자원·환경경제연구 제 12 권 제 4 호
Environmental and Resource Economics Review
Volume 12, Number 4, December 2003 : pp. 545~557

A Semiparametric Estimation of the Contingent Valuation Model*

Joo Heon Park**

〈Contents〉

Ⅰ. Introduction

The Contingent Valuation Method (CVM) has been recognized as a dominant methodology to estimate the value of non-market goods. For instance, the existence and other passive-use values of the environment can be estimated by using the CVM. Nevertheless, it is widely

* Received for publication: September. 10, 2003; Revision accepted: November, 4, 2003.
** Associate Professor, Department of Economics, Dongduk Women's University.

recognized that the CVM still has some unsolved bias problems such as the discredibility bias[1] and the misspecification bias. And many researches have been made to find out a scheme to reduce the bias. Especially there have been a lot of studies to reduce the discredibility bias by inducing the respondents to disclose their true willingness to pay (WTP) and some fruitful results actually have been produced. However, the misspecification bias problem of the CVM has rarely been discussed.

The purpose of this paper is to propose a semiparametric estimation to address the misspecification bias problem of a dichotomous choice CVM. The estimator is developed by adapting the well-known density weighted average derivative of regression function that was proposed by Powell *et al.* (1989).

Ⅱ. A Semiparametric Dichotomous Choice CVM and Estimation

A traditional contingent valuation method usually assumes that the amount a consumer is willing to pay for a non-market good is determined by a linear function:

$$y_i^* = x_i^T\beta + \varepsilon_i, \quad i = 1, 2, \cdots, N \tag{1}$$

where y^* is WTP, x is $(k-1)$ individual characteristics, β is a

1) The discredibility bias includes the embedding bias, strategic bias, hypothetical bias among others. See Diamond and Hausman (1994) for details.

parameter vector of interest and ε is the stochastic error term with $E[\varepsilon|x]=0$. In a dichotomous elicitation format, a survey respondent answers to a hypothetical question like "Are you willing to pay $\$t$ to buy a defined good?" A respondent would say "yes" if $y^* > t$ or answer "no" otherwise.

A semiparametric dichotomous choice CVM can be obtained by converting the above answering mechanism into a simple binary response model without a distributional assumption on ε as:

$$\begin{array}{ll} y_i = 1 \text{ if } y_i^* \geq t_i \\ y_i = 0 \text{ otherwise} \end{array} \Leftrightarrow \begin{array}{ll} y_i = 1 \text{ If } \varepsilon_i \geq t_i - x_i^T\beta \\ y_i = 0 \text{ otherwise} \end{array} \qquad (2)$$

Then the semiparametric regression function of the binary response (y) on the bidding price (t) and the individual characteristics (x) looks like:

$$g(t,x) \equiv E[y|t,x] = pr[\varepsilon \geq t - x^T\beta] = 1 - F_\varepsilon(t - x^T\beta) \qquad (3)$$

where $F_\varepsilon(\cdot)$ denotes an unknown cumulative distribution function of ε.

The model (3) is manifested as a restriction on the derivative of the regression function as:

$$\begin{aligned} \frac{\partial g(t,x)}{\partial x} &= \frac{\partial(1-F(t-x'\beta))}{\partial x} = f(t-x'\beta)\beta \\ \frac{\partial g(t,x)}{\partial t} &= \frac{\partial(1-F(t-x'\beta))}{\partial t} = -f(t-x'\beta) \end{aligned} \qquad (4)$$

as long as $g(\cdot)$ is continuous and differentiable at x. The equation (4) suggests that $\frac{\partial g}{\partial x}$ and $\frac{\partial g}{\partial t}$ are proportional to β for each value of

(t, x), thereby any weighted average of the derivatives $\frac{\partial g}{\partial x}$ and $\frac{\partial g}{\partial t}$ would also be proportional to β. Taking the density of t and x, $f(t, x)$ as a weighting function, the density weighted average derivative of a semiparametric regression function is defined as[2)]

$$\begin{aligned}
\delta_x &\equiv E\left[f_{t,x}(t, x)\frac{\partial(1-F_\varepsilon(t-x'\beta))}{\partial(x'\beta)}\right]\beta \\
&= E[f_{t,x}(t, x)f_\varepsilon(t-x'\beta)]\beta \\
&= \gamma\beta \\
\delta_t &\equiv E\left[f_{t,x}(t, x)\frac{\partial(1-F_\varepsilon(t-x'\beta))}{\partial t}\right] \\
&= -E[f_{t,x}(t, x)f_\varepsilon(t-x'\beta)] \\
&= -\gamma
\end{aligned} \tag{5}$$

Therefore, a moment condition of the parameters can be obtained by taking a ratio of the density weighted average derivatives as

$$\beta = -\frac{\delta_x}{\delta_t} \tag{6}$$

If a consistent estimator for the density weighted average derivative (δ_x, δ_t) is available, then β can also be estimated consistently by the analogy principle. Powell *et al.* (1989) proposed a nonparametric estimator of $\hat{\delta}_x$ and $\hat{\delta}_t$ as

2) The choice of density weighting is made because it permits an estimation whose properties can be analyzed and understood in a straightforward fashion. See Powell *et al.* (1989) for details.

$$\hat{\delta}_x = -\frac{2}{N(N-1)}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\left[\frac{1}{h}\right]^{k+1}K_x\left[\frac{t_i-t_j}{h},\frac{x_i-x_j}{h}\right]y_i \qquad (7)$$

$$\hat{\delta}_t = -\frac{2}{N(N-1)}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\left[\frac{1}{h}\right]^{k+1}K_t\left[\frac{t_i-t_j}{h},\frac{x_i-x_j}{h}\right]y_i$$

where K_x and K_t are the first derivatives of the kernel function with respect to x and t respectively. Therefore, a nonparametric estimator of the parameters of dichotomous choice CVM in (2) can be obtained as

$$\hat{\beta} = -\frac{\hat{\delta}_x}{\hat{\delta}_t} = -\frac{\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}K_x\left[\frac{t_i-t_j}{h},\frac{x_i-x_j}{h}\right]}{\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}K_t\left[\frac{t_i-t_j}{h},\frac{x_i-x_j}{h}\right]} \qquad (8)$$

Powell *et al.* (1989) proved the consistency and asymptotic normality of $\hat{\delta}_x$ and $\hat{\delta}_t$ by using a general result on the asymptotic behavior of U-statistics. Therefore, the limiting distribution of the estimator $\hat{\beta}$ can be easily derived by applying the so called 'δ-method' as in the following theorem.

Theorem

Given the regularity conditions on the density function of (t,x) and the kernel function $K(\cdot)$, if h obeys $Nh^{k+2}\rightarrow\infty$ and $Nh^{2p}\rightarrow 0$ as $N\rightarrow\infty$, then

$$\sqrt{N}(\hat{\beta}-\beta)\xrightarrow{d} N(0, L\Sigma_{\delta}L^{T})$$

where k is the dimension of $z(\equiv(t,x^T)^T)$; $p=(k+4)/2$ if k is even number or $p=(k+3)/z$ if k is odd number; and $\Sigma_\delta = 4E[r(w)r(w)^T] - 4\delta_Z\delta_z^T$

$$r(w) = f(z)\frac{\partial g(z)}{\partial z} - [y-g(z)]\frac{\partial f(z)}{\partial z}$$

$$L \equiv \left[\frac{\partial(\hat{\delta}_x/\hat{\delta}_t)}{\partial(\hat{\delta}_t,\ \hat{\delta}_x)}\right]_{\substack{\hat{\delta}_t=\delta_t \\ \hat{\delta}_x=\delta_x}}$$

〈Proof〉

Assuming $\delta_t \neq 0$, $\partial\hat{\delta}_x/\partial\hat{\delta}_t$ is differentiable around (δ_t, δ_x), expanding $\hat{\delta}_x/\hat{\delta}_t$ into a Taylor series around (δ_t, δ_x) and multiplying $\sqrt{N}$ to both sides provides

$$\sqrt{N}\left(\frac{\hat{\delta}_x}{\hat{\delta}_t} - \frac{\delta_x}{\delta_t}\right) = \left[\frac{\partial(\hat{\delta}_x/\hat{\delta}_t)}{\partial(\hat{\delta}_t,\ \hat{\delta}_x)}\right]_{\substack{\hat{\delta}_t=\delta_t \\ \hat{\delta}_x=\delta_x}} \sqrt{N}\begin{pmatrix}\hat{\delta}_t-\delta_t \\ \hat{\delta}_x-\delta_x\end{pmatrix} + o_p(1) \quad (9)$$

Powell *et al.* (1989) proved that $\sqrt{N}\begin{pmatrix}\hat{\delta}_t-\delta_t \\ \hat{\delta}_x-\delta_x\end{pmatrix} \xrightarrow{d} N(0, \Sigma_\delta)$.

Therefore, the theorem can be easily proved by applying the so called 'δ-method' to (9).

Ⅲ. Simulation

A finite sample Monte Carlo simulation was conducted to demonstrate the practical performance of the semiparametric estimator of the CVM proposed in this paper. For simulation, it was assumed that the latent variable, WTP(y^*) is determined by a linear model as

$$y_i^* = \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i, \quad (i = 1,2,\cdots,N) \tag{10}$$

with true values $\beta_1 = \beta_2 = 1$ held constant across the experimental designs. The observable indicator variable y_i takes one if $y_i^* \geq t_i$, zero otherwise in which the bidding price t_i's are generated randomly from the normal distribution with the same mean and variance of y_i^*. Two different error (ε_i) distributions were assumed - the standard normal distribution and the extreme value distribution with mean zero and variance one.

The standard multivariate normal density function was taken as a kernel function[3] and $h = 0.2$. The simulation was repeated for 100 times with the sample size N=1,000. <Table 1> summarizes the results from the simulation.

The result for semiparametric estimation suggests that the estimator is robust. It is evident that the estimates are not significantly biased away from the true values regardless of the distributional assumption on ε_i as shown in the 'MEAN' and 'MEDIAN' columns. Detailed simulation results should be available upon request.

3) A 'bias-corrected' kernel function should be used in order to remove the asymptotic bias embedded in the nonparametric kernel estimator. But Powell *et al.* (1989) suggests that the estimators using the 'bias-corrected' kernels are not systematically less biased than the estimators based on the standard kernel in a small sample.

〈Table 1〉 Monte Carlo Simulation Results of the estimator $\hat{\beta}$

$y_i^* = \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i,\ \varepsilon_i \sim N(0,1)$							
	TRUE	MEAN	SD	RMSE	LQ	MEDIAN	UQ
β_1	1.0000	0.9741	0.1534	0.1254	0.8643	0.9682	1.0709
β_2	1.0000	0.9585	0.1343	0.1167	0.8529	0.9400	1.0523
$y_i^* = \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i,\ \varepsilon_i \sim$ the extreme value distribution with mean 0 and variance 1							
	TRUE	MEAN	SD	RMSE	LQ	MEDIAN	UQ
β_1	1.0000	0.9894	0.1170	0.0935	0.8938	0.9825	1.0591
β_2	1.0000	1.0093	0.1058	0.0860	0.9319	1.0018	1.0835

Note : $x_1 \sim$ *Logistic distribution with mean* 0 *and variance* 1; $x_2 \sim$ *Uniform* (-2,2).
MEAN : sample mean; SD: standard deviation; RMSE : root-mean-squared error; LQ : lower quantile; MEDIAN : median; UQ : upper quantile.

Ⅳ. An Empirical Application-Dong River Dam Project in Korea

In this section, the semiparametric estimator in (8) is compared with the MLE to show the misspecification bias problem. Two methods were applied to estimate the WTP for preserving the Dong River area in Korea. It is important to note that the main purpose of the empirical analysis in this paper is to demonstrate the misspecification bias with real data rather than to provide a meaningful WTP estimate for the Dong River project.

The Korean government is planning to construct a mammoth multi-purpose dam on the lower reaches of the Dong River. The main purpose of the dam is to provide a better flood control measure for the Han River that runs through the city of Seoul where more than 10 million residents live.

But the backlash for such project is very strong and immediate. The antagonistic group worries about the environmental havoc including the collapse of limestone species, extermination of many endangered species as well as the substantial declines in agricultural production.

A simple semiparametric CV model is assumed as (11)

$$\ln WTP_i = \beta_0 + \beta_1 \ln EDU_i + \beta_2 \ln INCOME_i + \varepsilon_i \qquad (11)$$

where *EDU* is the education level measured by years of schooling and *INCOME* is the household average monthly after-tax income.

A CV survey data set collected to estimate the WTP for preserving the Dong River area was used. The data set contains 150 respondents who live in Seoul area out of 330 people who were interviewed during March and April 1999. Each respondent was asked to answer whether they are willing to pay for the bidding price to preserve the Dong River area given a randomly selected bidding price from 15 different bidding prices ranged from 1,000 Won to 20,000 Won.

<Table 2> shows the estimation results of WTP model (11) using the responses and individual characteristics. It appears that a misspecification might be the main cause for the difference of coefficient estimates between two estimation methods. Especially, the income elasticity estimated by MLE is around two times larger than the one estimated by the semiparametric method.

〈Table 2〉 Estimation Results of WTP Function

	β_0	β_1	β_2
Semiparametric estimation	-	1.8493	1.1822
Maximum likelihood estimation	-5.0355	1.7211	2.4754

V. Concluding Remarks

In a traditional CVM analysis, the parametric coefficient estimators would have a misspecification bias unless the model is correctly specified. This paper proposes a new semiparametric estimator based on the well-known density weighted average derivative of regression function. The estimator is robust to the misspecification because it does not depend upon any property of error distribution except that the conditional mean is zero. A Monte Carlo simulation confirms the robustness of the estimator. And the misspecification bias was also demonstrated using an empirical data set.

However, the estimator is only useful to estimate the coefficients associated with the continuous and differentiable covariates and requires a fairly large sample size which is a common problem of the nonparametric estimation methods.

There are many issues waiting further studies yet. Among them, a specification test statistic could be developed using the estimator proposed here, which can be useful to test the specification of the model.

〈Appendix〉

Powell *et al.* (1989) proposed an estimator of the density-weighted average derivative of a general regression function and applied it to the estimation of coefficients in single index models. The density-weighted average derivative vector is defined as:

$$\delta \equiv E\left[f(x)\frac{\partial g(x)}{\partial x}\right]$$

where $g(x)(=E[y|x])$ is the true regression function, y denotes a dependent variable, and $f(x)$ is the *pdf* of the independent variables x.

Suppose $g(x)$ can be written in a single index form as $g(x) = G(x'\beta)$ for some univariate function $G(\cdot)$, the density-weighted average derivative is proportional to β:

$$\delta \equiv E\left[f(x)\frac{\partial g(x)}{\partial x}\right] = E\left[f(x)\frac{dG(x'\beta)}{d(x'\beta)}\right]\beta = \gamma\beta$$

where γ is a proportional factor. Therefore, an estimator of the density-weighted average derivative can be used for estimation of β up to a scale. Powell *et al.* (1989) proposed a nonparametric estimator of the density-weighted average derivative δ as:

$$\delta = \frac{-2}{N(N-1)}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\left(\frac{1}{h}\right)^{k+1}K'\left(\frac{x_i - x_j}{h}\right)(y_i - y_j)$$

where $K'(\cdot)$ is the derivative of the kernel function, N is the sample size, h is a bandwidth for nonparametric estimation of regression function and k is the number of the independent variables.

Powell *et al.* (1989) also proved the consistency and asymptotic normality of the estimator as in the following theorem.

Theorem

Given the regularity conditions on the density function of x and the kernel function $K(\cdot)$, if h obeys $Nh^{k+2} \rightarrow \infty$ and $Nh^{2p} \rightarrow 0$ as $N \rightarrow \infty$, then

$$\sqrt{N}\,(\hat{\delta} - \delta) \xrightarrow{d} N(0, \Sigma_\delta)$$

where k is the dimension of x; $p = (k+4)/2$ if k is even number or $p = (k+3)/2$ if k is odd number; and $\sigma_\delta = 4E[r(z)r(z)^T] - 4\delta\delta'$, $r(z) = f(x)\frac{\partial g(x)}{\partial x} - [y - g(x)]\frac{\partial f(x)}{\partial x}$.

◎ References ◎

1. Chen, Heng Z. and Alan Randall, "Semi-nonparametric Estimation of Binary Response Models with an Application to Natural Resource Valuation," *Journal of Econometrics* 76, 1997, pp. 323∼340.
2. Cooper, Joseph C., "Flexible Functional Form Estimation of Willingness to Pay Using Dichotomous Choice Data," *Journal of Environmental Economics and Management* 43, 2002, pp. 267∼279.
3. Creel, Michael and John Loomis, "Semi-nonparametric Distribution Dichotomous Choice

Contingent Valuation," *Journal of Environmental Economics and Management* 32, 1997, pp. 341～358.

4. Diamond, P. A. and J. A. Hausman, "Contingent Valuation : Is Some Number Better Than No Number?," *Journal of Economic Perspectives* 8, 1994, pp. 45～64.
5. Powell, James L., Stock, James H. and Thomas M. Stoker, "Semiparametric Estimation of Index Coefficients," *Econometrica* 57, 1989, pp. 1435～1460.
6. Powell, James L. and Thomas M. Stoker, "Optimal Bandwidth Choice for Density-weighted Averages," *Journal of Econometrics* 75, 1996, pp. 291～316.
7. Silverman, B. W., *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London, 1986.

A Semiparametric Estimation of the Contingent Valuation Model

Joo Heon Park

A new semiparametric estimator of a dichotomous choice contingent valuation model is proposed by adapting the well-known density weighted average derivative of the regression function. A small sample behavior of the estimator is demonstrated very briefly by a simulation and the estimator is applied to estimate the WTP for preserving the Dong River area in Korea.

Keywords : Contingent Valuation Method, Willingness to Pay, Semiparametric Estimation
JEL Classification : C14, C35, Q261.

조건부가치평가모형의 준모수 추정

박 주 헌

양분형 조건부가치평가모형의 준모수적 추정 방법을 소위 회귀함수 1차 도함수의 밀도가중평균(density weighted average derivative of regression function) 추정을 응용하여 제안한다. 논문에서 제안된 준모수 추정량의 소표본 특성은 몬테칼로 시뮬레이션 결과를 제시함으로써 간접적으로 나타낸다. 또 추정량을 동강보존을 위한 지불용의액을 조사한 조건부가치평가자료에 실제 적용함으로써 현실 적용 가능성을 보여준다.

주제어 : 조건부가치평가, 지불용의액, 준모수추정