자원·환경경제연구 제 10권 제 4호
Environmental and Resource Economics Review
Volume 10, Number 4, December 2001 : pp. 515~545

조건부가치측정모형의 최소절대편차추정*

김 동 일**

〈차 례〉

Ⅰ. 머 리 말

환경재의 실제적 사용가치는 시장의 데이터를 통해서 측정할 수 있다. 비록 환경재가 시장에서 거래되지 않는 경우라 하더라도, 헤도닉가격모형, 여행비용모형, 회피행위모형 등, 시장에서 현시된 선호행위를 이용한 간접시장접근법으로 가치측정이 가능하다. 그러나 존재가치나 선택가치 등과 같이 시장의 행위에 반영되지 않는 환경재의 비사용가치는 시장의 데이터를 이용하여 가치를 측정

* 유익한 논평을 해 주신 Charles Manski 교수님께 깊이 감사드린다. 하지만 논문에 남은 오류는 모두 저자의 것이다.
** 홍익대학교 법경대학 경제학과.

할 수 없다. 환경재의 비사용가치를 측정할 수 있는 유일한 방법으로 받아들여지고 있는 것이 바로 조건부가치측정법(contingent valuation method)이다. 조건부가치측정은 설문조사를 통해서 이루어진다. 설문조사에서 응답자는 어떤 환경재의 제공과 관련된 가상적 질문에 가상적 선택을 하고, 응답자의 선택은 조건부가치측정모형에 따라 설명된다. 응답자가 그 환경재에 부여하는 가치, 즉 응답자의 지불의사액(willingness-to-pay)은 시장데이터를 통해 직접적으로 관찰되지는 않지만, 조건부가치측정 설문조사에 대한 응답자의 선택을 통해 간접적으로 추정될 수 있다. 조건부가치측정 설문조사는 투표형 질문(referendum question), 지불카드법(payment-card method), 개방형 질문(open-ended question) 등을 사용하여 이루어지는데, 이 연구에서는 투표형 질문과 지불카드법과 같이 응답자가 이산적 선택(discrete choice)을 하게 되는 조건부가치측정의 경우에 논의를 제한하기로 한다.

조건부가치측정모형은 원래 Hanemann (1984)이나 Sellar *et al.* (1986) 등에 의해 임의효용이론(random utility theory)에 근거한 이산선택모형(discrete choice model)으로서 개발되었다. 이산선택모형은 간접효용함수를 추정함으로써, 이산적 선택을 효용의 관점에서 설명하는 것을 목적으로 한다. 하지만 환경경제학에서 조건부가치측정을 통해 연구자가 궁극적으로 얻으려고 하는 것은, 설문조사를 행하여 이끌어낸 응답자의 선택을 설명하는 것이 아니라, 그 정보를 이용하여 응답자의 지불의사액을 추정하는 것이다. 이산선택모형으로 개발된 조건부가치측정모형에서는 지불의사액에 대한 추정이 간접효용함수의 추정 후에 이차적으로 이루어지고, 이 점 때문에 간접효용함수의 형태나 임의오차의 분포에 대하여 여러 가지 강한 제약이 주어진다.

이후 Cameron and James (1987)와 Cameron (1988)은 지불의사액을 간접효용의 차이에서 간접적으로 정의하는 것이 아니라 보상변화(compensating variation)로 보다 직접적으로 정의하고, 지불의사액으로 정의되는 조건부가치측정모형을 제시하였다. 이 경우 조건부가치측정 설문조사의 응답을 설명하는 조건부가치측정모형을 추정하는 것이 바로 지불의사액을 추정하는 것이 되기 때

문에, 간접효용함수를 적분하는 복잡한 우회적인 추정이 필요 없게 되어 지불의사액에 대한 추정이 보다 간단명료해졌다. 이전의 임의효용이론의 조건부가치측정모형에 비해 지불의사액의 조건부가치측정모형이 가지는 또 다른 중요한 장점은 설문조사에서 여러 가지 질문과 응답이 있을 때 이 정보를 보다 직접적으로 결합할 수 있다는 것이다. 이 연구는 지불의사액의 조건부가치측정모형을 통계학적으로 추정하는 문제에 관한 것이며, 용어의 단순화를 위해 이 연구에서는 조건부가치측정모형을 특별히 지불의사액의 조건부가치측정모형을 지칭하는 것으로 사용한다.

Cameron and James (1987)와 Cameron (1988) 이후 조건부가치측정모형은 대개 최우추정법(maximum likelihood method)으로 추정되었다.[1] 최우추정은 분포에 대한 강한 가정을 필요로 하는데, 추정량은 가정이 맞으면 일치적이고 효율적이나(consistent and efficient), 대신 분포가정에 민감하게 의존하여 분포가정의 오류에 강건하지(lobust) 못하다는 단점을 가진다. 이 연구는 Manski (1975, 1985)와 Lee (1992)의 이산선택모형의 준모수적 추정에 대응하는, 조건부가치측정모형의 새로운 추정법을 소개한다. 최소절대편차(least absolute deviations: LAD)추정이라 부를 이 추정은 분포에 대해 매우 약한 가정만을 요구한다. 분포에 대한 함수형태의 가정을 필요로 하지 않으며, 알려지지 않은 형태의 이분산성(heteroskedasticity)의 경우와 같이 독립적이고 동일한 분포가 아닌 경우에도 추정이 가능하다. 이러한 약한 분포가정 때문에 LAD추정은 최우추정보다 덜 효율적이지만, 또 바로 그 이유 때문에 분포가정의 오류에 대하여 보다 강건하다.

선형모형을 가정할 경우, 이산선택모형의 모수는 완전히 식별되지 않고, 단지 척도까지만 식별된다(identified up to scale). 그러나 모형이 선형이더라도, 조

1) 또 다른 접근으로서 Kriström (1990), Carson *et al.* (1992, 1994), Habb (1997) 등의 조건부가치측정모형의 비모수적 추정이 있다. 하지만 비모수적 추정의 경우, 지불의사액에 대한 경제학적 설명을 할 수 없고, 지불의사액이 개인의 특성에 의존하는 조건부분포를 가질 때 이를 추정에 반영할 수 없다는 단점을 가진다.

건부가치측정모형의 모수는 척도를 포함하여 완전히 식별될 수 있다. Cameron and James (1987)와 Cameron (1988)은 임의효용이론의 조건부가치측정모형을 최우추정법으로 추정할 때는 모수를 척도까지만 식별할 수 있으나, 지불의사의 조건부가치측정모형을 최우추정법으로 추정할 때는 설문조사에서의 제시금액의 존재로 인해 모수를 완전히 식별할 수 있음을 지적하였다. 조건부가치측정모형의 LAD추정이 Manski (1975, 1985)와 Lee (1992)의 이산선택모형의 준모수적 추정과 구별되는 근본적인 차이는, 바로 조건부가치측정모형과 이산선택모형이 가지는 식별성에서의 차이에서 비롯된다. 기술적인 면에서 조건부가치측정모형의 LAD추정에서 모수의 완전식별성(full identification)이 이루어지는 조건에 대한 이론적인 증명이 새롭게 필요하며, 또한 LAD추정을 위한 계산법에도 수정이 요구된다.

이 연구는 모두 여섯 절로 구성된다. 제Ⅱ절에서는 조건부가치측정모형의 LAD추정량을 정의하고, 추정량의 완전식별성과 강일치성(strong consistency)을 증명하며, 제시금액의 영향을 검토한다. 투표형 질문의 조건부가치측정에서는, 추가제시금액(follow-up offer)이 첫 번째 질문에 대한 대답에 의존하는데, 추가제시금액의 이러한 내생성의 영향을 LAD추정과 연관하여 논의한다. 제Ⅲ절에서는 LAD추정의 계산전략을 설명하고, 모의실험 연구를 통해 소표본 성질과 추가질문의 효율성 기여 등의 추정량의 운영특성을 평가한다. 제Ⅳ절에서는 연구자에게 알려지지 않은 이분산성이 존재할 때 LAD추정과 최우추정의 편의와 효율성을 모의실험 연구를 통해 비교한다. 제Ⅴ절에서는 LAD추정이 실제로 어떻게 적용되는지를 보여 주는 간단한 예가 제시된다. 마지막 절에 맺음말이 따른다.

Ⅱ. 최소절대편차추정

1. 정의와 해석

어느 환경재의 제공에 대하여 어느 개인이 일정한 금액, y^*를 기꺼이 지불할 용의가 있다고 가정하자. 그리고 그 지불의사액은 관찰되는 변수들 $x \in R^k$와 관찰되지 않는 임의오차 $\varepsilon \in R$의 선형함수로 다음과 같이 주어진다고 가정하자.

$$y^* = x'\beta + \varepsilon \tag{1}$$

여기서 $\beta \in B$는 모수로서 모수집단 $B \in R^k$에 속한다. 개인의 지불의사액은 관찰되지 않으며, 따라서 지불의사액모형 (1)을 직접적으로 추정하는 것은 불가능하다. 대신 그 개인에게 환경재가 제공되면 $t > 0$의 금액을 기꺼이 지불할 용의가 있는지를 묻는 설문조사를 한다고 가정하자. 설문에 대해 만약 지불의사액이 제시금액과 같거나 그보다 크면 "예", 그렇지 않으면 "아니오"라고 대답한다면, 설문조사에 대한 응답은 다음과 같이 주어진다.

$$\begin{aligned} y &= \begin{cases} 1 & \text{if} \quad y^* \geq t \\ 0 & o.w. \end{cases} \\ &= 1(y^* \geq t) \\ &= 1(x'\beta + \varepsilon \geq t) \end{aligned} \tag{2}$$

여기서 1( )은 표시함수(indicator function)로서 어떤 사건 A가 일어나면 $1(A)=1$, 그렇지 않으면 $1(A)=0$이며, $y=1$은 설문조사의 질문에 대한

응답이 "예"인 경우, $y=0$은 "아니오"인 경우를 각각 의미한다. 식 (2)는 조건부가치측정모형으로서 조건부가치측정 설문조사의 응답을 설명한다. 임의효용이론의 조건부가치측정모형과는 달리, 이 모형은 조건부가치측정 설문조사의 응답을 지불의사액과 제시금액을 비교하여 설명하고, 조건부가치측정모형 (2)를 추정하는 것이 바로 지불의사액모형 (1)을 추정하는 것이 된다. 응답자의 지불의사액은 직접적으로 관찰되지 않지만, 조건부가치측정의 설문조사에 대한 응답을 통해 드러나는 정보를 통해 간접적으로 추정될 수 있다.

전통적인 선형회귀모형의 최소제곱추정량(least squares estimator)과 마찬가지로 조건부가치측정모형의 LAD추정량은 적합기준(fitting criterion)으로 정의된다. 만약 b를 β의 추정량이라 하면, 설문조사 응답에 대한 예측은 $1(x'b \geq t)$이고 그 오차는 $y-1(x'b \geq t)$이다. 조건부가치측정 설문조사가 응답자 $i=1,\cdots,N$에 대하여 이루어진다면, 설문조사 응답에 대한 절대예측오차의 합 $D_N(b)$는 다음과 같이 주어진다.

$$D_N(b) = \sum_{i=1}^{N} |y_i - 1(x_i'b \geq t_i)| \tag{3}$$

조건부가치측정모형 (2)의 LAD추정량 $\widehat{b_N}$은 설문조사 응답에 대한 절대예측오차의 합을 최소화하는 $b \in B$로 정의되며, 조건부가치측정 설문조사의 응답을 가장 잘 예측한다는 점에서 최적추정량이다.

사실 LAD추정량이란 이름은 적절하지 않을지도 모른다. 왜냐하면, 최소화되는 오차의 놈(norm)은 보다 일반적으로 해석될 수도 있기 때문이다. 한 예를 들어보면 $(y-1(x'b \geq t))^2 = |y-1(x'b \geq t)|^2 = |y-1(x'b \geq t)|$이고, 따라서 제곱오차의 합, $\sum_{i=1}^{N}(y_i - 1(x_i'b \geq t_i))^2$를 최소화하는 추정량 역시 LAD추정량과 일치한다. 일반적으로 어떤 $p>0$에 대해서도 $|y-1(x'b \geq t)|^p = |y-1(x'b \geq t)|$이고, 따라서 오차가 L^p 놈으로 정의되는 한, 예측오차의 합을 최소화하는 추정량은 동일하다.[2)] 조건부가치측정모형의 이 독특한 특

성은 LAD추정량에 또 다른 의미를 부여한다. LAD추정량은 그 이름이 의미하는 바와 같이 L^1 놈에서 가장 최적인 예측변수일 뿐만 아니라, L^p 놈에서도 가장 최적인 예측변수이다. 이를테면, 조건부가치측정모형의 LAD추정은 전통적인 선형회귀모형의 최소제곱추정에 상대되는 것으로 해석될 수 있다.

조건부가치측정 설문조사에서는 흔히 설문응답자에게 여러 개의 질문이 주어지거나 또는 그렇게 해석될 수 있는 경우가 있다. 질문은 투표형 질문에서처럼 차례로 주어질 수도 있고, 지불카드법에서처럼 동시에 주어질 수도 있다. 설문조사에서 $t_1 < t_2 < \cdots < t_J$인 J개의 제시금액에 대한 질문이 주어진다면, 이 J개의 제시금액은 $J+1$개의 서로 중첩되지 않는 구간 (t_0, t_1)과 $\{[t_j, t_{j+1})\}_{j=1}^{J}$을 구성한다. 여기서 $t_0 = -\infty$이고 $t_{J+1} = \infty$이다. 제시금액보다 자신의 지불의사액이 같거나 크면 "예", 그렇지 않으면 "아니오"라고 설문응답자가 대답한다고 가정하면, 그의 응답은 다음과 같이 주어진다.

$$
\begin{aligned}
y_j &= 1(y^* \geq t_j) \qquad\qquad (4)\\
&= 1(x'\beta + \varepsilon \geq t_j)\\
&\quad j = 1, \cdots, J
\end{aligned}
$$

설문조사 데이터 $\{y_i, x_i, t_i\}_{i=1}^{N}$가 조건부가치측정모형 (4)에 따라 주어진다고 하자. 여기서 y_i와 t_i는 벡터로서 $(y_{1i}, \cdots, y_{Ji})$와 $(t_{1i}, \cdots, t_{Ji})$를 각각 나타낸다. β의 LAD추정량 $\widehat{b_N}$은 다음과 같은 표본거리함수(sample distance function), $D_N(b)$를 최소화하는 $b \in B$로 정의된다.

2) 응답 y와 표시함수 $1(\ )$은 1 또는 0의 값을 가지고, 따라서 $|y - 1(x'b \geq t)|$는 0 또는 1의 값을 가진다. 그런데 $p > 0$이면 $|0|^p = |0|$, $|1|^p = |1|$이고, 따라서 $|y - 1(x'b \geq t)|^p = |y - 1(x'b \geq t)|$이다.

$$D_N(b) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{J}|y_{ji} - 1(x_i'b \geq t_{ji})| \tag{5}$$

$$= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{J}\left\{1 - \frac{1}{2}\left[(2y_{ji}-1)\,sgn(x_i'b - t_{ji}) + 1\right]\right\} \tag{6}$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left|\sum_{j=1}^{J}y_{ji} - \sum_{j=1}^{J}1(x_i'b \geq t_{ji})\right| \tag{7}$$

여기서 부호함수 $sgn(\ )$은 $x \geq 0$이면 $sgn(x) = 1$, $x < 0$이면 $sgn(x) = -1$의 값을 가진다고 정의된다.[3] 이 추정량은 표본거리함수를 어떻게 표시하는가에 따라 여러 가지 의미로 해석될 수 있다. 이것은 (5)의 $D_N(b)$를 최소화한다는 점에서 LAD추정량이고, (6)의 $D_N(b)$를 최소화한다는 점에서 최대점수추정량(maximum score estimator)이다. 이산선택모형이나 조건부가치측정모형에서는 추정의 적합지표(indicator of goodness of fit)로 예측오차표(prediction error table)가 종종 제시되는데, LAD추정량은 L^p 놈으로 정의되는 설문응답의 예측오차를 최소화하고 성공적인 예측을 극대화하며, 그러한 점에서 최적이다. 이 추정은 (7)의 $D_N(b)$를 최소화하는 중위값회귀(median regression)로 해석될 수 있는데, 왜냐하면 일정한 조건 아래서 $\sum_{j=1}^{J}1(x_i'\beta \geq t_{ji})$는 $\sum_{j=1}^{J}y_{ji}$의 중위값이기 때문이다.[4] 표본거리함수에 대응하는 모집단함수 $D(b)$는 각각 다음과 같이 주어진다.

$$D(b) = E\left(\sum_{j=1}^{J}|y_j - 1(x'b \geq t_j)|\right)$$

$$= E\left(\sum_{j=1}^{J}\left\{1 - \frac{1}{2}\left[(2y_j - 1)sgn(x'b - t_j) + 1\right]\right\}\right)$$

$$= E\left(\left|\sum_{j=1}^{J}y_j - \sum_{j=1}^{J}1(x'b \geq t_j)\right|\right)$$

---

3) $y^* \in [t_l, t_{l+1})$일 경우, $l = 0, \cdots, J$에 대해서 $\sum_{j=1}^{J}y_j = \sum_{j=1}^{J}1(y^* \geq t_j) = l$이다.

4) 예비정리 1과 부록에 있는 예비정리 1의 증명을 참조하시오.

2. 식별과 강일치성

이산선택모형에서 경계점(threshold points) $\{t_j\}_{j=1}^{J}$는 추정되어야 할 모수이다. 모수 β와 $\{t_j\}_{j=1}^{J}$에 일정한 척도를 곱해도 $\{1(x'\beta \geq t_j)\}_{j=1}^{J}$는 변하지 않고, 따라서 모수는 완전히 식별되지 않고 척도까지만 식별된다(identified only up to scale). 그러나 조건부가치측정모형에서 경계점은 다른 의미를 가진다. 설문조사를 통해 지불의사액을 이끌어내기 위해 경계점의 숫자와 수치는 연구자에 의해 기술적으로 조정될 수 있고, 경계점의 이러한 독특한 성격 때문에 x와 $\{t_j\}_{j=1}^{J}$가 일정한 조건을 만족하면, 모수 β는 완전히 식별될 수 있다.

가정 1 $\{y_i^*, x_i, t_i\}_{i=1}^{N}$은 시불변하며 에르고드성을 가진다(stationary and ergodic).

가정 2 $\beta \in B$이고 $B \subset R^k$는 컴팩트(compact)하다.

가정 3 $Med(\varepsilon \mid x, t) = 0$

가정 4 (x, t_m)분포의 받침(support)은 적어도 어느 한 $t_m \in t$에 대해서는 R^{k+1}의 어떤 정당한 선형부분공간에도 담겨지지 않는다.

가정 5 $\beta_1 \neq 0$이고, (x_2, t)가 주어진 x_1의 조건부분포가 R의 받침 위에 연속적일 수 있도록, $x = (x_1, x_2')'$와 $\beta = (\beta_1, \beta_2')'$의 상응하는 분할(partitions)이 존재한다.

가정 6 x가 주어진 t의 조건부분포가 R^J의 받침 위에 연속적이다.

<가정 1>과 <가정 3>에서 언급된 것과 같이 LAD추정에서 지불의사액의 분포에 대한 가정은 매우 약하다. 분포에 대한 어떤 특정한 함수 형태의 가정을 필요로 하지 않으며, 알려지지 않은 형태의 이분산성과 같이 독립적이고 동일한 분포가 아닌 경우에도 추정이 가능하다. 대신에 식별성과 일치성을 위해 설명변수의 데이터가 충분히 상세하고 광범위할 것을 요구하는데, 이 조건은 <가정 5>에 명시되어 있다. 조건부가치측정모형의 추가제시금액은 LAD추정의 식별성과 일치성을 보장하는 또 다른 방법이다. 비록 <가정 5>를 충족시킬 만큼 설명변수의 데이터가 충분히 상세하고 광범위하지 않더라도, 제시금액이 <가정 6>을 만족하는 적절한 분포를 가지면 LAD추정이 가능한데, 이 조건은 연구자에 의해 언제나 기술적으로 조정될 수 있다. 예를 들어, 설명변수가 상수항만인 경우에도 모수가 식별되는 일치적인 추정이 가능하게 만들 수 있다. <가정 5> 또는 <가정 6>은 충분조건이고 필요조건은 아니다. 지불의사액과 모수공간에 대한 보다 엄격한 조건이 있으면, 이 가정은 더 완화될 수 있다. <가정 4>는 식별성을 위한 정칙성(non-singularity)의 가정이며, 적어도 하나의 제시금액은 x와 선형적으로 독립적이어야 함을 요구하고 있다.

예비정리 1 식 별 성

<가정 3-4>, 그리고 <가정 5> 또는 <가정 6>이 만족하면, $b \in B$, $b \neq \beta$에 대해서 $D(\beta) < D(b)$이다.

정리 1 강일치성

<가정 1-4>, 그리고 <가정 5> 또는 <가정 6>이 만족하면, $\widehat{b_N}$은 β에 거의 확실하게 수렴한다.

〈증 명〉

<가정 5> 또는 <가정 6>이 만족하면, 각 $b \in B$에 대해 $x'b - t_j$는 연속적으로 분포되고, 각 $b \in B$에 대해 $1(x'b \geq t_j)$는 1의 확률로 연속적이며, $\sum_{j=1}^{J} 1(x'b \geq t_j)$ 역시 그러하다. 또한 $\sum_{j=1}^{J} y_j$와 $\sum_{j=1}^{J} 1(x'b \geq t_j)$ 모두 모든 $b \in B$에 대해서 유계되어(bounded) 있다. Newey and McFadden (1994, Lemma 2.4)에 따르면, <가정 1>과 <가정 2>가 만족하면 이것은 $D(b)$가 연속적이며, $D_N(b)$가 $D(b)$로 거의 확실하게 고르게 수렴(converge uniformly) 함을 의미한다. <가정 2>에 따라 B는 컴팩트하고, <예비정리 1>에 따라 $D(b)$는 $\beta \in B$에서 유일하게 최소화되므로, Amemiya (1973, Lemma 3)의 정리에 따라 $\widehat{b_N}$은 β로 거의 확실하게 수렴한다.

Q.E.D.

Kim and Pollard (1990)는 Manski (1975, 1985)의 최대점수추정량의 수렴율이 $N^{\frac{1}{3}}$이며, $\sqrt{N}$-일치적이지 않음을 보였다. 조건부가치측정모형의 LAD추정은 최대점수 추정에 근거하므로, LAD추정량 역시 그러하리라고 추측된다.

3. 추가질문이 있는 투표형 질문의 조건부가치측정

투표형 질문의 조건부가치측정 설문조사에서는 첫 번째 질문 후 다른 제시금액에 대한 추가질문이 주어지기도 하는데, 이 경우 추가제시금액 t_2는 첫 번째 응답 y_1에 다음과 같이 의존한다.

$$t_2 = \begin{cases} t_{2u} & y_1 = 1\text{일 경우} \\ t_{2l} & y_1 = 0\text{일 경우} \end{cases} \tag{8}$$

여기서 t_1은 첫 번째 제시금액이고, t_{2l}과 t_{2u}는 추가질문에서 주어지는 조

건부제시금액(contingent offers)인데, $t_{2l} < t_1 < t_{2u}$를 만족하도록 설문조사 전에 미리 마련된다. 첫 번째 질문에 응답자가 "예"라고 대답하면, 두 번째 질문에서는 보다 높은 조건부제시금액 t_{2u}가 주어지고, "아니오"라고 대답하면 보다 낮은 조건부제시금액 t_{2l}이 주어진다. 추가질문이 응답자의 응답, 따라서 지불의사액에 의존하게 되는 이러한 내생성 때문에, 추가질문이 있는 투표형 질문의 조건부가치측정모형의 LAD추정량을 정의할 때는 특별한 주의를 기울여야 한다. 설문조사에서 제시되는 제시금액 t_2는 응답자의 지불의사액에 의존하므로, $Med(\varepsilon\,|x, t_1, t_{2|y_1=1})$와 $Med(\varepsilon\,|x, t_1, t_{2|y_1=0})$는 서로 다르다. 따라서 $Med(\varepsilon\,|x, t_1, t_2) = 0$이라는 가정하에 성립하는 정리 1로는 다음의 표본거리합수를 최소화하는 추정량의 일치성이 보장되지 않는다.

$$\frac{1}{N}\sum_{i=1}^{N}\{\,|y_{1i} - 1(x_i'b \geq t_{1i})| + |y_{2i} - 1(x_i'b \geq t_{2i})|\} \qquad (9)$$

응답자는 설문조사에서 두 번의 질문에만 대답하지만, 결과적으로는 3개의 제시금액 (t_{2l}, t_1, t_{2u})에 대한 질문에 대답하고, 그것들이 의미하는 네 구간 중 하나를 선택하는 셈이다. 설문조사의 응답으로부터 조건부제시금액 (t_{2l}, t_{2u})에 대한 응답 (y_{2l}, y_{2u})는 각각 다음과 같이 주어질 수 있다.

$y_{2u} = y_2 \qquad y_{2l} = 1 \qquad y_1 = 1$일 경우

$y_{2u} = 0 \qquad y_{2l} = y_2 \qquad y_1 = 0$일 경우

다시 말해, 응답자가 t_1을 받아들이면 보다 낮은 제시금액 t_{2l} 역시 받아들이고, t_1을 거부하면 보다 높은 제시금액 t_{2u} 역시 거부하리라고 가정하는 것이다.[5] 조건부제시금액 t_{2l}과 t_{2u}는 응답자의 지불의사액에 의존하지 않고 외

5) Cameron and Quiggin (1994)은 응답자의 지불의사액이 설문과정에서 변할 수도 있다고 가

생적으로 정해진다. 따라서 $Med(\varepsilon \mid t_{2l}, t_1, t_{2u})$는 잘 정의되고 다음을 최소화하는 추정량은 정리 1의 해당조건들을 만족할 경우 일치적인 LAD추정량이 될 수 있다.

$$\frac{1}{N}\sum_{i=1}^{N}\{|y_{2li}-1(x_i'b \geq t_{2li})| + |y_{1i} - 1(x_i'b \geq t_{1i})| + |y_{2ui} - 1(x_i'b \geq t_{2ui})|\} \tag{10}$$

4. 지불카드를 이용한 조건부가치측정

지불카드법에서는 응답자에게 (t_0, t_1)과 $\{[t_j, t_{j+1})\}_{j=1}^{J}$의, $J+1$개의 서로 중복되지 않는 구간을 가진 지불카드가 제시되고, 응답자의 최대지불의사액이 속하는 구간을 묻는 질문이 주어진다. 응답자가 한 구간을 표시하면, 그는 결과적으로 J개의 다른 질문을 동시에 대답하는 셈이다. 예를 들어 응답자가 $[t_j, t_{j+1})$의 구간을 표시하면, $l \leq j$에 대해서는 $y_l = 1$이며, 그 밖에는 $y_l = 0$이 된다. 표본거리함수는 아무런 문제없이 구성될 수 있지만, 그 대신 지불카드를 설계할 때 주의를 기울여야 한다. 여러 지불카드가 준비되어 임의로 배당되어야 하며, 그렇지 않으면 x가 상수를 포함할 경우에 식별성이 보장되지 않는다.

정하는 이변량의(bivariate) 조건부가치측정모형을 제시하였다. (4)의 조건부가치측정모형은 지불의사액이 설문과정에서 변하지 않음을 암묵적으로 의미한다.

Ⅲ. 운영특성

이 절은 LAD추정의 계산전략을 설명하고, 모의실험 연구를 통해 소표본 성질과 추가질문의 효율성 기여 등의 LAD추정량의 운영특성을 평가한다.

1. 모의실험의 설계

지불의사액은 다음과 같이 로그정규분포에 따라 분포되어 있다.

$$\log(y^*) = \beta_1 + \beta_2 \log(\text{income}) + \varepsilon \tag{11}$$
$$\varepsilon \sim N\left(0, \frac{1}{\sqrt{3}}\right)$$

따라서 y^*의 치역은 R^+에 제한된다. 목적함수는 t가 $\log(t)$로 대체되는 것 이외에는 앞에서의 선형함수의 경우와 같다. 설명변수로는 상수와 로그소득이 사용되었다. 소득데이터는 1998년 미국의 18세 이상 개인연간소득(1,000달러)의 경험분포(empirical distribution)로부터 임의로 추출되었다. 개인연간소득의 경험분포는 1998년 인구조사파일(Current Population Survey Annual Demographic File)의 개인소득, 나이, 개인가중치에 근거하여 만들어졌고, <표 1>에 요약되어 있다. β_1과 β_2는 모두 1이라고 임의로 가정하였는데, 이 모수값의 선택은 모의실험의 결과에 특별한 영향을 미치지 않았다. 설문조사의 질문방식으로는 투표형 질문이 사용되었는데, 설문조사의 첫 번째 제시금액은 $t_1 \sim U(0, 200]$을 만족하도록 임의로 추출되었고, 조건부제시금액은 $t_{2u} = 2t_1$, $t_{2l} = \frac{1}{2} t_1$가 성립하도록, 두 번째 제시금액 t_2는 (8)에 따라 각각 주어졌다. 지불의사액 y^*는 경험분포로부터 임의추출된 소득과 정규분포 $N\left(0, \frac{1}{\sqrt{3}}\right)$에

〈표 1〉 1998년 미국 개인연간소득 분포

%	소득(1,000달러)	%	소득(1,000달러)
10	2.5	60	23.0
20	5.9	70	30.0
30	9.3	80	38.6
40	13.0	90	54.8
50	18.0	100	603.3

주: 경험분포는 미국의 1998년 인구조사 파일에 근거하여 만들어졌다. 18세 이상의 개인은 총 19만 7,421명으로서, 이는 실제 9만 4,990의 데이터에 개인가중치를 적용한 것이다. 원래의 데이터에서 개인연간소득변수는 빚을 반영하여 음의 값을 가지기도 하지만, 여기서는 최소한 2,500달러의 소득을 가지도록 하였다.

서 임의추출된 오차 ε으로 구성되었고, 위에서 만들어진 제시금액과 결합하여 설문조사의 응답 y_1, y_2가 (4)를 만족하도록 만들어졌다. 각 모의실험 데이터는 1,000개의 관측치를 가지고, 각 실험은 100개의 데이터에 대해서 이루어졌다.

2. 계산법과 추정

이산선택모형의 최대점수추정의 계산법은 Manski and Thompson (1986)에 잘 설명되어 있다. 조건부가치측정모형의 LAD추정의 계산도 유사한 방법으로 할 수 있지만, 차이점은 모수 벡터가 완전히 식별되지 않는다는 점이다. Manski and Thompson (1986)의 계산법은 다음과 같이 요약될 수 있다. 임의의 직교방향(orthogonal directions)의 집합을 선택하고, 각 방향을 따라 차례로 그때까지의 최적점을 시작점으로 하여 선탐색(line search)을 행한다. 만약 게임중단반복수(end-game iteration number) N_e 동안에 허용한도(tolerance) e 이내에서 최소점의 탐색에 진전이 없으면 계산이 중단된다. 주어진 허용한도 하에서 계산의 정확성은 게임중단수를 변화시킴으로써 조정된다. Manski and

〈표 2〉 LAD추정
Manski and Thompson (1986) 계산법

		$N_e=5$	$N_e=10$	$N_e=20$	$N_e=50$	$N_e=100$
$\widehat{b}_1$	편 의	-42.2705	-33.0939	-12.2595	-19.5310	-20.4216
	표준편차	390.9188	327.1577	90.5053	195.2227	141.2837
	RMSE	393.4145	328.9499	91.4037	196.2812	142.8672
$\widehat{b}_2$	편 의	13.0635	8.9806	3.6216	5.7378	5.7368
	표준편차	121.4345	88.6813	26.9138	57.3273	39.5078
	RMSE	129.2432	95.0801	29.7954	60.8342	44.8421

Thompson (1986)에서 설명되어 있듯이, 선탐색은 매 방향마다 정확하게 이루어진다. 하지만 계산법은 여전히 운에 달려 있는데 이는 방향탐색(directional search)이 임의로 이루어지기 때문이다. <표 2>는 5에서 100 사이의 게임중단반복수로 Manski and Thompson (1986)의 계산법으로 LAD추정을 한 결과를 보여 준다. 큰 표준편차에서 보여지듯이 추정은 매우 불안정하며, 게임중단반복수를 늘려도 계산의 정확성은 개선되지 않았다.

이 계산의 결과를 다른 각도에서 들여다보기 위해서, 100 데이터 중에서 1 데이터를 임의로 골라 게임중단반복수 10으로 Manski and Thompson (1986)의 계산법에 따라 20번 추정을 하였다. <표 3>은 추정된 결과를 최소화된 절대편차의 순서에 따라 정리해 보여 준다. 추정에서 최소화된 절대편차가 가장 최소인 0.1093일 때 모수의 추정은 가장 정확하지만, 그 최소값은 20번의 추정 중에서 여섯 번만 얻어졌다. 추정이 가장 최악인 절대편차 0.2697의 값을 얻었을 때, 추정치는 참모수로부터 멀리 떨어져 버렸다. 끊임없이 시작점이 경신되는 Manski and Thompson (1986)의 계산법은 절대편차의 최소값에 도달하는데 충분한 확률적 기초를 제공하지 못하는 것처럼 보인다.

이 연구에서는 약간 다른 계산전략(algorithm)이 채택되었는데, 이 계산전략은 여러 차선의 최적점들을 유지하면서 최적화의 초점을 서서히 좁혀 가는,

〈표 3〉 LAD추정
Manski and Thompson (1986) 계산법
모의실험 데이터 No.93
$N_e = 10$

절대편차	$\widehat{b_1}$	$\widehat{b_2}$
0.2697	-692.8290	204.7866
0.2697	-641.5715	189.7326
0.2030	-12.6078	5.0896
0.1347	-1.6416	1.7792
0.1213	-0.5900	1.4548
0.1130	0.2362	1.2080
0.1107	0.5308	1.1223
0.1103	1.0651	0.9646
0.1097	1.0023	0.9836
0.1097	0.9778	0.9903
0.1097	0.9816	0.9889
0.1097	1.0345	0.9758
0.1097	1.0276	0.9774
0.1097	1.0333	0.9768
0.1093	0.9998	0.9846
0.1093	0.9979	0.9853
0.1093	0.9998	0.9846
0.1093	1.0006	0.9842
0.1093	1.0000	0.9841
0.1093	1.0007	0.9840

주: 보고된 절대편차는 $\frac{1}{J} D_N(b)$이고, 0에서 1 사이의 값을 가진다.

Szu and Hartley (1987)의 모의서냉(simulated annealing) 계산법의 원리에 바탕을 둔 것이다. 게임중단수, N_{e1}의 최소화과정이 시작점으로부터 N_1번 반복된다. 이 중 상위 N_{1n}의 최소화점을 고른 다음, 각 점에서 시작하여 게임중단

〈표 4〉 LAD추정 및 최우추정

		LAD추정	최우추정
$\widehat{b}_1$	편 의	0.0190	0.0049
	표준편차	0.2537	0.1006
	RMSE	0.2544	0.1007
$\widehat{b}_2$	편 의	-0.0039	-0.0009
	표준편차	0.0783	0.0312
	RMSE	0.0807	0.0316

수, N_{e2}의 최소화과정을 N_{2n}번 반복한다. 마지막으로 $N_{1n}N_{2n}$개의 추정치에서 최종 추정치를 고른다. 이렇게 보수적인 계산법은 계산에 매우 많은 시간을 필요로 하지만, 모의실험 연구에서 보다 좋은 추정결과를 제공하였다. <표 4>는 새로운 계산전략에 의한 LAD추정의 결과를 보여 준다.[6] 표는 또한 지불의사액의 로그정규분포를 올바르게 가정하는 최우추정의 결과도 함께 제시한다. 분포가 올바르게 가정되었기 때문에 최우추정은 편의를 가져오지 않으며, 효율적이다. LAD추정의 경우, 편의는 무시할 만큼 작았지만, 표준편차가 컸고, 따라서 제곱근평균제곱오차(root mean squared error: RMSE)의 관점에서 최우추정에 비해 상대적으로 비효율적이었다.

3. 표본 크기의 효과와 추가질문의 영향

<표 5>는 LAD추정과 최우추정의 소표본 성질을 평가한다. 실험은 표본크기를 달리하는 같은 모의실험 데이터에 대해서 이루어졌다. 표본크기가 커지면서 표준편차는 예상대로 작아지지만, LAD추정과 최우추정 두 경우 모두 표본크기가 100인 경우에서조차도 추정의 편의는 매우 작았다.

6) $N_{e1}=5$, $N_1=10$, $N_{e2}=5$, $N_{1n}=5$, 그리고 $N_{2n}=1$.

〈표 5〉 LAD추정 및 최우추정
소표본의 영향

			LAD추정	최우추정
$N=100$	$\widehat{b}_1$	편 의	-0.1306	-0.0215
		표준편차	0.7956	0.4060
		RMSE	0.8070	0.4066
	$\widehat{b}_2$	편 의	0.0344	0.0037
		표준편차	0.2434	0.1260
		RMSE	0.2784	0.1279
$N=200$	$\widehat{b}_1$	편 의	-0.0320	-0.0327
		표준편차	0.4893	0.2843
		RMSE	0.4905	0.2863
	$\widehat{b}_2$	편 의	0.0104	0.0105
		표준편차	0.1495	0.0883
		RMSE	0.1532	0.0947
$N=500$	$\widehat{b}_1$	편 의	0.0060	-0.0047
		표준편차	0.3556	0.1518
		RMSE	0.3557	0.1519
	$\widehat{b}_2$	편 의	0.0018	0.0019
		표준편차	0.1053	0.0474
		RMSE	0.1055	0.0476

투표형 질문의 조건부가치측정은 추가질문이 주어지면 통계학적으로 보다 효율적이게 된다. Hanemann *et al.* (1991)은 추가질문이 있는 경우의 최우추정의 효율성을 증명했고, 그를 뒷받침하는 경험적 증거를 제시하였다. 그러나 LAD추정의 경우에는 점근적 분포(asymptotic distribution)를 도출할 수 없으므로, 이론적으로 점근적 효율성을 증명하는 것은 불가능하다. 대신 모의실험을 통해 추가질문이 LAD추정의 효율성에 기여하는 정도를 평가하였다. <표 6>은 첫 번

〈표 6〉 LAD추정 및 최우추정
첫 번째 질문과 응답만 사용

		LAD추정	최우추정
$\widehat{b}_1$	편 의	0.0051	-0.0044
	표준편차	0.4781	0.1717
	RMSE	0.4782	0.1718
$\widehat{b}_2$	편 의	0.0014	0.0008
	표준편차	0.1413	0.0510
	RMSE	0.1414	0.0512

째 질문만 사용하는 LAD와 최우추정의 결과를 보여 준다. <표 4>와 <표 6>으로부터 LAD추정과 최우추정의 경우 모두 추가질문이 추정의 효율성을 개선하는 것은 명백하다. 그러나 LAD추정의 경우가 최우추정의 경우보다 추가질문의 효율성 기여가 상대적으로 크다. <표 6>에서 LAD추정과 최우추정의 표준편차는 <표 5>의 표본크기 200과 500일 경우에서 표준편차에 각각 비교될 수 있다. 다시 말해, 추가질문은 추정의 효율성을 LAD추정의 경우에는 5배, 최우추정의 경우에는 2배 증가시킨다고 할 수 있다.

Ⅳ. 이분산성

최우추정이 편의를 가지게 되는 경우에 LAD추정은 불편추정치를 줄 수 있다는 것이 LAD추정의 동기인데, 그러한 경우의 한 예로서 알려지지 않은 형태의 이분산성을 들 수 있다. 이 절에서는 이분산성이 존재할 때 조건부가치측정모형의 LAD와 최우추정의 편의와 효율성을 모의실험을 통해 비교한다.

1. 모의실험의 설계

지불의사액은 (11)에서와 같이 로그정규분포로 분포되지만, 지불의사액의 오차는 이제 다음과 같이 주어진다.

$$\varepsilon = \chi\eta \qquad \eta \sim N\left(0, \frac{1}{\sqrt{3}}\right) \tag{12}$$

여기서 비확률적(non-stochastic) 요소 χ는 확률적 요소 η의 척도를 조정한다. 만약 χ가 상수이면 분포는 등분산적(homoskedastic)이다. 제Ⅲ절에서의 지불의사액모형은 $\chi = 1$인 경우에 해당한다. 이분산성은 χ가 x에 의존할 때 존재한다. 다음과 같은 두 유형의 이분산성이 실험되었다.

$$\text{유형 I} \qquad \chi = \left(\frac{x'\beta}{m}\right)^{-1}$$

$$\text{유형 II} \qquad \chi = \left(\frac{x'\beta}{m}\right)^{-2}$$

여기서 $m = \beta_1 + \beta_2 \log(18)$이다. <표 1>에서 볼 수 있듯이, 1998년 미국개인연간소득의 중위값은 18(1,000달러)이고, 따라서 이분산성에 따라 다음이 성립한다.

$$\text{소득} \begin{pmatrix} < \\ = \\ > \end{pmatrix} 18\text{이면} \qquad \varepsilon \begin{pmatrix} > \\ = \\ < \end{pmatrix} \eta$$

이는 등분산성의 경우에 비교할 때, 소득이 중위값보다 크면 지불의사액의 곱셈오차의 척도가 작고, 소득이 중위값보다 작으면 지불의사액의 곱셈오차의

척도가 작다는 것을 의미한다. 이분산성의 정도는 유형 I의 경우가 유형 II의 경우보다 크다. 먼저와 같이 각각 1,000개의 관측치를 가지는 100개의 데이터가 각 유형의 이분산성분포에 대해 만들어졌다.

2. 효율성과 강건함

두 유형의 이분산성에 대하여, 조건부가치측정모형을 LAD추정과 최우추정법으로 추정하였다. 최우추정은 등분산성의 가정하에서 이루어졌다. <표 7>은 모의실험의 결과를 보여 준다.

LAD추정의 편의는 두 실험의 경우 모두 무시할 수 있을 만큼 작았으며, 이는 알려지지 않은 형태의 이분산성에 대한 LAD추정량의 강건함과 일치한다.

〈표 7〉 LAD추정 및 최우추정 이분산성의 영향

			LAD추정	최우추정
유형 I	$\widehat{b}_1$	편 의	0.0260	0.2876
		표준편차	0.2770	0.1435
		RMSE	0.2783	0.3340
	$\widehat{b}_2$	편 의	-0.0054	-0.0910
		표준편차	0.0801	0.0428
		RMSE	0.0843	0.3046
유형 II	$\widehat{b}_1$	편 의	0.0293	0.6735
		표준편차	0.3285	0.1919
		RMSE	0.3299	0.7304
	$\widehat{b}_2$	편 의	-0.0057	-0.2075
		표준편차	0.0943	0.0588
		RMSE	0.0989	0.7072

그러나 등분산성을 가정하는 최우추정은 두 유형의 이분산성에 대하여 모두 편향된 추정치를 주고, 그 편의의 정도는 유형 II의 이분산성의 경우가 보다 컸다. LAD추정의 상대적인 비효율성은 상대적으로 큰 표준편차에 나타나며, 표준편차가 RMSE의 지배적인 요소이었다. 이에 비해 최우추정의 경우는, 표준편차는 상대적으로 작았고, RMSE의 지배적인 요소는 편의이었다. 두 유형의 이분산성을 실험한 결과, 평균제곱오차의 기준에서 볼 때, LAD추정은 최우추정보다 더 나은 추정결과를 보여 주었다.

V. 응용 예

이 절에서는 LAD추정이 실제로 어떻게 적용되는가를 보여 주기 위해 간단한 응용 예를 제시한다. 데이터는 1990년 오스트레일리아 자원평가위원회(Australian Resource Assessment Commission)의 위촉으로 이루어진 Imber *et al.* (1991)의 조건부가치측정 설문조사의 데이터를 사용했는데, 그 조사는 오스트레일리아 카카두 보존지역(Kakadu Conservation Zone)에서의 광산운영에 따른 환경피해를 평가하기 위한 것이었다. 설문조사에서는 두 가지 각본이 사용되었는데, 이 연구에서 사용된 데이터는 소피해(minor damage) 각본에 대한 것이다.[7] 총 1,013의 응답자에게 추가질문이 있는 투표형 질문의 조건부가치측정 설문조사질문이 주어졌다. 제시금액은 2달러, 5달러, 20달러, 50달러, 100달러, 250달러 등이 사용되었다. 첫 번째 제시금액은 5달러에서 100달러까지의 금액 중에서 임의로 선택되었고, 두 번째 제시금액은 첫 번째 응답에 따라 한 단위 높거나 한 단위 아래의 제시금액이 사용되었다. 설문조사의 응답은 <표 8>에 요약되어 있다.

<표 9>는 LAD추정과 최우추정의 결과를 보여 준다. 로그선형의 지불의사액

7) 같은 데이터는 Cameron and Quiggin (1994, 1998)과 Haab (1998)에서도 사용되었다.

〈표 8〉 데이터 요약

제시금액			대답: 첫 번째 - 두 번째				합계
첫 번째	두 번째		예-예	예-아니오	아니오-예	아니오-아니오	
5달러	20달러	2달러	150	17	7	79	253
20달러	50달러	5달러	136	20	11	85	252
50달러	100달러	20달러	124	23	15	93	255
100달러	250달러	50달러	105	31	17	100	253

〈표 9〉 LAD추정 및 최우추정

	LAD추정	최우추정
상 수	4.2844	3.8011
	(0.6203)	(0.2944)
여 자	0.7789	1.3444
	(0.7347)	(0.4322)
σ		5.6432
		(0.4094)
logL		-1095.4188
절대편차	0.3972	0.4886
지불의사액의 중위값	115.1250	107.8888
	(30.7526)	(29.8005)

주: 괄호 안의 수치는 표준오차이다. 보고된 절대편차는 $\frac{1}{J}D_N(b)$이고, 0과 1 사이의 값을 가진다. LAD추정은 $N_1=20$, $N_{e1}=5$, $N_{1N}=5$, $N_{2n}=1$, $N_{e2}=5$의 방법으로 이루어졌다.

모형이 가정되었고, 지불의사액을 설명하는 변수로서 상수항과 여자 더미변수가 사용되었으며, 최우추정에서는 로그정규분포가 가정되었다. 보고된 지불의사액의 중위값은 모든 응답자의 중위값 지불의사액의 평균이고, 해당하는 점근적 표

준오차가 보고되었다. 앞에서 언급한 대로 LAD추정치의 표준오차에 대하여 분석적인 근사치를 제공하는 것은 불가능하다. 대신, 추정의 정확성에 대한 기술적 측도(descriptive measure)로서, LAD추정치 표준오차의 RMSE에 대한 부트스트랩(bootstrap) 추정치를 구했는데, 부트스트랩 추정치는 100 부트스트랩 시도로부터 계산되었다.

추정의 결과는 제Ⅳ절에서 보고된 모의실험 결과와 일치한다. 표준편차와 RMSE가 크다는 점에서, 최우추정에 비해 LAD추정은 상대적으로 비효율적이다. 그러나, 절대편차가 작다는 점에서 설문조사응답에 대한 보다 나은 예측을 제공한다.

Ⅵ. 맺음말

LAD추정은 조건부가치측정 데이터의 정보를 결합하는 방법에서 상대적으로 비효율적이고 보수적이지만, 모의실험에서 최우추정에 상응하는 고무적인 추정결과를 보였다. 많은 연구자들이 조건부가치측정모형의 최우추정이 지불의사액에 대한 가정에 대해 민감하다고 지적한다. 이렇게 분포에 대한 가정과 최우추정이 도전을 받을 때, LAD추정은 최우추정의 좋은 대안이 될 수 있으며, 또는 가치있는 제2의 의견으로 고려될 수 있을 것이다.

하지만 LAD추정을 현행 조건부가치측정 설문조사의 데이터에 적용하는 데에는 많은 제약이 따른다. 왜냐하면 이제까지 조건부가치측정 설문조사를 실행한 연구자들은, LAD추정의 가능성을 고려하지 않고 설문조사를 설계하였기 때문이다. 제시금액을 충분히 풍부하고 광범위한 분포에서 추출하면 LAD추정은 어떤 경우에도 효과적으로 이루어질 수 있고, 모수추정이나 비모수적 추정과 자유롭게 비교될 수 있다. 연구자들의 이해가 절실히 필요로 되는 대목이다.

〈부록 1〉 예비정리 1의 증명

증명은 크게 두 부분으로 이루어진다.

첫째, x와 t가 주어졌을 때, $\sum_{j=1}^{J} 1(x'b-t_j) \neq \sum_{j=1}^{J} 1(x'\beta-t_j)$를 만족하는 어떤 $b \in B$에 대해서도 $D(\beta) < D(b)$임을 보인다. 둘째, 어떤 $b \in B$, $b \neq \beta$에 대해서도 $\sum_{j=1}^{J} 1(x'b \geq t_j) \neq \sum_{j=1}^{J} 1(x'\beta \geq t_j)$가 양의 확률로 성립함을 보인다.

첫 번째 부분은 Lee (1992, Theorem 1)의 증명과 유사하지만, 다른 점은 여기서는 경계점이 모수가 아니라 외생변수로 취급된다는 것이다. 중위값 m을 가진 어떤 확률변수(random variable) Y에 대해서도 다음이 성립한다.

$$E(|Y-c|) - E(|Y-m|) \geq \begin{cases} 2(c-m)\left[P(Y \leq m) - \frac{1}{2}\right] & c > m \text{일 경우} \\ 2(m-c)\left[P(Y \geq m) - \frac{1}{2}\right] & c < m \text{일 경우} \end{cases}$$

여기서 강부등호는 $P(Y \leq m)$과 $P(Y \geq m)$이 $\frac{1}{2}$보다 클 경우에 성립한다. <가정 3>으로부터 다음이 성립함을 보여 줄 수 있다.

$$Med(y_j|x,t) = 1(x'\beta \geq t_j) \qquad \text{for} \quad j = 1, \cdots, J$$

$$Med\left(\sum_{j=1}^{J} y_j|x,t\right) = \sum_{j=1}^{J} 1(x'\beta \geq t_j)$$

왜냐하면 모든 $j = 1, \cdots, J$에 대해서 $y_j = 1(x'\beta + \varepsilon \geq t_j)$는 ε에 대해 증가하기 때문이다. 또한 $\{t_j\}$는 증가하므로, $\{y_j\}$와 $\{1(x'\beta \geq t_j)\}$는 모두 j에 대해 감소한다. 만약 $1(x'\beta \geq t_0) = 1$이고 $1(x'\beta \geq t_{J+1}) = 0$이라 가정하면, 어느 $r = 0, 1, \cdots, J$에 대해서 다음이 성립한다.

$$
\begin{aligned}
& \sum_{j=1}^{J} 1(x'\beta \geq t_j) = r \\
\Leftrightarrow\ & 1(x'\beta \geq t_r) = 1, \quad 1(x'\beta \geq t_{r+1}) = 0 \\
\Leftrightarrow\ & P(y_r = 0) < \frac{1}{2}, \quad P(y_{r+1} = 1) < \frac{1}{2} \\
\Leftrightarrow\ & P(\sum_{j=1}^{J} y_j < r) < \frac{1}{2}, \quad P(\sum_{j=1}^{J} y_j \geq r+1) < \frac{1}{2} \\
\Leftrightarrow\ & P(\sum_{j=1}^{J} y_j \geq r) > \frac{1}{2}, \quad P(\sum_{j=1}^{J} y_j \leq r) > \frac{1}{2}
\end{aligned}
$$

이것이 첫 번째 증명을 완료한다.

증명의 두 번째 부분은 Manski (1985, Lemma 2)의 증명과 유사하지만 여기서는 제시금액의 영향에 대한 언급이 덧붙여진다. 어느 j에 대해서도 $1(x'b \geq t_j) \neq 1(x'\beta \geq t_j)$이면 $\sum_{j=1}^{J} 1(x'b \geq t_j) \neq \sum_{j=1}^{J} 1(x'\beta \geq t_j)$이 성립하므로, 다음을 보여 주는 것만으로 충분하다.

$$
\begin{aligned}
& P\{(x, t) \mid 1(x'b \geq t_m) \neq 1(x'\beta \geq t_m)\} \\
&= P\{(x, t) \mid x'b < t_m \leq x'\beta\} + P\{(x, t) \mid x'\beta < t_m \leq x'b\} > 0
\end{aligned}
$$

$\beta_1 > 0$인 경우를 고려해 보자. $\beta_1 < 0$의 경우는 대칭적이다. <가정 5>가 만족한다고 하면, $b_1 < 0$이면 다음이 성립한다.

$$
P\{x'b < t_m < x'\beta\} = P\left\{ \frac{t_m - x_2'b_2}{b_1} < x_1, \frac{t_m - x_2'\beta_2}{\beta_1} < x_1 \right\}
$$

이것은 (x_2, t)가 주어진 x_1의 꼬리확률(tail probability)이며, <가정 5>가 만족하므로 양의 부호를 가진다. 만약 $b_1 = 0$이면 다음이 성립한다.

$$P\{x'b < t_m < x'\beta\} = P\left\{x_2'b_2 - t_m < 0,\ \frac{t_m - x_2'\beta_2}{\beta_1} < x_1\right\}$$
$$P\{x'\beta < t_m < x'b\} = P\left\{x_2'b_2 - t_m > 0,\ x_1 < \frac{t_m - x_2'\beta_2}{\beta_1}\right\}$$

<가정 4>에 따라 $P\{x_2'b_2 - t_m = 0\} < 1$이고, $\{x_2'b_2 - t_m < 0\}$ 또는 $\{x_2'b_2 - t_m > 0\}$ 중 하나가 양의 확률로 성립해야 한다. 그럴 경우 <가정 5>에 따라 $P\{x'b < t_m < x'\beta\}$ 또는 $\Pr\{x'\beta < t_m < x'b\}$ 중 하나가 양의 부호를 가져야 한다. 왜냐하면 그것은 x_1의 꼬리확률이기 때문이다. 만약 $b_1 > 0$이면 다음이 성립한다.

$$P\{x'b < t_m < x'\beta\} = P\left\{\frac{t_m - x_2'\beta_2}{\beta_1} < x_1 < \frac{t_m - x_2'b_2}{b_1}\right\}$$
$$P\{x'\beta < t_m < x'b\} = P\left\{\frac{t_m - x_2'b_2}{b_1} < x_1 < \frac{t_m - x_2'\beta_2}{\beta_1}\right\}$$

<가정 4>에 따라 $P\left\{(x, t) \mid \frac{t_m - x_2'\beta_2}{\beta_1} \neq \frac{t_m - x_2'b_2}{b_1}\right\} > 0$이 $b \in B$, $b \neq \beta$에 대해서 성립한다. 그러한 (x, t)에 대해서 $\left(\frac{t_m - x_2'\beta_2}{\beta_1}, \frac{t_m - x_2'b_2}{b_1}\right)$와 $\left(\frac{t_m - x_2'b_2}{b_1}, \frac{t_m - x_2'\beta_2}{\beta_1}\right)$의 어느 한 구간이 존재하며, x_1이 그 구간에 속할 확률은 <가정 5>에 따라 양이다.

<가정 6>이 만족한다고 하면 <가정 4>에 따라 $b \in B$, $b \neq \beta$에 대해 $P\{(x, t) \mid x'b \neq x'\beta\} > 0$이다. 그러한 (x, t)에 대해 $(x'b, x'\beta)$와 $(x'\beta, x'b)$ 중 어느 한 구간이 존재하며, t_m이 그 구간에 속할 확률은 <가정 6>에 따라 양이다.

Q.E.D.

◎ 참 고 문 헌 ◎

1. Amemiya, T., "Regression Analysis When the Dependent Variable is Truncated Normal," *Econometrica*, 41(6), 1973, pp. 997～1016.
2. Cameron, T. A., "A New Paradigm for Valuing Non-market Goods using Referendum Data: Likelihood Estimation by Censored Logistic Regression," *Journal of Environmental Economics and Management*, 15, 1988, pp. 355～379.
3. ______ and M. D. James, "Efficient Estimation Methods for 'Closed-ended' Contingent Valuation Surveys," *The Review of Economics and Statistics*, 69(2), 1987, pp. 269～276.
4. ________ and J. Quiggin, "Estimation using Contingent Valuation Data from a 'Dichotomous Choice with Collow-up' Questionnaire," *Journal of Environmental Economics and Management*, 27, 1994, pp. 218～234.
5. ______, "Estimation using Contingent Valuation Data from a 'Dichotomous Choice with Follow-up' Questionnaire: Reply," *Journal of Environmental Economics and Management*, 35, 1998, pp. 195～199.
6. Carson, R. T., Hanemann, W. M., Kopp, R. J., Krosnick, J. A., Mitchell, R. C., Presser, S., Ruud, P. A. and V. K. Smith, "Prospective Interim Lost Use Value due to DDT and PCB Contamination in the Southern California Bight," NRDA Inc. Report to NOAA, 1994.
7. Carson, R. T., Mitchell, R. C., Hanemann, W. M., Kopp, R. J., Presser, S. and P. A. Ruud, "A Contingent Valuation Study of Lost Passive Use Values Resulting from the Exxon Vladez Oil Spill," NRDA Inc. Unpublished Report to Attorney General of the State of Alaska, 1992.
8. Haab, T. C., "Estimation using Contingent Valuation Data from a 'Dichotomous Choice with Follow-up' Questionnaire: A Comment," *Journal of Environmental Economics and Management*, 35, 1998, pp. 190～194.
9. Habb, T. C. and K. E. McConnell, "Referendum Models and Negative Willingness

Topay: Alternative Solutions," *Journal of Environmental Economics and Management*, 32(3), 1997, pp. 251～270.

10. Hanemann, W. M., "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses," *American Journal of Agricultural Economics*, 66, 1984, pp. 332～341.
11. ________, Loomis, J. and B. Kanninen, "Statistical Efficiency of Double-bounded Dichotomous Choice Contingent Valuation," *American Journal of Agricultural Economics*, 73, 1991, pp. 1255～1263.
12. Imber, D., Stevenson, G. and L. Wilks, "A Contingent Valuation Survey of the Kakadu Conservation Zone," *RAC Research Paper* 3, Resource Assessment Commission, 1991.
13. Kim, J. and D. Pollard, "Cube Root Asymptotics," *Annals of Statistics*, 18(1), 1990, pp. 191～219.
14. Kriström, B., "A Non-parametric Approach to the Estimation of Welfare Measures in Discrete Response Valuation Studies," *Land Economics*, 66, 1990, pp. 135～139.
15. Lee, M.-J., "Median Regression for Ordered Discrete Response," *Journal of Econometrics*, 51, 1992, pp. 59～77.
16. Manski, C. F., "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 3, 1975, pp. 205～228.
17. ______, "Semiparametric Analysis of Discrete Response," *Journal of Econometrics*, 27, 1985, pp. 313～333.
18. ________ and T. S. Thompson, "Operational Characteristics of Maximum Score Estimation," *Journal of Econometrics*, 32, 1986, pp. 85～108.
19. Newey, W. K. and D. McFadden, "Large Sample Estimation and Hypothesis Testing," Volume 4 of Handbook of Econometrics, Chapter 36, 1994, pp. 2111～2245, Elsevier Science Publishers B.V.
20. Sellar, C., Chavas, J. P. and J. R. Stoll, "Specification of the Logit Model: The Case of Valuation of Nonmarket Goods," *Journal of Environmental Economics and Management*, 13, 1986, pp. 382～390.

21. Szu, H. and R. Hartley, "Fast Simulated Annealing," *Physics Letters A*, 122, 1987, pp. 157~162.

22. Tauchen, G., "Diagnostic Testing and Evaluation of Maximum Likelihood Models," *Journal of Econometrics*, 30, 1985, pp. 415~443.

ABSTRACT

The Least Absolute Deviations Estimation of the Contingent Valuation Model

Dongil Kim

This paper introduces the least absolute deviations estimation of the contingent valuation model, which corresponds to the semi-parametric estimation of discrete choice models by Manski (1975, 1985) and Lee (1992). The least absolute deviations estimation is more robust to mis-specified distributional assumptions in the estimation of the contingent valuation model, compared to the maximum likelihood estimation. The full identification and strong consistency of the estimation are proved and its application to different formats of contingent valuation survey data is discussed. Simulation studies are designed to evaluate its operational characteristics including computational strategies, small sample properties and the efficiency gain of a follow-up question. The bias and efficiency of least absolute deviations and maximum likelihood estimation are compared in the presence of heteroskedasticity.