

DEA 효율성 점수의 결정요인 분석방법 비교*

김 성 호**

A Comparison of Alternative Approaches to Determinants of DEA Efficiency Scores

Seongho Kim**

■ Abstract ■

Many papers have used a two-stage approach of first calculating DEA efficiency scores and then seeking to correlate these scores with various environmental variables. Most of the studies have not checked whether such a two-stage approach is statistically valid for identifying significant environmental variables. Recently Simar and Wilson (2007) (SW) introduce a sensible data generating process and bootstrap procedure based on truncated regression for the two-stage approach. Banker and Natarajan (2008) (BN) provide a statistical foundation for the two-stage approach comprising a DEA followed by an ordinary least squares or maximum likelihood estimation. Researchers have to identify an approach suitable for their research circumstances in terms of properties, merits, demerits, and robustness to plausible departures from its chosen data generating process. We summarize the foundations and properties of the two-stage procedures suggested by SW and BN. And we discuss merits and demerits of those procedures. Also using Monte Carlo simulation we assess their relative performance under several misspecified settings.

Keyword : Data Envelopment Analysis, Truncated Regression, Stochastic Frontier, Bootstrap Procedure

논문접수일 : 2010년 01월 12일 논문게재확정일 : 2010년 05월 03일
논문수정일(1차 : 2010년 04월 02일, 2차 : 2010년 04월 22일)

* 이 논문은 인하대학교의 지원에 의하여 연구되었음.

** 인하대학교 경영학부

1. 서 론

효율성에 영향을 미치는 요인에 관한 분석은 효율성 연구자 뿐 아니라 경영자나 정책담당자들의 중요한 관심대상이다. Byrnes et al.[10]은 미국의 서부 및 내륙지역에 있는 노천광의 생산효율성에 영향을 미치는 요인을 분석하기 위해 2단계 접근법을 사용하였다. 이 방법의 첫 번째 단계는 자료포락분석(data envelopment analysis : DEA)으로 효율성을 추정하는 것이고 두 번째 단계는 환경변수가 독립변수이고 전단계에서 구한 DEA효율성점수가 종속변수인 회귀모형(이하 효율성모형)을 추정하는 것이다. 이후 효율성 및 효율성모형의 추정을 위한 2단계 접근법(two-stage approach)은 요양소[32], 은행[2], 학교[34] 등 다양한 분야에 적용되어 왔다.¹⁾

DEA효율성점수는 생산단위의 실제효율성(true efficiency)을 나타내는 기술지표(descriptive measure)로 해석할 수도 있고, 생산단위의 실제효율성에 대한 추정치(estimate)로 해석할 수도 있다[30]. 효율성 결정요인 분석에 관한 대부분의 기존연구들은 DEA효율성점수를 생산단위의 실제효율성에 대한 기술지표로 해석하고 효율성모형을 보통최소제곱법(ordinary least squares : OLS)으로 추정하거나 또는 효율성모형을 토빗회귀모형으로 설정한 후 최대우도법(maximum likelihood : ML)으로 추정하였다. 한편 Grosskopf[23]는 DEA효율성점수를 기술지표로 보는 기존연구들이 효율성모형에 관한 통계적 추론의 타당성을 확보할 수 있는 자료생성과정(data generating process)²⁾의 내용을 제시하지 못하고 있다는 한계점을 지적하였다. 예를 들어 효율성을 추정하는데 사용된 변수들과 환경변수가 상관관계를 갖는다면 자료생성과정의 내용에 따라서 효율성모

형의 추정결과가 통계적으로 타당하지 못할 수도 있다. 최근 Simar and Wilson[37]과 Banker and Natarajan[6]은 Grosskopf가 지적한 기존연구의 한계점을 극복하기 위해 DEA효율성점수를 실제효율성에 대한 추정치로 해석하고 통계적 추론의 타당성을 확보할 수 있는 효율성모형의 자료생성과정과 추정방법을 제시하였다.

Simar and Wilson[37]은 DEA효율성점수가 1.0 (100%의 효율성을 가진 것으로 평가된 점수)인 경우는 실제로 존재하는 값에 대한 추정치가 아니고 유한표본 및 DEA 모형의 구조에 의해 만들어진 결과물이며 따라서 DEA효율성점수가 1.0인 평가대상을 분석대상에서 제외하는 절단회귀모형(truncated regression model)이 효율성모형의 추정을 위한 통계모형으로서 타당함을 주장하였다. 이와 더불어 Simar and Wilson[37]은 DEA효율성점수에 포함된 편향(bias), 효율성모형의 오차항간의 계열상관, 그리고 오차항과 공변량의 상관관계 등으로 인한 문제점을 극복할 수 있는 붓스트랩절차(bootstrap procedure)를 제시하였다. 한편 Banker and Natarajan[6]은 종속변수를 DEA효율성점수의 로그변환으로 설정하고 오차항이 비대칭확률오차와 대칭확률오차³⁾로 구성된 확률프론티어모형(stochastic frontier model)으로 효율성모형을 나타내었다. 그리고 회귀계수에 대한 OLS 추정치 및 ML 추정치가 일치성(consistency)을 가지고 있으며 또한 OLS 추정치는 점근정규성(asymptotic normality)을 가지고 있음을 증명하였다.

최근 McDonald[30]는 Banker and Natarajan[6](BN)의 연구결과에 대해서 DEA효율성점수의 결정

1) Simar and Wilson[37]에는 2단계접근법이 활용된 문헌 47편의 목록이 수록되어 있다.
2) 자료생성과정(data generating process)은 표본자료가 생성되는 과정을 설명하는 확률모형(probability model) 및 표본추출과정(sampling process)을 의미한다[38, p.424].

3) 오차항은 효율성에 영향을 주는 요인으로 모형에 포함된 환경변수(예를 들어 농장주변도시의 인구밀도, 경쟁농장의 수 등)를 제외한 나머지 확률적 요인들의 영향을 나타낸다. 확률프론티어모형의 오차항은 비대칭확률오차와 대칭확률오차로 구분된다. 비대칭확률오차는 한 쪽 방향으로의 변화가 효율성에 기여하는 확률적 요인(예를 들어 농장경영자의 지식, 판단력, 자질 등)의 영향을 나타내며 대칭확률오차는 적정 수준으로의 변화가 효율성에 기여하는 확률적 요인(예를 들어 해발고도, 강수량, 기온 등)의 영향을 나타낸다.

요인 분석방법에 통계적 기반을 제공한 매우 중요한 업적으로 평가한 반면 Simar and Wilson[37](SW)이 제시한 접근방법에 대해서는 가정된 자료생성과정 하에서는 타당할 수도 있지만 가정과 다른 자료생성과정에 대해서 강건하지 못함(not robust)을 지적하였다. McDonald의 생각과는 다르게 최근 SW의 붓스트랩절차는 환경[33], 공항[7], 농장[28], 제조업[26], 에너지[22], 신용조합[18] 등의 분야에서 기존에 사용되었던 방법 즉 효율성모형을 OLS로 추정하는 방법 또는 효율성모형을 토빗회귀모형으로 설정한 후 ML로 추정하는 방법의 한계점을 극복하기 위한 대안으로 사용된 바 있다. 이러한 배경에서 본 연구는 SW의 접근방법에 대한 McDonald의 지적을 체계적으로 검토해보고자 한다.

SW 방법과 BN 방법은 각각 서로 다른 가정을 전제로 하고 있으며 각각 서로 다른 특징을 가지고 있다. 따라서 DEA효율성점수의 결정요인을 분석하고자 하는 연구자는 SW 방법의 가정 및 특징이 자신의 연구상황에 적합한지 아니면 BN 방법의 가정과 특징이 더 적합한지를 판단해야 한다. 또한 자료생성과정이 가정과 다를 때 우월한 성과를 나타내는 방법이 연구자에게 바람직할 것이다. SW 방법이 가정과 다른 자료생성과정에서 열등하다면 McDonald의 SW 방법에 대한 비판적 지적을 받아들여야 할 것이다. 본 논문에서는 SW 방법과 BN 방법의 특징 및 장단점을 비교하고 모의실험으로 성과를 비교해보고자 한다. 이러한 비교결과는 DEA효율성점수의 결정요인을 분석하고자 하는 연구자들이 방법선택에 활용할 수 있는 유용한 정보가 될 것으로 기대한다.

본 논문의 구성은 다음과 같다. 제 2장에서는 SW 방법과 BN 방법의 기본틀이 되는 효율성모형과 DEA효율성점수에 대해서 설명한다. 제 3장에서는 DEA효율성점수의 결정요인을 분석하기 위한 SW 방법과 BN 방법을 구체적으로 설명하고 각 방법의 특징 및 장단점에 관한 비교결과를 제시한다. 제 4장에서는 모의실험을 사용하여 SW 방법과 BN 방법의 유향표본상의 성과를 비교하고 그 결과를 제시한다. 마지막으로 제 4장에서 결론을 제시한다.

2. 효율성모형과 DEA효율성점수

2.1 생산기술 및 효율성모형

생산기술(production technology)은 기술적으로 실행가능한 투입산출벡터(x, y)의 집합을 의미하는 생산가능집합(production possibility set)

$$T = \{(x, y) \mid x \text{로 } y \text{를 생산할 수 있음}\}$$

으로 나타낼 수 있다. 여기서 x 는 투입벡터이고 y 는 산출벡터이다. Farrell[17]의 기술효율성(technical efficiency)은

$$\delta = \max\{\mu \mid (x, \mu y) \in T\}$$

으로 정의되며 δ 는 생산단위가 현재의 투입수준을 유지하면서 효율적 상태에 도달하기 위해 산출을 비례적으로 몇 배 증가시켜야 하는가를 나타낸다.⁴⁾ Farrell기술효율성의 역수 $\theta (=1/\delta)$ 는 Shephard[36]가 정의한 거리함수(distance function)와 같다. 생산프론티어(production frontier) ∂T 는 T 에 속한 생산단위 중에서 효율적 상태에 있는 생산단위들로 구성된 집합이다. 생산단위 k 가 ∂T 에 속해 있으면 $\delta=1(\theta=1)$ 이고 ∂T 에서 벗어나 T 의 내부에 위치에 있으면 $\delta>1(0<\theta<1)$ 이다. 투입 x 에 대응되는 산출 y 의 효율적 수준 $y^\theta(x)$ 는 $y^\theta(x)=\delta y$ 또는 $y^\theta(x)=\theta^{-1}y$ 으로 나타낼 수 있으며 따라서 산출 y 는 $y=y^\theta(x)\theta$ 또는 $y=y^\theta(x)\delta^{-1}$ 으로 나타낼 수 있다.

단일산출을 가정하면 확률오차 및 환경변수의 영향을 반영한 생산기술과 효율성모형을 편리하게 설명할 수 있다.⁵⁾ 단일산출의 경우 산출 y 의 효율적

4) 경영자 또는 정책담당자가 투입요소의 사용량을 절약하기 위한 노력 보다는 산출량을 증가시키기 위한 노력에 더 관심을 가지고 있는 것으로 가정하여 산출방향을 선택하였다.

5) 다수산출로의 확장은 단일산출을 전제로 한 논의내용의 변형없이 Banker and Natarajan[6]의 각주 3 (p.58)에서 설명한 방식으로 직접적으로 이루어질 수 있다.

수준 $y^0(x)$ 는 전통적인 생산함수(production function)이며 $\theta (=1/\delta)$ 는 생산단위가 생산함수로부터 벗어나있는 정도를 나타내는 효율성점수이다. 확률 오차 및 환경변수의 영향을 반영한 생산기술과 효율성모형은

$$y = y^0(x)\theta, \theta = g[h(z), \nu, \omega] \text{ 또는}$$

$$y = y^0(x)\delta^{-1}, \delta = g[h(z), \nu, \omega] \quad (1)$$

으로 나타낼 수 있다. 여기서 z 는 환경변수벡터, $h(z)$ 는 환경변수의 영향을 나타내는 함수, ν 는 한 쪽 방향으로의 변화가 효율성에 기여하는 비대칭확률오차, ω 는 적정 수준으로의 변화가 효율성에 기여하는 대칭확률 오차를 나타낸다. 함수 $g[h(z), \nu, \omega]$ 는 환경변수의 영향 $h(z)$, 비대칭확률오차 ν , 대칭확률오차 ω 등과 효율성점수 θ (또는 δ)와의 관계를 나타내는 함수이다.

2.2 DEA효율성점수

실증연구에서는 관측된 표본자료 $\{(x_k, y_k) | k=1, 2, \dots, n\}$ 로부터 효율성점수 δ_k 의 값을 구해야 한다. 여기서 x_k 와 y_k 는 k 번째 생산단위의 투입벡터와 산출벡터이고 n 은 표본을 구성하는 생산단위의 수이다. 자료포락분석(data envelopment analysis: DEA)은 다수투입·다수산출 상황에서 Farrell기술 효율성점수를 표본자료로부터 비모수적으로 구하기 위해 Charnes et al.[14]가 처음 제시한 수리계획법 모형이다. 이후 DEA 모형은 다양한 일반화 모형으로 발전하였고 여기서는 가장 많이 활용되는 Banker et al.[5]의 DEA 모형을 사용한다. 표본이 n 개의 생산단위로 구성되어 있을 때 k 번째 생산단위의 DEA 효율성점수 $\hat{\delta}_k$ 는 다음과 같은 선형계획법모형으로 나타낼 수 있다.

$$\hat{\delta}_k = \max_{\mu, \lambda} \{ \mu | x_k \geq X\lambda, \mu y_k \leq Y\lambda, i' \lambda = 1, \lambda \in \mathbb{R}_+^n \} \quad (2)$$

여기서 $X = [x_1, \dots, x_n]$, $Y = [y_1, \dots, y_n]$

i 는 요소가 모두 1이고 크기가 $(n \times 1)$ 인 벡터

λ 는 크기 $(n \times 1)$ 의 중요도변수벡터

표본으로부터 식 (2)의 선형계획법모형으로 구한 k 번째 생산단위의 효율성점수 $\hat{\delta}_k$ 는 실제효율성 δ_k 에 대한 기술지표(descriptive measure)로 해석할 수도 있고, 실제효율성에 대한 추정치(estimate)로 해석할 수도 있다. 후자의 관점에서 Banker[4], Kneip et al.[25] 등은 식 (2)에 나타낸 DEA추정량이 유한 표본에서는 편의(bias)를 갖지만 표본크기가 커질수록 편의가 사라지는 일치성(consistency)을 가지고 있음을 증명하였다.

3. DEA효율성점수의 결정요인 분석 방법

3.1 Simar and Wilson의 접근방법

Simar and Wilson[37](SW)은 Farrell 기술효율성 δ_k 를 효율성의 지표로 사용하였다. SW는 식 (1)에서 생산함수 $y^0(x)$ 가 정의하는 생산가능집합을 공짜없음(no free lunches), 자유처분성(free disposability), 볼록성(convexity) 등을 만족하는 폐집합(closed set)으로 가정하였다. 그리고 식 (1)의 효율성모형을 $h(z)$ 은 선형함수로 가정하고 $g[h(z), \nu, \omega]$ 는 (ν 가 생략된) $g[h(z), \nu, \omega] = h(z) + \omega$ 로 가정하여

$$\delta_k = z_k \beta + \omega_k, k=1, 2, \dots, n \quad (3)$$

여기서 n 은 표본에 포함된 생산단위의 수

ω_k 는 오차항으로 평균 0, 표준편차 σ_ω 의 정규분포이고 $1 - z_k \beta$ 에서 좌측으로 절단된 비대칭확률변수

z_k 는 $(1, z_{1k}, z_{2k}, \dots, z_{pk})$ 의 환경변수벡터 p 는 환경변수의 수

β 는 추정하고자 하는 계수벡터

으로 나타내었다. 식 (3)에서 $z_k \beta$ 는 모형에 포함된

환경변수로 설명할 수 있는 효율성점수이고 ω_k 는 효율성에 부정적인 확률적 요인이 작용해서 나타난 점수이다. 추리통계의 관점에서 식 (3)의 효율성모형은 실제효율성점수 δ_k 의 관측값을 구할 수 없어 실행하기 어렵다. SW는 계수벡터 β 를 추정하기 위해 식 (3)의 효율성모형에서 관측할 수 없는 실제효율성점수 δ_k 를 표본에서 추정한 DEA 효율성점수 $\hat{\delta}_k$ 로 대체하여

$$\hat{\delta}_k = \mathbf{z}_k \beta + u_k, \quad k = 1, 2, \dots, n \quad (4)$$

여기서 u_k 는 비효율성으로 평균 0, 표준편차 σ_u 의 정규분포이고 $1 - \mathbf{z}_k \beta$ 의 좌측에서 절단된 확률변수 n, \mathbf{z}_k, β 는 식 (2)에서와 같음

으로 나타내었다. 식 (3)과 식 (4)는 오차항이 $1 - \mathbf{z}_k \beta$ 의 좌측에서 절단된 절단정규분포(truncated regression model)⁶⁾이다. SW가 절단회귀모형을 주장한 배경은 DEA 효율성점수가 1.0(100%의 효율성을 가진 것으로 평가된 점수)인 경우는 실제로 존재하는 값에 대한 추정치가 아니고 유한표본 및 DEA 모형의 구조에 의해 만들어진 결과물이며 따라서 DEA 효율성점수가 1.0인 평가대상은 분석에서 제외되어야 한다는 것이다.

일반적으로 최대우도법은 오차항 $u_k, k = 1, \dots, n$ 이 동일한 확률밀도함수를 갖는 서로 독립인 확률변수이고, 오차항(u_k)과 공변량(\mathbf{z}_k)은 서로 독립임을 가정한다. 만약 이와같은 최대우도법의 일반적인 가정이 충족된다면 절단회귀모형으로 설정한 식

(4)의 로그우도함수(log-likelihood function)는

$$\ln L_{tr}(\beta, \sigma_u) = \sum_{k=1}^m \ln \left\{ \frac{1}{\sigma_u} \phi \left(\frac{u_k}{\sigma_u} \right) \left[\Phi \left(-\frac{1 - \mathbf{z}_k \beta}{\sigma_u} \right) \right]^{-1} \right\} \quad (5)$$

여기서 m 은 표본에 포함된 생산단위 중에서 DEA 효율성점수가 1.0인

경우를 제외한 생산단위의 수

$$u_k = \hat{\delta}_k - \mathbf{z}_k \beta$$

$\phi(\cdot)$ 는 표준정규분포의 확률밀도함수

$\Phi(\cdot)$ 는 표준정규분포의 누적분포함수

로 나타낼 수 있다[21]. 절단회귀모형으로 설정한 효율성모형의 계수벡터 β 에 대한 ML 추정치 $\hat{\beta}_{ml}^{tr}$ 는 식 (5)의 로그우도함수를 최대화하는 β 의 값을 의미한다. 이 값은 일치성, 점근정규성, 점근효율성 등의 점근특성(asymptotic properties)을 가지며 따라서 피셔정보행렬(Fisher information matrix)에 기초한 표준적인 추론방법을 사용할 수 있다. 즉 $\hat{\beta}_{ml}^{tr}$ 의 표준오차를 피셔정보행렬⁷⁾의 역행렬의 주대각요소의 제곱근으로 추정할 수 있으며 이를 이용하여 계수벡터 β 의 신뢰구간을 구하거나 β 에 대한 가설을 검정할 수 있다.

한편 SW는 식 (4)의 효율성모형에 대해서 $\hat{\beta}_{ml}^{tr}$ 의 표준오차에 대한 추정치를 활용하여 추론한 결과의 타당성이 보장되지 않음을 지적하였다. SW가 제시한 문제점은 다음과 같다. 첫째, 유한표본에서 종속변수인 DEA 효율성점수($\hat{\delta}_k$)의 편이가 존재하는 문제점을 가지고 있으며, 둘째, 오차항(u_k)간의 상관관계 그리고 오차항(u_k)과 공변량(\mathbf{z}_k)의 상관관계 등이 존재하여 최대우도법의 일반적인 가정이 만족되지 않으며, 셋째, 이러한 편이 및 상관관계가 표본크기의 증가에 따라 사라지기는 하지만 그 소멸 속도가 느려 피셔정보행렬에 기초한 표준적인 추론방법의 타당성이 보장되지 못함을 지적하였다. SW

6) 본 논문에서는 절단회귀모형(truncated regression model)과 수정회귀모형(censored regression model)을 Greene[21]이 구분한 것과 같은 방식으로 구분하여 사용한다. Greene[21]은 종속변수의 값이 경계값(limit value)과 같은 관측값을 표본에서 제외한 경우를 절단회귀모형으로 부르고, 종속변수의 값이 경계값과 같은 관측값을 표본에 포함시킨 경우를 수정회귀모형으로 지적하였다. 수정회귀모형은 토빗모형(tobit model)으로 부르기도 한다.

7) 피셔정보행렬은 표본자료로부터 관측정보행렬(observed information matrix) 또는 경사벡터외적(outer product of gradients)의 형태로 추정될 수 있다[21].

는 이러한 문제점을 극복하기 위한 방안으로 절단 회귀모형에 기반한 다음과 같은 붓스트랩절차를 제시하였다.

1) 절단회귀모형으로 설정된 식 (4)의 효율성모형에 ML을 적용하여 모수 β, σ_u 의 ML 추정치 $\hat{\beta}_{ml}^{tr}, \hat{\sigma}_{u ml}^{tr}$ 을 구한다.

2) 다음의 절차 (2.1)~(2.4)를 L_1 회 반복하여 붓스트랩점수 $\hat{\delta}_{kb}^*$, $k = 1, \dots, n, b = 1, \dots, L_1$ 를 구한다.

(2.1) $1 - z_k \hat{\beta}_{ml}^{tr}$ 에서 좌측으로 절단된 정규분포 $N(0, \hat{\sigma}_{u ml}^{tr})$ 에서 $u_k, k = 1, \dots, n$ 을 추출한다.

(2.2) $\delta_k^* = z_k \hat{\beta}_{ml}^{tr} + u_k, k = 1, \dots, n$ 을 계산한다.

(2.3) $y_k^* = (\delta_k / \delta_k^*) y_k, k = 1, \dots, n$ 을 계산한다.

(2.4) 식 (2)에서 Y 를 $Y^* = [y_1^*, \dots, y_n^*]$ 으로 대체하여 효율성점수 $\hat{\delta}_{kb}^*, k = 1, \dots, n$ 을 구한다.

3) $\bar{\delta}_k = \hat{\delta}_k - bias(\hat{\delta}_k)$ 으로 편의수정추정치 $\bar{\delta}_k, k = 1, \dots, n$ 를 구한다. 여기서 $bias(\hat{\delta}_k) = (1/L_1) \sum_{b=1}^{L_1} \hat{\delta}_{kb}^* - \hat{\delta}_k$ 이다.

4) 효율성모형 $\bar{\delta}_k = z_k \beta + u_k, k = 1, 2, \dots, n$ 을 절단회귀모형으로 설정하고 ML을 적용하여 모수 β, σ_u 의 ML 추정치 $\tilde{\beta}_{boot}^{tr}, \tilde{\sigma}_{u boot}^{tr}$ 을 구한다.

5) ML 추정치 $\tilde{\beta}_{boot}^{tr}, \tilde{\sigma}_{u boot}^{tr}$ 의 신뢰구간을 구하기 위해 다음의 절차 (5.1)~(5.3)를 L_2 회 반복하여 모수 β, σ_u 의 붓스트랩추정치 $(\tilde{\beta}_b^{*tr}, \tilde{\sigma}_{u,b}^{*tr}), b = 1, \dots, L_2$ 를 구한다.

(5.1) $1 - z_k \tilde{\beta}_{boot}^{tr}$ 에서 좌측으로 절단된 정규분포 $N(0, \tilde{\sigma}_{u boot}^{tr})$ 에서 $u_k, k = 1, \dots, n$ 을 추출한다.

(5.2) $\delta_k^{*tr} = z_k \tilde{\beta}_{boot}^{tr} + u_k, k = 1, \dots, n$ 을 계산한다.

(5.3) $\delta_k^{*tr} = z_k \beta + u_k, k = 1, \dots, n$ 을 $1 - z_k \beta$ 에서 좌측으로 절단된 절단회귀모형으로 설정한 후 ML로 $\tilde{\beta}_b^{*tr}, \tilde{\sigma}_{u,b}^{*tr}$ 을 구한다.

6) 최대우도추정치 $\tilde{\beta}_{boot}^{tr}$ 의 $(1-\alpha) \times 100\%$ 신뢰구간의 하한은 $\tilde{\beta}_{boot}^{tr} + lb_\alpha^*$ 으로 구하고 상한은 $\tilde{\beta}_{boot}^{tr} + ub_\alpha^*$ 으로 구한다. 여기서 lb_α^* 와 ub_α^* 은 $\Pr[-ub_\alpha \leq (\tilde{\beta}^{*tr} - \tilde{\beta}_{boot}^{tr}) \leq -lb_\alpha] \approx 1-\alpha$ 로 구한다. $\tilde{\sigma}_{u boot}^{tr}$ 의 신뢰구간도 $\tilde{\beta}_{boot}^{tr}$ 의 신뢰구간을 구하는 것과 같은 방식으로 구할 수 있다.

3.2 Banker and Natarajan의 접근방법

Banker and Natarajan[6](BN)은 Shephard거리 함수 θ_k 를 효율성의 지표로 사용하였다. BN은 식 (1)에서 생산함수 $y^o(x)$ 이 x 에 대해서 단조증가오목(monotone increasing and concave)임을 가정하였다. 그리고 식 (1)의 효율성모형을 $h(z)$ 은 선형함수로 가정하고 $g[h(z), \nu, \omega]$ 은 $g[h(z), \nu, \omega] = \exp[-h(z) + \nu - \omega]$ 로 가정한 후 양변에 로그를 취하여

$$\begin{aligned} \ln \theta_k &= -M - z_k \beta + \nu_k - \omega_k \\ &= -M - z_k \beta + \xi_k, \quad k = 1, 2, \dots, n \end{aligned} \tag{6}$$

여기서 M 은 임의의 양의 실수

ν_k 는 오차항으로 $-M \leq \nu_k \leq M$ 에서 양의 확률을 가지며 $E(\nu_k) = 0$ 인 대칭 확률변수
 ω_k 는 오차항으로 $0 \leq \omega_k < \infty$ 에서 양의 확률을 갖는 비대칭 확률변수

$\xi_k (= \nu_k - \omega_k)$ 는 복합오차항

z_k 는 $(z_{1k}, z_{2k}, \dots, z_{pk})$ 이며 요소가 모두 양수인 환경변수벡터⁸⁾

β 는 추정하고자 하는 계수벡터

으로 나타내었다. 식 (6)에서 종속변수가 $\ln \theta_k \leq 0$ ($\Leftrightarrow 0 < \theta_k \leq 1$)이 되려면 $z_k \beta \geq 0$ 이어야 한다. BN은 $z_k \beta \geq 0$ 이 만족하도록 환경변수벡터 z_k 의 요소

8) SW의 효율성모형 식 (4)에서 환경변수벡터 z_k 는 첫 번째 요소로 1을 포함하고 있으나 위의 식 (6)에서 환경변수벡터 z_k 는 첫 번째 요소로 1을 포함하지 않고 환경변수로만 구성되어 있으며 편의상 같은 문자로 나타낸다.

는 모두 양수이며 값이 클 수록 생산단위의 비효율성이 커지게 측정된 것으로 가정하였다. 식 (6)에서 $\mathbf{z}_k\beta$ 는 모형에 포함된 환경변수로 설명할 수 있는 효율성점수이다. 그리고 비대칭확률오차 ω_k 는 한쪽 방향으로의 변화가 효율성에 기여하는 확률적 요인이 작용해서 나타난 점수이며 대칭확률오차 ν_k 는 적정 수준으로의 변화가 효율성에 기여하는 확률적 요인이 작용해서 나타난 점수이다. BN은 식 (6)의 효율성모형에서 관측할 수 없는 효율성점수 θ_k 를 표본에서 추정한 DEA효율성점수 $\hat{\theta}_k$ 로 대체하여

$$\begin{aligned}\ln \hat{\theta}_k &= -M - \mathbf{z}_k\beta + \nu_k - u_k \\ &= -M - \mathbf{z}_k\beta + \epsilon_k, \quad k = 1, 2, \dots, n\end{aligned}\quad (7)$$

여기서 ν_k 는 오차항으로 $-M \leq \nu_k \leq M$ 에서 확률을 가지며 $E(\nu_k) = 0$ 인 대칭확률변수
 u_k 는 오차항으로 $0 \leq u_k < \infty$ 에서 확률을 갖는 비대칭확률변수
 $\epsilon_k (= \nu_k - u_k)$ 는 복합오차항
 M, \mathbf{z}_k, β 는 식 (6)에서와 같음

으로 나타내었다. 식 (6)과 식 (7)은 대칭확률변수와 비대칭확률변수로 구성된 복합오차항을 갖는 확률프론티어모형(stochastic frontier model)이다.

BN은 식 (7)의 효율성모형에서 환경변수벡터(\mathbf{z}_k), 확률오차(ν_k, u_k), 투입벡터(\mathbf{x}_k) 등을 서로 독립(환경변수간의 상관관계와 투입변수간의 상관관계는 허용)이며 유한분산(finite variance)을 가지고 있는 확률변수로 가정하였다. 이러한 가정을 전제로 하여 BN은 계수벡터 β 에 대한 OLS 추정치 $\hat{\beta}_{ols}^{sf}$ 가 일치성 및 점근정규성 등의 특성을 가지고 있음을 증명하였다. OLS 추정치 $\hat{\beta}_{ols}^{sf}$ 는 식 (7)에서 잔차(residuals)의 제곱합을 최소화하는 β 의 값이며

$$(\hat{M}_{ols}^{sf}, \hat{\beta}_{ols}^{sf}) = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}' \ln \hat{\theta}\quad (8)$$

여기서 \mathbf{Z} 는 $(-1, -\mathbf{z}_k)$ 를 k 번째 행으로 갖는 $n \times (p+1)$ 행렬

p 는 환경변수의 수

$\hat{\theta}$ 는 $\hat{\theta}_k$ 를 k 번째 요소로 갖는 $(n \times 1)$ 벡터

으로 구할 수 있다.

점근정규성은 충분한 크기의 표본에서 통계적 추론의 타당성을 보장해주는 근거가 된다. BN은 식 (7)의 효율성모형의 OLS 추정치 $\hat{\beta}_{ols}^{sf}$ 에 대해서는 일치성과 함께 점근정규성이 만족됨을 증명하였으나 ML 추정치 $\hat{\beta}_{ml}^{sf}$ 에 대해서는 일치성만을 증명하고 점근정규성에 대한 증명은 제시하지 못하였다. 따라서 $\hat{\beta}_{ml}^{sf}$ 에 대한 통계적 추론은 타당성이 보장되지 못한다. ML 추정치를 구하기 위해서는 확률오차 ν_k 와 u_k 의 확률분포의 종류를 구체적으로 가정해야 한다. BN은 $\hat{\beta}_{ml}^{sf}$ 의 일치성을 증명할 때 모의실험으로 성과를 비교할 때 다른 분포모형을 사용하였다. 일치성증명에서 사용된 대칭확률오차 ν_k 의 분포는 평균 0, 표준편차 σ_ν , $-M$ 의 좌측과 M 의 우측으로 절단된 정규분포이고, 비대칭확률오차 u_k 의 분포는 모수가 2와 λ 인 감마분포이다. BN이 일치성증명에서 정규-감마모형(normal-gamma model)을 선택한 것은 Greene조건[19] 및 Bierens정리[8, 24]를 활용하여 비교적 간단히 증명할 수 있기 때문인 것으로 판단된다.

BN이 모의실험에서 사용한 대칭확률오차 ν_k 의 분포는 평균 0, 표준편차 σ_ν 의 정규분포이고, 비대칭확률오차 u_k 의 분포는 평균 0, 표준편차 σ_u 의 반정규분포이다. BN이 모의실험에서 정규-반정규모형(normal-halfnormal model)을 사용한 것은 정규-감마모형의 경우와 마찬가지로 Greene조건 및 Bierens정리로 일치성을 간단히 증명할 수 있으며 또한 계산이 정규-감마모형보다 간단하기 때문인 것으로 판단된다. 정규-반정규모형의 로그우도함수는

$$\ln L_{sf}(M, \beta, \lambda, \sigma) = \sum_{k=1}^n \ln \left\{ \frac{2}{\sigma} \phi \left(\frac{\epsilon_k}{\sigma} \right) \Phi \left(-\frac{\lambda \epsilon_k}{\sigma} \right) \right\}\quad (9)$$

여기서 $\epsilon_k = \ln \hat{\theta}_k - (-M - \mathbf{z}_k\beta)$

$$\sigma = (\sigma_v^2 + \sigma_u^2)^{1/2}$$

$$\lambda = \sigma_u / \sigma_v$$

$\phi(\cdot)$ 는 표준정규분포의 확률밀도함수

$\Phi(\cdot)$ 는 표준정규분포의 누적분포함수

로 나타낼 수 있다[1, 21]. 확률프론티어모형으로 설정한 효율성모형의 계수벡터 β 에 대한 ML 추정치 $\hat{\beta}_{ml}^*$ 는 식 (9)의 로그우도함수를 최대화하는 β 의 값을 의미한다. 앞서서도 언급하였듯이 $\hat{\beta}_{ml}^*$ 의 점근정규성은 아직 증명되지 않았으며 따라서 피서정보행렬에 기초한 표준적 추론결과의 타당성이 보장되지 않는다.

3.3 특징 및 장단점 비교

DEA 모형은 표본에서 구한 생산프론티어를 비교기준으로 생산단위의 효율성수준을 평가하는 선형계획법모형이라 할 수 있다. SW 모형과 BN 모형은 모두 환경변수가 효율성평가기준 즉 생산프론티어에 영향을 주지 않는다는 분리조건(separability condition)을 전제로 하고 있다. 분리조건이 전제되지 않으면 환경변수가 생산프론티어에 영향을 주게 되며 따라서 효율성점수를 구하는 DEA 모형 자체에 환경변수의 영향이 포함되어야 한다(환경변수의 영향이 포함된 DEA 모형에 관해서는 Syrjanen[40], Muñiz et al.[31], Camanho et al.[11] 등을 참조할 수 있다).

BN은 투입벡터의 함수로 정의된 생산프론티어가 환경변수의 영향을 받지 않도록 하기 위해, 즉 분리조건이 충족되도록 하기 위해 투입벡터와 환경변수 벡터가 서로독립(환경변수간의 상관관계와 투입변수간의 상관관계는 허용)임을 가정하였다. 이는 환경변수와 무관하게 투입수준을 결정하는 생산단위를 대상으로 BN 모형이 적용되어야 함을 의미한다. 일반적으로 환경변수에 따라 생산단위의 투입수준은 달라질 수 있다. 이러한 관점에서 McDonald [30]는 BN 모형이 투입변수와 환경변수가 상관관계를 가질 수 있는 상황으로 확장될 필요가 있음을 지적하였다. 한편 SW 모형에서는 관측값의 확률밀

도함수의 받침(support)을 생산가능집합과 환경변수공간의 데카르트곱에서 갖도록 정의하고 효율성 점수를 환경변수의 함수로 나타내었다([37] 가정 A1과 A2). 이를 통해서 환경변수가 생산프론티어에 영향을 주지 않으면서 평가대상의 효율성수준에 영향을 줄 수 있도록 하였다. 이는 투입변수 및 환경변수가 서로 독립이어야 한다는 가정을 전제로 하지 않고 분리조건이 충족되도록 하여 생산단위의 투입의사결정이 환경변수의 영향을 받는 경우에도 SW 모형을 적용할 수 있음을 의미한다.

BN이 제시한 식 (7)의 확률프론티어모형에서는 생산프론티어와 평가대상의 편차가 반영된 종속변수 $\ln \hat{\theta}_k$ 를 환경변수의 영향, 비대칭확률오차 및 대칭확률오차의 3개 요소로 설명하고 있다. 한편 SW가 제시한 식 (4)의 절단회귀모형에서는 생산프론티어와 평가대상의 편차가 반영된 종속변수 $\hat{\delta}_k$ 를 환경변수의 영향과 비대칭확률오차만으로 설명하고 있다. SW 모형은 생산단위의 투입의사결정이 환경변수의 영향을 받는 경우에도 적용될 수 있다는 장점을 갖는 반면 BN 모형에서 고려하고 있는 대칭확률오차가 포함되어 있지 않아 모형의 적용범위가 BN 모형과 비교하여 상대적으로 제한적이라는 한계점을 갖는다.

SW 모형에서는 환경변수의 값의 범위나 측정방식에 관한 조건이 없으며 따라서 SW 모형을 적용할 경우 환경변수의 값의 범위나 측정방식에 관한 검토가 필요치 않다. 한편 BN 모형은 $z_k \beta \geq 0$ 을 보장하기 위해 환경변수가 모두 양수이며 값이 클수록 생산단위의 비효율성이 커지게 측정된 것으로 가정하고 있다. 환경변수에 음수가 포함되어 있으면 '가장 작은 음수 $\times (-1)$ '을 더하고, 환경변수의 값이 작을수록 생산단위의 비효율성이 커지게 측정되어 있으면 '-1'을 곱하여 범위나 측정방식에 관한 조건을 충족시킬 수 있다. 그런데 식 (7)의 효율성 모형의 계수벡터에 대한 OLS 추정치와 ML 추정치는 '가장 작은 음수 $\times (-1)$ ' 더하기 변환에 대해서는 절편항의 추정치만 달라지고 '-1' 곱하기 변환에 대해서는 계수추정치에 부호만 달라진다. 따라서 BN

모형을 적용할 경우에도 SW 모형의 경우와 같이 환경변수의 값의 범위나 측정방식에 관한 검토가 필요치 않다.

BN 모형의 OLS 추정치 및 ML 추정치는 LIMDEP [20], STATA[39] 등의 통계패키지에 계산절차가 구현되어있어 비교적 간단하게 구할 수 있다. 특히 OLS 추정치는 모든 통계패키지에서 제공된다. 그러나 SW의 붓스트랩절차를 구현해 놓은 통계패키지는 아직 없다. 따라서 SW의 붓스트랩절차로 효율성 결정요인을 분석하려면 Gauss, MATLAB, R, S-Plus 등과 같이 프로그래밍언어의 체계를 갖고 있는 소프트웨어로 연구자가 직접 프로그램을 작성하여 실행해야 한다. 이는 Gauss, MATLAB, R, S-Plus 등의 명령어 및 구문에 관한 충분한 지식을 전제로 해야 하는 한계점을 갖는다.

4. 모의실험

제 3장에서는 DEA효율성점수의 결정요인을 분석하기 위한 Simar and Wilson[37](SW)의 접근방법과 Banker and Natarajan[6](BN)의 접근방법을 설명하고 이론적 특징 및 장단점을 비교하였다. 여기서는 실제효율성모형(true efficiency model)이 가정과 다를 때 SW 방법과 BN 방법이 유효표본에서 나타내는 성과를 모의실험으로 비교한다.

4.1 실험설계

모의실험으로 효율성모형 추정방법(즉 SW 방법과 BN 방법)의 성과를 비교하려면 종속변수인 DEA 효율성점수와 독립변수인 환경변수벡터의 관측값이 필요하며 관측값 자료를 생성하기 위한 실제생산기술(true production technology)과 실제효율성모형(true efficiency model)을 식 (1)의 생산함수와 효율성모형의 형태로 가정해야 한다. 본 논문에서는 Kumbhakar et al.[27], Caudill et al.[13], Alvarez et al.[3] 등이 제안한 3가지 효율성모형을 효율성모형 추정방법의 성과를 비교하기 위한 실제효율성모형

으로 사용하고 실제생산기술을 나타내기 위한 생산함수와 모든 모수값은 Banker and Natarajan[6]의 실험설계를 사용한다.

Kumbhakar, Ghosh, and McGuckin[27](KGM), Caudill, Ford, and Gropper[13](CFG), Alvarez, Amsler, Orea, and Schmidt[3](AAOS) 등이 제안한 3가지 효율성모형⁹⁾은 각각 다음 식 (10)~식 (12)로 나타낼 수 있다.

$$\ln\theta_k = \nu_k - \omega_k, \nu_k \sim N(0, \sigma_\nu^2), \omega_k \sim N^+(\mathbf{z}_k\beta, \sigma_\omega^2) \quad (10)$$

$$\ln\theta_k = \nu_k - \omega_k, \nu_k \sim N(0, \sigma_\nu^2), \omega_k \sim N^+[0, \{\sigma_\omega \exp(\mathbf{z}_k\beta)\}^2] \quad (11)$$

$$\ln\theta_k = \nu_k - \omega_k, \nu_k \sim N(0, \sigma_\nu^2), \omega_k \sim N^+[\exp(\mathbf{z}_k\beta), \{\sigma_\omega \exp(\mathbf{z}_k\beta)\}^2] \quad (12)$$

여기서 N^+ 의 위첨자 +는 정규분포를 0에서 좌측으로 절단하였음을 나타내며 \mathbf{z}_k, β 는 식 (3)에서와 같다. 식 (10)의 KGM모형에서 비대칭확률오차 ω_k 의 분포는 평균이 환경변수의 값에 따라 달라진

9) 생산단위의 효율성점수를 추정하는 방법은 생산함수의 모수적 형태를 가정해야 하는 확률프론티어분석(stochastic frontier analysis : SFA)과 생산함수의 모수적 형태를 가정할 필요가 없는 자료포락분석(data envelopment analysis : DEA)으로 구분할 수 있다 [16]. 효율성점수의 결정요인을 분석하는 방법은 모수적 2단계접근법, 비모수적 2단계접근법, 그리고 모수적 1단계 접근법 등으로 구분할 수 있다[40, 6]. 모수적 2단계 접근법은 환경변수가 독립변수이고 모수적 방법인 SFA로 추정된 효율성점수가 종속변수인 효율성모형을 추정하는 접근법이고, 비모수적 2단계 접근법은 환경변수가 독립변수이고 비모수적 방법인 DEA로 추정된 효율성점수가 종속변수인 효율성모형을 추정하는 접근법이다. 마지막으로 모수적 1단계 접근법은 생산함수와 효율성모형을 결합하고 이를 ML로 추정하는 방법이다. 본 논문의 관심은 비모수적 2단계 접근법으로 제시된 SW 방법과 BN 방법을 비교하는 것이며 이를 위해 SW 방법 및 BN 방법의 가정과 다른 효율성모형이 필요하다. KGM, CFG, AAOS모형은 모수적 1단계 접근법으로 제안된 효율성모형으로 SW방법과 BN 방법의 비교를 위한 모의 실험자료를 생성하는데 사용한다.

다. 식 (10)의 KGM모형은

$$\begin{aligned} \ln\theta_k &= -z_k\beta + \nu_k - \omega_k, \nu_k \sim N(0, \sigma_\nu^2), \\ \omega_k &\sim N(0, \sigma_\omega^2), \omega_k \geq -z_k\beta \end{aligned} \quad (13)$$

로 다시 나타낼 수 있으며 식 (13)의 KGM모형은 $z_k\beta \geq 0$ 을 요구하지 않는다는 점이 식 (6)의 BN 모형과 다르다. 식 (11)의 CFG모형에서 비대칭확률 오차 ω_k 의 분포는 분산이 환경변수의 값에 따라 달라진다. 식 (11)에서 ω_k 의 분포 $N^+[0, \{\sigma_\omega \exp(z_k\beta)\}^2]$ 는 $\exp(z_k\beta)N^+(0, \sigma_\omega^2)$ 와 같이 환경변수의 영향을 받는 부분 $\exp(z_k\beta)$ 와 그렇지 않은 부분 $N^+(0, \sigma_\omega^2)$ 의 곱으로 다시 나타낼 수 있으며 여기서 $N^+(0, \sigma_\omega^2)$ 은 생산단위의 경영능력의 분포로 해석할 수 있고 $\exp(z_k\beta)$ 은 경영능력이 생산단위를 효율적으로 경영하는데 발휘되는 정도로 해석할 수 있다[3]. 식 (12)의 AAOS모형에서 비대칭확률오차 ω_k 의 분포는 평균과 분산 모두 환경변수의 값에 따라 달라진다. 식 (12)에서 ω_k 의 분포 $N^+[\exp(z_k\beta), \{\sigma_\omega \exp(z_k\beta)\}^2]$ 는 $\exp(z_k\beta)N^+(1, \sigma_\omega^2)$ 와 같이 환경변수의 영향을 받는 부분 $\exp(z_k\beta)$ 와 그렇지 않은 부분 $N^+(1, \sigma_\omega^2)$ 의 곱으로 다시 나타낼 수 있으며 각각 CFG모형에서와 같은 방식으로 해석할 수 있다.

모의실험을 위한 관측값의 생성절차는 다음과 같다.

- 1) (투입변수) $x_k, k=1, 2, \dots, n$ 을 [1, 4]의 균등 분포에서 추출한다.
- 2) (환경변수) $z_k, k=1, 2, \dots, n$ 을 [0, 1]의 균등 분포에서 추출한다.
- 3) (대칭확률오차) $\nu_k, k=1, 2, \dots, n$ 을 $\sigma_\nu = 0.04$ 의 $N(0, \sigma_\nu^2)$ 에서 추출한다.
- 4) (비대칭확률오차) $\beta = 0.2, \sigma_\omega = 0.15$ 로 놓고 $\omega_k, k=1, 2, \dots, n$ 을 KGM모형은 $N^+(z_k\beta, \sigma_\omega^2)$ 에서, CFG모형은 $N^+[0, \{\sigma_\omega \exp(z_k\beta)\}^2]$ 에서, AAOS모형은 $N^+[\exp(z_k), \{\sigma_\omega \exp(z_k\beta)\}^2]$ 에서 추출한다.

- 5) (산출변수) $\ln y_k, k=1, 2, \dots, n$ 을 $\ln y_k = \ln(-37 + 48x_k - 12x_k^2 + x_k^3) + \nu_k - \omega_k$ 로 구한 후 $y_k = \exp(\ln y_k)$ 로 $y_k, k=1, 2, \dots, n$ 을 구한다.
- 6) (DEA 효율성점수) $x_k, y_k, k=1, 2, \dots, n$ 을 식 (2)의 DEA 모형에 적용해서 $\hat{\delta}_k, k=1, 2, \dots, n$ 을 구한다.

4.2 비교대상

본 논문에서는 모의실험을 사용하여 효율성모형을 추정하기 위한 4가지 방법의 성과를 비교한다. SW가 제시한 첫 번째 방법과 두 번째 방법은 각각 절단회귀모형으로 설정한 효율성모형을 ML로 추정하는 방법(식 (5)의 로그우도함수를 최대화하는 $\hat{\beta}_{ml}^{tr}$)과 절단회귀모형에 기초한 붓스트랩절차로 추정하는 방법(SW의 붓스트랩절차로 구하는 $\hat{\beta}_{boot}^{tr}$)이다. SW는 효율성모형에 대해서 $\hat{\beta}_{ml}^{tr}$ 의 표준오차로 추론한 결과는 편의 및 상관관계 문제로 인해 타당성이 보장되지 못하며 붓스트랩절차로 구한 $\hat{\beta}_{boot}^{tr}$ 와 $\hat{\beta}_{boot}^{tr}$ 의 표본분포를 대안으로 제시하였다. SW는 식 (3)을 실제효율성모형으로 사용하여 생성한 관측값에 $\hat{\beta}_{ml}^{tr}$ 와 $\hat{\beta}_{boot}^{tr}$ 를 적용한 후 신뢰구간이 실제모수를 포함하는 비율(coverages of confidence intervals: CCI)과 평균제곱오차제곱근(root mean square error : RMSE)으로 성과를 측정하였다. SW의 모의 실험에서는 $\hat{\beta}_{ml}^{tr}$ 보다 $\hat{\beta}_{boot}^{tr}$ 이 우월한 성과를 보여 SW의 예상과 일치하는 결과를 나타내었다. 여기서는 SW 방법의 가정과 다른 KGM 모형, CFG 모형, AAOS 모형 등을 실제효율성모형으로 사용하여 생성한 관측값에서도 SW의 예상과 일치하는 결과가 나타나 는가를 살펴본다.

모의실험으로 비교하고자 하는 세 번째 방법과 네 번째 방법은 BN이 제시한 방법으로 각각 확률 프론티어모형으로 설정한 효율성모형을 OLS와 ML로 추정하는 방법(식 (8)로 구하는 $\hat{\beta}_{ols}^{sf}$ 와 식 (9)의 로그우도함수를 최대화하는 $\hat{\beta}_{ml}^{sf}$)이다. BN은 식 (6)

을 실제효율성모형으로 사용하여 생성한 관측값에 $\hat{\beta}_{ols}^{sf}$ 와 $\hat{\beta}_{ml}^{sf}$ 를 포함한 여러 가지 방법을 적용한 후 평균절대편차비율(mean absolute deviation percentage : MADP)과 평균제곱오차제곱근비율(root mean square error percentage : RMSEP)로 성과를 측정하였다. BN의 모의실험에서는 $\hat{\beta}_{ols}^{sf}$ 와 $\hat{\beta}_{ml}^{sf}$ 이 다른 추정방법들보다 성과가 우월한 것으로 나타났다. 본 논문에서는 BN 방법의 가정과 다른 KGM 모형, CFG 모형, AAOS 모형 등을 실제효율성모형으로 사용하여 생성한 관측값에서 $\hat{\beta}_{ols}^{sf}$ 와 $\hat{\beta}_{ml}^{sf}$ 의 성과가 어떻게 나타나는가를 살펴본다.

4.3 실험결과

효율성모형 추정방법의 성과를 비교하기 위하여 4.1절에서 설명한 관측값 생성절차를 사용하여 표본 크기(n)가 100, 200, 400인 표본을 각각 1000세트 생성하였다. 각각의 표본에 4가지 추정방법 1) $\hat{\beta}_{ols}^{sf}$ (확률프론티어모형으로 설정한 효율성모형을 OLS로 추정하는 방법), 2) $\hat{\beta}_{ml}^{sf}$ (확률프론티어모형으로 설정한 효율성모형을 ML로 추정하는 방법), 3) $\hat{\beta}_{ml}^{tr}$ (절단회귀모형으로 설정한 효율성모형을 ML로 추정하는 방법), 4) $\hat{\beta}_{boot}^{tr}$ (절단회귀모형에 기초한 부스트랩절차로 추정하는 방법)을 적용하여 각 방법별로 1,000개의 β 에 대한 추정치를 구하였다. $\hat{\beta}_{boot}^{tr}$ 을 적용할 때 필요한 L_1 과 L_2 의 값은 Simar and Wilson[36]에서와 같은 $L_1 = 100$, $L_2 = 2000$ 을 사용하였다. 모의실험계산은 GLPK Package[29]와 R[41]로 프로그램을 작성하여 수행하였다.

[그림 1]은 효율성모형 추정방법의 표본분포에 대한 핵밀도추정(kernel density estimates)결과를 나타낸 것이다.¹⁰⁾ [그림 1]은 4개의 행과 3개의 열

로 배열된 12개의 밀도곡선그림으로 구성되어 있다. 4개의 행은 각각 효율성모형 추정방법에 대응되고 3개의 열은 자료생성에 사용된 각각의 실제효율성모형에 대응된다. 두 번째, 세 번째, 및 네 번째 행에 나타낸 $\hat{\beta}_{ml}^{sf}$, $\hat{\beta}_{ml}^{tr}$, $\hat{\beta}_{boot}^{tr}$ 는 모두 로그우도함수를 최대화하는 계산절차를 사용한다. 세 번째와 네 번째 행의 $\hat{\beta}_{ml}^{tr}$ 와 $\hat{\beta}_{boot}^{tr}$ 는 로그우도함수 최대화문제의 수렴이 1,000개 모두 정상적으로 이루어 졌으나 두 번째 행의 $\hat{\beta}_{ml}^{sf}$ 는 실제생산기술이 KGM일 때 86개 그리고 CFG일 때 40개가 정상적으로 수렴된 모수를 구하는데 실패하였다. [그림 1]의 각 밀도곡선그림에서 얇은실선은 표본크기 100, 점선은 표본크기 200, 굵은실선은 표본크기 400에 대응되며 얇은실선, 점선, 굵은실선의 밀도곡선의 차이는 표본크기의 증가에 따른 효과를 나타낸다. [그림 1]의 밀도곡선들을 살펴보면 표본크기가 증가함에 따라 밀도곡선의 중심위치는 변화가 없지만 산포는 감소하고 있음을 알 수 있다.

각각의 밀도곡선그림에서 수직방향의 점선은 실제모수값 $\beta = 0.2$ 를 나타내며 밀도곡선의 중심위치가 수직방향점선에 가깝고 분포의 산포가 적을수록 효율성모형 추정방법의 성과가 우수함을 나타낸다. [그림 1]의 밀도곡선들을 살펴보면 실제효율성모형에 관계없이 우월한 성과를 나타내는 추정방법은 없는 것으로 나타났다. 실제효율성모형이 KGM일 경우에는 SW가 제안한 $\hat{\beta}_{ml}^{tr}$ 와 $\hat{\beta}_{boot}^{tr}$ 이 우수한 것으로 나타났고 실제효율성모형이 AAOS일 경우에는 BN이 제안한 $\hat{\beta}_{ols}^{sf}$ 와 $\hat{\beta}_{ml}^{sf}$ 이 우수한 것으로 나타났다. 실제효율성모형이 CFG일 경우에는 네 가지 추정방법 모두 표본분포의 위치가 실제모수값과 달라 추정성과가 좋지 못한 것으로 나타났으며 표본분포의 산포는 크지만 중심위치가 상대적으로 모수에 가까운 SW방법($\hat{\beta}_{ml}^{tr}$ 와 $\hat{\beta}_{boot}^{tr}$)이 BN 방법($\hat{\beta}_{ols}^{sf}$ 와 $\hat{\beta}_{ml}^{sf}$)보다 다소 우수한 것으로 판단된다.

효율성모형에서 효율성에 영향을 주는 요소는 환경변수(z_k), 비대칭확률오차(ω_k), 대칭확률오차(ν_k) 등으로 구분할 수 있다. <표 1>은 [그림 1]에 나타낸

10) 핵밀도추정치는 R의 density 함수로 구하였고 핵(kernel)과 띠너비선택(bandwidth selection)은 density 함수에서 기본으로 설정되어 있는 가우스핵(Gaussian kernel)과 띠너비선택관련규칙(rule-of-thumb for bandwidth selection)을 사용하였다.

〈표 1〉 추정방법의 효율성모형과 실제효율성모형의 비교

모형			비대칭확률오차(ω_k)		환경변수의 영향 ($z_k\beta$)	대칭확률 오차(ν_k)의 분포
구 분	명칭	식	분포	좌측절단점		
추정방법의 효율성모형	BN	(6) ^a	$N(0, \sigma_\omega^2)$	$\omega_k \geq 0$	$z_k\beta > 0$	$N(0, \sigma_\nu^2)$
	SW	(3)	$N(0, \sigma_\omega^2)$	$\omega_k \geq 1 - z_k\beta$	$(-\infty, \infty)$	없음
실제 효율성모형	KGM	(10) ^b	$N(0, \sigma_\omega^2)$	$\omega_k \geq -z_k\beta$	$(-\infty, \infty)$	$N(0, \sigma_\nu^2)$
	CFG	(11)	$N(0, \sigma_\omega^2) \exp(z_k\beta)$	$\omega_k \geq 0$	$(-\infty, \infty)$	$N(0, \sigma_\nu^2)$
	AAOS	(12)	$N(1, \sigma_\omega^2) \exp(z_k\beta)$	$\omega_k \geq 0$	$(-\infty, \infty)$	$N(0, \sigma_\nu^2)$

주) ^a 정규-반정규모형의 특징을 요약한 것임.

^b 식 (10)과 동치인 식 (13)의 특성을 요약한 것임.

결과의 의미를 해석하기 위해 추정방법의 효율성모형과 실제효율성모형의 특징을 비교할 수 있는 형태로 정리한 것이다. 실제효율성모형 KGM, CFG, AAOS는 대칭확률오차를 포함하고 있으며 환경변수의 영향($z_k\beta$)에 범위제약이 없다는 공통점을 가지고 있으나 비대칭확률오차에 관한 조건에서 서로 구분된다. KGM은 비대칭확률오차(농장경영자능력의 영향)의 좌측절단점(효율성에 가장 크게 기여하는 수준)이 환경변수(주변인구밀도, 경쟁농장의 수)의 상태에 따라 달라진다. 반면 CFG와 AAOS는 비대칭확률오차의 좌측절단점이 환경변수의 상태에 따라 달라지지 않고 0으로 고정되어 있다. KGM에서는 비대칭확률오차의 분포(농장경영자능력의 분포)가 환경변수의 상태와 무관하지만 CFG와 AAOS에서는 비대칭확률오차의 분포가 환경변수의 상태에 따라 달라진다. CFG에서는 환경변수의 상태에 따라 비대칭확률오차 분포의 산포형태가 달라지고 AAOS에서는 산포형태와 평균수준이 함께 달라진다.

추정방법의 효율성모형 BN과 SW는 비대칭확률오차의 분포가 같지만 나머지 조건은 모두 다르다. BN에서는 대칭확률오차(기온, 강수량, 해발고도)가 포함되어 있으며 좌측절단점은 환경변수의 상태와 관계없이 0으로 고정되어 있고 환경변수의 영향은 0보다 큰 범위로 제약되어 있다. SW에서는 대칭확률오차가 포함되어 있지 않으며 좌측절단점은 환경변수의 상태에 따라 달라지고 환경변수의 영향에는 범위제약이 없다. [그림 1]에 나타난 결과에서 실제

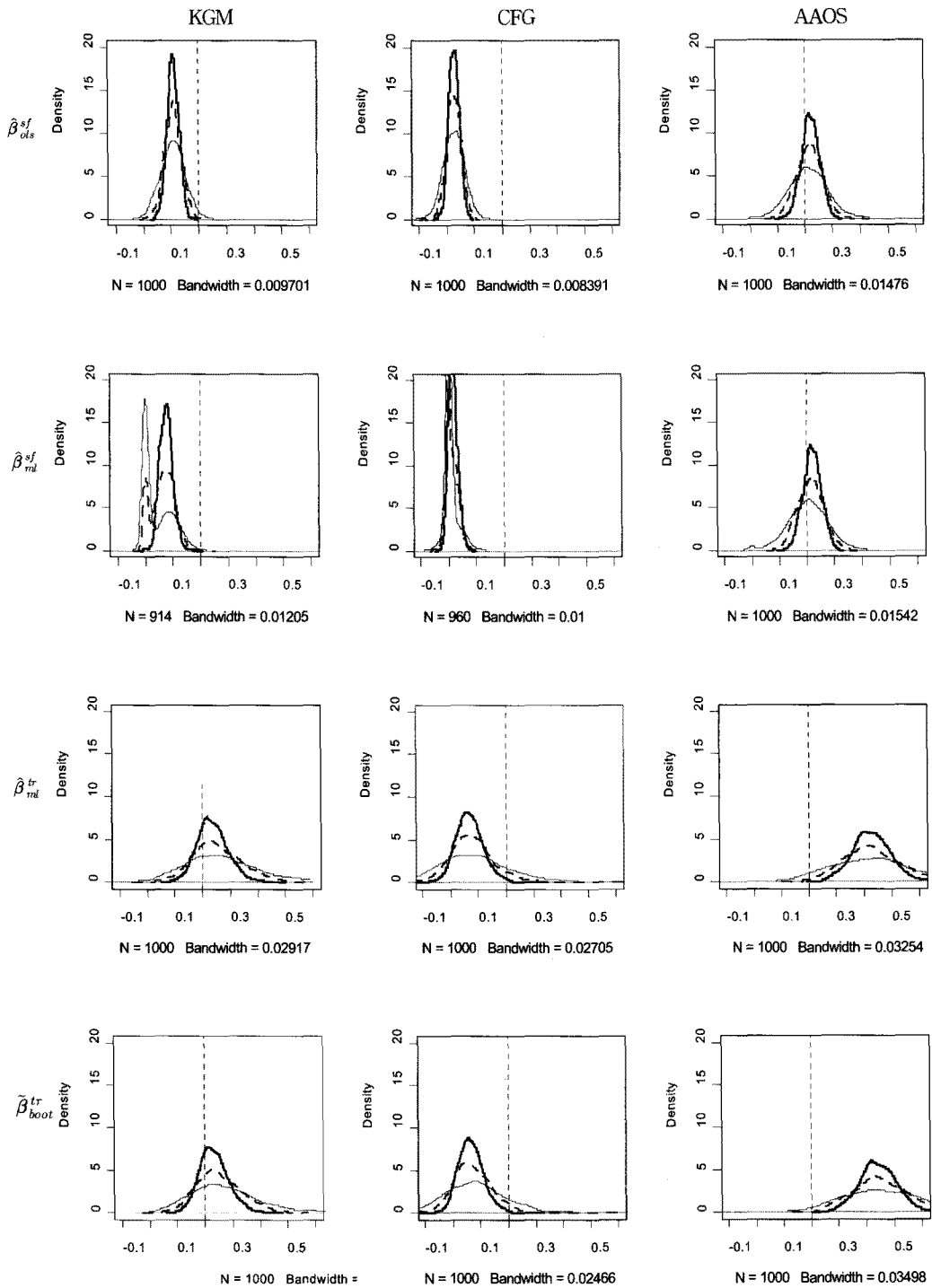
효율성모형이 KGM일 때 SW 방법이 BN 방법보다 우수하게 나타난 것은 대칭확률오차가 포함된 것의 영향 보다 좌측절단점(비대칭확률오차가 효율성에 가장 크게 기여하는 수준)이 환경변수의 상태에 따라 달라지는 특성이 더 중요하게 작용했음을 의미한다. 실제효율성모형 CFG에서는 환경변수의 상태에 따라 비대칭확률오차 분포의 산포형태가 달라지고 AAOS에서는 산포형태와 평균수준이 함께 달라진다. [그림 1]에 나타난 결과는 CFG와 AAOS의 차이점이 BN 방법과 SW 방법의 성과에 영향을 주고 있음을 보여준다. 환경변수의 상태가 비대칭확률오차 분포(농장경영자능력의 분포)의 산포형태와 평균수준에 모두 영향을 주는 상황에서는 BN 방법의 성과가 우월하게 나타났으나 산포형태에만 영향을 주는 상황에서는 BN 방법 및 SW 방법 모두 낮은 성과를 나타내었다.

〈표 2〉는 각 추정방법의 성과를 평균절대편차비율(MADP)

$$MADP = 100 \times \left(\frac{1}{1,000} \cdot \sum_{i=1}^{1000} |\hat{\beta} - \beta| \right) / \beta$$

로 평가한 결과이다. 여기서 β 는 실제모수이고 $\hat{\beta}$ 는 모수의 추정치이다.

〈표 2〉에 나타난 MADP로 평가한 효율성모형 추정방법의 성과는 [그림 1]에서의 판단과 전반적으로 일치하는 결과를 보여주고 있다. 실제효율성



[그림 1] 효율성모형 추정방법의 표본분포

<표 2> 효율성모형 추정방법의 성과 (MADP)¹¹⁾

실제 효율성 모형	표본 크기	효율성모형 추정방법			
		BN 방법		SW 방법	
		$\hat{\beta}_{ols}^{sf}$	$\hat{\beta}_{ml}^{sf}$	$\hat{\beta}_{ml}^{tr}$	$\hat{\beta}_{boot}^{tr}$
KGM	100	47.4	79.6	58.0	55.8
	200	46.8	69.6	39.2	37.7
	400	46.5	62.7	25.3	24.0
CFG	100	87.0	96.9	76.3	72.5
	200	87.1	95.7	65.5	66.3
	400	86.7	92.6	66.9	68.2
AAOS	100	26.5	27.6	119.4	135.7
	200	19.2	19.1	113.8	125.0
	400	15.3	15.1	113.2	121.2

모형이 KGM일 경우 표본크기가 100일 때 $\hat{\beta}_{ols}^{sf}$ 의 MADP를 제외하면 SW 방법의 MADP가 BN 방법의 MADP보다 적게 나타났고 실제효율성이 AAOS일 경우에는 BN 방법의 MADP가 SW 방법의 MADP보다 적게 나타났다. 실제효율성모형이 CFG일 경우에는 SW 방법의 MADP가 BN 방법의 MADP보다 다소 적게하게 나타나 [그림 1]에서의 판단과 일치하는 결과를 보여주고 있다.

SW는 $\hat{\beta}_{boot}^{tr}$ 와 $\hat{\beta}_{ml}^{tr}$ 을 효율성모형의 추정방법으로 제시하면서 종속변수인 DEA효율성점수의 편의가 포함되는 $\hat{\beta}_{ml}^{tr}$ 보다 붓스트랩으로 편의를 수정하는 $\hat{\beta}_{boot}^{tr}$ 이 더 우월함을 주장하였다. 그리고 <표 1>의 SW를 실제효율성모형으로 사용한 모의실험에서 $\hat{\beta}_{boot}^{tr}$ 이 $\hat{\beta}_{ml}^{tr}$ 보다 우월하다는 자신들의 주장과 일치하는 결과를 제시하였다. <표 2>에서 KGM을 실제효율성모형으로 사용한 경우의 결과는 SW의 주장을 뒷받침하고 있으나 CFG와 AAOS를 실제효율성모형으로 사용한 경우는 SW의 주장과 상반된 결과를 보여주고 있다. 이는 $\hat{\beta}_{boot}^{tr}$ 이 $\hat{\beta}_{ml}^{tr}$ 보다 우월하다는 주

장이 항상 성립하지는 않지만 실제효율성모형에서 비대칭확률오차의 좌측절단점(비대칭확률오차가 효율성에 가장 크게 기여하는 수준)이 환경변수의 상태에 따라 달라지는 경우에는 지지될 수 있음을 보여주는 결과라 할 수 있다.

BN은 $\hat{\beta}_{ols}^{sf}$ 와 $\hat{\beta}_{ml}^{sf}$ 을 효율성모형의 추정방법으로 제시하였다. BN은 $\hat{\beta}_{ols}^{sf}$ 와 $\hat{\beta}_{ml}^{sf}$ 의 성과에 관한 이론적 비교는 제시하지 않았으나 BN을 실제효율성모형으로 사용한 모의실험에서 $\hat{\beta}_{ols}^{sf}$ 와 $\hat{\beta}_{ml}^{sf}$ 이 서로 비슷한 성과를 보였으며 비교대상으로 고려되었던 다른 추정방법들보다는 우월함을 보여주는 결과를 제시하였다. <표 2>에서 AAOS를 실제효율성모형으로 사용한 경우는 $\hat{\beta}_{ols}^{sf}$ 와 $\hat{\beta}_{ml}^{sf}$ 이 서로 비슷한 성과를 보이고 있으며 KGM과 CFG를 실제효율성모형으로 사용한 경우는 $\hat{\beta}_{ols}^{sf}$ 이 $\hat{\beta}_{ml}^{sf}$ 보다 우월한 성과를 보여주고 있다. 이는 $\hat{\beta}_{ols}^{sf}$ 이 실제효율성모형에 관계없이 $\hat{\beta}_{ml}^{sf}$ 와 비슷한 성과를 나타내거나 또는 우월함을 보여주는 결과라 할 수 있다.

5. 결 론

DEA효율성점수의 결정요인 분석방법과 관련하여 최근 Simar and Wilson[37](SW)은 절단회귀모형을 기반으로한 붓스트랩절차를 제시하였고 Banker and Natarajan[6](BN)은 확률프론티어모형으로 나타난 효율성모형을 보통최소제곱법 또는 최대우도법으로 추정하는 방안을 제시하였다. McDonald[30]는 BN의 연구결과에 대해서 DEA효율성 점수의 결정요인 분석방법에 통계적 기반을 제공한 매우 중요한 업적으로 평가한 반면 SW가 제시한 접근방법에 대해서는 가정된 자료생성과정 하에서는 타당할 수도 있지만 가정과 다른 자료생성과정에 대해서 강건하지 못함(not robust)을 지적하였다. 현실의 자료생성과정은 BN 모형 및 SW 모형의 가정과 다를 수 있다. McDonald의 지적이 현실에 적용된 BN 모형과 SW 모형의 강건함을 직접적으로 비교한 것은 아니지만, 현실에 적용할 경우 SW 모형의 성과가 (SW

11) 효율성모형 추정방법의 성과를 평균제곱오차제곱근(RMSE) 또는 평균제곱오차제곱근비율(RMSEP)으로 평가한 경우는 평균절대편차비율(MADP)로 평가한 것과 같은 결과를 나타내었고 여기서는 생략하였다.

모형이 가정한 자료생성과정 하에서의 성과보다) 저하될 것이라는 점을 지적하여 BN 모형의 선택을 간접적으로 권하고 있다. 본 논문에서는 SW 방법과 BN 방법의 이론적 특징 및 장단점을 비교하고 모의실험으로 가정과 다른 자료생성과정에서의 성과를 비교하여 McDonald의 SW 방법에 대한 지적을 체계적으로 검토하였다.

BN 모형은 분리조건의 충족을 위해 투입벡터와 환경변수벡터가 서로 독립임을 가정하고 있으며 이는 환경변수와 무관하게 투입수준을 결정하는 생산단위를 대상으로 BN 모형이 적용되어야 함을 의미한다. SW 모형은 생산단위의 투입의사결정이 환경변수의 영향을 받는 생산단위에도 적용할 수 있다는 장점을 갖는 반면 BN 모형에서 고려하고 있는 대칭확률오차가 포함되어 있지 않아 모형의 적용범위가 제한되는 한계점을 갖는다. 모의실험에서는 가정과 다른 자료생성과정(실제효율성모형)에 관계없이 항상 우월한 성과를 나타내는 효율성모형 추정방법은 없는 것으로 나타났다. 실제효율성모형에 따라 BN 방법이 우월한 경우, SW 방법이 우월한 경우, 그리고 두 방법 모두 성과가 좋지 못한 경우가 나타났다. 이는 자료생성과정이 SW 방법의 효율성모형과 다를 경우 SW 방법의 성과가 BN 모형의 성과보다 상대적으로 적게 저하되는 경우도 있음을 보여주는 결과이다. 본 논문에서 제시한 비교 결과는 DEA 효율성 점수의 결정요인을 분석하고자 하는 연구자들에게 방법선택에 활용할 수 있는 유용한 정보가 될 것으로 기대한다.

참고 문헌

- [1] Aigner, D., C.A.K. Lovell, and P. Schmidt, "Formulation and Estimation of Stochastic Frontier Production Function Models," *Journal of Econometrics*, Vol.6, No.1(1977), pp.21-37.
- [2] Aly, H.Y., R. Grabowski, C. Pasurka, and N. Rangan, "Technical, Scale, and Allocative Efficiencies in U.S. Banking : An Empirical Investigation," *The Review of Economics and Statistics*, Vol.72, No.2(1990), pp.211-218.
- [3] Alvarez, A., C. Amsler, L. Orea, and P. Schmidt, "Interpreting and Testing the Scaling Property in Models where Inefficiency Depends on Firm Characteristics," *Journal of Productivity Analysis*, Vol.25, No.3(2006), pp.201-212.
- [4] Banker, R.D., "Maximum Likelihood, Consistency and Data Envelopment Analysis : A Statistical Foundation," *Management Science*, Vol.39, No.10(1993), pp.1265-1273.
- [5] Banker, R.D., A. Charnes, and W.W. Cooper, "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis," *Management Science*, Vol.30, No.9 (1984), pp.1078-1092.
- [6] Banker, R.D. and R. Natarajan, "Evaluating Contextual Variables Affecting Productivity Using Data Envelopment Analysis," *Operations Research*, Vol.56, No.1(2008), pp.48-58.
- [7] Barros, C.P. and P.U.C. Dieke, "Measuring the Economic Efficiency of Airports : A Simar-Wilson Methodology Analysis," *Transportation Research Part E : Logistics and Transportation Review*, Vol.44, No.6(2008), pp.1039-1051.
- [8] Bierens, H.J., *Topics in Advanced Econometrics*, Cambridge University Press, Cambridge, UK., 1994.
- [9] Bjurek, H., U. Kjulin, B. Gustafsson, and D.L. Bosworth, "Efficiency, Productivity and Determinants of Inefficiency at Public Day Care Centers in Sweden : Comment," *Scandinavian Journal of Economics*, Vol.94, No. Supplement (1992), pp.S173-S191.
- [10] Byrnes, P., R. Fare, S. Grosskopf, and C.A.K. Lovell, "The Effect of Unions on Productivity: U.S. Surface Mining of Coal," *Management Science*, Vol.34, No.9(1988), pp.1037-1053.

- [11] Camanho, A.S., M.C. Portela, and C.B. Vaz, "Efficiency Analysis Accounting for Internal and External Non-discretionary Factors," *Computers and Operations Research*, Vol.36, No.5 (2009), pp.1591-1601.
- [12] Carrington, R., N. Puthuchery, D. Rose, and S. Yaisawarng, "Performance Measurement in Government Service Provision : The Case of Police Services in New South Wales," *Journal of Productivity Analysis*, Vol.8, No.4(1997), pp.415-430.
- [13] Caudill, S.B., M.F. Jon, and D.M. Gropper, "Frontier Estimation and Firm-Specific Inefficiency Measures in the Presence of Heteroscedasticity," *Journal of Business and Economic Statistics*, Vol.13, No.1(1995), pp.105-111.
- [14] Charnes, A., W.W. Cooper, and E.L. Rhodes, "Measuring the Efficiency of Decision Making Units," *European Journal of Operational Research*, Vol.2, No.6(1978), pp.429-444.
- [15] Chilingirian, J.A., "Evaluating Physician Efficiency in Hospitals : A Multivariate Analysis of Best Practices," *European Journal of Operational Research*, Vol.80, No.3(1995), pp.548-574.
- [16] Coelli, T.J., D.S.P. Rao, C.J. O'Donnell, and G. E. Battese, *An Introduction to Efficiency and Productivity Analysis*, 2nd ed., New York : Springer Science, 2005.
- [17] Farrell, M.J., "The Measurement of Productive Efficiency," *Journal of the Royal Statistical Society, Series A, General*, Vol.120, No.3(1957), pp.253-281.
- [18] Glass, J.C., D.G. McKillop, and S. Rasaratnam, "Irish Credit Unions : Investigating Performance Determinants and the Opportunity Cost of Regulatory Compliance," *Journal of Banking and Finance*, Vol.34(2010), pp.67-76.
- [19] Greene, W.H., "Maximum Likelihood Estimation of Econometric Frontier Functions," *Journal of Econometrics*, Vol.13, No.1(1980), pp.27-56.
- [20] Greene, W.H., *LIMDEP Version 9/0 : Econometric Modeling Guide* Vol.2, New York : Econometric Software, Inc., 2007.
- [21] Greene, W.H., *Econometric Analysis 6th Edition*, NJ : Prentice Hall., 2008
- [22] Grosche, P., "Measuring Residential Energy Efficiency Improvements with DEA," *Journal of Productivity Analysis*, Vol.31, No.2(2009), pp.87-94.
- [23] Grosskopf, S., "Statistical Inference and Nonparametric Efficiency : A Selective Survey," *Journal of Productivity Analysis*, Vol.7, No.2/3 (1996), pp.161-176.
- [24] Gstach, D., "Another Approach to Data Envelopment Analysis in Noisy Environments : DEA+," *Journal of Productivity Analysis*, Vol.9, No.2(1998), pp.161-76.
- [25] Kneip, A., B.U. Park, and L. Simar, "A Note on the Convergence of Nonparametric DEA Estimators for Production Efficiency Scores," *Econometric Theory*, Vol.14, No.6(1998), pp.783- 793.
- [26] Kravtsova, V., "Foreign Presence and Efficiency in Transition Economies," *Journal of Productivity Analysis*, Vol.29, No.2(2008), pp. 91-102.
- [27] Kumbhakar, S.C., S. Ghosh, and J.T. McGuckin, "A Generalized Production Frontier Approach for Estimating Determinants of Inefficiency in U.S. Dairy Farms," *Journal of Business and Economic Statistics*, Vol.9, No.3(1991), pp.279-286.
- [28] Latruffe, L., S. Davidova, and K. Balcombe, "Application of a Double Bootstrap to Investigation of Determinants of Technical Efficiency of Farms in Central Europe," *Journal of Productivity Analysis*, Vol.29, No.2(2008), pp.183-191.

- [29] Makhorn, A., *GNU Linear Programming Kit : Reference Manual* (Version 4.39), Boston : Free Software Foundation, Inc., 2008.
- [30] McDonald, J., "Using Least Squares and Tobit in Second Stage DEA Efficiency Analyses," *European Journal of Operational Research*, Vol.197, No.2(2009), pp.792-798.
- [31] Muñiz, M., J. Paradi, J. Ruggiero, and Z. Yang, "Evaluating Alternative DEA Models used to Control for Non-discretionary Inputs," *Computers and Operations Research*, Vol.33, No.5 (2006), pp.1173-1183.
- [32] Nyman, J.A. and D.L. Bricker, "Profit Incentives and Technical Efficiency in the Production of Nursing Home Care," *Review of Economics and Statistics*, Vol.71, No.4(1989), pp.586-594.
- [33] Picazo-Tadeo, A.J., and A. Garcia-Reche, "What Makes Environmental Performance Differ Between Firms? Empirical Evidence from the Spanish Tile Industry," *Environment and Planning A*, Vol.39, No.9(2007), pp.2232- 2247.
- [34] Ray, S.C., "Resource-Use Efficiency in Public Schools : A Study of Connecticut Data," *Management Science*, Vol.37, No.12(1991), pp.1620-1628.
- [35] Sexton, T.R., S. Sleeper, and R.E. Taggart, Jr., "Improving Pupil Transportation in North Carolina," *Interfaces*, Vol.24, No.1(1994), pp.87-103.
- [36] Shephard, R.W., *Cost and Production Functions*, Princeton, NJ : Princeton University Press, 1953.
- [37] Simar, L. and P.W. Wilson, "Estimation and Inference in Two-Stage, Semi-Parametric Models of Production Processes," *Journal of Econometrics*, Vol.136, No.1(2007), pp.31-64.
- [38] Simar, L. and P.W. Wilson, "Statistical Inference in Nonparametric Frontier Models : Recent Developments and Perspectives," in H.O. Fried, C.A.K. Lovell, and S.S. Schmidt, eds., *The Measurement of Productive Efficiency and Productivity Growth*, New York : Oxford University Press, 2008, pp.421-521.
- [39] StatsCorp., *Stata Statistical Software : Release 10.0* College Station, TX : Stata Corporation, 2007.
- [40] Syrjanen, M.J., "Non-discretionary and Discretionary Factors and Scale in Data Envelopment Analysis," *European Journal of Operational Research*, Vol.158, No.1(2004), pp.20-33.
- [41] Venables, W.N., D.M. Smith and the R Development Core Team, *An Introduction to R-Notes on R : A Programming Environment for Data Analysis and Graphics*(Version 2.9.0), 2009, <http://cran.r-project.org/>.
- [42] Wang, H.-J., and P. Schmidt, "One-Step and Two-Step Estimation of the Effects of Exogenous Variables on Technical Efficiency Levels," *Journal of Productivity Analysis*, Vol.18, No.2 (2002), pp.29-144.