

Generalized Estimation Methods for Non-i.i.d. Binary Data: An Application to Dichotomous Choice Contingent Valuation

SOO-IL KIM

sikim@keei.re.kr

Dept. of Energy Policy Studies, Korea Energy Economics Institute
665-1 Naeson 2-dong, Euiwang-si, Kyonggi-do, 437-713, Korea

Timothy C. Haab

haab.1@osu.edu

AED Economics, The Ohio State University
2120 Fyffe Road, Columbus, Ohio United States 43210

*Selected Paper prepared for presentation at the American Agricultural Economics
Association Annual Meeting, Providence, Rhode Island, July 24-27, 2005*

*Copyright 2005 by SOO-IL KIM and Timothy C. Haab. All rights reserved. Readers may
make verbatim copies of this document for non-commercial purposes by any means,
provided that this copyright notice appears on all such copies.*

Generalized Estimation Methods for Non-i.i.d. Binary Data: An Application to Dichotomous Choice Contingent Valuation

SOO-IL KIM and Timothy C. Haab

Abstract: We challenge the assumption of i.i.d random utility across alternatives embedded in typical applications of logit models to dichotomous choice contingent valuation data. Using a Gumbel mixed distribution which nests a number of traditional models, we show that the logistic distribution is not a suitable distribution for contingent valuation analysis.

1. Introduction

Including dichotomous choice contingent valuation (DCCV), widely different fields of economics have used the binary data based on the random utility model. The binary choice in DCCV is “yes (or one)” if the random utility after environmental change is still greater than that of the current state, and “no (zero)” otherwise. With the assumption of *i.i.d.* type I extreme value for the distribution of the unobserved term, the random utility models can be estimated through a simple logit model (See Haab and McConnell 2002).

The simplicity and robustness of the estimation model, however, are the result of strong assumptions or constraints on the decision model rather than the natural outcome of correct specification of the model¹. The main problem is that *i.i.d.* assumption across alternatives can be far from the real choice situation. First, since the state after environmental change is uncertain to the respondent in spite of surely increasing environmental quality, the variance of the additive error term in the proposed state may be different from that in the current random utility. Second, the existence of alternative project that respondents prefer but the researcher does not consider in the CV survey, can lead the respondent to refuse the proposed project even though respondent agrees with the change in environmental quality. In this situation, the simple logit is not suitable estimation model and yields an incorrect measure of parameters or welfare change.

Undoubtedly, there has been a series of studies to relax the *i.i.d.* assumption in the logit model. For example, the heteroskedastic extreme value model has been suggested in the transportation (Bhat 1995) and marketing literatures (Allenby and Ginter 1995) to incorporate heteroskedasticity across alternatives into the multinomial or conditional logit models. However, no literature has paid attention to the strict assumption of identical error distributions across alternatives in the choice set. Since alternatives are only two and the scale and level of the utility is immaterial, variance-covariance parameters cannot be estimated with the binary choice by conventional models including generalized extreme values such as nested logit, paired combinatorial model, etc².

¹ We focus only on the logit model. When we assume the bivariate normal distribution, the choice probability still follows the univariate normal. The probit model, however, can identify only one parameter related with error distribution because of normalization. For details, see Train (2003).

² Note that these models have only $J(J - 1)/2 - 1$ covariance parameters after normalization, where J is the number of total alternatives

In this paper, we relax the identical and independent disturbance assumption of the random utility model by utilizing Gumbel mixed model. Gumbel mixed model is an asymptotic bivariate distribution of maxima, which provides us a scale parameter of one state normalized by the scale parameter of the other state and an association parameter related with correlation coefficient. The derivation of Gumbel mixed model is explained in the section 2. In section 3, we introduce two estimation techniques with Gumbel mixed model: approximation using Gaussian quadrature and simulation using mixed logit model. The generalized estimation procedures are applied to two contingent valuation studies in the section 4. Section 5 applies the generalized estimation method into the alternative choice model (willingness to pay function from expenditures difference), in which we found the same result as the random utility model. The section 6 concludes the analysis.

2. Gumbel Mixed Model of Bivariate Extreme Values Distribution

Including Gumbel (1960, 1961), Gumbel and Mustafi (1967) and Tiago de Oliveira (1980, 1983), a series of papers has introduced several bivariate extreme value distributions including the Gumbel mixed model which is one of differentiable bivariate extreme value distributions³. Let $F(\varepsilon_0, \varepsilon_1)$ be a asymptotic distribution of bivariate extreme values of maxima for ε_0 and ε_1 with Gumbel margins, $F(z)$. The asymptotic distribution of bivariate maxima is defined as

$$(1) \quad F(\varepsilon_0, \varepsilon_1) = [F(\varepsilon_0)F(\varepsilon_1)]^{k(\tau)}$$

where $k(\cdot)$ is called the dependence function representing the asymptotic connection between ε_0 and ε_1 , and τ is reduced difference defined as $\varepsilon_0/\theta_0 - \varepsilon_1/\theta_1$. θ_i is a scale factor and the location factor is assumed to be equal to zero.

Different bivariate distributions are derived using different dependence functions, of which the Gumbel mixed model has $k(\tau|\lambda) = 1 - \lambda \exp(\tau)/(1 + \exp(\tau))^2$, where λ is an

³ For other examples of parametric families of bivariate extreme value distributions, see Kotz and Nadarajah, 2000. Applications of Gumbel mixed model can be found in the hydrological engineering studies (Yue 2000, Yue et al. 1999).

association parameter⁴. Plugging the dependence function into equation (1) and using the definition of Gumbel margins, $F_{\varepsilon_i}(z) = \exp(-\exp(-z/\theta_i))$, the Gumbel mixed model is expressed as

$$(2) \quad F(\varepsilon_0, \varepsilon_1 | \Gamma) = \exp \left[- \left(\exp \left(-\frac{\varepsilon_0}{\theta_0} \right) + \exp \left(-\frac{\varepsilon_1}{\theta_1} \right) \right) + \frac{\lambda}{\exp(\varepsilon_0/\theta_0) + \exp(\varepsilon_1/\theta_1)} \right]$$

where Γ is a parameter set of scale factor (θ_0, θ_1) and association factor (λ) . Note that the expected value and the variance of independent extreme value ε_i are $E(\varepsilon_i) \approx 0.57722\theta_i$ and $Var(\varepsilon_i) = \theta_i^2 \pi^2 / 6$. Figure 1 shows the contour of the Gumbel mixed bivariate distribution function with $\lambda = 0.5$. For $\lambda = 0$, the joint distribution is independent such that $F(\varepsilon_0, \varepsilon_1) = F(\varepsilon_0)F(\varepsilon_1)$. The correlation coefficient is a function of the association parameter λ .

From the Gumbel mixed distribution, several important distributions are derived; probability density function, conditional distribution and distribution of reduced difference. The probability density function is derived by differentiating (2) with respect to ε_0 and ε_1 such that

$$(3) \quad f(x, y) = \frac{F(x, y)}{\theta_1 \theta_2} \left[\frac{2\lambda e^{x/\theta_1 + y/\theta_2}}{(e^{x/\theta_1} + e^{y/\theta_2})^3} + \left\{ e^{-x/\theta_1} - \frac{\lambda e^{x/\theta_1}}{(e^{x/\theta_1} + e^{y/\theta_2})^2} \right\} \left\{ e^{-y/\theta_2} - \frac{\lambda e^{y/\theta_2}}{(e^{x/\theta_1} + e^{y/\theta_2})^2} \right\} \right].$$

The contour of probability function is shown in Figure 2. As can be seen in Figure 1 and 2, the bivariate extreme value distribution is upper-right skewed.

[Figure 1-2 located here]

The conditional cumulative distribution function of the Gumbel mixed model is

$$(4) \quad F_{\varepsilon_0|\varepsilon_1}(\varepsilon_0) = F(\varepsilon_0, \varepsilon_1) \left\{ \exp \left[\exp(-\varepsilon_1/\theta_1) \right] - \lambda \frac{\exp \left[2\varepsilon_1/\theta_1 + \exp(-\varepsilon_1/\theta_1) \right]}{\left[\exp(\varepsilon_0/\theta_0) + \exp(\varepsilon_1/\theta_1) \right]^2} \right\}$$

⁴ The logistic model, one of differentiable bivariate extreme value distribution, is derived using the difference function of $k(\tau | \lambda) = [1 + \exp(-\tau/(1-\lambda))]^{1-\lambda} / [1 + \exp(-\tau)]$. Unfortunately, the logit model, i.e. the generalized extreme value with two alternatives, cannot identify the association factor λ .

from $f_{\varepsilon_0|\varepsilon_1} = f(\varepsilon_0, \varepsilon_1) / f_{\varepsilon_1}(\varepsilon_1)$ (Yue 2000). The distribution function of reduced difference is derived as (Tiago de Oliveira 1980)

$$(5) \quad D(\tau | \lambda) = \frac{\exp(\tau)}{1 + \exp(\tau)} \frac{(1 + \exp(\tau))^2 - \lambda}{(1 + \exp(\tau))^2 - \lambda \exp(\tau)}$$

Figure 3 and 4 show the cumulative distribution and probability function of the reduced difference with various λ . The probability density function of reduced difference is symmetric around zero mean. For $\lambda = 0$, i.e. independent case, the conditional distribution (4) reduces to be a univariate type I extreme value and the difference distribution (5) becomes a logistic distribution.

[Figure 3-4 located here]

3. Random Utility and the Probability of Binary Choice

3.1. Random Utility Model

A standard random utility consists of two parts; a systematic component observable to researcher and an error component that is known to respondent but not necessarily. Let the random utility of individual n with alternative i be $U_{in} = V_i(I_n, z_n) + \varepsilon_{in}$, where V_i is the systematic component and ε_i is the error component. Alternatives in the binary choice set are the proposed state representing to accept the policy and the current state without change indicating to reject the policy⁵. The systematic part is the function of respondent's income (I_n) and the vector of respondent's characteristics and choice attributes (z_n). The probability of choosing the state one is;

$$(6) \quad P_{1n} = P(U_{0n} < U_{1n}) = P(V_{0n} + \varepsilon_{0n} < V_{1n} + \varepsilon_{1n}) = P(\varepsilon_{0n} < v_n + \varepsilon_{1n})$$

where $v_n = V_{1n} - V_{0n}$. Further progress in estimation is feasible by specifying a parametric form for both of the systematic component and the error distribution in equation (6). The systematic component is usually assumed linear in parameters even though only linearity in income is sufficient. The error component in the standard additive random utility for discrete choice case is assumed independent and identical distribution over states

⁵ We assumes that 'to be uncertain' responses are grouped as 'no' response for conservative reason. For details of 'uncertain' response issue, see Carson et al. 1998; Groothuis and Whitehead 1998.

(Karlström 1999). With *i.i.d.* type I extreme value (or Gumbel) distribution, the derivation of the logistic distribution for the difference of two identical extreme values is straightforward. In addition, McFadden (1974) also shows that the logit formula for the choice probabilities implies extreme value distribution for the random utility.

The classical assumption about the additive error components, however, may be wrong because of several reasons. For example, uncertainty in the future, reliability on the implementation and result of the project, etc, can be possible source that respondent accept the proposed and current states differently. More uncertainty in the proposed state introduces larger variance of the distribution. The proposed state is random and unobservable even to the respondent, which is different from the unobservability of the current state. Another possibility of violation of the assumption is that, if respondents have alternative options instead of the proposed policy that may be unknown to researcher, the response of reject represents either staying without change (the current state) or changing through other process (the future state possibly for different level of environmental quality)⁶. In addition to the possible heteroskedasticity due to unknown alternatives, the current and proposed states may be correlated if the unknown alternatives have the same goal of the environmental change with the proposed policy in the survey. Because of those reasons, we name the current state as the reference state to avoid misinterpretation.

3.2. Approximation of the Log Likelihood

Regardless of the heteroskedasticity and correlation, the choice probability in the equation (6) can be expressed as an integration of the conditional distribution over marginal distribution;

$$(7) \quad P_{1n} = \int_{\varepsilon_1 = -\infty}^{\varepsilon_1 = +\infty} F_{\varepsilon_0 | \varepsilon_1} (v_n + \varepsilon_1) f(\varepsilon_1) d\varepsilon_1 .$$

⁶ Train (2003) defined three characteristics that alternatives in the choice set should satisfy: exclusiveness, exhaustiveness and countable finiteness. To vote for and vote against are mutually exclusive and finite. For exhaustiveness, the current state without change includes not only the state without change but also all possible changes except the policy proposed in the survey. Furthermore, NOAA panel report (Arrow et al. 1993) recommends the reminder of substitute commodities among guideline for designing contingent valuation questions, such as other comparable natural resources or the future state of the same resource to assure that respondents have the alternatives clearly in mind (Haab and McConnell, 2002). Haab and Hicks (1999) has broadly surveyed the choice set issues in recreation demand modeling.

The equation (7) is a general expression of the choice probability nesting a simple logit, heteroskedastic extreme values and common scale factor model. Since the general expression of bivariate extreme value model in equation (7) does not have the closed form for the integration, we need to approximate or simulate the choice probability.

As already shown in the multinomial heteroskedastic case by Bhat (1995), the approximation procedure utilizing Gaussian quadrature can provide a fast and highly accurate proxy of the choice probability⁷. Define a new variable such that $u = \exp(-\exp(-w))$, thus $w = -\ln(-\ln u)$ and $du = \exp(-w)\exp(-\exp(-w))dw$ ⁸. The new variable u is the form of cumulative distribution of extreme value and has the support of $[0,1]$. This transformation enables us to approximate the choice probability much easier through Gaussian-Legendre quadrature. Let $\varepsilon_1 = \theta_1 w_1$ and $\gamma = \theta_1 / \theta_0$, then the conditional density and marginal probability functions are $F_{\varepsilon_0|\varepsilon_1}(v_n + \varepsilon_1) = F_{\varepsilon_0|\varepsilon_1}(v_n + \gamma w_1)$ and $f(\varepsilon_1) = f(w_1) / \theta_1$. The arguments in the conditional probability is normalized by θ_0 . Plugging the new variable u into the choice probability function, the choice probability becomes $P_{1n} = \int_{u=0}^{u=1} G(v_n, u) du$ since $d\varepsilon_1 = \theta_1 dw_1$, where $G(v_n, u) = F_{\varepsilon_0|\varepsilon_1}(v_n - \gamma \ln(-\ln u))$. The integration can be substituted by Gaussian-Legendre quadrature such as

$$\int_{u=0}^{u=1} G(v_n, u) du \approx \sum_{l=1}^L \xi_l G(v_n, u_l) = \hat{P}_{1n}$$

where ξ_l and u_l are L weights and support points (abscissas) of Gaussian-Legendre quadrature. The log likelihood function is approximated as

$$\log L = \sum_{n=1}^N y_n \log \sum_{l=1}^L \xi_l G(v_n, u_l) + (1 - y_n) \log \left\{ 1 - \sum_{l=1}^L \xi_l G(v_n, u_l) \right\}.$$

3.3. Simulation of the Log Likelihood

The mixed logit model introduced into recreation model by Train (1998, 1999) directly simulates the choice probability in the equation (6). Let the true random utility to

⁷ Alternatively, Allenby and Ginter (1995) also suggest the Bayesian estimation procedure for heteroskedastic extreme values.

⁸ Bhat (1995) uses the transformation of $u = \exp(-w)$ with the support of $[0, \infty]$ and applies a Gaussian-Laguerre quadrature.

be $U_{in} = \zeta'_{in} z_{in}$, where $z'_{in} = (x'_{in}, d_i)$ and $\zeta'_{in} = (\beta'_{in}, \varepsilon_{in})$. By rescaling the utility upward sufficiently (s) and adding an *i.i.d.* extreme value terms on both sides, the resulting choice probability is expressed such as

$$P_{1n} = \int \left(\frac{\exp\left[\left(\zeta'_{1n} / s\right) z_{1n}\right]}{\sum_{j=0,1} \exp\left[\left(\zeta'_{jn} / s\right) z_{jn}\right]} \right) f(\zeta) d\zeta$$

where $f(\zeta)$ is a joint density of β_{jn} and ε_{jn} ⁹. Suppose that coefficients of systematic part of utility are invariant across individual ($\beta_{in} = \beta_i$) and the joint density of $f(\zeta)$ is a bivariate distribution of ε_{0n} and ε_{1n} . Then the mixed logit model becomes

$$(8) \quad P_{1n} = \int L_{1n}(\zeta) f(\varepsilon_0, \varepsilon_1) d\langle \varepsilon_0, \varepsilon_1 \rangle$$

where

$$(9) \quad L_{1n}(\zeta) = \frac{\exp\left[v_n / s + (\varepsilon_{1n} - \varepsilon_{0n}) / s\right]}{1 + \exp\left[v_n / s + (\varepsilon_{1n} - \varepsilon_{0n}) / s\right]}.$$

Because of the equivalence to the logit smoothed-AR simulator (McFadden 1989), the estimation of the mixed model follows the simulation procedure of ‘logit kernel probit’ (Ben-Akiva and Bolduc 1996) adjusted simply for the bivariate extreme values. However, since the random draw from a bivariate extreme value distribution is unavailable, we employ an importance sampling procedure with Halton sequence to simulate the random draw from bivariate extreme values. The importance sampling provides simulated random variables with correlation and heteroskedasticity by transforming the original density, named target density, into a density from which it is easy to draw, named a proposal density (Train 2003).

Let $g(\varepsilon_i)$ be a univariate extreme value distribution. Define the weight as

⁹ The mixed logit model, usually, has employed a joint distribution of parameters β in the systematic component of random utility. The probability function of the random parameter is defined to be

$$P_{in} = \int \left(\frac{\exp(\beta' x_{in})}{\sum_j \exp(\beta' x_{jn})} \right) \phi(\beta) d\beta$$

where $\phi(\cdot)$ is the distribution function of parameters which can be flexibly assumed such as a normal (Provencher and Bishop 2004), lognormal (Bhat 2000), uniform or triangular (Train 2001) distribution. The mixed logit model can be applied to any choice model with any degree of accuracy by assuming appropriate distribution (Train 2003, McFadden and Train 2000). By assuming that parameters have an individual and alternative specific randomness, the mixed model relaxes the IIA assumption and represents any pattern of substitution among alternatives.

$$(10) \quad \frac{f(\varepsilon_0, \varepsilon_1)}{g(\varepsilon_0)g(\varepsilon_1)} = \Psi(\varepsilon_0, \varepsilon_1) \exp \left\{ (\varepsilon_0/\theta_0 + \varepsilon_1/\theta_1) + \frac{\lambda}{\exp(\varepsilon_0/\theta_0) + \exp(\varepsilon_1/\theta_1)} \right\}$$

where $f(\varepsilon_0, \varepsilon_1)$ is the joint density in the mixed logit model and

$$\Psi(x, y) = \frac{2\lambda e^{x/\theta_0 + y/\theta_1}}{(e^{x/\theta_0} + e^{y/\theta_1})^3} + \left\{ e^{-x/\theta_0} - \frac{\lambda e^{x/\theta_0}}{(e^{x/\theta_0} + e^{y/\theta_1})^2} \right\} \left\{ e^{-y/\theta_1} - \frac{\lambda e^{y/\theta_1}}{(e^{x/\theta_0} + e^{y/\theta_1})^2} \right\}.$$

Using the fact that $\varepsilon_i = \theta_i w_i$ and $g(\varepsilon_i) = (1/\theta_i)g(w_i)$, the choice probability of mixed logit model becomes

$$\hat{P}_{1n} = \int L_{1n}(\varepsilon) \Psi(w_0, w_1) \exp \left\{ (w_0 + w_1) + \frac{\lambda}{\exp(w_0) + \exp(w_1)} \right\} d\langle w_0, w_1 \rangle,$$

since multiplying the integrand of equation (8) by $g(\varepsilon)/g(\varepsilon)$ does not change the original choice probability.

Application of importance sampling to the mixed logit model is as follows: (1) Take draws for w_0 and w_1 from a standard extreme value distribution and construct two-dimensional independent random variables. In this first step and through the repetition, Halton sequence is used to draw standard extreme values¹⁰. (2) For this draw, calculate the logit formula, L_{1n} with prespecified scaling factor (s), and the weight function of the equation (10). (3) Repeat two steps enough times and take the average of the result,

$\hat{P}_{1n} = \frac{1}{R} \sum_r P_{1n}$, which is an unbiased estimate of the choice probability with correlation

and heteroskedasticity. The probability of choosing the alternative zero is $\hat{P}_{0n} = 1 - \hat{P}_{1n}$. The simulated log likelihood function becomes

$$\log L = \sum_{n=1}^N y_n \log \hat{P}_{1n} + (1 - y_n) \log (1 - \hat{P}_{1n}).$$

3.4. Estimation of Welfare Change

We define the expected welfare change (willingness to pay for the environmental change) as the expected maximum income that equates the expected random utility in two

¹⁰ Halton sequence reduces the number of draws and the simulation error associated with a given number of draws. The simulation error with 125 Halton draws is smaller than even with 2000 random draws. See Bhat (1999), Train (1999, 2003) and Greene (2002).

states¹¹. Assume that the systematic component of the random utility is linear in the income and the marginal utility of income is constant (α) across individuals and states, i.e. no income effect, then the expected willingness to pay becomes

$$(11) \quad E(WTP_n) = \frac{1}{\alpha} v_n + \frac{1}{\alpha} E(\varepsilon_{1n} - \varepsilon_{0n}).$$

Due to the linearity assumption of the income, the income variable is not included in v_n . While the expectation of the error term of the logit model is zero by including a constant term in the systematic component, the expected value of error terms in equation (11) is not zero. As explained below, it is much convenient for estimating equation (11) to remain the expectation term.

In general case, we can estimate the expectation of error differences through a simulation procedure using estimated relative scale and association parameters since the exact moment of error difference in Gumbel mixed model is unknown. Note that the expected value of error difference is the integration of random variables over Gumbel mixed bivariate probability;

$$E(\gamma w_{1n} - w_{0n}) = \int (\gamma w_{1n} - w_{0n}) f(w_{1n}, w_{0n}) d\langle w_{1n}, w_{0n} \rangle.$$

By reapplying importance sampling procedure with Halton sequence, the expected willingness to pay becomes

$$E(WTP_n) = x'_n \frac{\tilde{\beta}}{\tilde{\alpha}} + \frac{1}{\tilde{\alpha}} \hat{E}(\gamma w_{1n} - w_{0n}).$$

since $E(\varepsilon_{1n} - \varepsilon_{0n}) / \alpha$ is equivalently $E(\gamma w_{1n} - w_{0n}) / \tilde{\alpha}$.

Except the general case, however, the expected willingness to pay can be exactly calculated. In the heteroskedastic case, the expectation of $\varepsilon_{1n} - \varepsilon_{0n}$ is approximately $0.57722 \cdot (\theta_1 - \theta_0)$, providing the final expression of the expected willingness to pay as $E(WTP_n) \approx v_n / \tilde{\alpha} + 0.57722(\gamma - 1) / \tilde{\alpha}$. For identical cases, the willingness to pay is simply $E(WTP_n) = v_n / \tilde{\alpha}$.

¹¹ A series of papers has investigated the correct welfare measure consistent with the microeconomic theory. The welfare measurement is incorrect if we estimate the models using the incorrect choice set (Kaoru et al. 1995). More seriously, a large difference of amount of money in a cost-benefit analysis has been found even though the welfare estimates from different model are similar (Hau 1986, Herriges and Kling 1999, Karlström 1999).

4. Applications to Dichotomous Choice Contingent Valuation Study

To test classical assumption, we apply Gumbel mixed model to the previous dichotomous choice CV studies. The study used in the estimation includes the sewage treatment in Barbados and the wastewater disposal system in Montevideo, Uruguay (McConnell and Ducci, 1989)¹². Observations in data were 1276 for Montevideo and 426 for Barbados data. For mixed logit model, the rescaling factor s was set to be 0.3 and the simulation was iterated 125 times. The association parameter (λ) was constrained to be between zero and one, and the relative scale factor (γ) was restricted to be nonnegative in the CML procedure of Gauss program¹³.

Table 1 and Table 2 show the estimation results of random utility model with Barbados and Montevideo data, respectively. The results consist of three sets; simple logit model in the second column, the result of bivariate extreme values in the third to sixth column and the result of mixed logit in the last four columns.

[Table 1 located here]

In Table 1, the first part of results of each estimation method is for the constrained model with independent and identical error ($\gamma = 1$, $\lambda = 0$) that is theoretically equivalent to the simple logit model. The constrained bivariate extreme value model with $\gamma = 1$ and $\lambda = 0$ provides exactly same parameter estimates with the simple logit model while the constrained mixed logit model with $\gamma = 1$ and $\lambda = 0$ has slightly different estimates from the simple logit model. However, both estimation models fail to estimate parameters with the constraint of $\lambda = 0$ due to too large relative scale estimate. When the correlation is allowed in estimation, i.e. in general model and constrained model with $\gamma = 1$, the association parameter is different from zero but not statistically significant. In addition, the relative scale parameter is not statistically different from one. LR statistics fail to reject the constraints for homoskedasticity or independence. Barbados data shows that the assumption of independent and identical distribution is suitable for estimation of random utility.

[Table 2 located here]

¹² The data is available in Haab and McConnell (2002).

¹³ The positive constraint can be assigned in the model by transforming the parameter such as exponential term. However, the estimation results for other parameters are not different and zero estimate of relative scale implies extreme difference of scale terms.

Table 2 also shows that bivariate extreme value model with constraints of $\gamma = 1$ and $\lambda = 0$ provides the estimation result closer to logit model than mixed logit model does. In both of bivariate extreme value and mixed logit models, association parameter estimates are statistically indifferent from zero except one case of mixed logit model. The relative scale estimates, however, are zero implying that the variance of the reference state is extremely larger than that of the proposed state. The association parameter estimates are not statistically different from zero. LR statistics fails to reject the constraint of independence ($\lambda = 0$), but heteroskedasticity is statistically significant.

[Table 3 located here]

Table 3 shows the sample average of the expected willingness to pay from Table 1 and 2¹⁴. RD indicates the random utility model and ED represents the expenditure difference model result explained in the next section. In spite of similar parameter estimates among different constrained models, the welfare measure from the change of environmental quality varies enormously depending on the relation of error terms. For instance, the sample average of the expected willingness to pay in Montevideo is estimated around -28 ~ -26 when $\gamma = 1$ is imposed, but it is estimated -81 ~ -65 without the constraint of $\gamma = 1$. Unfortunately, due to the failure of estimation, the willingness to pay cannot be estimated in two heteroskedastic models of random utility with Barbados data. However, the expected willingness to pay with the homoskedasticity constraint is also similar to logit model since independence has been found in most cases.

5. Expenditure Difference Model

An alternative model of explaining respondent's choice in dichotomous choice contingent valuation is the willingness to pay function derived from the expenditure functions. Let the minimum expenditure of individual n be $m_{0n} = m(q^0, u_n^0)$ at the reference state and $m_{1n} = m(q^1, u_n^0)$ at the proposed state where q^i is the environmental quality at state i . Then, the willingness to pay function is defined as a difference of two expenditure

¹⁴ Since the purpose of reporting willingness to pay is to compare the result from each estimation and decision model, monetary units are ignored in the table. Furthermore, by the assumption of linear function and infinite range of error distribution, the expected willingness to pay can be negative value.

function: $WTP(u^0) = m(q^0, u^0) - m(q^1, u^0)$. Like the random utility, the expenditure function consists of a systematic component (m_n^*) and an unobservable random component (η_n). The logistic distribution of the willingness to pay function implies that the underlying distribution of expenditure functions is the *i.i.d.* type I extreme value distribution. As can be recognized, the exactly same problems as random utility model arise in the willingness to pay function model.

While the random utility is derived from the utility maximization, the expenditure function is the minimized cost, requiring that the extreme value of the expenditure function is the smallest extreme value. The smallest extreme value is easily derived from the dual relation of $\min(Z_i) = -\max(-Z_i)$ (Tiago de Oliveira, 1983). The joint distribution function for minima with Gumbel reduced margins is

$$\Omega(\eta_0, \eta_1) = 1 - F(-\eta_0) - F(-\eta_1) + F(-\eta_0, -\eta_1),$$

where $F(\cdot)$ and $F(\cdot, \cdot)$ are marginal and joint distributions of maxima. The probability density function of bivariate extreme values of minima is $\omega(\eta_0, \eta_1) = f(-\eta_0, -\eta_1)$ by definition. From the dual relationship, we can derive the marginal distribution function and probability function as $\Omega(\eta_i) = 1 - F(-\eta_i)$ and $\omega(\eta) = f(-\eta)$. Note that the expected value of η is $E(\eta_i) \approx -0.57722\theta_i$. The conditional distribution of minima is derived from the conditional distribution of maxima in equation (9) such as

$$\Omega(\eta_1 | \eta_0) = \int_{\eta=-\infty}^{\eta=\eta_1} \omega(\eta | \eta_0) d\eta = 1 - F(-\eta_1 | -\eta_0)$$

since $\omega(\eta_1 | \eta_0) = f(-\eta_1 | -\eta_0)$. The distribution and probability functions of reduced difference of minima are identical to that of maxima.

The estimation of the willingness to pay function follows the same procedures in random utility model. The choice probability of expenditure difference is expressed as

$$P_{1n} = P(m_{1n}^* + b_n + \eta_1 < m_{0n}^* + \eta_0) = \int_{\eta_0=-\infty}^{\eta_0=+\infty} \Omega(m_n^* - b_n + \eta_0 | \eta_0) \omega(\eta_0) d\eta_0.$$

where $m_n^* = m_{0n}^* - m_{1n}^*$. The choice probability can be approximated by Gaussian quadrature such as $\hat{P}_{1n} = \sum_{l=1}^L \xi_l H(m_n^*, u_l)$ through appropriate transformation. The mixed logit model

simulates the choice probability as $P_{1n} = \int L_{1n}(\zeta) \omega(\eta) d\eta$ where $\eta = \langle \eta_0, \eta_1 \rangle$ and $L_{1n} = \left\{ 1 + \exp \left[- \left(m_n^* - b_n \right) / s - \left(\eta_{n0} - \eta_{n1} \right) / s \right] \right\}^{-1}$. The simulation procedure of the bivariate probability of minima is the same as the bivariate probability of maxima. However, all parameters are normalized by θ_1 rather than by θ_0 such as $\tilde{\beta} = (\beta_0 - \beta_1) / \theta_1$ for the systematic part, $\tilde{\beta}_b = 1 / \theta_1$ for the minus bid value and $\tilde{\gamma} = \theta_0 / \theta_1$ for the relative scale factor. Due to the bid variable, the expenditure difference model is able to identify both scale parameters. Finally, the expected willingness to pay is estimated as $E(WTP_n) = m_n^* + E(\eta_0 - \eta_1)$.

Table 4 and Table 5 report the estimation results of expenditure difference model using the same data in the random utility model. Note that the relative scale factor γ is estimated as θ_1 / θ_0 rather than θ_0 / θ_1 to enable the comparison with random utility model. Parameter estimates of the expenditure difference model are statistically duplicates of the random utility model, i.e. assumption of underlying distribution such as maxima or minima does not affect the estimation result. Table 4 shows that Barbados study satisfies the classical assumption of logit model in terms of parameter estimates and LR test statistics. Montevideo data in Table 5, however, rejects the homoskedasticity constraint but fails to reject the independence constraint in both estimation models as the random utility model (Table 2). However, in spite of that the relative scale estimate is statistically significantly less than one, parameter estimates of systematic component of expenditure difference are seemingly equivalent with that of the random utility model¹⁵. The expected willingness to pay in expenditure difference model is reported in Table 3.

[Table 4-5 located here]

6. Conclusions

We challenge the theoretical and technical background of the simple logit model often used for estimating willingness to pay from dichotomous choice contingent valuation applications. The simple logit model assumes that the respondent's evaluations of the two

¹⁵ Note that the parameters are normalized by θ_1 not by θ_0 .

states are stochastically independent and homoskedastic. However, when random utilities across states of the world are heteroskedastic or correlated, we cannot derive the logit model theoretically consistent with random utilities. In this paper, we suggest generalized estimation methods by utilizing Gumbel mixed model to relax restrictive assumptions of the standard random utility model. Nested within this generalized model are the heteroskedastic logit model and the simple logit. The nesting structure allows for straightforward tests of the homoskedastic-independent error assumptions. In addition to the random utility model, expenditure difference (willingness to pay) model was estimated using Gumbel mixed bivariate distribution of minima. Again, this model has nested within it a number of standard logit-expenditure difference models.

Estimation results from several existing data including Barbados and Montevideo data show that correlation between two states is usually minimal, but Montevideo data presents extremely different scale of error terms across states implying that the *i.i.d.* extreme values, i.e. logistic distribution for the difference of error terms, may not be a suitable distribution. Serious problem arises in estimation of welfare measure. Heteroskedasticity or correlation provides willingness to pay estimate different from estimate of the simple logit, thus different policy implication in benefit-cost analysis.

In spite of the simplicity and profound theory of binary choice logit model, much careful consideration is required to apply the model into contingent valuation studies. Various estimation models do not suggest different decision process but indicate that due to the nature of decision process, the estimation result from simple logit model could be incorrect. However, as we mentioned before, if random utility or expenditure function follows the bivariate normal distribution, the binary choice probability is still a normal distribution. The similar estimation result between probit and logit models arises the question about the comparison between models with bivariate normal and bivariate extreme value distributions. Decision of which estimation model should be used in practice is based on the how researchers define the choice situation and choice set, but when they employ the logit model, the assumption of the model should be tested in priori.

References

- Allenby, G.M. and J.L. Ginter, 1995, "The effects of In-store Displays and Feature Advertising on Consideration Sets," *International Journal of Research in Marketing*, 12, 67-80.
- Arrow, K., R. Solow, E. Leamer, P. Portney, R. Rander, and H. Schuman, 1993, "Report of the NOAA Panel on Contingent Valuation." *Federal Register*: 4601-14.
- Ben-Akiva, M. and D. Bolduc, 1996, "Multinomial Probit with a Logit Kernel and a General Parametric Specification of the Covariance Structure," Working Paper, Department of Civil Engineering, MIT.
- Bhat, C.R., 1995, "A Heteroskedastic Extreme Value Model of Intercity Travel Mode Choice," *Transportation Research B*, 29(6), 471-483.
- Bhat, C.R., 1997, "Covariance Heterogeneity in Nested Logit Models: Econometric Structure and Application to Intercity Travel," *Transportation Research B*, 31(1), 11-21.
- Bhat, C.R., 1999, "An analysis of evening commute stop-making behavior using repeated choice observation from a multi-day survey," *Transportation Research B* 33, 495-510.
- Bhat, C.R., 2000, "Incorporating observed and unobserved heterogeneity in urban work mode choice modeling," *Transportation Science*, 34, 228-238
- Carson, R.T., W.M. Hanemann, R.J. Kopp, J.A. Krosnick, R.C. Mitchell, S. Presser, P.A. Ruud, and V.K. Smith, 1994, "Prospective Interim Lost Use Value Due to DDT and PCB Contamination in the Southern California Bight," NOAA Contract No. 50-DGNC-1-00007.
- Chen, H.Z. and S.R. Cosslett, 1998, "Environmental Quality Preference and Benefit Estimation in Multinomial Probit Models: A Simulation Approach," *American Journal of Agricultural Economics*, 80, 512-520.
- Chu, C., 1981, "Structural Issues and Sources of Bias in Residential Location and Travel Choice Models," Ph.D. Thesis, Northwestern University.
- Chu, C., 1989, "A Paired Combinational Logit Model for Travel Demand Analysis," *Proceedings of Fifth World Conference on Transportation Research* 4, 295-309.
- Greene, W.H., 2002, *NLOGIT Version 3.0 Reference Guide*, Econometric Software Inc., New York.
- Groothuis and Whitehead, "Does Don't Know Mean No? Analysis of 'Don't Know' Responses in Dichotomous Choice Contingent Valuation Question," unpublished, 1998.
- Gumbel, E.J., 1960, "Multivariate Extremal Distributions", *Bull. Inst. Internat. De Statistique*, 37(2), 471-475.

- Gumbel, E.J., 1961, "Bivariate Logistic Distribution," *Journal of American Statistical Association*, 56(294), 335-349.
- Gumbel, E.J. and C.K. Mustafi, 1967, "Some Analytical Properties of Bivariate Extreme Distributions," *Journal of the American Statistical Association*, 62(318), 569-588.
- Haab, T.C. and R. Hicks, 1999, "Choice Set Considerations in Models of Recreation Demand: History and Current State of Art," *Marine Resource Economics*, 14, 255-270.
- Haab, T.C. and K.E. McConnell, 2002, *Valuing Environmental and Natural Resources: The Econometrics of Nonmarket Valuation*, Edward Elgar Publisher.
- Hanemann, W.M., 1984, "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses," *American Journal of Agricultural Economics*, 66, 332-341.
- Hau, T.D.-K., 1986, "Distributional Cost-Benefit Analysis in Discrete Choice," *Journal of Transportation Economics and Policy*, 20, 313-338.
- Hensher, D., J. Louviere and J. Swait, 1999, "Combining Sources of Preference Data," *Journal of Econometrics*, 89, 197-221.
- Herriges, J.A. and C.L. Kling, 1999, "Nonlinear Income Effects in Random Utility Models," *The Review of Economics and Statistics*, 81, 62-73.
- Kaoru, Y., V.K. Smith and J.L. Liu, 1995, "Using Random Utility Models to Estimate the Recreational Value of Estuarine Resources," *American Journal of Agricultural Economics*, 77, 141-151.
- Karlström, A., 1999, "On the Role of Income in Welfare Evaluations with RUM," 38th Annual Meeting of the Western Regional Science Association, California.
- Kotz, S. and S. Nadarajah, 2000, *Extreme Value Distributions: Theory and Applications*, Imperial College Press.
- McConnell, K.E. and J.H. Ducci, 1989, "Valuing Environmental Quality in Developing Countries: Two Case Studies," presented at the ASSA annual meetings, Atlanta, GA.
- McFadden, D., 1974, "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics* ed. by Zarembka, P., Academic Press, New York.
- McFadden, D., 1989, "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, 57, 995-1026.

- McFadden, D., 1999, "Computing Willingness-To-Pay in Random Utility Models," in *Trade, Theory, and Econometrics: Essays in Honour of John S. Chipman* ed. by Moore, J., R. Riezman and J. Melvin, Routledge.
- McFadden, D. and K. Train, 2000, "Mixed MNL Models for Discrete Response," *Journal of Applied Econometrics*, 15, 447-470.
- Provencher, B. and R.C. Bishop, 2004, "Does Accounting for Preference Heterogeneity Improve the Forecasting of a Random Utility Model? A Case Study," *Journal of Environmental Economics and Management*, 48(1), 793-810.
- Smith, V.K., 1996, "Pricing What is Priceless: A Status Report on Non-market Valuation of Environmental Resources," in *International Yearbook of Environmental and Resource Economics 1997/1998: A Survey of Current Issues* ed. by Folmer, H. and T. Tietenberg, New Horizons in Environmental Economics Series, Edward Elgar.
- Straud, A.H. and D. Secrest, 1966, *Gaussian Quadrature Formulas*, Series in Automatic Computation, Prentice-Hall, London.
- Tawn, J.A., 1988, "Bivariate Extreme Value Theory: Models and Estimation," *Biometrika*, 75(3), 397-415.
- Tiago de Oliveira, J., 1980, "Bivariate Extremes: Foundations and Statistics," in *Multivariate Analysis V: proceeding of the fifth international symposium on multivariate analysis* ed. by P.R. Krishnaiah, North-Holland Publishing Company.
- Tiago de Oliveira, J., 1983, "Bivariate Models for Extremes; Statistical Decision," in *Statistical Extremes and Application* ed. by J. Tiago de Oliveira, NATO ASI Series C: Mathematical and Physical Science Vol. 131. D. Reidel Publishing Company.
- Train, K.E., 1998 "Recreation Demand Models with Taste Differences over People," *Land Economics*, 74 (2), 230-239.
- Train, K.E., 1999, "Mixed Logit Models of Recreation Demand," in *Valuing the Environment Using Recreation Demand Models* ed. By C. Kling and J. Herriges, Edward Elgar Press.
- Train, K.E., 2001, "A comparison of hierarchical Bayes and maximum simulated likelihood for mixed logit", Working Paper, Department of Economics, University of California, Berkeley.
- Train, K.E., 2003, *Discrete Choice Methods with Simulation*, Cambridge University Press.
- Whitehead, J.C., T.C. Haab and J-C. Huang, 1998, "Part-Whole Bias in Contingent Valuation," *Southern Economic Journal*, 65, 160-168.

Yue, S., 2000, "The Gumbel Mixed Model Applied to Storm Frequency Analysis," *Water Resource Management*, 14, 377-389.

Yue, S., T.B.M.J. Quarda, B. Bobée, P. Legendre, and P., Bruneau, 1999, "The Gumbel Mixed Model for Flood Frequency Analysis," *Journal of Hydrology*, 226(1-2), 88-100.

Figure 1: Distribution Function of Gumbel Mixed Model of Maxima with $\lambda = 0.5$

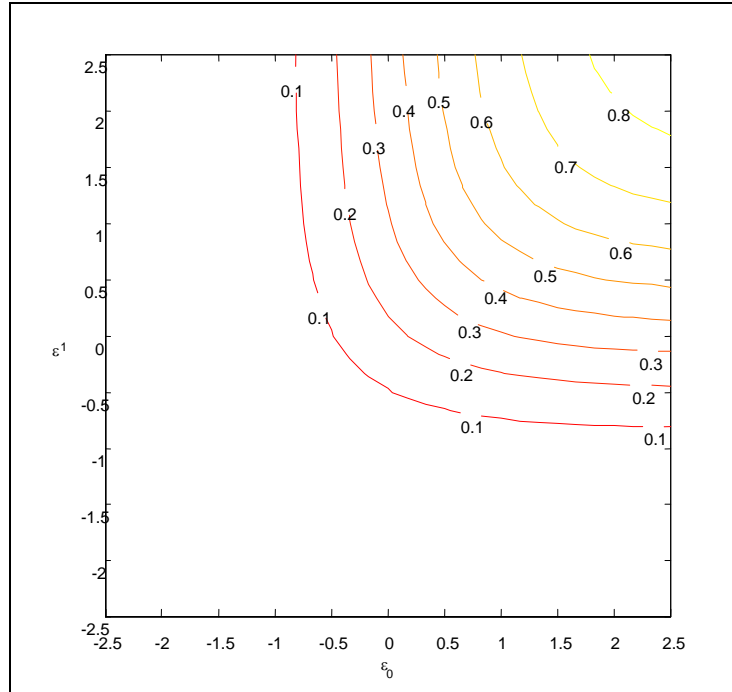


Figure 2: Probability Function of Gumbel Mixed Model of Maxima with $\lambda = 0.5$

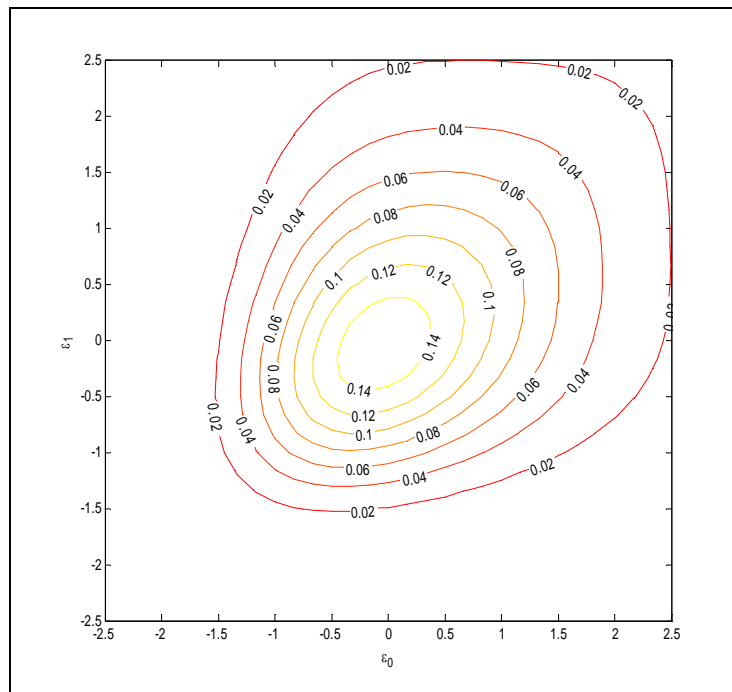


Figure 3: Distribution Function of Reduced Difference of Extreme Values

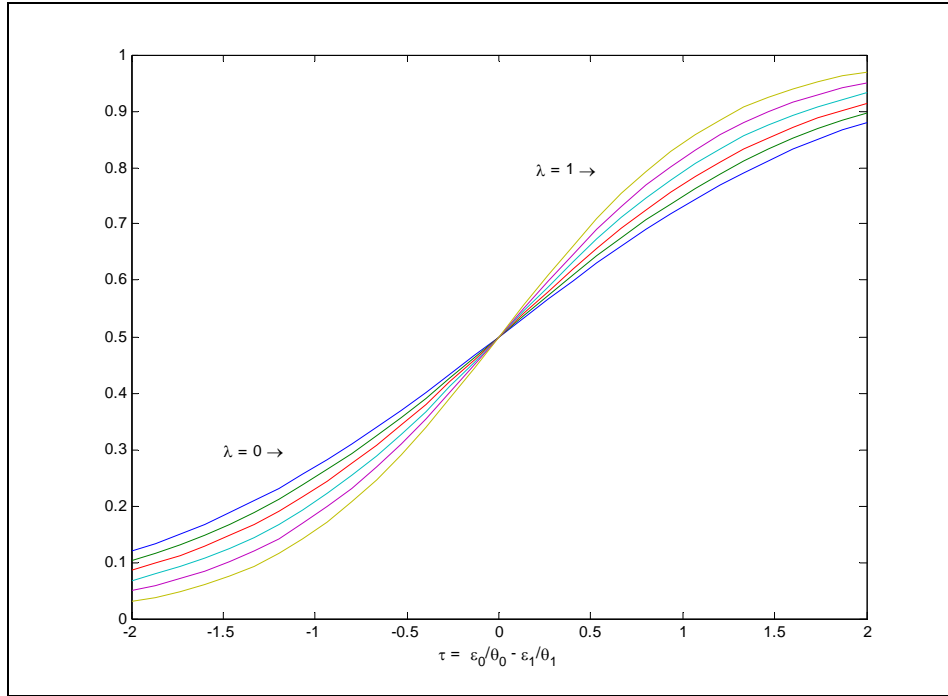


Figure 4: Probability Function of Reduced Difference of Extreme Values

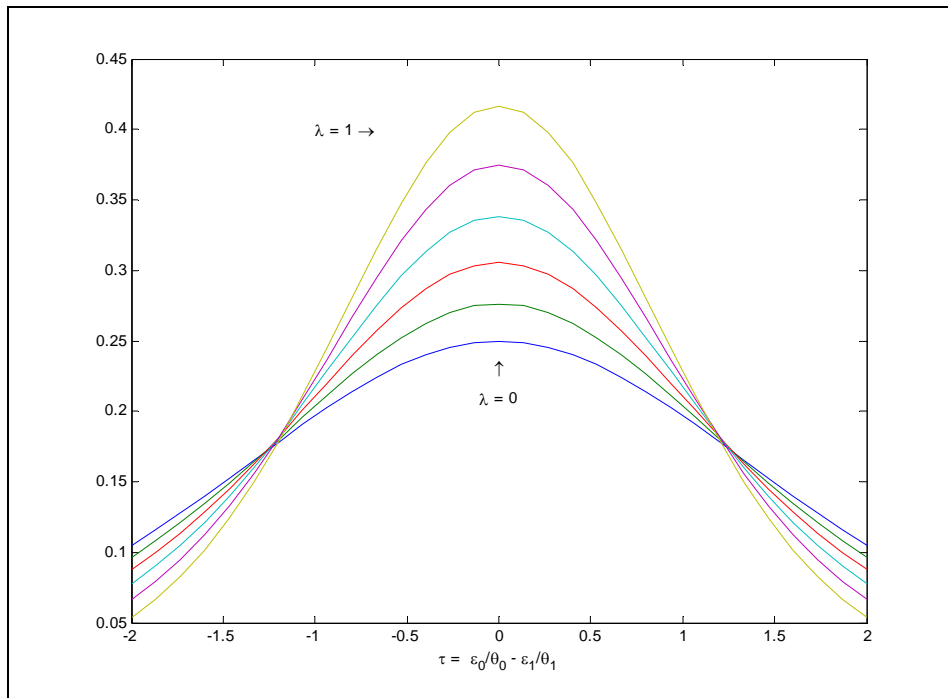


Table 1. Estimation Result of Random Utility Model with Barbados Data

	Logit	Constrained Bivariate			Bivariate Extreme Value Model	Constrained Mixed Logit			Mixed Logit Model
		$(\gamma = 1, \lambda = 0)$	$(\lambda = 0)^*$	$(\gamma = 1)$		$(\gamma = 1, \lambda = 0)$	$(\lambda = 0)^*$	$(\gamma = 1)$	
Constant	0.6789	0.6790		0.5959	0.6700	0.4718		0.5803	0.6646
	0.5261	0.5262		0.4233	0.5861	0.5207		0.4771	0.7226
Income	0.0549	0.0549		0.0412	0.0498	0.0645		0.0532	0.0812
	0.0210	0.0210		0.0178	0.0375	0.0238		0.0233	0.0703
City	0.4099	0.4100		0.2804	0.3401	0.4527		0.3432	0.5297
	0.2929	0.2929		0.2304	0.3602	0.3150		0.3137	0.6748
Age	-0.0285	-0.0285		-0.0219	-0.0272	-0.0265		-0.0250	-0.0386
	0.0090	0.0090		0.0083	0.0209	0.0098		0.0092	0.0232
Bid	0.0387	0.0387		0.0311	0.0389	0.0408		0.0368	0.0554
	0.0063	0.0063		0.0082	0.0283	0.0078		0.0068	0.0355
γ	-	1.0000		1.0000	1.2648	1.0000		1.0000	1.6949
	-	-		-	0.8285	-		-	0.9574
λ	-	0.0000		0.6099	0.5638	0.0000		0.3516	0.2623
	-	-		0.3739	0.4613	-		0.5122	0.6528
Log likelihood	-160.841	-160.841		-160.211	-160.122	-160.782		-160.678	-160.268

*Function calculation is failed.

Table 2. Estimation Result of Random Utility Model with Montevideo Data

	Logit	Constrained Bivariate			Bivariate Extreme Value Model	Constrained Mixed Logit			Mixed Logit Model
		$(\gamma = 1, \lambda = 0)$	$(\lambda = 0)$	$(\gamma = 1)$		$(\gamma = 1, \lambda = 0)$	$(\lambda = 0)$	$(\gamma = 1)$	
Constant	-0.6140	-0.6139	-0.1297	-0.4860	-0.1297	-0.6985	-0.1360	-0.6565	-0.0994
	0.1295	0.1295	0.0822	0.1444	0.0823	0.1336	0.0924	0.1617	0.0819
Income	0.0944	0.0944	0.0743	0.0778	0.0743	0.1108	0.0812	0.1074	0.0708
	0.0192	0.0192	0.0138	0.0228	0.0138	0.0183	0.0146	0.0286	0.0077
Bid	0.0100	0.0100	0.0053	0.0078	0.0053	0.0104	0.0063	0.0096	0.0063
	0.0014	0.0014	0.0008	0.0022	0.0008	0.0015	0.0009	0.0017	0.0009
γ	-	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000
	-	-	-	-	-	-	-	-	-
λ	-	0.0000	0.0000	0.5313	0.1909	0.0000	0.0000	0.2277	1.0000
	-	-	-	0.4445	0.0000	-	-	0.2478	-
Log likelihood	-726.719	-726.720	-723.611	-726.244	-723.611	-726.174	-723.625	-725.782	-723.432

Table 3. Sample Average of Welfare Measure for Environmental Quality Change

	Logit	Constrained Bivariate			Bivariate Extreme Value Model	Constrained Mixed Logit			Mixed Logit Model
		$(\gamma = 1, \lambda = 0)$	$(\lambda = 0)$	$(\gamma = 1)$		$(\gamma = 1, \lambda = 0)$	$(\lambda = 0)$	$(\gamma = 1)$	
Barbados Data									
RU ¹	-2.0701	-2.0688	.	-0.2368	1.6110	-2.1668	.	-1.1496	1.6930
ED ²	-2.0701	-2.0709	7467.3236	-0.2374	2.8939	-3.4701	3.7356	-0.0794	2.9927
Montevideo Data									
RU ¹	-26.8148	-26.8149	-81.6514	-25.7220	-81.6283	-27.9171	-65.7042	-27.2881	-65.1584
ED ²	-26.8148	-26.8189	.	-25.7257	.	-27.3092	-65.3239	-26.4295	-65.3239

¹ RU represents the random utility model.

² ED represents the expenditure difference model.

Table 4. Estimation Result of Expenditure Difference Model with Barbados Data

	Logit	Constrained Bivariate			Bivariate Extreme Value Model	Constrained Mixed Logit			Mixed Logit Model
		$(\gamma = 1, \lambda = 0)$	$(\lambda = 0)$	$(\gamma = 1)$		$(\gamma = 1, \lambda = 0)$	$(\lambda = 0)$	$(\gamma = 1)$	
Constant	0.6789	0.6789	0.3512	0.5958	0.5283	0.5658	1.1122	0.6474	0.7489
	0.5261	0.5260	0.4404	0.4233	0.4392	0.5402	1.1487	0.4580	0.7095
Income	0.0549	0.0549	0.0419	0.0412	0.0393	0.0569	0.1063	0.0494	0.0649
	0.0210	0.0210	0.0155	0.0178	0.0173	0.0216	0.0513	0.0185	0.0396
City	0.4099	0.4099	0.3047	0.2804	0.2687	0.4682	0.9152	0.3478	0.4112
	0.2929	0.2929	0.2342	0.2304	0.2228	0.2981	0.6835	0.2690	0.4621
Age	-0.0285	-0.0285	-0.0244	-0.0219	-0.0215	-0.0281	-0.0594	-0.0249	-0.0330
	0.0090	0.0090	0.0075	0.0083	0.0082	0.0088	0.0346	0.0081	0.0235
Bid	0.0387	0.0387	0.0351	0.0311	0.0307	0.0385	0.0845	0.0366	0.0510
	0.0063	0.0063	0.0055	0.0082	0.0083	0.0046	0.0390	0.0053	0.0297
γ	-	1.0000	455.6437	1.0000	1.2726	1.0000	2.4492	1.0000	1.4920
	-	-	42029.933	-	0.8559	-	1.1828	-	0.8958
λ	-	0.0000	0.0000	0.6099	0.5630	0.0000	0.0000	0.4651	0.4246
	-	-	-	0.3740	0.4635	-	-	0.3311	0.3744
Log likelihood	-160.841	-160.841	-160.171	-160.211	-160.121	-160.320	-159.267	-159.667	-158.996

Table 5. Estimation Result of Expenditure Difference Model with Montevideo Data

	Logit	Constrained Bivariate			Bivariate Extreme Value Model*	Constrained Mixed Logit			Mixed Logit Model
		($\gamma = 1, \lambda = 0$)	($\lambda = 0$)*	($\gamma = 1$)		($\gamma = 1, \lambda = 0$)	($\lambda = 0$)	($\gamma = 1$)	
Constant	-0.6140	-0.6140	-0.4860	-0.6757	-0.1988	-0.6238	-0.1988		
	0.1295	0.1295	0.1444	0.1392	0.1462	0.1312	0.1462		
Income	0.0944	0.0944	0.0778	0.1064	0.0824	0.1110	0.0824		
	0.0192	0.0192	0.0228	0.0214	0.0153	0.0222	0.0153		
Bid	0.0100	0.0100	0.0078	0.0104	0.0064	0.0082	0.0064		
	0.0014	0.0014	0.0022	0.0015	0.0009	0.0016	0.0009		
γ	-	1.0000	1.0000	1.0000	0.1009	1.0000	0.1009		
λ	-	-	-	-	0.1913	-	0.1911		
	-	0.0000	0.5313	0.0000	0.0000	0.6041	0.0000		
Log likelihood	-726.719	-726.719	-726.243	-726.269	-723.976	-724.497	-723.976		
			0.4445	-	-	0.3006	.		

*Function calculation is failed.