 v_i 는 평균이 0, 분산이 σ_v^2 인 정규분포를 가정한다. 비효율성 u_i 는 평균이 0, 분산이 σ_u^2 인 half-normal 분포를 가정한다. (σ_u^2, σ_v^2) 의 값으로는 Martins-Filho와 Yao (2015)에서 고려한 값을 사용하였다.

$$(\sigma_u^2, \sigma_v^2) = (1.379, 0.501), (0.988, 0.642), (0.551, 0.799)$$

이러한 (σ_u^2, σ_v^2) 의 값은 (σ, λ) 의 값으로는

$$(\sigma, \lambda) = (1.371, 1.659), (1.277, 1.241), (1.153, 0.841)$$

에 해당한다. 표본크기(n)는 100, 200, 300을 고려하였고, 각각의 표본크기에서 1,000번을 반복하여 추정량을 계산한 다음 추정량의 성능을 평가하였다.

추정량의 모든 계산은 R 프로그램을 이용하여 하였다. 단조성 제약 조건이 없는 프런티어 함수 추정량 계산을 위해서 스플라인 방법, local constant 방법, local linear 방법의 세 가지 방법을 사용하여 조건부 평균함수를 추정하였다. 본 논문에서 제안한 추정량 계산을 위해 동일하게 세 가지 추정방법을 사용

Table 4.1. Performance comparison result of Model 1 when $(\sigma, \lambda) = (1.371, 1.659)$

		$y_i = 1 + \log(1 + x_i) + \epsilon_i, (\sigma, \lambda) = (1.371, 1.659)$							
		Unconstrained			Constrained				
		sp	lc	ll	sp	lc	ll	kk	
$g(\cdot)$	$n = 100$	IBIAS2	1.091	0.884	0.447	0.680	0.820	0.449	3.294
		IVAR	14.701	8.434	8.631	9.233	8.439	8.512	10.978
		MISE	15.791	9.318	9.078	9.913	9.259	8.962	14.272
	$n = 200$	IBIAS2	0.262	0.450	0.138	0.193	0.403	0.138	1.680
		IVAR	6.904	4.344	4.376	4.744	4.300	4.321	6.275
		MISE	7.166	4.794	4.515	4.937	4.703	4.459	7.955
	$n = 300$	IBIAS2	0.280	0.348	0.157	0.207	0.318	0.156	1.485
		IVAR	5.031	3.135	3.133	3.485	3.098	3.113	4.702
		MISE	5.311	3.484	3.290	3.693	3.416	3.269	6.187
σ	$n = 100$	BIAS	-0.084	-0.039	-0.040	-0.051	-0.038	-0.040	-0.098
		VAR	0.036	0.037	0.037	0.036	0.037	0.037	0.033
		MSE	0.043	0.038	0.038	0.039	0.038	0.038	0.043
	$n = 200$	BIAS	-0.043	-0.024	-0.023	-0.029	-0.023	-0.023	-0.061
		VAR	0.020	0.020	0.020	0.020	0.020	0.020	0.019
		MSE	0.022	0.021	0.021	0.021	0.021	0.021	0.023
	$n = 300$	BIAS	-0.041	-0.025	-0.023	-0.029	-0.025	-0.023	-0.053
		VAR	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		MSE	0.017	0.016	0.016	0.016	0.016	0.016	0.018
λ	$n = 100$	BIAS	0.024	0.109	0.120	0.106	0.112	0.120	0.029
		VAR	0.838	0.970	1.012	1.014	0.990	1.003	0.821
		MSE	0.837	0.981	1.025	1.024	1.002	1.017	0.822
	$n = 200$	BIAS	0.005	0.031	0.035	0.030	0.032	0.035	-0.008
		VAR	0.350	0.351	0.354	0.354	0.350	0.352	0.337
		MSE	0.349	0.351	0.355	0.355	0.351	0.353	0.337
	$n = 300$	BIAS	-0.040	-0.010	-0.005	-0.013	-0.008	-0.004	-0.036
		VAR	0.240	0.246	0.248	0.250	0.245	0.248	0.231
		MSE	0.241	0.245	0.248	0.250	0.245	0.248	0.232

하여 조건부 평균함수를 추정하되 2절에서 설명된 단조성을 만족하는 버전을 사용하였다. 단조성 제약 조건이 없는 상황에서 조건부 평균함수 추정을 위해 스플라인 방법을 사용할 때 knot는 0과 1사이의 등간격의 knot를 사용하되 Hastie와 Tibshirani (1990)에 소개되어 있는 generalized cross validation 방법을 사용하여 internal knot의 개수를 선택하였다. 또한 동일한 상황에서 국소다항식 적합을 위해서는 식 (4.1)의 AIC타입의 measure를 최소화하는 평활량(h)를 선택하였다.

$$AIC(h) = \log \left(\sum_{i=1}^n (y_i - \hat{m}_{lp}(x_i; h))^2 \right) + 2 \left(\sum_{i=1}^n L_{ii}(h) \right) / n. \quad (4.1)$$

이 때, $L_{ii}(h)$ 는 평활량이 h 로 주어져 있을 때 linear smoother $\hat{m}_{lp}(\cdot; h)$ 의 smoothing matrix의 i 번째 대각원소이다. 단조성을 만족하는 조건부 평균함수의 추정량 계산을 위한 internal knot의 개수나 평활량은 단조성 제약조건이 없는 상황을 위해 선택된 것을 동일하게 사용했다. 한편, Keshvari와 Kuoosmanen (2013)의 추정량 계산을 위해 필요한 $\hat{m}_{inls}(\cdot)$ 은 R 프로그램의 기본함수인 isoreg를 이용하여 계산하였다. 마지막으로 pseudo-likelihood에 기반한 (σ, λ) 의 추정을 위한 최적화에는 R package 'nloptr'를 사용하였다.

Table 4.2. Performance comparison result of Model 1 when $(\sigma, \lambda) = (0.988, 0.642)$

		$y_i = 1 + \log(1 + x_i) + \epsilon_i, (\sigma, \lambda) = (0.988, 0.642)$							
		Unconstrained			Constrained				
		sp	lc	ll	sp	lc	ll	kk	
$g(\cdot)$	$n = 100$	IBIAS2	0.523	0.672	0.225	0.399	0.612	0.233	2.656
		IVAR	14.429	8.917	9.066	9.577	8.916	8.987	11.307
		MISE	14.952	9.588	9.291	9.976	9.529	9.220	13.963
	$n = 200$	IBIAS2	0.174	0.421	0.110	0.155	0.375	0.110	1.525
		IVAR	7.504	5.004	5.050	5.418	4.996	4.979	6.867
		MISE	7.678	5.425	5.159	5.573	5.371	5.089	8.392
	$n = 300$	IBIAS2	0.242	0.343	0.166	0.202	0.316	0.165	1.437
		IVAR	5.565	3.812	3.804	4.100	3.783	3.789	5.349
		MISE	5.807	4.155	3.970	4.302	4.099	3.954	6.786
σ	$n = 100$	BIAS	-0.056	-0.020	-0.021	-0.030	-0.019	-0.021	-0.071
		VAR	0.032	0.033	0.033	0.033	0.033	0.033	0.030
		MSE	0.035	0.034	0.034	0.033	0.034	0.034	0.035
	$n = 200$	BIAS	-0.031	-0.016	-0.015	-0.020	-0.015	-0.015	-0.047
		VAR	0.019	0.019	0.019	0.019	0.019	0.019	0.018
		MSE	0.020	0.019	0.019	0.019	0.019	0.019	0.020
	$n = 300$	BIAS	-0.033	-0.020	-0.019	-0.024	-0.020	-0.019	-0.045
		VAR	0.015	0.015	0.015	0.015	0.015	0.015	0.014
		MSE	0.016	0.015	0.015	0.015	0.015	0.015	0.016
λ	$n = 100$	BIAS	0.079	0.113	0.119	0.112	0.115	0.120	0.077
		VAR	0.619	0.686	0.693	0.665	0.691	0.695	0.582
		MSE	0.625	0.698	0.706	0.677	0.703	0.708	0.587
	$n = 200$	BIAS	0.021	0.032	0.035	0.033	0.033	0.035	0.018
		VAR	0.256	0.262	0.264	0.264	0.264	0.262	0.252
		MSE	0.257	0.263	0.265	0.265	0.265	0.263	0.252
	$n = 300$	BIAS	-0.024	-0.008	-0.006	-0.010	-0.007	-0.006	-0.022
		VAR	0.181	0.185	0.188	0.186	0.186	0.188	0.179
		MSE	0.181	0.185	0.188	0.186	0.185	0.188	0.179

각각의 프런티어 함수 추정방법의 성능을 평가하기 위해 유한차원 모수인 σ 와 λ 에 대해서는 bias, variance, MSE를 계산하였고 무한차원 모수의 프런티어 함수 $g(\cdot)$ 에 대해서는 식 (4.2)와 같이 함수의 정의역에 대해서 추정량 $\hat{g}(x)$ 의 bias의 제곱, variance, MSE 함수를 적분한 값(GBIAS2, GIVAR, GMISE)을 계산하였다.

$$\begin{aligned}
 \text{MISE} &= \frac{1}{1000} \sum_{j=1}^{1000} \text{ISE} \left(\hat{g}^{(j)} \right) := \frac{1}{1000} \sum_{j=1}^{1000} \left[\frac{1}{101} \sum_{i=0}^{100} \left(\hat{g}^{(j)}(z_i) - g(z_i) \right)^2 \right] \\
 &= \frac{1}{101} \sum_{i=0}^{100} \left(g(z_i) - \bar{g}(z_i) \right)^2 + \frac{1}{101} \sum_{i=0}^{100} \left[\frac{1}{1000} \sum_{j=1}^{1000} \left(\hat{g}^{(j)}(z_i) - \bar{g}(z_i) \right)^2 \right] \\
 &\equiv \text{GBIAS2} + \text{GIVAR}.
 \end{aligned}
 \tag{4.2}$$

이 때, $\{z_i, i = 0, \dots, 100\}$ ($z_0 = 0, z_{100} = 1$)는 $[0, 1]$ 를 100등분하는 grid이고, $\hat{g}^{(j)}(\cdot)$ 는 j 번째 데이터 샘플로부터 추정된 프런티어 함수의 추정량이며, $\bar{g}(z_i) = 1000^{-1} \sum_{j=1}^{1000} \hat{g}^{(j)}(z_i)$ 이다.

Tables 4.1–4.6는 두 가지 시뮬레이션 모델의 다양한 (σ_u^2, σ_v^2) 시나리오에서의 추정량 성능비교 결과를

Table 4.3. Performance comparison result of Model 1 when $(\sigma, \lambda) = (0.551, 0.799)$

		$y_i = 1 + \log(1 + x_i) + \epsilon_i, (\sigma, \lambda) = (0.551, 0.799)$							
		Unconstrained			Constrained				
		sp	lc	ll	sp	lc	ll	kk	
$g(\cdot)$	$n = 100$	IBIAS2	0.091	0.673	0.202	0.307	0.614	0.212	1.893
		IVAR	13.592	8.088	8.377	8.939	8.096	8.246	10.689
		MISE	13.683	8.761	8.579	9.246	8.710	8.459	12.583
	$n = 200$	IBIAS2	0.105	0.500	0.154	0.190	0.453	0.158	1.233
		IVAR	7.025	4.661	4.689	5.042	4.631	4.650	6.563
		MISE	7.130	5.161	4.843	5.233	5.085	4.808	7.795
	$n = 300$	IBIAS2	0.028	0.270	0.057	0.080	0.245	0.059	1.079
		IVAR	5.166	3.524	3.491	3.810	3.512	3.478	4.989
		MISE	5.194	3.794	3.548	3.890	3.757	3.537	6.068
σ	$n = 100$	BIAS	0.001	0.030	0.028	0.021	0.031	0.028	-0.016
		VAR	0.026	0.026	0.026	0.026	0.026	0.026	0.024
		MSE	0.026	0.027	0.027	0.026	0.027	0.027	0.024
	$n = 200$	BIAS	0.012	0.025	0.025	0.021	0.025	0.025	-0.003
		VAR	0.014	0.015	0.015	0.015	0.015	0.015	0.014
		MSE	0.015	0.015	0.015	0.015	0.015	0.015	0.014
	$n = 300$	BIAS	0.006	0.016	0.017	0.012	0.016	0.017	-0.005
		VAR	0.011	0.011	0.011	0.011	0.011	0.011	0.011
		MSE	0.011	0.011	0.011	0.011	0.011	0.011	0.011
λ	$n = 100$	BIAS	0.215	0.215	0.222	0.220	0.218	0.222	0.207
		VAR	0.417	0.392	0.426	0.425	0.396	0.427	0.388
		MSE	0.462	0.438	0.475	0.473	0.443	0.476	0.430
	$n = 200$	BIAS	0.136	0.138	0.139	0.139	0.139	0.139	0.133
		VAR	0.175	0.179	0.179	0.177	0.179	0.179	0.176
		MSE	0.194	0.198	0.198	0.196	0.198	0.198	0.193
	$n = 300$	BIAS	0.085	0.091	0.093	0.088	0.091	0.093	0.089
		VAR	0.119	0.122	0.123	0.122	0.122	0.123	0.119
		MSE	0.126	0.130	0.132	0.129	0.130	0.131	0.126

보여주고 있다. unconstrained 부분은 Fan 등 (1996)의 추정량을 $m(x)$ 의 스플라인 추정량(sp), local constant 추정량(lc), local linear 추정량(ll)에 기반하여 계산했을 때 추정성능 평가결과를 보여주고 있다. constrained 부분은 본 논문에서 제안하고 하는 추정량을 $m(x)$ 의 단조성을 만족하는 스플라인 추정량, local constant 추정량, local linear 추정량에 기반하여 계산했을 때의 추정성능 평가결과를 보여주고 있다. 아울러, constrained 부분의 kk column은 Keshvari와 Kuosmanen (2013)의 방법에 근거한 추정량의 추정성능과 연관된 값을 보여주고 있다. 시뮬레이션 결과에 따르면 Model의 종류나 (σ_u^2, σ_v^2) 시나리오와 관계없이 본 논문에서 제안한 프런티어 함수($g(x)$)의 추정량이 단조성 제약조건을 고려하지 않은 Fan 등 (1996)의 프런티어 함수 추정량보다 추정성능이 우수한 것으로 나타났으며, 단조성 제약조건을 고려하긴 하지만 추정량이 계단함수로 주어지는 단점이 있는 Keshvari와 Kuosmanen (2013)보다도 추정성능이 우수한 것으로 나타났다. Keshvari와 Kuosmanen (2013)의 프런티어 함수 추정량은 계단함수로 나타난다는 단점 때문에 단조성을 고려하지 않은 Fan 등 (1996)의 프런티어 함수 추정량보다 추정성능 자체는 더 나쁜 것으로 나타났다. 이러한 현상이 발생하는 이유를 다음과 같이 생각해 볼 수 있다. 표본의 크기가 점점 커져가면 단조성 제약조건 없이 얻어진 함수추정량이 자료를 따라가는 특성

Table 4.4. Performance comparison result of Model 2 when $(\sigma, \lambda) = (1.371, 1.659)$

		$y_i = 3 + 2 \tan^{-1}(10(x_i - 0.5)) + \epsilon_i, (\sigma, \lambda) = (1.371, 1.659)$							
		Unconstrained			Constrained				
		sp	lc	ll	sp	lc	ll	kk	
$g(\cdot)$	$n = 100$	IBIAS2	0.228	1.521	1.683	0.645	1.690	2.148	2.252
		IVAR	16.117	10.808	11.568	11.030	10.581	10.695	16.269
		MISE	16.345	12.328	13.251	11.676	12.271	12.843	18.521
	$n = 200$	IBIAS2	0.415	1.150	1.275	0.429	1.103	1.297	3.198
		IVAR	7.966	6.027	6.418	5.912	5.817	5.873	10.144
		MISE	8.381	7.177	7.693	6.341	6.919	7.169	13.342
	$n = 300$	IBIAS2	0.397	0.927	1.052	0.401	0.881	1.064	2.528
		IVAR	5.599	4.396	4.683	4.332	4.191	4.248	7.695
		MISE	5.996	5.322	5.736	4.733	5.072	5.313	10.224
σ	$n = 100$	BIAS	-0.006	0.011	0.005	0.014	0.015	0.009	-0.100
		VAR	0.026	0.026	0.026	0.026	0.026	0.025	0.020
		MSE	0.026	0.026	0.026	0.026	0.026	0.025	0.030
	$n = 200$	BIAS	-0.050	-0.042	-0.046	-0.038	-0.042	-0.044	-0.136
		VAR	0.020	0.020	0.020	0.020	0.020	0.020	0.017
		MSE	0.023	0.022	0.022	0.021	0.021	0.022	0.036
	$n = 300$	BIAS	-0.045	-0.041	-0.044	-0.037	-0.040	-0.041	-0.114
		VAR	0.015	0.015	0.015	0.015	0.015	0.015	0.013
		MSE	0.017	0.017	0.017	0.017	0.017	0.017	0.026
λ	$n = 100$	BIAS	0.210	0.213	0.210	0.218	0.214	0.207	0.174
		VAR	0.422	0.398	0.409	0.421	0.398	0.394	0.317
		MSE	0.466	0.443	0.453	0.468	0.443	0.437	0.347
	$n = 200$	BIAS	-0.010	0.000	-0.007	0.008	-0.001	-0.007	-0.090
		VAR	0.352	0.346	0.346	0.344	0.346	0.345	0.290
		MSE	0.352	0.346	0.345	0.344	0.346	0.344	0.298
	$n = 300$	BIAS	-0.048	-0.042	-0.048	-0.030	-0.038	-0.043	-0.101
		VAR	0.232	0.237	0.235	0.242	0.237	0.234	0.205
		MSE	0.234	0.239	0.237	0.243	0.238	0.235	0.215

으로 인해 제약조건을 만족하는 경우가 많아져 제약조건 부여로 인한 성능차이가 줄어들게 된다. 하지만, 참프런티어 함수가 smooth한 경우에 Keshvari와 Kuosmanen (2013)의 추정량과 같이 계단형으로 주어진 함수추정량은 표본의 크기가 커져도 추정량의 non-smoothness로 인해 발생한 편이가 어느 정도 계속 남아 있는 것이 가능하고 이것이 성능저하의 주원인이 될 수 있다.

하지만, 프런티어 함수외에 SFM 모형의 추정해야 할 다른 모수인 (σ, λ) (또는 (σ_u^2, σ_v^2))에 대해서는 본 논문에서 제안한 방법이 Fan 등 (1996)의 방법에 비해 의미있는 추정성능의 차이를 나타내지 않았다. 이를 통해 SFM 모형과 같은 준모수적 모형(semiparametric model)의 추정의 소표본상황에서 적절한 제약조건을 통해 무한차원 모수(프런티어 함수)의 추정성능을 향상 시키는 방법이 모형내의 다른 유한 차원 모수들(비효율성과 오차의 분산)의 추정성능의 향상을 가지고 오지는 않을 수도 있음을 알 수 있었다. 한편, (σ, λ) 의 추정과 관련하여 본 논문의 추정방법과 Keshvari와 Kuosmanen (2013) 방법의 결과를 비교하면 두 추정방법이 비슷한 성능을 나타내긴 하지만 σ 의 경우는 제안된 방법이 더 우수하며, λ 의 경우는 Keshvari와 Kuosmanen (2013)의 방법이 약간 더 우수한 성능을 나타내는 것을 볼 수 있었다.

Table 4.5. Performance comparison result of Model 2 when $(\sigma, \lambda) = (0.988, 0.642)$

		$y_i = 3 + 2 \tan^{-1}(10(x_i - 0.5)) + \epsilon_i, (\sigma, \lambda) = (0.988, 0.642)$							
		Unconstrained			Constrained				
		sp	lc	ll	sp	lc	ll	kk	
$g(\cdot)$	$n = 100$	IBIAS2	0.780	1.866	2.120	0.860	1.721	2.194	5.133
		IVAR	16.676	11.591	12.222	11.694	11.290	11.367	16.742
		MISE	17.457	13.457	14.342	12.553	13.012	13.561	21.875
	$n = 200$	IBIAS2	0.288	1.053	1.174	0.364	1.013	1.221	2.493
		IVAR	8.552	6.591	7.020	6.628	6.378	6.496	10.573
		MISE	8.840	7.644	8.194	6.991	7.390	7.718	13.066
	$n = 300$	IBIAS2	0.357	0.890	1.001	0.380	0.850	1.023	2.139
		IVAR	6.203	4.951	5.220	4.956	4.768	4.847	8.221
		MISE	6.560	5.842	6.221	5.336	5.618	5.870	10.360
σ	$n = 100$	BIAS	-0.065	-0.044	-0.051	-0.040	-0.040	-0.046	-0.167
		VAR	0.032	0.032	0.032	0.032	0.032	0.031	0.025
		MSE	0.036	0.034	0.035	0.034	0.033	0.033	0.053
	$n = 200$	BIAS	-0.036	-0.030	-0.033	-0.028	-0.029	-0.031	-0.112
		VAR	0.019	0.018	0.019	0.019	0.018	0.018	0.016
		MSE	0.020	0.019	0.020	0.020	0.019	0.019	0.028
	$n = 300$	BIAS	-0.037	-0.033	-0.036	-0.030	-0.032	-0.034	-0.097
		VAR	0.015	0.014	0.014	0.015	0.015	0.014	0.013
		MSE	0.016	0.016	0.016	0.016	0.016	0.016	0.022
λ	$n = 100$	BIAS	0.063	0.079	0.075	0.093	0.084	0.075	0.005
		VAR	0.601	0.586	0.600	0.629	0.582	0.590	0.577
		MSE	0.604	0.591	0.604	0.637	0.588	0.595	0.576
	$n = 200$	BIAS	0.017	0.019	0.016	0.022	0.018	0.017	-0.020
		VAR	0.259	0.255	0.257	0.261	0.254	0.258	0.222
		MSE	0.259	0.255	0.257	0.262	0.254	0.258	0.222
	$n = 300$	BIAS	-0.030	-0.025	-0.031	-0.020	-0.025	-0.029	-0.049
		VAR	0.178	0.178	0.176	0.182	0.178	0.177	0.164
		MSE	0.179	0.178	0.177	0.182	0.178	0.178	0.166

5. 결론 및 추후연구방향

확률 프런티어 모형하에서 프런티어 함수를 비모적으로 추정하는 방법은 연구되었지만, 단조성 제약 조건 하에서 추정하는 방법은 많이 연구되어지지 않았다. 본 논문에서는 Keshvari와 Kuosmanen (2013) 방법에서 얻어지는 단조성을 만족시키는 프런티어 함수 추정량이 계단 함수로 주어지는 것 때문에 가지게 되는 단점을 보완하는 새로운 프런티어 함수 추정량을 제안하였다. 제안된 프런티어 함수 추정량은 소표본 시뮬레이션 연구에서 단조성이 고려되지 않은 Fan 등 (1996)의 추정량과 단조성은 만족하지만 계단함수로 주어지는 Keshvari와 Kuosmanen (2013)의 추정량보다 우수한 추정성능을 나타내었다. 하지만, 이런 추정 성능상의 장점으로 모형내의 다른 추정 모수 (σ, λ) 의 추정에 이익이 발생하지 않았다.

본 연구에서 사용된 방법은 기본적으로 단조성을 만족하는 프런티어 함수 추정뿐만 아니라 단조성과 오목성을 동시에 프런티어 함수 추정으로 확장될 수 있다. 조건부 평균 함수의 제약조건을 부여할 때, 원래 함수뿐만 아니라 한 번 미분한 함수도 증가함수가 되도록 제약조건을 걸면 된다. 아울러, 프런티어 함수 추정 문제에게 투입변수가 여러 개인 상황이 흔한 상황인 것을 고려할 때, 제안된 방법론을 다차원 투입변수에 적용할 수 있는 방법으로 확장하는 추후 연구가 필요하다.

Table 4.6. Performance comparison result of Model 2 when $(\sigma, \lambda) = (0.551, 0.799)$

		$y_i = 3 + 2 \tan^{-1}(10(x_i - 0.5)) + \epsilon_i, (\sigma, \lambda) = (0.551, 0.799)$							
		Unconstrained			Constrained				
		sp	lc	ll	sp	lc	ll	kk	
$g(\cdot)$	$n = 100$	IBIAS2	0.228	1.521	1.683	0.645	1.690	2.148	2.252
		IVAR	16.117	10.808	11.568	11.030	10.581	10.695	16.269
		MISE	16.345	12.328	13.251	11.676	12.271	12.843	18.521
	$n = 200$	IBIAS2	0.144	0.959	1.036	0.288	1.035	1.242	1.242
		IVAR	8.175	6.247	6.624	6.278	6.020	6.116	10.276
		MISE	8.319	7.206	7.661	6.566	7.055	7.358	11.518
	$n = 300$	IBIAS2	0.103	0.653	0.725	0.196	0.691	0.863	1.088
		IVAR	5.784	4.598	4.850	4.571	4.442	4.482	7.861
		MISE	5.887	5.251	5.576	4.767	5.133	5.345	8.949
σ	$n = 100$	BIAS	-0.006	0.011	0.005	0.014	0.015	0.009	-0.100
		VAR	0.026	0.026	0.026	0.026	0.026	0.025	0.020
		MSE	0.026	0.026	0.026	0.026	0.026	0.025	0.030
	$n = 200$	BIAS	0.007	0.012	0.010	0.014	0.013	0.012	-0.061
		VAR	0.014	0.014	0.014	0.014	0.014	0.014	0.013
		MSE	0.014	0.014	0.014	0.015	0.014	0.014	0.016
	$n = 300$	BIAS	0.004	0.005	0.004	0.008	0.007	0.006	-0.049
		VAR	0.011	0.011	0.011	0.011	0.011	0.011	0.010
		MSE	0.011	0.011	0.011	0.011	0.011	0.011	0.012
λ	$n = 100$	BIAS	0.210	0.213	0.210	0.218	0.214	0.207	0.174
		VAR	0.422	0.398	0.409	0.421	0.398	0.394	0.317
		MSE	0.466	0.443	0.453	0.468	0.443	0.437	0.347
	$n = 200$	BIAS	0.135	0.131	0.132	0.134	0.130	0.129	0.119
		VAR	0.178	0.175	0.173	0.178	0.174	0.174	0.161
		MSE	0.196	0.192	0.191	0.196	0.191	0.191	0.175
	$n = 300$	BIAS	0.085	0.084	0.083	0.086	0.084	0.084	0.085
		VAR	0.118	0.116	0.115	0.118	0.117	0.117	0.111
		MSE	0.125	0.123	0.122	0.126	0.124	0.124	0.118

References

Daouia, A., Noh, H., and Park, B. U. (2016). Data envelope fitting with constrained polynomial splines, *Journal of the Royal Statistical Society-Series B*, **78**, 3–30.

Deprins, D., Simar, L., and Tulkens, H. (1984). Measuring labor-efficiency in post offices, *The Performance of Public Enterprises: Concepts and Measurement*, 243–267.

Du, P., Parmeter, C. F., and Racine, J. S. (2013). Nonparametric kernel regression with multiple predictors and multiple shape constraints, *Statistica Sinica*, **23**, 1347–1371.

Fan, Y., Li, Q., and Weersink, A. (1996). Semiparametric estimation of stochastic production frontier models, *Journal of Business & Economic Statistics*, **14**, 460–468.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*, Chapman and Hall/CRC, New York.

Jeong, S. O. and Simar, L. (2006). Linearly interpolated FDH efficiency score for nonconvex frontiers, *Journal of Multivariate Analysis*, **97**, 2141–2161.

Keshvari, A. and Kuosmanen, T. (2013). Stochastic non-convex envelopment of data: applying isotonic regression to frontier estimation, *European Journal of Operational Research*, **231**, 481–491.

Kuosmanen, T. and Kortelainen, M. (2012). Stochastic non-smooth envelopment of data: semi-parametric frontier estimation subject to shape constraints, *Journal of Productivity Analysis*, **38**, 11–28.

- Martins-Filho, C. and Yao, F. (2015). Semiparametric stochastic frontier estimation via profile likelihood, *Econometric Reviews*, **34**, 413–451.
- Nadaraya, E. A. (1965). On nonparametric estimates of density functions and regression curves, *Theory of Probability and Its Applications*, **10**, 186–190.
- Noh, H. (2014). Frontier estimation using kernel smoothing estimators with data transformation, *Journal of the Korean Statistical Society*, **43**, 503–512.
- Racine, J. S. (2016). Local polynomial derivative estimation: analytic or Taylor? In G. González-Rivera, R. C. Hill, T. H. Lee (Eds.), *Essays in Honor of Aman Ullah (Advances in Econometrics)*, (Volume 36, 617–633), Emerald Group Publishing Limited.
- Schumaker, L. L. (2007). *Spline Functions: Basic Theory* (3rd ed.), Cambridge University Press.
- Watson, G. S. (1964). Smooth regression analysis, *The Indian Journal of Statistics. Series A*, **26**, 359–372.

확률프런티어 모형하에서 단조증가하는 매끄러운 프런티어 함수 추정

윤단비^a · 노호석^{a,1}

^a숙명여자대학교 통계학과

(2017년 7월 17일 접수, 2017년 8월 23일 수정, 2017년 8월 31일 채택)

요약

생산성 평가를 위해서는 주어진 생산 자료를 기반으로 투입 대비 최대산출량을 나타내는 최대산출량을 나타내는 생산 프런티어 곡선에 대한 정보가 필요한 경우가 많다. 이러한 프런티어 함수를 확률프런티어 모형하에서 추정하는 경우에 초기에는 프런티어 함수의 특정한 모수적 형태를 가정하는 경우가 많았다. 그러나 최근에는 프런티어 함수를 프런티어 함수가 기본적으로 만족해야 하는 단조성이나 오목성등을 만족하도록 하면서 비모수적 방법으로 추정하는 방법들이 많이 이루어졌다. 하지만, 이러한 방법들에서 얻어지는 추정량들은 프런티어 함수를 조각적 선형함수 또는 계단함수로 추정하는 특징 때문에 추정의 효율이 떨어지거나 프런티어 함수가 해석이 용이하지 않은 불연속점을 가지는 문제를 가지게 된다. 본 논문에서는 이러한 문제를 해결하기 위해 확률프런티어 모형에서 단조증가하는 매끄러운 프런티어 함수 추정법을 제시하고 제안된 추정방법이 기존의 추정방법에 비해서 가지는 추정 효율의 장점을 시뮬레이션을 통해 예시하였다.

주요용어: 프런티어 함수, 단조성 제약조건, 생산효율, 확률프런티어모형

이 논문은 2014년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. 2014 R1A1A2059875).

¹교신저자: (04310) 서울특별시 용산구 청파로47길 100, 숙명여자대학교 통계학과.

E-mail: word5810@gmail.com