

프론티어 모델을 이용한 수익성분석

임승범¹⁾, 강창완²⁾, 서명록³⁾, 최용석⁴⁾

요 약

최근 활발히 행하여지는 금융 CRM(Customer Relationship Management)은 은행의 수익성 제고를 위하여 개별고객의 주요특성(거래이력, 가치 등)을 파악하고, 이를 근거로 유사한 고객들을 분류하여 고객관리 방안을 찾는데 그 목적이 있다. 본 연구에서는 B은행의 실제 CRM을 통하여 수익성을 높일 수 있는 마케팅 시사점을 도출하고자 하며 이러한 마케팅의 도출과 목표가 되는 고객을 어떻게 선정할 것인가의 질문에 대한 방법으로 계량경제학 분야에서 기업단위연구의 생산효율성을 측정하기 위하여 사용되어지고 있는 SFM(Stochastic Frontier Model)과 OLS(Ordinary Linear Model)의 방법을 사용하였다.

주요용어 : Frontier Model, OLS, CRM.

1. 서론

최근 들어 금융, 보험, 카드회사와 그 이외의 다양한 분야에서 고객관계관리가 이루어지고 있다. 이는 국내 대부분의 기업들이 성숙기를 맞이하고 있기 때문에 CRM을 통하여 기존 고객에 대한 이해도 증진 및 충성도를 제고시켜 수익성을 극대화하기 위한 노력을 기울이고 있음을 보여주고 있다. 금융권 기업의 CRM에서도 수익성 관점에서 목표마케팅, 고객세분화, 신용평가 및 이탈정도 등에 관심을 두고 있으며 이러한 분석을 통하여 수익성을 높일 수 있는 마케팅의 시사점을 도출하려고 한다. 일반적으로 마케팅 시사점의 도출은 유사한 특성을 가진 고객들 사이에서 나타나는 평균적인 경향과 각 개인별 실제 경향을 비교함으로써 그 잠재력을 파악하고 이를 바탕으로 마케팅 활동이 수행되어야 한다. 그리고 마케팅 내용이 결정되어지고 나면 그러한 활동의 목표가 되는 고객을 선별하여야 하는데 이러한 마케팅의 도출과 목표가 되는 고객을 선정하는데 있어서 어려움이 있다. 이에 본 논문에서는 마케팅 시사점 도출과 목표고객 선정에 있어 은행의 수익자료를 기준으로 설명하고자 한다. 은행의 수익자료와 고객의 특성자료로서 모형을 구축하여 예측되는 수익을 나타내고, 입력변수가 유사한 집단 내에서 예측되는 수익보다 적은 수익을 주는 정도를 측정하고자 한다. 이렇게 측정되어진 수익의 차이는 같은 속성을 가지는 다른 고객에 비하여 비효율적인 수익을 제공하고 있다고 할 수 있는데, 다르게 말하면 그 차이만큼 더 줄 수도 있다고 볼 수 있다. 이러한 차이를 수익에 대한 추가제공 가능성(up-selling potential)이라고 한다. 이와 같은 조건에서 추가제공 가능성을 측정하는 모형으로 계량경제학 분야에서 기업단위연구 중 생산의 효율성을 측정하기 위하여 사용되어 지는 SFM 방법을 적용하고, 이를 전통적인 회귀분석(OLS)에 의한 방법과 비교하였다. SFM방법에서는 주어진 동일한 입력에 최적의 산출량(Frontier)을 추정하는 모형으로서, 이를 은행의 수익자료에 적용하여 보면 유사한 고객들간의 최적의 수익금액을 Frontier로 볼 수 있다. 여기서 고객별 실

- 1) 동의대학교 자연과학대학 정보통계학과 석사과정
- 2) 동의대학교 자연과학대학 정보통계학과 교수
- 3) 부산대학교 자연과학대학 통계학과 석사과정
- 4) 부산대학교 자연과학대학 통계학과 교수

제 수익금액이 Frontier보다 적은 경우를 수익에 대하여 추가제공 가능성이 있다고 볼 수 있으며 이러한 모형으로부터 각 고객별 추가제공 가능성을 추정하고, 순서화하여 그 가능성이 큰 고객부터 마케팅의 목표고객으로 선정할 수 있다. 이러한 방법으로 금융권 기업에서의 고객의 수익금액과 영향변수들(인구통계적 정보, 은행이용정보 등)을 통하여 수익의 증가를 기대할 수 있는 마케팅모형을 제공하고자 한다.

2. OLS에 의한 수익의 추가제공 가능성의 추정

각 고객별 수익금액과 영향변수에 의한 마케팅모형은 전통적인 접근 방법으로서 회귀분석에 의한 모형이 있으며 이의 일반적인 형태는 다음과 같이 표현된다.

$$Y_i = f(X_i) + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2) \quad (1)$$

여기서 Y_i 는 수익금액에 대한 i 번째 고객의 관측치를 나타내며, X_i 는 $1 \times k$ 벡터로서 총 K 개의 설명변수에 대한 i 번째 고객의 관측치를 나타낸다. ε_i 는 에러항목으로서 일반적으로 평균 0과 표준편차 σ 를 가진 정규분포를 가정하며, 함수 $f(X_i)$ 는 선형이며 회귀식의 모수는 OLS에 의하여 추정된다. (1)과 같은 수익금액의 모형이 구축되는 경우 추정된 식을 바탕으로 수익금액의 예측값 \hat{Y}_i 과 실제값 Y_i 를 비교하여 만약 예측값이 측정값에 비하여 높다면 그 고객은 예상했던 것에 비하여 수익금액이 낮은 것으로 판단한다. 이와 같은 방법으로 $\hat{Y}_i - Y_i$ 로서 수익의 추가제공 가능성을 추정하고 이러한 차이 값이 높은 고객의 순서대로 수익금액 증대를 위한 마케팅활동의 목표고객을 선정한다.

3. SFM에 의한 수익의 추가제공 가능성의 추정

SFM은 개별고객의 수익에 대한 추가제공 가능성을 유사한 특성을 가진 고객을 비교하여 그들에게서 기대 되어지는 수익(Frontier에 근접한 고객)과 기대 수익에 미달하는 고객을 판별함으로써 그러한 가능성을 측정하는 방법이다. 고객을 판별함에 있어서 사용되어지는 중요한 개념이 Frontier인데 이것은 OLS 방법에서의 회귀모형과 유사 하지만 비정규적이며 비대칭적인 에러항목이 추가된 형태라고 할 수 있는데(Schmit, P. and Lovell, C.A.K, 1979) 이를 모형화 하면 다음과 같다.

$$Y_i = g(X_i) - u_i + v_i, \quad v_i \sim N(0, \sigma^2) \quad (2)$$

OLS에 의한 식 (1)과 SFM에 의한 (2)의 차이는 v_i 이외에 다른 에러항목인 u_i 가 추가되었다는 점인데, u_i 는 확률변수로서 항상 양의 값을 가진다고 가정된다. 이는 실제값 Y_i 가 u_i 에 의하여 Frontier인 $g(X_i) + v_i$ 에 비하여 같거나 적어도 아래에 위치하도록 한다. 따라서 u_i 는 개별 고객이 Frontier에서 미달하는 정도 즉 수익의 추가제공 가능성을 나타낸다. u_i 는 금융기업의 통제범위 내에 존재하고 있는 요소로서 금융기업의 통제범위 밖에 존재하는 요소를 나타내는 v_i 와 구분되며, u_i 와 v_i 는 서로 독립적인 것으로 가정된다. 여기서 u_i 를 최대우도추정법에 의하여 추정하는 경우 모수적 접근방법은 u_i 에 대한 분포를 가정하여 각종의 모수를 추정하게 된다. u_i 에 대한 분포가정으로 대표적인 것은 half-normal(Aigner et al. 1977), exponential (Meesun and Van den Broech, 1977), truncated(at 0) normal(Stevenson, 1980) 분포 등이 사용되어진다.

4. SFM 사용의 전제조건

상기 절들에서 제시한 각각의 방법을 통하여 목표고객을 선정하는 경우, OLS는 $\hat{Y}_i - Y_i$ 를 계산하여 이 값이 높은 순서대로 선택하며, SFM의 경우 추가제공 가능성 u_i 가 가장 높은 고객의 순서대로 선택하게 된다. 따라서 OLS에 의한 추정식과 SFM에 의한 Frontier가 서로 평행한 경우에는 어떠한 방법을 사용하더라도 동일한 고객이 목표고객으로 선정된다. 그러나 u_i 가 이분산적인 경우, 즉 독립변수로 사용되는 하나 또는 이상의 변수들에 따라 체계적으로 변화하는 경우에는 기울기가 평행하지 않게 되는데 이러한 현상은 대규모의 횡단자료의 경우에 흔히 나타나는 것이라고 할 수 있다. u_i 가 이분산적인가에 대한 검토는 OLS에 의하여 계산된 잔차의 왜도(skewness)가 긍정적으로 왜곡되어 있다면 이는 자료의 등분산성을 의미하는 $\sigma=0$ 를 의미하게 되어 이분산적이지 않음을 나타내고 OLS 방법에 의해서만 최우추정치가 제시되고 SFM에 의한 추정치는 제시되지 않는다(Waldman 1982). 그러나 SFM이 정의되는 경우 먼저 각각의 대안적인 모델에 대한 모수 β 가 추정되며 수익의 추가제공 가능성인 u_i 가 계산되어 지는데 이는 전체 에러항목($\epsilon_i = v_i - u_i$)에 대한 u_i 의 조건부확률로 계산되는데(Johndrow et al. 1982), 이러한 조건부확률의 평균값이 u_i 의 추정치로 산출된다.

5. 적용사례

2002년 8월부터 12월까지 이루어진 국내 특정 B은행의 CRM에서 수신에 의한 수익을 주는 고객의 모집단 22만 명의 고객자료 중에서 외국인과 수익금액의 누락값 또는 오류가 있는 자료 그리고 수익금액이 음의 값을 갖는 자료를 제외하고 최종적으로 69367명의 고객을 대상으로 분석을 하였다. 추출된 69367명의 자료들 중에서 수익자료는 2001년 8월부터 2002년 7월까지의 자료를 기준으로 하였고 나머지 기타 독립변수들은 고객데이터베이스에서 1~2년을 기준으로 작성하였다. 이러한 수익자료와 가장 관련 있는 변수를 채택하게 위하여 의사결정나무와 다중 회귀분석을 실시하여 공통적으로 유의한 변인들만 채택하였으며 본 연구에 사용된 각종 변수들에 대한 측정방법과 기초통계분석 결과를 <표 1>에 제시하였다.

<표 1> 변수의 요약 및 기초분석 결과

변수	변수설명	Mean ± S.D
총 거래건수	1년간 은행과 거래한 총 건수 (방문, ATM, 인터넷, 폰뱅킹 등)	37.43 ± 37.76
실제거래기간	1년 중 실제 거래가 존재하였던 달의 횟수 (단위: 개월)	11.64 ± 1.51
카드보유기간	2년간 카드 보유기간 (단위:개월)	5.36 ± 8.88
평잔	1년간 평균 잔액 (단위:원)	48323086.33 ± 85738327.40
수익금액	1년간 총 수익 금액에 로그를 취한 값 (단위:원)	10.88 ± 1.34

종속변수는 수익금액으로서 연간 총 수익금액에 로그를 취한 값을 사용하였는데 최대 14.95와 최소 7.06의 분포를 보이고 있으며, 총 거래건수의 경우에는 최대 649건에서 최소 0건의 분포를 보이고 있다. 실제거래기간의 경우 1년을 기준으로 정하여 12개월 중 실제로 거래가 존재하였던 달을 측정하였는데 최대 12개월에서 최소 1개월의 분포를 나타내고 있다. 카드보유기간은 측정시점으로부터 카드를 보유하고 있었던 기간을 측정한 것으로 최대 24개월에서 최소 0개월의 분포를 보이고 있으며, 평잔은 최대 887,306,912원에서 최소 0원까지의 분포가 나타나고 있다. <표 2>에서는 <표 1>의 자료를 사용하여 앞에서 언급하였던 OLS 및 u_i 의 분포가정별

SFM의 결과를 비교하여 제시하였다. 이 결과는 경제학에서 사용되는 LIMDEP 프로그램의 결과로서 u_i 가 이분산적인 경우 OLS 및 SFM 방법에 의한 모수가 추정되어지고, 그렇지 못한 경우 OLS에 의한 결과만 제시되어진다. 본 연구에서는 u_i 가 이분산적으로 나타나 OLS 및 SFM에 의한 추정결과 모두가 출력되었는데 여기서 사용되어진 독립변수 모두가 $p=0.0001$ 에서 유의한 것으로 나타났다. 또한 OLS에 대한 SFM 적용의 타당성분석 결과로서 로그우도값을 비교하였는데 SFM의 3가지 방법 모두가 OLS에 의한 로그우도값인 -98715.03 보다 크게 나타나 적합도가 좋은 것으로 나타났다. 이러한 결과는 모수의 수를 고려한 적합도 평가 기준인 AIC(Akaike Information Criteria)와 BIC(Bayesian Information Criteria)에서도 OLS의 경우 각각 -98720.03, -98727.13으로서 SFM에 의한 3가지 방법보다 적합도가 떨어지는 것으로 나타났다.

<표 2> OLS 및 SFM에 의한 추정결과

	OLS		SFM					
	β값	p-값	Half-Normal		Exponential		Truncated Normal	
	β값	p-값	β값	p-값	β값	p-값	β값	p-값
상수항	8.9930	0.0001	9.7852	0.0000	9.4673	0.0000	9.7852	0.0000
총 거래건수	0.0043	0.0001	0.0043	0.0000	0.0043	0.0000	0.0043	0.0000
실제거래기간	0.0897	0.0001	0.0897	0.0000	0.0897	0.0000	0.0897	0.0000
카드보유기간	0.0573	0.0001	0.0573	0.0000	0.0573	0.0000	0.0573	0.0000
평균	0.7584E-8	0.0001	0.7728E-8	0.0000	0.7694E-8	0.0000	0.7728E-8	0.0000
σ_u^2	-	-	0.9858	-	0.2249	-	0.9858	-
σ_v^2	-	-	0.6501	-	0.7833	-	0.6501	-
$\lambda (= \sigma_u/\sigma_v)$	-	-	1.2314	0.0000	-	-	1.2314	0.0000
$\sigma (= \sqrt{\sigma_u^2 + \sigma_v^2})$	-	-	1.2790	0.0000	-	-	1.2790	0.0000
μ/σ_u	-	-	-	-	-	-	-0.81E-18	1.0000
θ	-	-	-	-	2.1084	0.0000	-	-
σ_r	-	-	-	-	0.8850	0.0000	-	-
-Log Likelihood	-98715.03		-98424.32		-98776.25		-98424.33	
-AIC	-98720.03		-98431.32		-98784.25		-98432.33	
-BIC	-98727.13		-98441.26		-98795.61		-98443.69	

AIC = Log Likelihood - (Number of Estimated Parameter)

BIC = Log Likelihood - (1/2 × Log N) × (Number of Estimated Parameter)

여기서 SFM방법에 의한 추정이 가능하다는 것은 수익의 추가제공 가능성인 u_i 가 존재하여 은행 고객의 수익금액 변동이 있다는 것을 나타낸다고 할 수 있다. 결과적으로 고객의 수익금액에 대한 모형을 바탕으로 수익의 추가제공 가능성을 추정하는데는 OLS(회귀분석)에 의한 방법 보다는 SFM 방법이 더욱 유용하다고 할 수 있다. 특히 SFM 방법 중에서도 u_i 의 분포를 Half-Normal로 가정하여 모수 및 u_i 를 추정한 모형이 가장 좋은 것으로 나타났다. <표 2>에서 추정된 수익금액에 대한 독립변수의 효과를 살펴보면 총 거래건수, 실제거래기간, 카드보유기간 끝으로 평균이 증가하는 경우 수익금액은 증가하는 것으로 나타났다.

수익의 추가제공 가능성인 u_i 의 분포가정별 계산방법과 로그우도함수는 본 연구에서 최적 모형으로 채택되어진 Half-Normal에 의한 계산법만 제공하도록 한다. u_i 의 분포를 Half-Normal 이라고 가정한 계산식은 예러항목 ε_i 에 대한 조건부확률의 평균값으로 계산되어지는데 (Jondrow et al. 1982)

$$E[u|\varepsilon] = \frac{\sigma\lambda}{1+\lambda^2} \left[\frac{\phi(\varepsilon\lambda/\sigma)}{1-\Phi(\varepsilon\lambda/\sigma)} - \varepsilon\lambda/\sigma \right] \quad (3)$$

위와 같이 나타나며 여기서 $\lambda = \sigma_u/\sigma_v$ 이고, $\phi(\cdot)$ 는 표준정규분포의 확률밀도함수이며, $\Phi(\cdot)$ 는

표준정규분포의 누적밀도함수를 의미한다. 또한 로그우도함수는 아래와 같이 표현된다.

$$\sum_i \ln L_i = - \ln \sigma - 1/2 \ln(2/\pi) - (\epsilon_i/\sigma)^2 + \ln \Phi[-\epsilon_i/\sigma] \quad (4)$$

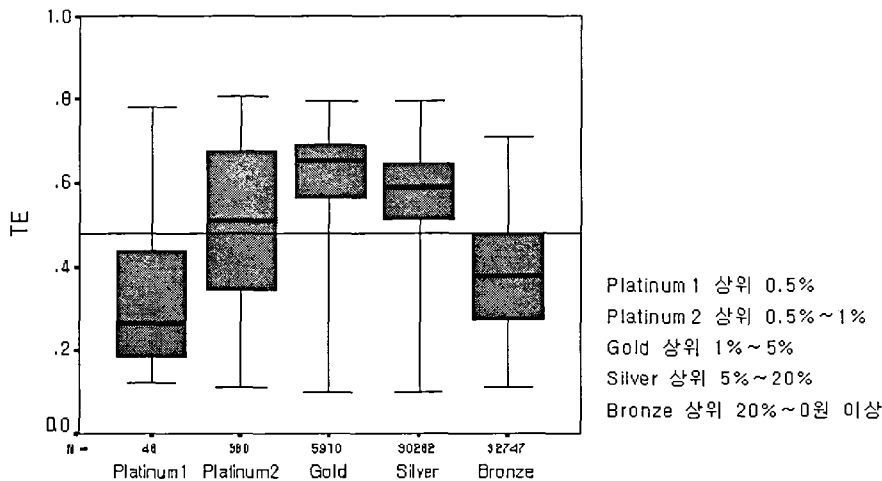
Farrell(1957)에 의하면 기업의 효율성(firm efficiency)은 TE(technical efficiency)와 AE(allocative efficiency)라는 2가지 측면에서 측정할 수 있는데 TE는 주어진 입력에 최고의 산출량을 얻는데 관련된 효율성이고, AE는 입력에 사용될 최적의 크기에 관련된 효율성을 나타낸다. 여기서 본 연구의 관심은 금융권 기업에 관련되어 주어진 입력에 대한 최적의 수익(frontier)을 나타내는 정도 또는 최적의 수익으로부터 미달되는 정도를 측정하는 것이다. 그러므로 TE는 개별 고객에 대하여 관측된 수익금액이 최적의 수익금액으로부터 어느 정도 미달되는지를 또는 수익에 대하여 추가제공 가능성을 측정하는 지표가 되며 다음과 같이 표현되어진다(Aigner, 1977).

$$TE_i = \frac{\text{observed output}}{\text{frontier output}} = y_i / \exp(X_i\beta) = \exp(X_i\beta - u_i) / \exp(X_i\beta) = 0 < \exp(-u_i) \leq 1 \quad (5)$$

여기서 frontier output에 exp를 하는 이유는 종속변수 y에 로그를 취하여서 추정된 함수이므로 원래대로 환산하기 위해서이다. 식 (5)에서 결국 수익에 대한 추가제공 가능성의 지표인 TE는 u_i 의 함수가 되고 크기는 0에서 1사이의 값을 가지게 되는데 이 값이 작을수록 추가 제공성이 큰 고객을 의미하게 된다. 위의 식 (3)을 통하여 B은행의 고객(N=69,367)에 대하여 u_i 를 계산하였고, 식 (5)를 통하여 각 고객별 TE를 계산하였다. B은행은 개인별 연간 총 수익금액의 크기 별로 고객을 Platinum I ~LeadII 까지 8개 그룹으로 고객을 선별하는데 본 연구에서는 수익금액이 0원 이상인 고객을 대상으로 하였기에 Platinum I ~Bronze까지 5개 그룹의 고객에 대해서 분석하였고 각 그룹별 설명과 TE의 기술통계량 및 상자그림을 <표 3>과 <그림 1>과 같이 나타내었다.

<표 3> 전체 고객 및 그룹별 TE의 기술통계 및 그룹별 up-selling 마케팅의 평균 기대수익

	N	최소값	최대값	평균	표준편차	평균이하n	TE하위10%	평균기대수익(W)
All Customer	69367	0.0982	0.8048	0.4789	0.1507	31918	6997	
Platinum I	48	0.1229	0.7774	0.3163	0.1643	40	23	8,029,631
Platinum II	380	0.1131	0.8048	0.4990	0.1953	168	59	4,997,998
Gold	5910	0.0982	0.7943	0.6032	0.1311	865	229	2,332,439
Silver	30282	0.1040	0.7944	0.5657	0.1068	5783	443	594,807
Bronze	32747	0.1123	0.7089	0.3761	0.1166	25062	6237	33,586



<그림 1> 그룹별 TE의 상자그림

<그림 1>에서 가로 참조선은 전체 고객에서의 TE의 평균을 나타내고 이 값보다 적은 고객의 수는 31918명으로 전체 고객의 46%가 평균 TE보다 낮은 수익을 제공하는 것으로 나타난다. <표 3>에서 평균 TE보다 낮은 고객을 그룹별 비율을 살펴보면 Platinum 1이 83.3%, Platinum 2가 44.2%, Gold 14.6%, Silver 19.1%, Bronze는 76.5%로 나타나 P1, B, 그리고 P2 그룹에서 Frontier 보다 낮은 수익을 주는 고객의 비율이 높다는 것을 알 수 있다. 여기서 전체 고객에 대하여 TE값을 기준으로 하위 10%(6237명)를 심각한 수익미달 고객이라고 판단할 때, 6237명 내 각 그룹별 고객의 수는 <표 3>과 같이 분포되었다. 각 그룹별로 고객에 대한 up-selling 마케팅을 하였을 경우 그룹별 기대되어지는 평균 수익도 계산하였다. 수익 증가를 위한 고객 마케팅을 실시할 경우 마케팅비용의 규모와 마케팅 초점이라는 2가지 면에서 검토되어진다. 우선 마케팅비용의 규모를 고려하는 것인데, B그룹에서 P1그룹으로 갈수록 고객 당 마케팅비용은 큰 폭의 차이로 인하여 P1 그룹의 고객을 많이 선정할 경우 목표고객의 수가 줄어들게 된다. 두 번째는 마케팅의 초점을 수익금액에 둘 것인지 아니면 비효율적인 수익을 주는 고객의 수를 줄이는데 맞출 것인지에 관한 문제이다. 이러한 2가지 문제 중에서 어떤 것이 우선적인지는 금융권 기업마다 당면한 상황이 다르지만 <표 3>과 <그림 1>을 통하여 합리적인 결정에 도움을 줄 수 있을 것이다.

6. 결론

본 연구에서는 금융권 기업의 CRM에서 핵심이 되는 수익금액에 대한 목표마케팅의 시사점 도출 및 목표고객의 선정의 방법으로 통계적 기법인 OLS와 OLS내의 잔차 u_i 와 v_i 로 세분화하여 이 중에서 u_i 의 분포가 규명 가능한 경우 계량경제학의 모형인 SFM의 적용이 가능한 것과 OLS보다 모형의 적합성이 보다 나음을 알 수 있었다. 또한 SFM 방법을 통하여 구하여진 수익에 대한 추가 제공 가능성인 u_i 를 추정할 수 있었고 이러한 결과를 토대로 금융 CRM에서 수익의 증가를 위한 목표마케팅의 시사점 도출 및 대상 고객의 선별이 가능함을 볼 수 있었다.

참고문헌

- [1] Farrell, M. J. (1957), "The Measurement of Productive Efficiency", Journal of the Royal Statistical Society, Series A, General, 120, p. 253-281.
- [2] Aigner, D. J., Lovell, C. A. K. and Schmidt, P. J. (1977), "Formulation and Estimation of Stochastic Frontier Production Function Model", Journal of Econometrics, 6, p.21-37.
- [3] Jondrow, J., C. Lovell, C. A. K., Materov, I. S. and Schmidt, P. (1982), "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model", Journal of Econometrics, 19, p. 233-238.
- [4] Meesun, W. and van den Broech, J (1977), "Efficiency Estimation from Cobb-Douglas Production Function with Composed Error", International Economic Review, 18(2), p.435-444.
- [5] Schmidt, P. and Lovell, C. A. K (1979), "Estimating Technical and Allocative Inefficiency Relative to Stochastic Production and cost Frontiers", Journal of Econometrics, 9, p. 343-366.
- [6] Stevenson, R. E (1980), "Likelihood Functions for Generalized Stochastic Frontier Estimation", Journal of Econometrics, 13, p. 58-66
- [7] 주영혁, 전종근 (2001), "이동통신이용고객 추가이용(up-selling)가능성 및 측정방법에 관한 연구", 마케팅연구, 16(3), p. 147-166.