ESSAYS ON THE TEMPORAL INSENSITIVITY, OPTIMAL BID DESIGN AND GENERALIZED ESTIMATION MODELS IN THE CONTINGENT VALUATION STUDY

DISSERTATION

Presented in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy in

the Graduate School of The Ohio State University

By

SOO-IL KIM, M.A.

* * * * *

The Ohio State University
2004

Dissertation Committee:

Professor Timothy C. Haab, Adviser

Professor Stephen R. Cosslett

Professor Alan Randall

Approved by

____

Adviser

Graduate Program in Agricultural,
Environmental, and Development Economics

ABSTRACT

In spite of theoretical background, technical simplicity and popularity in application, contingent valuation studies have several issues remained in debating among environmental economists. This dissertation aims to provide answers to some of issues in dichotomous choice contingent valuation: the temporal structure of willingness to pay, practical guideline for survey design and generalized estimation method.

The first essay entitled "*Temporal Insensitivity of Willingness to Pay and Implied Discount Rates*" proposes the temporal willingness to pay (*TWTP*) as an alternative definition of the present value of willingness to pay. In the survey of contingent valuation, a respondent compares *TWTP* with the present value of randomly assigned cost. *TWTP* enables the test for consistency of respondent's valuation with respect to payment schemes. Using a sequential test suggested by Haab et al (1999), the insensitivity of *TWTP* is tested on the data of oyster reef restoration programs in the Chesapeake Bay. The test result shows that *TWTP* is insensitive to the offered payment schedule or on the length of the stream of benefits of the project, which implies consistent willingness to pay for the environmental project. However, discount rates estimated from the data vary significantly across project lengths and time span between offered payment schedules.

The second essay, "*Optimal, Robust kk and Uniform Experimental Designs in Binary Choice Model: Analytical and Empirical Comparison of Efficiency and Bias*" suggests a practical alternative design named a uniform design, to existing optimal or robust bid designs in contingent valuation. The uniform design draws cost assigned to respondent from a predetermined uniform distribution. Analytics and simulations show that the uniform design has lower bound of efficiency at 84 percent of D-optimum. Simulations demonstrate that the uniform design outperforms optimal designs when initial information is poor and outperforms robust designs when true values of parameters are known.

The third essay, "*Generalized Estimation Methods and Implication of the Result in Dichotomous Choice Contingent Valuation Model*" challenges the theoretical and technical background of the simple logit model. Standard logit model in contingent valuation assumes *i.i.d* error distribution between initial and proposed states. Relaxing the restrictive assumption in the simple logit model requires a generalized estimation technique that utilizes a Gumbel mixed model. Estimation results show that correlation between two states is usually minimal, but homoskedastic errors are rejected in many cases. Heteroskedasticity or correlation provides willingness to pay estimate different from estimate of the simple logit, thus different policy implication in benefit-cost analysis.

Dedicated to my parents

ACKNOWLEDGMENTS

This dissertation began several years ago when Dr. Timothy C. Haab, as my advisor, encouraged me to seek the answer about comparing design effect on the value elicitation method in environmental valuation. During the journey toward my doctorate degree, his supporting and mentoring made him more than an advisor. My committee members, Dr. Alan Randall and Dr. Stephen R. Cosslett, have also guided me on this journey and provided insightful suggestions from the very beginning of the study.

I thank anonymous attendants in AAEA and HERE seminars for their helpful comments. I wish to thank my friends and colleagues in AED Department who have provided support and suggestions during my graduate study. Lastly, I acknowledge my parents for believing in and supporting me not only for this ultimate degree but also in my whole life.

VITA

March 26, 1971…………………………….. Born – Kyoungki, Korea

February, 1997………………………………B.A. Economics, Korea University

February, 1999………………………………M.A. Economics, Korea University

2000 – present……………………………….Graduate Research Associate,
The Ohio State University

FIELDS OF STUDY

Major Field: Agricultural, Environmental and Development Economics

Minor Field: Environmental and Resource Economics

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ESSAY 1

TEMPORAL INSENSITIVITY OF WILLINGNESS TO PAY AND IMPLIED DISCOUNT RATES

ABSTRACT

Two interrelated anomalies associated with eliciting willingness to pay for environmental change over time have been reported: insensitivity of willingness to pay to payment schedules and variation in discount rates over time. This essay proposes an alternative definition of the temporal insensitivity with respect to the temporal willingness to pay (*TWTP*) rather than the present value of willingness to pay (*PVWTP*). Insensitivity of *TWTP* implies that subject in the survey responds consistently to value elicitation questions regardless of payment schedule. Using a sequential test provided by Haab et al (1999), the insensitivity of *TWTP* is tested on the data of oyster reef restoration programs in the Chesapeake Bay. The test result shows that for this case, *TWTP* is insensitive to the offered payment schedule or to the length of the stream of benefits of the project, which provides consistent willingness to pay for the environmental project. However, discount rates estimated from the response vary significantly across project lengths and time span between offered payment schedules.

1.1 INTRODUCTION

Two interrelated anomalies associated with elicited willingness to pay for public goods over time have been noted: insensitivity of willingness to pay (*WTP*) to payment schedules and variation in discount rates over time. Insensitivity of willingness to pay to payment schedules is also known as temporal embedding. Temporal embedding effect has been argued to depend on situation and commodity specifics (Crocker and Shogren 1993), money specifics (Thaler 1981), or respondents specifics (Stevens et al. 1997)[1].

Stevens et al. (1997) define two types of temporal embedding effects: strong insensitivity and weak insensitivity to payment schedule. Strong insensitivity to payment schedule indicates the inability of respondents to differentiate between a series of payments and a lump sum payment on the project. Let WTP_L be the lump sum *WTP* for a project, and WTP_t be the annual payment of *t*-th year in an annual payment scheme, then strong insensitivity is defined as $WTP_L = WTP_1 = WTP_2 = \cdots = WTP_T$, where *T* is the terminal period of the temporal payment scheme. Alternatively, weak insensitivity implies the inequality of individual *WTP* between two temporally differentiated payment schemes but with abnormally high implied discount rates[2]. Following the above notation, weak insensitivity is defined as $WTP_L \neq WTP_t$ and $WTP_L = \sum_{t=1}^{T} WTP_t \left(1+r\right)^{-(t-1)}$, where *r*

[1] In addition to the temporal embedding effect, scope and scale embedding effects have been reported in the contingent valuation studies. Moral satisfaction (Kahneman and Knetsch 1992; Diamond and Hausman 1994), symbolic bias (Mitchell and Carson 1989), or design and analysis product (Smith 1992; Hanemann 1994) are known to be responsible for scope and scale embedding effects.

[2] In fact, "high" discount rate is still questionable in the sense how high is high. For example, the discount rate in the market varies from 1% for savings accounts to over 30% for some types of credit card debt. In some developing or under developed countries, more than 100% discount rates have been reported.

is the discount rate implied in the equality of lump sum *WTP* and the discounted sum of annual *WTP*. Note that strong insensitivity algebraically implies an infinite discount rate.

Kahneman and Knetsch (1992) find evidence of strong insensitivity of median *WTP* wherein respondents showed the same median WTP_L and WTP_t of five-year payment for funding a toxic waste treatment facility. Strong insensitivity may represent inconsistency in respondents' behavior or misunderstanding the survey questions. On the other hand, a series of papers (Rowe et al. 1992, Stevens et al. 1997, Ibáñez and McConnell 2001, Bond et al. 2002) find weak insensitivity of *WTP* with discount rates ranging from two digits to several thousand percent[3]. For example, Ibáñez and McConnell (2001) estimated *WTP* for reduction in pathogen discharge in Columbia using an intertemporal random utility model with assumption of constant discount rate. In the survey, either a lump sum payment or three monthly installments were randomly assigned to respondents. The estimation results showed a wide range of mean *WTP* and the discount rate was as high as 5,102%. Bond et al. (2002) also argued that the implicit discount rates were high relative to the market discount rate and the explicit discount rates were generally insignificant in the study of a federal Steller sea lion recovery program in Alaska using three temporal treatments of one, five, and fifteen years. Generally, strong insensitivity has been rejected in empirical tests but weak insensitivity has been widely observed.

[3] Relatively high implicit discount rates have been reported in experimental research as well. Harrison and Johnson (2002) and Harrison et al. (2002) report 28.1 percent individual discount rates in average over all subjects in a field experiment in Denmark. Coller, Harrison and Rutström (2002) provide a similar experimental result.

Previous studies, however, have imposed strong assumptions on the underlying decision process of a subject valuing a proposed environmental project, which should be tested before eliciting *WTP* and deriving discount rates from dichotomous choice CV studies with a temporal dimension. Specifically, the typical temporal CV study assumes that the present value of willingness to pay (*PVWTP*) is constant across all offered payment schemes (i.e. consistent *PVWTP*) and that the variance of the conditional distribution of *PVWTP* is invariant to the payment schedule (i.e. homoskedasticity of *PVWTP* function)[4]. If the distribution is homoskedastic and the mean of *PVWTP* is consistent, the simply pooled data enables the researcher to estimate the implied discount rate directly from parameter estimates of payment scheme variables by taking the ratio of them. However, if *PVWTP* is not consistent, then we cannot compare two different present values, and if the distribution of conditional *PVWTP* is heteroskedastic across different payment schedules, the variance is unidentified because the parameter is the product of the heteroskedastic variance and discount factor. Identification and estimation of the discount rate by varying the payment scheme relies critically on the assumption of a consistent and homoskedastic *PVWTP* independent of payment context.

To provide context for the following methodological development, section 1.2 briefly describe the application utilizing a unique mail survey about a proposed oyster reef restoration program encompassing several states around Chesapeake Bay. Section 1.3 defines an alternative temporal-dimensioned valuation process of environmental

[4] Haab et al. (1999) test the consistency of *WTP* under real and hypothetical formats, reporting that if the heteroskedasticity due to different question format is corrected, the estimated *WTP* is consistent across different question formats. On the other hand, Huhtala (2000) investigates the heterogeneous preference in the contingent valuation method by distinguishing preferences according to the respondent's attitude on environmental policy. Heterogeneity in preference explains the inconsistent *WTP*

project, named as temporal willingness to pay (*TWTP*), to alleviate restrictive assumptions about the *PVWTP*. Based on *TWTP*, the temporal insensitivity of *WTP* to payment scheme implies that respondent consistently evaluates the environmental project regardless of payment scheme. In spite of different definition of willingness to pay, value of cost stream follows the typical definition of present value, which simplifies the derivation of implicit discount rate from value elicitation response. Section 1.4 explains the present value of cost and implied discount rates. Section 1.5 describes the basic estimation model and reports initial estimation result. Section 1.6 explains the test procedure for the consistency and homoskedasticity of *TWTP* following the sequential method proposed by Haab et al. (1999). According to the test and estimation result, *TWTP* of a value elicitation survey on oyster reef restoration programs in the Chesapeake Bay does not depend on the payment scheme (consistent *TWTP* across the payment scheme) or the benefit stream offered in the survey (indifferent *TWTP* to the build-up phase). Based on the homoskedasticity test of the conditional distribution, estimation results show that implied discount rates vary significantly across the length of project life and time span between offered payment schedules.

1.2 THE OYSTER REEF RESTORATION PROGRAM IN THE CHESAPEAKE BAY

The tall reefs in the Chesapeake Bay are the main habitat for Bay oysters, the most harvested seafood species in the Chesapeake Bay. Due to intense harvest over more than one hundred years, however, very few reefs remain in the Bay and the Chesapeake's oyster population has fallen to less than one percent of their historic maximum levels. In 2002, as part of the Marine Recreational Fisheries Statistics Survey, the National Marine Fisheries Service conducted a random digit dial (RDD) telephone survey over several states around Chesapeake Bay to assess respondents' attitudes toward oysters and oyster reef restoration in the Chesapeake Bay. Among 8,077 people contacted in the RDD survey, a follow-up mail survey sponsored by the Chesapeake Bay Foundation was sent to 1,710 respondents who agreed to participate in a follow-up mail survey. The mail survey provided a brief explanation of the role and benefits of oysters in the Bay, and then asked referendum question about willingness to pay as well as attitude and preference towards the Chesapeake Bay, general knowledge of oyster reefs and socio-demographic questions (for details, refer to the appendix A) [5].

The follow-up mail survey consisted of two temporal versions of the hypothetical oyster reef restoration project which were randomly assigned to respondents (A for five year and B for ten year). Both projects aim to restore 10,000 acres of oyster habitat and 1,000 acres of artificial reef until the terminal period of the project at a constant rate. Thus, the ten-year (five-year) restoration program accumulates at a rate of 100 (200) acres of reef restoration and 1,000 (2,000) acres of habitat preservation per year. For each restoration program, one of three temporal payment schemes was randomly offered to respondents: a one-time (lump sum) payment on the next year's state tax return (payment schedule 1), an annual payment on state tax returns over the life of the project (payment

[5] The referendum type question was adapted for eliciting the value of the oyster reef restoration instead of open-ended question because data from open-ended question has incentive compatible problem against other advantages.

schedule 2) and a permanent annual payment on the state tax return (payment schedule 3). The final survey consisted of a 2x3 design (2 project lengths and 3 payment schemes). Figure 1.1 shows the whole structure of the survey design and Appendix A provides the actual questionnaire in the survey.

A hypothetical referendum question for the randomly assigned restoration project and payment scheme was asked of respondents. For example, the question for the five-year project scenario with one-time payment reads as follows;

> The restoration program is estimated to cost your household a total of **$___**. Your household would pay this as a special one time tax added to next year's State income tax. **If an election were to be held today and the total cost to your household was $___ would you vote for or against the 5 year restoration program** (Check one)?
>
> - ❑ I would vote for the program
> - ❑ I would vote against the program
> - ❑ I do not know whether I would vote for or against the program

For annual and perpetuity payment types, the questions were worded appropriately for the value of cost and payment type. Each design of the payment type had a set of three bid points, one of which was randomly assigned to each respondent. The payment values used in one-time payment were selected from the set of {50, 150, 300}. A 25% discount rate was applied to calculate annual value of cost in both annual and perpetuity payment types as in Figure 1.1. If there is no starting point bias or anchoring effect, the discount rate used in the survey design does not affect the decision mechanism by the property of dichotomous choice contingent valuation model[6]. Survey result of response and some of demographic variables are provided in the section 1.3.4.

---

[6] Since the discount rate changes the actual value of cost per year thus the range of bid set, different set of payment affects the bias and efficiency of parameter estimates (See Essay 2).

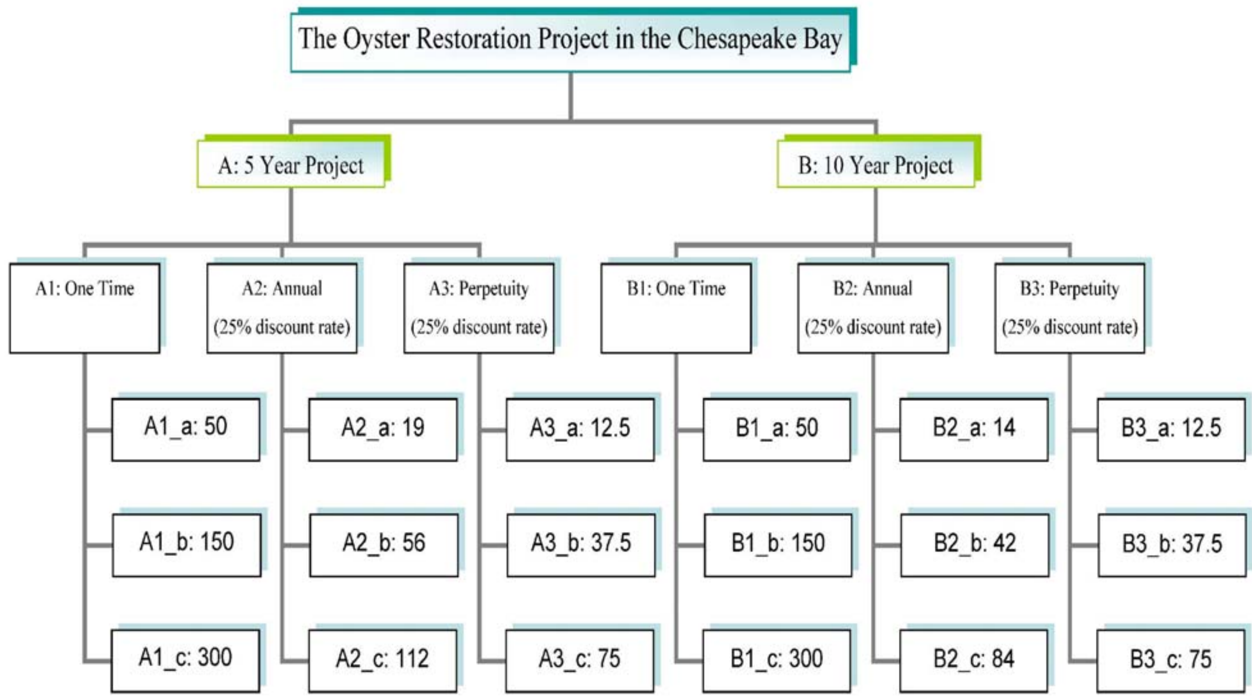

Figure1.1: The Structure of the Experimental Design for the Oyster Restoration Program in the Chesapeake Bay

1.3 VALUES OF BENEFIT STREAM AND COST STREAM

1.3.1 Present Value of Willingness to Pay vs. Temporal Willingness to Pay

Environmental projects, by their nature, include temporal dimension of benefits and costs which may or may not be considered by researcher. When a respondent is asked the valuation question in the form of dichotomous choice, he or she will compare the benefit stream from the project with the cost stream required for the project. Suppose that a proposed project consists of a stream of annual benefits B_t, $t = 1, 2, \ldots, T_B$ and an associated stream of annual costs, C_t, $t = 1, 2, \ldots, T_C$, where T_B represents the life of the benefits of the project and T_C is the life of the costs. The project is fully described by the benefit-cost pairing (B_t, C_t). The benefit is subjective due to respondent's experience and reliability of the program, uncertainty of the future, different cognizance about the benefit stream, etc. The cost is of monetary value and implicit because respondents have their own discount factor consisting of interest rate, uncertainty, etc. The basic assumption is that respondents have a well defined value of B_t with which they can compare the monetary value of costs[7]. The binary response to the value elicitation question will be one if the *WTP* for the benefit stream of the project exceeds the value of cost stream and zero otherwise.

In the previous literature, *PVWTP* has represented the value of the benefit stream based on the time separable annual *WTP*. The assumption of time-separability, however,

[7] A well-defined range of value is enough for comparison. Moreover, the value of benefit does not have to be a monetary unit. The decision, then, will be made by comparing the benefit and cost in terms of the same but any plausible unit from a respondent. For convenience and simplicity, respondents are assumed to have well defined monetary value of benefit stream since we only observe the monetary value of cost.

requires more serious assumptions to simplify the model, the problem which has not received enough attention from researchers in contingent valuation studies. To see the problem in more detail, suppose that respondent *i* has a stream of *WTP* in each period as a function of the benefit and individual specific covariates ($x_{i,t}$); $WTP_{i,t} = x'_{i,t}\beta_t + \varepsilon_{i,t}$, where $\varepsilon_{i,t}$ is an additive error term of respondent *i* at period *t* which is unknown to researcher. Then, *PVWTP* can be expressed as the discounted sum of *WTP*;

$$PVWTP_i = \sum_t \frac{x'_{i,t}\beta_t}{\left(1+r\right)^{t-1}} + \sum_t \frac{\varepsilon_{i,t}}{\left(1+r\right)^{t-1}} .$$

Two temporal structures are detected in this formulation. First, *WTP* at each period varies depending on the perceived benefit stream and individual specific covariates that may or may not vary over time. Representing the individual specific variables by the current value generates more uncertainty in the error term since they are not realized at the moment of survey. Second, the temporal structure in the error term has classical issues in time series data. Even if the respondents are assumed to have constant covariates, such a formulation requires strong assumptions about the temporal relation of error terms. Furthermore, unless the life of benefit from the project (T_B) is constant across respondent, the estimation model does not have *i.i.d.* error distribution[8].

The complexity of the temporal structure of *PVWTP* motivates an alternative definition of the valuation process such that respondents have a value for the whole benefit stream (i.e. *WTP* for the full benefit stream) rather than a value stream for each

---

[8] T_B can be given in the survey explicitly by the researcher. However, a respondent perceives implicitly the terminal period of the benefit stream, which can be seriously different across individuals. *PVWTP* has different time-span of discounting for each respondent.

period (i.e. *WTP* of each period). Let π be the value of the benefit stream that respondent may have at the moment of survey. The value of the benefit stream, named as a temporal willingness to pay (*TWTP*), is a function of the benefit stream and current individual specific covariates,

$$TWTP_i = f\left(\pi, x_i, \beta\right) + \varepsilon_i \tag{1.1}$$

where $f\left(\cdot\right)$ is a systematic component observed by researcher and ε_i is an unobserved random error with mean zero. The error term can be conditional on the project type and payment schemes but is invariant across individuals. *TWTP* is time-dependent in the sense that it could be different in the different moment of the survey due to information, uncertainty, time-varying covariates, etc[9]. The error term in the *TWTP* function, however, is independent of time since the structure of *TWTP* is static at the moment of survey.

In summary, since *TWTP* is a lump-sum value that an individual may have for the environmental project at decision moment, *TWTP* model does not require researcher to sum the discounted errors across time or impose restrictions on the temporal relation of multi-period error terms. Individual specific covariates are not discounted but affect the implied discount rate through the estimation. *TWTP* provides a reasonable and realistic valuation structure about how an individual thinks of the environmental project proposed in the survey, without assuming the time-separable *WTP* stream at each period.

The benefit stream of the oyster reef restoration program includes explicitly a multi-period build-up phase for the five- and ten-year versions and constant benefit

[9] However, Carson et al. (1997) showed that CV estimates exhibited no significant sensitivity to the timing of interviews.

stream with the assumption of no degradation after the completion of the project. Let the terminal period of the project be L and the accumulation rate of the benefit be b over the life of the project, then the benefit stream is $B_t = bt$ if $t \le L$ and $B_t = bL$ after L. Note that L is not the life of benefit but the terminal period of project. The expected benefit at the choice moment is a discounted sum of benefit stream with the implicit discount rate of benefit such that

$$\pi(B, L, r_b) \equiv \sum_{t=1}^{T_B} \frac{B_t}{\left(1+r_b\right)^{t-1}}$$

where r_b is a discount rate for the benefit stream and B_t is a measure of the benefits in period t. For simplicity, assume an infinite benefit stream ($T_B = \infty$). Summing over the infinite time horizon, the discounted benefit stream becomes

$$\pi(B, L, r_b) = b\sum_{t=1}^{L} \frac{t}{\left(1+r_b\right)^{t-1}} + \tilde{B} \sum_{t=L+1}^{\infty} \frac{1}{\left(1+r_b\right)^{t-1}} = b\gamma_1 + \tilde{B}\gamma_L \tag{1.2}$$

where $\gamma_1 = \sum_{t=1}^{L} \frac{t}{\left(1+r_b\right)^{t-1}} = \left(\frac{(1+r_b)^L - 1}{r_b^2 (1+r_b)^{L-2}} - \frac{L}{r_b\left(1+r_b\right)^{L-1}} \right)$, $\gamma_L = \sum_{t=L+1}^{\infty} \frac{1}{\left(1+r_b\right)^{t-1}} = \frac{1}{r_b (1+r_b)^{L-2}}$

and $\widetilde{B} = b \cdot L$.

With an alternative definition of the willingness to pay for the benefit from environmental change, the insensitivity of *WTP* is defined such that *TWTP* does not change due to the payment schedules. Insensitivity of *TWTP* to payment schedule implies the consistency of respondent's valuation of environmental project. Therefore, the assumption in the existing literature that *PVWTP* is same across payment schemes can be tested by comparing *TWTP* in different payment schedules. The classical definition of (strong or weak) insensitivity of *WTP* depends on the consistency test explained in the section 1.4.

1.3.2 Present Value of Cost and Implicit Discount Rates

The value of the cost stream is typically derived in terms of the present value. Since the annual amount of cost is constant over payment schedule ($C_{t,j} = C_j$), the general form of present value of cost (*PVC*) is

$$PVC = \sum_{t=1}^{T_C} \frac{C_j}{\left(1+r_C\right)^{t-1}} = C_j \beta_C^j$$

where *j* represents the payment schedule, T_C is the terminal period of the cost and r_C is discount rate of the cost stream. The discount rate or discount factor of the cost stream can be different from the discount rate of benefit stream since it depends on different factor such as market discount rate, belief on the financial market, uncertainty, etc. *PVC* has parameter values specific to the payment schedule. For a lump sum payment scheme with C_1 in period 1, *PVC* becomes

$$PVC_1 = C_1. \tag{1.3}$$

implying that $\beta_C^1 = 1$. *PVC* of annual payments with C_2 over T_C years is

$$PVC_2 = C_2 \left(\frac{1+r_C}{r_C} \right) \left(1 - \frac{1}{\left(1+r_C\right)^{T_C}} \right). \tag{1.4}$$

and $\beta_C^2 = \left(1+r_C\right)\left(1-\left(1+r_C\right)^{-T_C}\right)/r_C$. Finally, *PVC* of perpetual payment of C_3 is

$$PVC_3 = C_3 \left(\frac{1+r_C}{r_C} \right). \tag{1.5}$$

with $\beta_C^3 = (1+r_C)/r_C$ when the discount rate is positive. When the coefficients of (1.3), (1.4) and (1.5) are identified, then the discount rate of cost stream is estimated by the ratio of any pair of coefficient estimates. Abnormally high discount rate implies the conventional weak insensitivity of *WTP*.

1.4 ESTIMATION MODELS

1.4.1 Linear *TWTP* Model

Unlike *PVWTP*, *TWTP* is flexible in the functional form of systematic component[10]. The simplest case is assuming a linear function of $f(\cdot)$ in equation (1.1) such as

$$TWTP_i = \beta_B \pi(B, L, r_b) + x_i'\beta + \varepsilon_i .$$

With a linear *TWTP* function and a normal error distribution with mean zero, the probability that a respondent *i* will vote for a program *k* given the payment version *j* is

$$\begin{aligned} P(i \text{ vote for } k \mid j) &= P\left(TWTP_{i,k} \geq PVC_{i,j}\right) \\ &= P\left(\beta_B \pi_k + x_{i,k}'\beta_k + \varepsilon_{i,k,j} \geq C_j \beta_C^j\right) \\ &= \Phi\left(\tilde{\beta}_B \pi_k + x_{i,k}'\tilde{\beta}_k - C_j \tilde{\beta}_C^j\right) \end{aligned} \tag{1.6}$$

where $\tilde{\beta}_B$ and $\tilde{\beta}_k$ are parameters normalized by the standard error, $\sigma_{k,j}$, and $\Phi(\cdot)$ is the standard normal cumulative distribution function. Equation (1.6) is the standard probability of a vote for the project in a probit referendum model and the probability of

[10] Due to the discounted summation, *PVWTP* model cannot be estimated except the case of linear function.

vote against is defined as the complement to the probability of vote for.[11] Generally, the variance of the error term is conditional on the project version (*k*) and payment scheme (*j*). Note that the probability in (1.6) cannot identify $\tilde{\beta}_B$ because the variable of benefit stream is invariant to project length and payment scheme.

To identify the coefficient of benefit streams, it is necessary to pool data across project versions and use a dummy variable to capture the difference of benefit stream across the different projects. Using a dummy variable for project version (d_k), the probability conditional on payment schedule is

$$P\left(i \text{ vote for} \mid j\right) = \Phi\left(\sum_{k=1}^{K} d_k \tilde{\beta}_k + x_i'\tilde{\beta} - C_j \tilde{\beta}_C^j\right). \tag{1.7}$$

The estimation model assumes that *TWTP* varies in the mean across the project version but the error term may have different variance across payment schemes. Pooling data over payment schedules provide conditional probabilities that vary over project and payment scheme:

$$P\left(i \text{ vote for} | k\right) = \Phi\left(\tilde{\beta}_B \pi_k + x_i'\tilde{\beta} - \sum_{j=1}^{J} d_j C_j \tilde{\beta}_C^j\right). \tag{1.8}$$

where d_j is a dummy indicator for payment type *j*. Parameter $\tilde{\beta}_B$ is not identified in equation (1.6) but $\tilde{\beta}_C^j$ provides information about the discount rate of cost stream. If the model is homoskedastic across project and payment scheme and if *TWTP* is different only in the mean across the project version, then the data can be pooled over all project versions and payment types, and probability function of equation (1.6) is simplified to be

[11] See Haab and McConnell (2002) for details.

$$P(i \text{ vote for}) = \Phi\left(\sum_{k=1}^{K} d_k \tilde{\beta}_k + x_i' \tilde{\beta} - \sum_{j=1}^{J} d_j C_j \tilde{\beta}_C^j \right) \tag{1.9}$$

The consistency of *TWTP* in (1.7), (1.8) and (1.9) and constant variance in (1.8) and (1.9) are strong assumptions that will be subsequently relaxed and tested.

If the survey design includes more than three different benefit streams with the same final level, the discount rate of benefit stream can be estimated from the ratio of parameter estimates of the dummies for the project. Some primary conditions are; the data should be homoskedastic across different project version, the difference of benefit stream among projects is measurable and the discount rate of benefit stream is constant. Let t_1 and t_2 be terminal periods of two projects. The ratio of two parameter estimates for dummies of project version provides the information of the discount rate of benefit stream such as

$$\begin{aligned} \frac{\tilde{\beta}_{t_1}}{\tilde{\beta}_{t_2}} = \frac{\gamma_{t_1}}{\gamma_{t_2}} &= \left(\frac{(1+r_b)^{t_1} - 1}{r_b^2 (1+r_b)^{t_1-2}} - \frac{t_1}{r_b \left(1+r_b\right)^{t_1-1}} \right) \cdot \left(\frac{(1+r_b)^{t_2} - 1}{r_b^2 (1+r_b)^{t_2-2}} - \frac{t_2}{r_b \left(1+r_b\right)^{t_2-1}} \right)^{-1} \\ &= \left(\frac{\left(1+r_b\right)^{t_1+1} - \left(1+\left(t_1+1\right) r_b\right)}{\left(1+r_b\right)^{t_2+1} - \left(1+\left(t_2+1\right) r_b\right)} \right) \left(1+r_b\right)^{t_2-t_1} \end{aligned} . \tag{1.10}$$

The discount rate of benefit stream is solution of nonlinear equation (1.10). Unfortunately, the oyster reef restoration data does not have enough project versions to apply the equation (1.10) for estimation of discount rate of benefit stream.

1.4.2 Exponential *TWTP* Model

An exponential *WTP* has been widely used for modeling positive *WTP*. Assuming a log normal error distribution, the functional expression for exponential *TWTP* is $TWTP_{i,k} = \exp\left(\beta_B \pi_k + x'_{i,k}\beta_k + \varepsilon_{i,k,j}\right)$, where $\varepsilon_{i,k,j}$ is a normal error distribution with mean zero and variance $\sigma^2_{k,j}$. By taking the natural log on both of *TWTP* and *PVC*, the binary response variable is one if $\beta_B \pi_k + x'_{i,k}\beta_k + \varepsilon_{i,k,j} \geq \ln\left(PVC_j\right)$, and zero otherwise. The probability of vote for in the exponential *TWTP* is

$$P\left(i \text{ vote for } k \mid j\right) = P\left(\varepsilon_{i,k,j} \geq \ln\left(C_j \beta_C^j\right) - x'_{i,k}\beta_k\right) = \Phi\left(x'_{i,k}\tilde{\beta}_k - \frac{1}{\sigma_{kj}}\left[\ln\left(C_j\right) + \ln\left(\beta_C^j\right)\right]\right).$$

The parameter estimate of $\ln C_j$ is the inverse of standard deviation of error term as in the conventional probit model. Since β_C^j is invariant to project length and payment scheme, the split sample model cannot identify the discount factor of cost stream. Thus, except *TWTP* of one-time payment scheme for which the discount factor β_C^j is one, *TWTP* are not identified across different payment schedules. Nonidentification problem arises in the conditional probability corresponding to (1.7), which is expressed as

$$P\left(i \text{ vote for} \mid j\right) = \Phi\left(\sum_{k=1}^{K} d_k \tilde{\beta}_k + x'_i \tilde{\beta} - \frac{1}{\sigma_j}\left[\ln\left(C_j\right) + \ln\left(\beta_C^j\right)\right]\right)$$

where σ_j is assumed to be homoskedastic across different project version.

With the assumption of consistent *TWTP*, a pooled model can identify the discount factor of the cost stream. Consider the exponential *TWTP* substituted into the equation (1.9),

$$P(i \text{ vote for}) = \Phi\left(\sum_{k=1}^{K} d_k \tilde{\beta}_k + x_i'\tilde{\beta} - \frac{1}{\sigma_j}\sum_{j=1}^{J} d_j \left[\ln(C_j) + \ln(\beta_C^j)\right]\right). \quad (1.11)$$

Since $\ln \beta_C^{j=1} = 0$, it is natural to use dummies for j = 2 and 3 in the model by dropping $d_{j=1}$. The pooled model can identify the discount factor without further assumption of σ_j. If the error term is homoskedastic across payment scheme, then the equation (1.11) is simplified to

$$P(i \text{ vote for}) = \Phi\left(\sum_{k=1}^{K} d_k \tilde{\beta}_k + x_i'\tilde{\beta} - \frac{1}{\sigma}\ln(C) - \frac{1}{\sigma}\sum_{j=1}^{J} d_j \ln(\beta_C^j)\right).$$

Again, without $\beta_C^{j=1} = 1$, i.e. lump-sum payment schedule, the model cannot estimate *TWTP*.

1.4.3 Initial Estimation

Table 1.1 reports the response in each scenario and each cost amount shown in Figure 1.1. A and B indicate the project versions of 5- and 10-year and 1, 2 and 3 represent payment schemes of one-time, annual payment and perpetuity, respectively. Among 1,710 who participated in the mail survey, 577 respondents replied to the survey questionnaire resulting in a 33.7 percent response rate. After dropping incomplete responses, 519 observations were used for the estimation and tests of insensitivity. For a conservative estimate of *TWTP*, the 'I don't know' response is assumed to be 'vote against' response (Carson et al. 1998, Groothuis and Whitehead 1998).

As can be seen in Table 1.1, the proportion of respondents voting for the project violates the monotonicity of probability distribution in two cases: from A1a to A1b (case

1) and from B3b to B3c (case 2). These problematic features of data may distort estimation result and temporal insensitivity test because they arise in the potential tail of the distribution. In nonparametric estimation, the pooled adjacent violators algorithm (PAVA) has been suggested by Kriström (1990) and Haab and McConnell (1997) to provide a self-consistent bounded estimator such as Turnbull estimator for the inconsistent data. In parametric estimation, however, the violation of monotonic probability distribution may distort estimation result and temporal sensitivity test because the violation arises in the potential tail of the distribution[12].

Table 1.2 shows the summary of demographic variables of respondents. *SEX* is a dummy variable that is one for female. *HS*, *AGE* and *EDUC* are the size of household, age and education variables. *RE* is an ordinal variable for ranking the role of oysters among food, economy, environment and fish habitat. *RE* = 1 represents that respondent thinks environment is the most important role oysters play in the Chesapeake Bay and *RE* = 4 shows that environment is the least important.

The estimation model utilizes the linear function of *TWTP* such that, for instance, the conditional probability on the payment schedule (equation 1.7) is

$$\begin{aligned}&P(\text{i votes for } | j)\\&=\Phi\left(\tilde{\beta}_{j1}+\tilde{\beta}_{j2}FIVE+\tilde{\beta}_{j3}RE+\tilde{\beta}_{j4}HS+\tilde{\beta}_{j5}SEX+\tilde{\beta}_{j6}AGE+\tilde{\beta}_{j7}EDUC-\tilde{\beta}_{C}^{j}C_{j}\right)\end{aligned}$$

where *FIVE* is a dummy indicator that equals one if individual *i* receives the five year restoration plan and zero otherwise. The model was estimated using Gauss 5.0.

---

[12] Since the estimate of mean *WTP* is sensitive to the design of payment set (Cooper and Loomis 1992, Kanninen, 1995, Roach et al. 2002), Kanninen (1995) recommends to avoid obviously excessive payment amount. Illogical response to the excessive payment generates seriously biased estimate of *WTP*.

Table 1.3 shows the estimation results of split and pooled data corresponding to the estimation model of equations (1.6) and (1.7). The first six columns show the estimates of split sample for each project and payment type with potential heteroskedasticity and different *TWTP* in each scenario. The last three columns are estimation results of pooled data with assumption that *TWTP* between five and ten-year projects is different in the mean of *TWTP*. FEE1, FEE2A, FEE2B and FEE3 represent payment vectors for one-time, annual payment for five years, annual payment for ten years and perpetuity-type payment, respectively. At the bottom of the table is reported the mean of log likelihood value of each estimation. As explained before, due to the violation of the monotonicity of probability function in the response of B3 category, parameter estimate of FEE3 has a negative sign. Note that the parameter estimates of *FIVE*, a dummy for project A, are insignificant across all payment types in the last three columns of Table 1.3. Statistically, willingness to pay for ten-year project are same to that of five-year project which has faster provision of the identical final target quality.

Table 1.4 and 1.5 report the estimation results of equation (1.8) pooling all possible combination of payment types in five and ten-year project, respectively. Table 1.6 shows the estimation result of same combination for the data pooled over two project versions, i.e. equation (1.9). Note that, in Table 1.5 and 1.6, estimates of FEE3 are smaller than that of FEE2B. The scaled model and estimation result of the model are explained in the next section.

		A									B								
		1			2			3			1			2			3		
		a	b	c	a	b	c	a	b	c	a	b	c	a	B	c	a	b	c
Vote for	333 (64.2)	17 (60.7)	33 (73.3)	12 (42.9)	21 (77.8)	22 (64.7)	13 (59.1)	16 (84.2)	25 (69.4)	9 (50)	21 (77.8)	26 (53.1)	7 (28)	19 (86.4)	25 (64.1)	12 (57.1)	15 (71.4)	25 (67.6)	15 (71.4)
Response	519	28	45	28	27	34	22	19	36	18	27	49	25	22	39	21	21	37	21
Total	1710	103	137	103	78	101	77	77	102	77	102	138	104	77	102	77	77	102	76

Table 1.1: Responses (Rates) to the Referendum Question in Each Category

	Average	Std. Dev.	Minimum	Maximum
Size of Household	2.7553	1.3768	1	12
Age	49.7919	14.4101	15	90
Education	14.9075	2.6797	8	20
	Female	**Male**		
Sex	273 (52.6)	246 (47.4)		
	1	**2**	**3**	**4**
Ranking of Environment	399 (76.9)	65 (12.5)	36 (6.9)	19 (3.7)

Parenthesis reports percentage.

Table 1.2: Demographic Variables of the Chesapeake Bay Study

	A			B			AB		
	A1	A2	A3	B1	B2	B3	AB1	AB2	AB3
Const	-0.8726 (1.0970)	2.1975 (1.2577)	0.9891 (1.4054)	-1.3436 (1.2341)	0.3837 (1.3487)	1.0987 (1.2841)	-1.0353 (0.7997)	1.3334 (0.9163)	1.1250 (0.8959)
FIVE	—	—	—	—	—	—	0.2757 (0.1936)	-0.0821 (0.4350)	-0.0198 (0.2183)
RE	-0.5018* (0.2050)	-0.1448 (0.1674)	-0.3698 (0.2059)	-0.0978 (0.1819)	-0.0784 (0.3007)	-0.0160 (0.1951)	-0.2667* (0.1327)	-0.1177 (0.1454)	-0.1781 (0.1320)
HS	0.2077 (0.1367)	-0.0918 (0.1477)	0.0296 (0.0864)	-0.0664 (0.1057)	-0.1031 (0.1330)	-0.1551 (0.1076)	0.0421 (0.0812)	-0.1169 (0.0977)	-0.0620 (0.0678)
SEX	0.1370 (0.2842)	0.5178 (0.3229)	-0.0473 (0.3483)	-0.0341 (0.2914)	0.0503 (0.3258)	-0.7667* (0.3308)	0.0567 (0.1992)	0.1954 (0.2202)	-0.3929 (0.2289)
AGE	0.0429* (0.0121)	0.0074 (0.0130)	-0.0053 (0.0105)	0.0274* (0.0109)	0.0008 (0.0125)	-0.0076 (0.0119)	0.0316* (0.0077)	0.0004 (0.0086)	-0.0057 (0.0077)
EDUC	-0.0226 (0.0504)	-0.1074 (0.0624)	0.0554 (0.0673)	0.0973 (0.0610)	0.0566 (0.0549)	0.0485 (0.0576)	0.0351 (0.0374)	-0.0068 (0.0393)	0.0426 (0.0419)
FEE1	0.0033* (0.0015)	—	—	0.0068* (0.0017)	—	—	0.0048* (0.0011)	—	—
FEE2A	—	0.0059 (0.0042)	—	—	—	—	—	0.0059 (0.0041)	—
FEE2B	—	—	—	—	0.0098 (0.0060)	—	—	0.0096 (0.0059)	—
FEE3	—	—	0.0152* (0.0072)	—	—	-0.0009 (0.0071)	—	—	0.0068 (0.0048)
Observations	101	83	73	101	82	79	202	165	152
Mean ln(*L*)	-0.563545	-0.585194	-0.556581	-0.572385	-0.587763	-0.556881	-0.590213	-0.600678	-0.586453

Numbers in parentheses are standard error of parameter estimate. * indicates significance at 95% confidence level.

Table 1.3: Estimation Results for Split and Pooled over Project Version

	1+2+3		1+3		1+2		2+3	
	Scaled	**Pooled**	**Scaled**	**Pooled**	**Scaled**	**Pooled**	**Scaled**	**Pooled**
Const	0.2993 (0.8034)	0.4940 (0.6379)	-0.3034 (0.9232)	-0.0230 (0.6874)	-0.0082 (0.6664)	0.3011 (0.7991)	1.4791 (1.0287)	1.3088 (0.8444)
RE	-0.3956* (0.1390)	-0.2759* (0.1053)	-0.4685* (0.1732)	-0.3510* (0.1384)	-0.4227* (0.1564)	-0.3031* (0.1267)	-0.2886 (0.1555)	-0.2244 (0.1264)
HS	0.0734 (0.0848)	0.0355 (0.0628)	0.1196 (0.0977)	0.0605 (0.0687)	0.1247 (0.1069)	0.0775 (0.0983)	-0.0018 (0.0790)	-0.0068 (0.0728)
SEX	0.2607 (0.2124)	0.2028 (0.1684)	0.2163 (0.2481)	0.1795 (0.2059)	0.2597 (0.2419)	0.2585 (0.2038)	0.1028 (0.2642)	0.1447 (0.2206)
AGE	0.0229* (0.0080)	0.0140* (0.0062)	0.0282* (0.0093)	0.0164* (0.0070)	0.0342* (0.0095)	0.0254* (0.0086)	0.0001 (0.0087)	0.0016 (0.0076)
EDUC	-0.0114 (0.0393)	-0.0100 (0.0317)	0.0097 (0.0449)	0.0224 (0.0358)	-0.0421 (0.0390)	-0.0458 (0.0375)	0.0002 (0.0512)	-0.0139 (0.0429)
FEE1	0.0041* (0.0012)	0.0032* (0.0010)	0.0041* (0.0014)	0.0034* (0.0012)	0.0037* (0.0013)	0.0029* (0.0012)	—	—
FEE2A	0.0063 (0.0042)	0.0064* (0.0029)	—	—	0.0049 (0.0045)	0.0058 (0.0032)	0.0074 (0.0045)	0.0071* (0.0034)
FEE3	0.0108 (0.0065)	0.0103* (0.0046)	0.0092 (0.0076)	0.0112* (0.0052)	—	—	0.0156* (0.0060)	0.0117* (0.0053)
Scale Factors	0.4645	—	—	—	0.4850	—	0.3957	—
	0.4266	—	0.5620	—	—	—	—	—
Observations	257		174		184		156	
Mean ln(L)	-0.599198	-0.601452	-0.590015	-0.592957	-0.591264	-0.595141	-0.592860	-0.594799

Numbers in parentheses are standard error of parameter estimate. * indicates significance at 95% confidence level

Table 1.4: Estimation Results for Pooled and Scaled Data of 5-year Project

	1+2+3		1+3		1+2		2+3	
	Scaled	**Pooled**	**Scaled**	**Pooled**	**Scaled**	**Pooled**	**Scaled**	**Pooled**
Const	-0.3385 (0.9077)	0.0748 (0.7161)	-0.4325 (1.0360)	0.0486 (0.8465)	-0.9441 (1.0400)	-0.6070 (0.8829)	0.8622 (0.9387)	0.8098 (0.9067)
RE	-0.0521 (0.1436)	-0.0322 (0.1148)	-0.0414 (0.1533)	-0.0248 (0.1276)	-0.0876 (0.1647)	-0.0520 (0.1481)	-0.0568 (0.1604)	-0.0538 (0.1567)
HS	-0.1209 (0.0810)	-0.1087 (0.0636)	-0.1211 (0.0900)	-0.1200 (0.0740)	-0.0889 (0.0927)	-0.0702 (0.0808)	-0.1434 (0.0825)	-0.1355 (0.0802)
SEX	-0.2152 (0.2174)	-0.2630 (0.1733)	-0.3293 (0.2493)	-0.3990 (0.2098)	-0.0149 (0.2439)	-0.0486 (0.2099)	-0.3598 (0.2336)	-0.3308 (0.2252)
AGE	0.0146 (0.0080)	0.0091 (0.0063)	0.0169 (0.0090)	0.0114 (0.0074)	0.0215* (0.0092)	0.0159* (0.0079)	-0.0020 (0.0087)	-0.0019 (0.0084)
EDUC	0.0834* (0.0422)	0.0556 (0.0324)	0.0797 (0.0500)	0.0541 (0.0401)	0.0967* (0.0491)	0.0738 (0.0404)	0.0474 (0.0404)	0.0461 (0.0389)
FEE1	0.0067* (0.0013)	0.0054* (0.0011)	0.0064* (0.0015)	0.0055* (0.0013)	0.0070* (0.0014)	0.0059* (0.0013)	—	—
FEE2B	0.0110* (0.0052)	0.0088* (0.0040)	—	—	0.0117* (0.0059)	0.0111* (0.0044)	0.0072 (0.0048)	0.0072 (0.0046)
FEE3	0.0024 (0.0071)	0.0055 (0.0046)	0.0021 (0.0071)	0.0047 (0.0051)	—	—	0.0040 (0.0054)	0.0036 (0.0053)
Scale Factors	0.3180	—	—	—	0.4341	—	0.0755	—
	0.5864	—	0.4982	—	—	—	—	—
Observations	262		180		183		161	
Mean ln(*L*)	-0.591832	-0.593972	-0.589986	-0.591607	-0.583979	-0.587389	-0.587562	-0.587629

Numbers in parentheses are standard error of parameter estimate. * indicates significance at 95% confidence level

Table 1.0.6: Estimation Results for Pooled and Scaled Data of 10-year Project

	1+2+3		1+3		1+2		2+3	
	Scaled	**Pooled**	**Scaled**	**Pooled**	**Scaled**	**Pooled**	**Scaled**	**Pooled**
Const	-0.0422 (0.6104)	0.2748 (0.4772)	-0.5469 (0.7104)	-0.0615 (0.5691)	-0.3877 (0.6948)	-0.0870 (0.5859)	1.1265 (0.6408)	1.0837 (0.6168)
FIVE	0.1686 (0.1640)	0.1038 (0.1336)	0.2127 (0.1749)	0.1361 (0.1419)	0.2264 (0.1841)	0.1904 (0.1714)	-0.0434 (0.1958)	-0.0452 (0.1927)
RE	-0.2293* (0.0999)	-0.1595* (0.0760)	-0.2361* (0.1160)	-0.1632 (0.0903)	-0.2522* (0.1132)	-0.1901* (0.0950)	-0.1609 (0.1000)	-0.1537 (0.0966)
HS	-0.0353 (0.0601)	-0.0426 (0.0447)	-0.0037 (0.0684)	-0.0300 (0.0508)	-0.0074 (0.0717)	-0.0179 (0.0615)	-0.0768 (0.0559)	-0.0739 (0.0545)
SEX	0.0324 (0.1541)	-0.0192 (0.1184)	-0.0195 (0.1802)	-0.0911 (0.1446)	0.1220 (0.1700)	0.0960 (0.1431)	-0.1159 (0.1606)	-0.1003 (0.1544)
AGE	0.0182* (0.0057)	0.0102* (0.0043)	0.0231* (0.0066)	0.0134* (0.0051)	0.0248* (0.0066)	0.0178* (0.0056)	-0.0014 (0.0058)	-0.0013 (0.0056)
EDUC	0.0369 (0.0287)	0.0251 (0.0222)	0.0468 (0.0338)	0.0376 (0.0272)	0.0270 (0.0319)	0.0166 (0.0265)	0.0232 (0.0295)	0.0217 (0.0284)
FEE1	0.0053* (0.0009)	0.0041* (0.0008)	0.0051* (0.0010)	0.0043* (0.0009)	0.0051* (0.0010)	0.0042* (0.0008)	—	—
FEE2A	0.0082* (0.0040)	0.0072* (0.0027)	—	—	0.0081* (0.0041)	0.0081* (0.0029)	0.0061 (0.0032)	0.0059 (0.0030)
FEE2B	0.0090 (0.0052)	0.0084* (0.0035)	—	—	0.0079 (0.0054)	0.0082* (0.0039)	0.0095* (0.0043)	0.0093* (0.0040)
FEE3	0.0063 (0.0049)	0.0076* (0.0032)	0.0035 (0.0057)	0.0077* (0.0036)	—	—	0.0081* (0.0037)	0.0074* (0.0037)
Scale Factors	0.4744	—	—	—	0.4750	—	0.0780	—
	0.5495	—	0.6813	—	—	—	—	—
Observations	519		354		367		317	
Mean ln(*L*)	-0.609624	-0.611483	-0.608022	-0.611704	-0.603025	-0.606419	-0.601857	-0.601939

Numbers in parentheses are standard error of parameter estimate. * indicates significance at 95% confidence level

Table 1.6: Estimation Results for Pooled and Scaled Data of 5 and 10-year Project

1.5 TEMPORAL SENSITIVITY TEST OF *TWTP* AND IMPLICIT DISCOUNT RATES

1.5.1 Sequential Test

PVWTP model recovers the mean of annual *WTP*'s for each payment schedules and derives implicit discount rate by equalizing *PVWTP* of annual payment to *WTP* of lump-sum payment scheme. The procedure assumes that *PVWTP*'s are consistent and error terms are invariant across payment schemes[13]. This essay relaxes the assumption of consistency and homoskedasticity and tests them using a sequential test proposed by Swait and Louviere (1993) and adapted by Haab, Huang and Whitehead (1999) to a contingent valuation framework. Implicit discount rate is calculated based on the test result.

The null hypothesis of a sequential test consists of two steps of separate tests:

$$H_0 = \begin{Bmatrix} \beta_B^l = \beta_B^m \\ \sigma_l = \sigma_m \end{Bmatrix}.$$

where *l* and *m* indicate different payment schedules. In the first stage, the composite hypothesis tests the consistency of *TWTP* ($H_0^A = \{\beta_B^l = \beta_B^m\}$), without restriction on variances across payment schemes. Note that parameters in the first stage are coefficients of *TWTP* with different payment schedules. The second stage tests the hypothesis of invariance of the error term across payment schemes ($H_0^B = \{\sigma_l = \sigma_m\}$). In both stages, LR (Likelihood ratio) provides simple test statistics.

[13] Violation of consistency causes unidentification of *PVWTP* and discount rates, and violation of homoskedasticity leads to unidentification of *PVWTP*. For example, the result reported in Bond et al (2002) does not seem to be consistent.

The testing procedure can be conducted as follows. To test the consistency hypothesis H_0^A, the unconstrained model is split sample data reported in Table 1.3. The restricted model constrains parameters of *TWTP* to be equal across payment schemes without restriction on the variance. Following Haab, Huang and Whitehead (1999), the restricted model, named as rescaled model, can be estimated by normalizing the variance of one sub-sample (lump-sum payment scenario) to be one and estimating the relative variances of the other two sub-samples. The positive standard deviation for the pooled data is defined as

$$\sigma_j = \sigma \exp\left(\delta' w_j\right) = \sigma \exp\left(\delta_2 d_2 + \delta_3 d_3\right) \qquad (1.12)$$

where σ is the standard error of lump-sum payment scenario. A straightforward method to estimate the restricted model is a probit model with heteroskedasticity (Limdep 7.0 provides such a model). Estimation result is reported in the column titled rescaled of Tables 1.4, 1.5 and 1.6.

If the first stage hypothesis is rejected in the LR test, then stop the procedure. Rejection of the first hypothesis indicates that *TWTP* is sensitive to the payment scheme, i.e. inconsistency of *TWTP*. Respondent changes his or her value of the environmental project depending on the payment scheme.

Conditional on the failure to reject the first hypothesis, the second step is to test heteroskedasticity across payment schedules. The unrestricted conditional model in the second stage is the rescaled model used as the restricted model in the first stage. The restricted conditional model is the pooled data model stacking all samples with equal parameters in *TWTP* and dummies for payment scheme. Tables 1.4, 1.5 and 1.6 also reports the estimation results of restricted conditional model under the title of pooled.

Parameter estimates of payment schedule imply the information of the discount rate of cost stream. With failure to reject the second stage of the sequential test, the discount rate is easily derived from the ratio of parameters. Recall the normalized parameters of *PVC* defined in equations (1.3), (1.4) and (1.5) such that $\tilde{\beta}_C^1 = \frac{1}{\sigma}$, $\tilde{\beta}_C^2 = \frac{1+r_C}{\sigma r_C}\left[1-\frac{1}{\left(1+r_C\right)^{T_C}}\right]$, and $\tilde{\beta}_C^3 = \frac{1+r_C}{\sigma r_C}$. If the error variance is constant across payment type, then an implicit discount rate is

$$r_C = \frac{\tilde{\beta}_C^1}{\tilde{\beta}_C^3 - \tilde{\beta}_C^1} \tag{1.13}$$

or the solution to the nonlinear function of

$$\frac{\tilde{\beta}_C^2}{\tilde{\beta}_C^1} = \left(\frac{1+r_C}{r_C}\right)\left[1-\frac{1}{\left(1+r_C\right)^{T_C}}\right]. \tag{1.14}$$

If the discount rate varies on time intervals, $\tilde{\beta}_C^2$ and $\tilde{\beta}_C^3$ provide different implicit estimates of the discount rate.

The rejection of the second hypothesis, however, shows the heteroskedasticity of *TWTP* across payment version although *TWTP* is time-consistent. Since the structure of the standard deviation is defined to be $\sigma_j = \sigma \exp\left(\delta_2 d_2 + \delta_3 d_3\right)$, ratios of parameter estimates are

$$\frac{\tilde{\beta}_C^3}{\tilde{\beta}_C^1} = \frac{\sigma_1}{\sigma_3}\left(\frac{1+r_C}{r_C}\right) = \frac{1}{\exp\left(\hat{\delta}_3 d_3\right)}\left(\frac{1+r_C}{r_C}\right)$$

and

$$\frac{\tilde{\beta}_C^2}{\tilde{\beta}_C^1} = \frac{\sigma_1}{\sigma_2}\left(\frac{1+r_C}{r_C}\right)\left[1-\frac{1}{\left(1+r_C\right)^{T_C}}\right] = \frac{1}{\exp\left(\hat{\delta}_2 d_2\right)}\left(\frac{1+r_C}{r_C}\right)\left[1-\frac{1}{\left(1+r_C\right)^{T_C}}\right].$$

Using estimation result of the rescale factor, $\exp\left(\delta_2 d_2 + \delta_3 d_3\right)$, implied discount rates can be calculated.

1.5.2 Sequential Test Results

Each project version (A and B) has been tested for consistency and homoskedasticity of data pooling all three payment schemes (one time vs. annual payment vs. perpetuity) and through pairwise comparisons of each of the payment schemes (one time vs. annual; one time vs. perpetuity; annual vs. perpetuity). The composite hypothesis was also tested with data pooled over five and ten-year projects using a dummy variable (*FIVE*).

Table 1.3 provides the split sample estimates for each project version A, B, and AB (5 year, 10 year, and 5-10 year combined). The log likelihood of the unrestricted model is $\ln\left(L_u\right) = \sum_{j=1}^{J} n_j \cdot \ln\left(L_j\right)$, where n_j is the number of observations and $\ln\left(L_j\right)$ is the mean of log likelihood in payment scenario j. For instance, the unrestricted log likelihood of A1 (5-year one-time payment) versus A2 (5-year annual payment) is (-.564*101-.585*83=-105.52). The restricted log likelihood are the log likelihood of rescaled model for relevant combination of payment schedules in Table 1.4, 1.5 and 1.6. For A1 versus A2, the restricted log-likelihood value is -.591*184 = -108.74. The LR test statistic for the hypothesis of A1 and A2 is $LR = -2(\ln(L_r) - \ln(L_u)) = 6.44$. The likelihood ratio statistics is distributed as χ^2 with degrees of freedom (df) equal to the number of restrictions (in A1 versus A2, df = 7). With 95% confidence level, the consistency hypothesis of A1 and A2 fails to be rejected.

Conditional on failure to reject the hypothesis of consistent *TWTP*, the scaled model represents the unrestricted conditional model for the test of homoskedastic errors distribution across payment schemes. The restricted conditional model is the estimation result of pooled data in Table 1.4, 1.5 and 1.6. For A1 and A2, the restricted conditional log likelihood value is -.595*184 = -109.51, thus the LR statistic is 1.43 with df = 1. The test for A1 and A2 shows that respondents value the oyster reef restoration project consistently and the error term unobservable to researcher is identically distributed.

Table 1.7 reports results of the sequential test for all possible combinations. LR1 is the test statistics for the first stage, consistency of *TWTP* across payment schemes, and LR2 is for the homoskedasticity conditional on the first stage. Except AB1+3, all combinations of payment schedules fail to reject the consistency hypothesis, $\beta_B^j = \beta_B^k$. Except AB1+3 for which the second stage test is not necessary, the test results of the second hypothesis shows that the variance of *TWTP* is not statistically different across the payment type. Consequently, test results demonstrate that the value of oyster reef restoration program depends only on the benefit stream and individual specific variables but not on the payment schedule determined by researcher and that discount rate of cost stream can be derived from pooling data across payment schedules.

Based on the result of sequential test, Table 1.8 reports various implied discount rates from all possible combinations using equation (1.13) and (1.14). Due to the violation of monotonic probability in B3 scenario, the long term discount rate r_{13} in ten year project is problematically high (5,647%). Furthermore, the ill defined data generates larger parameter estimate of FEE3 than FEE2B, which means that we are unable to calculate the discount rate for this case. Except the case in which estimates are insignificant or cannot be calculated due to negative coefficients, the numerical solutions for implied discount rates range from 20% to more than 100%[14]. Discount rates are still relatively high but much lower than previous studies. Interestingly, for the five year project, the long term discount rate, r_{13}, is much lower than short term discount rate, r_{1A}, implying that data shows hyperbolic discount rates[15].

Finally, Table 1.9 shows the result of average and 95 percent interval of expected *TWTP* estimated through Krinsky-Robb (K-R) procedure (Haab and McConnell 2002). The first column is *TWTP* of one-time payment scheme using the estimation result in Table 1.3. *TWTP*'s of other model are estimated based on consistency and homoskedasticity of *TWTP*, except the combination annual and perpetuity payments. In Table 1.9, *TWTP* ranges between \$263 and \$277 for the five-year project and between \$159 and \$177 for ten-year project. The difference of *TWTP* between two project

---

[14] Note that the discount factor defined in equation (1.5) is derived under the restriction that $r > 0$.

[15] The hyperbolic discount rate implies that larger discount rate is applied to near-term returns than to distant-term returns (Cropper and Laibson 1999).

versions is approximately $100. The result is expected because five-year project provides the benefit with faster rate than ten-year project does. When the data is pooled over five and ten-year projects, the difference reduces to approximately $30 and is statistically insignificant.

		1+2+3	1+3	1+2	2+3
A					
	LR1	15.75	10.23	6.61	6.57
	LR2	1.16	1.02	1.43	0.60
B					
	LR1	10.12	8.79	1.72	4.81
	LR2	1.12	0.58	1.25	0.02
AB					
	LR1	17.84	13.75*	5.95	5.07
	LR2	1.93	—	2.49	0.05

* Rejected in 90% confidence interval in Chi-squared distribution with d.o.f of seven.

Table 1.7: Test Result of Insensitivity to Temporal Payment Schedules

	1+2+3	1+3	1+2	2+3
A				
†r_{13}	0.46	0.45	—	—
‡r_{1A}	0.94	—	0.98*	—
§r_{3A}	0.22	—	—	0.20
B				
†r_{13}	56.47*	N/A	—	—
‡r_{1B}	1.62	—	1.14	—
§r_{3B}	N/A	—	—	N/A
AB				
†r_{13}	1.20	—	—	—
‡r_{1A}	1.29	—	1.02	—
‡r_{1B}	0.96	—	1.05	—
§r_{3A}	0.87	—	—	0.38*
§r_{3B}	N/A	—	—	N/A
††r_{AB}	0.43	—	1.31	0.12*

N/A indicates that coefficient of Perpetuity is less than that of other payment schedule.
* One of coefficients of FEE is not significantly different from zero.
† Calculated using coefficients of One time and Perpetuity in pooled data.
‡ Calculated using coefficients of One time and Annual in pooled data.
§ Calculated using coefficients of Annual and Perpetuity in pooled data.
†† Calculated using coefficients of 5 and 10 year Annual payments in pooled data.

Table 1.8: Implicit Discount Rates

		One Time Project		1+2+3		1+3		1+2	
A									
	TWTP	263.98		268.50		263.28		276.70	
	95% KR	(170.48	629.97)	(186.81	517.92)	(184.53	510.33)	(182.52	645.28)
B									
	TWTP	176.47		163.91		159.92		176.03	
	95% KR	(135.35	223.47)	(122.99	221.41)	(115.46	215.18)	(134.99	231.88)
AB									
	*TWTP**	233.49		218.68		216.99		233.82	
	95% KR	(177.78	318.74)	(167.45	296.22)	(165.13	294.56)	(175.86	324.17)
	*TWTP***	181.82		198.22		189.01		194.98	
	95% KR	(126.74	249.03)	(148.86	270.05)	(139.13	257.57)	(139.71	271.92)

* Temporal willingness to pay for five-year project
** Temporal willingness to pay for ten-year project

Table 1.9: Mean of *PVWTP* and 95% Interval by Krinsky-Robb Procedure

1.6 CONCLUSIONS

Previous studies have defined and tested the insensitivity of willingness to pay to temporal payment schedules in terms of the present value of willingness to pay. In spite of the simple concept of insensitivity, those studies have imposed restrictive assumption that the willingness to pay is time-separable and the present value of willingness to pay is identical across different payment schemes. The simple and widely used concept of present value may not be suitable in the binary decision process of CV studies. In this essay, the insensitivity to payment schedule is redefined in terms of the temporal willingness to pay. Different from the classical definition of the temporal embedding effect, the insensitivity of temporal willingness to pay to payment schedule demonstrates the consistency of valuing behavior. Using a sequential test proposed by Haab et al. (1999), assumptions such as consistency and homoskedasticity of willingness to pay are tested before deriving implied discount rate.

The sequential test with oyster reef restoration program in Chesapeake Bay shows that holding the length of the project constant, temporal willingness to pay is statistically identical across the payment types. In holding the payment scheme constant, however, temporal willingness to pay does not vary significantly across project versions. That is, in spite of fast supply of environmental benefit, temporal willingness to pay for five-year project is statistically same with that of ten-year project. Respondents may consider the change in the environment but do not care how fast the benefit is supplied once the project is implemented.

Homoskedasticity of the error distribution across payment types confirms using the pooled data to derive implied discount rate of the cost stream. Estimated discount rates shows relatively high and significantly varying across payment schemes and project versions. However, five-year project show consistently high discount rate in short term and low discount rate in long term. Unfortunately, the response rate of vote for in ten-year project with perpetuity payment scheme violates the monotonicity of probability function. Due to small number of payment points, the violation could harm the estimation result seriously.

The benefit stream scenario was not enough to identify the discount rate of the benefit stream. Elaborate design of the benefit stream and payment schedule will provide more informative and exact result about the temporal structure and discount rate of benefit and cost streams. For example, more than three scenarios of the benefit stream enable researcher to estimate the discount rate of the benefit stream based on the sequential test. Individual discount rate with covariates for discount function can be estimated with more observations. Other functional form of willingness to pay or distribution is also recommended for future study with careful application of test procedures.

ESSAY 2

OPTIMAL, ROBUST *kk* AND UNIFORM EXPERIMENTAL DESIGNS IN BINARY CHOICE MODEL: ANALYTICAL AND EMPIRICAL COMPARISON OF EFFICIENCY AND BIAS

ABSTRACT

While bid (payment) design affects the efficiency and bias of parameter and welfare estimates in dichotomous choice contingent data, the contingent valuation literature does not provide well-established guidelines for practical bid design. In this essay, bid design utilizing a predetermined uniform distribution is proposed as a practical and robust alternative to existing optimal or naïve bid designs. Analytics and simulations show that the uniform design has lower bound of efficiency at 84 percent of D-optimum. The uniform design outperforms optimal designs when initial information is poor and outperforms naïve designs when true values of parameters are known. Simulation based on the existing data demonstrates that the uniform design provides higher efficiency and less bias than other designs even under flexible model specification such as exponential willingness to pay function.

2.1 INTRODUCTION

Binary response experiments have been widely used in fields as different as biology and economics. For example, in biological assay studies, clinical trial participants receive a randomly assigned 'dose,' and then observed at some point in the future for their 'response'. In many cases, the response variable takes the form of a binary indicator: alive or not, cancer-free or not. The varying dose information combined with the binary response variable form the necessary information to estimate the dose-response function. In economics, the contingent valuation method (CVM) closely mimics the biological assay framework. CVM measures consumer willingness to pay (WTP) for goods or services for which traditional markets do not exist: these are often public goods. Hypothetical markets, in which survey participants must decide whether to purchase a good or service (binary response) at a randomly offered bid (dose), act as a proxy for market based decisions. The dose-response function estimated from the survey responses gives a measure of WTP (or demand) for the good or service.

A pressing question in such dichotomous choice contingent valuation studies becomes, what is the optimal set of bids from which offered prices should be drawn and offered to subjects to get the most information about the population willingness to pay for the good or service of interest? Similarly, biological assay researchers must choose the optimal set of 'doses' to apply to the sample of participants to provide the most information about the population response function.

Such examples of experimental studies and environmental economics describe the unique statistical problem of designing the experiment. In the linear regression, the optimal design is to establish the limits of the support of the covariate and choose an equal number of observations from both of endpoints of the support (Casella and Berger 2002 pp 547 - 8). However, the simple design in the linear regression case cannot be applied to binary data since estimation result from binary data hinge critically on experimental design and unknown true parameters.

Experimental design points are often chosen based on an ad hoc design or based on an optimal design rule that requires prior knowledge of the true response function. The bias of parameter estimates is analytically a function of experimental points and unknown true parameters (Copas 1988), and the choice of experimental points results in dramatically different point estimate (Cameron and Huppert 1991, Cooper and Loomis, 1992, Kanninen 1995). Although parameter estimates converge asymptotically to the true parameter, the standard deviation of parameter estimates still depend on both experimental design points and unknown true parameters (e.g. Abdelbasit and Plackett 1983, Sitter 1992).

This essay proposes a practical and viable alternative to existing experimental designs. While the proposed design has applications to many fields, the essay focuses on the problem of designing the optimal bid set in dichotomous choice contingent valuation. The new experimental design, named the uniform design, draws upon the work of Lewbel et al. (2003) which assumes a continuous bid distribution to solve an identification problem in nonparametric estimation of willingness to pay in contingent valuation.

Boyle et al. (1988) suggest a similar continuous bid design known as the “method of complementary random numbers,” that constructs an empirical cumulative distribution function by utilizing prior information on the distribution of WTP. The difference between the uniform design and the method of complementary random numbers is that the uniform design selects random bid points from a predetermined uniform distribution not from the empirical distribution. Researchers can implement the bid design simply by deciding the range of the uniform distribution based on prior information of the mean and variance of willingness to pay.

The primary goal of the new design is to overcome the serious dependence of optimal designs on the true parameters. Efficacy of the proposed design is measured by the relative size of information matrix. For analytical reason, we assume that the true distribution is a logistic distribution. Compared with other designs, including optimal designs, the uniform design dramatically reduces the risk from poor information and the cost of deriving an extensive optimal design.

2.2 OVERVIEW OF EXISTING OPTIMAL DESIGNS

Suppose that a public project (*G*) enhances the environmental quality and an individual has willingness to pay (WTP_i) for implementing the project. Contingent valuation study draws the information of welfare change from the project by directly asking questions about the willingness to pay to the individual. Due to incentive compatibility, a dichotomous choice question stylized as “*Would you be willing to pay* $\$b_i$ *for G*?” is a typical form in the study, rather than an open-ended question such as

"*How much would you be willing to pay for G*?" Therefore, to implement a CV study, a researcher needs to design the set of b_i: the value (b_i) of bid (payment) points, the number of observations at each point (n_i) and total number (*J*) of bids[16]. For randomly assigned cost, b_i, a subject indicates whether b_i is acceptable or not. The binary response for the dichotomous choice question is one if WTP_i is greater than b_i, and zero otherwise.

For tractable analysis, assume that WTP_i for *G* has a constant mean (μ) and an additive *i.i.d.* error component (ε_i) with zero mean and constant variance (σ^2): $WTP_i = \mu + \varepsilon_i$. Let $F(\cdot)$ be a logistic distribution function and $f(\cdot)$ be a logistic probability function of the error term ε. Then, the probability of binary response of one is

$$\Pr_i(yes) = \Pr_i[\mu + \varepsilon_i > b_i] = F(\beta(\mu - b_i)) = F(\alpha - \beta b_i) \qquad (2.1)$$

where $F(z) = \exp(z)[1+\exp(z)]^{-1}$ and $\{\mu, \beta\}$ or $\{\alpha, \beta\}$ are parameters of interest. Usually, parameterization of the model by either of $\{\mu, \beta\}$ or $\{\alpha, \beta\}$ does not change properties of estimate, so this essay keeps parameters $\{\mu, \beta\}$ for analysis. Note that logistic distribution has unique property which simplifies the analysis, $\partial F_i / \partial \theta = f_i = F_i[1 - F_i]$.

The log likelihood of probability (2.1) is expressed as

[16] The final number of observation at each point, n_i could not be decided in prior of the survey of CV studies by researcher when the survey is in the mail format. Instead, the researcher can decide how many survey letters will be distributed with each bid point. In other survey formats such as in-person interview, n_i can be optimally designed.

$$\log L = \sum_i \left\{ (1-y_i)\ln\left[1-F\left(\beta(\mu-b_i)\right)\right] + y_i \ln F\left(\beta(\mu-b_i)\right) \right\} \tag{2.2}$$

where y_i is binary response vector. The maximum likelihood estimate (MLE) is a solution to the set of nonlinear equations of the first derivative of the log likelihood function (the score function) set to zero. From equation (2.2) and the property of logistic distribution, the score function becomes

$$S(\mu,\beta) = \sum_i (y_i - F_i)\begin{pmatrix} \beta \\ \mu - b_i \end{pmatrix}. \tag{2.3}$$

Define the weight w_i as

$$w_i \equiv f_i = \frac{\exp\left[\beta(\mu-b_i)\right]}{\left\{1+\exp\left[\beta(\mu-b_i)\right]\right\}^2},$$

then, the Hessian matrix of the logit model, the second derivative of (2.2), simplifies to be $H(\mu,\beta) = -\sum_i w_i \begin{pmatrix} \beta \\ \mu - b_i \end{pmatrix}\begin{pmatrix} \beta & \mu - b_i \end{pmatrix}$. The Fisher's information matrix is the negative of Hessian matrix such that

$$I(\mu,\beta) = \begin{bmatrix} \sum_i w_i\beta^2 & \sum_i w_i\beta(\mu-b_i) \\ \sum_i w_i\beta(\mu-b_i) & \sum_i w_i(\mu-b_i)^2 \end{bmatrix} \tag{2.4}$$

and the asymptotic variance-covariance matrix of estimates is the inverse of the information matrix.

ML estimate from (2.3) is a consistent estimate when the model is specified correctly. Thus, the main concern of optimal designs is to choose J, b_i and n_i to get the most efficient estimate under some statistical criteria. Since the Fisher information matrix is the lower bound of variance-covariance matrix, the optimality of design derives from some properties of information matrix. For instance, A-optimal design minimizes the trace of the inverse of information matrix, i.e., trace of variance-covariance matrix. Since the trace of the variance-covariance matrix is the summation of its diagonal entries that are variances of corresponding parameter estimate, A-optimal design minimizes the summation of the variance of all parameter estimates. A-optimal design results in a two-point symmetric design in the class of symmetric designs (Sitter and Wu 1993a, Mathew and Sinha 2001).

C-optimal and Fiducial designs minimize the variance or the asymptotic variance of the summary statistic of interest, such as mean or median of willingness to pay. Using Slutsky's theorem and the delta method, the asymptotic variance of estimated median is[17]

$$\operatorname{var}\left(\hat{\mu}=\frac{\hat{\alpha}}{\hat{\beta}}\right)=\left(\frac{1}{\hat{\beta}}\right)^2\left\{\operatorname{var}(\hat{\alpha})+\frac{\hat{\alpha}^2}{\hat{\beta}^2}\operatorname{var}(\hat{\beta})-2\frac{\hat{\alpha}}{\hat{\beta}}\operatorname{cov}(\hat{\alpha},\hat{\beta})\right\}.$$

C-optimality suggests a single optimal design point that is equal to the true mean or median (Wu 1988, Ford et al. 1992). However, when *WTP* function consists of a constant term and covariates, the single point is merged into the constant and the variance cannot be estimated. C-optimal design cannot identify parameter estimates of *WTP* function.

Instead of the asymptotic confidence interval, Fiducial design minimizes the length of the fiducial interval proposed by Finney (1971) using Fieller's theorem. Fieller's theorem shows the exact confidence set (parabola) of the ratio of normal random variables given desired confidence level and the roots of the parabola are the endpoints of

[17] The asymptotic variance of estimated mean is same with that of median when the model is linear.

the confidence set (See Appendix B). Fiducial interval is generally superior to the asymptotic confidence interval (Sitter and Wu 1993b). Alberini (1995) provides the expression of the square of the length of the fiducial interval as

$$(1-g)^{-2}\frac{4t}{\hat{\beta}^2}\left[\operatorname{var}(\alpha)-2\hat{\mu}\operatorname{cov}(\alpha,\beta)+\hat{\mu}^2\operatorname{var}(\beta)-g\left(\operatorname{var}(\alpha)-\frac{\operatorname{cov}^2(\alpha,\beta)}{\operatorname{var}(\beta)}\right)\right]$$

where $g = t^2 \operatorname{var}(\beta)/\hat{\beta}^2$ and t is the value of the standard normal variate for the corresponding probability mass. Fiducial design consists of two or three points depending on the sample size and confidence level (Abdelbasit and Plackett 1983, Alberini 1995).

D-optimality minimizes the volume of the confidence ellipsoid of parameter estimates. Since the determinant of a matrix represents the volume of the matrix in *k*-dimensional space, the volume of the confidence ellipsoid, i.e. the volume of variance-covariance matrix, is inversely proportional to the determinant of Fisher's information matrix. From the equation (2.4), the determinant of information matrix becomes

$$\det\left[I(\mu,\beta)\right]=\beta^2\left\{\sum_{i=1}^{N}w_i\sum_{i=1}^{N}w_i(\mu-b_i)^2-\left[\sum_{i=1}^{N}w_i(\mu-b_i)\right]^2\right\} \quad (2.5)$$

where N is the total number of observations. D-optimality turns out to maximize the determinant of the information matrix in equation (2.5). Note that D-optimality considers the entire volume of variance-covariance matrix including off-diagonal elements while A-optimal designs focus only on the summation of diagonal elements of variance-covariance matrix. D-optimal design has two design points symmetric with respect to μ (Rosenberger and Kalish Technical Report 33 Department of Statistics Pennsylvania

State University 1978, Abdelbasit and Plackett 1983, Minkin 1987, Ford et al. 1992, Nyquist 1992, Mathew and Sinha 2001).

Optimal designs, except the MSE-based design, typically consist of one, two or three bid values that depend on the correct model specification and true parameters of the underlying response function. The fundamental paradox of the optimal bid design literature is that to achieve optimality requires knowledge of the true parameters and distribution. If such information is available, estimation is unnecessary (Haab and McConnell 2002). The information required for the design is exactly the information to be estimated. Because all existing designs require some initial information about the parameters, the efficacy of each design hinges on the quality of that prior information. Poor initial information about the true parameter values results in a loss of efficiency[18] relative to the efficiency obtained from the optimal design applied with perfect information[19] (Abdelbasit and Plackett 1983).

An obvious solution for efficiency loss due to poor initial estimate is a sequential design using the consistency of estimates (Abdelbasit and Plackett 1983, Minkin 1987, Nyquist 1992). Sequential designs divide the experiments into a series of sub-experiments. The bid design is updated after each iteration based on estimates of the parameters garnered from the previous stage. Consequently, sequential designs have more design points than optimal designs. The total efficiency of a sequential design is the summation of efficiencies at all stages. Successive updates improve the efficiency of the

[18] Efficiency is defined to be a ratio of the determinant of the information matrix with poor information to the optimal determinant. This is defined and discussed in detail in section 2.3.

[19] See for example Abdelbasit and Plackett (1983) who derive the efficiency losses for D-optimal and Fiducial designs with less than perfect information.

design for poor initial estimates as Abdelbasit and Plackett (1983) argue "increasing the number of subjects at each level of the stimulus does not necessarily compensate for a poor initial estimate, but is more likely to do so if the number of design points in increased". The procedure can be designed more efficiently by considering how good the initial estimates turns out to be once the previous estimation is conducted (Minkin 1987). In spite of intuitive appeal, however, the practicality of a sequential method is still in question in contingent valuation applications.

Alternatively, Sitter (1992) introduces a minimax procedure to obtain robust designs to prevent the efficiency loss due to the uncertainty of the initial parameter values. The minimax procedure minimizes over possible design, the maximum of some optimality function over a region of the parameter space. Restricting the possible designs to the set of *kk*-designs, Sitter reports several tables under Fieller, C- and D-optimal criteria, of robust design points and the space between adjacent design points for rectangular region of μ and β representing the experimenter's uncertainty about the initial estimates[20]. For instance, Table 2.1 from Sitter's Table 1 shows optimal bid points and space between points under D-optimality[21]. The robust design has more design points over a wider range than other optimal design criteria. Sitter (1992) argues that "the less knowledge of the parameter values one has prior to the experiment, the more spread out the design should be and the more design points should be used." Although Sitter's design is robust to poor initial parameter estimates and the implementation for a specific

[20] A *kk*-design has k design points symmetric around μ and allocates an equal number of observations to each point.

[21] The efficiency of the robust design is calculated assuming that true parameters are known in both robust and D-optimal designs while the original table includes the 'worst' situation.

application is straightforward, the robust design relies heavily on the initial information and more seriously on the experimenter's confidence about information.

Usually, optimal designs assume an unbounded symmetric error distribution for the population. A series of articles provide some optimal design schemes under asymmetric error distributions. Ford et al. (1992) derive C- and D-optimal points in the case of complementary log-log and skewed logit models as well as the case in which the design region is bounded. Cooper (1993) shows optimal bid designs in the case of gamma or log-normal error distributions using MSE criterion. However, properties of optimal designs with asymmetric error distributions are not known well yet. Furthermore, MSE designs vary seriously depending on the underlying distribution and need intensive calculation over all possible bundles of bid values and observations in each bid. In other literature, Crooker and Herriges (2004) show the simulation result that the semi-nonparametric (SNP) technique estimates the model better in terms of MSE as the range of number of bids becomes wider, while the generalized maximum entropy (GME) technique does better with much fewer bids.

In addition to the minimum variance (optimal efficiency) of estimate, bias of estimate with small sample has been another issue in optimal designs. Note that although the ML estimator from equation (2.2) is consistent, the finite sample properties of the estimate are usually unknown. In specific application to CV studies, Cooper and Loomis (1992) demonstrates that the estimate of mean *WTP* is sensitive to sample design through simulation using the bid points grouped into upper, middle and lower values. The simulation results also shows that an incorrect assumption about the underlying distribution exacerbates the sensitivity of *WTP* to bid design in small samples. Due to sensitivity of estimate to bid values, Kanninen (1995) suggests a general rule-of-thumb placing bids within 15th and 85th percentiles of true *WTP* to avoid obviously excessive bids. Alberini (1995) shows that an optimal design better for estimating the median tends to perform worse for the mean or vice versa if the distribution is asymmetric.

As a special case of the small sample, Copas (1988) derives the closed form of bias for logit estimate. By expanding the score function (2.3) to the proper order of Taylor series, the bias of s*th* parameter is

$$bias_s \simeq \frac{1}{2}\sum_j \sum_k \sum_l H^{sj} H^{kl} L_{jkl} \tag{2.6}$$

where the bias is defined as $bias \equiv E\left(\hat{\theta} - \theta\right)$, H^{jk} is the inverse of $H = \left\{H_{jk}\right\}$ and L_{jkl} is the element of the Hessian matrix of the score function[22]. In a simple case of single covariate in logistic regression, the bias of estimate is

$$bias = \frac{\sum x_i^3 w_i \left(2p_i - 1\right)}{2\left(\sum x_i^2 w_i\right)^2} \tag{2.7}$$

where p_i is the probability of bid point *i* (Copas 1988). The choice of bid points affects the bias of estimates through *H* and *L*. The bias is decreasing as the number of observation increases.

[22] The exact expressions for *H* and *L* are provided in Copas (1988) or Kanninen (1995), and also in Appendix B.

β_U / β_L		μ_Δ 0	.5	1.0	1.5	2.0	2.5	3.0	3.5
1.0	*J*	2	2	2	3	3	3	4	4
	h	3.09	3.15	3.35	2.53	2.99	3.46	2.99	3.17
	Eff	1	.9993	.9883	.7500	.6165	.4847	.4907	.4075
1.25	*J*	2	2	3	3	3	4	4	4
	h	2.75	2.67	1.86	2.30	2.77	2.31	2.68	3.03
	Eff	.9789	.9673	.8638	.8074	.6815	.6291	.5216	.4369
1.5	*J*	2	2	3	3	4	4	4	5
	h	2.50	2.41	1.69	2.15	1.86	2.23	2.57	2.23
	Eff	.9342	.9119	.8544	.8368	.7702	.6543	.5518	.4724
2.0	*J*	2	2	3	4	4	5	5	6
	h	2.12	2.02	1.50	1.40	1.76	1.61	1.85	1.72
	Eff	.8187	.7795	.8158	.8547	.7979	.7134	.6149	.5321

Table 2.1: Efficiency of Equi-spaced *kk*-Designs (Sitter's Robust Design)

2.3 DETERMINANT AND EFFICIENCY OF D-OPTIMAL AND *kk*-DESIGNS

2.3.1 Determinant of General *J*-points Design

In this section are analytically derived the determinant and efficiency of the D-optimal design and robust design as special cases of general *kk*-design. *WTP* and underlying error distribution follow the same assumptions in the previous section; a constant mean μ of *WTP* and an *i.i.d.* additive error term of logistic distribution with mean zero and a constant variance of σ^2. From equation (2.5), the general expression of the determinant of information matrix is

$$\det\left[I(\mu,\beta)\right]_J = \frac{(n\beta)^2}{2}\sum_{i=1}^{J}\sum_{j=1}^{J} w_i w_j \left(b_i - b_j\right)^2$$

where *J* is the number of design points and observations are equally allocated in each point by *n* (= *N*/*J*). Obviously, the determinant depends on the relationship of each pair of two bid points; the squared distance and the weight evaluated at each point. Intuitively, increasing the distance between two points increases the determinant but due to product of weights w_i and w_j the full effect will be mixed.

Define $t_i = \beta\left(\mu - b_i\right)$ as the normalized point of *i*-th observation. By substituting t_i into the equation, the determinant becomes

$$\det\left[I(\mu,\beta)\right]_J = \frac{n^2}{2}\sum_{i=1}^{J}\sum_{j=1}^{J} w_i w_j \left(t_i - t_j\right)^2 . \quad (2.8)$$

Suppose the researcher chooses actual bid points by selecting normalized design points d_i from a standard logistic distribution and then calculates the actual bid points based on the design points and the prior information about the true parameter values[23]:

$$b_i = \mu_0 + d_i / \beta_0 \tag{2.9}$$

where μ_0 is the initial information of the population mean and β_0 is the initial information of the inverse of standard deviation. Note that d_i is equivalent to $\ln\left[p_i/\left(1-p_i\right)\right]$ in Abdelbasit and Plackett (1983). By substituting (2.9) into t_i, the normalized point t_i becomes

$$t_i = \beta\left(\mu-\mu_0\right) - d_i\left(\beta/\beta_0\right). \tag{2.10}$$

Note that d_i is in fact the design point by researcher and initial information, μ_0 and β_0, distorts the design points through $\beta\left(\mu-\mu_0\right)$ and β/β_0. From equations (2.8) and (2.10), μ_0 does not affect the squared distance, $\left(t_i - t_j\right)^2 = \left\{\frac{\beta}{\beta_0}\left(d_i - d_j\right)\right\}^2$. Furthermore, since w_i is symmetric in terms of t_i, the deviation of μ_0 from μ affects the determinant symmetrically if the design is symmetric. However, poor information of β (β/β_0) affects the determinant through either of inward or outward deviation of design point d_i.

Substitute equation (2.10) into (2.8), then the determinant of J points design is

$$\det_J = \frac{n^2 \exp\left\{2\beta\left(\mu-\mu_0\right)\right\}}{2} \sum_{i=1}^{J}\sum_{j=1}^{J} \exp\left\{\frac{\beta}{\beta_0}\left(d_i + d_j\right)\right\}\left\{\frac{\beta}{\beta_0}\frac{\left(d_i - d_j\right)}{A_i A_j}\right\}^2 \tag{2.11}$$

[23] Most practitioners directly choose b_i when implementing a CVM survey. For generality in design, the optimal design literature focuses on choosing the normalized bid points, d_i. Conditional on the prior information, there is a one-to-one mapping between normalized bid points and actual bid points.

where $A_k = \exp\left\{\frac{\beta}{\beta_0} d_k\right\} + \exp\left\{\beta(\mu - \mu_0)\right\}$. Holding $\left|d_i - d_j\right|$ constant, only one pair of d_i and d_j which is symmetric around $\beta(\mu - \mu_0)$, maximizes the summation component in the determinant. Given true information of μ and β, the optimal distance is $\left|d_i - d_j\right| = 3.09$, which is D-optimal design point. If the mean is known correctly ($\mu_0 = \mu$), the symmetric design, $d_i + d_j = 0$, always yields greater determinant than asymmetric ones given the distance[24]. Holding $d_i + d_j$ constant, expanding the distance increases the determinant first but decrease it after the critical point of distance.

2.3.2 D-optimal Design and Efficiency

Suppose that $\pm d_0$ are two symmetric design points and observations are equally assigned to them. From equation (2.11), the determinant of 2-point symmetric design becomes

$$\det\left[I(\mu,\beta)\right] = \left[\frac{Nd_0}{AB}\left(\frac{\beta}{\beta_0}\right)\exp\left\{\beta(\mu - \mu_0)\right\}\right]^2 \qquad (2.12)$$

where $A = \exp\left(\frac{\beta}{\beta_0} d_0\right) + \exp\left\{\beta(\mu - \mu_0)\right\}$ and $B = \exp\left(-\frac{\beta}{\beta_0} d_0\right) + \exp\left\{\beta(\mu - \mu_0)\right\}$.

Equation (2.12) is identical with the determinant in Abdelbasit and Plackett (1983) by substituting d_0 with p_0. The determinant of two-point bid design is maximized when initial estimates are correct as $\mu_0 = \mu$ and $\beta_0 = \beta$ and two points are placed optimally at

[24] This is consistent result with previous literatures showing that 2-point symmetric design is optimal under several criteria including D-optimality (Rosenberger and Kalish 1978, Ford et al. 1992). However, when the initial estimate of μ is not correct, some of $d_i + d_j \neq 0$ increase the determinant.

$d_0 = 1.54$. The optimal distance $\pm d_0$ corresponds to the probability mass of $p_0 = 0.824$ and $1 - p_0 = 0.176$. Optimal design point, i.e. optimal probability mass *p*, maximizing the determinant differs depending on the underlying distribution, for example, $p_0 = 0.872$ when the underlying distribution is normal. See Ford et al. (1992) for the optimal probability mass point of various distributions. The maximum value of determinant is

$$\det\left[I(\mu,\beta)\right]_{\mu_0=\mu,\beta_0=\beta} = \left[N \cdot C \cdot d_0\right]^2 = 5.01 \cdot 10^{-2} N^2 \tag{2.13}$$

where $C^{-1} = \left[1+\exp(d_0)\right] \cdot \left[1+\exp(-d_0)\right]$. Using the notation p_0 of Abdelbasit and Plackett (1983), the maximum determinant in the equation (2.13) is expressed as

$$\left[N p_0 (1-p_0) \ln\left(\frac{p_0}{1-p_0}\right)\right]^2.$$

Following Abdelbasit and Plackett (1983), the efficiency of a design is defined as the ratio of the determinant of a design at μ_0 and β_0 to the maximum determinant of D-optimal design. Therefore, from equations (2.11) and (2.13), the general expression of the efficiency of *J*-point design becomes

$$Eff_J = \frac{1}{2}\left(\frac{n \exp\{\beta(\mu-\mu_0)\}}{N \cdot C \cdot d_0}\right)^2 \sum_{i=1}^{J}\sum_{j=1}^{J} \exp\left\{\frac{\beta}{\beta_0}\left(d_i + d_j\right)\right\}\left\{\frac{\beta}{\beta_0}\frac{\left(d_i - d_j\right)}{A_i A_j}\right\}^2. \tag{2.14}$$

From equations (2.11) and (2.12), the efficiency of D-optimal design is

$$Eff_D = \left[\left(\frac{\beta}{\beta_0}\right)\frac{\exp\{\beta(\mu-\mu_0)\}}{A \cdot B \cdot C}\right]^2. \tag{2.15}$$

The efficiency shows relative increase of confidence volume of parameter estimates due to poor information.

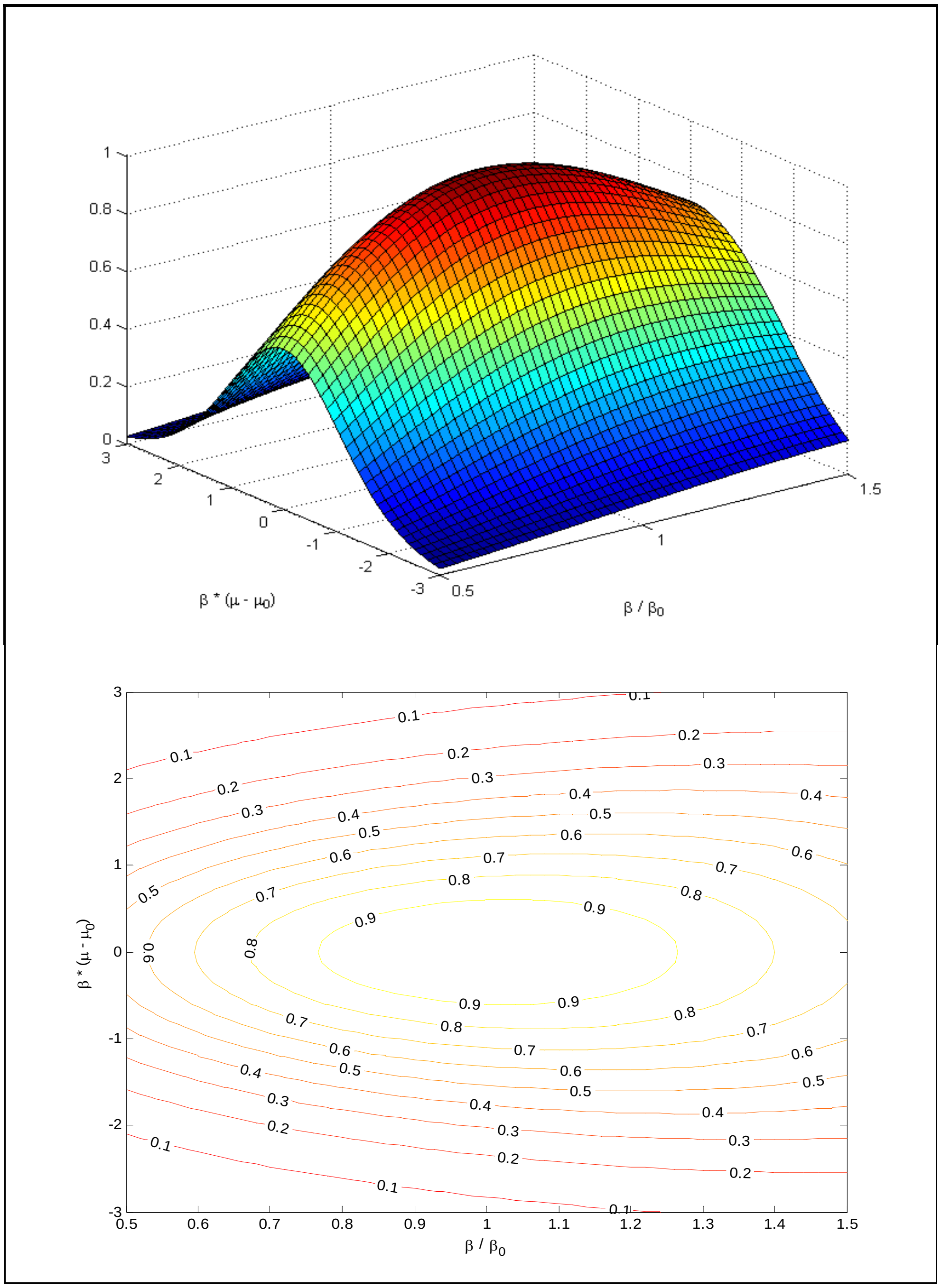

Figure 2.1: The Efficiency of D-optimal Design at Poor Initial Estimates

Figure 2.1 shows graphically the efficiency of D-optimal design under poor initial estimates. The efficiency is a function of $\beta(\mu-\mu_0)$ and β/β_0. As can be seen in Figure 2.1, the effect of poor estimates of mean is symmetric. Overestimating β, i.e. β_0 is larger than true β, is more serious to efficiency than underestimating. Note that the efficiency is calculated using β/β_0 while Abdelbasit and Plackett (1983) uses β_0/β in the Table 5 of their paper. In other expression, underestimating σ is serious problem to efficiency than overestimating σ. As the size of β is larger, i.e., the true variance is smaller, the effect of poor information is more serious. Parameterization the model by α and β does not change those properties.

In addition to efficiency, this essay defines the relative efficiency of a design as the ratio of the determinant of a design with μ_0 and β_0 to the determinant of D-optimal design evaluated at the same μ_0 and β_0 such that

$$Rff_J = \frac{\det_J}{\det_D} = \frac{Eff_J}{Eff_D}.$$

The relative efficiency shows how slow a design loses the efficiency compared with D-optimal design. Using equations (2.14) and (2.15), the relative efficiency of J-point design becomes

$$Rff_J = \frac{1}{2}\left(\frac{nAB}{N\cdot d_0}\right)^2\left(\frac{\beta}{\beta_0}\right)^{-2}\sum_{i=1}^{J}\sum_{j=1}^{J}\exp\left\{\frac{\beta}{\beta_0}\left(d_i+d_j\right)\right\}\left\{\frac{\beta}{\beta_0}\frac{\left(d_i-d_j\right)}{A_iA_j}\right\}^2. \qquad (2.16)$$

The equations (2.11), (2.14) and (2.16) are the general forms of determinant, efficiency and relative efficiency of *J*-point designs which include 2-point D-optimal design, *kk*- and robust designs in the below as special cases.

2.3.3 Equi-spaced *kk*-Designs (Sitter's Robust Design)

As noted, the determinant depends on the absolute distance between points. In the case of equally spaced *kk*-design, the distance between adjacent points except out of the lowest and highest is of the same length, which simplifies the analysis further. The determinant and efficiency of the equally spaced *kk*-designs are derived from the general forms of equation (2.11), (2.14) and (2.16). Let h_J be the distance between adjacent points and suppose that design points are arranged in the order from the lowest. Then, the relationship between any two design points can be expressed using the distance and the orders of two points such that $d_i - d_j = (i-j)h_J$, $d_i + d_j = (i+j-J-1)h_J$, and $d_i = \left[i-(J+1)/2\right]h_J$. Plugging them into equations (2.11) and (2.14), the determinant and efficiency of equi-spaced *kk*-design are

$$\det_J = \frac{n^2 \exp\{2\beta(\mu-\mu_0)\}}{2}\sum_{i=1}^{J}\sum_{j=1}^{J}\exp\left\{\frac{\beta}{\beta_0}(i+j-J-1)h_J\right\}\left\{\frac{\beta}{\beta_0}\frac{(i-j)h_J}{\tilde{A}_i\tilde{A}_j}\right\}^2 \quad (2.17)$$

and

$$Eff_J = D\sum_{i=1}^{J}\sum_{j=1}^{J}\exp\left\{\frac{\beta}{\beta_0}(i+j-J-1)h_J\right\}\left\{\frac{\beta}{\beta_0}\frac{(i-j)h_J}{\tilde{A}_i\tilde{A}_j}\right\}^2, \quad (2.18)$$

respectively, where $\tilde{A}_i = \exp\left\{\frac{\beta}{\beta_0}\left[i-(J+1)/2\right]h_J\right\} + \exp\left\{\beta(\mu-\mu_0)\right\}$ and

$$D = \frac{1}{2}\left(\frac{n\exp\left\{\beta(\mu-\mu_0)\right\}}{N \cdot C \cdot d_0}\right)^2 .$$

Table 2.1 in section 2.2 shows *J* and *h* of equi-spaced *kk*-design suggested by Sitter's robust design at various β_U / β_L and μ_Δ [25]. The efficiency, however, is calculated as the ratio of determinant of the robust design with correct initial estimates against the maximum determinant of D-optimal design using equation (2.18). Considering that D-optimal design is the simplest and optimal robust design, the efficiency implies the efficiency loss by employing more design points with different length. Efficiency loss from more design points in Table 2.1 is not too serious. As explained in the general model, there is an optimal length between design points given the number of points. For example, when the total number of design points is three, the efficiency is maximized at *h* = 1.86 and is decreasing as the length between two points is either longer or shorter. Similarly, in Abdelbasit and Plackett (1983), the three-point D-optimal design has the point at -1.85, 0 and 1.85 and the maximum efficiency is 86 percent.

To facilitate the comparison of robust design with D-optimal design, one example from the Table 2.1 are randomly chosen; the design with length (*h*) of 2.23 and design points (*J*) of 4, which is robust at $\beta_U / \beta_L = 1.5$ and $\mu_\Delta = 2.0$. Figure 2.2 shows the efficiency defined in equation (2.18) of the sample robust design with poor initial information. As can be seen in Figure 2.2 and Table 2.1, the efficiency of robust design

[25] β_U / β_L and μ_Δ are allowance level of errors in initial estimates when they are nor reliable. For the minimax design and notations, see Sitter (1992).

with $J = 4$ and $h = 2.23$ is 65.4 percent if initial estimates are correct. The efficiency of robust design is not unimodal such as that of D-optimal design but shows the symmetric effect of μ_0 given β / β_0. Furthermore, the efficiency increases when β / β_0 is smaller than one and μ_0 is correct to μ.

The relative efficiency of equi-spaced *kk*-design is just the ratio of equation (2.17) to equation (2.12) or the ratio of (2.18) to (2.15). Figure 2.3 shows the relative efficiency of the robust design with $J = 4$ and $h = 2.23$. The robust design has relative advantage over D-optimal design as the initial estimate is poor except when β / β_0 is large. Especially, if the initial estimate of the mean is too far from the true mean, the robust design always provide greater determinant than D-optimal design, which is also symmetric. Other robust designs have the same properties of efficiency and relative efficiency.

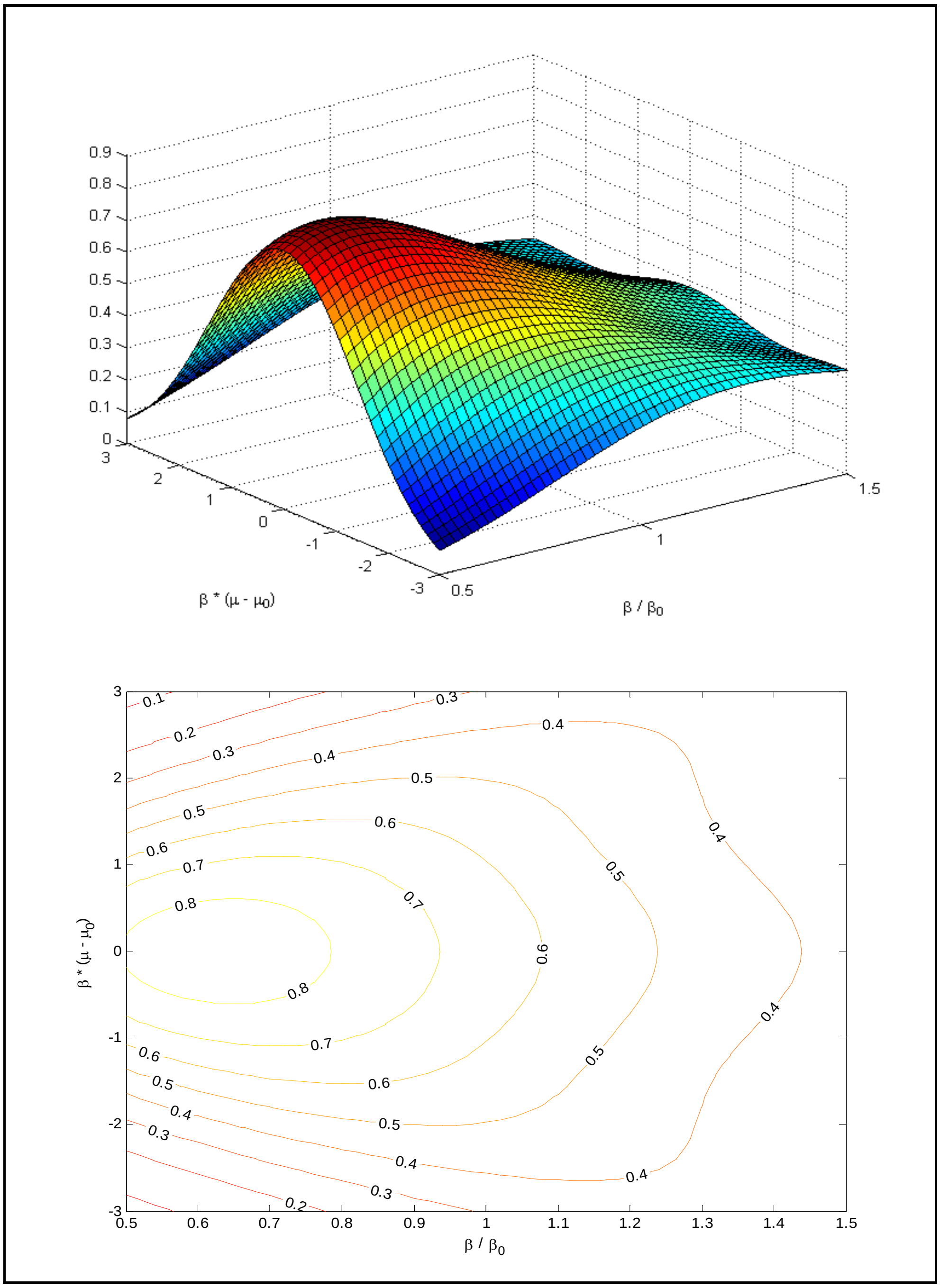

Figure 2.2: Efficiency of a Robust Design with h = 2.23 and J = 4 against D-optimum

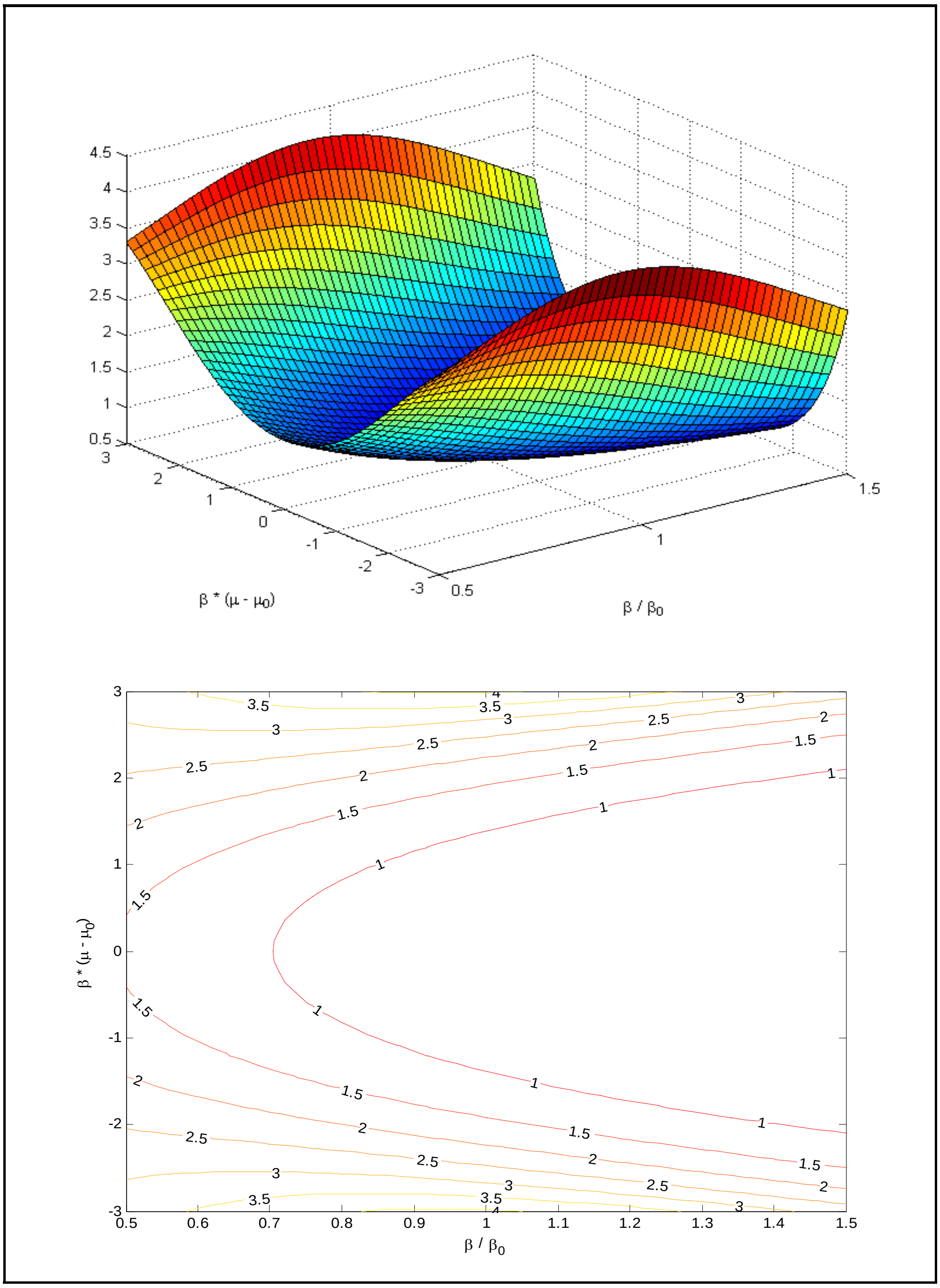

Figure 2.3: Relative Efficiency of a Robust Design with $h = 2.23$ and $J = 4$

2.4 DETERMINANT AND EFFICIENCY OF CONTINUOUS UNIFORM DESIGN

The uniform design randomly draws design points from a predetermined continuous uniform distribution. Due to the randomness of *b*, the uniform design does not have the closed form of the determinant of information matrix. Let the asymptotic distribution of *b* be $h(b)$ and take the limit on the information matrix as $J \to \infty$ so that the summation is replaced by the integral and n_j by $h(b)db$. Then, using $dt = -\beta db$ and assuming a uniform distribution for $h(b)$, the asymptotic information matrix becomes

$$I(\mu,\beta) = \begin{bmatrix} \beta \int w(t)dt & \frac{1}{\beta}\int w(t)tdt \\ \frac{1}{\beta}\int w(t)tdt & \frac{1}{\beta^3}\int w(t)t^2dt \end{bmatrix}.$$

The asymptotic determinant of information matrix is

$$\det I(\mu,\beta) = \frac{1}{\beta^2}\left[\left\{\int w(t)dt\right\}\left\{\int w(t)t^2dt\right\} - \left\{\int w(t)tdt\right\}^2\right]. \tag{2.19}$$

The asymptotic value of determinant depends on true variance and the range of uniform distribution. As the true variance is greater (*β* is smaller), the determinant becomes larger. That is, as the true willingness to pay is distributed widely, the uniform design provides more information.

Let the range of bid distribution be $\left[\underline{b},\ \overline{b}\right]$. The typical way of choosing the range is to utilize initial information such as $\underline{b}_0 = \mu_0 - r/\beta_0$ and $\overline{b}_0 = \mu_0 + r/\beta_0$ where r is researcher's choice of the range. Note that the last term in the right hand side of the equation (2.19) is

$$\int_{\underline{b}}^{\bar{b}} w(t)\,t\,dt = \left\{ t - \frac{t}{1+\exp(t)} - \ln\left[1+\exp(t)\right] \right\}_{t=\underline{b}}^{\bar{b}} . \tag{2.20}$$

By the definition of $w(t)$, equation (2.20) is zero if the range is symmetric around the mean of *t*. The other two integration terms in the right hand side of the equation (2.19) also becomes algebraically

$$\int_{\underline{b}}^{\bar{b}} w(t)\,dt = \left(\frac{-1}{1+\exp(t)} \right)_{t=\underline{b}}^{\bar{b}}$$

and

$$\int_{\underline{b}}^{\bar{b}} w(t)\,t^2 dt = \left\{ t\left\{ \frac{t\exp(t)}{1+\exp(t)} - 2\log\left[1+\exp(t)\right] \right\} - 2\sum_{\eta=1}^{\infty} \frac{\left[-\exp(t)\right]^{\eta}}{\eta^2} \right\}_{t=\underline{b}}^{\bar{b}} .$$

Note that $w(t)$ and $w(t)t^2$ are symmetric around zero.

For comparability, determinant of D-optimal design is also transformed into asymptotic expression. The exact determinant of D-optimum is the square of the rectangular area with the height of $p_0(1-p_0)\log\{p_0(1-p_0)\}$ and the width of *N*. By taking the limit as $N \to \infty$ and normalizing to the same range of the uniform distribution of uniform design, the asymptotic determinant of D-optimum becomes

$$\left[Np_0(1-p_0)\left\{ \log\left(\frac{p_0}{1-p_0} \right) \right\} \right]^2 \simeq \left[\int_{\underline{b}}^{\bar{b}} p_0(1-p_0)\left\{ \log\left(\frac{p_0}{1-p_0} \right) \right\} dt \right]^2 = \left(0.1\bar{b}\right)^2 \tag{2.21}$$

because $p_0 = .824$. As the exact value of determinant depends only on the sample size *N*, the asymptotic determinant of D-optimum depends only on the range.

The asymptotic efficiency of uniform design is defined as a ratio of the asymptotic determinant of uniform design in equation (2.19) to the asymptotic determinant of D-optimal in equation (2.21) such as

$$Eff_U = \left(\frac{1}{0.1\bar{b}\beta} \right)^2 \left[\left\{ \int_{\underline{b}}^{\bar{b}} w(t)\,dt \right\} \left\{ \int_{\underline{b}}^{\bar{b}} w(t)\,t^2 dt \right\} - \left\{ \int_{\underline{b}}^{\bar{b}} w(t)\,t dt \right\}^2 \right]. \qquad (2.22)$$

Computational examination in the personal computer shows that the asymptotic efficiency of the uniform design increases as the range becomes wide but decreases after the critical point. The maximum efficiency is 84 percent of the D-optimum and the optimal range of the uniform distribution is approximately $[-2.72,\ 2.72]$. The optimal range is between 6.2th and 93.8th percentiles in the logistic distribution. Note that D-optimal design has design points at 17.6th and 82.4th percentiles. When the uniform design has the range of two D-optimal points, the efficiency is 60 percent of D-optimum[26]. In addition, simulation also demonstrates that given the range, the symmetric design is always optimal.

Since the optimal range of the uniform design given true information is $[-2.72,\ 2.72]$, the best choice of the uniform design is $\underline{b}_0 = \mu_0 - 2.72/\beta_0$ and $\bar{b}_0 = \mu_0 + 2.72/\beta_0$. Then, the normalized term of two endpoints of uniform distribution is

$$\bar{t} = \beta(\mu - \mu_0) + 2.72\frac{\beta}{\beta_0} \text{ and } \underline{t} = \beta(\mu - \mu_0) - 2.72\frac{\beta}{\beta_0}.$$

[26] The result that uniform design has wide range of bid is consistent with previous literatures suggesting wider range for the robust estimate. However, the uniform design provides much wider than others; see, for example, Kanninen (1995) and Alberini (1995).

Plugging two normalized endpoints into equation (2.22) provides the efficiency of optimal uniform design depending on initial information. Figure 2.4 shows the effects of poor initial information on the asymptotic efficiency of optimal uniform design. The efficiency of the uniform design is maximized to be 84 percent with correct initial information where D-optimal design has the maximum determinant. Reminding the efficiency of D-optimal design in Figure 2.1, the asymptotic efficiency of the optimal uniform design is relatively flat. Thus, poor initial information is not as serious in the uniform design as in the D-optimal design.

The asymptotic relative efficiency of optimal uniform design is derived from the asymptotic expression of the determinant of D-optimal design in equation (2.12) and the asymptotic determinant of uniform design in equation (2.19). Equivalently, the relative efficiency can be derived using the asymptotic efficiency of D-optimal design and uniform design. This essay, however, demonstrates the asymptotic relative efficiency of optimal uniform design graphically in Figure 2.5 by comparing Figure 2.1 and Figure 2.4. According to the relative efficiency, the uniform design outperforms D-optimal design especially when the initial information of μ is poor. The minimum relative efficiency of the uniform design is 84 percent at the point of $\mu_0 = \mu$ and $\beta_0 = \beta$ where efficiency of the design has the maximum value. The uniform design has less relative efficiency than robust design when initial information is too much poor. However, uniform design guarantees the lower bound of relative efficiency at 84 percent while the robust design loses the efficiency more if β / β_0 is great.

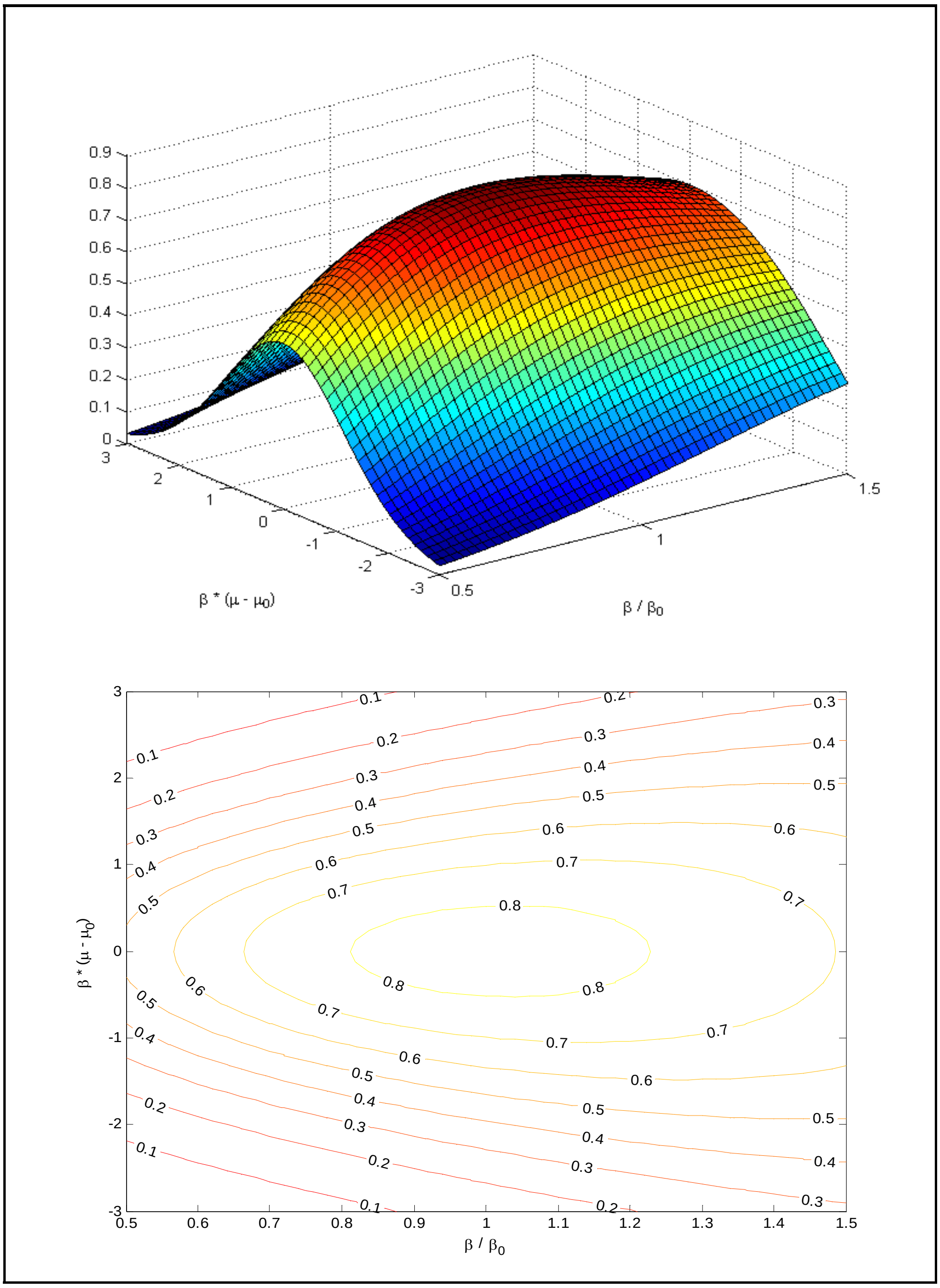

Figure 2.4: The Asymptotic Efficiency of Optimal Uniform Design

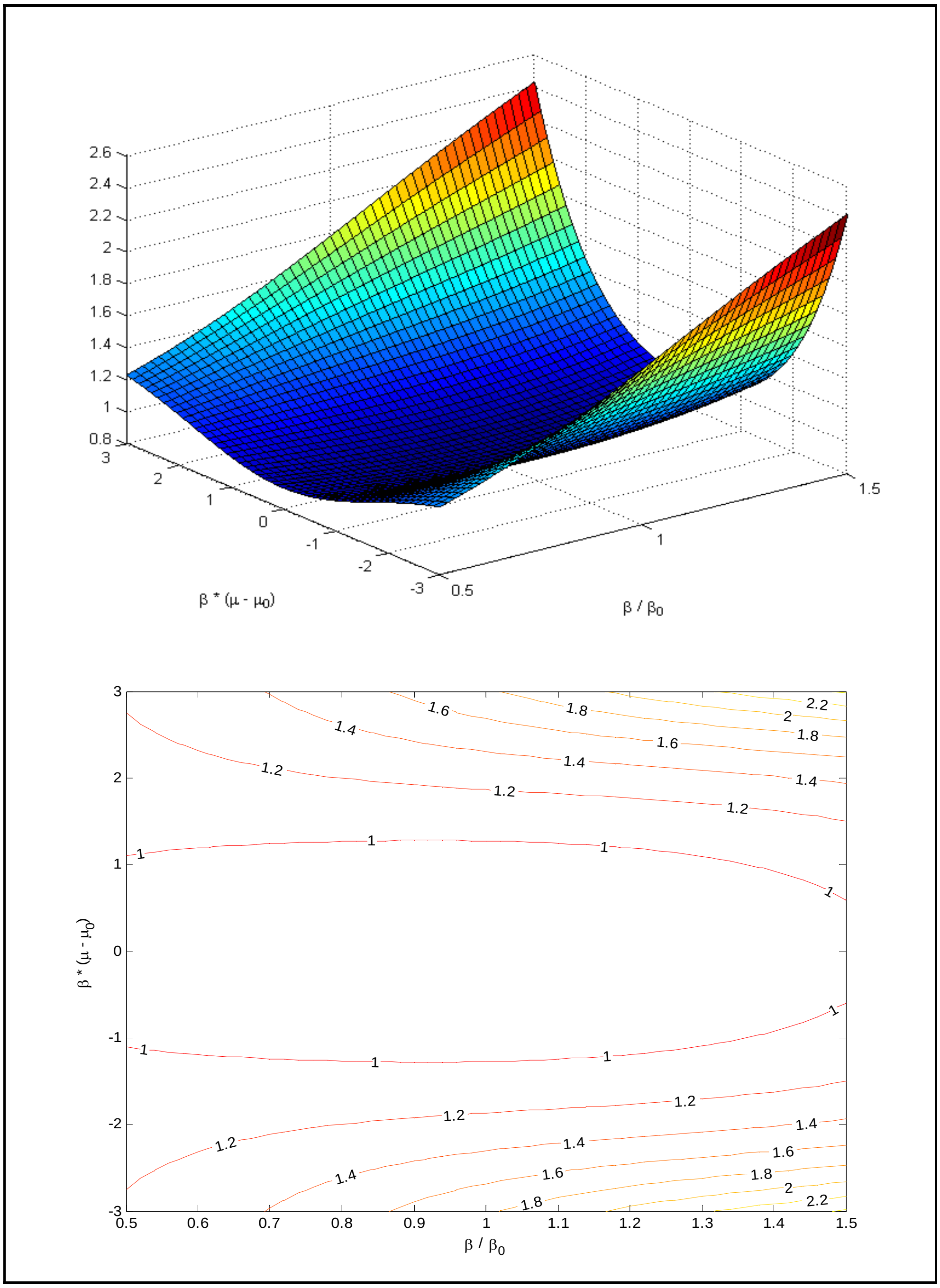

Figure 2.5: The Relative Asymptotic Efficiency of Optimal Uniform Design

2.5 BIAS OF β IN D-OPTIMAL, *kk* AND UNIFORM DESIGNS WITH KNOWN μ

The bias of ML estimate in the logit model is a function of the second and third derivatives of the log likelihood (See the function 2.6 and 2.7). In this section is analyzed simply the bias of single covariate case assuming that the true mean μ is known. Consider the bid point, $b_i = \mu - d_i / \beta_0$ and normalized point $t_i = \beta\left(\mu - b_i\right) = d_i\left(\beta / \beta_0\right)$ with true mean of μ. By substituting the bid points into the equation (2.7), the bias with single covariate becomes

$$bias = \frac{\beta_0 \sum d_i^3 w_i\left(2p_i - 1\right)}{2\left(\sum d_i^2 w_i\right)^2} \tag{2.23}$$

since $x_i = \mu - b_i = d_i / \beta_0$ with $\mu_0 = \mu$.

Suppose an equi-spaced *kk*-design with *J* points. By substituting design point, $d_i = \left[i - \left(J+1\right)/2\right]h_J$, into equation (2.23), the bias of β becomes

$$bias_J = \frac{\beta_0 \sum \Psi_i h_J\left[i - \left(J+1\right)/2\right]\left(2p_i - 1\right)}{2n\left(\sum \Psi_i\right)^2}. \tag{2.24}$$

where $\Psi_i = w_i\left\{\left[i - \left(J+1\right)/2\right]h_J\right\}^2$. The bias of *J*-point design is hardly simplified further, but the property of the bias from the design can be found through simulation. The size of the bias in *J*-point design is inversely related with the sample size. Simply, points far from zero may bias the estimate while those close to zero have opposite contribution. Since the uniform design depends on the random sampling from the uniform distribution, the asymptotic bias of uniform design with small sample is simulated.

Figure 2.6 shows the bias from the uniform design and *kk*-designs with $k = 2$ and 4 with $N = 100$. The uniform design implements 100 iterations of the 100 random draws from the uniform distribution. The *x*-axis represents the range of the support for the uniform distribution in the uniform design and the *kk*-design with $k = 2$. For the *kk*-design with $k = 4$, outer two points are allocated at $\pm x$ and the inner two points are at $\pm x/3$ to make distance between two adjacent points to be equal. In all designs, bias of β increases as the support of the bid widens, specifically, the bias of the *kk*-design with $k = 2$ increases faster than other designs. The bias of the uniform design is always smaller than the bias of *kk*-design with $k = 4$ within the specified range of the simulation. Interestingly, beyond the point of 2.5, the uniform design has smaller bias than even D-optimal design.

For the special case of two symmetric bid points such as $\pm d/\beta_0$ with $d > 0$, the bias of 2-point design can be simplified to be

$$bias_2 = \frac{\beta_0}{2Nd}\left[\exp\left(\frac{\beta}{\beta_0}d\right) - \exp\left(-\frac{\beta}{\beta_0}d\right)\right] \tag{2.25}$$

since $w(t) = w(-t)$, $p(t) = 1 - p(-t)$, and $2p(t) - 1 = -\left[2p(-t) - 1\right]$ from properties of the logistic probability and weight function. By substituting D-optimal design points, equation (2.25) represents the bias of β in the D-optimal design when the true μ is known. In 2-point design case, the bias of β can be shown to be of the order of $O\left(n^{-1}\right)$. The bias of β is always overestimated since β is positive and the bias has the same sign as β (Copas 1988). Furthermore, since $\dfrac{\partial bias_2}{\partial\left(d/\beta_0\right)} > 0$, the bias of β is an increasing function of d/β_0. The bias is also bounded below by $bias_2 > \beta/2n = \beta/N$ by L'Hopital's theorem. This is graphically shown in Figure 2.6 such that the bound of the bias for the *kk*-design with k = 2 is 0.01. The bound of the bias depends on the true variance of willingness to pay and sample size. Intuitively, as β is larger, i.e. as the variance is smaller, precise estimation is more difficult.

Figure 2.7 presents the bias of β in D-optimal design, robust design with $J = 4$ and $h = 2.23$, and the optimal uniform with various sample size. The design points are allocated at the optimal point of each design with correct information. The sample size differs from 50 representing the small sample to 1,000 for the large sample size. As shown algebraically, bias in all designs decreases as the samples size increases. The magnitude of the bias is in the order of *kk*-design, the uniform and D-optimal design. Figure 2.8 shows the effect of poor initial information on bias of β in D-optimal design, robust design with $J = 4$ and $h = 2.23$, and the optimal uniform. True β is set to be one and the x-axis represents poor initial estimate of β_0. At $\beta_0 = 1$, the bias of the uniform design corresponds the bias at $x = 2.72$ of uniform design in the Figure 2.6 and the bias of the D-optimal design responds to the bias at $x = 1.54$ of two-point design. Usually, uniform design has larger bias than D-optimal design but smaller than robust design except that β_0 is extremely small.

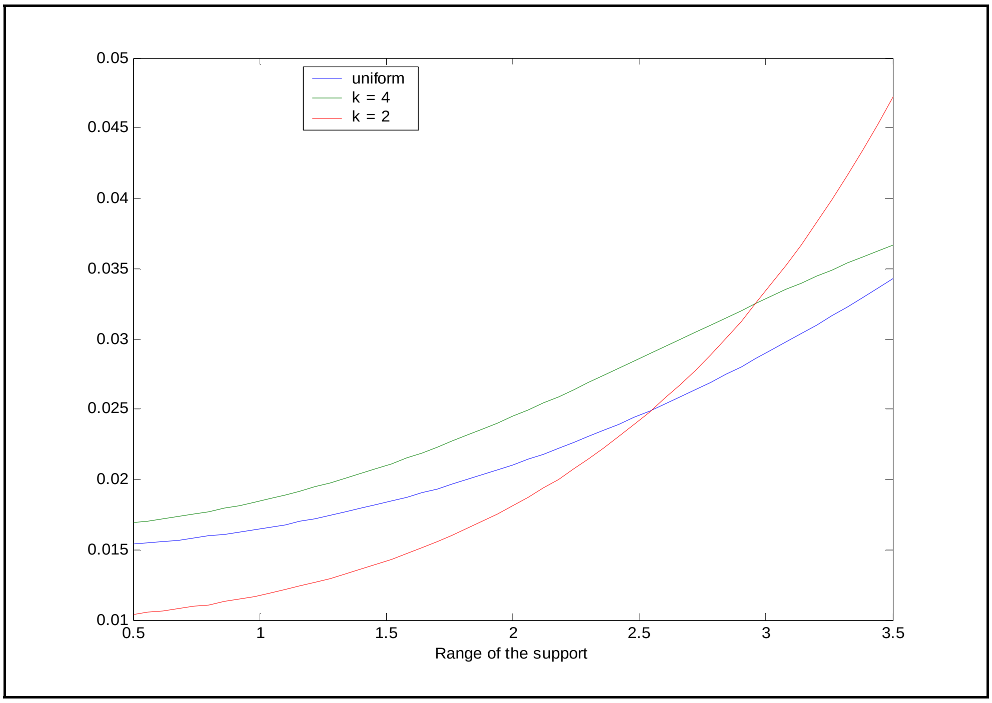

Figure 2.6: Bias of β in *kk* ($k = 2$ and 4) and Uniform Designs, $N = 100$

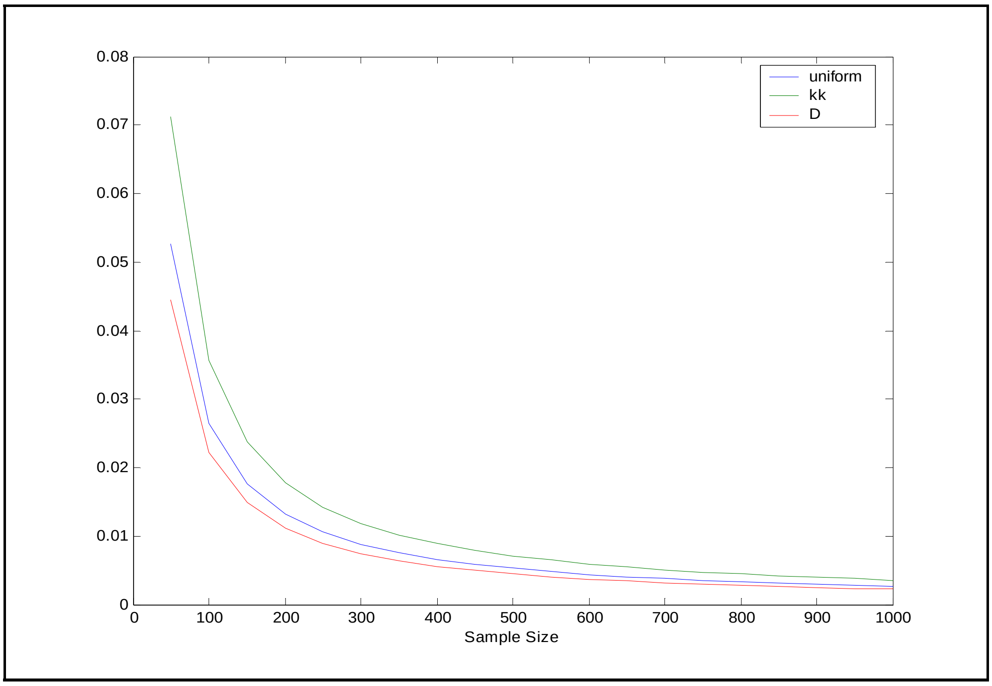

Figure 2.7: Bias of β in D-optimal, Robust with $J = 4$ and $h = 2.23$, and Uniform Designs with Different Sample Size

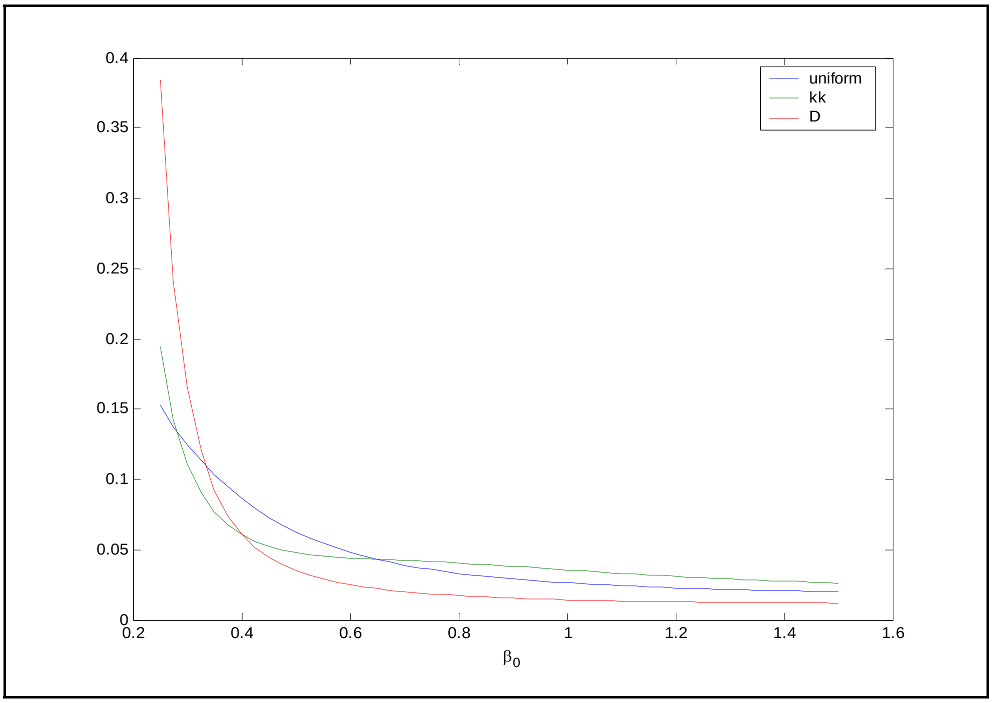

Figure 2.8: Bias of β in D-optimal, Robust with $J = 4$ and $h = 2.23$, and Uniform Designs with Poor Initial Estimate, $N = 100$

2.6 MONTE CARLO SIMULATIONS

In this section, a series of Monte Carlo simulation compares the relative performance of D-optimal, equi-spaced *kk*, and uniform designs. The simulation scenarios cover the optimal uniform range (scenario 1), asymptotic properties with a large sample (scenario 2), various sample size (scenario 3), poor initial estimates of μ and β (scenario 4 and 5, respectively) and flexible error distributions such as beta distribution (scenario 6). Scenarios 1, 2 and 3 assume that the true parameters are known in allocating bid points in D-optimal, *kk*- and uniform designs.

The basic model is a constant willingness to pay; $WTP_i = \mu + \varepsilon_i$, where $\mu = 100$ and ε_i is logistically distributed with zero mean and the standard deviation (σ) of 30. The parameters in estimation are μ and β (= $1/\sigma$) as in equation (2.1). D-optimal design consists of two points at $\mu_0 \pm 1.54/\beta_0$. Sitter's robust design with J = 4 and h = 2.23 represents the equi-spaced *kk*-design by allocating bids at $\mu_0 \pm 3.345/\beta_0$ and $\mu_0 \pm 1.115/\beta_0$. With random number seed of 710602, the simulation is conducted using Gauss 5.0 of Aptech Systems Inc. and CML Version 1.0.35 for maximum likelihood calculation.

Scenario 1 reported in Table 2.2 estimates binary model with uniform design by drawing bids from a uniform distribution with various ranges. The range of the uniform distribution differs from $\pm 1.71/\beta$ to $\pm 3.72/\beta$. Table 2.2 shows the result of 100 iterations with the sample size of 320. The sample averages of μ and σ are 100.0989 and 29.9176, respectively. The parenthesis reports the standard error in 100 iterations and *Eff* represents the efficiency calculated as the ratio of the determinant of uniform design to D-optimum. *RMSE* (Root Mean Squared Error) and *MAE* (Mean Absolute Error) of μ are also reported for comparison. The efficiency is maximized to be 83.95 percent when the bid points are drawn within the range of $[\mu - 2.72/\beta, \mu + 2.72/\beta]$, the result which is consistent with analytical demonstration and confirms the optimal uniform design. The narrow range of the uniform distribution is more serious to the efficiency than the wide range. Interestingly, the standard error of μ increases with range while the standard error of σ decreases.

d	1.72	2.22	2.72	3.22	3.72
μ	100.4421 (4.1466)	99.9719 (4.2302)	99.9065 (4.3130)	99.6234 (4.6786)	99.7744 (4.7611)
σ	30.4492 (5.1925)	30.0612 (3.6852)	30.2802 (3.2124)	30.1802 (3.2274)	29.7865 (3.1162)
Eff	66.9283	80.3383	83.9492	81.3053	74.5273
RMSE	54.3148	54.3250	54.3360	54.3784	54.3729
MAE	41.5063	41.5022	41.5266	41.5581	41.5661

* Results of D-optimal are μ = 100.0832 (4.3723), σ = 29.7660 (2.5358), *RMSE* = 54.3474 and *MAE* = 41.5301, and *kk*-design are μ = 99.8898 (4.9126), σ = 30.0528 (3.3509), *RMSE* = 54.3878 and *MAE* = 41.5965.

Table 2.2: Different Range of Uniform Distribution with 100 Iterations, N = 320

Table 2.3 shows the results of all designs with 1000 observations. Based on analytics and simulation result in scenario 1, the optimal uniform design draws bid points from the optimal uniform distribution of $\left[\mu - 2.72/\beta, \mu + 2.72/\beta\right]$. The second column titled "Actual" reports the sample mean and inverse of the standard deviation (β). Parentheses show the standard error of estimates reported by the Gauss program. Since the scenario assumes that true parameters are known, the determinant of D-optimal design is the maximum value of the determinant. The efficiency of the uniform and *kk*-designs are 83.2300 and 62.8604 percent of D-optimal design, respectively, all being consistent to but slightly lower than analytical solutions.

For parameter estimate of mean willingness to pay (μ), the result of uniform design is closest to the true or actual value of μ, and furthermore, uniform design has the smallest the standard error of parameter estimate. Since the uniform design provides the minimum bias and standard error of μ, both *RMSE* and *MAE* are the lowest in the uniform design among three designs. Interestingly, the simple simulation provides counter-result of A- and C-optimalities when considering the variance of both parameter estimates of μ and β [27]. Although the standard error of β estimate in the uniform design is largest, the difference in magnitude is still ignorable and consequently the summation of variances is minimized in the uniform design.

[27] In fact, the uniform design is posited in the opposite side of the design spectrum from the optimal designs in the sense that the uniform design consists of bid points as many as the number of observations but optimal designs allocate all of them at one or two design points. Although the simulation draws the sample observation without iterations, the uniform design with a large sample provides potential superiority under A- and C-optimalities

	Actual	D-optimal		kk		Uniform	
μ	99.3460	97.9300	(2.5669)	97.2631	(2.8061)	98.9095	(2.3491)
β	0.0333	0.0314	(0.0017)	0.0348	(0.0020)	0.0331	(0.0021)
Eff		100		62.8604		83.2300	
RMSE		54.4952		54.5166		54.4786	
MAE		41.2281		41.2599		41.1967	

Table 2.3: Estimation Results of D-optimal, *kk* and Uniform Designs with $N = 1000$

Table 2.4 reports estimation results of the scenario 3 varying the sample size from 80 to 640 with 100 iterations. Hereafter, the parenthesis reports the standard error in 100 iterations. The efficiency is measured from the average of the determinant in iterations. The simulation result shows 65.26 ~ 66.91 percent of efficiency for *kk*-design with h = 2.23 and $J = 4$, which is around the analytical solution of 65.43 percent. The efficiency of *kk*-design, analytically, does not depend on the sample size but is inversely related with the number of different bids (See the equation 2.18). The efficiency of the uniform design is 83.38 ~ 86.28 percent around the asymptotic efficiency of 84 percent, the result which is independent of the sample size.

	Actual	D-optimal		*kk*		Uniform	
	$N = 80$						
μ	99.4097	99.1955	(9.1829)	99.2402	(10.8442)	98.6798	(8.5866)
σ	30.2086	31.2072	(6.4858)	29.3343	(5.9843)	29.2926	(6.2882)
Eff		100		66.9077		83.3844	
RMSE		55.1554		55.3460		55.0535	
MAE		42.0547		42.3707		41.9465	
	$N = 160$						
μ	99.7602	99.9925	(5.9714)	99.1349	(7.0857)	99.5768	(5.8555)
σ	30.1389	30.3131	(4.9260)	29.5214	(4.0622)	29.6650	(4.3637)
Eff		100		66.1130		85.6703	
RMSE		54.8221		54.9164		54.7977	
MAE		41.9145		42.0114		41.8692	
	$N = 320$						
μ	99.9744	99.9537	(4.7597)	99.8970	(4.8828)	100.1410	(4.7288)
σ	29.9883	30.6379	(3.3750)	29.5047	(2.9103)	29.9610	(3.0482)
Eff		100		65.2646		86.2760	
RMSE		54.4653		54.4988		54.4770	
MAE		41.6550		41.6558		41.6691	
	$N = 640$						
μ	100.0564	100.0261	(3.1681)	100.2129	(3.3033)	100.1940	(3.0738)
σ	29.8866	29.9448	(1.9328)	29.8000	(1.8732)	29.9058	(2.3439)
Eff		100		65.3181		85.0355	
RMSE		54.2470		54.2615		54.2374	
MAE		41.5070		41.5275		41.4963	

Table 2.4: 100 Iterations of Scenario 1 with $N = 80$, 160, 320 and 640

While the efficiency is independent of the sample size, the simulated standard error of estimates, however, decreases in all designs as the sample size increases. Furthermore, uniform design provides the minimum variance of the mean estimate and of the sum of standard errors of both parameter estimates confirming the counter-evidence in Table 2.3 except one case of $N = 640$. Therefore, the simulation results varying the sample size with 100 iterations support the potential usefulness of the uniform design under even C-, Fiducial interval and A-optimality criteria. Under *RMSE* and *MAE* criteria, the results uphold the outperformance of the uniform design at least in estimating μ, except only one case of $N = 320$. Table 2.4 roughly shows decreasing tendency of the bias of estimate σ in all bid designs as the number of observations increases as shown in Figure 2.7. The decreasing tendency of the bias is also detected in the estimate of μ but not clear.

Scenarios 4 and 5 investigate the performance of bid designs with poor initial information of μ and β holding the sample size at 320. The initial information of μ varies from 55 to 145 corresponding to [1.5, -1.5] of $\beta(\mu - \mu_0)$. Rather than information of β, for convenience, information of σ varies between 10 and 60 which also corresponds 0.3 and 2 of β / β_0. *Rff* represents the relative efficiency in terms of percentage. By definition, *Rff* of D-optimal design is always 100 percent.

Table 2.5 shows the effect of poor information of μ assuming known standard deviation. Sample mean of willingness to pay (μ) in the simulation is 100.0989 and standard deviation is 29.9176. Among 100 iterations, one iteration step with $\mu_0 = 145$ reports failure in calculating function of D-optimal design. Information of μ has symmetric effect on the relative efficiency of *kk*- and uniform designs, consistent to the analytical comparison. Uniform design provides the acceptable efficiency with good initial information and superior relative efficiency at the relatively extreme circumstance.

Table 2.6 provides estimation result with poor information of σ. Sample value of μ is 100.2090 and actual σ is 30.0047. Analytically, poor information of σ affects the relative efficiency asymmetrically (see Figure 2.3 and 2.5). For *kk*-design with $J = 4$ and $h = 2.23$, the relative efficiency decreases as β / β_0 increases, i.e. σ_0 becomes larger. The uniform design has the lowest relative efficiency at correct initial estimate of σ_0 and the relative efficiency increases as the poorness increases to any direction.

Unfortunately, properties of estimation bias are not clearly found in poor initial estimates scenario as in the analytical comparison. However, *RMSE* and *MAE* show that the uniform design performs fairly well with poor information of μ and σ. Except $\mu_0 =$ 125, uniform design yields the non-worst, usually best estimation result in terms of *RMSE* and *MAE* in Table 2.5 and outperforms D-optimal and *kk*-designs when σ_0 is larger than the true in Table 2.6. Uniform design also provides the best result under A- and C-optimality when the initial estimate of σ is poor.

	D-optimal		*kk*		Uniform	
	$\mu_0 = 55$					
μ	100.3421	(4.3694)	100.3697	(5.8708)	100.2821	(5.2056)
σ	29.2950	(4.0077)	30.1820	(3.0038)	29.9249	(4.0637)
Rff	100		113.7679		110.5209	
RMSE	54.3777		54.4384		54.4259	
MAE	41.5198		41.6233		41.5847	
	$\mu_0 = 75$					
μ	100.5234	(4.2747)	100.0102	(5.3408)	100.1249	(4.2405)
σ	29.7526	(2.9618)	29.9169	(3.4280)	30.1654	(3.1692)
Rff	100		76.6577		92.2807	
RMSE	54.3500		54.4036		54.3387	
MAE	41.5370		41.5892		41.5148	
	$\mu_0 = 100$					
μ	100.0832	(4.3723)	99.8898	(4.9126)	99.9065	(4.3130)
σ	29.7660	(2.5358)	30.0528	(3.3509)	30.2802	(3.2124)
Rff	100		65.9432		83.9492	
RMSE	54.3474		54.3878		54.3360	
MAE	41.5301		41.5965		41.5266	
	$\mu_0 = 125$					
μ	100.4928	(4.6248)	100.2123	(4.4947)	98.9705	(4.8026)
σ	29.9357	(2.9464)	30.2059	(3.5096)	30.2015	(3.6458)
Rff	100		76.2574		89.3220	
RMSE	54.3433		54.3787		54.3811	
MAE	41.5363		41.5666		41.5526	
	$\mu_0 = 145$*					
μ	98.4429	(10.9916)	99.6627	(5.0660)	100.1229	(5.2359)
σ	30.9991	(15.0630)	30.0419	(3.5605)	29.8914	(3.3762)
Rff	100		113.6570		113.0370	
RMSE	55.2745		54.4147		54.4067	
MAE	42.1239		41.6157		41.5679	

* 1 function calculations failed in the D-optimal design

Table 2.5: Poor Initial Estimates of μ with 100 iterations

	D-optimal		*kk*		Uniform	
	$\sigma_0 = 10$					
μ	99.9413	(3.5821)	100.1924	(3.2139)	100.2011	(3.3338)
σ	31.0174	(7.4705)	30.2099	(4.6208)	29.9297	(7.7508)
Rff	100		188.2141		102.9199	
RMSE	54.4542		54.4497		54.4528	
MAE	41.5098		41.5051		41.4916	
	$\sigma_0 = 20$					
μ	100.2520	(3.7448)	100.0511	(3.9763)	100.2147	(3.8841)
σ	30.0553	(3.6172)	29.7537	(2.9083)	30.1323	(4.5760)
Rff	100		106.0529		88.1675	
RMSE	54.4621		54.4779		54.4660	
MAE	41.5331		41.5537		41.5404	
	$\sigma_0 = 30$					
μ	100.4169	(4.1858)	99.8175	(5.3151)	100.0356	(3.4741)
σ	29.9782	(2.9953)	29.6584	(3.0420)	29.9387	(3.4376)
Rff	100		65.2806		84.1582	
RMSE	54.5028		54.5577		54.4695	
MAE	41.5942		41.6461		41.5504	
	$\sigma_0 = 40$					
μ	100.5465	(4.9207)	99.6000	(6.4361)	99.7180	(3.9807)
σ	30.0485	(2.6084)	29.4623	(3.3248)	29.9082	(3.0037)
Rff	100		51.8393		89.4349	
RMSE	54.5438		54.6527		54.5005	
MAE	41.6851		41.7485		41.5897	
	$\sigma_0 = 50$					
μ	100.5412	(5.8683)	99.3532	(7.4830)	100.5217	(5.3780)
σ	30.0761	(2.5784)	29.1542	(3.3584)	29.8827	(3.0294)
Rff	100		51.4139		106.9764	
RMSE	54.6406		54.7920		54.5640	
MAE	41.7976		41.8727		41.6712	
	$\sigma_0 = 60$					
μ	99.9107	(8.0184)	99.2053	(8.2823)	99.7317	(5.1448)
σ	29.6054	(2.7518)	29.2356	(3.7721)	30.0000	(3.4945)
Rff	100		63.6623		150.8424	
RMSE	54.8710		54.9194		54.6104	
MAE	42.0027		42.0048		41.6670	

Table 2.6: Poor Initial Estimates of σ with 100 iterations

One of interesting questions about the existing designs is how they perform if the true distribution is unknown and asymmetric because optimal bid designs can be optimal only when the underlying assumptions are correct. Optimal bid points are hardly known in the asymmetric distribution. The reliance on the prior assumption is also serious in the *kk*-design. Scenario 6 assumes that true error distribution is a beta distribution with various shape parameters to compare the performance of bid design in the case of unknown asymmetric error distribution. The beta distribution is either right- or left-skewed depending on shape parameters, *a* and *b*. However, since the estimation model is specified as logit, the scenario represents misspecification of the error distribution. The true mean and standard error are assumed to be known for bid design.

Table 2.7 shows the simulation result with shape parameters (2, 3), (2.5, 2.5) and (3, 2). In this simple simulation, surprisingly, D-optimal design has the largest determinant no matter what the shape of distribution is in terms of relative efficiency. The relative efficiency of the uniform design shows almost 87 percent of D-optimal design and that of *kk*-design is slightly higher than 72 percent. D-optimal and uniform designs show a tendency that when the distribution is left- (right-) skewed, they under- (over-) estimate the mean, while *kk*-design shows the result in the opposite way. Uniform design is superior in terms of C-optimality but the *kk*-design yields better result in terms of A-optimality. The properties of estimation result from asymmetric error distribution are analyzed more in detail using log-normal distribution and the actual survey data in the next section.

	Actual	D-optimal		kk		Uniform	
	*(2, 3) (120, 33.0797**)						
μ	120.3063	119.0280	(5.1557)	121.1151	(5.4924)	118.6580	(5.0230)
σ	33.1084	41.3910	(4.4445)	33.6435	(3.0979)	39.4347	(4.5767)
Rff		100		72.2039		87.1280	
RMSE		60.1123		60.1492		60.1171	
MAE		49.8730		50.0022		49.8346	
	*(2.5, 2.5) (150, 33.7618**)						
μ	150.1951	150.3924	(5.4428)	150.8241	(6.0330)	150.1364	(5.2994)
σ	33.6002	41.7015	(4.8726)	34.6713	(3.0551)	39.3621	(5.0721)
Rff		100		72.4775		87.2118	
RMSE		60.9814		60.0449		60.9943	
MAE		50.5794		50.6072		50.5687	
	*(3, 2) (180, 33.0797**)						
μ	179.8880	180.2808	(5.2549)	179.6522	(5.0183)	182.1175	(5.1080)
σ	33.1867	40.2857	(4.0691)	34.4068	(3.0565)	38.8170	(4.4010)
Rff		100		72.8120		86.3334	
RMSE		60.2527		60.2417		60.2774	
MAE		50.0462		50.0626		49.9507	

* The first parenthesis represents the shape parameter (a, b) of beta distribution and the second shows the true mean and standard error.

** The standard error is normalized as that of logistic distribution by multiplying $\sqrt{3}/\pi$ to the standard error of beta distribution.

Table 2.7: Flexible Beta for Error Distribution

2.7 AN APPLICATION TO ALBEMARLE AND PAMLICO SOUNDS DATA

While previous results provide insight into the potential usefulness of the uniform design, analytical and simulation results in previous sections depend on a known distributional form and simple parametric specification. This section compares D-optimal, *kk*- and uniform designs as well as the original design by simulating true willingness to pay from the actual survey data. The focus of comparison is on the performance of different designs when nonnegative willingness to pay function has covariates and the error distribution is asymmetric.

Huang, Haab and Whitehead (1997) studied the willingness to pay for the water quality improvement in the Albemarle and Pamlico Sounds in eastern North Carolina. The original data consisted of double bounded dichotomous questions. However, in this section, only responses to the first question were considered for design comparison. True willingness to pay was simulated as follows. First, under the assumption of exponential willingness to pay function and log normal error distribution, a probit model was implemented to estimate parameters of willingness to pay[28]. Willingness to pay for the water quality improvement in Albemarle and Pamlico Sounds was

$$\ln\left(WTP\right) = 3.8623 + 0.1034 \cdot INC - 0.3580 \cdot D + \varepsilon \text{ and } \varepsilon \sim N\left(0, 0.3047^{-2}\right)$$

where *INC* is income level and *D* is a dummy variable for Pamlico sound only. The expected willingness to pay, $E\left(WTP\right) = \exp\left(\overline{x}'\beta + .5\sigma^2\right)$, was \$12340.51 in the sample. The median of willingness to pay was \$56.60 and the mean of the expected log

[28] To facilitate *kk* design of $J = 4$ in design comparison, first two observations were dropped since the original data includes 726 observations.

willingness to pay, $E\left(x'\beta\right)$, was \$3.99. Next, the true individual willingness to pay was simulated by adding a random error from normal distribution to the deterministic log willingness to pay assuming that the estimation result in the first step is true parameters. The sample average of willingness to pay, $Average\left(WTP\right)$, was \$4682.27. Finally, the simulated true willingness to pay was used to generate the sample dichotomous response for each bid design. The simulated response is one if $\ln\left(WTP\right) > \ln\left(bid\right)$, and zero, otherwise.

Bid set of D-optimal, *kk* with *J* = 4 and *h* = 2.23, and uniform designs were constructed assuming that the true parameters were known. Initial information used in designs was the mean and standard error of log willingness to pay; $E\left(x'\beta\right) = \mu = 3.9941$ and $\sigma = 0.3047^{-1}$. Also, to adjust the analytical solution of the logit model for the normal distribution, the standard logit variates of *kk*-design and the uniform design in the previous section were multiplied by $\sqrt{3}/\pi$. Thus, *kk*-design had bid points of $\mu \pm 0.61\sigma$ and $\mu \pm 1.84\sigma$, and the support of optimal uniform design was $\left[\mu - 1.50\sigma,\ \mu + 1.50\sigma\right]$. The D-optimal bid points were $\mu \pm 1.14\sigma$ following previous studies. Optimal points and range of uniform design were transformed to nonnegative dollar amount by taking exponential. Finally, the dollar value of bids in the D-optimal design was {\$1.29, \$2288.12} and bid amount of *kk*-design were randomly selected from {\$0.13, \$7.22, \$408.14, \$23077.07}. The optimal uniform design had a uniform distribution of [\$0.40, \$7448.07]. In addition, the original design in Huang, Haab and Whitehead (1992) consisted of {\$100, \$200, \$300, \$400}, which corresponded from 4.6052 to 5.9915 of the expected log willingness to pay. Note that log value of bids in the original design is higher than the mean of the expected log willingness to pay.

Table 2.8 shows the estimation results of the Albemarle and Pamlico Sounds data. Since the log willingness to pay is a linear model and error term is symmetric in terms of log value, D-optimal design was expected to provide the maximum determinant of information matrix. However, uniform design yields the largest determinant followed by the D-optimal, the original and *kk*-designs. The original design has also the determinant larger than the *kk*-design although the original design is a one-sided design (i.e., all bids are greater than the mean of expected log willingness to pay). The result strongly supports that uniform design outperforms other designs under D-optimal criterion when the error distribution is asymmetric.

The uniform design also outperforms other bid designs in terms of variance of estimate, median willingness to pay and *RMSE*. The summation of the individual variance of estimates is minimized in uniform design demonstrating that the uniform design yields the best result under the A-optimality. The simulated confidence interval of the median willingness to pay shows that the uniform design still performs well under the C-optimality. The uniform design provides the second best result of the expected willingness to pay following the *kk*-design, the result which is, in fact, the closest value to the sample average.

	True	D-optimal	*kk*	Original	Uniform
Constant	3.8623	4.1112 (.3951)*	3.4456 (.4700)*	4.0511 (.3660)*	4.1953 (.3483)*
INC	0.1034	-0.0458 (.0893)	0.0511 (.1028)	0.0877 (.0579)	-0.0266 (.0793)
D	-0.3580	-0.2909 (.3708)	-0.1640 (.4117)	-0.1708 (.2361)	-0.4584 (.3320)
ln(Bid)	0.3047	0.3424 (.0174)*	0.2968 (.0182)*	0.4179 (.0896)*	0.3249 (.0218)*
det(I)		9.1766e+7	4.2038e+7	5.4386e+7	11.2556e+7
Mean	12340.51 (4682.27)**	3256.79	9943.71	1235.81	5576.14
Median	56.60	45.80 (31.56 64.61)	34.03 (22.49 50.52)	70.59 (41.72 112.59)	48.94 (34.36 66.30)
RMSE		46946.28	47209.22	47040.18	46943.93
MAE		6909.15	12881.86	5307.73	8909.99

* Estimates are statistically significant with 95% confidence level.
** The sample average of *WTP*

Table 2.8: Estimation Result with Albemarle and Pamlico Sounds Data

2.8 CONCLUSIONS

This essay introduces a new bid design utilizing a predetermined uniform distribution. The new design assumes continuity and randomness of bid points. Both analytically and through Monte Carlo simulations, this essay compares the efficiency and relative efficiency of the uniform design with D-optimal design and one of Sitter's robust designs. D-optimality is chosen to represent optimal criterion because of its popularity and usefulness. Sitter's robust design is a member of symmetric designs, in which design points are selected depending on researcher's belief about the correctness of information. Uniform design assumes continuity and randomness of bid points.

By construction, optimal bid designs provide optimal efficiency under the ideal situation that the underlying true distribution and parameters are known. Optimal design consisting usually of one, two or three design points, however, depends too seriously on the knowledge about true information that is in fact to be estimated. Unknown true parameter values and uncontrollable response rate of the survey make it difficult to employ optimal designs in the study.

In contrast, robust or ad hoc designs in the actual studies reduce the risk from their reliance on initial information by dispersing optimal design points into more points. Uniform design goes further by randomizing all design points. Analytics and simulations show that uniform bid design provides higher efficiency than the robust designs under ideal conditions, and outperforms the optimal design with poor initial information. In simulation results, uniform design also outperforms other designs under the A- and C-optimalities. Ultimately, uniform design reduces the dependence of optimal designs on design structure and poor information.

It is easy for researcher to implement uniform design in any specific application. Since a design independent of the poor initial information is unavailable, the uniform bid design offers a practical and robust alternative to existing bid designs for researchers facing strict budget constraints, or performing a pre-survey to gather better information for the next stage. Uniform design provides binary data continuously sorted by bid value, enabling the researcher to apply more flexible non- and semi-parametric estimation techniques (Lewbel et al. 2003). Although we focus on design problem in dichotomous choice contingent valuation, the adjustments of the design for other studies are straight forward and in most cases simply notational.

ESSAY 3

GENERALIZED ESTIMATION METHODS AND IMPLICATION OF THE RESULT IN DICHOTOMOUS CHOICE CONTINGENT VALUATION MODEL

ABSTRACT

This essay challenges the theoretical and technical background of the simple logit model often used for estimating willingness to pay from dichotomous choice contingent valuation. The simple logit model assumes that the respondent's evaluations of the two states are stochastically independent and homoskedastic. Relaxing restrictive assumptions suggests a generalized estimation technique that utilizes a Gumbel mixed 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. Estimation results show that correlation between two states is usually minimal, but homoskedastic errors are rejected in many cases, i.e. logistic distribution for the difference of error terms, may not be a suitable distribution. Heteroskedasticity or correlation provides willingness to pay estimate different from estimate of the simple logit, thus different policy implication in benefit-cost analysis.

3.1 INTRODUCTION

Dichotomous choice contingent valuation (CV) has been most widely used in eliciting welfare measures (willingness to pay) from environmental projects, thus enabling benefit-cost analysis. Value elicitation questions ask respondent to show his or her utility or willingness to pay in the binary choice setting. Given a specific cost, a subject's binary response will be one if the utility after environmental change is still greater than that of the current state, and zero otherwise. Equivalently, a binary response is one if the willingness to pay is greater than the cost offered, and zero otherwise.

The decision models consistent with economic theory are, among others, the random utility model and the willingness to pay model. The random utility and willingness to pay function consist of a systematic part observable to the researcher and an unobservable error component. With appropriate assumptions about the distribution of the unobserved term, the random utility and willingness to pay models can be simply estimated by logit or probit. For instance, the standard additive random utility model assumes a constant variance between the initial and the proposed states. Using *i.i.d.* type I extreme value (or normal) distribution for each state, the standard additive random utility is estimated through a simple logit (or probit) model. Nice properties and theoretical backgrounds of those models have helped researchers to easily conduct the task of estimation and to focus on other valuable issues.

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 model. The main problem arising against the advantage in estimation is the possibility of losing the realism of actual choice situation. Suppose that we want to estimate the welfare change from enhancing environmental quality. First, the state after environmental change is uncertain to the respondent although the environmental quality is surely increased, which illustrates the possibility that the additive error term in the proposed state may be different from that in the current random utility in terms of variance. Second, if the environmental project is a debated issue in the relevant population, there could be several alternatives that respondents may prefer but the researcher does not consider in the CV survey. The unknown (to the researcher) alternatives can lead the respondent to refuse the proposed project although respondent agrees with the change in environmental quality. Consequently, the simple logit or probit may not be suitable in some situations of decision and yield 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 in CV has paid attention to the strict assumption of identical error distributions across alternatives in the choice set. CV studies have usually assumed and tested heteroskedasticity only across individuals or different groups. Unfortunately, generalized logit models such as nested logit or paired combinatorial logit are not applicable to the contingent valuation since the choice set in CV consists of only two alternatives, the case which reduces some of generalized logit models to the simple logit.

Therefore, this essay relaxes the constant disturbance assumption of the random utility and willingness to pay models in standard contingent valuation after reviewing the value elicitation questions of CV studies. Additive error terms in initial and proposed states can be independent and identical, independent but not identical, dependent but identical, or dependent and not identical. For all possible relations of error terms, a generalized estimation model is suggested by utilizing Gumbel mixed bivariate extreme value distribution (Gumbel 1960, 1961, Gumbel and Mustafi 1967, Tiago de Oliveira 1980, 1983). The generalized model, named as a bivariate extreme value model, covers a heteroskedastic logit (Bhat 1995, Allenby and Ginter 1995), correlated alternatives case and the simple logit. This essay also introduces a mixed logit model with extreme value distributions as an alternative model to cover all specific cases. In addition, unknown alternative case is directly estimated by assuming that different policies for the same target of environmental quality have constant effect on random utility or expenditure.

The generalized estimation models (bivariate extreme value and mixed logit models) show interesting results under various constraints when they are applied to several existing CV data. Error terms of two states are, in most cases, independent but not identical. The extremely different scale factor may imply that the extreme value distribution, i.e. logistic distribution for the difference of two random utilities or expenditures, is not suitable distribution although it provides similar result to probit model. More importantly, heteroskedasticity or correlation provides welfare measures (willingness to pay) different from the estimate of the simple logit. Although parameter estimates are not much different in magnitude, the expected willingness to pay of generalized estimation model can draw opposite conclusion in benefit-cost analysis of environmental project.

3.2 CHOICE MODELS AND WELFARE MEASURE IN CONTINGENT VALUATION

3.2.1 Environmental Issues and Choice Scenarios

Dichotomous choice contingent valuation (CV) study addresses a binary choice question to respondent with randomly assigned cost, i.e. to vote for and to vote against, or to accept and to reject[29]. Alternatives in the choice set consist of the proposed state representing to accept the policy and the current state without change indicating to reject the policy. The following example shows the environmental issues and choice sets in previous CV study.

> Example: In 1994, Carson et al. estimated the welfare measure from the environmental damage due to the deposition of PCB and DDT on the ocean floor off the coast of Los Angeles through several outfall pipes. Chemical sediment does not harm humans but endangers some species of fish. After explaining the problem extensively with instruments including maps and cards, the survey asked a binary choice question about a speed-up program to recover two species of fish earlier than natural processes.

[29] To be uncertain or unsure is also a recommended option in addition to yes and no options. This essay 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.

The payment vehicle for the randomly assigned cost was one time additional amount on the state income tax.

If respondents have no alternative except the speed-up program to recover fast species of fish, it may be reasonable to assume that utilities in the proposed and current states are independent and identical. Alternatively, choice set in the example can represents different state of nature. Uncertainty in the future, reliability on the implementation and result of the project, etc, may cause the difference between the distribution of utilities in the proposed and current states. The proposed state is random and has unobservable part even from the perspective of respondent itself. More uncertainty in the proposed state introduces larger variance of the distribution.

The example shows not only the possibility of heteroskedasticity but also the potential misspecification of binary choice model. If recovering endangered species of fish is a serious issue to residents in Los Angeles, respondents may consider other options to speed up recovering them that may be unknown to researcher, rather than speed-up program proposed in the survey[30]. Consequently, the response of reject in the contingent valuation survey, by nature, represents either staying without change or changing through other process (or possibly in different level). By this reason, the current state is named as the reference state against the proposed state to avoid misinterpretation. Although

[30] 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.

unknown alternative is not always the case, the possibility of existence increases when the project is suggested for the local and debated environmental issues.

Unknown alternative introduces at least two possible cases. First, although distributions of the random utility with alternatives are independent and identical, the variance of the reference state can be greater than that of the proposed state since the reference state includes stochastic component of random utility with unknown alternative. Second, since the unknown alternative may be a competing process for the same goal of the environmental project proposed in the survey, the error term of the random utility with unknown alternative can be correlated with that of the utility at the proposed state.

3.2.2 Choice Probability of Random Utility Model

Hanemann (1984) introduced the theory-consistent random utility model into the dichotomous choice contingent valuation using the framework originally developed by McFadden (See Haab and McConnell 2002). Given two alternatives (accept or reject) in CV, respondent chooses the alternative providing maximum utility under the relevant constraints. The resulting indirect utility function is well defined by a random utility function. Since the conventional random utility is assumed random to the investigator, a standard random utility consists of two parts; a systematic component observable to researcher and an error component which may be known to respondent but not necessarily.

Let the random utility of individual *n* at the state *i* be

$$U_{in} = V_i\left(I_n, z_n\right) + \varepsilon_{in}$$

where V_i is the systematic component, ε_i is the error component, the subscript $i = 0$ represents the reference state and $i = 1$ represents the proposed state. The systematic part is a function of the respondent's income (I_n) and vector of respondent's characteristics and choice attributes (z_n). The income at the proposed state is the amount that is detracted by the assigned cost, b_n . Karlström (1999) summarizes assumptions defining the standard additive random utility model for a discrete choice case.

Definition: A discrete choice random utility model that satisfies the following assumptions is a standard additive random utility model;

A1. weak complementarity, i.e. only own prices and qualities affect the conditional utility associated with alternative *i*,

A2. additive disturbances,

A3. identical distribution in the initial state and the proposed state, and

A4. a finite amount of money for restoring utilities for any finite change.

Note that *i.i.d.* assumption is imposed not only across individual but also between states.

The probability of choosing the proposed state is the probability that the random utility in the state one is greater than that of the state zero;

$$P_{1n} = P\left(U_{0n} < U_{1n}\right) = P\left(V_{0n} + \varepsilon_{0n} < V_{1n} + \varepsilon_{1n}\right) = P\left(\varepsilon_{0n} < v_n + \varepsilon_{1n}\right) \quad (3.1)$$

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 (3.1). The systematic component is usually assumed linear in parameters although only linearity in income is sufficient. In addition, a typical estimation model of the random utility assumes the *i.i.d.* error distribution in equation (3.1) such as *i.i.d.* type I extreme value (or Gumbel) distribution or normal distribution, resulting the choice probability to be a logistic or normal distribution. The derivation of the logistic distribution from difference of two identical extreme values is straightforward. In addition to the relation of extreme value distribution to the logit formula, McFadden (1974) also shows the analysis that the logit formula for the choice probabilities implies extreme value distribution for the random utility.

The choice probability in equation (3.1) also can be expressed using the mixed logit model that is initially applied into recreation model by Train (1998, 1999). A mixed logit can be derived from a random coefficients model (RCM) or error-component model. Let the true random utility to be $U_{in} = \varsigma'_{in} z_{in}$, where $z'_{in} = (x'_{in}, d_i)$ and $\varsigma'_{in} = (\beta'_{in}, \varepsilon_{in})$. Respondent will accept the proposed policy when $U_{1n} > U_{0n} = \varsigma'_{1n} z_{1n} > \varsigma'_{0n} z_{0n}$. 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

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

where $f(\varsigma)$ is a joint density of β_{jn} and ε_{jn} [31]. Note that the innocuous scale factor *s* does not affect the choice probability. Rescaling procedure is solely used for attaining the

[31] 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

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

approximation of the true model or for the degree of smoothing (as in a smoothed AR simulator). The mixed logit model approximates any random utility model to any degree of accuracy (Train 2003, McFadden and Train 2000).

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

$$P_{1n} = \int L_{1n}(\varsigma) f(\varepsilon_0, \varepsilon_1) d\langle \varepsilon_0, \varepsilon_1 \rangle \tag{3.2}$$

where

$$L_{1n}(\varsigma) = \frac{\exp\left[v_n / s + (\varepsilon_{1n} - \varepsilon_{0n}) / s\right]}{1 + \exp\left[v_n / s + (\varepsilon_{1n} - \varepsilon_{0n}) / s\right]}. \tag{3.3}$$

The choice probability of the mixed logit model is exactly same to the logit-smoothed AR simulator with two alternatives, the model which has been suggested by McFadden (1989). Ben-Akiva and Bolduc (1996) named the model by 'logit-kernel probit' applying to the probit. In fact, the logit-smoothed AR simulator can be applied to any choice model by assuming appropriate distribution about error terms (Train 2003).

The mixed logit model (3.2) is equivalent to the choice probability (3.1) when we assume the same error distribution. Either of choice probability in equation (3.1) or (3.2) is estimated by maximizing the likelihood function,

---

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 logit probability in the integral is derived conditional on β. 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.

$$\log L = \sum_{n=1}^{N} y_n \log P_{1n} + \left(1 - y_n\right) \log\left(1 - P_{1n}\right). \tag{3.4}$$

If *i.i.d.* assumption is violated, a simple logistic distribution cannot be applied to (3.4). The choice probability of (3.1) and mixed logit model of (3.2), however, allow the flexible error distribution for estimation. Two differences from previous models deserve to be noted. While previous literatures have considered heteroskedasticity only across individuals or group, equation (3.1) can estimate the model with heteroskedasticity across alternatives, including simple logit model as a special case. The mixed logit probability has been applied for random parameters of systematic component in the multinomial case such as mode choice in transportation or site choice in recreation. By allowing randomness in the error term like equation (3.2), contingent valuation study can get the benefit of flexible mixed logit model. However, note that the correlation between different states arises in the estimation model rather than in the behavioral model. If choice set is well defined and the random utility is specified well enough to capture all sources of correlation among alternative explicitly, the simple logit model will provide consistent estimate of random utility difference and welfare measure.

3.2.3 Welfare Measure

In welfare measure, three definitions of the hicksian variation induced from environmental change are proposed (Karlström 1999);

D1. the expected amount of money to keep the random utility constant for each individual, D2. the expected amount of money to keep utility at the expected utility for each individual, and

D3. the (deterministic) amount of money to keep the expected utility constant.

A series of papers has investigated the correct welfare measure consistent with the microeconomic theory. However, 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 although the welfare estimates from different model are similar (Hau 1986, Herriges and Kling 1999, Karlström 1999).

In spite of the importance of investigating different welfare estimate from different definition, this essay adapts the conventional definition of willingness to pay to calculate the welfare change in the random utility. The expected willingness to pay for the environmental change is defined as the expected maximum income that equates the expected random utility in two states. Although individual is assumed to have a deterministic utility known at the time of decision, at least the utility level of the proposed state is stochastic not only to researcher but also to the respondent due to the nature of the CV scenario. If the alternative at the state zero represents the reference utility including all other possibilities, then the reference state is also stochastic to the respondent.

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, to derive the welfare change for the representative individual. The willingness to pay that equates the expected random utilities in both states is

$$WTP_n = E[U_{1n}] - E[U_{0n}] = \frac{1}{\alpha}\{v_n + E(\varepsilon_{1n}) - E(\varepsilon_{0n})\}$$

where the income variable is not included in v_n [32]. Note that the expectation is conditional expectation. By taking an unconditional expectation to the willingness to pay, the welfare measure of individual *n* owing to the environmental change can be expressed as

$$E(WTP_n) = \frac{1}{\alpha}v_n + \frac{1}{\alpha}E(\varepsilon_{1n} - \varepsilon_{0n}) \qquad (3.5)$$

In previous literatures using symmetric distributions such as logistic or normal, the expected mean of error terms is zero by including a constant term in the systematic component. However, as explained in the next section, the expected value of error terms is not zero but it is much easier to remain the expectation term in equation (3.5) if asymmetric distributions are employed.

3.3 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

---

[32] A typical specification of the systematic component in the random utility assumes a linear function as $V_{in} = x_n'\beta_i + \alpha I_n$, where I_n is the income of individual *n*. Let the systematic utility of the reference state be $V_{0n} = x_n'\beta_0 + \alpha I_n$, and $V_{1n} = x_n'\beta_1 + \alpha(I_n - b_n)$ be for the proposed state, where b_n is bid value offered to individual *n*. Then, the utility difference in the logistic distribution is $V_{1n} - V_{0n} = x_n'(\beta_1 - \beta_0) - \alpha b_n$.

distributions including the Gumbel mixed model which is one of differentiable bivariate extreme value distributions (for 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).

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 probability density function and the cumulative distribution function of Gumbel margin are, respectively,

$$f(\varepsilon_i) = \frac{1}{\theta_i} \exp\left(-\frac{\varepsilon_i}{\theta_i}\right) \exp\left(-\exp\left(-\frac{\varepsilon_i}{\theta_i}\right)\right) \tag{3.6}$$

and

$$F_{\varepsilon_i}(z) = \int_{\varepsilon_i=-\infty}^{z} f(\varepsilon_i)\, d\varepsilon_i = \exp\left(-\exp\left(-\frac{z}{\theta_i}\right)\right). \tag{3.7}$$

The expected value and the variance of ε_i are

$$E(\varepsilon_i) \approx 0.57722\theta_i \text{ and } Var(\varepsilon_i) = \frac{\theta_i^2 \pi^2}{6}.$$

The asymptotic distribution of bivariate maxima is defined as

$$F(\varepsilon_0, \varepsilon_1) = \left[F(\varepsilon_0) F(\varepsilon_1)\right]^{k(\tau)} = \exp\left[-\left\{\exp\left(-\frac{\varepsilon_0}{\theta_0}\right) + \exp\left(-\frac{\varepsilon_1}{\theta_1}\right)\right\} k(\tau)\right]$$

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

Different bivariate distributions are derived using different dependence functions, of which the Gumbel mixed model has

$$k(\tau \mid \lambda) = 1 - \frac{\lambda \exp(\tau)}{\left(1 + \exp(\tau)\right)^2} \tag{3.8}$$

where λ is an association parameter[33]. The parameter λ indicates the association between the two extremes. By plugging (3.8) into the asymptotic distribution, the Gumbel mixed logit model becomes

$$F(\varepsilon_0, \varepsilon_1 \mid \Gamma) = \exp\left[-\left(\exp\left(-\frac{\varepsilon_0}{\theta_0} \right) + \exp\left(-\frac{\varepsilon_1}{\theta_1} \right) \right) + \frac{\lambda}{\exp(\varepsilon_0 / \theta_0) + \exp(\varepsilon_1 / \theta_1)} \right] \tag{3.9}$$

where Γ is a parameter set of scale factor (θ_0, θ_1) and association factor (λ). Figure 3.1 shows the contour of the Gumbel mixed bivariate distribution function with $\lambda = 0.5$. The probability density function is derived by differentiating (3.9) with respect to ε_0 and ε_1 such that

$$f(x,y) = \frac{F(x,y)}{\theta_1 \theta_2} \left[\frac{2\lambda e^{x/\theta_1 + y/\theta_2}}{\left(e^{x/\theta_1} + e^{y/\theta_2}\right)^3} + \left\{ e^{-x/\theta_1} - \frac{\lambda e^{x/\theta_1}}{\left(e^{x/\theta_1} + e^{y/\theta_2}\right)^2} \right\} \left\{ e^{-y/\theta_2} - \frac{\lambda e^{y/\theta_2}}{\left(e^{x/\theta_1} + e^{y/\theta_2}\right)^2} \right\} \right]. \tag{3.10}$$

The contour of probability function is shown in Figure 3.2. As can be seen, the bivariate extreme value distribution is upper-right skewed.

---

[33] The logistic model, one of differentiable bivariate extreme value distribution, is derived using the difference function of $k(\tau \mid \lambda) = \left[1 + \exp(-\tau/(1-\lambda))\right]^{1-\lambda} / \left[1 + \exp(-\tau)\right]$. The logistic model is the simple version of the generalized extreme value distribution widely used in the transportation and recreational site choice literatures. Unfortunately, the logit model, i.e. the generalized extreme value with two alternatives, cannot identify the association factor λ (See the Appendix C).

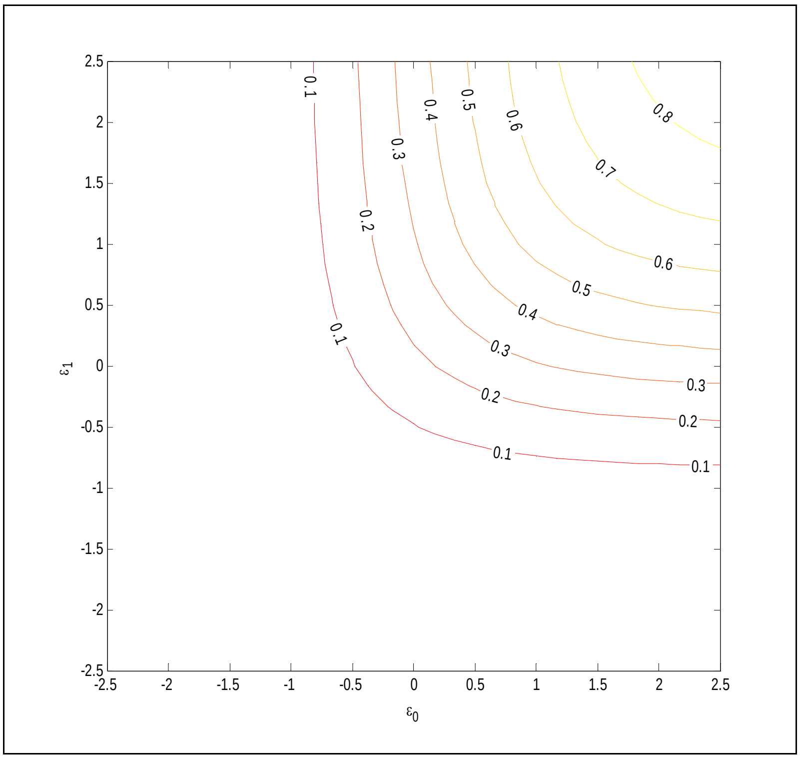

Figure 3.1: Distribution Function of Gumbel Mixed Model of Maxima with $\lambda = 0.5$

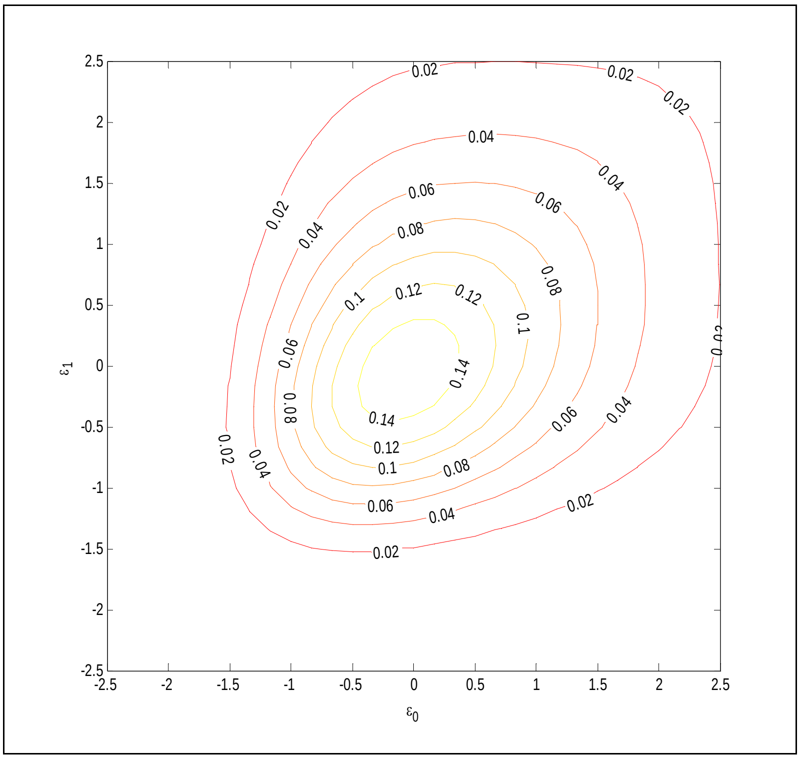

Figure 3.2: Probability Function of Gumbel Mixed Model of Maxima with $\lambda = 0.5$

For $\lambda = 0$, the joint distribution is independent such that $F(\varepsilon_0, \varepsilon_1) = F(\varepsilon_0) F(\varepsilon_1)$, and generally, the inequality $F(\varepsilon_0, \varepsilon_1) > F(\varepsilon_0) F(\varepsilon_1)$ holds for dependent case $\lambda > 0$. The correlation coefficient is a function of the association parameter λ;

$$\rho(\lambda) = \frac{6}{\pi^2}\left[\arccos\left(1 - \frac{\lambda}{2}\right)\right]^2 \quad (0 \le \rho \le 2/3).$$

When the correlation coefficient is greater than 2/3, the mixed model can not be used. If both random variables are available, the association parameter is estimated from the correlation coefficient such that $\hat{\lambda} = 2\left(1 - \cos\left(\pi \cdot \sqrt{\hat{\rho}/6}\right)\right)$ where $\hat{\rho}$ is the estimated correlation from data. Since binary data cannot provide the correlation coefficient, the association parameter is directly estimated from data.

From the Gumbel mixed distribution, several important distributions are derived; among others conditional distribution and distribution of reduced difference. The conditional cumulative distribution function of the Gumbel mixed model is

$$F_{\varepsilon_0|\varepsilon_1}(\varepsilon_0) = F(\varepsilon_0, \varepsilon_1)\left\{\exp\left[\exp\left(-\frac{\varepsilon_1}{\theta_1}\right)\right] - \lambda \frac{\exp\left[2\frac{\varepsilon_1}{\theta_1} + \exp\left(-\frac{\varepsilon_1}{\theta_1}\right)\right]}{\left[\exp\left(\frac{\varepsilon_0}{\theta_0}\right) + \exp\left(\frac{\varepsilon_1}{\theta_1}\right)\right]^2}\right\} \tag{3.11}$$

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 to be (Tiago de Oliveira 1980)

$$D(\tau \mid \lambda) = \frac{\exp(\tau)}{1 + \exp(\tau)} \frac{(1 + \exp(\tau))^2 - \lambda}{(1 + \exp(\tau))^2 - \lambda \exp(\tau)}. \tag{3.12}$$

For $\lambda = 0$, i.e. independent case, the conditional distribution (3.11) reduces to be a univariate type I extreme value and the difference distribution (3.12) becomes a logistic distribution. Since the argument in the difference distribution function is a reduced difference, equation (3.12) can be applied to the estimation model only if utilities have the same variance. The probability density function of (3.12) is

$$\zeta(\tau) = \frac{\exp(\tau)}{\left[1+\exp(\tau)\right]^2}\left\{\frac{\left(1+\exp(\tau)\right)^4 - \lambda^2 \exp(2\tau) - \lambda\left(1-\exp(2\tau)\right)^2}{\left[\left(1+\exp(\tau)\right)^2 - \lambda \exp(\tau)\right]^2}\right\}.$$

Figure 3.3 and 3.4 show the cumulative distribution and probability function of the reduced difference with various λ. The probability density function of reduced difference is symmetric, $\zeta(\tau) = \zeta(-\tau)$ with zero mean. Note that the independence condition reduces the difference distribution function to be a logistic function.

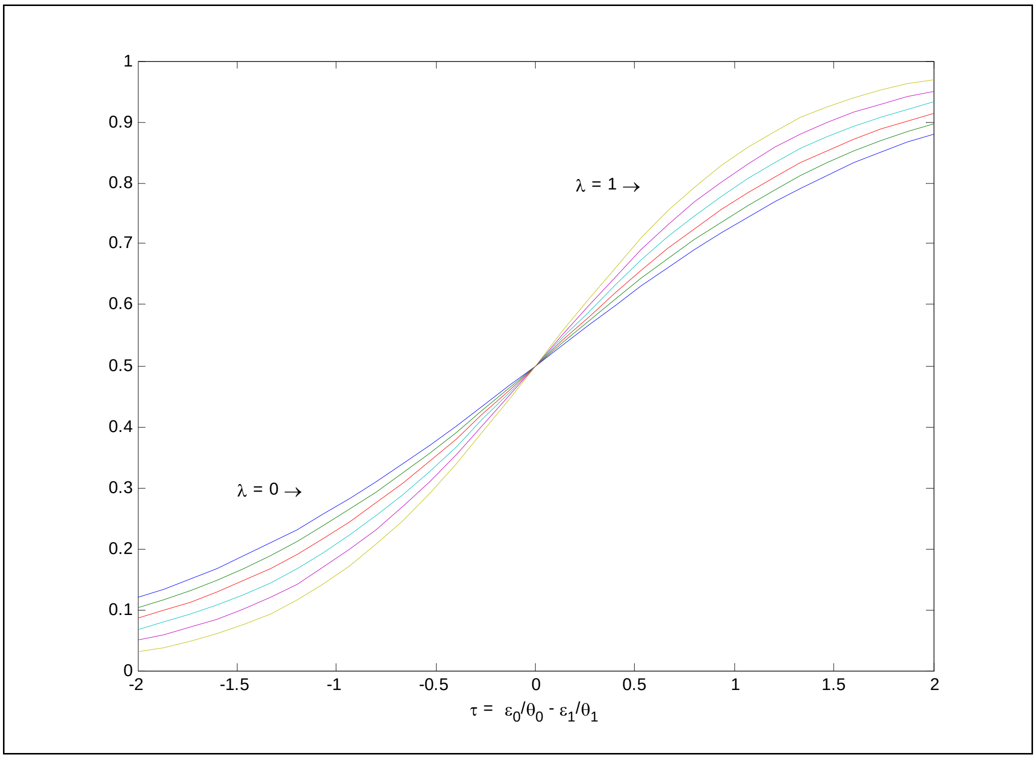

Figure 3.3: Distribution Function of Reduced Difference of Extreme Values

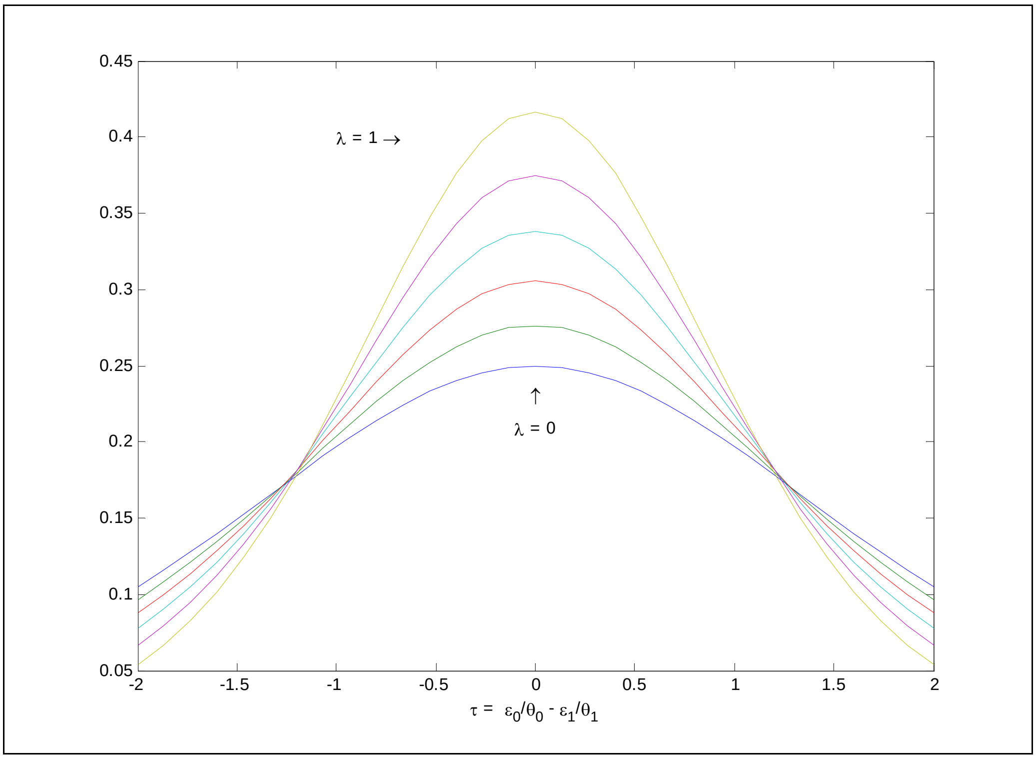

Figure 3.4: Probability Function of Reduced Difference of Extreme Values

3.4 ESTIMATION OF RANDOM UTILITIES WITH BINARY DATA

3.4.1 Bivariate Extreme Value Model

The choice probability in the equation (3.1), $P_{1n} = P\left(\varepsilon_0 < v_n + \varepsilon_1\right)$, can be expressed as an integration of the conditional distribution over marginal distribution. From equations (3.11) and (3.6), the choice probability becomes

$$P_{1n} = \int_{\varepsilon_1=-\infty}^{\varepsilon_1=+\infty} F_{\varepsilon_0|\varepsilon_1}\left(v_n + \varepsilon_1\right) f\left(\varepsilon_1\right) d\varepsilon_1 . \qquad (3.13)$$

Since the model estimates only the difference of two random utilities, one of them should be normalized such that parameters in systematic component of utility at the reference state are set to be zero and the scale factor of error term to be one ($\theta_0 = 1$). Equivalently, parameters in systematic component are estimated as difference of functions normalized by the standard deviation of the reference state and the scale factor of the proposed state is estimated as a relative scale term.

The bivariate extreme value model of (3.13), unfortunately, does not have the closed form for the integration, thus requiring approximation or simulation techniques for estimation of the choice probability. Gaussian quadrature approximates the choice probability with high accuracy and implements estimation with fast speed[34]. A simulation method also provides intuitively easy way for estimating the choice probability. Both of methods are explained in detail at the section IV. Let the choice probability of choosing proposed state be $\hat{P}_{1n}$ from either of approximation or simulation method. The probability

---

[34] In heteroskedastic extreme values, Bhat (1995) shows that Gaussian quadrature generates a highly accurate estimates for integrals rather than simulation. Alternatively, Allenby and Ginter (1995) also suggest the Bayesian estimation procedure for heteroskedastic extreme values.

of choosing the reference state is the complementary probability of P_{1n} such that $\hat{P}_{0n} = 1 - \hat{P}_{1n}$. Therefore, the log likelihood function simply becomes

$$\log L = \sum_{n=1}^{N} y_n \log \hat{P}_{1n} + (1 - y_n) \log \left(1 - \hat{P}_{1n}\right). \tag{3.14}$$

The willingness to pay is estimated using equation (3.5). Following assumptions of a linear random utility and constant marginal utility of income, the expected willingness to pay is

$$E(WTP_n) = x_n' \frac{\tilde{\beta}}{\tilde{\alpha}} + \frac{1}{\alpha} E(\varepsilon_{1n} - \varepsilon_{0n})$$

where $\tilde{\beta} = (\beta_1 - \beta_0)/\theta_0$ and $\tilde{\alpha} = \alpha / \theta_0$. The expectation of $\varepsilon_{1n} - \varepsilon_{0n}$ is not easy to calculate because the mean of error terms are not zero and they are correlated. Rather than deriving the expected value of error terms, this essay estimates the expectation of error differences through a simulation procedure with estimated relative scale and association parameters.

Correlated extreme values

Suppose that the error components are homoskedastic but correlated. Restriction of identical variance on Gumbel mixed model transforms bivariate distribution into a reduced difference distribution in equation (3.12). By substituting the equation (3.12) into the decision model, the probability of choosing the proposed state becomes

$$P_{1n} = \frac{\exp(v_n/\theta)}{1 + \exp(v_n/\theta)} \frac{\left[1 + \exp(v_n/\theta)\right]^2 - \lambda}{\left[1 + \exp(v_n/\theta)\right]^2 - \lambda \exp(v_n/\theta)} \tag{3.15}$$

where θ is the common scale factor. The log likelihood function for the mixed model with constant variance is obtained by substituting equation (3.15) into the log likelihood (3.4). Since the mean of the difference is zero, the willingness to pay is, unsurprisingly, the same with the simple formula in the logit model; $E\left(WTP_n\right) = x_n'\tilde{\beta} / \tilde{\alpha}$.

Heteroskedastic extreme values

The choice probability with Gumbel mixed model nests the heteroskedastic extreme values and simple logit models which assume independent error terms. When two error terms are independent ($\lambda = 0$), the conditional distribution (3.11) becomes a marginal extreme value distribution; $F_{\varepsilon_0|\varepsilon_1}\left(\varepsilon_0\right) = F\left(\varepsilon_0\right)$. Therefore, from the equation (3.13), the choice probability with heteroskedasticity can be simplified to be

$$P_{1n} = \int_{\varepsilon_1=-\infty}^{\varepsilon_1=+\infty} F\left(v_n + \varepsilon_1\right) f\left(\varepsilon_1\right) d\varepsilon_1 . \tag{3.16}$$

Substitute equations (3.6) and (3.7) into (3.16) and define $w = \varepsilon_1 / \theta_1$ and $\gamma = \theta_1 / \theta_0$, then the probability function becomes

$$P_{1n} = \int_{w=-\infty}^{w=+\infty} \exp\left[-\exp\left(-v_n / \theta_0 + \gamma w\right)\right]\left[\exp\left(-w\right)\right]\exp\left[-\exp\left(-w\right)\right] dw . \tag{3.17}$$

The equation (3.17) is the simplest version of the heteroskedastic extreme values model of Bhat (1995) and Allenby and Ginter (1995). In the independent but non-identical extreme value distribution, the heteroskedasticity extreme value model outperforms the multinomial logit and other generalized logit models (Bhat 1995). The equation (3.17), however, does not have the closed form of solution for the integral, requiring us to implement the approximation or simulation of the choice probability.

Note that $J - 1$ scale parameters are identified due to normalization in the heteroskedastic extreme values where J is the total number of alternatives, while other generalized logit models identify only $J(J-1)/2-1$ scale parameters[35]. Since the contingent valuation study has only two alternatives, the choice probabilities of (3.13) and (3.17) can identify the relative scale and association parameters, but other generalized models have the identification problem. The scale parameter is estimated as a relative scale, γ. More uncertainty in the future, i.e. more variance of random utility, is a reasonable nature in decision process, implying that γ is possibly greater than one but not necessarily.

As shown in the previous section, when bivariate extreme value distribution is independent, the expectation of a random variable is $E(\varepsilon_i) \approx 0.57722\theta_i$. Thus, 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

$$E(WTP_n) \approx x_n' \frac{\tilde{\beta}}{\tilde{\alpha}} + 0.57722 \frac{1}{\tilde{\alpha}}(\gamma - 1).$$

***i.i.d* extreme values**

Now assume identical disturbance of random utility in both states in addition to independence. By substituting the distribution (3.7) and density (3.6) functions into the choice probability (3.16), the choice probability with *i.i.d* error distributions becomes

$$P_{1n} = \int_{s=-\infty}^{\infty} \frac{1}{\theta} \exp\left(-\exp\left\{-\frac{1}{\theta}(v_n + \varepsilon_1)\right\}\right) \exp\left(-\frac{\varepsilon_1}{\theta}\right) \exp\left(-\exp\left(-\frac{\varepsilon_1}{\theta}\right)\right) d\varepsilon_1 .$$

---

[35] In general, multinomial probit identifies $J - 2$ free standard deviations and $(J - 1)(J - 2)/2$ free correlations., therefore a total of $J(J - 1)/2 - 1$ covariance parameters (Greene 2002, Train 2003).

Let $t = \exp\left(-\varepsilon_1 / \theta\right)$ and $dt = -\exp\left(-\varepsilon_1 / \theta\right) d\varepsilon_1 / \theta$, then the choice probability can be expressed as a logistic distribution such that

$$P_{1n} = \int_{t=0}^{\infty} \exp\left(-t\left(1+\exp\left\{-v_n / \theta\right\}\right)\right) dt = \left[1+\exp\left\{-v_n / \theta\right\}\right]^{-1}. \qquad (3.18)$$

The logistic distribution can be derived by constraining $\lambda = 0$ in the reduced difference distribution of Gumbel mixed model in equation (3.12). Since P_1 is the standard cumulative density function of the logistic distribution, the variance of $\varepsilon = \theta w$ is $\theta^2 \pi^2 / 3$. Note that when error terms are homoskedastic the choice probability always has a closed form of distribution as in equations (3.15) and (3.18).

The estimation technique for this simple case is straightforward and the estimation result is consistent under the correct model specification. All parameters are identified up to the normalized difference such that $\tilde{\beta} = \left(\beta_1 - \beta_0\right) / \theta$ and $\tilde{\alpha} = \alpha / \theta$, but the scale factor cannot be identified. With linear specification of random utility, the willingness to pay is the difference between systematic terms of the random utility except the income, multiplied by the inverse of the marginal utility of income because the expected value of difference between two identical errors is zero; $E\left(WTP_n\right) = x_n' \tilde{\beta} / \tilde{\alpha}$

3.4.2 Approximating Log Likelihood of Bivariate Extreme Value Model

The bivariate extreme value model in equation (3.14) integrates a conditional distribution over one-dimensional random variable, $\int_{\varepsilon_1=-\infty}^{\varepsilon_1=+\infty} F_{\varepsilon_0|\varepsilon_1}\left(v_n + \varepsilon_1\right) f\left(\varepsilon_1\right) d\varepsilon_1$. As Bhat (1995) already showed that a Gaussian-Laguerre quadrature outperforms a simulation in multinomial heteroskedastic case, Gaussian quadrature can provide a fast and highly accurate likelihood function even in this bivariate case by appropriately transforming the function. Define a transformation such that $u = \exp(-\exp(-w))$, thus $w = -\ln(-\ln u)$ and $du = \exp(-w)\exp(-\exp(-w))dw$ [36]. The new variable *u* is the form of cumulative distribution of extreme value and has the support of $[0,1]$. This transformation enables the approximation 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 $F_{\varepsilon_0|\varepsilon_1}(v_n + \varepsilon_1) = F_{\varepsilon_0|\varepsilon_1}(v_n + \theta_1 w_1)$ and $f(\varepsilon_1) = f(w_1)/\theta_1$. The arguments in the conditional probability is normalized by the standard deviation of θ_0. Plugging the new variable *u* into the function and defining

$$G(v_n,u) = F_{\varepsilon_0|\varepsilon_1}\left(v_n - \gamma \ln(-\ln u)\right),$$

the choice probability is expressed as

$$P_{1n} = \int_{u=0}^{u=1} G(v_n,u)\,du\ .$$

since $d\varepsilon_1 = \theta_1 dw_1$. The integration is approximated by Gaussian-Legendre quadrature such as

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

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

where ξ_l and u_l are L weights and support points (abscissas) of Gaussian-Legendre quadrature. The points and weights for approximation are reported in Straud and Secrest (1966). Using $L = 40$ which is the maximum points provided in *Gauss 5.0* program, the log likelihood function is approximated as

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

3.4.3 Mixed Logit Model with Gumbel Mixed Extreme Value Distribution

Recall the choice probability of mixed logit model in equation (3.2),

$$P_{1n} = \int L_{1n}\left(\varsigma\right) f\left(\varepsilon_0, \varepsilon_1\right) d\left\langle \varepsilon_0, \varepsilon_1 \right\rangle$$

where the probability density function $f\left(\varepsilon_0, \varepsilon_1\right)$ is joint density function of bivariate extreme values in equation (3.10). By construction, the mixed logit model is equivalent to the choice probability in equation (3.1) and consequently the mixed logit model nests all three simple cases; the correlated extreme values ($\lambda \neq 0$ but $\theta_0 = \theta_1$), the heteroskedastic extreme values ($\lambda = 0$ but $\theta_0 \neq \theta_1$) and simple logit models ($\lambda = 0$ and $\theta_0 = \theta_1$). The mixed model for choice probability (3.2), however, is different from the typical random coefficients model in the sense that the model allows the flexibility only in the original error component, thus captures the heterogeneity or correlation of utility across alternatives in CV.

Owing to equivalence to the logit smoothed-AR simulator, the estimation of the mixed model follows the simulation procedure of 'logit kernel probit' adjusted simply for the bivariate extreme values. Following Train (2003), the simulation procedure is generally: (1) Draw a 2-dimensional random vector of ε from a Gumbel mixed bivariate extreme values. Label the draw as $\varepsilon_n^r = \left\langle \varepsilon_{0n}^r, \varepsilon_{1n}^r \right\rangle$. (2) Using this draw, calculate the utility difference $U_{1n}^r - U_{0n}^r = v_n + \varepsilon_{1n}^r - \varepsilon_{0n}^r$. (3) The logit formula of equation (3.3) in the mixed logit model is calculated from the utility difference and with a scale factor *s* specified by the researcher. (4) Repeat steps (1)-(3) many times ($r = R$), and then the simulated probability is the average of them, $\hat{P}_{1n} = \frac{1}{R}\sum_{r=1}^{R} L_{1n}\left(\varepsilon^r \mid \beta\right)$. The simulated log likelihood function is $\log L = \sum_{n=1}^{N} y_n \log \hat{P}_{1n} + \left(1 - y_n\right)\log\left(1 - \hat{P}_{1n}\right)$. However, since random drawing from a bivariate extreme value distribution in the step (1) is unavailable, this essay employs an importance sampling procedure with Halton sequence to simulate the random draw from bivariate extreme values.

The expected willingness to pay in the mixed logit model can be estimated by the same formula of bivariate extreme value model. When the variance is identical across states, the expected willingness to pay is simply the difference of systematic component divided by the income parameters. When there exists heteroskedasticity, the expected difference of error terms is added into the expected value of the systematic part. In general case, the expected difference of error term is simulated using parameter estimates of relative scale and association. Note that rescale process for deriving logit formula does not change the willingness to pay since the welfare measure is estimated by the ratio of parameter estimates.

3.4.4 Simulating Log Likelihood of Mixed Logit Model

In spite of intuitively simple procedure in simulation, it is not easy to draw random variables ε_0 and ε_1 from the Gumbel mixed distribution. Alternatively, 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. Suppose that there is a density, $g(\varepsilon)$, that can be handled easily. Since multiplying the integrand of equation (3.2) by $g(\varepsilon)/g(\varepsilon)$ does not change the original choice probability, the choice probability of mixed logit model becomes

$$P_{1n} = \int L_{1n}(\varsigma)\frac{f(\varepsilon)}{g(\varepsilon)}g(\varepsilon)d\varepsilon .$$

The choice probability is simulated by random drawing from $g(\varepsilon)$, calculating the logit formula with a weight $f(\varepsilon)/g(\varepsilon)$ for each draw.

Let $g(\varepsilon_i)$ be a univariate extreme value distribution. Using the joint density of Gumbel mixed model given in equation (3.10), the weight $f(\varepsilon)/g(\varepsilon)$ is calculated as

$$\frac{f(\varepsilon_0,\varepsilon_1)}{g(\varepsilon_0)g(\varepsilon_1)} = \Psi(\varepsilon_0,\varepsilon_1)\exp\left\{(\varepsilon_0/\theta_0+\varepsilon_1/\theta_1)+\frac{\lambda}{\exp(\varepsilon_0/\theta_0)+\exp(\varepsilon_1/\theta_1)}\right\} \quad (3.19)$$

where

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

Normalizing $\theta_0 = 1$ and 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 is simulated as

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

Application of importance sampling to the mixed logit model in this essay 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 useful to draw standard extreme values[37]. Using Halton draws for the given sample size, the standard extreme value distribution is recovered from the inverse of cumulative distribution of extreme value. (2) For this draw, calculate the logit formula, L_{1n}, and the weight function of equation (3.19) with prespecified scaling factor in the logit formula (s). (3) Repeat two steps enough times and average 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. Note that by construction, R repetition is equivalent with R Halton draws. The probability of choosing the alternative zero is $\hat{P}_{0n} = 1 - \hat{P}_{1n}$. The simulated log likelihood function becomes

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

[37] 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).

3.4.5 Simulating Expected Willingness to Pay

In both of bivariate extreme value and mixed logit models, the expected difference of error terms is not zero if error variables are heteroskedastic. The expected difference is approximately $0.57722 \cdot (\theta_1 - \theta_0)$. However, it is not easy to take expectation on error difference if error terms are heteroskedastic and correlated. Importance sampling can be easily reapplied to simulate the expected willingness to pay.

Since $E(\varepsilon_{1n} - \varepsilon_{0n}) / \alpha$ is equivalently $E(\gamma w_{1n} - w_{0n}) / \tilde{\alpha}$, 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 applying importance sampling procedure with Halton sequence to the expected error difference, the expected willingness to pay becomes

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

Note that, except the case of heteroskedasticity, the expected error difference can be exactly calculated without relying on the simulation.

3.5 EXPENDITURE DIFFERENCE AND GUMBEL MIXED MODEL

3.5.1 Expenditure Difference Model

An alternative model of the random utility is based on the willingness to pay function derived from the expenditure functions. Expenditure function is a dual function of the indirect utility function. Let the minimum expenditure of individual *n* be

$m_{0n} = m\left(q^0, u_n^0\right)$ at the reference state and $m_{1n} = m\left(q^1, u_n^0\right)$ at the proposed state given the same utility level (u_n^0) where q^i is the environmental quality at state *i*. As a random utility, the expenditure function consists of a systematic component (m_n^*) and an unobservable random component (η_n); $m_n = m_n^* + \eta_n$. The willingness to pay function is defined to be

$$WTP\left(u^0\right) = m\left(q^0, u^0\right) - m\left(q^1, u^0\right).$$

The binary response to the dichotomous choice question is one if the willingness to pay is greater than bid amount, and zero otherwise. Alternatively, the response of accept implies that the respondent agrees to pay the cost when the expenditure at the proposed state plus the bid amount is still less than the expenditure at the reference state. Typically, the willingness to pay function is estimated by first assuming an appropriate distribution for the unobserved component and then applying a form of probit or logit model.

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. However, unlike the random utility model, the expenditure function is derived from a minimization problem implying that extreme value of the expenditure function is the smallest value such as the type I smallest extreme value distribution. As can be recognized, the exactly same problems of random utility model arise in the willingness to pay function model if we assume a logistic distribution. The general form of choice probability in expenditure difference model is

$$
\begin{aligned}
P_{1n} &= P\left(b_n < m_{0n} - m_{1n}\right) \\
&= P\left(\eta_1 < m_{0n}^* - m_{1n}^* - b_n + \eta_0\right)
\end{aligned}.
$$

Error terms η_0 and η_1 may have heteroskedastic variance or be correlated.

3.5.2 Gumbel Mixed Model of Minima

From the dual relation of $\min\left(Z_i\right) = -\max\left(-Z_i\right)$, the joint distribution function for minima with Gumbel reduced margins is

$$
\Omega\left(\eta_0,\eta_1\right) = 1 - F\left(-\eta_0\right) - F\left(-\eta_1\right) + F\left(-\eta_0,-\eta_1\right), \tag{3.20}
$$

where $F\left(\cdot\right)$ and $F\left(\cdot,\cdot\right)$ are marginal and joint distributions of maxima defined in equation (3.7) and (3.9) (Tiago de Oliveira, 1983). From the relation between maxima and minima in equation (3.20), the probability density function of bivariate extreme values of minima is defined as

$$
\omega\left(\eta_0,\eta_1\right) \equiv \frac{\partial^2 \Omega\left(\eta_0,\eta_1\right)}{\partial\eta_0 \partial\eta_1} = \frac{\partial^2 F\left(-\eta_0,-\eta_1\right)}{\partial\eta_0 \partial\eta_1} = f\left(-\eta_0,-\eta_1\right). \tag{3.21}
$$

Figure 3.5 and 3.6 show the joint distribution and probability function of minima, respectively. Note that the probability function of minima is symmetric function of probability of maxima around zero, thus the tail of minima is lower-left skewed.

Since a bivariate distribution of maxima satisfies the boundary conditions $F\left(-\infty,y\right) = F\left(x,-\infty\right) = F\left(-\infty,-\infty\right) = 0$, the bivariate distribution of minima also satisfies the boundary conditions, $\Omega\left(-\infty,\eta_1\right) = \Omega\left(\eta_0,-\infty\right) = 0$. From the definition of marginal distribution of maxima, $F\left(x,\infty\right) = F\left(x\right)$ and $F\left(\infty,y\right) = F\left(y\right)$, the marginal distribution of minima is defined as $\Omega(\eta_0) = \Omega(\eta_0, \infty) = 1 - F(-\eta_0)$ and

$\Omega(\eta_1) = 1 - F(-\eta_1)$ such that

$$\Omega(z) = 1 - \exp\left(-\exp\left(\frac{z}{\theta_i}\right)\right).$$

The probability density function of marginal distribution is also easily derived as

$$\omega(\eta) = \frac{\partial\Omega(\eta)}{\partial\eta} = -\frac{\partial F(-\eta)}{\partial\eta} = f(-\eta). \tag{3.22}$$

Owing to the relation between maxima and minima, the expected value of η is easily derived to be $E(\eta_i) \approx -0.57722\theta_i$.

The conditional distribution of minima is derived from the conditional distribution of maxima in equation (3.11). Since the conditional probability function of minima is expressed as

$$\omega(\eta_1 \mid \eta_0) = \frac{\omega(\eta_0, \eta_1)}{\omega(\eta_0)} = \frac{f(-\eta_0, -\eta_1)}{f(-\eta_0)} = f(-\eta_1 \mid -\eta_0),$$

the conditional distribution of minima is

$$\Omega(\eta_1 \mid \eta_0) = \int_{\eta=-\infty}^{\eta=\eta_1} \omega(\eta \mid \eta_0)\, d\eta = -F(-\eta \mid -\eta_0)\Big|_{\eta=-\infty}^{\eta=\eta_1} = 1 - F(-\eta_1 \mid -\eta_0). \tag{3.23}$$

Note that the distribution and probability functions of reduced difference of minima are identical to that of maxima.

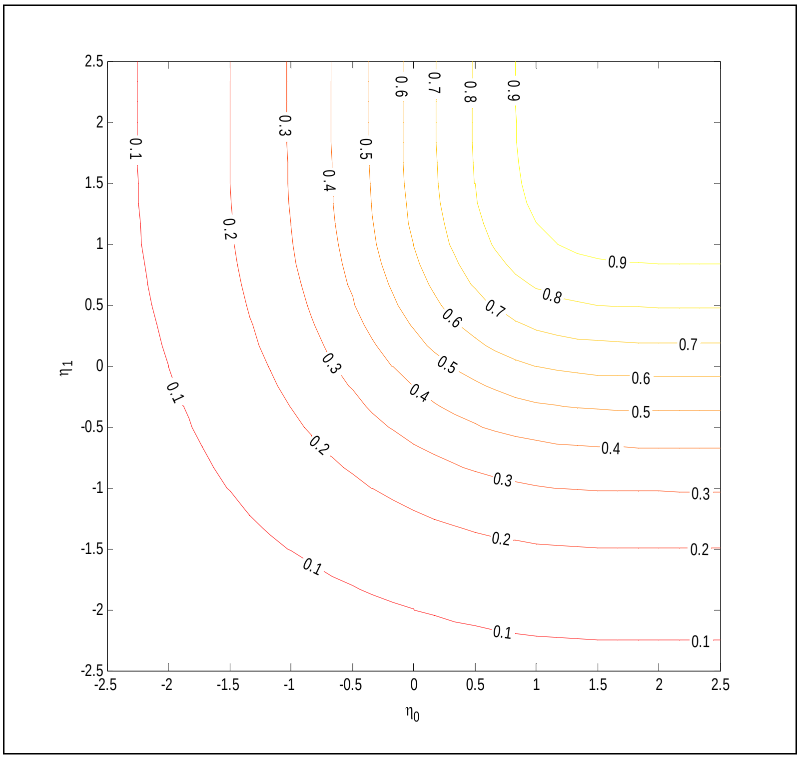

Figure 3.5: Distribution Function of Gumbel Mixed Model of Minima with $\lambda = 0.5$

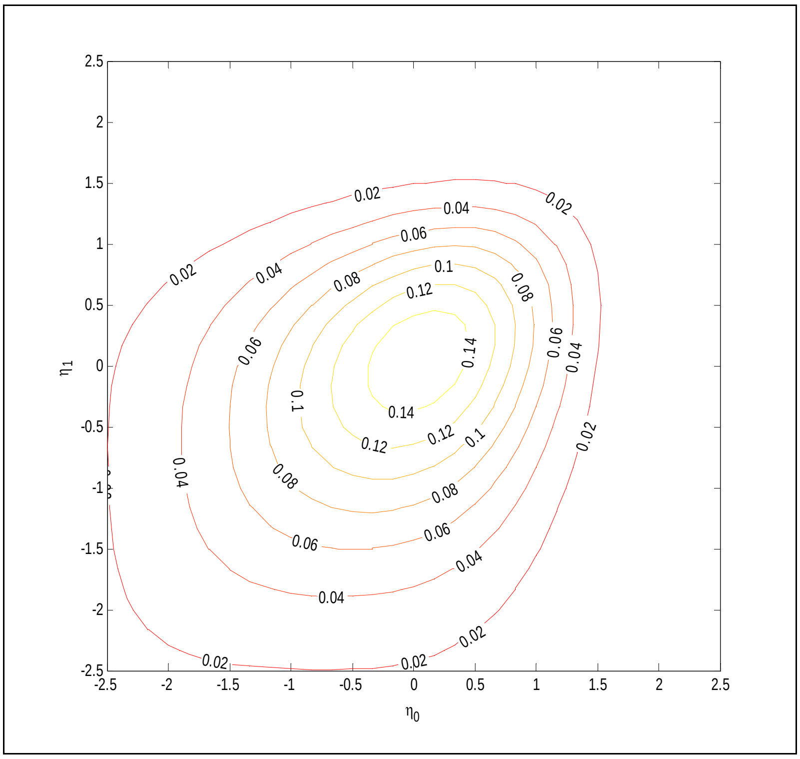

Figure 3.6: Probability Function of Gumbel Mixed Model of Minima with $\lambda = 0.5$

3.5.3 Bivariate Extreme Value and Mixed Logit Models of Expenditure Difference

Note that the respondent accepts to pay the cost when the expenditure plus bid amount at the proposed state is still smaller than the expenditure at the reference state. Following the same logic in random utility model, the choice probability of expenditure difference is expressed as

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

where $m_n^* = m_{0n}^* - m_{1n}^*$. By substituting equation (3.23) into the choice probability function, the choice probability of expenditure difference is

$$P_{1n} = \int_{\eta_0=-\infty}^{\eta_0=+\infty} \left\{1 - F\left(-\left(m_n^* - b_n + \eta_0\right) \mid -\eta_0\right)\right\} f\left(-\eta_0\right) d\eta_0 .$$

The choice probability can be approximated by Gaussian quadrature. Define a transformation such that $1 - u = \exp\left(-\exp\left(w\right)\right)$, thus $w = \ln\left(-\ln\left(1-u\right)\right)$ and $du = \exp\left(w\right)\exp\left(-\exp\left(w\right)\right)dw$. Let $\eta_0 = \theta_0 w_0$ and $\gamma = \theta_0 / \theta_1$, thus $d\eta_0 = \theta_0 dw_0$. Plugging new variables into the function and defining

$$H\left(m_n^*, u\right) = 1 - F\left(-m_n^* - b_n - \gamma \ln\left(-\ln\left(1-u\right)\right) \mid -\eta_0\right),$$

the approximated choice probability is

$$\hat{P}_{1n} = \sum_{l=1}^{L} \xi_l H\left(m_n^*, u_l\right).$$

Expenditure difference model has an estimation model corresponding to each case of random utility model. The choice probability of *i.i.d* error distribution is

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

Recalling the cumulative distribution function and probability density function of the type I smallest extreme value are, respectively, $\Omega(z) = 1 - F(-z)$ and $\omega(\eta) = f(-\eta)$, the probability function with homoskedastic variance is

$$P_{1n} = \int_{\eta_0=-\infty}^{\eta_0=+\infty} \left[1 - F\left(-\frac{m_n^* - b_n + \eta_0}{\theta} \right) \right] f(-\eta_0)\, d\eta_0 . \qquad (3.24)$$

By substituting $t = \exp(\eta_0 / \theta)$ and $dt = \exp(\eta_0 / \theta) / \theta d\eta_0$, the probability function can be further simplified to

$$P_{1n} = \frac{\exp\left[\left(m_n^* - b_n \right) / \theta \right]}{1 + \exp\left[\left(m_n^* - b_n \right) / \theta \right]},$$

which is a simple logit model. In the heteroskedastic case, define again $w = \eta_0 / \theta_0$, then from the equation (3.24), heteroskedastic expenditure model can be expressed such as

$$P_{1n} = \int_{w=-\infty}^{w=+\infty} \left[1 - F\left(-\frac{m_n^* - b_n + \theta_0 w}{\theta_1} \right) \right] f(-w)\, dw .$$

The choice probability of heteroskedastic expenditure is approximated using Gaussian quadrature.

The general expression equivalent to the mixed logit model (3.2) is

$$P_{1n} = \int L_{1n}(\varsigma)\, \omega(\eta)\, d\eta$$

where $L_{1n} = \left\{ 1 + \exp\left[-\left(m_n^* - b_n \right) / s - \left(\eta_{n0} - \eta_{n1} \right) / s \right] \right\}^{-1}$ and $\eta = \langle \eta_0, \eta_1 \rangle$. The bivariate probability of minima is simulated by the same way of bivariate probability of maxima. In importance sampling, the target density is joint density of minima and the proposal densities are standard extreme value densities. From (3.21) and (3.22), importance sampling is implemented by

$$\frac{\omega(\eta_0,\eta_1)}{\omega(\eta_0)\omega(\eta_1)}=\frac{f(-\eta_0,-\eta_1)}{f(-\eta_0)f(-\eta_1)}.$$

The consistent estimate of the parameter in the expenditure function is estimated by ML with approximated or simulated log likelihood function

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

Assume further that the expenditure function is linear in parameters; $m_{in}^{*}=x_n'\beta_i$. Then, as in the random utility model, parameters are identified up to the difference of two expenditure functions. The parameter estimates are $\tilde{\beta}=(\beta_0-\beta_1)/\theta_1$ 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. Note that all parameters are normalized by θ_1 rather than by θ_0. Owing to the bid variable, the expenditure difference model is able to identify both scale parameters. The willingness to pay is defined as

$$E(WTP_n)=m_n^{*}+E(\eta_0-\eta_1)$$

which is estimated through the same simulation method of the random utility model. Since the expected univariate extreme value of minima is $E(\eta_i)\approx-0.57722\theta_i$, the willingness to pay of heteroskedasticity is estimated as follows

$$E(WTP_n)\approx x_n'\frac{\tilde{\beta}}{\tilde{\beta}_b}+0.57722\frac{1}{\tilde{\beta}_b}(\tilde{\gamma}-1).$$

3.6 APPLICATIONS OF RANDOM UTILITIES AND EXPENDITURE DIFFERENCE MODELS

Gumbel mixed distribution of bivariate extreme values has been applied to the random utility model and expenditure difference model. The data used in the estimation includes wastewater disposal system in Montevideo, Uruguay, and the sewage treatment in Barbados (McConnell and Ducci, 1989)[38]. In Barbados study, the households were asked through in-person interview, if they would be willing to pay the given amount of money in increased water bill for the installation of a sewage system. Observations in each data were 1276 for Montevideo and 426 for Barbados data.

Log likelihood function is approximated in the bivariate extreme value model and is simulated in the mixed logit model using Gauss program version 5.0 with CML library. For mixed logit model, the rescaling factor s is set to be 0.3 and the simulation is repeated 125 times. In both of random utility and expenditure difference models, the association parameter (λ) is constrained to be between zero and one. For the relative scale factor (γ), CML procedure in Gauss assigned nonnegative constraint although it is positive in theory[39]. Also, to make the results to be comparable, the inverse of relative scale, θ_1 / θ_0, was estimated in expenditure difference model rather than θ_0 / θ_1.

Tables 3.1 and 3.2 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

[38] The data is available in Haab and McConnell (2002).

[39] 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.

sixth column and the result of mixed logit in the last four columns. The first part of results in bivariate extreme values and mixed logit is estimation result with constraints of independent and identical error ($\gamma = 1$, $\lambda = 0$). These general models with constraints of $\gamma = 1$ and $\lambda = 0$ are theoretically equivalent to the simple logit model. In the following columns are estimation results with heteroskedastic only ($\lambda = 0$), correlation only ($\gamma = 1$), and without constraint for full flexibility.

In Table 3.1, constrained bivariate extreme value model with $\gamma = 1$ and $\lambda = 0$ provides exactly same parameter estimates with the simple logit model. The constrained mixed logit model with $\gamma = 1$ and $\lambda = 0$, however, has estimates different from the simple logit model. Unfortunately, both of estimation models fail to estimate parameters due to too large relative scale estimate when the association parameter is fixed to be one and only heteroskedasticity is allowed. 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 both of bivariate extreme value and mixed logit models. In addition, relative scale parameter is not statistically different from one. LR statistics for homoskedasticity or independence fail to reject the constraints. Barbados data shows that the assumption of independent and identical distribution is suitable for estimation of random utility.

Table 3.2 shows the estimation result of random utility model with Montevideo data. Like the Barbados data, bivariate extreme value model with constraints of $\gamma = 1$ and $\lambda = 0$ shows the estimation result closer to logit model than mixed logit model does. Both of bivariate extreme value and mixed logit models provide association parameter estimate statistically indifferent from zero except one case of mixed logit model. The relative scale estimate at convergence is zero implying that the scale of reference state is extremely larger than that of proposed state. Bivariate model with constraint of $\gamma = 1$ shows that dependence between error terms is not statistically significant. LR statistics fails to reject the constraint of independence ($\lambda = 0$), but heteroskedasticity is statistically significant in both of bivariate extreme value and mixed logit models.

Table 3.3 and Table 3.4 report the estimation results of expenditure difference model with the same data set. 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. By construction, however, parameters of systematic part of expenditure difference are normalized by θ_1 not by θ_0. Thus, the parameter of bid value in expenditure difference model represents the inverse of standard error θ_1 while that of random utility model implies the marginal utility of income normalized by θ_0. As random utility model, the expenditure difference model reports the estimation result with constraints of *i.i.d.* ($\gamma = 1$ and $\lambda = 0$), independence ($\lambda = 0$), and homoskedasticity ($\gamma = 1$).

Table 3.3 presents the estimation result of expenditure difference model with Barbados data. When the relative scale parameter is estimated, the result shows that the inverse of relative scale is greater than one, implying $\theta_1 > \theta_0$, but statistically indifferent from one except the constrained bivariate extreme value model with $\lambda = 0$. The estimate of association parameter is not statistically different from zero. The results of expenditure difference model with Barbados data show that Barbados study satisfies the classical assumption of logit model in terms of parameter estimates and LR test statistics.

Table 3.4 reports the estimation result of expenditure difference model with Montevideo data. Unfortunately, the estimation fails in the general bivariate extreme value model and constrained bivariate with $\lambda = 0$ due to extremely large θ_0. LR test fails to reject the independence constraint in both estimation models. However, LR statistics are 5.48 for homoskedasticity constraint in the mixed logit model, which is significant with 95% confidence. Note that the heteroskedasticity is statistically significant in the corresponding bivariate models of the random utility model (Table 3.2). Consequently, Montevideo data demonstrates statistically insignificant dependence but significant heteroskedasticity. However, although the relative scale estimate is statistically significantly less than one, parameter estimates of systematic component of expenditure difference are seemly equivalent with that of the random utility model. Remind that the expenditure difference model is normalized by θ_1 while the random utility model is normalized by θ_0. When $\gamma = \theta_1 / \theta_0 < 1$, parameter estimates in the expenditure difference model should be greater than that of the random utility model. The parameter estimates in the mixed logit model, however, does not show decreasing tendency from the result of bivariate extreme value model.

Interestingly, the 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. When error components between two states are independent and identical, parameter estimates are almost similar in both of bivariate extreme values and mixed models since two models are different only in interpretation of choice probability but indistinguishable in estimation. Although estimation results present independence of error components in both data, the variance of the reference state of Montevideo data is greater than that of the proposed state.

In spite of similar parameter estimates, the welfare measure from the change of environmental quality varies enormously depending on the relation in error terms. Table 3.5 shows the sample average of the expected willingness to pay from Table 3.1 to Table 3.4[40]. Unfortunately, due to estimation failure, the willingness to pay cannot be estimated in two heteroskedastic models of random utility with Barbados data and in two cases of the bivariate extreme value model for the expenditure difference with Montevideo data. Approximation method (the bivariate extreme value model) generally provides better estimation result than simulation method (the mixed logit model) when the error terms are constrained with $\gamma = 1$ and $\lambda = 0$. Furthermore, expected willingness to pay with homoskedasticity constraint ($\gamma = 1$) is also similar to logit model since independence has been found in most cases. However, the sample average of the expected willingness to pay with heteroskedasticity is quite different from the sample average of the expected willingness to pay under homoskedasticity. 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 $\gamma = 1$.

[40] 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.

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

* Parameter estimates are reported followed by standard error of estimate.

** Function calculation is failed.

Table 3.1: 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	−0.6140 0.1295	−0.6139 0.1295	−0.1297 0.0822	−0.4860 0.1444	−0.1297 0.0823	−0.6985 0.1336	−0.1360 0.0924	−0.6565 0.1617	−0.0994 0.0819
Income	0.0944 0.0192	0.0944 0.0192	0.0743 0.0138	0.0778 0.0228	0.0743 0.0138	0.1108 0.0183	0.0812 0.0146	0.1074 0.0286	0.0708 0.0077
Bid	0.0100 0.0014	0.0100 0.0014	0.0053 0.0008	0.0078 0.0022	0.0053 0.0008	0.0104 0.0015	0.0063 0.0009	0.0096 0.0017	0.0063 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 0.4445	0.1909 0.0000	0.0000 –	0.0000 –	0.2277 0.2478	1.0000 .
Log likelihood	-726.719	-726.720	-723.611	-726.244	-723.611	-726.174	-723.625	-725.782	-723.432

* Parameter estimates are reported followed by standard error of estimate.

Table 3.2: 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	0.6789	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

[*] Parameter estimates are reported followed by standard error of estimate.

Table 3.3: 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	-0.6140 0.1295	-0.6140 0.1295		-0.4860 0.1444		-0.6757 0.1392	-0.1988 0.1462	-0.6238 0.1312	-0.1988 0.1462
Income	0.0944 0.0192	0.0944 0.0192		0.0778 0.0228		0.1064 0.0214	0.0824 0.0153	0.1110 0.0222	0.0824 0.0153
Bid	0.0100 0.0014	0.0100 0.0014		0.0078 0.0022		0.0104 0.0015	0.0064 0.0009	0.0082 0.0016	0.0064 0.0009
γ	-	1.0000 -		1.0000 -		1.0000 -	0.1009 0.1913	1.0000 -	0.1009 0.1911
λ	-	0.0000 -		0.5313 0.4445		0.0000 -	0.0000 -	0.6041 0.3006	0.0000 .
Log likelihood	-726.719	-726.719		-726.243		-726.269	-723.976	-724.497	-723.976

* Parameter estimates are reported followed by standard error of estimate.

** Function calculation is failed.

Table 3.4: Estimation Result of Expenditure Difference Model with Montevideo Data*

	Logit	Constrained Bivariate			Bivariate Extreme Value Model	Constrained Mixed Logit			Mixed Logit Model
		$(\gamma = 1, \lambda = 0)$	$(\lambda = 0)$	$(\gamma = 1)$		$(\gamma = 1, \lambda = 0)$	$(\lambda = 0)$	$(\gamma = 1)$	
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

[*] Parameter estimates are reported followed by standard error of estimate.
[1] RU represents the random utility model.
[2] ED represents the expenditure difference model.

Table 3.5: Sample Average of Welfare Measure for Environmental Quality Change[*]

3.7 UNKNOWN ALTERNATIVES (MISSING ALTERNATIVE IN THE MULTINOMIAL LOGIT MODEL)

Now suppose the choice situation that individual considers three alternatives, one of which is the current state (i = 0), the second is the proposed state (i = 1) and the last is another alternative for the same environmental improvement (i = A). Alternative A which is unknown to researcher implies that respondent agrees the environmental change but through different way. If the alternative A turns out to be impossible to implement, i.e. if no other options are available except the proposed one, it may be reasonable to think that more respondent would accept the proposed project rather than still remain in the current state. The random utility from the unknown alternative consists of the individual specific components that are same across all alternatives, the alternative specific components that may be similar with the proposed alternative but different with respect to at least the process, and unobservable random error term.

If error terms of random utility are *i.i.d.* type I extreme values across alternatives, the probability to accept the proposed state becomes a multinomial logit,

$$P_{1n} = P\left(U_{1n} > U_{0n}, U_{1n} > U_{An}\right) = \frac{\exp\left(V_1/\theta\right)}{\exp\left(V_1/\theta\right) + \sum_{j=0,A} \exp\left(V_j/\theta\right)}.$$

Without further assumption, the probability function cannot be useful in the estimation because choice of the alternative A is unknown to researcher. Since the unknown alternative is supposed to have the same target of environmental change but through a different process, assume that the random utility from the unknown alternative is same with that of the proposed state with respect to the marginal effect of parameters but different only in the expected value of random utilities, i.e. constant term[41]. Then the systematic part of the random utility of alternative A is expressed as $V_A = \phi + V_1$. The probability function of response of "yes" becomes

$$P_{1n} = \frac{1}{1 + \exp(\phi / \theta) + \exp(-v_n / \theta)}$$

and finally the log likelihood function for the unknown alternative model is

$$\ln L = \sum_{n=1}^{N} y_n \ln \left\{ \frac{1}{1 + \exp(\phi / \theta) + \exp(-v_n / \theta)} \right\} + (1 - y_n) \ln \left\{ \frac{\exp(\phi / \theta) + \exp(-v_n / \theta)}{1 + \exp(\phi / \theta) + \exp(-v_n / \theta)} \right\}$$

where θ is common standard deviation of error terms and $\tilde{\phi} = \phi / \theta$ is the normalized difference of the expected random utility between the proposed policy and the unknown alternative.

As ϕ goes to negative infinity, the probability of accepting the proposed state becomes the typical logit model. When $\phi = 0$, i.e. the unknown alternative provides the same utility as the proposed project, the probability becomes $P_{1n} = \left[2 + \exp(-v_n / \theta) \right]^{-1}$. If the unknown alternative provides much higher utility, the proposed project will be rarely accepted, i.e. the probability choosing the proposed policy becomes lower as the unknown alternative provides higher utility. Interestingly, the unknown alternative model has some similarities with the misclassification model in the sense that some portion of the response is classified into wrong category. However, the unknown alternative model is different from the misclassification model in that the unknown alternative model

[41] Although the assumption of the same marginal effect but different expected value is surely restrictive to the model, the assumption is similar with the constant treatment effect, which has been widely used in labor economics for unconfoundedness condition (for details, see Imbens 2004).

corrects the misclassification with strong assumption while the typical misclassification model estimates the probability of misclassification directly by plugging it in the log likelihood function. Misclassification in the unknown alternative model implies the unidentification of utilities in no response.

The willingness to pay can easily but also carefully be calculated. Note that alternative in the contingent valuation study is the state itself while typical multinomial logit model has alternatives in the given state[42]. Since there exist two alternatives after the change in the unknown alternative model, three kinds of welfare measure arise from the model. The first and traditional willingness to pay is

$$E\left(WTP_n \mid i=1\right)=\frac{1}{\alpha}x_n'\left(\beta_1-\beta_0\right)=x_n'\frac{\tilde{\beta}}{\tilde{\alpha}},$$

which is interpreted as the willingness to pay for the environmental change conditional on the project proposed in the survey, i.e. willingness to pay for the project. In fact, this is the typical interpretation of the willingness to pay in previous CV literatures. The second, also conditional definition is the willingness to pay for the unknown alternative;

$$E\left(WTP_n \mid i=A\right)=\frac{\phi}{\alpha}+\frac{1}{\alpha}x_n'\left(\beta_1-\beta_0\right)=\frac{\tilde{\phi}}{\tilde{\alpha}}+x_n'\frac{\tilde{\beta}}{\tilde{\alpha}}.$$

By definition, if the estimate of $\tilde{\phi}$ is positive, then the unknown alternative provide higher expected random utility, suggesting that policy maker needs to consider another

[42] Alternatively, given the current state (before change), respondent has to make a choice between proposed state and unknown state (after change). In this way of interpretation, the willingness to pay can be estimated as difference between the log sum of after change and of before change. With *i.i.d* type I extreme values, the log sum formula is well known for calculating the willingness to pay in the multinomial logit (Ben Akiva 1973, McFadden 1973, 1978, 1981, Domencich and McFadden 1975).

process of achieving the goal. The final definition of welfare measure is the willingness to pay for the environmental change itself such that

$$E\left(WTP_n\right)=P_{1n|1,A}\cdot E\left(WTP_n\mid i=1\right)+P_{An|1,A}\cdot E\left(WTP_n\mid i=A\right).$$

Since the probability of choosing the proposed state given the environmental change is $P_{1n|1,A}=\left[1+\exp\left(\phi/\theta\right)\right]^{-1}$, the unconditional willingness to pay becomes

$$E\left(WTP_n\right)=x_n'\frac{\tilde{\beta}}{\tilde{\alpha}}+\frac{\tilde{\phi}}{\tilde{\alpha}}\left[1+\exp\left(-\tilde{\phi}\right)\right]^{-1}.$$

Note that $\left[1+\exp\left(-\tilde{\phi}\right)\right]^{-1}$ is the conditional logistic probability of choosing unknown alternative and $\tilde{\phi}/\tilde{\alpha}$ is the constant difference of willingness to pay of unknown alternative from the proposed policy. Therefore, unconditional willingness to pay is the conditional willingness to pay for the project plus the weighted constant estimate.

Table 3.6 reports the estimation result of unknown alternative model with Barbados and Montevideo data sets. The parameter estimate of $\tilde{\phi}/\tilde{\alpha}$, constant2, is extremely small and insignificant in Barbados data, the result which is consistent with the result in Table 3.1 and 3.3 that show *i.i.d* extreme value distribution. Barbados data implies that respondent may not consider other alternatives thus the decision model becomes a simple logit. However, the estimation result from Montevideo data is quite different from the results in Table 3.2 and 3.4 since the general models do not impose the strong assumption on the functional form and parameter estimates are normalized by different standard deviation. The estimate of willingness to pay in Montevideo, therefore, provides different result from the general model.

Barbados		Montevideo	
Parameter	Estimate	Parameter	Estimate
Constant	0.6789 0.5261	Constant	-0.8614 0.2357
Income	0.0549 0.0210	Income	0.3334 0.0905
Bid	0.0387 0.0063	Bid	0.0125 0.0022
City	0.4099 0.2929		-
Age	-0.0285 0.0090		-
Constant2	-13.2911 550.4758	Constant2	-0.4415 0.3100
Log likelihood	-160.841		-714.887
Sample Average of WTP**	-2.0701		29.1442

* Parameter estimates are reported followed by standard error of estimate.
** Willingness to pay for the proposed state

Table 3.6: Estimation Result of Unknown Alternatives Model*

3.8 CONCLUSIONS

This essay challenges the theoretical and technical background of the simple logit model often used for estimating willingness to pay from dichotomous choice contingent valuation applications. The survey questionnaire in a dichotomous choice contingent valuation asks a respondent to compare two states of the world: the proposed state and reference state. The proposed state represents the future with environmental change by the proposed policy and the reference state represents the future with all other possibilities including current state. The simple logit model assumes that the respondent's evaluations of the two states are stochastically independent and homoskedastic. However, it is possible and in many cases likely, that uncertainty on the part of the respondent, poor questionnaire design or simple inherent heterogeneity across states of the world may lead to heteroskedastic and correlated errors across states of a given individual. For instance, respondent has some degree of uncertainty about the proposed policy while the current state is deterministic. In addition, the reference state consists not only of the no-change-state but also of all other possibilities including change under unknown alternatives.

By relaxing restrictive assumptions of the standard random utility model, this essay suggests a generalized estimation technique that includes a number of existing models as special cases. To identify heteroskedasticity and correlation between the reference and proposed states, a Gumbel mixed model of maxima, a member of the class of bivariate extreme value distributions, has been employed into the random utilities. Nested within this generalized model are the heteroskedastic logit model and the simple logit. In addition to the random utility model, the essay also develops an expenditure difference model estimated with a Gumbel mixed bivariate distribution of minima. Again, this model has nested within it a number of standard logit-expenditure difference models. The nesting structure allows for straightforward tests of the homoskedastic-independent error assumptions.

Estimation results from several existing data including Barbados and Montevideo data show that correlation between two states is usually minimal, but homoskedastic errors are rejected in many cases. Montevideo data presents extremely different scale of error terms across states implying that the extreme value distribution, 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 refine the choice situation and 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. Decision of which estimation model should be used in practice is solely in the researcher on the basis how he or she defines the choice situation and choice set.

BIBLIOGRAPHY

Abdelbasit, K.M. and R.L. Plackett, 1983, "Experimental Design for Binary Data," *Journal of the American Statistical Association* 78(381), 90-98.

Alberini, A., 1995, "Optimal Designs for Discrete Choice Contingent Valuation Surveys: Single-Bound, Double-Bound, and Bivariate Models," *Journal of Environmental Economics and Management*, 28, 287-306.

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., 1973, "The Structure of Travel Demand Models," Ph.D. Thesis, MIT.

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.

Bond, C.A., K. Giraud and D. Larson, 2002, "Temporal Payment Issues in Contingent Valuation Analysis," *presented in AAEA meeting* in Long Beach, California.

Boyle, K.J., M. P. Welsh, and R. C. Bishop, 1988, "Validation of Empirical Measures of Welfare Change: Comment," *Land Economics* 64(1), 94-98.

Cameron, T.A. and D.D. Huppert, 1991, "Referendum Contingent Valuation Estimates: Sensitivity to the Assignment of Offered Values," *Journal of the American Statistical Association*, 86(416), 910-918.

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.

Carson, R.T., W.M. Hanemann, R.J. Kopp, J.A. Krosnick, R.C. Mitchell, S. Presser, P.A. Ruud, and V.K. Smith with M. Conaway and K. Martin, 1997, "Temporal Reliability of Estimates from Contingent Valuation." *Land Economics*, 73 (2): 151-63.

Carson, R.T., W.M. Hanemann, R.J. Kopp, J.A. Krosnick, R.C. Mitchell, S. Presser, P.A. Ruud, and V.K. Smith with M. Conaway and K. Martin, 1998, "Referendum Design and Contingent Valuation: The NOAA Panel's No-Vote Recommendation," *Review of Economics and Statistics*, 80, 335-338.

Casella, G. and R.L. Berger, 2002, *Statistical Inference* (2nd ed.), Duxbury Press.

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.

Chen, H.Z. and A. Randall, 1997, "Semi-nonparametric Estimation of Binary Response Models with an Application to Natural Resource Valuation," *Journal of Econometrics*, 76, 323-340.

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.

Coller, M., G.W. Harrison, E.E. Ruström, 2002, "Dynamic Consistency in the Laboratory," unpublished.

Cooper, J.C., 1993, "Optimal Bid Selection for Dichotomous Choice Contingent Valuation Surveys," *Journal of Environmental Economics and Management*, 24, 25-40.

Cooper, J.C. and J. Loomis, 1992, "Sensitivity of Willingness-to-Pay Estimates to Bid Design in Dichotomous Choice Contingent Valuation Models," *Land Economics* 68(2), 211-24.

Copas, J.B., 1988, "Binary Regression Models for Contaminated Data," *Journal of the Royal Statistical Society. Series B* (*Methodological*), 50(2), 225-265.

Crocker, T.D., and J.F. Shogren. 1993. "*Dynamic Inconsistency in Valuing Environmental Goods*." *Ecological Economics*, 7, 239-54.

Crooker, J.R. and J.A. Herriges, 2004, "Parametric and Semi-Nonparametric Estimation of Willingness-to-Pay in the Dichotomous Choice Contingent Valuation Framework", *Environmental and Resource Economics*, 27, 451-480.

Cropper, M., D. Laibson, 1999, "The Implication of Hyperbolic Discounting for Project Evaluation" in Discounting and Intergenerational Equity by Weyant. Washington, DC: Resources for the Future.

Diamond, P.A. and J.A. Hausman. 1994. "Contingent Valuation: Is Some Number Better Than No Number?" *Journal of Economic Perspectives*, 8 (4), 45-64.

Domencich, T. and D. McFadden, 1975, *Urban Travel Demand-A Behavioral Analysis*, Amsterdam: North Holland Publishers.

Finney, D.J., 1971, *Probit Analysis* (3rd ed.), Cambridge University Press.

Ford, I., B. Torsney and C.F.J. Wu, 1992, "The Use of a Canonical Form in the Construction of Locally Optimal Designs for Non-Linear Problems," *Journal of the Royal Statistical Society. Series B* (*Methodological*), 54(2), 569-583.

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., J. Huang, J.C. Whitehead., 1999, "Are Hypothetical Referenda Incentive Compatible? A Comment," *Journal of Political Economy*, 107 (1), 186-196.

Haab, T.C. and K.E. McConnell, 2002, *Valuing Environmental and Natural Resources: The Econometrics of Nonmarket Valuation*, Edward Elgar Publisher.

Haab, T.C. and K.E. McConnell, 1997, "Referendum Models and Negative Willingness to Pay: Alternative Solutions," *Journal of Environmental Economics and Management*, 32, 251-270.

Hanemann, W.M., 1984, "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses," *American Journal of Agricultural Economics*, 66, 332-341.

Hanemann, W.M., 1994, "Valuing the Environment Through Contingent Valuation." *Journal of Economic Perspectives*, 8 (4), 19-43.

Harrison, G.W., and E. Johnson, 2002, "Estimating Individual Discount Rates for the United States: Inferences from a Natural Field Experiment," unpublished.

Harrison, G.W., M.I. Lau, M.B. Williams, 2002, "Estimating Individual Discount Rates in Denmark: A Field Experiment," Forthcoming, *American Economic Review*, 92.

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.

Huang, J., T.C. Haab and J.C. Whitehead, 1997, "Willingness to Pay for Quality Improvements: Should Revealed and Stated Preference Data Be Combined?" *Journal of Environmental Economics and Management*, 34, 240-255.

Huhtala, A., 2000, "Binary Choice Valuation Studies with Heterogeneous Preferences Regarding the Program Being Valued," *Environmental and Resource Economics* 16, 263-279.

Ibáñe, A.M., and K. McConnell, 2001, "Scheduling Payments and Discount Rates: Implication for Contingent Valuation in Developing Countries," unpublished.

Imbens, G.W., 2003, "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review," NBER Technical Working Paper 294, National Bureau of Economic Research, Inc.

Júdex, L., R. de Andrés, C. P. Hugalde, E. Urzainqui and M. Ibañez, 2000, "Influence of Bid and Subsample Vectors on the Welfare Measure Estimate in Dichotomous Choice Contingent Valuation: Evidence from a Case-study," *Journal of Environmental Management*, 60, 253-265.

Kahneman, D., and J.L. Knetsch, 1992, "Valuing Public Goods: The Purchase of Moral Satisfaction," *Journal of Environmental Economics and Management*, 22(1), 55-70.

Kalish, L.A. and J.L. Rosenberger, 1978, "Optimal Designs for the Estimation of the Logistic Function," Technical Reports #33, Department of Statistics Pennsylvania State University.

Kanninen, B.J., 1995, "Bias in Discrete Response Contingent Valuation," *Journal of Environmental Economics and Management*, 28, 114-125.

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.

Kriström, B., 1990, "A Non-Parametric Approach to the Estimation of Welfare Measure in Discrete Response Valuation Studies." *Land Economics* 66, 135-139.

Lewbel, A., O. Linton and D. McFadden, 2002, "Estimating Features of a Distribution from Binomial Data", Manuscript, Boston College.

Mathew, T. and B.K. Sinha, 2001, "Optimal Designs for Binary Data under Logistic Regression," *Journal of Statistical Planning and Inference*, 93, 295-307.

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 Zaremmbka, P., Academic Press, New York.

McFadden, D., 1978, "Modeling the Choice of Residential Location," in *Spatial Interaction Theory and Planning Models* ed. by Karlqvist, A., et al., Amsterdam: North-Holland.

McFadden, D., 1981, “Econometric Models of Probabilistic Choice,” in *Structural Analysis of Discrete Data with Econometric Applications* ed. by Manski, C.F. and D. McFadden, MIT Press, Cambridge, MA.

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.

Minkin, S., 1987, “Optimal Designs for Binary Data,” *Journal of the American Statistical Association*, 82(400), 1098-1103.

Mitchell, R.C., and R.T. Carson, 1989, *Using Surveys to Value Public Goods: The Contingent Valuation Method*. Washington, DC: Resources for the Future.

Nyquist, H., 1992, “Optimal Designs of Discrete Response Experiments in Contingent Valuation Studies,” *The Review of Economics and Statistics* 74(3), 559-563.

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.

Read, D., 2001, “Is Time-Discounting Hyperbolic or Subadditive?” *Journal of Risk and Uncertainty*, 23 (1), 5-32.

Rowe, R.D., W.D. Shaw, and W. Schulze, 1992, “NESTWCCA Oil Spill.” In *Natural Resource Damages: Law and Economics*, eds. K. Ward and J. Duffield. New York: John Wiley and Sons.

Sitter, R.R., 1992, “Robust Designs for Binary Data,” *Biometrics*, 48(4), 1145-1155.

Sitter, R.R., and C.F.J. Wu, 1993, “Optimal Designs for Binary Response Experiments: Fieller, D, and A Criteria,” *Scandinavian Journal of Statistics*, 20, 329-342

Sitter, R.R., and C.F.J. Wu, 1993, “On the Accuracy of Fieller Intervals for the Binary Response Data,” *Journal of the American Statistical Association*, 88(423), 1021-1025.

Smith, V.K., 1992, “Comment: Arbitrary Values, Good Causes and Premature Verdicts.” *Journal of Environmental Economics and Management*, 22(1), 71-89.

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.

Stevens, T., N. DeCoteau, and C. Willis, 1997, "Sensitivity of Contingent Valuation to Alternative Payment Schedules." *Land Economics*, 73 (1), 140-148.

Straud, A.H. and D. Secrest, 1966, *Gaussian Quadrature Formulas*, Series in Automatic Computation, Prentice-Hall, London.

Swait, J. and J. Louviere, 1993, "The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models." *Journal of Marketing Research*, 30, 305-14.
Tawn, J.A., 1988, "Bivariate Extreme Value Theory: Models and Estimation," *Biometrika*, 75(3), 397-415.

Tawn, J.A., 1988, "Bivariate Extreme Value Theory: Models and Estimation," *Biometrika*, 75(3), 397-415.

Thaler, R., 1981, "Some Empirical Evidence on Dynamic Inconsistency." *Economic Letters*, 8: 201-7.

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., 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.

Wu, C.F.J., 1988, "Optimal Design for Percentile Estimation of a Quantal Response Curve" in *Optimal Design and Analysis of Experiments*, Dodge, Y., V. Federov and H.P. Wynn (eds), Elsevier, Amsterdam, 213-223.

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.

APPENDICES

A: Survey Questionnaire for Five-year Project (One-time Payment)

Chesapeake Bay Attitude and Preference Survey: **Page 1 of 4**

1) Is a national goal of protecting nature and preventing pollution very important, somewhat important, or not at all important to you (Check one)?

- ❑ Very Important
- ❑ Somewhat Important
- ❑ Not Important

2) On a scale of 1 to 10, with 1 being that pollution controls have gone too far and cost more than they are worth and 10 being that that pollution controls can never be too strict and improvements should be made regardless of cost, how would you rate current pollution controls in the United States (please circle one)?

1	2	3	4	5	6	7	8	9	10

3) Do you belong to or contribute to any environmental organizations on a regular basis (such as The Sierra Club, The Nature Conservancy, The Chesapeake Bay Foundation)? (please circle one)

Yes No

4) Do you consider yourself an environmentalist?

Yes No

5) How would you rate the overall quality of the Chesapeake Bay (check one)?

- ❑ Excellent/Pristine
- ❑ Very Good
- ❑ Good
- ❑ Fair
- ❑ Poor

6) Is a regional goal of protecting water quality in the Chesapeake Bay very important, somewhat important, or not at all important to you (check one)?

- ❑ Very Important
- ❑ Somewhat Important
- ❑ Not Important

7) Do you own a boat (circle one)?

Yes No

8) On a scale of 1 to 10, with 1 being that programs to protect the Chesapeake Bay have gone too far and cost more than they are worth and 10 being that that programs to protect the Chesapeake Bay can never be too strict and these programs should be made regardless of cost, how would you rate current protection programs in the Chesapeake Bay region?

1	2	3	4	5	6	7	8	9	10

9) Over the past year, about how many times have you participated in each of these activities on the Chesapeake or its tributaries?

Activity	Number of times participating in activity (check the appropriate box)										
	0	1	2	3	4	5	6	7	8	9	More than 10
Boating/Jet Skiing/Skiing											
Swimming											
Fishing											
Bird Watching											
Beach Going											

Please list any other activities in or around Chesapeake Bay waters you have participated in over the last year.

____

____

10) Many different types of seafood are harvested from the Chesapeake Bay. These include Striped Bass, Blue Crabs, and Oysters. On a scale of 1 to 10, with 1 being not important and 10 being extremely important, how important do you think seafood harvested from the Chesapeake Bay is to the economy of the Chesapeake Bay region?

1	2	3	4	5	6	7	8	9	10

11) On a scale of 1 to 10, with 1 being not important and 10 being extremely important, how important do you think seafood harvested from the Chesapeake Bay is to the water quality in the Chesapeake Bay?

1	2	3	4	5	6	7	8	9	10

A1_a

Some Information About Oysters and the Chesapeake Bay: Please Read.

Oysters are a keystone species in Chesapeake Bay. An adult oyster is capable of filtering 50 gallons of water a day. At its peak, the Bay's oyster population could filter in three to four days as much water as is in the bay. Likewise, oyster reefs are home to more than 300 species of Bay life. Oysters were once the most valuable commercial fishery on the Bay. However, overharvest, disease, and pollution diminished the population to 1 percent of historic levels in little over a century.

During the 20th century, oysters were the most harvested species in the Bay. Because heavy harvest, loss of reef habitat, pollution, and disease the Chesapeake's oyster population today is thought to be only one percent of what it was just over a century ago. Bay oysters used to grow in tall reefs that elevated oysters from the silty bottom into food-rich currents above. Reefs provided far more nooks and crannies for creatures to hide in than flatter beds do .In the 19th century, oyster reefs were so large that they were considered navigational hazards. After 120 years of intense harvest, very few reefs remain in the Bay.

What Role Do Oysters Play in the Chesapeake Bay Region?

Food: Each year, more than 500,000 pounds each of oysters and rockfish are consumed from the waters of the Chesapeake

Economic: Oystering was the most valuable commercial fishery in the Bay until the mid-1980s, when it was overtaken by crabbing.

Environmental: Oysters purify the Chesapeake Bay as they filter the water for their food. Dirt, nutrients, and algae can cause problems in Bay waters. Oysters filter these things, and either eat them or shape them into small packets, which are deposited on the bottom where they are not harmful. The oysters in the Bay could once filter the entire Bay in three to six days. Filtering now takes almost one full year.

Fish Habitat: Oyster bars (reefs) are among the best places to fish. The hard surfaces of oyster shells and the nooks between the shells provide places where a host of small animals can live. Hundreds of animals use oyster bars: grass shrimp, amphipods, bryozoans, anemones, barnacles, oyster drills, hooked mussels, mud crabs, and red beard sponge to name a few. Many of these serve as food for larger animals including striped bass, weakfish, black drum, Croakers, and blue crabs.

Questions About Oysters and the Chesapeake Bay

12) Do you find this information helpful in understanding the role oysters play in the Bay (Circle One)?

Yes No

13) The information mentions four roles oysters play in the Chesapeake Bay: Food, Economic, Environmental and Fish Habitat. Please rank these four roles in order of which you think is most important, with 1 being the most important and 4 being the least important:

____Food
____Economic
____Environmental
____Fish Habitat

14) To the best of your knowledge, are there more, about the same, or less oysters in the Chesapeake Bay as 10 years ago?

- ❑ More
- ❑ Less
- ❑ About the same

15) Over the course of a year, about how often do you eat oysters or food containing oysters?

- ❑ Once a week
- ❑ Once a month
- ❑ Once every two months
- ❑ Once a year
- ❑ Less than once a year

16) Many types of projects exist for the improvement of water quality and habitat in the Chesapeake Bay. These include, but are not limited to, reducing pollution in the Bay, restoring habitat for Bay species, managing fisheries for conservation, and developing policies to ensure the sustainability of the Bay. On a scale of 1 to 10, with 1 being no support and 10 being strong support, how would you rate your level of support for Chesapeake Bay

1	2	3	4	5	6	7	8	9	10

17) How effective do you think programs have been to restore the Chesapeake Bay (check one)?

- ❑ Very Effective
- ❑ Somewhat Effective
- ❑ Not very effective
- ❑ Not effective at all

A1_a

Oyster reefs play an important ecological role in the Chesapeake Bay. Oysters cluster together to create a hard surface on the Bay bottom and a three-dimensional reef habitat for many species. Oyster reefs have declined in the Chesapeake Bay due to harvest pressure, oyster diseases, and pollution. Harvesting techniques have reduced many three-dimensional reefs to flat surfaces.

18) On a scale of 1 to 10, with 1 being a minor problem and 10 being a major problem, how would you rate declining oyster populations as a threat to the health of the Chesapeake Bay (Circle one)?

1	2	3	4	5	6	7	8	9	10

In 2000, the Bay's scientific community agreed on a strategy to jump-start the natural machinery that sustains the Bay's oysters. To restore the oyster population, the strategy calls for rebuilding oyster habitat, stocking it with hardy oysters, and establishing sanctuary areas that provide a continuous, long-term supply of healthy oysters.

19) On a scale of 1 to 5, with 1 being no support and 5 being strong support, how would you rate your level of support for oyster sanctuary and reef creation programs (Circle one)?

1	2	3	4	5

Current proposals are to create and protect 10,000 acres of new oyster sanctuary and 1,000 acres of new oyster reef within these sanctuaries over the next **5 years**. Oyster sanctuaries will be established in 10-25 acre parcels spread throughout the Bay and estuaries. These oyster sanctuaries are protected regions in which the harvesting of oysters is not allowed. Other types of fishing will be allowed in these sanctuaries.

If the restoration project were started today, 10,000 acres of oyster sanctuary and 1,000 acres of constructed reef would be created over the next **5 years** at a rate of 2,000 acres of sanctuary and 200 acres of reef created per year. It is expected that upon completion these 10,000 acres of sanctuary and 1,000 acres of artificial reef will increase oyster populations ten-fold over current levels, and allow the water of the bay to be filtered approximately once a month as opposed to the current once per year. The benefits of the restoration: increased water filtering and improved fish habitat would begin with the first completed reef and continue to increase until completion of the project. The benefits would then continue at that level into the future.

20) Do you favor, oppose or are you indifferent to the use of oyster sanctuaries (harvest-free areas) as a means of increasing oyster population in the Bay?

- ❑ Favor
- ❑ Oppose
- ❑ Neither favor or oppose

21) Do you consider 10,000 acres of oyster sanctuary including 1,000 acres of constructed reef over the next **5 years** to be (Check one):

- ❑ a major improvement
- ❑ a minor improvement
- ❑ no improvement

We are now interviewing people to find out how they would vote if this program were on the ballot in a statewide election. The plan is to create 10,000 acres of oyster sanctuary including 1,000 acres of constructed reef over the next **5 years**. Here's how it would be paid for. **Each taxpayer would pay a one-time additional amount on their next state income tax return.** This is the only payment that would be required, and all payments would go into a special fund that could only be used for the program to restore oyster reefs.

22) In general, do you think a special fund for oyster reef restoration is:

- ❑ a good idea
- ❑ a bad idea
- ❑ I don't know

The program would only be carried out if people are willing to pay this one time tax. There are reasons why you might vote for the restoration program and reasons why you might vote against it. Upon completion, the restoration program will allow the waters of the Chesapeake to be filtered on a monthly basis as opposed to the current annual basis. Further, fish habitats will be enhanced. Your household might prefer to spend the money to solve other social and environmental problems instead. Or, the program might cost more money than your household wants to spend for this.

23) The restoration program is estimated to cost your household a total of **$__**. Your household would pay this as a special one time tax added to next year's state income tax. **If an election were to be held today and the total cost to your household was $__ would you vote for or against the 5 year restoration program** (Check one)?

- ❑ I would vote for the program
- ❑ I would vote against the program
- ❑ I do not know whether I would vote for or against the program

A1_a

24) What is the maximum amount you would be willing to pay for the Oyster Reef restoration project?

$____________________

The next few questions are about you and your household. The responses are for statistical purposes only. They will not be associated with your name in any way. All responses will be anonymous and confidential.

25) How long have you lived in your current state?

_____________Years_______________Months

26) Do you plan to move from this location in the next 5 years?

Yes No

26) How long have you lived in your current county?

_____________Years______________Months

27) Approximately how far do you live from the closest point of access to the Chesapeake Bay?

___________________Miles

28) What is your Zip Code?

_ _ _ _ _

29) How many people, including yourself, normally live in your household?

____

30) How many are under the age of 18?

____

31) Are you male or female? (circle one)

Male Female

32) What is your race or ethnic background?

- ❑ White
- ❑ Black
- ❑ Hispanic
- ❑ Other__________________

33) What is your political affiliation (check one)?

- ❑ Republican
- ❑ Democrat
- ❑ Independent
- ❑ Other_____________________________

34) What year were you born?

19 ________

35) What is the highest level of education you have completed?

- ❑ Some High School
- ❑ High School
- ❑ Some College/Junior College
- ❑ Associates Degree
- ❑ Bachelors Degree
- ❑ Master's degree
- ❑ Doctorate
- ❑ Professional Degree

36) Are you currently:

- ❑ Employed full-time
- ❑ Employed part-time
- ❑ Retired
- ❑ Unemployed

37) To the best of your recollection, what was your total household income over the past year?

- ❑ Less than $15,000
- ❑ $15,000-$24,999
- ❑ $25,000-$34,999
- ❑ $35,000-$49,999
- ❑ $50,000-$74,999
- ❑ $75,000-$99,999
- ❑ $100,000-$150,000
- ❑ greater than $150,000

Thank you for completing this survey. To return it, place it in the stamped self-addressed envelope that accompanied the survey and mail it back to us. If for some reason you do not have the envelope, please mail the survey to:

Chesapeake Bay Survey C/O Timothy C. Haab
AEDE, 2120 Fyffe Road
The Ohio State University
Columbus, Ohio, 43210

A1_a

Survey Questionnaire for Five-year Project (Annual Payment)

Oyster reefs play an important ecological role in the Chesapeake Bay. Oysters cluster together to create a hard surface on the Bay bottom and a three-dimensional reef habitat for many species. Oyster reefs have declined in the Chesapeake Bay due to harvest pressure, oyster diseases, and pollution. Harvesting techniques have reduced many three-dimensional reefs to flat surfaces.

18) On a scale of 1 to 10, with 1 being a minor problem and 10 being a major problem, how would you rate declining oyster populations as a threat to the health of the Chesapeake Bay (Circle one)?

1	2	3	4	5	6	7	8	9	10

In 2000, the Bay's scientific community agreed on a strategy to jump-start the natural machinery that sustains the Bay's oysters. To restore the oyster population, the strategy calls for rebuilding oyster habitat, stocking it with hardy oysters, and establishing sanctuary areas that provide a continuous, long-term supply of healthy oysters.

19) On a scale of 1 to 5, with 1 being no support and 5 being strong support, how would you rate your level of support for oyster sanctuary and reef creation programs (Circle one)?

1	2	3	4	5

Current proposals are to create and protect 10,000 acres of new oyster sanctuary and 1,000 acres of new oyster reef within these sanctuaries over the next **5 years**. Oyster sanctuaries will be established in 10-25 acre parcels spread throughout the Bay and estuaries. These oyster sanctuaries are protected regions in which the harvesting of oysters is not allowed. Other types of fishing will be allowed in these sanctuaries.

If the restoration project were started today, 10,000 acres of oyster sanctuary and 1,000 acres of constructed reef would be created over the next **5 years** at a rate of 2,000 acres of sanctuary and 200 acres of reef created per year. It is expected that upon completion these 10,000 acres of sanctuary and 1,000 acres of artificial reef will increase oyster populations ten-fold over current levels, and allow the water of the bay to be filtered approximately once a month as opposed to the current once per year. The benefits of the restoration: increased water filtering and improved fish habitat would begin with the first completed reef and continue to increase until completion of the project. The benefits would then continue at that level into the future.

20) Do you favor, oppose the use of oyster sanctuaries (harvest-free areas) as a means of increasing oyster population in the Bay?

- ❑ Favor
- ❑ Oppose
- ❑ Neither favor or oppose

21) Do you consider 10,000 acres of oyster sanctuary including 1,000 acres of constructed reef over the next **5 years** to be (Check one):

- ❑ a major improvement
- ❑ a minor improvement
- ❑ no improvement

We are now interviewing people to find out how they would vote if this program were on the ballot in a statewide election. The plan is to create 10,000 acres of oyster sanctuary including 1,000 acres of constructed reef over the next **5 years**. Here's how it would be paid for. **Each taxpayer would pay an additional amount on their state taxes over the next 5 years.** This is the only payment that would be required, and all payments would go into a special fund that could only be used for the program to restore oyster reefs.

22) In general, do you think a special fund for oyster reef restoration is:

- ❑ a good idea
- ❑ a bad idea

The program would only be carried out if people are willing to pay this annual tax. There are reasons why you might vote for the restoration program and reasons why you might vote against it. Upon completion, the restoration program will allow the waters of the Chesapeake to be filtered on a monthly basis as opposed to the current annual basis. Further, fish habitats will be enhanced. Your household might prefer to spend the money to solve other social and environmental problems instead. Or, the program might cost more money than your household wants to spend for this.

23) The restoration program is estimated to cost your household a total of **$_____ per year**. Your household would pay this as an annual tax over the next **5 years** added to your state income tax. **If an election were to be held today and the cost to your household was $_____ per year for the next 5 years, would you vote for or against the 5 year restoration program?** (Check one)?

- ❑ I would vote for the program
- ❑ I would vote against the program
- ❑ I do not know whether I would vote for or against the program

A2_a

Survey Questionnaire for Five-year Project (Perpetuity Payment)

Oyster reefs play an important ecological role in the Chesapeake Bay. Oysters cluster together to create a hard surface on the Bay bottom and a three-dimensional reef habitat for many species. Oyster reefs have declined in the Chesapeake Bay due to harvest pressure, oyster diseases, and pollution. Harvesting techniques have reduced many three-dimensional reefs to flat surfaces.

18) On a scale of 1 to 10, with 1 being a minor problem and 10 being a major problem, how would you rate declining oyster populations as a threat to the health of the Chesapeake Bay (Circle one)?

1	2	3	4	5	6	7	8	9	10

In 2000, the Bay's scientific community agreed on a strategy to jump-start the natural machinery that sustains the Bay's oysters. To restore the oyster population, the strategy calls for rebuilding oyster habitat, stocking it with hardy oysters, and establishing sanctuary areas that provide a continuous, long-term supply of healthy oysters.

19) On a scale of 1 to 5, with 1 being no support and 5 being strong support, how would you rate your level of support for oyster sanctuary and reef creation programs (Circle one)?

1	2	3	4	5

Current proposals are to create and protect 10,000 acres of new oyster sanctuary and 1,000 acres of new oyster reef within these sanctuaries over the next **5 years**. Oyster sanctuaries will be established in 10-25 acre parcels spread throughout the Bay and estuaries. These oyster sanctuaries are protected regions in which the harvesting of oysters is not allowed. Other types of fishing will be allowed in these sanctuaries.

If the restoration project were started today, 10,000 acres of oyster sanctuary and 1,000 acres of constructed reef would be created over the next **5 years** at a rate of 2,000 acres of sanctuary and 200 acres of reef created per year. It is expected that upon completion these 10,000 acres of sanctuary and 1,000 acres of artificial reef will increase oyster populations ten-fold over current levels, and allow the water of the bay to be filtered approximately once a month as opposed to the current once per year. The benefits of the restoration: increased water filtering and improved fish habitat would begin with the first completed reef and continue to increase until completion of the project. The benefits would then continue at that level into the future.

A3_a

20) Do you favor, oppose the use of oyster sanctuaries (harvest-free areas) as a means of increasing oyster population in the Bay?

- ❑ Favor
- ❑ Oppose
- ❑ Neither favor or oppose

21) Do you consider 10,000 acres of oyster sanctuary including 1,000 acres of constructed reef over the next **5 years** to be (Check one):

- ❑ a major improvement
- ❑ a minor improvement
- ❑ no improvement

We are now interviewing people to find out how they would vote if this program were on the ballot in a statewide election. The plan is to create 10,000 acres of oyster sanctuary including 1,000 acres of constructed reef over the next **5 years**. Here's how it would be paid for. **Each taxpayer would pay an annual tax added to their state income taxes.** This is the only payment that would be required, and all payments would go into a special fund that could only be used for the program to restore oyster reefs.

22) In general, do you think a special fund for oyster reef restoration is:

- ❑ a good idea
- ❑ a bad idea

The program would only be carried out if people are willing to pay this annual tax. There are reasons why you might vote for the restoration program and reasons why you might vote against it. Upon completion, the restoration program will allow the waters of the Chesapeake to be filtered on a monthly basis as opposed to the current annual basis. Further, fish habitats will be enhanced. Your household might prefer to spend the money to solve other social and environmental problems instead. Or, the program might cost more money than your household wants to spend for this.

23) The restoration program is estimated to cost your household a total of **$_____ per year**. Your household would pay this as an annual amount added to your state tax return. **If an election were to be held today and the cost to your household was $_____ per year, indefinitely, would you vote for or against the 5-year restoration program** (Check one)?

- ❑ I would vote for the program
- ❑ I would vote against the program
- ❑ I do not know whether I would vote for or against the program

B: Technical Note on Some Useful Results

B.1 Fieller's Theorem (Casella G. and R.L. Berger, 2002)

Fieller's theorem (Fieller 1954) is to get an exact confidence set on a ratio of normal means. Given a random sample (x_i, y_i) from a bivariate normal distribution with parameter $(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$, a confidence set on $\theta = \mu_y / \mu_x$ can be formed as follows. Define $\bar{Z}_\theta = \bar{y} - \theta\bar{x}$, then $\bar{Z}_\theta$ is normal with mean zero and variance

$$V_\theta = \frac{1}{n}\left(\sigma_y^2 - 2\theta\rho\sigma_y\sigma_x + \theta^2\sigma_x^2\right).$$

It can be shown that $\bar{Z}_\theta / \sqrt{\hat{V}_\theta} \sim t_{n-1}$ and the set

$$\left\{\theta : \frac{\bar{z}_\theta^2}{\hat{v}_\theta} \le t_{n-1,\alpha/2}^2\right\}$$

defines a 1-α confidence set for θ. This set defines a parabola in θ and the roots of the parabola give the endpoints of the confidence set.

B.2 Determinant and efficiency of *J*-point *kk*-design and D-optimal design

When the number of bid points is *J* and observations are equally distributed in *J* points of bid, the determinant of information matrix is

$$\begin{aligned}\det\left[I(\mu,\beta)\right]_J &= n^2\beta^2\left[\sum_{j=1}^{J} w_j \sum_{j=1}^{J} w_j\left(\mu - b_j\right)^2 - \left\{\sum_{j=1}^{J} w_j\left(\mu - b_j\right)\right\}^2\right] \\ &= n^2\beta^2\sum_{i=1}^{J-1}\sum_{j=1}^{J-i} w_i w_{i+j}\left(b_i - b_{i+j}\right)^2\end{aligned}$$

Through simple manipulation, the double summation can be expressed as

$$\det\left[I(\mu,\beta)\right]_J = \frac{(n\beta)^2}{2}\begin{bmatrix} w_1w_1(b_1-b_1)^2 + w_1w_2(b_1-b_2)^2 + & \dots & +w_1w_J(b_1-b_J)^2 + \\ \vdots & \ddots & \vdots \\ +w_Jw_1(b_J-b_1)^2 + w_Jw_2(b_J-b_2)^2 + & \dots & +w_Jw_J(b_J-b_J)^2 \end{bmatrix}$$

$$= \frac{(n\beta)^2}{2}\sum_{i=1}^{J}\sum_{j=1}^{J} w_iw_j(b_i-b_j)^2.$$

Using $t_i = \beta(\mu - b_i)$, the determinant becomes

$$\det\left[I(\mu,\beta)\right]_J = \frac{n^2}{2}\sum_{i=1}^{J}\sum_{j=1}^{J} w_iw_j(t_i-t_j)^2.$$

Let the design point be $b_i = \mu_0 - d_i/\beta_0$. Then, since $t_i = \beta(\mu-\mu_0) + d_i\beta/\beta_0$, the contribution of a pair of any two points to the determinant in the double summation is

$$w_iw_j(t_i-t_j)^2 = \frac{\exp\{t_i+t_j\}\{t_i-t_j\}^2}{\left\{\left[1+\exp\{t_i\}\right]\left[1+\exp\{t_j\}\right]\right\}^2}$$
$$= \exp\left\{\frac{\beta}{\beta_0}(d_i+d_j)\right\}\left\{\frac{\beta}{\beta_0}\frac{(d_i-d_j)\exp\{\beta(\mu-\mu_0)\}}{A_iA_j}\right\}^2.$$

where $A_i = \exp\left\{\frac{\beta}{\beta_0}d_i\right\} + \exp\{\beta(\mu-\mu_0)\}$. Thus, the determinant *J*-point *kk*-design becomes

$$\det\left[I(\mu,\beta)\right]_J = \frac{n^2}{2}\sum_{i=1}^{J}\sum_{j=1}^{J}\exp\left\{\frac{\beta}{\beta_0}(d_i+d_j)\right\}\left\{\frac{\beta}{\beta_0}\frac{(d_i-d_j)\exp\{\beta(\mu-\mu_0)\}}{A_iA_j}\right\}^2.$$

When the design points are two and allocated at the D-optimal points such as $\{d_0, -d_0\}$ where $d_0 = \ln(p_0(1-p_0))$ and $p_0 = 0.824$, the determinant of the information matrix is simplified to be

$$\det\left[I(\mu,\beta)\right]=\left[\frac{Nd_0}{AB}\left(\frac{\beta}{\beta_0}\right)\exp\left\{\beta(\mu-\mu_0)\right\}\right]^2$$

where $A=\exp\left(\frac{\beta}{\beta_0}d_0\right)+\exp\left\{\beta(\mu-\mu_0)\right\}$ and $B=\exp\left(-\frac{\beta}{\beta_0}d_0\right)+\exp\left\{\beta(\mu-\mu_0)\right\}$. The result can be derived by substituting two optimal points $b=\mu_0\pm\beta_0^{-1}\ln\left(p_0(1-p_0)\right)$ into the determinant expression (Abdelbasit and Plackett 1983). The determinant of D-optimal design is maximized when $\mu=\mu_0$ and $\beta=\beta_0$. The maximum value of determinant is

$$\det\left[I(\mu,\beta)\right]_{2,\mu=\mu_0,\beta=\beta_0}=\left[Np_0(1-p_0)\left\{\ln\left(\frac{p_0}{1-p_0}\right)\right\}\right]^2.$$

The efficiency of D-optimal design with poor information is,

$$Eff_{\mu,\beta}=\left[\left(\frac{\beta}{\beta_0}\right)\frac{\exp\left\{\beta(\mu-\mu_0)\right\}}{p_0(1-p_0)AB}\right]^2.$$

For the *J*-point *kk*-design, the efficiency is

$$Eff_J=\left(\frac{n\exp\left\{\beta(\mu-\mu_0)\right\}}{\sqrt{2}Np_0(1-p_0)d_0}\right)^2\sum_{i=1}^{J}\sum_{j=1}^{J}\exp\left\{\frac{\beta}{\beta_0}\left(d_i+d_j\right)\right\}\left\{\frac{\beta}{\beta_0}\frac{\left(d_i-d_j\right)}{A_iA_j}\right\}^2.$$

The general expression of determinant and efficiency of *J*-point *kk*-design can be applied for the equi-spaced *kk*-design such as Sitter's robust design. Assume that the bids are designed to be equi-spaced. Let h_J be the distance between adjacent bid points and suppose that bids are numbered by size from low to high. Then, $d_{\frac{J+1}{2}}=0$, $d_i-d_j=(i-j)h_J$, $d_i+d_j=\left(i+j-(J+1)\right)h_J$, and $d_i=\left[i-(J+1)/2\right]h_J$, the determinant is

$$\det\left[I(\mu,\beta)\right]_J = \frac{n^2}{2}\sum_{i=1}^{J}\sum_{j=1}^{J}\exp\left\{\frac{\beta}{\beta_0}(i+j-J-1)h_J\right\}\left\{\frac{\beta}{\beta_0}\frac{(i-j)h_J\exp\{\beta(\mu-\mu_0)\}}{\tilde{A}_i\tilde{A}_j}\right\}^2$$

where $\tilde{A}_i = \exp\left\{\frac{\beta}{\beta_0}\left[i-(J+1)/2\right]h_J\right\}+\exp\{\beta(\mu-\mu_0)\}$.

B.3 The general form of bias in the logit regression (Copas 1988)

Taylor expansion of the score function is

$$0 = S_j(\hat{\theta}) = S_j(\theta)+(\hat{\theta}-\theta)' H_j + \tfrac{1}{2}(\hat{\theta}-\theta)' L_j(\hat{\theta}-\theta).$$

where S_j is j^{th} component of the score, H_j is the j^{th} column of the Hessian matrix, and L_j is third derivative of the log-likelihood. The expectation of the above equation gives

$$bias' H_j \simeq \frac{1}{2} tr\left(H^{-1}L_j\right) = h_j$$

since $E\left[S_j(\theta)\right]=0$ and $E\left[(\hat{\theta}-\theta)'(\hat{\theta}-\theta)\right] = Var(\hat{\theta}) = -H^{-1}$, where $bias \equiv E\left[\hat{\theta}-\theta\right]$.

An approximate expression for the bias of maximum likelihood estimate is

$$bias \simeq H^{-1}h$$

where h is the vector of h_j. The sth element of bias is

$$bias_s \simeq \frac{1}{2}\sum_j\sum_k\sum_l H^{sj}H^{kl}L_{jkl}$$

where H^{jk} is the inverse of $H=\{H_{jk}\}$. For the single-bound logit model, each element is calculated as

$$H_{jk} = -\sum_i x_{ij}x_{ik}p_i(1-p_i) = -\sum_i x_{ij}x_{ik}w_i = -X'\Omega X$$

and

$$L_{jkl} = -\sum_i x_{ij} x_{ik} x_{il} w_i \left(1 - 2p_i\right).$$

As a simple example with single covariate, the bias is

$$bias = \frac{\sum x_i^3 w_i \left(2p_i - 1\right)}{2\left(\sum x_i^2 w_i\right)^2}.$$

The bias of β in the symmetric two-point design with known mean can be derived as follows. Suppose that two-point design is symmetric around known mean and observations are equally assigned into each bid. Two design points are $b = \mu \mp d / \beta_0$ where $d > 0$. Let $\pm x = \mu - b = \pm d / \beta_0$. The bias of two-point symmetric designs, now, can be expressed as

$$bias_2 = \frac{n\left\{x^3 w(\beta x)\left[2p(\beta x) - 1\right] + (-x)^3 w(-\beta x)\left[2p(-\beta x) - 1\right]\right\}}{2\left\{n\left[x^2 w(\beta x) + (-x)^2 w(-\beta x)\right]\right\}^2}.$$

Using that $w(\beta x) = w(-\beta x)$, $p(-\beta x) = 1 - p(\beta x) = \left[1 + \exp(\beta x)\right]^{-1}$ and $2p(\beta x) - 1 = -\left[2p(-\beta x) - 1\right]$, the bias reduces to be

$$bias_2 = \frac{\left[\exp(\beta x) - 1\right]\left[\exp(\beta x) + 1\right]}{2Nx\exp(\beta x)} = \frac{\exp(\beta x) - \exp(-\beta x)}{2Nx}.$$

The bias of the two-point symmetric design is an increasing function of *x*. The first derivative of bias with respect to *x* is

$$\frac{\partial bias_2}{\partial x} = \frac{(\beta x - 1)\exp(2\beta x) + (\beta x + 1)}{2Nx^2 \exp(\beta x)}.$$

By construction, $x > 0$ and $\beta > 0$, the first derivative of bias is positive since $(\beta x - 1)\exp(2\beta x) > -1$. Furthermore, since both of the numerator and denominator go to zero as $x \to 0$, using L'Hopital's theorem, the bias is bounded below by $bias_2 > \beta / N$. However, the bias is unbounded when $x \to \infty$.

For the equi-spaced *J*-point *kk*-design, the bias can be expressed as

$$bias_J = \frac{\beta_0 \sum \left\{ h_J \left[i - (J+1)/2 \right] \right\}^3 w_i (2p_i - 1)}{2\left(\sum \left\{ h_J \left[i - (J+1)/2 \right] \right\}^2 w_i \right)^2},$$

using $d_i = \left[i - (J+1)/2 \right] h_J$. Let $\Psi_i \equiv d_i^2 w_i = w_i \left\{ h_J \left[i - (J+1)/2 \right] \right\}^2$, then the numerator is

$$h_J \sum \Psi_i \left[i - (J+1)/2 \right] \left\{ \frac{\exp\left(\frac{\beta}{\beta_0} h_J \left[i - (J+1)/2 \right] \right) - 1}{\exp\left(\frac{\beta}{\beta_0} h_J \left[i - (J+1)/2 \right] \right) + 1} \right\}.$$

The denominator is simply

$$\left(\sum d_i^2 w_i \right)^2 = \left(\sum \Psi_i \right)^2$$

The bias of the equi-spaced *J*-point *kk*-design is

$$bias_J = \frac{\beta_0 h_J}{2n\left(\sum \Psi_i \right)^2} \sum \Psi_i \left[i - (J+1)/2 \right] \left\{ \frac{\exp\left(\frac{\beta}{\beta_0} h_J \left[i - (J+1)/2 \right] \right) - 1}{\exp\left(\frac{\beta}{\beta_0} h_J \left[i - (J+1)/2 \right] \right) + 1} \right\}.$$

C: Gumbel Mixed Model and GEV

C.1 Gumbel mixed bivariate extreme values and reduced extreme values difference

Suppose a sequence of *i.i.d.* pair of random utilities $\{(U_{0i'}, U_{1i'})\}$, $i' = 1, 2, \ldots, m$. Then the bivariate extreme value distribution $\Psi^m(\tau_m + \theta_m v_0, \tau'_m + \theta'_m v_1)$ is an asymptotically approximated distribution of the pair of $(\max U_{0i'} \le \tau_m + \theta_m v_0, \max U_{1i'} \le \tau'_m + \theta'_m v_1)$ with the margins of the Gumbel distribution, where τ_m, τ'_m, θ_m and θ'_m are location and dispersion sequences. By uniform convergence of a continuous function, $\Psi^m(v_0, v_1)$ is approximated by

$$F\left(\frac{v_0 - \tau}{\theta}, \frac{v_1 - \tau'}{\theta'}\right).$$

The asymptotic distribution function is

$$F(v_0, v_1) = \left[F(v_0) F(v_1)\right]^{k(v_1 - v_0)} = \exp\left[-\{\exp(-v_0) + \exp(-v_1)\} k(v_1 - v_0)\right],$$

where $k(\cdot)$ is the dependence function representing the asymptotic connection between $\max U_{0i'}$ and $\max U_{1i'}$. For details of derivation and analytical properties, see Gumbel and Mustafi (1967) and Tiago De Oliveira (1975). Some useful relationships in a bivariate distribution are the boundary conditions $F(-\infty, y) = F(x, -\infty) = F(-\infty, -\infty) = 0$ and the definition of the margins as $F(x, \infty) = F(x)$ and $F(\infty, y) = F(y)$.

The mixed model, one of differentiable bivariate extreme value distributions, has the dependence function defined as

$$k(\tau \mid \lambda) = 1 - \frac{\lambda \exp(\tau)}{\left(1 + \exp(\tau)\right)^2}$$

where τ is reduced difference $v_0 - v_1$. The distribution function of Gumbel mixed model is

$$F(v_0, v_1) = F(v_0) \cdot F(v_1) \exp\left\{ -\lambda \left[\frac{1}{\ln F(v_0)} + \frac{1}{\ln F(v_1)} \right]^{-1} \right\}$$

where the marginal distribution is $F(z) = \exp\left[-\exp(-z)\right]$. The exchangeable distribution, $F(v_0, v_1) = F(v_1, v_0)$, such as the mixed model or logistic model has the symmetric dependence function, $k(-\tau) = k(\tau)$. The parameter λ indicates the association between the two extremes. For $\lambda = 0$, the joint distribution is independent such that $F(v_0, v_1) = F(v_0) F(v_1)$, and generally, the inequality $F(v_0, v_1) > F(v_0) F(v_1)$ holds for dependent case $\lambda > 0$. The correlation coefficient can be expressed as a function of the association parameter λ;

$$\rho(\lambda) = \frac{6}{\pi^2} \left[\arccos\left(1 - \frac{\lambda}{2} \right) \right]^2 \quad (0 \le \rho \le 2/3).$$

When the correlation coefficient is greater than 2/3, the mixed model can not be used.

The reduced difference distribution function is derived using the fact

$$D(w \mid \lambda) = \frac{\exp(\tau)}{1 + \exp(\tau)} + \frac{k'(\tau)}{k(\tau)}.$$

From the dependence function of Gumbel mixed model, the second term in the right-hand side of difference function becomes

$$\frac{k'(\tau)}{k(\tau)}=\frac{\lambda\exp(\tau)\left[\exp(\tau)-1\right]}{\left(1+\exp(\tau)\right)\left\{\left[\exp(\tau)+1\right]^2-\lambda\exp(\tau)\right\}}.$$

Therefore, the distribution function of the difference can be simplified as

$$D(\tau\,|\,\lambda)=\frac{\exp(\tau)}{1+\exp(\tau)}\frac{\left[\exp(\tau)+1\right]^2-\lambda}{\left[\exp(\tau)+1\right]^2-\lambda\exp(\tau)}.$$

The probability density function of reduced difference can be derived by differentiating the difference distribution with respect to τ,

$$\zeta(\tau)=\frac{\partial D(\tau)}{\partial\tau}=\frac{\exp(\tau)}{\left[1+\exp(\tau)\right]^2}+\frac{k''(\tau)k(\tau)-\left[k'(\tau)\right]^2}{\left[k(\tau)\right]^2}.$$

Since $k(\tau)$ is symmetric and differentiable, dependence function satisfies

$$k(\tau)=k(-\tau),\ \left[k'(\tau)\right]^2=\left[k'(-\tau)\right]^2 \text{ and } k''(\tau)=k''(-\tau).$$

This implies the second part of the probability function of reduced difference is symmetric, thus the probability density function is symmetric, $\zeta(\tau)=\zeta(-\tau)$. Algebraic description of the probability density function of reduced difference is as follows. Since the last term of the right-hand side is

$$\frac{k''(\tau)k(\tau)-\left[k'(\tau)\right]^2}{\left[k(\tau)\right]^2}$$
$$=-\frac{\lambda\exp(\tau)}{\left(1+\exp(\tau)\right)^2}\left\{\frac{\left[\exp(2\tau)+1\right]\left[\exp(\tau)-1\right]^2-2\exp(2\tau)\left[4-\lambda\right]}{\left[\left(1+\exp(\tau)\right)^2-\lambda\exp(\tau)\right]^2}\right\},$$

by rearranging terms, the probability density function becomes

$$\zeta(\tau)=\frac{\exp(\tau)}{\left[1+\exp(\tau)\right]^2}\left\{\frac{\left(1+\exp(\tau)\right)^4-\lambda^2\exp(2\tau)-\lambda\left(1-\exp(2\tau)\right)^2}{\left[\left(1+\exp(\tau)\right)^2-\lambda\exp(\tau)\right]^2}\right\}.$$

Bivariate distribution function of the logistic model is

$$F(x,y\mid\lambda)=\exp\left\{-\left[\exp\left(-\frac{x}{1-\lambda}\right)+\exp\left(-\frac{y}{1-\lambda}\right)\right]^{1-\lambda}\right\}\text{ for }0\leq\lambda\leq 1,$$

since the dependence function is defined as

$$k(\tau\mid\lambda)=\frac{\left[1+\exp\left(-\tau/(1-\lambda)\right)\right]^{1-\lambda}}{1+\exp(-\tau)}.$$

The distribution of the difference is

$$D(\tau\mid\lambda)=\left[1+\exp\left(-\frac{\tau}{1-\lambda}\right)\right]^{-1}.$$

The logistic model is independent for $\lambda=0$ and dependent for $\lambda=1$ with diagonal case. As can be seen, the logistic model is the simple version of the generalized extreme value distribution that has been widely used in the transportation and recreational site choice literatures.

Since the mixed and logistic models converge in the same independent case, they are called nonseparated models. The choice of the appropriate model is statistically important and also affects the estimation result. Tiago de Oliveira (1983) suggests the decision rule using the rejection region of independence test in each model. So far, the logistic model of the multivariate case has been used in the environmental economics but there is no application of the mixed model.

C.2 Nested logit and paired combinatorial models with GEV distribution

The nested logit model is obtained by assuming the error term (the vector of unobserved utility) has a type of generalized extreme value distribution

$$F\left(\varepsilon_{jn}\right)=\exp\left(-\sum_{k=1}^{K}\left(\sum_{j\in B_k}\exp\left(-\varepsilon_{jn}/\lambda_k\right)\right)^{\lambda_k}\right).$$

As well known, for this generalized extreme value distribution, any two alternatives in the same nest are correlated but these in different nests are uncorrelated. The choice probability for alternative *i* is

$$P_{in}=\frac{\exp\left(V_{in}/\lambda_k\right)\left[\sum_{j\in B_k}\exp\left(V_{jn}/\lambda_k\right)\right]^{\lambda_k-1}}{\sum_{l=1}^{K}\left[\sum_{j\in B_l}\exp\left(V_{jn}/\lambda_l\right)\right]^{\lambda_l}}.$$

Parameter λ_k is a measure of the degree of independence in unobserved utility among the alternatives in the nest *k*.

Now suppose that the current state and the proposed change of the environment can be recognized either of in the same nest or different nests. In the case of recognizing them in the same nest, there exists only one nest and the probability of choosing one collapses to

$$P_{1n}=\frac{\exp\left(V_{1n}/\lambda_k\right)}{\sum_{j\in B_k}\exp\left(V_{jn}/\lambda_k\right)}=\frac{\exp\left(V_{1n}/\lambda_k\right)}{\exp\left(V_{1n}/\lambda_k\right)+\exp\left(V_{0n}/\lambda_k\right)}.$$

When they are in different nests, the probability is

$$P_{1n}=\frac{\left[\exp\left(V_{1n}/\lambda_k\right)\right]^{\lambda_k}}{\sum_{l=1}^{2}\left[\exp\left(V_{jn}/\lambda_l\right)\right]^{\lambda_l}}=\frac{\exp\left(V_{1n}\right)}{\exp\left(V_{1n}\right)+\exp\left(V_{0n}\right)}.$$

Therefore, λ_k is not identified and the parameters are indistinguishable between two cases.

The general paired combinatorial model with *J* alternatives is proposed by Chu (1981, 1989) as follows

$$P_{in} = \frac{\sum_{j \neq i} \exp\left(V_{in} / \lambda_{ij}\right)\left[\exp\left(V_{in} / \lambda_{ij}\right) + \exp\left(V_{jn} / \lambda_{ij}\right)\right]^{\lambda_{ij}-1}}{\sum_{k=1}^{J-1} \sum_{l=k+1}^{J} \left[\exp\left(V_{kn} / \lambda_{kl}\right) + \exp\left(V_{ln} / \lambda_{kl}\right)\right]^{\lambda_{kl}}} .$$

where a parameter λ_{ij} indicates the degree of independence between alternatives *i* and *j*. For the contingent valuation, since only two alternatives explicitly exist, the general model can be simplified to be

$$\begin{aligned} P_{n1} &= \frac{\exp\left(V_{1n} / \lambda_{12}\right)\left[\exp\left(V_{1n} / \lambda_{12}\right) + \exp\left(V_{2n} / \lambda_{12}\right)\right]^{\lambda_{12}-1}}{\left[\exp\left(V_{1n} / \lambda_{12}\right) + \exp\left(V_{2n} / \lambda_{12}\right)\right]^{\lambda_{12}}} \\ &= \frac{\exp\left(V_{1n} / \lambda_{12}\right)}{\left[\exp\left(V_{1n} / \lambda_{12}\right) + \exp\left(V_{2n} / \lambda_{12}\right)\right]} \end{aligned} .$$

The probability of paired combinatorial model with $J = 2$ is equivalent to the generalized nested logit with two alternatives, resulting the simple logit model. Since alternatives are only two and the scale and level of the utility is immaterial, no covariance parameter can be estimated for the dichotomous choice contingent valuation in the generalized model such as nested logit and paired combinatorial model. Note that these models have only $J\left(J-1\right)/2-1$ covariance parameters after normalization.