Department of Agricultural and Resource Economics
Division of Agriculture and Natural Resources
University of California at Berkeley

Working Paper No. 798

THE STATISTICAL ANALYSIS OF DISCRETE-RESPONSE CV DATA

by

W. Michael Hanemann

and

Barbara Kanninen

California Agricultural Experiment Station
Giannini Foundation of Agricultural Economics
June, 1996

11 **The Statistical Analysis of Discrete-Response CV Data**

by

W. Michael Hanemann

and

Barbara Kanninen

FINAL VERSION
June, 1996

To appear in *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EC, and Developing Countries*. edited by Ian J. Bateman and Ken G. Willis. Oxford: Oxford University Press (forthcoming).

TABLE OF CONTENTS

1. INTRODUCTION AND OVERVIEW 1

2. ECONOMIC THEORY AND STATISTICAL MODELS 2
- 2.1 Types of Discrete Response Format 2
- 2.2 The Economic Foundations of a Statistical Model 4
- 2.3 Utility Theoretic Restrictions 11
- 2.4 Welfare Evaluation 18
- 2.5 Summary 23

3. STATISTICAL ESTIMATION AND DESIGN 23
- 3.1 The Maximum Likelihood Approach 23
- 3.2 Asymptotic Properties of Maximum Likelihood Estimators 25
  - 3.2.1 Consistency 25
  - 3.2.2 Efficiency 26
  - 3.2.3 Calculating the Variance of WTP 26
- 3.3 Checking Model Misspecification 28
  - 3.3.1 Model Checking 30
  - 3.3.2 Hypothesis Tests 33
  - 3.3.3 Goodness of Fit Measures 34
- 3.4 Alternatives to Maximum Likelihood Estimation 35
- 3.5 Optimal Experimental Design 37
  - 3.5.1 Bias and Variance in the Logit Model 38
  - 3.5.2 Basic Principles of Optimal Design 39
  - 3.5.3 Sequential Design 41
  - 3.5.4 Concluding Comments on Optimal Design 41

4. ADDITIONAL TOPICS 42
- 4.1 Valuing Multiple Items 42
  - 4.1.1 Hedonic Utility Formulations 43
  - 4.1.2 Multivariate WTP Distributions 45
  - 4.1.3 Contingent Ranking and Related Techniques 49
  - 4.1.4 Random Coefficient Models 58
- 4.2 Double-Bounded CV 62
- 4.3 Extending the Response Probability Model 69
- 4.4 Nonparametric and Semiparametric Models 75
  - 4.4.1 Single-Bounded Data 75
  - 4.4.2 Double-Bounded Data 77
  - 4.4.3 Nonparametric Welfare Measurement 78
  - 4.4.4 Concluding Observations on Nonparametric Estimation 80

4.5 Analysis of Nonresponse Bias 81
  - 4.5.1 Unit Nonresponse 81
  - 4.5.2 Item Nonresponse 83

Appendix: Maximum Likelihood Algorithms 90

11 **The Statistical Analysis of Discrete-Response CV Data**

MICHAEL HANEMANN AND BARBARA KANNINEN[1]

1. INTRODUCTION AND OVERVIEW

In recent years, increasing attention has been given to the statistical aspects of contingent valuation (CV) survey design and data analysis. The main reason for the growing interest in statistical issues is the shift in CV practice from using an open-ended question to ask about willingness to pay (WTP) to using a closed-ended question. The open-ended format confronts respondents with a question on the lines of "What is the most you would be willing to pay for ..?" The closed-ended format uses a question like "If it cost $x to obtain... , would you be willing to pay that amount?" The closed-ended format was introduced into CV by Bishop and Heberlein (1979). Since the mid-1980's it has gained widespread acceptance as the preferred way to cast the valuation question, a position that was recently endorsed by NOAA's Blue Ribbon Panel on CV (Arrow et al., 1993). For reasons mentioned in Section 5, we share this preference against the open-ended approach and in favor of the closed ended approach. However, the closed-ended approach creates a heavy demand for statistical technique. Statistical issues are not pressing with the open-ended format because the data essentially speak for themselves—the survey responses yield a direct measure of WTP which requires little or no further analysis. With the closed-ended format, by contrast, the CV responses are not dollar amounts but answers of "yes" or "no," and one obtains a WTP value from these responses by introducing a statistical model that links them to the dollar amounts which people faced in the survey.

Since the CV responses are binary variables, one needs a statistical model appropriate for a discrete dependent variable. Such models have come to play an important role in many areas of micro-economics since the 1970s. A major reason was the growth in the availability of disaggregated survey data on specific choices by individual agents—when economic behavior is observed at that level, the outcomes often turn out to be discrete variables, or mixed discrete/continuous variables, rather than purely continuous variables. Some of the most important statistical models for discrete dependent variables used in field like labor economics were first developed in the field of biometrics—logit and probit are obvious examples—but they found ready application to microeconomic data. Discrete response CV is no exception—logit and probit play a key role in the analysis of CV data. Moreover, there is a direct analogy between the response-generating mechanism in some biometrics applications and that in a CV survey. A common setting in biometrics is the dose-response study, where the stimulus is an administered dose of some substance and the outcome is a discrete measure of health status (e.g., alive or dead). The aim is to measure the probability distribution of susceptibility to the substance in question. In a CV survey, the monetary amount presented to subjects (the "bid") can be thought of as the dose, and the yes/no reply as the response; the equivalent of susceptibility is the WTP amount. Another fruitful analogy is with reliability analysis and life testing in industrial engineering, where equipment is stressed until it fails. Sometimes one does not know when it failed, only that it failed after more than 7,000 hours, say, or less than 10,00 hours. The aim is to estimate a probability distribution for the true life of the item—analogous to the true WTP in a CV setting—using information on upper or lower bounds. Statistically, the data look just like responses from a discrete-response CV survey. In CV the bids are subject to the control of the

[1] We gratefully acknowledge helpful comments and suggestions from Anna Alberini, Ian Bateman, Richard Carson, Anni Huhtala, and Kerry Smith, as well as excellent research assistance by Craig Mohn. We are deeply indebted to Ian Bateman for his great kindness and forbearance.

researcher, just as the dose or stress are subject to control by the researcher in dose-response and life-testing experiments. Thus, one can apply to the selection of CV bids statistical techniques for optimal experimental design that have been developed in biometrics and reliability analysis. These analogies with other branches of applied statistics have substantially transformed the statistical side of CV research in recent years.

In this chapter we survey the statistical issues associated with discrete-response CV. Our aim is to make the reader aware of the recent developments in the design and analysis of CV surveys. A major theme is the close interplay between economics and statistics. The CV responses are analyzed using statistical models. But, the models must make sense from the point of view of economic theory. This places significant restrictions on the statistical models that can be used. These implications were first developed by Hanemann (1984a), but his analysis was incomplete and incorrect in some particulars. Here we extend his analysis and correct the errors. We show that many of the models in the current literature violate some of the restrictions of economic theory. Another major theme is the value of richness in the statistical models in order to accommodate heterogeneity of preferences in the population of interest and, also, heterogeneity of response behavior in a survey setting. Richness typically requires introducing additional parameters into the model; the question is how this should be done. Again, there are interesting interactions between the statistics and the economics. Structural models provide the most information from an economic point of view, but they can be fragile if the structure is misspecified. Nonparametric models are more robust and offer greater flexibility in the shape of the response function, but they provide less economic information. Moreover, which economic welfare measure one wants to use can affect the choice of statistical methodology; for example, nonparametric models generally have more trouble providing an estimate of mean than median WTP.

This chapter consists of three main sections. Section 2 presents the economic theory of responses to CV survey questions, set within the framework of a random utility maximization model, and shows how this generates statistical models for the survey responses. Section 3 reviews the statistical issues that arise in the estimation of economic welfare measures from data on CV responses using the maximum likelihood method; it also describes procedures for selecting the bids to be used in CV surveys and for interpreting the results after the data have been collected, based on the principles of optimal experimental design. Section 4 considers some advanced topics in the statistical design and analysis of discrete-response CV data that are both of practical interest and on the current frontiers of research. Section 5 offers some brief conclusions.

2. ECONOMIC THEORY AND STATISTICAL MODELS

2.1 Types of Discrete Response Format

We start by reviewing the different types of closed-ended question that could be used in a CV survey. In what follows, we denote the object of the valuation by q.[2] The CV study could involve just two levels of q or multiple levels of q, depending on whether the goal is to value a single program, to value multiple programs, or indeed to estimate an entire valuation function.

When valuing a single program, one asks people about their WTP to obtain a change from q^0 (the

[2] Depending on the circumstances, q could be a scalar or a vector.

status quo) to some particular alternative q^1. With closed-ended questions, there are *single-bounded* and *multiple-bounded* ways to do this. The single-bounded approach is the original format of Bishop and Heberlein (1979) where respondents were presented with a specific dollar cost to obtain a commodity and asked whether they would be willing to pay this amount. With a public program, the question can be framed in terms of whether they would vote for or against the program if it involved a specified increase in their taxes.[3] Their response provides qualitative information in the form of a bound on their WTP for the item—a lower bound if they answer "yes," and an upper bound if they answer "no." The double-bounded format, proposed by Hanemann (1985a) and Carson (1985) and first applied by Carson and Steinberg (1990) and Hanemann, Loomis and Kanninen (1991), follows up on the initial question with a second question, again involving a specific dollar cost to which they can respond with a "yes" or a "no." The amount presented in this second bid depends on their response to the first bid: if they answered "no" to the first bid the second bid is some lower amount, while if they answered "yes" it is some higher amount. Consequently, if they answer "yes" to one of the questions and "no" to the other, this provides both upper and lower bounds on their WTP for the change. Similarly, the triple-bounded format has an additional follow-up question with a bid amount that depends on the responses to the first two questions (Bateman, Langford and Rasbash, this volume; Cooper and Hanemann, 1995). The additional bids can lead to sharper bounds on the estimate of WTP.

Other formulations of the survey question also generate bounds on WTP. For example, subjects can be shown a *payment card* listing various dollar amounts and asked to circle the one that comes closest to their own value (Mitchell and Carson, 1981). The response can be interpreted not as an exact statement of WTP but, rather, as an indication that the WTP lies somewhere between the highest number below the amount circled and the smallest number above it (Cameron and James, 1987). Equivalently, they can be shown a list of dollar amounts and asked to mark off the amounts they are sure they would be willing to pay and those they are sure they would not (Welsh and Bishop, 1993). Similarly, they can be presented with alternative dollar *ranges* and asked to identify the one that comes closest to their own value. These approaches all yield upper and lower bounds on WTP for the change in q.

When dealing with multiple programs, one can proceed similarly, comparing each program q^1, q^2, q^3, ... in turn with the baseline status quo, q^0, using either a single- or multiple-bounded format. This is like valuing a single program, but repeated many times over. Alternatively, one can ask respondents to assess several programs simultaneously. With *contingent ranking*, introduced by Beggs et al. (1981) and Chapman and Staelin (1982), respondents are presented with a set of programs each involving a specific action and a specific cost, and asked to rank them.[4] For example, they might be offered programs to save 5,000 acres of wetlands at a cost of \$30 per taxpayer, 10,000 acres at a cost of \$60 per taxpayer, and 15,000 acres at a cost of \$150 per taxpayer, and asked to rank these from most to least preferred.

In each case, whether it is yes/no or a ranking, the response to the CV question provides only qualitative information about WTP. Thus, from the raw responses alone, one cannot obtain a quantitative

---

[3] Because of this, the closed-ended format is sometimes called *referendum* CV. Strand (1981) appears to have been the first CV study to use the frame of voting.

[4] If q is a vector, one must ask about not only different *levels* of q but also different *combinations* of the elements of q, in order to span the full range covered by the multi-attribute valuation function. When valuing a single program, by contrast, it matters little whether q is a scalar or a vector because one is simply asking for a global valuation of q^1 versus q^0.

measure of WTP. To obtain that, one must embed the data in a model relating the responses to the monetary stimuli that induced them. How this is done—what models to use, and how then to derive quantitative measures of value—is the subject of this section.

2.2 The Economic Foundations of a Statistical Model

Statistically, the CV responses are discrete dependent variables since they are measured on a nominal or ordinal scale. There is a variety of statistical models for analyzing such data.[5] While we consider specific models below, their common structure can be summarized as follows. In the general case, the CV responses can assume a finite number of values, which we index j = 1,..,M. For the i^{th} observed response, the probability that it takes a particular value can be expressed as some function

$$\Pr\{\text{ response}_i = j \} = H_j\,(A_i;\, Z_i;\, \gamma) \tag{1}$$

where A_i is the bid on that occasion, Z_i represents other covariates describing the subject, the item being valued, or any other pertinent aspect of the survey, and γ is a vector of parameters to be estimated from the data. In order for the probabilities to be well defined, the right hand side (RHS) of (1) must return a value between zero and one, and it must sum to unity over all possible outcomes j = 1,..,M. In binary response models where there are just two possible outcomes, "yes" and "no," (1) reduces to

$$\begin{aligned} &\Pr\,\{\text{response is "yes"}\} = H(A;\, Z;\, \gamma) \\ &\Pr\,\{\text{response is "no"}\} = 1 - H(A;\, Z;\, \gamma). \end{aligned} \tag{1'}$$

It is common to write the $H(\cdot)$ function as the composition of two subfunctions

$$H(A;\, Z;\, \gamma) \equiv 1 - F[T(A;\, Z;\, \gamma)], \tag{2}$$

which then permits the statistical response model to be cast in the form

$$\text{Response} \;=\; \begin{cases} \text{"yes"} & \text{if } T(A;\, Z;\, \gamma) - \eta \geq 0 \\ \text{"no"} & \text{otherwise,} \end{cases} \tag{3}$$

where $T(\cdot)$ is some function of A and Z, η is some random variable with cumulative distribution function (cdf) $F(\cdot)$, and γ represents both coefficients associated with $T(\cdot)$ and parameters of the cdf. The composition ensures that the RHS of (1') returns a value within the range [0,1]. As we show below, different discrete response models involve different formulas on the RHS of (1) or (1').

This statistical perspective can be distinguished from what might be called the economic perspective, which requires that the survey responses be economically meaningful in the sense that they constitute a utility-maximizing response to the survey question. To satisfy both perspectives, one wants

---

[5] Econometric texts with good coverage of these models include Maddala (1983), Amemiya (1985), and Greene (1993). More specialized references include Collett (1991) and Morgan (1992); the latter calls these "quantal response" models.

to formulate a statistical model for the CV responses that is consistent with an economic model of utility maximization. In this section, we sketch how that is done.

We assume an individual consumer with a utility function defined over both market commodities, denoted x, and some non-market item which is to be valued, denoted q; perhaps there are other arguments that shift her preferences for x and q such as attributes of the individual or the market goods, which we denote s. The corresponding indirect utility function depends on the prices of the market goods, p; the individual's income, y, and her characteristics, s; and the non-market item q.[6] The other key component of the indirect utility function is a stochastic component representing the notion of random utility maximization (RUM). It is the RUM concept which provides the link between a statistical model of observed data and an economic model of utility maximization. In a RUM model it is assumed that, while the individual knows her preferences with certainty and does not consider them stochastic, they contain some components which are unobservable to the econometric investigator and are treated by the investigator as random (Hanemann, 1984b). These unobservables could be characteristics of the individual and/or attributes of the item; they can stand for both variation in preferences among members of a population and measurement error. We represent the stochastic component of preferences by ε without specifying whether it is a scalar or a vector, so that the indirect utility function is $v(p,q,y,s,\varepsilon)$.

To fix ideas, we focus on the valuation of a single program using the single-bounded approach.[7] Thus, the individual is confronted with the possibility of securing a change from q^0 to $q^1 > q^0$. We assume she regards this as an improvement, so that $v(p,q^1,y,s,\varepsilon) \geq v(p,q^0,y,s,\varepsilon)$. She is told this change will cost \$A, and she is then asked whether she would be in favor of it at that price. By the logic of utility maximization, she answers "yes" only if $v(p,q^1,y\text{-}A,s,\varepsilon) \geq v(p,q^0,y,s,\varepsilon)$, and "no" otherwise. Hence,

$$\text{(4)} \qquad \Pr\{\text{response is "yes"}\} = \Pr\{v(p,q^1,y-A,s,\varepsilon) \geq v(p,q^0,y,s,\varepsilon)\}.$$

An equivalent way to express this same outcome uses the compensating variation measure, which is the quantity C that satisfies

$$\text{(5)} \qquad v(p,q^1,y-C,s,\varepsilon) = v(p.q^0,y,s,\varepsilon).$$

Thus, $C = C(p,q^0,q^1,y,s,\varepsilon)$ is her maximum WTP for the change from q^0 to q^1. It follows that she answers "yes" if the stated price is *less* than this WTP, and "no" otherwise. Hence, an equivalent condition to (4) is

$$\text{(6a)} \qquad \Pr\{\text{response is "yes"}\} = \Pr\{ C(p,q^0,q^1,y,s,\varepsilon) \geq A \}.$$

In a RUM model, $C(p,q^0,q^1,y,s,\varepsilon)$ itself is a random variable—while the respondent's WTP for the change in q is something that she herself knows, it is something that *the investigator* does not know but treats as a random variable. Let $G_C(\cdot)$ be what the investigator assumes is the cdf of C, and $g_C(\cdot)$ the corresponding

---

[6] Following the suggestion in McFadden and Leonard (1993), the variable y could be *supernumerary* income that remains after allowing for committed expenditures on market or non-market goods; i.e., $y \equiv \bar{y} - \kappa$, where $\bar{y}$ is full income and κ is committed expenditure. The parameter κ could be known exogenously or it could, in principle, be estimated from the CV responses.

[7] The double bounded approach and the valuation of multiple items are taken up in Section 4.

density function (the investigator will estimate the parameters from the CV data); then, (6a) becomes

(6b) $$\Pr\{\text{response is "yes"}\} = 1 - G_C(A).$$

Equations (4) and (6a,b) constitute not only an economic model of respondent behavior but also a statistical model, since the RHS defines a particular form of the H(·) function in (1'), viz. $H(A;Z;\gamma)) \equiv 1 - G_C(A)$. In effect, this represents the integrability condition for the single-bounded case: the statistical model (1') is consistent with an economic model of maximizing behavior if and only if the RHS of (1') can be interpreted as the cdf of a random WTP function, $C(p,q^0,q^1,y,s,\varepsilon)$.[8] To simplify the notation, we now suppress the arguments (p,s) and write the utility and WTP functions as $v(q,y,\varepsilon)$, and $C(p,q^0,q^1,y,\varepsilon)$.

There are two ways to formulate a statistical model with this property. One approach, associated with Cameron (1988), is to directly specify a particular cdf for the individual's random WTP. Let E{C} = μ (in a regression model one would have $\mu = X\beta$), let var{C} = σ^2, and let G(·) be the cdf of the standardized variate $z = (C-\mu)/\sigma$; then

(7) $$\Pr\{\text{response is "yes"}\} = 1 - G_C(A) = 1 - G\left(\frac{A - \mu}{\sigma}\right).$$

For example, when $G(x) = \Phi(x)$, the standard normal cdf, one has a probit model

(8) $$\Pr\{\text{response is "yes"}\} = \Phi\left(\frac{\mu - A}{\sigma}\right),$$

while $G(x) = [1 + e^{-x}]^{-1}$, the standard logistic, yields a logit model

(9) $$\Pr\{\text{response is "yes"}\} = \frac{1}{1 + e^{\frac{A-\mu}{\theta}}}$$

where $\theta \equiv \sigma\sqrt{3}/\pi$.[9]

The other approach, emphasized by Hanemann (1984a), starts by specifying a particular indirect utility function $v(q,y,\varepsilon)$ and a particular cdf for ε, and then constructs the corresponding $G_C(\cdot)$. For example, the following Box-Cox utility function nests many of the models used in the existing literature

(10a) $$u_q = \alpha_q + \beta_q\left[\frac{y^\lambda - 1}{\lambda}\right] + \varepsilon_q, \qquad q = 0 \text{ or } 1$$

where $\alpha_1 \geq \alpha_0$ and $\beta_1 \geq \beta_0 > 0$.[10] This can be regarded as a form of CES utility function in q and y. The

---

[8] A specific implication of the RUM hypothesis is that $\partial\Pr\{\text{"yes"}\}/\partial A = \partial H(A)/\partial A = -g_C(A) \leq 0$: regardless of the form of the utility function, the "demand function" for the change in q slopes downward.

[9] The cdf of a logistic with scale parameter $\tau > 0$ and location parameter μ is $F(x) = [1 + e^{-(x-\mu)/\tau}]^{-1}$; this has a mean and median of μ, and a variance of $\tau^2\pi^2/3$. In the standard logistic, $\mu = 0$ and $\tau = 1$. In (9), G(·) is taken to be the cdf of $z = (C-\mu)/\theta$.

[10] The coefficients α_q, β_q and λ can be functions of (p,s). Since $y > 0$, the problems raised by Burbidge et al. (1988) with the Box-Cox transformation do not arise.

corresponding formula for WTP is

$$C = y - \left[\frac{\beta_0 y^{\lambda}}{\beta_1} - \frac{\lambda\alpha}{\beta_1} + \frac{\beta_1 - \beta_0}{\beta_1} - \frac{\lambda\eta}{\beta_1} \right]^{\frac{1}{\lambda}} \tag{10b}$$

where $\alpha \equiv \alpha_1 - \alpha_0$ and $\eta \equiv \varepsilon_1 - \varepsilon_0$.[11] McFadden and Leonard (1993) proposed a restricted version of this model with $\beta_1 = \beta_0 \equiv \beta > 0$, yielding

$$C = y - \left[y^{\lambda} - \frac{\alpha}{b} - \frac{\eta}{b} \right]^{\frac{1}{\lambda}} \tag{11}$$

where $b \equiv \beta/\lambda$. This includes as special cases the two utility models used by Hanemann (1984a), the linear model where $\lambda = 1$

$$u_q = \alpha_q + \beta y + \varepsilon_q$$
$$C = \frac{\alpha + \eta}{\beta} , \tag{12}$$

and the logarithmic model where $\lambda = 0$

$$u_q = \alpha_q + \beta \ln y + \varepsilon_q$$
$$C = y\left[1 - e^{-(\frac{\alpha + \eta}{\beta})}\right] . \tag{13}$$

The model is completed by specifying a probability distribution for η. Let $G_\eta(\cdot)$ be this cdf; the response probability formulas for the models (10) - (13) are given by:[12]

$$\Pr\{\text{response is "yes"}\} = 1 - G_\eta\left[\frac{\beta_0 y^{\lambda}}{\lambda} - \frac{\beta_1 (y - A)^{\lambda}}{\lambda} + \frac{\beta_1 - \beta_0}{\lambda} - \alpha \right] \tag{10'}$$

$$\Pr\{\text{response is "yes"}\} = 1 - G_\eta[by^{\lambda} - b(y-A)^{\lambda} - \alpha]. \tag{11'}$$

$$\Pr\{\text{response is "yes"}\} = 1 - G_\eta[-\alpha + \beta A] \tag{12'}$$

$$\Pr\{\text{response is "yes"}\} = 1 - G_\eta\left[-\alpha - \beta \ln\left(1 - \frac{A}{Y}\right)\right] . \tag{13'}$$

---

[11] Note that α_1 and α_0 are not separately identifiable here, only their difference α. Similarly, in this model the cdfs of ε_1 and ε_0 are not separately identifiable, only the cdf of η. Depending on the model structure, there will generally be some identifiability restrictions such as these.

[12] The models in this paragraph all involve an additively random structure, of the general form $u_q = \bar{v}(q,y) + \varepsilon_q$, q = 0 or 1. Given this structure, the individual responds "no" if $\Delta\bar{v} - \eta \geq 0$ where $\Delta\bar{v} \equiv [\bar{v}(q^0,y) - \bar{v}(q^1,y\text{-}A)]$ and $\eta \equiv (\varepsilon_1 - \varepsilon_0)$. Hanemann (1984a) called $\Delta\bar{v}$ a *utility difference*; it corresponds to t(A;Z) in (3). This utility difference formulation does *not* apply when the random term enters the utility model nonlinearly, as in (19) below.

The linear and logarithmic models have been widely used in CV because of their simplicity. However, they impose quite stringent restrictions on the shape of the WTP function: the linear model (12) implies that the income elasticity of C is zero; the logarithmic model (13) implies that it is unity. The Box-Cox models are more flexible in this regard, since they permit other values for the income elasticity of C. The income elasticity is negative when $\lambda > 1$, positive when $\lambda < 1$, and is approximately equal to $(1-\lambda)$. Though more flexible than the linear and logarithmic models, the Box-Cox models provide a relatively restricted shape for the graph of WTP against income; more flexible functional forms (with additional parameters) may be needed. Nevertheless, for convenience we will frequently use (11) in this chapter to illustrate points of modelling methodology.

Observe that, when one uses the linear model with $G_\eta(\cdot)$ a standard normal cdf, (12') becomes

$$\text{Pr}\ \{\text{response is "yes"}\} = \Phi(\alpha - \beta A), \tag{14}$$

which is a probit model that becomes identical to the model in (8) when $\alpha = \mu/\sigma$ and $\beta = 1/\sigma$. This can be verified by inspection of the formula for C: since η has zero mean and unit variance, it follows directly from (12) that $E\{C\} \equiv \mu = (\alpha/\beta)$ and $\text{var}\{C\} \equiv \sigma^2 = 1/\beta^2$. Similarly, when $G_\eta(\cdot)$ is the standard logistic cdf, (12') becomes

$$\text{Pr}\ \{\text{response is "yes"}\} = \frac{1}{1 + e^{-\alpha + \beta A}}\ , \tag{15}$$

which is a logit model that becomes identical to the model in (9) when $\alpha = \mu/\theta$ and $\beta = 1/\theta$.[13] [14] These examples illustrate the point that, with an appropriate choice of distributions, the two alternative approaches to generating a statistical model for the CV responses yield the same formula. This result holds generally. There is no essential difference between the two approaches because any formula for an indirect utility function, $v(q, y, \varepsilon)$, implies a corresponding formula for WTP, $C(q^0, q^1, y, \varepsilon)$. The converse is also true since, from (3), one has

$$\begin{aligned} C &= y - m[q^1, v(q^0, y, \varepsilon), \varepsilon] \\ &\equiv y - K(q^0, q^1, y, \varepsilon) \end{aligned} \tag{16}$$

where $m(q, u, \varepsilon)$ is the expenditure function associated with $v(q, y, \varepsilon)$, and $K(q^0, q^1, y, \varepsilon)$ is the income-compensation function which, as Hurwicz and Uzawa (1971) showed, provides a complete representation of consumer preferences. Thus, $C(\cdot)$ and $v(\cdot)$ are alternative representations of the same preference ordering; any probability distribution of u implies a corresponding probability distribution of C, and conversely.

---

[13] In this case, since η has a variance of $\pi^2/3$, (12) implies that $\text{var}\{C\} \equiv \sigma^2 = \text{var}\{\eta\}/\beta^2 = \theta^2\pi^2/3$.

[14] The probit model arises if ε_1 and ε_0 are normally distributed; if they are iid normal with mean zero and variance 0.5, their difference, η, is standard normal. Similarly, the logit model arises if the ε's are extreme value variates. The cdf of an extreme value variate with location parameter ζ and scale parameter $\tau > 0$ is $F(\varepsilon) = \exp[-\exp(-(\varepsilon-\zeta)/\tau)]$; the mean is $\zeta + 0.5772\tau$ and the variance is $\tau^2\pi^2/6$. In the standard extreme value, $\zeta = 0$ and $\tau = 1$. If ε_1 and ε_0 are independent extreme value variates with separate location parameters ζ_1 and ζ_0 and common scale parameter τ, $\eta = \varepsilon_1 - \varepsilon_0$ has a logistic distribution with location parameter $\mu = \zeta_1 - \zeta_0$ and scale parameter τ. The standard logistic arises when $\zeta_1 = \zeta_0$ and $\tau = 1$.

There are two important implications. First, just as economic theory places certain restrictions on $v(\cdot)$, it also places restrictions on $C(\cdot)$. These must be reckoned with whatever the approach to model formulation. The existing literature—our own work included—has not always paid adequate attention to these restrictions. We discuss them further in the next section.

Second, an essential feature of RUM models is that the stochastic and deterministic components are commingled. In conventional economic analysis, one generally formulates a deterministic economic model and then adds a random term for the purpose of statistical estimation. The random term is assumed to arise outside the economic model and it plays no role in the economic analysis once the model has been estimated. In a RUM model, by contrast, the stochastic component is an essential part of the economic model and it plays a substantive role in the use of the model for both prediction of behavior and evaluation of welfare.

We have already seen an example of this in connection with the variance parameter $\sigma^2 \equiv \mathrm{Var}\{C\}$. Assuming a regression setting where $\mu_i = X_i\beta$ for the i^{th} respondent, the response probability formula in (7) can be written

$$\text{(7')} \qquad \Pr\{i^{th} \text{ respondent says "yes"}\} = 1 - G\left(\frac{A_i}{\sigma} - \frac{X_i\beta}{\sigma}\right),$$

which implies that the term $1/\sigma$ serves as the "coefficient" of the bid amount in the statistical model of respondent behavior. Thus, what might have been thought a purely statistical parameter also has a behavioral significance. This is a general result, not limited to any particular RUM specification.[15] [16] Some insight into the reason for it can be obtained from the Box-Cox utility model; (10b) implies that $\mathrm{var}\{C\}$ is a function of not only the stochastic terms in the utility function but also the structural parameters, such as α, β_q, or λ. Moreover, in (10), (11) and (13), $\mathrm{var}\{C\}$ is a function of income, y, so that C has a heteroscedastic distribution. It is inappropriate, therefore, to think of $\mathrm{var}\{C\}$ as merely a statistical constant. Many other examples of the commingling of stochastic and deterministic components in RUM models will be encountered below.

---

[15] Holding all else constant, changes in σ affect response behavior since, from (7), $\partial\Pr\{\text{"yes"}\}/\partial\sigma \geq (\leq)\ 0$ according as $A_i \geq (\leq)\ \mu_i$. Note that (7') assumes a homoscedastic distribution of WTP where σ is the same across all survey responses. This is not a necessary feature of RUM models; one can formulate models where σ varies as a function of attributes of the individuals, attributes of the item being valued, and/or attributes of the survey administration. Examples appear in several parts of this chapter.

[16] Assuming a homoscedastic WTP distribution, Cameron (1988) points out that σ can be identified from the CV responses only because there is variation in A_i. If the same bid were used for all respondents, σ would be unidentified. In that case, one could still identify the ratio β/σ if X_i varied across respondents; but, one needs variation in A_i to obtain separate estimates of σ and β. Even with variation in A_i, there still can be circumstances that create an identification problem. Suppose subjects believe that the price of the item will actually turn out be different from what is stated in the CV survey and their subjective perception of price is some function of A, denoted $\psi(A)$. Then, the RHS of (7) becomes $1 - G[\psi(A_i)/\sigma - (\mu_i/\sigma)]$. Depending on the form of $\psi(\cdot)$, it may not be possible to identify σ. This happens, for example, if $\psi(A) = \theta A$ for some $\theta > 0$. If something like this is possible, it reinforces the behavioral interpretation of the parameter σ.

In their original paper, Bishop and Heberlein (1979) used the following model for the CV responses:

$$\text{(17)} \qquad \Pr\{\text{response is "yes"}\} = \frac{1}{1 + e^{-\alpha + \beta \ln A}} \ .$$

Like (15), this is a logit model where the response probability is independent of income; the difference is that (15) uses A, whereas (17) uses (ln A). Hanemann (1984a) asserted that (17) was not consistent with economic theory because it was not a valid RUM model. As we now show, that assertion was incorrect.

The definition in (16) implies that $C(q^0, q^1, y, \varepsilon)$ is independent of y if and only if $m_u(q^1, u^0, \varepsilon) = m_u(q^0, u^0, \varepsilon)$. As shown in Chapter 3, this comes about if

$$\text{(18a)} \qquad u = T[y + \theta(q,\varepsilon), \varepsilon]$$

for some increasing function $T(\cdot)$. The resulting formula for C is

$$\text{(18b)} \qquad C = \theta(q^1,\varepsilon) - \theta(q^0,\varepsilon),$$

which implies response probabilities of the form

$$\text{(18')} \qquad \Pr\{\text{response is "yes"}\} = \Pr\{C \geq A\} = \Pr\{\theta(q^1,\varepsilon) - \theta(q^0,\varepsilon) \geq A\}.$$

Any RUM model with response probabilities independent of income must be nested within this structure. By inspection, the linear model (12) fits the structure in (18a), but not other versions of the Box-Cox model. The Bishop-Heberlein model (17) corresponds to the particular version of (18) where

$$\text{(19a)} \qquad C = e^{(\alpha + \varepsilon)/\beta} = e^{\alpha/\beta}\upsilon \ ,$$

where $\upsilon \equiv \exp(\varepsilon/\beta)$. This formula for C is generated by the utility model

$$\text{(19b)} \qquad \begin{aligned} v(q^0, y, \varepsilon) &= y + \delta \\ v(q^1, y, \varepsilon) &= y + \delta + \exp\frac{\alpha + \varepsilon}{\beta} \end{aligned}$$

for some arbitrary δ. The resulting response probability is

$$\text{(19')} \qquad \Pr\{\text{response is "yes"}\} = 1 - G_\varepsilon(- \alpha + \beta \ln A)$$

where $G_\varepsilon(\cdot)$ is the cdf of ε. This model (19) combined with various distributions for $G_\varepsilon(\cdot)$ appears commonly in the CV literature. To obtain (17), Bishop and Heberlein used the logistic distribution for $G_\varepsilon(\cdot)$, which makes υ and C log-logistic. If ε is standard normal, υ and C are lognormal, and (19') becomes

$$\text{(20)} \qquad \Pr\{\text{response is "yes"}\} = \Phi(\alpha - \beta \ln a).$$

If ε has the standard extreme value distribution, υ and C have a cdf that corresponds to the survivor function of a two-parameter Weibull distribution, so that (19') becomes[17]

$$\text{(21)} \qquad \Pr\{\text{response is "yes"}\} = 1 - \exp[\, - e^{\alpha - \beta \ln a}\,].$$

The utility model (19), especially in its log-logistic, lognormal and Weibull versions (17), (20) and (21) respectively, appears frequently in the CV literature.[18] Another very common model is the linear Box-Cox model, (12), in its probit and logit versions (14) and (15). These particular models are so popular because they can readily be estimated with canned programs available in well-known statistical packages such as LIMDEP, SAS, STATA and SYSTAT.[19] While the models (12) and (19) may seem somewhat similar, they are fundamentally different because the random term η enters (12) *additively*, whereas the random term ε enters (19) *nonlinearly*. Thus, if one uses a log-logistic, lognormal or Weibull cdf for $G_\eta(\cdot)$ in (12), the result is *not* (17), (20), or (21) but, rather, models with a different response probability formula. Those are indicated in Appendix Table 1, which summarizes the various parametric statistical models that have been discussed in this section. For reasons we now explain, these models must be modified further in order to make them fully consistent with economic theory.

2.3 Utility Theoretic Restrictions

For convenience of exposition, we continue to focus on the valuation of a change from q^0 to q^1 using the single-bounded approach. We showed in the previous section that the statistical model for these CV responses can be cast in the form of (1) or (6). The economic integrability condition is that the RHS in these equations be interpretable as the survivor function of a random WTP distribution. We now examine some of the implications for the stochastic specification of the RUM models, focusing in particular on the upper and lower bounds on C that are implied by economic theory.

Economic theory implies that a person's maximum willingness to pay for an item is bounded by her income; in terms of the WTP function, the constraint is that

$$\text{(22a)} \qquad C(q^0, q^1, y, \varepsilon) \leq y.$$[20]

---

[17] The two-parameter Weibull with scale parameter $\theta > 0$ and shape parameter $\gamma > 0$ has a survivor function $S(x) = \exp[-(x/\theta)^\gamma]$. Setting $\gamma = 1$ produces the exponential distribution; setting $\gamma = 2$ produces the Rayleigh distribution. When ε has the standard extreme value distribution, the cdf of C corresponds to a Weibull survivor function with $\theta = e^{\alpha/\beta}$ and $\gamma = -\beta$.

[18] In the literature on reliability and life testing these are referred to as survival or life distributions (Kalbfleisch and Prentice, 1980; Nelson, 1982). Being distributions for a non-negative random variable, they are commonly used to model the length of life; hence the name.

[19] Those packages parametrize the log-logistic, lognormal and Weibull distributions in terms of $\mu \equiv E\{\ln C\} = \alpha/\beta$ and $\sigma = 1/\beta$. In order to convert the parameter estimates reported by those packages to our parametrization, one sets $\alpha = \mu/\sigma$ and $\beta = 1/\sigma$.

[20] Two qualifications should be noted. First, we observed earlier that the relevant income variable could be *supernumerary* income rather than full income. Second, the individual's preferences could be such that her WTP has an upper bound which is *less* than y: $C \leq C^{max} < y$. If so, C^{max} should be

In terms of the response probability formula, the constraint is that

(22b) $$\Pr\{\text{response is "yes"}\} = 0 \qquad \text{when } A \geq y.$$

In terms of the RUM indirect utility function, conditions sufficient to ensure (22) are that

(23a) $$\lim_{y \to 0} v(q, y, \varepsilon) = -\infty$$

or that, given any q' and any y' > 0, there exists no q" such that

(23b) $$v(q'', 0, \varepsilon) = v(q', y', \varepsilon).$$ [21]

However, most of the models in the existing CV literature violate these restrictions. In the Box-Cox family, only the logarithmic model (13) satisfies (22). To make the other Box-Cox models satisfy (22) one needs to modify the cdf of η so that, focusing on (11),

(24) $$\eta \leq \eta^{max} \equiv -\alpha + by^{\lambda}.$$

Similarly, the nonlinear model (19) fails to satisfy (22) unless one modifies the cdf of ε so that

(25) $$\varepsilon \leq \varepsilon^{max} \equiv -\alpha + \beta \ln y.$$

There are two ways to generate a cdf that satisfies (24) or (25)—truncating the distribution, and inserting a probability mass (a "spike"). We illustrate these for the nonlinear model (19).[22] Truncating at ε^{max} means changing the distribution of ε from whatever it was originally, say $G_\varepsilon(\cdot)$, to a new cdf, say $\bar{G}_\varepsilon(\cdot)$, defined by[23] [24]

---

substituted for y in the discussion below. In principle, C^{max} could be estimated from the data along with the other model parameters -- see, for example, Ready and Hu (1995).

[21] This implies that the market goods x in the underlying direct utility function $u(x,q,\varepsilon)$, taken as a group, are essential.

[22] For the Box-Cox model, substitute η for ε, η^{max} for ε^{max}, and (by^{λ}) for $\beta \ln y$.

[23] This was suggested by Boyle, Welsh and Bishop (1988) for the Bishop-Heberlein model (17), which they proposed truncating from above at some value $C^{max} < y$.

[24] Ready and Hu (1995) suggest an alternative to truncation which they call "pinching". They pinch the WTP distribution down to some $C^{max} < y$ by replacing the original response probability function that lacks a finite upper bound, denoted by $\tilde{\Pr}$, with a new response probability, denoted by $\bar{\Pr}$, defined by

$$\bar{\Pr}\{\text{response is "yes"}\} = [1 - (A/C^{max})] \cdot \tilde{\Pr}\{\text{response is "yes"}\}.$$

They apply this to the Bishop-Heberlein model (17) and find that it fits their data better than the truncated model. However, when one does this, the utility model is no longer given by (17). The shrinking factor implicitly changes both the utility model and the formula for C, not just the cdf of ε.

$$(26) \qquad \bar{G}_\varepsilon(\varepsilon) = \begin{cases} G_\varepsilon(\varepsilon) \;/\; G_\varepsilon(-\alpha + \beta \ln y) & \varepsilon \leq -\alpha + \beta \ln y \\ 1 & \varepsilon > -\alpha + \beta \ln y\,. \end{cases}$$

Inserting a probability mass means replacing ε with a new random term $\bar{\varepsilon} \equiv \min(\varepsilon, \varepsilon^{max})$, which has a cdf $\bar{G}_\varepsilon(\cdot)$ defined by

$$(27) \qquad \bar{G}_\varepsilon(\varepsilon) = \begin{cases} G_\varepsilon(\varepsilon) & \varepsilon < -\alpha + \beta \ln y \\ 1 & \varepsilon \geq -\alpha + \beta \ln y. \end{cases}$$

This is also called *censoring* the distribution $G_\varepsilon(\cdot)$. The censored distribution has a spike at ε^{max} since, by construction, $\Pr\{\bar{\varepsilon} = \varepsilon^{max}\} = 1 - G_\varepsilon(-\alpha + \beta \ln y)$.[25] In both cases, the response probability formula becomes

$$(28) \qquad \Pr\{\text{response is "yes"}\} = 1 - \bar{G}_\varepsilon(-\alpha + \beta \ln A).$$

A graphical illustration of truncation and censoring may be useful. Figure 1(a) shows the graph of the response probability as a function of the bid amount, A. It has been drawn for a model that fails to satisfy (22b). Figure 1(b) illustrates how this is remedied by truncating the distribution, thus "shrinking" the response probability graph to the horizontal axis at A = y. Figure 1(c) illustrates censoring via the insertion of a spike at A = y. While both approaches produce a response function that satisfies (22b), they have different implications for model estimation. Inserting a spike at ε^{max} is a simple adjustment that can be performed *after* the original response model has been estimated, whereas truncation (and, for that matter, pinching) requires re-estimation of the modified response probability model.[26]

Turning from the upper to the lower bound on C, when we defined WTP in (5) we assumed that the increase in q would be viewed as *an improvement*. Of course, it is an empirical question how people feel about the change in q. For the purpose of further analysis, we now consider three cases: (1) the individual views the change as an improvement; (2) the individual views the change as an improvement or is indifferent to it (i.e., places zero value on it); (3) the individual could view the change as a good thing, or as a bad thing, or be indifferent. We start with the first case which we refer to as the *canonical case*. This is because, as shown below, models for the other two cases can always be constructed by appropriate modification of this canonical case.

In the canonical case, the individual sees the change as an improvement and, with probability one,

$$(29a) \qquad C(q^0, q^1, y, \varepsilon) \geq 0.$$

---

[25] Replacing ε by $\bar{\varepsilon}$ is equivalent to replacing C by $\bar{C} \equiv \min\{\exp[(\alpha+\varepsilon)/\beta], y\}$, a Tobit-like model. There is some confusion over terminology in the CV literature; Ready and Hu (1995) refer to (27) as a truncated model, and to the truncated model (26) as a normalized truncated model.

[26] If the CV survey used bids that exceeded respondents' incomes, then even under censoring it would be necessary to re-estimate the response probability function; but not otherwise.

In terms of the response probability formula, the constraint is that

(29b) $\qquad$ Pr {response is "yes"} = 1 $\quad$ when A = 0.

In terms of the RUM indirect utility function, this is equivalent to requiring that $u^1 \equiv v(q^1, y, \varepsilon)$ *first-order stochastically dominates* $u^0 \equiv v(q^0, y, \varepsilon)$. This stochastic dominance is automatically satisfied by the nonlinear model (19), given that $\alpha > 0$. For the nonlinear family, therefore, the response probability formula (28) meets *both* of the restrictions on the range of C that are imposed by economic theory. However, the Box-Cox family fails to satisfy (29) unless the lower support of the additive stochastic term η is adjusted appropriately. Existing Box-Cox models use distributions for $G_\eta(\cdot)$ which are defined over either $(-\infty, \infty)$ or $[0, \infty)$. The former give C a lower support of $C^{min} = -\infty$. The latter make C^{min} positive. Neither produces $C^{min} = 0$.

With (11), for example, when η has a lower support of $\eta^{min} = 0$, then $C^{min} = y - [y^\lambda - (\alpha/b)]^{(1/\lambda)} > 0$; in this case, therefore, it is necessary to restrict $G_\eta(\cdot)$ so that $\eta^{min} = -\alpha$.[27] There are two ways to accomplish this, both involving truncation. One is to employ a distribution for η with a finite, negative lower support, such as the beta. The beta density function is

$$g_\eta(\eta) = \frac{1}{B(r,s)} \frac{\left(\eta - \eta^{min}\right)^{r-1}\left(\eta^{max} - \eta\right)^{s-1}}{\left(\eta^{max} - \eta^{min}\right)^{r+s-1}}$$

where r > 0 and s > 0 are parameters, and B(r,s) is the beta function.[28] In this case, one sets $\eta^{min} = -\alpha$ in order to satisfy (29), while setting $\eta^{max} = -\alpha + by^\lambda$ to satisfy (22). Then, instead of (11'), the response probability formula becomes, for $0 \le A \le y$,

(30) $$Pr\{\text{"}yes\text{"}\} = \int_{by^\lambda - b(y-A)^\lambda - \alpha}^{by^\lambda - \alpha} \left(\frac{(\eta + \alpha)^{r-1} (by^\lambda - \alpha - \eta)^{s-1}}{B(r,s)\ (by^\lambda)^{r+s-1}} \right) d\eta \ .$$

The other way is truncation of a cdf defined over $(-\infty, \infty)$ such as the normal or logistic.[29] In order to satisfy (22), the cdf would already have been modified at the upper end by truncation or censoring, as in (26) or (27). To satisfy (29) it would now be truncated from below at $-\alpha$. If truncation had been used at the upper end to satisfy (22), the cdf would now be *doubly* truncated to satisfy (29) as well. If $G_\eta(\cdot)$ is the untruncated cdf, the doubly truncated cdf is $[G_\eta(\eta) - G_\eta(-\alpha)]/[G_\eta(-\alpha + by^\lambda) - G_\eta(-\alpha)]$ and, instead of (11'), the response probability formula becomes, for $0 \le A \le y$,

(31) Pr {response is "yes"} = $1 - \{[G_\eta(-\alpha + by^\lambda - b(y-A)^\lambda) - G_\eta(-\alpha)] / [G_\eta(-\alpha + by^\lambda) - G_\eta(-\alpha)]\}$.

If censoring had been used at the upper end to satisfy (22), the cdf would be truncated at the lower end

---

[27] More generally, with any additively random model $\bar{v}(q,y) + \varepsilon_q$, where $\eta \equiv \varepsilon_1 - \varepsilon_0$, it is necessary to restrict $G_\eta(\cdot)$ so that $\eta^{min} = \bar{v}(q^0,y) - \bar{v}(q^1,y) < 0$ in order to have $C^{min} = 0$.

[28] Special cases include the uniform distribution (r=s=1), the arcsine (r=s=0.5), and the half triangular (r=1 and s=2, or r=2 and s=1).

[29] There does not appear to be any way to adjust a $G_\eta(\cdot)$ defined over $[0, \infty)$, such as the lognormal, log-logistic or Weibull, so that it generates a Box-Cox model satisfying (29).

to satisfy (29). With the combination of censoring and truncation, the response probability formula becomes

$$(32) \quad Pr\{\text{"}yes\text{"}\} = \begin{cases} 1 - \{[G_\eta(-\alpha + by^\lambda - b(y-A)^\lambda) - G_\eta(-\alpha)] / [1 - G_\eta(-\alpha)]\} & 0 \le A < y \\ 0 & A \ge y. \end{cases}$$

To summarize, in the canonical case where the individual sees the change as an improvement, economic theory imposes the restriction that $0 \le C \le y$. One way or another, this is violated by most of the RUM models in the existing literature, including those listed in Appendix Table 1. Appendix Table 2 presents some models that do satisfy this requirement, based on (28) and (30) - (32); we shall refer to these as canonical models.

Now suppose one wants to allow for indifference—with some positive probability, the individual has a zero WTP for the change in q. Indifference is equivalent to a probability mass at C = 0. There are at least two ways to introduce this. One is to combine a canonical model, denoted $G_C(\cdot)$, with a degenerate distribution for C consisting of a spike at zero, γ, so as to form the mixture distribution[30]

$$(33) \qquad \Pr\{C \le x\} \equiv \bar{G}_C(x) = \gamma + (1-\gamma)\, G_C(x) \qquad x \ge 0.$$

By construction, $\gamma \equiv \Pr\{C = 0\}$. This formulation can be interpreted as representing a population which consists of two distinct types: a group of people, amounting to 100γ% of the population, who are simply indifferent, for whom $\Pr\{C = 0\} = 1$; and another group which has a varying but positive WTP for the change in q, with a cdf given by $G_C(\cdot)$. For the mixture to make sense, the lower support of $G_C(\cdot)$ should be zero. Hence, $G_C(\cdot)$ should be a canonical model like those listed in Appendix Table 2, rather than a model like those listed in Appendix Table 1. With the mixture approach, the response probability created by the mixture model, which we denote $\overline{\Pr}$, is related to the response probability associated with the canonical model, which we denote $\tilde{\Pr}$, as follows:

$$(33') \qquad \begin{aligned} \overline{\Pr}\{\text{response is "no"}\} &= \gamma + (1-\gamma)\, \tilde{\Pr}\{\text{response is "no"}\} \\ \overline{\Pr}\{\text{response is "yes"}\} &= (1-\gamma)\, \tilde{\Pr}\{\text{response is "yes"}\}. \end{aligned}$$

Any of the formulas in Appendix Table 2 could be used as the formula for $\tilde{\Pr}$. The "downweighting" in the response probability $\overline{\Pr}$ is illustrated in Figure 2; Figure 2(a) depicts the response probability graph associated with a canonical model, while Figure 2(b) shows the response probability graph associated with the mixture model.

---

[30] For a CV application see Kriström, Nyquist and Hanemann (1992). Schmidt and Witte (1989) apply this model to recidivism, calling it a "split population" survival model. In one version of their model, γ is treated as a constant to be estimated from the data; in another, they make γ a function of covariates, Z, by writing $\gamma = [1 + e^{\theta Z}]^{-1}$, where θ are coefficients to be estimated. Werner (1994) shows that, if γ does depend on covariates in the true model, and the canonical WTP distribution also depends on the same covariates (e.g. through α), estimating a misspecified model that treats γ as a constant seriously biases both the estimate of γ and the coefficient estimates for the covariates in the canonical WTP distribution.

The other way to introduce indifference is to use a response probability model with censoring at C = 0. In the case of the Box-Cox model (11), for example, truncation generated the response probability formula in (31). Now, one starts with a random term η defined over $(-\infty, -\alpha + by^{\lambda}]$—for example a normal or logistic distribution that has been modified at the upper end by truncation or censoring, as in (26) or (27)—and one censors rather than truncates this from below. Thus, one replaces η with $\bar{\eta} \equiv \max(\eta, -\alpha)$. The distribution of $\bar{\eta}$ has a spike at $-\alpha$ since, by construction, $\Pr\{\bar{\eta} = -\alpha\} = G_{\eta}(-\alpha)$. This is equivalent to replacing C as defined in (11) with $\bar{C} \equiv \max(C,0)$. This use of censoring to represent indifference is illustrated in Figure 2(c). Response probability formulas for such censored models are given in Appendix Table 3.

The first of these methods for putting a spike at C = 0, the mixture approach (33), can be applied to the nonlinear utility model (19) as well as the Box-Cox model (11); the second approach, based on censoring, requires an additive error and thus can only be applied to the Box-Cox model. The difference is that the former introduces an extra parameter, γ, to account for the spike, while the latter uses the model parameters α, β, and λ to represent both the spike and the rest of the response probability function. The first approach implies a population consisting of two distinct types, as described above, while the second assumes a single group with homogeneous preferences.

Lastly, we consider the case where the individual could view the change in q as either good or bad. As in the previous case, there are several ways to model this. One approach uses a mixture distribution combining *two* models, one representing positive preferences ($C \geq 0$) and the other negative preferences ($C \leq 0$). The positive preferences are represented by a canonical cdf $G_{+}(\cdot)$ defined over [0, y] with $G_{+}(0) = 0$. The negative preferences are represented by a cdf $G_{-}(\cdot)$ defined over the negative domain with $G_{-}(0) = 1$.[31] In addition, there can be a spike at zero to allow for indifference. The resulting mixture distribution is

$$(34) \qquad \Pr\{C \leq x\} \equiv G_C(x) = \gamma_1 G_{-}(x) + \gamma_2 G_{+}(x) + (1 - \gamma_1 - \gamma_2).$$

Here, one can think of the population as consisting of three groups—a fraction γ_1 who dislike the change, a fraction γ_2 who like it, and the remainder who are indifferent.[32]

It is important to note that, when the individual dislikes the change, -C as defined in (5) represents minimum *willingness to accept* to suffer the change, rather than maximum WTP to avoid it.[33] The exception is any utility model like (18) that has no income effects, such as the linear Box-Cox model (12).

---

[31] If one takes the Box-Cox model (11) and sets $\alpha < 0$ instead of $\alpha > 0$, this would generate a model of negative preferences. This does not work with the nonlinear model (19) since that formula yields $C \geq 0$ regardless of the sign of $(\alpha + \varepsilon)/\beta$.

[32] The parameters γ_1 and γ_2 can be taken as constants or made functions of covariates, in the manner described in footnote 30.

[33] As defined in (5), C is the Hicksian compensating variation measure. When C < 0, WTP to avoid the change is measured by the equivalent variation, rather than the compensating variation. The equivalent variation is the quantity E that satisfies $v(q^1, y, \varepsilon) = v(q^0, y + E, \varepsilon)$. It is shown in chapter 3 in this volume that, if C' is the compensating variation for a change from q' to q", and E" the equivalent variation for the reverse change from q" to q', then E" = -C'.

With no income effects, the (negative) WTP and WTA coincide. Therefore, with dispreference in that model, -C must satisfy the restriction that $-C \leq y$ since it measures WTP as well as WTA. Otherwise, in models with non-zero income effects such as the general Box-Cox model (11), when $C < 0$, there is no restriction that $-C \leq y$ since -C measures only WTA. Hence, C is potentially unbounded from below in the negative domain.

Therefore, another approach to modelling both positive and negative preferences is to use the Box-Cox family with a distribution for η that ranges from $-\infty$ to $-\alpha + by^{\lambda}$. This ensures $C^{max} \leq y$ while leaving C unbounded in the negative left tail. This model is intermediate between the Box-Cox models listed in Appendix Tables 1 and 2—like those in Appendix Table 2, η is bounded from above while, like those in Appendix Table 1, it is unbounded from below. Response probability formulas for such models are presented in Appendix Table 4.[34]

The two approaches to modelling both positive and negative preferences are illustrated in Figure 3. The upper panel illustrates the mixture approach, (34); the lower panel illustrates the modified Box-Cox model. In general when one uses a mixture distribution, the mixing parameters—γ in (33), γ_1 and γ_2 in (34)—are either known exogenously or else treated as unknowns to be estimated from the response data along with the other model parameters. The mixture approach provides the most flexibility—it allows for *different* parameters (and different covariates) in the nonstochastic component of the RUM, depending on whether the individual likes or dislikes q. By contrast, the modified Box-Cox model imposes the *same* deterministic tastes on the whole population, leaving the burden of differentiating between preference and dispreference to the distribution of the stochastic component, η.

In this section, we have emphasized the importance of modifying statistical models for CV responses to incorporate restrictions imposed by economic theory on the range of C. Without this, the statistical models are not in fact consistent with the economic hypothesis of utility maximization. As outlined in Appendix Tables 2-4, there are several different possible restrictions, depending on what one assumes about indifference or negative preferences. Two points should be emphasized. First, it is an empirical question which is the appropriate set of restrictions to impose—economic theory *per se* cannot prescribe whether people like, dislike or are indifferent to the change in q. The analyst either has to know ahead of time which of these to assume, taking that as a maintained assumption, or she has to make a point of collecting data which can resolve this, for example by asking respondents directly whether they like or dislike the change.[35] Indeed, if there is a chance that respondents dislike the change, one should ask about this *at the beginning of the survey*, since this permits the CV question to be framed appropriately—those who like the change can be asked their WTP to secure it, while those who dislike the change can be asked their WTP to avoid it.[36] The responses to the initial question can be used to

---

[34] We have noted that, in these models, $C < 0$ measures WTA rather than WTP. Using the result mentioned in the previous footnote, one obtains the equivalent variation measure, E, that corresponds to WTP when $E < 0$ by replacing $(\alpha + \eta)$ in these formulas with $-(\alpha + \eta)$.

[35] Failure to allow for the possibility of negative preferences towards the item being valued has figured in several CV disputes (e.g., Cameron and Quiggin, 1994).

[36] Duffield (1993) and Huhtala (1994) are examples of CV studies which follow this approach. Note that it entails asking the compensating variation for those who like the change, and the equivalent variation for those who dislike it. This difference should be reflected in the response probability models applied to

estimate γ_1 and γ_2 in (34). Similarly, if one believes that people either like the change or might be indifferent, as in (33), one could start off assuming that respondents like it but then ask those who say "no" to the CV bid a follow-up question on the lines of "would you be willing to pay anything at all for this change in q?"[37] The proportion who say "no" provides a direct estimate of γ in (33).

Secondly, the restrictions are inherently non-parametric. As illustrated in Figures 1-3, they constrain the lower and upper ends of the empirical response probability function. Therefore, once one has decided which restrictions to impose, they are consistent with many different statistical models. Although we have focused so far on the particular *parametric* models in Appendix Tables 1 - 4, other parametric models can be used as long as they satisfy (22b) and, if one assumes positive preferences, (29b). Moreover, semiparametric and nonparametric models can also be used. As discussed in Section 4.5, these models effectively censor the WTP distribution at the lowest and highest observed bids, which has the effect of guaranteeing that they satisfy (22b) and (29b).

What are the practical consequences if one disregards the restrictions on the range of C? As shown below, there are two possible effects—the formulas for welfare measures may be wrong and/or the estimates of model parameters may be biased due to the incorrect specification of the response probability model. The welfare measures are discussed in the following section. Estimation of model parameters is discussed in Section 3.

2.4 Welfare Evaluation

So far we have discussed how to formulate statistical models for the CV responses that are consistent with utility theory. We turn now to the question of how one derives useful measures of monetary value once the statistical models have been estimated.[38] As we noted earlier, by virtue of the RUM hypothesis, rather than being a fixed number, a person's WTP for the change is actually a random variable with a cdf given by $G_C(\cdot)$. The survey responses permit one to estimate this WTP distribution. Indeed, this is a basic difference between conventional regression models for continuous dependent variables (e.g., responses to open-ended CV questions) and quantal response models. Conventional regression models estimate the *conditional mean* of the dependent variable given the regressors. Quantal response models for discrete dependent variables estimate the *conditional distribution* of the dependent variable given the regressors, i.e. the entire cdf. The question then arises how one might summarize this distribution for practical purposes. As we show in this section, that is a nontrivial question.

The literature has generally focused on two summary statistics. One is the *mean* of the estimated WTP distribution

$$C^+ \equiv E\{\ C(q^0, q^1, y, \varepsilon)\ \}. \qquad (35)$$

---

the data.

[37] In practice, finding a way to ask this requires some care since one wants to avoid any appearance of pressuring the respondent to express a positive value.

[38] In this section, when we refer to parameters such as α and β, we mean the *estimates* of these parameters. The statistical properties of these parameter estimates are discussed in Section 3.

The other, advocated by Hanemann (1984a), is the *median* of the estimated WTP distribution, i.e., the quantity C* such that

$$1 - G_C\,(C^*) = 0.5. \tag{36}$$

It follows from (6a) that Pr {response is "yes"} = 0.5 when A = C*, i.e. there is a 50:50 chance that the individual would be willing to pay at least C*. In addition, Hanemann (1989) has suggested giving consideration to *other* quantiles of the $G_C(\cdot)$ distribution. The θ-percentile of the WTP distribution, C_θ, satisfies

$$\theta \;=\; 1 - G_C\,(C_\theta); \tag{37}$$

i.e., there is a 100θ% probability that the individual would be willing to pay at least C_θ.[39] By way of example, for the Box-Cox model (11) the mean is

$$C^+ \;=\; y \;-\; E\left(\left[y^\lambda - \frac{\alpha}{b} - \frac{\eta}{b}\right]^{\frac{1}{\lambda}}\right) \tag{38}$$

while the median is

$$C^* = y - \left[y^\lambda - \frac{\alpha}{b} - \frac{\eta^*}{b}\right]^{\frac{1}{\lambda}} \tag{39}$$

where η* is the median of distribution of η. In the linear model (12) where λ = 1, one has $C^+ = (\alpha + E\{\eta\})/\beta$ and $C^* = (\alpha + \eta^*)/\beta$. If $E\{\eta\} = \eta^*$ in that model, $C^+ = C^*$. Otherwise, the two welfare measures are numerically different.

Deciding which measure is appropriate involves considerations of both statistics and economics. Suppose that the survey data came from repeated questioning of a single individual—while this is fairly impractical, one could imagine it happening. In that case, even though we were estimating a single individual's WTP, it still would be a random variable for us as outside observers because of the RUM hypothesis. The issue would then be just one of representation—what is the best way to summarize the probability distribution? The answer depends on the statistical loss function: with a sum-of-squared-errors loss function, the mean is the optimal measure of central tendency; with a sum-of-absolute errors loss function, the median is optimal. For this reason, the mean is more sensitive to skewness or kurtosis in the WTP distribution (Stavig and Gibbons, 1977). This could be important because most RUM models with non-negative preferences imply a skewed distribution of WTP. As shown below, it can often happen that the point estimate of the median is more robust, or has a much smaller sampling error, than the point estimate of the mean.

Now consider the more realistic situation where the survey data come from questioning different individuals in a population. In that case, the summary measure of the WTP distribution would be

[39] The analogs of C_θ in the biometrics literature are what is known as the $ED_{100\theta}$ or $LD_{100\theta}$. These are the dose levels or concentrations at which, on average, 100θ% of subjects respond; *E* denotes "effective," and *L* denotes "lethal." The distribution of which they are the quantiles is called a tolerance distribution. Much of the attention in the biometrics literature is focused on the medians ED_{50} or LD_{50}.

multiplied by the number of people in the population to produce an estimate of aggregate value. Thus, the choice of a summary statistic implies a particular approach to the aggregation of welfare across the population. The mean, C^+, is equivalent to adopting the Kaldor-Hicks potential compensation principle. Suppose there are both positive and negative preferences for the change in q, as in the response probability models listed in Appendix Table 4. Then, $C^+ > 0$ if and only if those who are better off as a result of the change in q could fully compensate those who are worse off, and still gain by it. The Kaldor-Hicks criterion is commonly used, but it can lead to logical inconsistencies and it has been severely criticized on ethical grounds (Little, 1957). As a means of aggregating values, the median C* is equivalent to applying the principle of majority voting: the change is desirable if a majority of the population would vote for it. Using a lower quantile of the WTP distribution would correspond to super-majority voting; for example, $C_{0.67}$ would correspond to requiring a two-thirds majority vote. It is known that majority voting rules do not satisfy Pareto efficiency; but they still may be considered ethically superior. In view of these welfare-theoretic implications, choosing a measure of central tendency is essentially a value judgment. Moreover, different circumstances may call for different welfare measures. For example, Carson, Flores and Hanemann (1992) make a distinction between benefit-cost analysis and natural resource damage assessment, recommending the use of median WTP for the former and mean WTA for the latter because of the difference in the implied property right and the legal requirement in the latter case (but not the former) to restore all those who were injured to their original position. These are judgments that the researcher must make.

The reason why C^+ is more sensitive than C* to skewness in the WTP distribution becomes evident when one considers graphical representations of these welfare measures. The graphical approach permits one to calculate the welfare measures even with a non-parametric model that has no closed-form representation for the response probability function. The median, C*, can be read directly from the empirical response probability function—it is the dollar amount that corresponds to a 50% probability of saying "yes" (see Figure 4a). The graphical representation of C^+ comes from a standard result in statistics about the relation between the mean of a random variable and the integral of its cdf (Parzen, 1960):

(40a) $$C^+ = \int_{-\infty}^{0} G_C(A)dA + \int_{0}^{\infty}[1 - G_C(A)]dA.$$

This corresponds to the shaded areas over/under the empirical response probability function in Figure 4b. When C is restricted to being non-negative, the formula for the mean reduces to

(40b) $$C^+ = \int_{0}^{\infty} [1 - G_C(A)]dA.$$

which corresponds to the shaded area under the response probability function in Figure 4c.

Two implications follow from the graphical representation of welfare measures. The first concerns the relationship between the two welfare measures. Whereas the median depends on the location of the response probability graph at a particular point, viz., the 50% probability level, the mean depends on the location of response probability graph throughout its entire length, from tail to tail. When $C \geq 0$, while small differences in the right tail of the distribution have essentially no effect on the median, they can affect the mean greatly.[40] This explains why the relation between the mean and median can vary with

[40] Boyle, Welsh and Bishop (1988) refer to this as the "fat tails" problem.

the specification of the WTP distribution. By way of illustration, consider the various versions of the nonlinear utility model (19). With the lognormal model (20), application of (36) and (40b) yields the following formulas for median and mean

(41a) $$C* = e^{\frac{\alpha}{\beta}}$$

(41b) $$C^{+} = e^{\frac{\alpha}{\beta}} \exp\left[\frac{1}{\beta^2}\right] .$$

With the log-logistic model (17), while the formula for the median is the same as for the lognormal

(42a) $$C* = e^{\frac{\alpha}{\beta}} \ ,$$

the mean turns out to be

(42b) $$C^{+} = \begin{cases} e^{\alpha/\beta}\ \Gamma[1 + (1/\beta)]\ \Gamma[1 - (1/\beta)] & \text{if} \quad \beta > 1 \\ \infty & \text{if} \quad \beta \le 1. \end{cases}$$

With the Weibull model (21), one has

(43a) $$C* = e^{\alpha/\beta}\ (\ln 2)^{-1/\beta}$$

(43b) $$C^{+} = \begin{cases} e^{\alpha/\beta}\ \Gamma[1 - (1/\beta)] & \text{if} \quad \beta > 1 \\ \infty & \text{if} \quad \beta \le 1. \end{cases}$$

In each case, the ratio C^+/C^* depends crucially on the value of β, which is related to $\omega^2 \equiv \text{var}\{\ln c\}$. In the lognormal case, $\beta = 1/\omega$, and C^+ grows larger with ω but remains finite as long as $\beta > 0$. In the log-logistic case, $\beta = \pi/\omega\sqrt{3}$ and C^+ blows up if $\beta \le 1$, i.e., if $\omega^2 \ge \pi^2/3 \approx 3.29$. In the Weibull case, $\beta = \pi/\omega\sqrt{6}$ and C^+ blows up if $\omega^2 \ge \pi^2/6 \approx 1.65$. In terms of Figure 4c, the integral corresponding to the shaded area fails to converge when $\beta \le 1$. This is yet another example of the fact that the stochastic specification of a RUM can have substantive economic implications, since different probability distributions constrain the relation between mean and median in such different ways.

The second implication concerns the restrictions on the range of C. Welfare formulas such as those in (41a,b) and (42a,b), which were first presented in Hanemann (1984a), or in (43a,b) are in fact incorrect since they ignore the restriction that $C \le y$. When this restriction is imposed, as in Appendix Tables 2 or 3, it will probably not affect the formula for the median, C*, but it certainly will affect the formula for the mean, C^+.[41] This can be seen by decomposing the formula in (40b)

[41] The discussion in this paragraph focuses on the case where $C \ge 0$. Analogous results hold for the case of negative preferences (i.e., for the probability response models in Appendix Table 4).

(44a) $$C^+ = \int_0^y Pr\{\text{"yes"}\}\, dA + \int_y^\infty Pr\{\text{"yes"}\}\, dA$$

(44b) $$= \int_0^y Pr\{\text{"yes"}\}\, dA \quad .$$

When one takes account of the restriction that C ≤ y, the second term on the RHS of (44a) vanishes by virtue of (22b). Equivalently, (40b) can be manipulated into the form

(45) $$C^+ = y - \int_0^y Pr\{\text{"no"}\}\, dA.$$

Hence, C^+ is always finite and bounded from above by the person's income. Corrected formulas for C^+ and C* that incorporate the restrictions on the range of C are included along with the response probability formulas in Appendix Tables 2-3.

Although the existing analytic formulas for C^+ in Hanemann (1984a,1989) incorrectly ignore the restrictions on the range of C, not all of the empirical estimates of C^+ in the literature should be considered erroneous because some researchers used other formulas or procedures when calculating C^+. This started, in fact, with Bishop and Heberlein who truncated the WTP distribution at the highest bid in their survey, A^{max}, when estimating mean WTP. Instead of (44a), they calculated

(46) $$\tilde{C}^+ \equiv \int_0^{A^{max}} Pr\{\text{"yes"}\}\, dA.$$

Since $A^{max} < y$, $\tilde{C}^+$ is a conservative estimate of C^+. Computer programs to calculate $\tilde{C}^+$ are quite readily available, and this welfare measure has been used quite widely in the literature. We show in Section 4.5 that non-parametric estimators of the response probability function estimate the mean in essentially the same way, since they truncate the cdf of C at the highest observed data point. Instead of A^{max}, other researchers have proposed truncating the cdf at some amount, C^{max}, that represents the analyst's assessment of an upper bound on individuals' WTP, perhaps based on non-sample data.[42] [43]

A different type of approach was adopted by Werner (1994), who advocated partitioning the population into groups and computing the average of the medians for each group. The groups could be identified in various ways—according to administrative or political criteria, for example, or on the basis of socio-demographic, behavioral or attitudinal covariates. Suppose there are K groups; each group has a WTP distribution with a cdf, $G_k(C_k)$, that has a mean C_k^+ and a median C_k^*. Depending on how the groups are defined, these distributions would be obtained by fitting response probability functions to the data for each group separately or by fitting a single response probability function to data pooled over all groups but parametrized on group indicator covariates. Werner's statistic, the average of group

---

[42] See Boyle, Welsh and Bishop (1988), and Ready and Hu (1995). Carson et al. (1992) test for the presence of a $C^{max} < y$ and find they cannot reject the hypothesis that $C^{max} = 0.02y$ with their data.

[43] Related approaches are constrained estimation and Bayesian estimation where, at the stage of model estimation rather than when calculating the welfare measure, one imposes a constraint on the location of the right-hand tail of the response probability function or one introduces prior information about the location of this tail.

(conditional) medians, is $C^W \equiv (1/K) \Sigma C^*_k$. Of course, if there is only one group, this measure corresponds to the overall median, C*; if each group has only a single member, it corresponds to the overall mean, C^+. Otherwise, she establishes that C^W generally lies between C* and C^+ and that, for a given sample size, while estimates of C^W do not have quite as low a variance as C*, they have a considerably lower variance than estimates of C^+.

2.5 Summary

Most of the existing CV literature uses response probability models based on either the Box-Cox utility model (11), including its linear and logarithmic versions, or the nonlinear utility model (19). As summarized in Appendix Table 1, depending on the stochastic specification the former gives rise to probit and logit response probability models, while the latter generates lognormal, log-logistic and Weibull models. Most of these models violate restrictions on the range of C from economic theory. This problem can be avoided by using the modified response probability models listed in Appendix Tables 2-4. These generally satisfy the economic restrictions by either censoring (introducing a spike) or truncating the corresponding response probability model in Appendix Table 1. Censoring may not affect parameter estimation -- one can get the same parameter estimates for the censored and uncensored models -- but it does change the formula for the mean WTP and it could affect the median. Truncation invariably affects both parameter estimation and the formulas for mean as well as median WTP. How much difference this makes in practice is an empirical question that remains to be investigated. The choice between the response probability models listed in Appendix Tables 2, 3, and 4 involves a judgment by the analyst as to whether to allow for indifference and/or dispreference for the item in question. The choice of welfare measure -- mean, median, or some other quantile of the WTP distribution -- also calls for a judgment by the analyst that can involve both ethical and statistical considerations. The mean is the conventional measure in benefit-cost analysis and reflects the Kaldor-Hicks potential compensation criterion; the median may be more realistic in a world where decisions are based on voting and there is concern for the distribution of benefits and costs. From a statistical point of view, the mean is generally far more sensitive than the median to the choice of a response probability model or the method of estimation.

3. STATISTICAL ESTIMATION AND DESIGN

This section discusses practical issues in the estimation of CV data based on the maximum likelihood method. Section 3.1 reviews maximum likelihood procedures. Section 3.2 discusses the asymptotic properties of maximum likelihood estimators and approaches for calculating the variance of estimated WTP. Section 3.3 discusses procedures for checking model misspecification including hypothesis testing and goodness of fit. Section 3.4 discusses optimal experimental design.

3.1 The Maximum Likelihood Approach

The method of maximum likelihood seeks the values of the unknown parameters that are most likely to have generated the data that were observed. In the present context, the data consist of yes/no responses from survey participants. Let y_i denote the response of the i^{th} individual—we can think of y_i = 1 as "yes", and y_i = 0 as "no." In addition, we observe a vector of explanatory variables for each respondent that includes the bid amount presented to her, A_i, and any covariates that might have been recorded, such as income (if the utility model exhibits non-zero income effects), socio-demographic variables, attitudinal variables, etc. We now follow the notational convention in statistics by expressing all the exogenous variables for the i^{th} respondent as a vector x_i. Finally, we represent the unknown parameters, such as α, β, and λ, in a single vector θ. The response probability formulas discussed in Section 2, including those in Appendix Tables 1 - 4, can each be thought of as a likelihood function for the i^{th} observation—they express the likelihood (or probability) of observing the response that was observed, y_i, given the exogenous variables x_i and the (unknown) parameters, θ. Following the convention in statistics, we now refer to these response probabilities using the notation $P(y_i \mid x_i, \theta)$. Since the responses from a random sample represent independent observations, the likelihood function for the overall set of responses $\{y_1,...,y_n\}$ given the corresponding set of explanatory variables $\{x_1,...,x_n\}$ and the true but unknown parameters θ is simply the product of the individual likelihood functions for each observation in the sample. Taking the logarithm, the log-likelihood function for the sample is the sum of the individual likelihoods[44]

$$(47) \qquad L(\theta \mid y_1,...y_n \, , \, x_1,...x_n) = \sum_{i=1}^{n} \ln P(y_i \mid x_i, \theta) \; .$$

In forming equation (47), one substitutes for $P(y_i \mid x_i, \theta)$ the specific formula for the response probability function that the analyst has selected. Maximum likelihood estimation of the model parameters involves maximizing (47) with respect to θ; the maximum likelihood estimator is denoted $\hat{\theta}$.

With the single-bounded approach, the CV survey collects one response from each participant, and the log-likelihood function can be expressed as a series of Bernoulli trials:

$$(48) \qquad L = \sum_{i=1}^{n} y_i \ln P_i + (1-y_i) \ln (1-P_i)$$

where $P_i = P(y_i \mid x_i, \theta)$, the i^{th} individual's response probability. Maximization of the log-likelihood function (48) yields a set of first-order conditions:

$$(49) \qquad \frac{\partial L(\theta)}{\partial\theta} = \sum_{i=1}^{N} \frac{\partial \ln P(y_i \mid x_i, \theta)}{\partial\theta} = 0 \; .$$

In general, these first order conditions are a set of nonlinear equations that require iterative numerical solution techniques. Solution algorithms are discussed further in the appendix to this chapter.[45]

Two practical issues that the researcher must address when using maximum likelihood estimation are the choice of starting values for the parameter vector θ, and the convergence criterion. There is no generally accepted procedure for determining starting values. Fortunately, logit and probit models are well-behaved and generally converge quite quickly regardless of the starting values. With more complicated models, such as those in Appendix Tables 2 - 4 or some of the models to be discussed in Section 4, it may be useful to first estimate a simpler version of the model that takes the form of a logit or probit model and use the resulting coefficients as starting values for the more complicated model. Otherwise, starting values can be set on the basis of the researcher's intuition and knowledge of the data. In some

[44] In what follows, we use $L = L(\theta)$ to denote the *log*-likelihood function.

[45] See Judge et al. (1988) Appendix B or Greene (1993) Chapter 12 for discussions of numerical optimization algorithms. Gill, Murray and Wright (1981) provide an excellent general reference for the issues discussed in the next few paragraphs.

cases, trial and error is the only option.

The iterative estimation procedure ends when a convergence criterion is met. The best criterion is usually based on evaluating the first derivative vectors: when they are zero, the maximum has been reached. Some software packages assume convergence when there is only a very small change between successive estimates of the value of the likelihood function, or successive estimates of the vector of parameters. However, this can halt estimation prematurely. Even with fairly simple likelihood functions, it may be difficult to obtain convergence, especially in models with a large number of parameters. In this case, one can try changing the starting values for the parameters. When starting values are far from the true values, the algorithm for picking the next iteration may move in a direction that leads away from the optimum, or sometimes the function is flat in the region of the starting values and the algorithm cannot find an improving direction so that parameters barely change at each iteration. Another useful trick is to scale the data so that the parameters are expected to be of the same order of magnitude. The GAUSS manual states that 90% of technical support calls about failure to converge are solved by re-scaling the data appropriately.

Many econometric software packages offer a variety of binary response specifications, including logit, probit and Weibull, that use analytical derivatives and converge quite quickly. Estimation of these models involves little more on the part of the user than a one line command. Some packages, such as GAUSS, LIMDEP and STATA, allow the user to program his or her own likelihood function and derivatives, and to specify the starting points, algorithm and convergence criterion. The main reason for programming one's own likelihood function is to estimate models that are more complex than those in the standard software packages, such as those in Appendix Tables 2 - 4 or the models to be discussed in Section 4.

3.2 Asymptotic Properties of Maximum Likelihood Estimators

Once we leave the world of the standard normal linear model, properties such as unbiasedness are very difficult to evaluate. The linear model allows straightforward evaluation of the distributions of estimators and test statistics. For most other situations, however, we resort to *Monte Carlo* simulations or asymptotic evaluations of the limiting distributions of estimators at sample sizes of infinity. Two asymptotic properties of maximum likelihood estimators are of particular interest to us: one is the asymptotic behavior of the estimator itself; the second is the asymptotic behavior of the estimator standardized by sample size so that its limiting behavior is nondegenerate.

3.2.1 Consistency

The maximum likelihood estimator, $\hat{\theta}$, converges in probability to the true parameter value, θ, a property known as *consistency*.[46] This is a large sample property and does not ensure unbiasedness, especially in small samples.[47]

---

[46] See Amemiya (1985) or Huber (1981) for discussions of the asymptotic properties of maximum likelihood estimators.

[47] Anderson and Richardson (1979) and Griffiths, Hill and Pope (1987) found biases in numerical simulations of logit and probit models with small samples.

Small sample results are available from Copas (1988) for the logit model (15). He uses a Taylor expansion to derive analytically the bias of maximum likelihood parameter estimates (Cox and Hinkley, 1974). The bias of the *j*th parameter estimate can be defined implicitly using a Taylor expansion of the *j*th element of the score vector of a log-likelihood function:

$$L_{\theta j} = L_{\hat{\theta} j} + H_j (\theta - \hat{\theta}) + \frac{1}{2} (\theta - \hat{\theta})' M_j (\theta - \hat{\theta}) = 0 \tag{50}$$

where *H* is the jth column of the Hessian matrix and *M* is the jth matrix of third derivatives of the log-likelihood function. The bias is equal to the expected value of ($\hat{\theta}$ - θ). In general, the expectation of the expressions in equation (50) involve covariances between H and $\hat{\theta}$ cannot be simplified. In the case of the logit model, and only for this case, however, H is a constant and is not dependent on the particular response vector. Using this property, Copas obtains an approximate expression for the bias of the *s*th maximum likelihood parameter estimate for the logit model:

$$b_s \cong \sum_j \sum_k \sum_l H^{sj} H^{kl} M^{jkl} \tag{51}$$

where H^{jk} is the inverse of $H = \{H_{jk}\}$. Expressions for *H* and *M* are presented in Kanninen (1995).

The bias is a function of the maximum likelihood estimate, $\hat{\theta}$ and the bid vector. Equation (51) provides an analytical formula for the small sample bias of the maximum likelihood estimator in the logit model. This can be used to examine the effect of the choice of bids on the bias of the maximum likelihood estimator. However, analytical formulas for small sample bias do not appear to be available for other binary response models besides logit. We might surmise that probit models would have similar bias properties to the logit model but we are not aware that this has yet been proven in the literature, nor are we aware of small sample results for any of the other models listed in Appendix Tables 1 - 4.

3.2.2 Efficiency

The second asymptotic property of maximum likelihood estimators is the convergence of $\hat{\theta}$ scaled by the sample size, N: $\sqrt{N}$ ($\hat{\theta}$ - θ) converges in distribution to a normal distribution with mean 0 and variance $[I(\hat{\theta})/N]^{-1}$, where $I(\hat{\theta})$ is the Fisher Information matrix (the negative of the expectation of the Hessian matrix), also known as the *Cramer-Rao Lower Bound*. For the binary response model, the Fisher information matrix is:

$$I(\theta) = \sum_{i=1}^{N} \frac{f^2}{F\ (1-F)}\ x_i x_i' \tag{52}$$

where f is the p.d.f. and F the c.d.f. of the response vector, and x_i is the vector of independent variables including the bid value and a one for the constant term. The Cramer-Rao inequality states that no consistent estimator can have lower variance than the inverse of the Fisher information matrix. This implies that $\hat{\theta}$ is an *efficient* estimator. We refer to the inverse of $[I(\hat{\theta})/N]$ as the asymptotic variance of the maximum likelihood estimator.

3.2.3 Calculating the Variance of WTP

Although we know the asymptotic distribution of the maximum likelihood estimator for θ, we do not necessarily know the asymptotic distribution of the maximum likelihood estimator of the welfare measures C^* or C^+ since those are *functions* of the elements of θ. For example, we might be interested in the parameter $\mu = \alpha/\beta$ which corresponds to both C* and C^+ for the simple logit and probit models (14) and (15), and to C* for censored versions of those models as well as for nonlinear models based on (19). While the maximum likelihood estimators of α and β are asymptotically normal, the distribution of a function of them such as μ is *not* asymptotically normal. Sometimes, this can be tackled by reparametrizing the likelihood function; as noted earlier, the logit and probit models can be parametrized in terms of $\mu = \alpha/\beta$ and $\sigma = 1/\beta$ instead of α and β (Cameron, 1988). Because the maximum likelihood estimator satisfies an invariance principle, both parameterizations yield exactly the same estimate of $\hat{\mu}$.[48] But, this does not apply to the *variance* of the maximum likelihood estimate of $\hat{\mu}$. One obtains different estimates of this variance depending on whether the model is parametrized in terms of (μ,σ) or (α,β). The former provides a direct estimate of var($\hat{\mu}$) via the appropriate element of the inverse of the information matrix; the latter provides an asymptotic variance matrix for $(\hat{\alpha},\hat{\beta})$ from which one must somehow extract an estimate of var($\hat{\mu}$).[49]

When one cannot obtain the asymptotic distribution of the welfare measure via reparametrization, there are three ways to proceed. One approach is to apply the delta method based on a truncated Taylor series expansion which yields the following convergence in distribution of a function $f(\theta)$:

$$\sqrt{N}\ [f(\hat{\theta}) - f(\theta)] \rightarrow N[\ 0,\ f'(\hat{\theta})\ I(\hat{\theta})\ f'(\hat{\theta})'\ ]\ . \tag{53}$$

For the logit or probit models (14) and (15), the resulting asymptotic variance of $\mu = \alpha/\beta$ is

$$Var\left(\frac{\alpha}{\beta}\right) \cong \frac{1}{\beta^2}\left[\sigma_{\alpha\alpha} + \left(\frac{\alpha}{\beta}\right)^2 \sigma_{\beta\beta} - 2\left(\frac{\alpha}{\beta}\right)\sigma_{\alpha\beta}\right] \tag{54}$$

where $\sigma_{\alpha\alpha}$ and $\sigma_{\beta\beta}$ are the variances of α and β and $\sigma_{\alpha\beta}$ their covariance. In practice, one uses the maximum likelihood estimates of the coefficients $\hat{\alpha}$ and $\hat{\beta}$ and the variance matrix $[I(\hat{\theta})/N]^{-1}$. Cox (1990) shows that the resulting estimate of the var($\hat{\mu}$) is the same as that which arises from asymptotic theory if the model is parametrized directly in terms of (μ,σ). However, two qualifications should be noted. First, since it is an asymptotic approximation, this variance formula may be unreliable with small samples. Second, it forces the confidence intervals for $\hat{\mu}$ or other quantiles of the WTP distribution to be symmetric, which may be implausible. The other methods discussed below do not possess this feature.

In the biometrics literature, Finney (1964) popularized the use of Fieller's (1954) theorem on ratios of normally distributed random variables to obtain a confidence intervals for the ED_{50}. Fieller's theorem generates the following 100(1-γ)% confidence interval for $\mu = \alpha/\beta$, $ci_\gamma(\mu)$,

[48] The invariance principle states that the maximum likelihood estimator of a function is the function of the maximum likelihood estimator.

[49] Many standard software packages force one to parametrize the response probability model in terms of (α,β) rather than (μ,σ), which could be an argument against using them. Note that, in cases where the welfare measure is a more complicated function of the model coefficients than the simple ratio (α/β), there may be no convenient reparametrization that permits direct estimation of the welfare measure even if one does program one's own likelihood function.

$$(55) \quad ci_\gamma(\mu) = \frac{1}{1-g}\left[\left(\frac{\alpha}{\beta} - g\frac{\sigma_{\alpha\beta}}{\sigma_{\beta\beta}}\right) \pm \frac{z_{\gamma/2}}{\beta}\left\{\sigma_{\alpha\alpha} - 2\sigma_{\alpha\beta}\frac{\alpha}{\beta} + \sigma_{\beta\beta}\left(\frac{\alpha}{\beta}\right)^2 - g\left(\sigma_{\alpha\alpha} - \frac{\sigma_{\alpha\beta}^2}{\sigma_{\beta\beta}}\right)\right\}^{\frac{1}{2}}\right]$$

where $z_{\gamma/2}$ is the upper ($\gamma/2$) point of the standard normal distribution and $g \equiv (z_{\gamma/2})^2\sigma_{\beta\beta}/\beta^2$. Again, one uses maximum likelihood estimates of coefficients and variance/covariance terms to construct the confidence interval. The more β is significantly different from zero, the smaller the term g. If g were zero, the confidence interval for $\hat{\mu}$ generated by (55) would be exactly the same as that generated by the asymptotic variance in (54). When g is small (say $g < 0.05$) the two confidence intervals are similar. They become different as g increases, and the confidence interval in (55) becomes increasingly asymmetric. Recent literature, summarized by Morgan (1992), suggests that the Fieller interval is generally preferable to the confidence interval generated by (54).

The Fieller approach to confidence intervals has been relatively neglected in the CV literature in favor of other methods involving Monte Carlo simulation or bootstrapping. The former simulates the asymptotic distribution of the coefficients (i.e., a multivariate normal distribution with means given by the maximum likelihood estimates of the coefficients and a variance-covariance matrix given by $[I(\hat{\theta})/N\,]^{-1}$), taking repeated random draws of coefficient vectors from this distribution and using them to generate an empirical distribution for the welfare measure, from which a confidence interval is computed. This approach was proposed by Krinsky and Robb (1986) and first applied to CV by Park, Loomis and Creel (1991). Bootstrapping and jackknife procedures (Efron & Tibshirani, 1993) simulate the distribution of the explanatory variables in the data set (i.e., the CV bid and any other covariates), using the actual data sample of *N* observations to simulate this distribution. Bootstrapping creates multiple simulated data sets, each formed by sampling *N* times with replacement from the actual data. The jackknife creates *N* simulated data sets, each formed by dropping one observation from the actual data. Given *N* simulated data sets for the explanatory variables, both approaches use the actual estimated maximum likelihood coefficients to generate a set of *N* quantal responses, and then apply maximum likelihood to these simulated data to obtain a new set of coefficient estimates, from which the welfare measure is computed. This is repeated over many simulated data sets to generate an empirical distribution for the welfare measure from which a confidence interval is computed. This was first applied to CV by Duffield and Patterson (1991). Cooper (1994) compares these methods for constructing confidence intervals for the welfare measure μ, along with the delta method (54), and finds they all perform fairly well, with the ranking of methods varying according to the size of the sample, the specification of the response probability distribution (e.g., logistic versus Weibull), and whether or not one allows for dispreference or indifference in the formula for the welfare measure. Poe et al. (1994) use bootstrapping to test for differences in WTP distributions (obtained, for example, through different elicitation procedures); they bootstrap the convolution distribution of the difference $\Delta \equiv (WTP_1 - WTP_2)$ and use the resulting confidence interval to test the null hypothesis that $\Delta = 0$.

3.3 Checking Model Misspecification

Fitting a response probability model and estimating a welfare measure is often only part of the work. The researcher may want to know whether the results are vulnerable to misspecification in the fitted model. This could arise in several ways, including specifying the wrong response probability model, omitting relevant regressors, or errors in variables used as regressors. Moreover, there could be heterogeneity of preferences among the respondents so that different observations correspond to different response probability models. Since the true model is unknown, misspecification is always a possibility. But, its consequences are considerably more severe in nonlinear models such as logit or probit than in conventional linear regression models. In the linear model, the ordinary least squares estimator possesses a robustness which makes it still consistent (although biased) even if the model is misspecified. In nonlinear models, by contrast, the maximum likelihood estimator is not consistent if the model is misspecified. This could undermine the reliability of both point estimates and confidence intervals for model coefficients and welfare measures. How serious this is depends on the circumstances.

Given misspecification of the form of the response probability model, Ruud (1983) showed analytically that, despite this, maximum likelihood estimates of slope coefficients in a quantal response model (but not the intercept) could still be consistent under fairly general conditions, except for a common scaling factor. Hence, *ratios* of slope coefficients could be estimated consistently. This result, however, would not help with welfare measures, since those depend on the intercept α as well as the slope coefficient β. A simulation analysis by Horowitz (1993) suggests that errors in estimating these are likely to be small as long as the assumed WTP distribution has the same *qualitative* shape as the true distribution, e.g., they both are unimodal and homoskedastic. But otherwise, for example when one fits a logit model (15) to data generated from a true distribution that departs significantly from an ogive (S-shaped) curve because it is bi-modal or heteroscedastic, maximum likelihood estimation of a misspecified model can bring large errors in coefficient estimates and response predictions.[50] [51]

Heteroscedasticity in quantal response models was considered by Yatchew and Griliches (1985) who found that, for small departures from homoscedasticity in a probit model, there is only a rescaling effect as long as the variance is uncorrelated with the explanatory variables. But, they found that there could be a substantial bias in coefficient estimates when the variance is correlated with the explanatory variables in question.

Omitting relevant regressors affects the estimates of both the intercept and the slope coefficients. With the latter, Yatchew and Griliches show that there is a combination of two effects. The first effect, as in linear regression, is that the coefficients of the included variables are in general biased, the bias being equal to the true coefficient of the omitted variable multiplied by the coefficient from a regression of the omitted variable on the included variable. In addition, unlike linear regression, there is a second effect resulting from the nonlinearity of the quantal response model. This is a rescaling of the slope coefficient which applies regardless of the first effect. Thus, with missing variables in a quantal response model the coefficients of the included variables may be biased even if the omitted variables are uncorrelated with included variables (see also Lee, 1982). Moreover, in contrast to linear regression, the biases depend on the shape of the conditional distribution of the omitted variable, not just the conditional mean.

---

[50] Bi-modality can arise from heterogeneity of preferences. Heteroscedasticity can arise for reasons discussed in footnotes 15 and 16.

[51] We showed in Section 2 that most of the response probability models in the existing literature fail to satisfy bounds on the WTP distribution imposed by economic theory and therefore are misspecified to some degree. Horowitz's findings suggest that the consequences might not be too severe since the response probability models in Appendix Table 1 and those in Appendix Tables 2 - 4 have similar shapes.

Errors in variables in quantal response models has been investigated by Yatchew and Griliches, Carroll et al. (1984), Kao and Schnell (1987), and many others; Carroll et al. (1995) provide a comprehensive review of the recent literature. When explanatory variables are measured with error, the maximum likelihood estimates of the intercept and slope coefficients are inconsistent. While the bias depends on the specific distribution of the measurement errors, there is a general tendency for maximum likelihood estimation to *attenuate* the coefficient of the variables measured with error, i.e. shrink them towards zero. The coefficients of other variables *not* measured with error will also be affected, often in an offsetting manner. In the case of a probit model with normal measurement errors, the maximum likelihood coefficient estimates are scaled down in a manner that leads to a consistent estimate for α/β but not for the mean and other quantiles of the WTP distribution. In small samples, however, even the estimate of α/β may be seriously biased.

3.3.1 Model Checking

While it is well known that there exists a variety of diagnostic techniques for checking model adequacy in the case of linear regression (Cook and Weisberg 1982, 1994), it is less well known that analogous techniques exist for logit and other quantal response models. These have so far received too little attention from CV researchers. In both linear and nonlinear models, the two main pillars of diagnostic testing are the analysis of residuals and the assessment of influence.

Residuals are used to identify ill-fitting data -- observations that do not gibe with the estimated model. There are several ways to define residuals for nonlinear models. The raw residual is $r_i = [y_i - E(y_i)] = [y_i - P(y_i \mid x_i, \theta)]$. For a binary response model where $y_i = 0$ or 1, r_i takes one of two possible values, $- P(y_i \mid x_i, \theta)$ or $1 - P(y_i \mid x_i, \theta)$. Since a large difference between y and E(y) should matter less when y has low rather than high precision, it is conventional to divide r by the standard deviation of y, producing the standardized residual

(56) $$r_S = [y - E(y)] \, / \, s.e.(y).$$

In the binomial case with n_j observations per bid, the standardized residual is simply

(56a) $$r_{Sj} = \frac{y_j - n_j \hat{P}_j}{\sqrt{n_j \hat{P}_j (1 - \hat{P}_j)}} \ ,$$

the normalized difference between the observed and predicted number of "yes" responses at the j^{th} bid. This is known as the *Pearson residual* since the sum of the squares of these residuals

(57) $$X^2 = \Sigma_j r^2_{Sj}$$

is Pearsons' chi-squared statistic for measuring the goodness of fit of a quantal response model. These residuals thus measure the contribution that each observation makes to the overall fit of the model. An alternative is the *standardized Pearson residual*, in which the raw residual r is normalized by its standard error in order to allow for the variation in $E(y)$ as well as y:

(58) $$r_P = [y - E(y)] \, / \, s.e.[y - E(y)].$$

By construction, these residuals have unit variance. Another approach is to find some transform of the responses, T(y), which renders them approximately normally distributed and then construct the analog of the standardized Pearson residual, which is known as the *Anscombe residual*:

(59) $$r_A = [T(y) - E(T(y))] / s.e.[T(y) - E(T(y))].$$ [52]

A different type of residual can be formed from the *deviance statistic*

(60) $$D = -2 [L^c - L^f] ,$$

where $L^c \equiv \Sigma_i l_i^c$ is the log-likelihood function for the "current" model that is being estimated, with l_i^c the log-likelihood term for the i^{th} (grouped) observation, and $L^f \equiv \Sigma_i l_i^f$ is the log-likelihood for the "full" model, i.e., a model that would have as many parameters as observations and therefore would fit the data perfectly. D measures the model's overall discrepancy with the data and thus can be considered an alternative to Pearson's chi-squared statistic. The signed square root of the i^{th} observation's contribution to this statistic,

(60a) $$d_i = - \mathrm{sgn}[y_i - E(y_i)] \cdot \{2(l_i^c - l_i^f)\}^{1/2},$$

where $D = \Sigma_i d_i^2$, is known as the *deviance residual*. The *standardized deviance residual* is

(61) $$r_{Di} = d_i / s.e.(d_i)$$

where $s.e.(d_i)$ is (an approximation to) the standard error of d_i. Another way of using the deviance statistic to define a residual is to compare the change in the value of the deviance when each of the observations in turn is omitted from the data set. The signed square root of the change in deviance for the i^{th} observation, known as the *likelihood residual*, can be approximated by a linear combination of the standardized Pearson residual and the standardized deviance residual for that observation (Collett, 1991).

Although the Anscombe residuals, standardized deviance residuals and likelihood residuals have very different formulas, the values that they take are often remarkably similar, and they tend to rank extreme observations in the same way (Pierce and Schafer, 1986). By contrast, the standardized Pearson residuals are not as closely approximated by a normal distribution and may not rank extreme observations correctly. Since the Anscombe residuals are much more difficult to compute than others, there is no great advantage in using them. The standardized deviance residuals or the likelihood residuals do not suffer from this disadvantage and should routinely be considered.

Residuals are used for two main purposes -- model checking and the identification of outliers (observation checking). First, by plotting residuals in various ways one can check for the omission of relevant variables, the inclusion of inappropriate variables, the use of inappropriate transformations for included variables, or the use of an inappropriate functional form or probability distribution in the quantal response model. Second, through plots or statistical tests of residuals one can identify observations that are discordant with the rest of the data with a view to investigating the causes of the discrepancy or handling them as outliers. Chesher and Irish (1987) and Collett (1991) are useful sources on this topic.

An observation is influential if its deletion from the data would cause major changes in coefficient estimates, confidence intervals, or other statistics of interest. An observation that is an outlier need not be

[52] Formulas for the standard errors in the denominators of (58) and (59) as well as for the transformation T(y) are discussed in Pierce and Schafer (1986) and Collett (1991).

influential, and *vice versa*. An influential observation that is not an outlier occurs when the observation distorts the form of the fitted model so much that there is only a small residual for that observation (the classic example is extreme observations in ordinary least squares regression). Since residuals *per se* do not necessarily reveal the presence of influential observations, additional diagnostic techniques are needed. In linear regression, the key tool for assessing influence is the *hat* matrix $H = X(X'X)^{-1}X'$ where X is the matrix of observations on the explanatory variables. H maps the observed value of the dependent variable, y, into the predicted value, $\hat{y}$: $\hat{y} = Hy$. Its diagonal elements, h_{ii}, measure the effect that the observed y_i has on the fitted $\hat{y}_i$, and are known as the *leverage* of the i^{th} observation (Cook and Weisberg, 1982). In nonlinear models estimated by maximum likelihood there is no analogous exact formula linking y to $E(y|\hat{\theta})$, but Pregibon (1981) obtained some useful results based on the *case deletion* principle which considers the change in the maximum likelihood coefficients when one observation is omitted. Pregibon developed a convenient approximation analogous to the hat matrix which can be used to measure the leverage of an observation on the estimated coefficients or some linear combination of them. The details can be found in Collett (1991), Morgan (1992) or Santner and Duffy (1989). Another approach uses infinitesimal perturbation of the maximum likelihood normal equations rather than case deletion to assess influence. This has been applied by James and James (1983) to assess influence on the estimate of the ED_{50} in a logit dose-response model, and is discussed in Morgan (1992).

In summary, despite its desirable asymptotic properties, maximum likelihood is still sensitive to errors in the data or deviations from the assumed model. There are three main ways to deal with this -- modifying the estimation procedure to make it less sensitive to influential or outlying observations, dropping outliers from the data, or modelling the outlier generating process itself.

Approaches that offer better protection than maximum likelihood against violations of the model assumptions include robust estimation (Huber, 1981), bounded-influence estimation (Hampel, 1974), and resistant fitting (Hoaglin Mosteller and Tukey, 1983). Technically, *robust estimation* refers to techniques that are robust against deviations from the assumed stochastic distribution (e.g. the random elements come from a distribution with a much heavier tail than the normal distribution), while the other two terms refer to techniques that are robust against more general violations of the assumptions such as gross errors in the variables or misspecification of the functional form. In *M-estimation*, one of the main methods of robust regression, coefficient estimates are chosen so as to minimize $\Sigma_i\ \psi(r_i)$ for some loss function $\psi(\cdot)$. Ordinary least squares and least-absolute-deviations are special cases where $\psi(r) = r^2$ and $\psi(r) = |r|$, respectively; the first is highly sensitive to large values of r_i, the latter highly insensitive. Most M-estimates use some loss function that is a compromise between those two; the Huber loss function is equal to r^2 when r is near 0, but close to $|r|$ when r is far from 0. This has the effect of downweighting observations that are discordant with the rest of the data. The limiting form of downweighting is deleting the observation entirely. M-estimation can thus be regarded as a generalization of more traditional approaches in statistics such as removing outliers when they are detected or *trimming*, in which one automatically removes a fixed proportion of the sample from one or both tails. The *influence function* of an estimator measures the rate of change in the estimator with respect to the proportion of observations that come from a contamination point rather than from the assumed model. Robust estimators generally have an influence function that is continuous and bounded. *Bounded-influence estimators* are those with minimum variance subject to a specified upper bound on the value of the influence function.

Robust and resistant estimators were first developed for linear regression but have subsequently been extended to nonlinear models. Early applications to logit were Hamilton et al. (1977) for trimmed estimation of the LA_{50}, Miller and Halpern (1980) for robust estimation, and Pregibon (1981, 1982) for bounded influence estimation. See Kunsch et al. (1989) and Morgan (1992) for some more recent developments. Nonparametric estimation, discussed below in Section 4.4, can be seen as yet another way of estimating a probability response model that protects against error in specifying a parametric model.

Finally, Copas (1988) has explored the possibility of dealing with outliers by modeling the contamination process explicitly. With a binary dependent variable, outliers must take a simple form -- the recorded response is "yes" when the model yields a high probability of "no," or conversely. Thus, Copas proposes a "transposition" model where some fraction of the responses, γ, is mis-classified:

(62) Pr{recorded response is "yes"} = $(1 - \gamma)$ Pr{true response is "yes"} + γPr{true response is "no"}

where γ is a parameter to be estimated from the data along with the parameters of the true response probability model.[53] This may be regarded as a structural approach to the problem of contamination, as opposed to that of resistant estimation.[54]

3.3.2 Hypothesis Tests

Given maximum likelihood coefficient estimates, formal hypothesis tests can be used to test the inclusion of a set of variables in the model. The classical approaches to hypothesis testing are the Wald Test, which relies on parameter estimates from the unrestricted model only, the Likelihood Ratio Test, which relies on parameter estimates from the restricted and unrestricted models and the Lagrange Multiplier Test which relies only on parameter estimates from the restricted model. The choice of test is generally based on ease of computation, which for most practitioners, depends on the statistics provided by the software package used.

Most software packages provide t-statistics for each estimated coefficient. The t-test is equal to the square root of the Wald test and is used to test the null hypothesis that a coefficient, β_0, is equal to some value, z:

(63)
$$t = \frac{\hat{\beta}_0 - z}{s.e.(\hat{\beta}_0)}$$

where s.e.($\hat{\beta}_0$) is equal to the asymptotic standard error of $\hat{\beta}_0$. The critical value for a two-tailed t-test with 95% confidence is 1.96; for a one-tailed t-test with 95% confidence it is 1.645.

The Likelihood Ratio Test compares the maximized log-likelihood function under the hypothesized restriction (L_R) with the unrestricted maximized likelihood function (L). The greater the difference between these two values, the less likely the restriction is to be true. This test is more flexible than the t-statistic because it can be used to compare any number of restrictions simultaneously. The test statistic is:

(64)
$$LRT = -2\ (L_R - L)$$

---

[53] Observe that this is a mixture model like (33’), but with a different structure.

[54] There is some debate in the literature as to which is preferable. Copas (1988) and Jennings (1986) argue that the techniques of resistant estimation do not carry over well from the linear model to quantal response models; Pregibon (1988) disagrees.

which has a χ^2 distribution under the null hypothesis with the number of degrees of freedom equal to the number of restrictions. Rejection of the null hypothesis simultaneously rejects the set of restrictions being tested. The likelihood ratio test can be used to test for a variety of misspecification problems such as the inclusion of variables, functional specifications or heteroscedasticity of known form.[55]

Lagrange multiplier tests, or score tests, have also been shown to be useful for testing a variety of misspecification problems. These tests compare the value of the score function, or first-derivative of the likelihood function, to see if it is significantly different from zero and are convenient because they require only estimates of the score function under the null hypothesis (the restricted model). Davidson and MacKinnon (1984) and McFadden (1987) show that Lagrange multiplier tests for binary response models can be computed using linear regressions of the residual values from the restricted models.

3.3.3 Goodness of Fit Measures

Goodness of fit measures are used to assess how well a model fits the observed data. We have already mentioned the Pearson chi-square statistic (57), which compares fitted and observed response frequencies, and the deviance statistic (60), which compares the maximized loglikelihood function with the "full" model, as measures of goodness of fit. Under the null hypothesis that the fitted model is the true one, both test statistics are asymptotically distributed χ^2 with M-K degrees of freedom, where M is the number of groups defined by the bid values and categorical variables and K is the number of estimated parameters in the model. The test statistics are attractive as goodness of fit measures because they can use statistical inference to explicitly test the fitted model. Unfortunately, when a large number of groups are defined, with a limited number of observations per group, the asymptotic χ^2 distribution is a poor approximation to the true distribution and statistical inference cannot be used. The approximation is especially poor when the data are ungrouped, i.e., there is a single observation per group (Collett, 1991, McFadden, 1976).[56]

Other approaches to measuring goodness of fit do not provide statistical tests but can be used as indicators of model adequacy, perhaps for comparing different models. The McFadden pseudo-R^2 is widely used and is available in most standard software packages. It can be written as:

$$R^2 = 1 - \frac{L_{\max}}{L_0} \tag{65}$$

where L_0 is the log-likelihood in the null case (where all coefficients other than the constant are assumed to be zero) and L_{max} is the unrestricted log-likelihood. There is no commonly accepted threshold value for the pseudo-R^2 that denotes a satisfactory or well-specified model; higher values are preferred. Ben-Akiva

---

[55] Note that this procedure cannot be used to compare different model specifications that are non-nested. Horowitz (1983) specifically addresses non-nested hypothesis testing for discrete choice models. See also Gourieroux, Monfort and Trognon (1983), Kent (1986), Pesaran and Deaton (1978), MacKinnon (1983) and Davidson and MacKinnon (1981).

[56] Furthermore, Collett (1991) has shown that, in this case, the deviance statistic does not measure goodness of fit because the "full" model will always equal exactly zero. The deviance statistic then depends only on the maximized log-likelihood function and not the actual observations.

and Lerman (1985) suggest that a shortcoming of the pseudo-R^2 measure is that it increases, or at least does not decrease, when new variables are added to the model. However, if the new variables do not add sufficient explanatory power to the model, one might prefer a more parsimonious specification. An *adjusted* pseudo-R^2 measure addresses this concern:

$$\bar{R}^2 = 1 - \frac{L_{\max} - A}{L_0} \quad . \tag{66}$$

Two proposals for the adjustment factor, *A*, have been offered. Ben-Akiva and Lerman suggest setting *A* equal to the number of parameters in the model, K; Horowitz (1982) proposes setting it equal to K/2.

The classification procedure is another common approach to evaluating goodness of fit. One generates a 2x2 classification table of hits and misses by comparing the predicted and actual outcomes. Prediction typically uses the simple rule that an outcome is predicted to be positive when the predicted response probability is greater than 0.5, and negative otherwise. The diagonal elements of the table display the numbers of correctly predicted positive and negative outcomes and the off-diagonal elements contain the numbers of mis-classified outcomes. There is no specific requirement for the minimum number of correct predictions. The main disadvantage of this technique is that the prediction rule is simplistic and can greatly misrepresent situations where the predicted probability is close to 0.5.

3.4 Alternatives to Maximum Likelihood Estimation

Alternatives to maximum likelihood estimation that are sometimes used in the CV literature include minimum chi-squared estimation, the generalized linear model and quasi-maximum likelihood estimation. Minimum chi-squared estimation is based on the Pearson chi-square statistic (57), or some transformed version of it, taken as an overall measure of goodness of fit. It was first developed by Berkson (1953) for logit, where the transform $z = \psi(P) \equiv \ln[P/(1-P)]$ applied to (15) yields a simple linear model

$$z_j = \alpha - \beta A_j. \tag{15'}$$

Let $\hat{z}_j = \psi(\hat{P}_j)$, where $\hat{P}_j$ is the observed proportion of yes responses for bid A_j. Then, (15') suggests an ordinary least squares regression of $\hat{z}_j$ on A_j

$$\hat{z}_j = \alpha - \beta A_j + \upsilon_j \tag{15''}$$

where $\upsilon_j \equiv \hat{z}_j - z_j$ is a random error term, with $E\{\upsilon_j\} = 0$.[57] However, var (υ_j) is not a constant, and therefore the model is heteroscedastic and must be estimated by weighted least squares. Provided that $\hat{P}_j$ is not too near 0 or 1, var(υ_j) is asymptotically equal to $V(P_j) \equiv [n_j P_j(1-P_j)]^{-1}$, where n_j is the number of observations associated with bid A_j. A consistent estimate of $V(P_j)$ is given by $V(\hat{P}_j) \equiv [n_j\hat{P}_j(1-\hat{P}_j)]^{-1}$. The weighted least squares estimator of (15'') minimizes

$$\Sigma_j\, (\hat{z}_j - z_j)^2 / V(\hat{P}_j).$$

[57] A similar transform can be applied to the logit version of the Box-Cox model (11) by substituting $[\alpha - by^\lambda + b(y-A_j)^\lambda]$ for $[\alpha - \beta A_j]$ in (15') and (15''). This leads to a nonlinear least squares regression.

The expression being minimized is a transformed version of the Pearson chi-squared statistic (57), and the resulting estimate of α and β is known as the minimum chi-squared estimator. This estimator is consistent and asymptotically normal, and has the same asymptotic covariance matrix as the maximum likelihood estimator. Following Berkson, Amemiya (1980) showed that it has a smaller mean square error than the maximum likelihood estimator. Minimum chi-squared estimators have been developed for probit and other quantal response models where there exists some appropriate transform of the response variable that generates a linear (or simple nonlinear) least squares regression like (15"); details can be found in Maddala (1983).

By construction, when an appropriate transformation exists, the minimum chi-squared estimator is simpler to estimate than the maximum likelihood estimator, since it involves weighted least squares regressions (sometimes iterated, if either the transformation $\psi(\cdot)$ or var(υ) depend on the parameters to be estimated) and has the same asymptotic covariance matrix as the maximum likelihood estimator. Since the approximation to the asymptotic variance breaks down when $\hat{P}_j = 0$ or 1, the method cannot be applied to individual data, just to grouped data where the observed proportions lie between 0 and 1. This is not a problem with single-bounded CV where multiple subjects receive the same bid, A_j, unless one wants to include covariates in the response probability model as determinants of α and/or β. In that case, one needs enough data so that there are multiple observations for each given value of the covariates, a requirement that is difficult to satisfy unless covariates take only a small number of discrete values.

In the generalized linear model (GLM), introduced by Nelder and Wedderburn (1972) and discussed in detail in McCullagh and Nelder (1989), it is assumed that the responses $\{y_1,...,y_n\}$ are independent observations drawn from the *exponential family* of distributions. The distribution of a random variable Y belongs to the exponential family if the density function takes form

$$f_Y(y) \;=\; exp\;[\;(y\zeta - b(\zeta))/a(\phi) \;+\; c(y,\phi)\;] \tag{67}$$

where $a(\cdot)$, $b(\cdot)$, and $c(\cdot)$ are specified functions, and ζ and ϕ are parameters. It follows from (67) that $E(Y) \equiv \mu = b'(\zeta)$, where b' is the first derivative of $b(\cdot)$, and $var(Y) = a(\phi)b''(\zeta)$, where b" is the second derivative of $b(\cdot)$. The parameter ζ is known as the natural parameter of the distribution. The parameter ϕ is known as the scale or dispersion parameter. In general, the variance can be a function of the mean, via its dependence on ζ. Different specifications of $a(\cdot)$, $b(\cdot)$, and $c(\cdot)$ give rise to different distributions. For example, if $a(\phi) = \phi \equiv \sigma^2$, $b(\zeta) = \zeta^2/2$, and $c(y,\zeta) = \{y^2/\sigma^2 + \ln\,(2\pi\sigma^2)\}/2$, one has the normal distribution with $\mu = \zeta$. Other members of this family include the Poisson, gamma, and binomial distributions. Covariates may enter the model through the natural parameter ζ, which can be written as a function $\zeta = \zeta(X)$. In a GLM, this is assumed to be a linear function, $\zeta = \Sigma\, X_k\beta_k$, and in this context ζ is referred to as a *linear predictor*. Also, in GLM terminology, the relation between the mean, μ, and the linear predictor, ζ, is known as the *link function*: $\zeta = g(\mu)$. It follows that the inverse of the link function coincides with the derivative of the $b(\cdot)$ function: $\mu = g^1(\zeta) = b'(\zeta)$. The logit model, for example, when viewed as a GLM, has $E\{Y\} \equiv \mu$, the link function is simply the transform $\zeta = \psi(\mu\pi) = \ln[\mu/(1-\mu)]$, and the linear predictor is $\zeta \equiv \alpha - \beta A$.

Given the GLM structure, when the method of scoring -- see (A.6) in the Appendix -- is applied to a GLM model, the coefficient estimates obtained at each Newton-Raphson iteration turn out to be the same as the regression coefficients that one would obtain in a weighted least squares regression with a certain weight matrix and a certain dependent variable regressed on the covariates $X_1,...,X_K$. Hence, the maximum likelihood estimator can be obtained via a series of weighted least squares regressions, which significantly simplifies the computation.[58] In the case of the logit model, the weighted least squares regression is essentially the same as that in minimum chi-squared estimation.

Since the original work of Nelder and Wedderburn (1972) there have been several refinements and extensions of the GLM structure. The *overdispersion* model of Williams (1982) provides a generalization of the binomial and Poisson distributions. The binomial distribution imposes a tight link between mean and variance: var (Y) = $\pi(1-\pi)$. However, it sometimes happens that data on binary outcomes possess a larger variance than permitted by the binomial model. This extra-binomial variation may be due to the omission of explanatory variables or other forms of misspecification captured in the model's random component. If it is not possible to eliminate the extra-binomial variation by including extra covariates, the alternative is to model it explicitly. For this purpose, Williams introduces an additional parameter, ω, the overdispersion parameter. His model has the same linear predictor $\zeta \equiv \alpha - \beta A$ and the same link function $\zeta = \ln[\pi/(1-\pi)]$ as in logit, but the variance is specified as $\text{var}(Y) = \omega\pi(1-\pi)$, where ω is estimated along with α and β within the GLM framework. Williams iterates on ω, α and β until the extra-binomial variation is eliminated, based on a goodness of fit statistic such as the Pearson chi-square statistic (57). Under the assumption of binomial residual dispersion, this statistic has an expected value equal to the degrees of freedom (N-K), where N is the number of distinct groups and K the number of parameters estimated; reaching this value is the convergence criterion.[59]

Another refinement is the method of estimation known as *quasi-likelihood* (Wedderburn, 1974), which extends GLM to statistical models that are *not* a member of the exponential family but have first and second moments with the same type of structure, i.e, $E(Y) \equiv \mu = b'(\zeta)$ and $\text{var}(Y) = a(\phi)b''(\zeta)$ where $a(\cdot)$ and $b(\cdot)$ are specified functions. Nothing else is specified about the distribution other than these two moments. Therefore, if one proceeds to estimate the model as though applying maximum likelihood to an exponential family, this would be a form of specification error. Although *in general* the desirable properties of maximum likelihood estimators are lost when a model is incorrectly specified, this turns out to be an exception where applying maximum likelihood in a situation where it is not strictly valid still produces consistent and asymptotically normal estimators.[60] The quasi-likelihood approach has come to serve as the basis for various recent extensions of GLM, including random-coefficient versions of GLM discussed in Section 4.1.4.

3.5 Optimal Experimental Design

In quantal response models, the bias and variance of $\hat{\theta}$ are functions of not only the underlying parameters of the WTP distribution and the sample size, but also the explanatory variables, including the bid values. This is evident from (51) and (54) for the logit model but it also applies to other binary

---

[58] This is implemented in the statistical packages GLIM and GENSTAT.

[59] Langford (1994) applies the overdispersion model to (15) and finds that, while it has little effect on the point estimate of $\mu = \alpha/\beta$, it increases the covariance matrix for (α,β) and widens the confidence interval for μ. Langford concludes that a conventional logit model ignoring the overdispersion would seriously understate the true confidence interval for median WTP.

[60] For the general theory of quasi-maximum likelihood estimation see Gourieroux, Montfort and Trognon (1984), who use the term pseudo-maximum likelihood estimation, or White (1994).

response models as well. This has led researchers to pursue the study of optimal experimental design as an additional technique for improving the efficiency of estimates of welfare measures obtained from discrete response CV studies. The goal of optimal experimental design is to find the bid values that provide the maximum possible information about the parameters of the WTP distribution, given the size of the sample. Similar issues have been studied extensively in the bioassay and applied statistics literatures for some time (Abdelbasit and Placket 1983, Chaloner 1986, Finney 1978, Minkin 1987, Tsutakawa 1972, 1980, Wu 1985, 1988). More recently, a literature has developed on optimal design for CV experiments (Alberini 1995b, Alberini and Carson 1993, Cameron and Huppert 1991, Cooper 1993, Duffield and Patterson 1991, Kanninen 1993a, 1993b, Nyquist 1992). Most of this literature has focused on the logit model (15) because of its simplicity and because, in this case, it is possible to obtain analytical results. None of the existing design literature has dealt with the models presented in Appendix Tables 2 - 4, but it is reasonable to assume that the many of the qualitative properties of the optimal bid design for the logit model would carry over to these other models. Our discussion, therefore, focuses on bid design for the logit model. We start by reviewing the impact of bid values on the bias and variance of estimators and then go on to discuss the broad principles of optimal bid design for this model.

3.5.1 Bias and Variance in the Logit Model

Table 1 presents estimates of bias and variance for parameter values and welfare measures in the logit model for several alternative bid designs. The biases of $\hat{\alpha}$ and $\hat{\beta}$ are calculated using the formulas in (51). The bias of estimated median WTP is the difference between estimated and actual median WTP: $(-\hat{\alpha}/\hat{\beta}) - (-\alpha/\beta)$, where $\hat{\alpha} = \alpha + \text{bias}(\hat{\alpha})$ and $\hat{\beta} = \beta + \text{bias}(\hat{\beta})$. The asymptotic variance of estimated median WTP is calculated using (54). The true values of the parameters (α, β) are set at (2.5, -0.01), which yields a median WTP value of \$250. Five bid designs are compared. In each case, it is assumed that, while respondents in the CV survey are presented with a single bid, this is drawn with equal probability from one of several possible bid values. The first three examples involve two-point bid designs -- i.e., there are two possible bid values, half the respondents receive one bid value and half the other.[61] The first design places the bid values fairly close to the median of the WTP distribution at \$200 and \$300 (38th and 62nd percentiles). The second design has bid values of \$100 and \$400 (18th and 82nd percentiles). The third design is a "tails only" design with bids in the tails of the WTP distribution: \$5 and \$500 (8th and 92nd percentiles). The fourth design combines the first and third designs to create a four-point design; the aim is to investigate how multiple-point designs perform relative to two-point designs. All of these designs are symmetric, in that the bid values are spaced symmetrically around the median WTP. The fifth design is an asymmetric, one-sided design involving three bid values, all greater than the median.

As Copas (1988) predicted on the basis of (51), the maximum likelihood estimates of both parameters tend to be overstated in a small sample. Because the welfare measures involve the ratio of parameters, the bias in the estimates of the welfare measures is substantially smaller than the biases in the estimates of the individual parameters. Consequently, so long as the bid design does not boost the precision of one parameter at the expense of the other, the estimate of the welfare measure may not be too biased.

---

[61] With a two-parameter model like logit, two design points are sufficient to identify the two parameters of the response probability distribution. See Kanninen (1995) for a similar analysis of bias where more than two design points are used. Previous studies of the bias in WTP estimation have involved empirical examples using actual data (Cooper and Loomis, 1992) and Monte Carlo simulations (Kanninen and Kriström 1993, Cooper and Loomis 1993) rather than the analytic approach employed here.

The bias and variance of the individual parameters and the welfare measures all increase as the bid values are placed further out into the tails of the distribution. The best design is the first one where the bid values are fairly close to the middle of the distribution; the worst design is the one that places the bids in the tails.[62] [63] The combined design, where some bids are placed in the middle and others are placed in the tails, performs better than the "tails only" design in terms of variance but not bias. The one-sided design does not perform well for either bias or variance.

The biases decrease with sample size. Table 1 uses a sample size of 250. The effects of increasing sample size are easily calculated by dividing the bias by the proportionate increase in sample size. For example, if sample size were 500, the bias of $\hat{\alpha}$ and β would be half of the values in the table. The corresponding biases in the estimated welfare measure do not decrease in exactly the same proportion as for the parameter estimates, but in practice they decrease roughly proportionately. It follows that, if there were no financial constraints, the simplest solution for avoiding bias due to poor bid design is to have a large sample.

The sample size and the specific value of β play crucial roles in the value of asymptotic variance. As shown in Kanninen (1993a), the asymptotic variance is inversely proportional to $N\beta^2$; this result extends directly to non-optimal bid designs such as those presented here. As sample size increases, the asymptotic variance decreases proportionately. Also, as β increases, the asymptotic variance decreases. This is because β is inversely proportional to the standard deviation of the WTP distribution. If β is small, a large asymptotic variance of WTP is unavoidable.

Provided that one wants to estimate median WTP and the response probability model is correctly specified, these results suggest that choice of bid design may not have much impact on the point estimate of welfare measures. It can, however, have a substantial effect on the variance of the estimated welfare measure. This is especially problematic in small samples. Consequently, our recommendations are (1) increase the sample size, which has the effect of reducing both bias and variance; and (2) keep the bids away from the tails, because excessively high or low bids inflate the variance of the estimated welfare measure. While the sample size is obviously constrained by the researcher’s budget, we would typically recommend samples of at least 500 respondents, and preferably closer to 1,000, to keep confidence intervals down to a reasonable size.

3.5.2 Basic Principles of Optimal Design

Optimal design may seem odd to an econometrician used to taking the data as given. But, in a discrete response CV experiment the researcher has the opportunity to influence the data by setting the bids. To do this optimally, the researcher must first make an assumption about the form of the underlying WTP distribution and the value of its parameters. Moreover, the researcher must decide which particular

---

[62] In addition to losing statistical information when bid values are located in the tails, excessive high or low bid values may strike respondents as implausible and thereby undermine the credibility of the CV survey. They can also distort welfare estimates if outliers occur in the data (Kanninen, 1995).

[63] Desvousges et al. (1993) are sometimes cited as evidence that about a third of respondents say "yes" to a bid as high as $1,000. This is extremely unrepresentative of the general experience in CV, and reflects the peculiar circumstances in which that survey was administered (Hanemann, 1996); in fact, such high bids typically receive a very small proportion of "yes" responses.

welfare measure she wishes to estimate, and must select a particular criterion for the precision of the estimate. Those decisions will reflect her interests and priorities. For example, she might be concerned to reduce the variance of the estimate of median WTP. This implies minimizing the asymptotic variance of the median, given in (54) for the logit model, with respect to the bid values. In the experimental design literature, minimizing a function of parameter estimates is known as a C-optimal design criterion.[64] Alternatively, the researcher could choose a D-optimal design, based on maximizing the determinant of the Fisher information matrix, (52). In the case of the logit model, this corresponds to jointly minimizing the confidence intervals around the parameters α and β. This criterion would be of most use for a researcher who is primarily interested in obtaining efficient parameter estimates, perhaps for predictive purposes or for benefits transfer.

Optimal bid designs for the minimum variance (C-optimal) and D-optimal design criteria are presented in Table 2. These are expressed in terms of percentiles of the underlying true WTP distribution. Of course, since that distribution is unknown, one has to rely on estimates or guesses of the distribution in order to identify the bid points. This is the basic paradox of optimal experimental design in nonlinear models. The optimal design depends on the true parameter values. Therefore, one must know these in order to engineer the best design for collecting the data. But, there would be no point in collecting the data if one already knew the true parameter values. As indicated, one resolves this paradox in practice by relying on prior information. In addition, it highlights a second important role for optimal design analysis, namely interpreting the results after the data have been collected and analyzed, by helping one to understand, *ex post,* which data points were influential or contributed the most statistical information.

The design results in Table 2 have several interesting features. For example, the designs are all quite simple, consisting of no more than two different bid values each. The single-bounded C-optimal design is especially simple, consisting of a single bid point located at what one believes to be the median of the WTP distribution. Note that such a design does not identify the two underlying parameters, α and β. But then it was not the goal of the optimal design to estimate these parameters -- the aim was to estimate their ratio, α/β. The result tells us something about where the most informative bids are located: the bid values near the center of the WTP distribution are best for obtaining an efficient estimate of median WTP.[65] Conversely, we note that no optimal design contains bids in the outer 12% of each tail of the distribution. The outer tails of the normal and logistic distributions are quite flat, and consequently, responses to bids lying on these parts of the distribution offer very little information about the WTP distribution. If one wanted to obtain reliable point estimates of response probabilities in the tails of the distribution, the sample size would have to be very large: for example, we expect to have to collect 100 observations at the 95th percentile in order to receive just five positive responses. This gets even worse as one moves further out into the tails.

---

[64] Other examples of C-optimal designs in the CV literature are the minimization of the variance of a nonparametric estimator of truncated mean WTP, (44) (Duffield and Patterson, 1991) and the minimization of the mean squared error of mean WTP (Cooper, 1993).

[65] Alberini and Carson (1993) derived optimal designs for this case using higher order Taylor series approximations to the asymptotic distribution of median WTP. They found that, while a first-order approximation (leading to a one-point design) was generally unreliable, there was little gain from using more than a third-order Taylor series term. They obtained two-point designs that tend to be within one standard deviation of the center of the WTP distribution.

3.5.3 Sequential Design

One way to resolve the paradox of optimal experimental design is to conduct the survey in stages, using data from the earlier stages to optimize the design in the later stages. Most CV researchers do this, at least to some degree. For example, a common practice is to employ a two-step procedure: first, perform a pre-test to obtain preliminary information about the WTP distribution, then survey the full sample. The pre-test results are used to design optimal bids for the full survey based on the parameter values estimated from the pre-test data. Typically, however, pre-test sample sizes are small relative to the size of main sample; this allocation of survey effort generates only a modest quantity of initial information.[66]

To take better advantage of the optimal design results available, a sequential estimation procedure should be used, preferably with more than two stages (Kanninen 1993a, Nyquist 1992). After each iteration of the survey, provided the survey instrument has not changed, the WTP distribution can be estimated using all available observations so that a new (optimal) bid design can be derived for use with the subsequent iteration of the survey. This approach can asymptotically attain the efficiency of the theoretical optimal design results.

Kanninen (1993a) demonstrates the efficiency of sequentially updating the bid design by performing a simulation where the initial bid design is poor, but is updated after a set number of observations is collected. If, for example, the sample size is 500 and two iterations are performed, then each iteration collects 250 observations; if three iterations are performed, each iteration collects 166 observations, etc. Table 3 presents the results of this experiment, comparing the simulated mean squared error with the asymptotic mean squared error that would be obtained from a C-optimal design. For a total sample of 500 observations split equally among iterations, the results show that the mean squared error decreases monotonically with the number of iterations. The rate of improvement slows after three iterations. With five iterations, the single-bounded model has a mean squared error that is only 23% larger than the C-optimal design.

Two lessons can be learned from the CV sequential design literature. First, there are substantial efficiency gains from increasing the number of iterations of data collection and bid design update. Second, the use of a sequential procedure relieves concerns arising from the fact that optimal design results call for just two bid values. With a sequential design, only two bid values are employed at each iteration, but the values change between iterations as information is accumulated. How many bid values are ultimately employed depends on how many iterations are performed.[67] How much the bid values differ between iterations depends on how close the initial assessment of parameter values was to the true distribution.

3.5.4 Concluding Comments on Optimal Design

Because optimal bid designs tend to require only one or two bid points, CV researchers are

---

[66] Pre-testing is also essential for honing the questionnaire and checking its field implementation. In our experience, inadequacies of pre-testing are often the greatest flaw in academic CV studies. Not only should there be a reasonably large sample for a pre-test (say, 100-200 observations), but also there needs to be at least two and perhaps three pre-tests for a high quality survey.

[67] Alberini (1995a) investigates how many design points are needed to obtain a powerful Pearson chi-square test.

sometimes reluctant to employ them in practice. There are clearly reasons to be concerned about using two-point designs. First, because there is *a priori* uncertainty about parameter values, it is unwise to limit the range of bid points collected. Second, because optimal bid designs vary depending on the functional form of the underlying distribution, it is risky to base the optimal design on a particular assumption about functional form when it is unknown *a priori*. For both reasons, one wants to modify optimal design results somewhat for practical application.

As discussed above, one solution is to conduct the CV survey sequentially. This approach gives the researcher flexibility for updating the bid designs so that the final bid design, given several iterations, can be close to optimal. The researcher maintains the ability of adapting to new information about the shape and position of the WTP distribution. If several iterations are performed, the data will contain several different bid points because of the different iterations of the survey, but by the end the researcher will have achieved an optimal set of bids to maximize efficiency.

The other solution is to hedge one's bets by making several guesses at the true parameter values, and then using multiple bids corresponding to the optimum designs for the various possible parameter values. By doing this, one might end up with, say, four to six bid points. As a matter of general practice, we would caution against having many more bid points than this. While it may seem attractive to multiply the number of bids in the face of uncertainty about the true WTP distribution, we believe that this is generally an unwise practice for two reasons. First, for a fixed sample size, having more bid points means fewer observations and therefore less information at each point. Second, as noted earlier, bid points from the tails of a distribution are generally uninformative -- for the logistic distribution, we observed that this ruled out bid points in the outer 12% tails of the distribution. Collecting observations in those tails typically wastes the observations; to obtain useful and reliable data so far out in the tails of the distribution, one would need an inordinately large sample.

*** CONTINUED ***

4. ADDITIONAL TOPICS

In this section we consider some additional statistical issues connected with discrete-response CV. We have chosen five topics that interest us and are at the current frontiers of research, although we treat them briefly because of space limitations. The topics are valuing multiple items, the double-bounded CV format, extensions of the response probability model, nonparametric response probability models, and non-response bias.

4.1 Valuing Multiple Items

So far we have focused on valuing a single item. In this section, we discuss the statistical issues that arise when respondents value multiple items. We also briefly consider situations where the same individual provides multiple valuations of the same item; this is then considered at greater length in Section 4.2.

From one point of view, valuing multiple items adds nothing of substance and raises no new issues: one merely takes each item by itself and fits a separate response probability model for that item, using the methods discussed above. However, if the analyst believes that the response probability models are related because they share parts of the utility function in common, there is a distinct advantage to treating the items simultaneously. One can obtain a more efficient estimate of the model parameters by pooling the responses, or at least analyzing them in a linked manner, thus taking advantage of the additional information embodied in the valuations of the other items. In this section we discuss some ways for doing this. Our discussion is divided into three parts. Section 4.1.1 focuses on ways of formulating hedonic models that can be applied to the valuation of multiple items. Section 4.1.2 focuses on the stochastic specification of such models. Section 4.1.3 deals with the ways in which valuations of multiple items can be elicited, including contingent ranking and related techniques. Section 4.1.4 considers the connection with random coefficient models.

4.1.1 Hedonic Utility Formulations

The various items that respondents are asked to value could be either different things (e.g., preserving forests, preserving wetlands, preserving lakes, etc.) or different quantities of the same thing (e.g., preserving 1 forest, preserving 2 forests, preserving 3 forests, etc). In either case, the utility function associated with the k^{th} item will be written $u_k = v_k(q^k,y,\varepsilon_k)$. The baseline utility in the absence of the item is $v_0(q^0,y,\varepsilon_0)$, which may or may not vary with k. The elements that are related or shared in common may involve the deterministic or the stochastic components of $v_k(q^k,y,\varepsilon_k)$, or both. We start with models where the commonality involves just the deterministic components of the utility function. In effect, this defines a *hedonic* utility model. Such models can take various forms. For example, suppose that the underlying utility model is the Box-Cox model (11) where β and λ are the same for all k, while α varies across programs according to some function $\alpha_k = \phi(q^k)$. Thus, the utility associated with the k^{th} program is

(68) $$u_k = \phi(q^k) + \beta\left[\frac{y^\lambda - 1}{\lambda}\right] + \varepsilon_k$$

and the individual's WTP for the k^{th} program is

(68')
$$C_k = y - \left[y^{\lambda} - \frac{\phi(q^k) - \phi(q^0)}{b} - \frac{\eta_k}{b} \right]^{\frac{1}{\lambda}}$$

where $\eta_k \equiv \varepsilon_k - \varepsilon_0$ and $b \equiv \beta/\lambda$. Given the specification of a joint distribution for the η_k's, response probability formulas can readily be generated from (68').[1] If the programs involve different quantities of the same item, q^k is a scalar and $\phi(\cdot)$ is a univariate, and presumably quasiconcave, function. If the programs involve different items, q^k is a vector and $\phi(\cdot)$ is a multivariate quasiconcave function. For this purpose, one could use a standard formula from the literature such as the translog multivariate utility function or some other flexible form. Hoehn (1991) uses a quadratic approximation to represent the change in $\phi(\cdot)$:[2]

(69)
$$\phi(q^k) - \phi(q^0) = e + g\Delta^k + \Delta^{k\prime}H\Delta^k,$$

where $\Delta^k \equiv q^k - q^0$ is the vector of changes in q, e and g are vectors of coefficients, and H is a matrix of coefficients whose off-diagonal elements, h_{ij}, $i \neq j$, reflect the utility interactions among the separate program components. If $h_{ij} < 0$, $i \neq j$, the i^{th} and j^{th} components of q are substitutes for one another, and an increase in q_j lowers the individual's WTP for an improvement in q_i; if $h_{ij} > 0$, $i \neq j$, they are complements, and an increase in q_j raises the WTP for an improvement in q_i.

If there are only a few discrete values of the q_i's, it may be simpler to replace $\phi(q)$, viewed as a function of continuous variable(s), with a step function consisting of a set of constant terms and dummy variables. Suppose, for example, that the programs involve three different items (e.g., i = 1 is forests, i = 2 is wetlands, and i = 3 is lakes), each of which can be either left in a baseline condition or improved by some specific amount. The change in an item is represented by a scalar index Δ_i, where $\Delta_i = 1$ if the item is improved, and $\Delta_i = 0$ otherwise. The k^{th} program consists of some combination $\Delta^k = (\Delta_1^k, \Delta_2^k, \Delta_3^k)$. Given that there are three resources and two possible states for each resource, there are $2^3 = 8$ possible programs. There are several ways to represent $\phi(\cdot)$. One model in the literature sets

(70)
$$\phi(q^k) = \Sigma_i\, g_i\Delta_i^k + h_{12}\, \Delta_1^k\Delta_2^k + h_{13}\, \Delta_1^k\Delta_3^k + h_{23}\, \Delta_2^k\Delta_3^k,$$

where the g_i's and h_{ij}'s are coefficients to be estimated. Another model adds a three-way interaction term

(71)
$$\phi(q^k) = \Sigma_i\, g_i\Delta_i^k + h_{12}\, \Delta_1^k\Delta_2^k + h_{13}\, \Delta_1^k\Delta_3^k + h_{23}\, \Delta_2^k\Delta_3^k + h_{123}\, \Delta_1^k\Delta_2^k\Delta_3^k.$$

Both models satisfy the normalization that $\phi(0,0,0) = 0$. The formulation in (70) is closer to Hoehn's quadratic approximation in (69).[3] The formulation in (71) is what is known as a fully saturated model. Using (71), the baseline utility associated with no improvement in any resource (denoted k = 000) is

---

[1] The restrictions on stochastic specification discussed in Section 2.3 would still need to be imposed in order to ensure that $0 \leq C_k \leq y$.

[2] In addition, he uses the version of (68) where $\lambda = 1$.

[3] Hoehn and Loomis (1993) use a version of (70) where $\lambda = 1$ and apply this to a case where there are *two* discrete levels of q_i^k in addition to the baseline q_i^0.

$$u_{000} = \beta\left[\frac{y^{\lambda} - 1}{\lambda}\right] + \varepsilon_{000} \ , \tag{71a}$$

while the utility associated with a program which improves just forests (k = 100) is

$$u_{100} = g_1 + \beta\left[\frac{y^{\lambda} - 1}{\lambda}\right] + \varepsilon_{100} \ , \tag{71b}$$

that associated with a program which improves both forests and wetlands (k = 110) is

$$u_{110} = g_1 + g_2 + h_{12} + \beta\left[\frac{y^{\lambda} - 1}{\lambda}\right] + \varepsilon_{110} \ , \tag{71c}$$

and that associated with a program which improves all three resources (k =111) is

$$u_{111} = g_1 + g_2 + g_3 + h_{12} + h_{13} + h_{23} + h_{123} + \beta\left[\frac{y^{\lambda} - 1}{\lambda}\right] + \varepsilon_{111} \ . \tag{71d}$$

The utilities associated with the other four possible programs (k= 001, 010, 011, and 101) are calculated similarly. The corresponding formulas for the individual's WTP to have improvement program k = *xxx* instead of the baseline k = 000 are obtained by substituting (71) into (68'), from which response probability formulas can be generated.

What is gained by using the step function hedonic model in (71)? Suppose that, instead of using the formulation in (71), one simply used (68) but with a separate constant $\alpha_k \equiv \phi(q^k)$ for each of the seven improvement programs, yielding

$$u_k = \alpha_k + \beta\left[\frac{y^{\lambda} - 1}{\lambda}\right] + \varepsilon_k \ . \tag{72}$$

In this case, it would be natural to impose the normalization that $\alpha_{000} = 0$. Fitting (72) to CV data covering all of the improvement programs provides estimates of the constants α_{100}, α_{110}, α_{111}, α_{101}, α_{001}, α_{010} and α_{011}. A comparison of (72) with (71b,c) shows that $\alpha_{100} = g_1$, $\alpha_{010} = g_2$, and $\alpha_{110} = g_1 + g_2 + h_{12}$. Hence, $h_{12} = \alpha_{110} - \alpha_{100} - \alpha_{010}$. The other parameters h_{13}, h_{23} and h_{123} can similarly be obtained as linear combinations of the α_k's. Thus, there is a one-to-one mapping between the seven α_k's in (72) and the seven parameters (g's and h's) in (71). It follows that (72) is equivalent to the fully saturated model in (71). By contrast, the model in (70) is not equivalent to (72) because it has fewer parameters -- given estimates of the α_k's from (72), the h's in (70) are over-identified.

Within the framework of the Box-Cox model (68), the formulation in (71) can be regarded as the most general, non-parametric representation of the $\phi(\cdot)$ function that can be estimated from responses to CV questions about the seven programs. Other formulations such as (70) are restrictions of this general model. If the components of q really can take on more than two discrete values associated with the programs in the CV survey, estimating (68) with some particular parametric representation of $\phi(\cdot)$ has the advantage that it allows one to predict the individual's WTP for a range of programs different from those covered in the survey. But, this advantage comes at the cost of imposing a particular parametric representation on $\phi(\cdot)$.

4.1.2 Multivariate WTP Distributions

We have focused so far on ways to model the deterministic component of $v(q^k,y,\varepsilon_k)$ that allow for linkage in preferences for the various items. We have said nothing about the stochastic specification of the ε_k's. Existing empirical models of multi-item valuation generally assume that these are independently and identically distributed (iid) -- i.e., there is *no* linkage among the stochastic components of preferences for the various items. There are two reasons why this may be inadequate. First, there may be unobserved attributes of the programs which induce some correlation in preferences for them (two programs which the analyst is treating as separate may actually be related because of factors that the analyst has overlooked, and this will appear to the analyst as inducing correlation among people's preferences for them). Second, there may be unobserved attributes of the respondents which induce correlation in preferences for programs (an individual who particularly likes one type of program may be more inclined to like (or dislike) another type). In these cases, one needs to introduce some joint density for the ε_k's, $f_\varepsilon(\varepsilon_0,\varepsilon_1,..,\varepsilon_K)$, with a non-diagonal covariance matrix. How one might do that is the subject of this section.

Continuing with the Box-Cox model, the dependence among the ε_k's in (68) carries over to the η_k's and the C_k's. The joint density $f_\varepsilon(\varepsilon_0,\varepsilon_1,..,\varepsilon_K)$ induces a joint density $f_\eta(\eta_1,..,\eta_K)$ and, via (68'), a joint density $g_C(C_1,..,C_K)$. The latter is used to represent the response probability formula for the suite of program valuations. Suppose, for example, that there are K = 3 programs which are evaluated using the single-bounded, discrete-response format. Let A_k denote the dollar cost presented to the individual for the k^{th} program. Suppose, the individual answers "no" to programs 1 and 3, but "yes" to program 2. From (6a,b), the probability of these responses is given by

$$\text{(73)} \qquad \Pr\{\text{"no" to 1 and 3, "yes" to 2}\} = \Pr\{C_1 \le A_1,\ C_2 \ge A_2, \text{ and } C_3 \le A_3\}$$
$$= \int_0^{A_1} \int_{A_2}^{C_2^{max}} \int_0^{A_3} g_C(C_1,C_2,C_3)\ dc_1\ dc_2\ dc_3$$

where $C_k^{max} \le y$ is the upper bound on the individual's WTP for program k. Let $G_C(C_1,..,C_K)$ be the joint distribution function associated with the density $g_C(C_1,..,C_K)$, and let $G_{(k)}(C_1,..,C_K)$ denote the partial derivative of this joint distribution with respect to the k^{th} argument: $G_{(k)}(C_1,..,C_K) \equiv \partial G_C(C_1,..,C_K)/\partial C_k$. An equivalent way of expressing (73) is

$$\text{(73')} \qquad \Pr\{C_1 \le A_1,\ C_2 \ge A_2, \text{ and } C_3 \le A_3\} = \int_{A_2}^{C_2^{max}} G_{(2)}(A_1,C_2,A_3)\ dc_2.$$

As noted above, the existing CV literature on multi-item valuation generally assumes independence among the C_k's (Loomis, Hanemann and Wegge 1990, Hoehn 1991, Hoehn and Loomis 1993). In that case, the response probability formula for the suite of responses is the product of the marginal response probabilities for the individual programs

$$\text{(74)} \qquad \Pr\{C_1 \le A_1,\ C_2 \ge A_2, \text{ and } C_3 \le A_3\} = \Pr\{C_1 \le A_1\}\cdot\Pr\{C_2 \ge A_2\}\cdot\Pr\{C_3 \le A_3\}.$$

With stochastic dependence among the C_k's, however, the factoring into marginal probabilities does not apply, and one must deal directly with the joint response probability. In general, this involves a K-dimensional integral, as in (73). The question is whether there are stochastic specifications that yield tractable formulations for this integral while still allowing for a general covariance structure among the random components of preferences. Achieving this is not simple.

A natural starting point is to look for some multivariate generalization of the logistic distribution that retains a closed-form representation of the multivariate cdf for the η_k's. Arnold (1992) reviews a variety of multivariate distributions that have the univariate logistic as their marginal distribution. But, most of these entail severe restrictions on the covariance structure. For example, the simplest formulation of a multivariate logistic cdf is

$$F_\eta(\eta_1,..,\eta_K) \;=\; (1 + \textstyle\sum_k e^{-\eta_k})^{-1} \;, \tag{75}$$

but this imposes a correlation among values for pairs of programs of $\rho(\eta_k,\eta_l) = 0.5$. Arnold describes some other multivariate logistic distributions, but these still involve correlation coefficients restricted to a subset of the interval [-1,1] that is likely to be unsuitable for CV applications. Wu, Cosslett and Randall (1992) use McFadden's (1978) Generalized Extreme Value distribution for the ε_k's to generate a nested logistic distribution for the η_k's. Suppose, for example, that $(\varepsilon_0, \varepsilon_1,..,\varepsilon_K)$ can be partitioned in a two-level clustering. In the top level, there are R groups; in the r^{th} group, there are S_r elements. The multivariate cdf for the ε_k's then takes the form

$$F_\varepsilon(\varepsilon) = \exp\left[-\sum_{r=1}^{R}\left(\sum_{s=1}^{S_r} \exp(-\varepsilon_S)^{\frac{1}{\theta_r}}\right)^{\theta_r}\right] \tag{76}$$

where the θ_r's are parameters that determine within-group correlation.[4] The correlation coefficients are (Johnson and Kotz 1972, p.256):

$$\mathrm{corr}(\varepsilon_k,\varepsilon_l) = \begin{cases} 1 - \theta_r^2 & \text{if } \varepsilon_k \text{ and } \varepsilon_l \text{ are both in the } r^{th} \text{ group,} \\ 0 & \text{if } \varepsilon_k \text{ and } \varepsilon_l \text{ are in different groups.} \end{cases}$$

Since there is usually a restriction that $\theta_r \leq 1$, this does not allow for any negative correlation among the ε_k's. More generally, there does not appear to be any multivariate logistic distribution that would allow for a full range of positive or negative correlation among preferences for programs.

When K = 2, the problem can be solved by using Mardia's (1970) method for generating a bivariate distribution with specified marginals and an unrestricted correlation. Let $F_i(\eta_i)$ be the marginal distribution for η_i, i = 1, 2. Let $G(x,y;\rho)$ be some bivariate cdf with an unrestricted correlation coefficient, ρ, and with marginal distributions $G_X(x)$ and $G_Y(y)$. One creates the transformed variables $\eta_1^* = G_X^{-1}[F_1(\eta_1)] \equiv J_1(\eta_1)$ and $\eta_2^* = G_Y^{-1}[F_2(\eta_2)] \equiv J_2(\eta_2)$ and endows them with the specified bivariate distribution, thereby inducing a joint distribution for η_1 and η_2, $G[J_1(\eta_1), J_2(\eta_2); \rho]$. For example, one could use the standard logistic distribution, $F_i(\eta) = (1 + e^{\eta})^{-1}$, combined with a standard bivariate normal for $G(x,y;\rho)$. This is convenient way to generate bivariate distributions with a general correlation structure, but it does not readily extend to multivariate distributions for K > 2.[5]

In the multivariate case, the best approach is to use a multivariate normal distribution. While this requires numerical evaluation it permits a general covariance structure that offers maximum flexibility in modelling preferences for multiple programs. Until recently, this would not have been regarded as a

---

[4] The distribution in (75) is essentially a special case of (76) with R = 1. It should be emphasized that the distributions of stochastic components discussed in Sections 4.1.2-4 may all need to be modified by censoring or truncation, along the lines described in Section 2.3, in order to satisfy the economic restriction that $0 \leq C_k \leq y$ and perhaps also to ensure some pattern of stochastic ordering among the u_k's.

[5] Lee (1983) applies this method to create a generalized selectivity bias model; Ward (1995) uses it for double-bounded CV.

feasible option. While many software packages will evaluate a bivariate normal integral, numerical integration methods are required for $K > 2$ that until recently were considered impractical.[6] For multivariate probit models (discussed further in Section 4.1.3), numerical approximations were used based on formulas developed by Clark (1961) for the moments of the maximum of (non-iid) normal random variables; this maximum is not itself a normal variate, but the approximation treats it as though it were. In practice, however, the approximation turns out to be poor when individual variances are highly dissimilar or there are negative correlations (Horowitz, Sparmonn and Daganzo, 1984), and this approach is rarely used now. Lerman and Manski (1981) proposed an alternative approach based on sampling from multivariate normal distributions via Monte Carlo simulation and evaluating multinomial choice probabilities based on the proportion of outcomes in the empirical simulation. Since the empirical cumulative distribution (ecdf) converges to the true cdf (given certain regularity conditions), any function of the ecdf will converge to the corresponding function of the cdf itself. The Lerman and Manski method combined simulation of the likelihood function for given parameter values with iteration over parameter values to achieve numerical maximization of the likelihood function. In application, this proved to be almost as burdensome as numerical quadrature of the integrals. A breakthrough occurred when McFadden (1989) and Pakes and Pollard (1989) found a different method of simulating multinomial probabilities that required far fewer simulations yet provided consistent and almost efficient parameter estimates. This has led to the development of a variety of highly efficient simulation algorithms; the leading algorithm is currently the Geweke-Hajivassiliou-Keane (GHK) simulator.[7] This has been applied to the analysis of contingent ranking data by Hajivassiliou and Ruud (1994) and Layton (1995).

So far, we have focused on dependent multivariate distributions for the ε_k's, η_k's or C_k's. The C_k's can be thought of as latent variables which generate the observed binary responses via the crossing of a threshold, as in (3) or (6a,b). Instead of modelling dependence among these latent continuous variables, an alternative is to model dependence among the binary response variables directly. This type of model, known as multivariate logit or simultaneous logit, was introduced by Cox (1972), Nerlove and Press (1973) and Schmidt and Strauss (1975), and has recently been the focus of much attention in the biometrics literature. Let Y_k denote the binary response for the k^{th} program in a single-bounded CV format with bid A_k, and let $y_k = 0$ or 1 denote the realization of this variable. The quadratic exponential model (Cox, 1972; Zhao and Prentice, 1990) characterizes the joint distribution of the Y_k's as a log-linear model with third order and higher terms zeroed out

$$\text{(77)} \qquad \Pr\{Y_1 = y_1,..,Y_K = y_K\} = (1/\Delta) \exp\left(\Sigma_k\, \gamma_k y_k + \Sigma_{k<l}\, \gamma_{kl} y_k y_l \right)$$

where Δ is a normalization constant which ensures that the probabilities sum to unity. This implies that

$$\ln\left[\frac{Pr\left\{Y_k = 1 \mid Y_l = y_l,\ Y_{k'} = 0,\ k' \neq k,l\right\}}{Pr\left\{Y_k = 0 \mid Y_l = y_l,\ Y_{k'} = 0,\ k' \neq k,l\right\}}\right] = \gamma_k + \gamma_{kl} y_l$$

and

6 Chapman and Hanemann (1993) use the bivariate normal distribution combined with the Box-Cox model (72) in a CV study where $K = 2$.

7 For further discussion of these approaches see Hajivassiliou (1993), Keane (1993), and Hajivassiliou and Ruud (1994).

$$\ln\left[\frac{Pr\{Y_l = 1 \mid Y_k = y_k,\ Y_{k'} = 0,\ k' \neq k,l\}}{Pr\{Y_l = 0 \mid Y_k = y_k,\ Y_{k'} = 0,\ k' \neq k,l\}}\right] = \gamma_l + \gamma_{kl} y_k \ .$$

Hence, in this model, γ_k is the log of the odds ratio for Y_k conditional on $Y_{k'} = 0$ all $k' \neq k$

$$\ln\left[\frac{Pr\{Y_k = 1 \mid Y_{k'} = 0,\ k' \neq k\}}{Pr\{Y_k = 0 \mid Y_{k'} = 0,\ k' \neq k\}}\right] = \gamma_k \ ,$$

which, in the Box-Cox model (72), would take the form

$$\gamma_k = \alpha_k + b(y-A_k)^\lambda - by^\lambda.$$

The γ_{kl} term measures the association between Y_k and Y_l; it characterizes how the conditional distribution of Y_k depends on Y_l,

$$\ln\left[\frac{Pr\{Y_k = 1, Y_l = 1 \mid Y_{k'} = 0,\ k' \neq k,l\} \cdot Pr\{Y_k = 0, Y_l = 0 \mid Y_{k'} = 0, k' \neq k,l\}}{Pr\{Y_k = 1, Y_l = 0 \mid Y_{k'} = 0,\ k' \neq k,l\} \cdot Pr\{(Y_{k'} = 0, Y_l = 1 \mid Y_{k'} = 0, k' \neq k,l\}}\right] = \gamma_{kl}$$

and, equally, how the conditional distribution of Y_l depends on Y_k. Y_k and Y_l are positively associated if $\gamma_{kl} > 0$, and negatively if $\gamma_{kl} < 0$. The parameters to be estimated are b, λ, the α_k's (or an underlying parametric function, $\phi(q^k)$, which generates them) and the γ_{kl}'s, which may themselves be expressed as functions of covariates. Estimation is by maximum likelihood (Fitzmaurice and Laird, 1993) or the method of moments (Liang, Zeger and Qaqish, 1992).

Though it is relatively tractable, this approach to creating a stochastic dependence among the evaluations of multiple programs has the disadvantage that one cannot directly recover from (77) an explicit multivariate distribution of the C_k's. Moreover, when applied to different elicitation formats that generate different types of discrete response such as single-bounded versus double-bounded (discussed in Section 4.2) or referendum-style voting versus ranking (discussed immediately below), the approach in (77) implies a separate and distinct underlying distribution of the C_k's for each format. In effect, this method of modelling stochastic dependence provides a statistical association among the dependent variables but not a structural association, a criticism made in a different context by Heckman (1978) and Maddala (1983). In contrast, directly specifying a dependent multivariate WTP distribution yields a structural model that can be applied consistently across different elicitation formats.

4.1.3 Contingent Ranking and Related Techniques

In this section we discuss contingent ranking and two related techniques, contingent choice and mixed choice/ranking. In the referendum CV format, the respondent is presented with a set of programs, each offering some set of environmental amenities q^k at a cost A_k, and votes "yes" or "no" on each program. An alternative is to have the respondent *rank* the programs from most to least preferred. This is known as contingent ranking. It was introduced by Beggs et al. (1981) and Chapman and Staelin (1982) for automobile and college choices, respectively, and first applied to environmental valuation by Rae (1982, 1983); recent applications are summarized in Layton (1995). In this case, the response is not a set of 1's and 0's but a set of integers from 1 to K giving each program's rank. Index the items so that k =

1 is rated highest and k = K lowest. Given the RUM setting, the probability of this ranking is

$$\text{(78)}\quad \Pr\{\text{item 1 ranked 1}^{st}, \text{ item 2 ranked 2}^{nd}, \ldots, \text{ item K ranked last}\} = \Pr\{v_1(q^1,y\text{-}A_1,\varepsilon_1) \geq v_2(q^2,y\text{-}A_2,\varepsilon_2) \geq \cdots \geq v_K(q^K,y\text{-}A_K,\varepsilon_K)\}.$$

Suppose the utility functions have an additively random structure as described in footnote 12, $u_k = \bar{v}_k(q^k,y\text{-}A_k) + \varepsilon_k$; the Box-Cox models (11) and (68) would be examples. Define the utility differences $\delta_k \equiv \bar{v}_k(q^k,y\text{-}A_k) - \bar{v}_{k+1}(q^{k+1},y\text{-}A_{k+1})$ and let $\bar{\eta}_k \equiv \varepsilon_{k+1} - \varepsilon_k$, k = 1,..,K-1. The ranking response can then be expressed in terms of $f_{\bar{\eta}}(\bar{\eta}_1,..,\bar{\eta}_{K-1})$, the joint density of the $\bar{\eta}_k$'s, as

$$\text{(79)}\quad \Pr\{\text{item 1 ranked 1}^{st}, \text{ item 2 ranked 2}^{nd}, \ldots, \text{ item K ranked last}\} = \int_{-\infty}^{\bar{\delta}_1} \cdots \int_{-\infty}^{\bar{\delta}_{K-1}} f_{\bar{\eta}}(\bar{\eta}_1,\ldots,\bar{\eta}_{K-1})\, d\bar{\eta}_1 \ldots d\bar{\eta}_{K-1} \quad .$$

Before discussing the statistical models that can be generated from (78) or (79), four points should be noted about the economic implications of ranking responses. First, with ranking responses as with referendum choices, the response probabilities depend essentially on differences in utilities. Therefore, there will generally be some identifiability restrictions like those noted in footnote 11. Second, with contingent ranking and related techniques, it is not necessary that the baseline status quo (receive q^0, pay $A_0 = 0$) be included among the items for respondents to consider. There may be practical advantages in doing this, for example with the mixed choice/ranking model to be described below, but it is not required in order to obtain a valid measure of individual preferences. Of course, if the status quo is not included, one cannot recover from (78) or (79) an estimate of the parameters uniquely associated with the utility function $v_0(q_0,y,\varepsilon_0)$; but, one will be able to recover an estimate of those associated with $v_k(q^k,y,\varepsilon_k)$, $k \neq 0$.[8] Third, regardless of whether or not the status quo is included as an item to be ranked, it is *not* generally possible to characterize the ranking response in terms of the multivariate WTP distribution. With referendum responses, as we noted in Section 2.2, it makes no difference whether one follows the approach to model formulation associated with Hanemann (1984a) based on the specification of indirect utility functions, or the approach associated with Cameron (1988) which directly specifies a WTP distribution. Both approaches lead to exactly the same response probability formula for referendum responses. This does not carry over to contingent ranking or contingent choice. In those cases, the responses generally cannot be expressed in term of the WTP's for the individual items (the C_k's). For example, suppose that K = 2 and the individual ranks the items such that

$$v_1(q^1,y\text{-}A_1,\varepsilon_1) \;\geq\; v_2(q^2,y\text{-}A_2,\varepsilon_2).$$

This is *not* equivalent to the condition that

$$C_1 - A_1 \;\geq\; C_2 - A_2.$$

---

[8] If one uses some form of hedonic model, it may be possible to extrapolate the parameters of $v_0(q_0,y,\varepsilon_0)$ from those of the utility functions $v_k(q^k,y,\varepsilon_k)$, $k \neq 0$. Freeman's (1991) objection to the use of rating or ranking for preference elicitation is essentially that the omission of a status quo alternative prevents one from recovering $v_0(q_0,y,\varepsilon_0)$.

Therefore, the joint density of C_1 and C_2 cannot usefully be employed to characterize the probability of the ranking responses.[9]

Fourth, when one thinks of the observed responses in terms of an underlying RUM model, there is an obvious link between ranking and *choice*. In a standard choice exercise, respondents are presented with K items and asked to choose one of them. Choice thus corresponds to a *partial ranking* where just one item is ranked, the most preferred:[10]

(80) $\Pr\{k^{th} \text{ item chosen}\} = \Pr\{k^{th} \text{ item ranked first}\} = \Pr\{v_k(q^k,y-A_k,\varepsilon_k) \geq v_l(q^l,y-A_l,\varepsilon_l),\ l = 1,..,K\}$.

Assume the utility functions have an additively random structure like that leading to (79). For simplicity, let k = 1 be the item chosen. Define the utility differences $\delta_k \equiv \bar{v}_1(q^1,y-A_1) - \bar{v}_k(q^k,y-A_k)$ and let $\eta_k \equiv \varepsilon_k - \varepsilon_1$, k = 2,..,K; the choice probability can then be expressed in terms of $f_\eta(\eta_2,..,\eta_K)$, the joint density of the η_k's, as

$$(81) \qquad \Pr\{\text{item 1 chosen}\} = \int_{-\infty}^{\delta_2} \cdots \int_{-\infty}^{\delta_K} f_\eta(\eta_2,\ldots,\eta_K)\, d\eta_2 \ldots d\eta_K \ .$$

If there are only 2 items, item 1 with (q^1, A_1) and item 0 with $(q^0, A_0 = 0)$, (81) reduces to (4). That is to say, referendum-like voting of the type considered in sections 2 and 3 is simply a choice among K = 2 alternatives; conversely, the contingent choice model is a generalization of voting to more than two courses of action. If the ε_k's are multivariate normal, then $f_\eta(\eta_2,..,\eta_K)$ is also multivariate normal and (81) is a multivariate generalization of the probit model in (14), known as multinomial probit (Hausman and Wise, 1978). If the ε_k's are iid extreme value variates with scale parameter τ, the η_k's are iid logistic variates with scale parameter τ and (81) is a multivariate generalization of the logit model in (15), known as multinomial logit (McFadden, 1974). In that case, the joint distribution function, $F_\eta(\eta_2,..,\eta_K)$, takes the form[11]

$$(82a) \qquad F_\eta(\eta_2,..,\eta_K) = \prod_{k=2}^{K} \left[1 + e^{-\eta_k/\tau}\right]^{-1}$$

and the choice probability in (81) has a closed-form representation[12]

$$(82b) \qquad \Pr\{\text{item 1 chosen}\} = e^{\bar{v}_1(q^1,y-A_1)/\tau} \Big/ \sum_{k=1}^{K} e^{\bar{v}_k(q^k,y-A_k)/\tau}$$

---

[9] An exception occurs when there are no income effects, as in (18a,b). In that case, the two conditions are equivalent and the ranking response can be characterized in terms of the multivariate WTP distribution. The general non-equivalence of the two characterizations arises because of the non-additivity of WTP and WTA -- the sum of the WTP (WTA) for a change from q^0 to q^1 plus the WTP (WTA) for a change from q^1 to q^2 does *not* in general equal the WTP (WTA) for the change from q^0 to q^2.

[10] This equivalence requires a regularity condition which is described in Barbera and Pattanaik (1986).

[11] Note the difference between this and (75), in which the η_k's are not independent.

[12] In addition to identification problems with the utility differences, there may also be a problem associated with the scaling parameter τ; conventionally one imposes the restriction that $\tau = 1$.

$$= \left[1 + \sum_{k=2}^{K} e^{\bar{v}_k(q^k, y-A_k)/\tau \; - \; \bar{v}_1(q^1, y-A_1)/\tau} \right]^{-1} .$$

Let P(k|C) denote the probability that item k is the one chosen when the alternatives available are the set C. The multinomial logit choice probabilities possess the property that, for any sets C and T, where C $\subseteq$ T, and for any alternatives k, l $\in$ C,

$$\frac{P(i \mid C)}{P(j \mid C)} = \frac{P(i \mid T)}{P(j \mid T)} . \tag{83}$$

This is the *independence of irrelevant alternatives (IIA)* property of choice (Luce, 1959). McFadden (1974) and Yellot (1977) establish that RUM choice probabilities satisfy IIA if and only if the u_k's are independent extreme value variates with a common scale parameter. However, as Debreu (1960) first observed, IIA is not necessarily a plausible or reasonable property for choice behavior since it generates what is known as the "red bus/blue bus" paradox (Ben-Akiva and Lerman, 1985). The generalized extreme value distribution (76) relaxes IIA partially, but not within individual groups; the multivariate normal distribution eliminates it entirely.

Just as choice is a form of ranking, the converse also holds: ranking can be seen as a series of choices. Suppose there are four items and the individual ranks items 1, 2, 3, and 4 in descending order. Given some regularity conditions, the joint probability of this ranking response can be expressed as a product of conditional choice probabilities

(84) Pr {item 1 ranked 1st, item 2 ranked 2nd, item 3 ranked 3rd, item 4 ranked last} =

$$P(1 \mid 1,2,3,4) \cdot P(2 \mid 2,3,4) \cdot P(3 \mid 3,4)$$

where the first term on the RHS is the conditional probability that 1 is chosen out of {1,2,3,4}, conditional on 2 being chosen out of {2,3,4} and 3 being chosen out of {3,4}; the second term is the conditional probability that 2 is chosen out of {2,3,4} conditional on 3 being chosen out of {3,4}; and the third term is the unconditional probability that 3 is chosen out of {3,4}. In general, however, these conditional probabilities involve complex expressions because the conditional probability of the utility associated with the most favored item depends on the ordering of the other items. The *cascading choice theorem* of Luce and Suppes (1965) establishes that, when the u_k's satisfy the IIA property, these conditional probabilities are *independent* of the ordering of the less favored choices. Hence, (84) holds as a simple product of *unconditional* choice probabilities on the RHS. This comes about if and only if the individual u_k's are independent extreme value variates with a common scale parameter (Strauss, 1979; Yellot, 1980). In that case the RHS of (84) involves the product of a set of multinomial logit choice probabilities; with the additively random formulation used above, this takes the form

(85) Pr {item 1 ranked 1st, item 2 ranked 2nd, ... , item K ranked last} =

$$\prod_{k=1}^{K-1} \left[e^{\bar{v}_k(q^k,\, y-A_k)/\tau} \Big/ \sum_{l=k}^{K} e^{\bar{v}_l(q^l,\, y-A_l)/\tau} \right]$$

where one conventionally sets $\tau = 1$.[13]

To summarize, if the ε_k's in an additively random model have the extreme value distribution, the ranking response probabilities take the form in (85). In all other cases, including additive models where the ε_k's follow the normal or generalized extreme value distribution, the ranking response probabilities take the form given in (78) or (79). Until recently, applications of contingent ranking used the logit formulation. As noted above, the multivariate normal distribution affords a more general covariance structure for the u_k's and, while there is no closed form representation for the choice probabilities in (79), Layton (1995) finds that simulation techniques such as GHK are eminently practical.

Compared to contingent choice, contingent ranking offers two advantages. First, some parameters of the utility function not estimable with data on only the single most preferred choice may be estimable with data on the complete ranking; this can occur when the distribution of the most preferred item in the sample data is highly skewed. Second, rankings provide more statistical information than choice, which leads to tighter confidence intervals around parameter estimates.[14] In practice, however, it turns out that the extra information comes at a cost; the general experience is that some rankings are often inconsistent and they tend to become increasingly noisy as the lower ranked items are included, causing attenuation in the estimated coefficients of explanatory variables. When one thinks of the cognitive burden involved, this should not be a surprise. Ranking every item in the set is harder than identifying the best item. Some items may be easier to rank than others and, as Ben-Akiva et al. (1992) suggest, people may use different decision criteria for different rankings (e.g., the weights on attributes change when they are selecting the best item versus lower-ranked items). Growing respondent fatigue could explain why lower rankings might be less reliable. Alternatively, while the top- and bottom-ranked items might be easy to discern, it could be harder to discriminate among the items in the middle.

Chapman and Staelin (1982) framed the issue as a trade-off between reducing sampling variance when one incorporates more of the rankings versus greater bias due to increasing unreliability. They suggested systematically varying the number of ranks included in the analysis and stopping when there was a significant decrease in goodness of fit or significant evidence of changing estimates of coefficients. In their empirical study, which involved the linear model (12) applied to rankings of 3 or 4 items, they found it desirable to restrict the analysis to the ranking of the top two items. Hausman and Ruud (1987) estimate a similar utility model for rankings of 8 items and apply both a Durbin-Hausman specification test for the IIA property and a consistency test for constant ratios of attribute coefficients; both tests reject the logit ranking model (85). The same result was obtained by Ben-Akiva at el. (1992) when they analyzed rankings of four items. To deal with this, Hausman and Ruud propose a modification of (85) whereby the scale parameter, τ, varies across rankings. This allows for differences in cognitive burden, decision criterion, or carefulness among the rankings. Let τ_k be the scaling parameter associated with the k^{th} ranking, k = 1,..,K-1, where τ_{K-1} is normalized to unity; smaller values of τ_k correspond to less noisy

[13] The ranked probit and logit models in (79) and (85) are not be confused with ordered probit and logit models, presented below in equation (88).

[14] For an exploration of both these issues, see Waldman (1983).

rankings. Hausman and Ruud's rank-ordered heteroscedastic logit model is[15] [16]

(86) Pr {item 1 ranked 1st, item 2 ranked 2nd, ... , item K ranked last} =

$$\prod_{k=1}^{K-1} \left[e^{\bar{v}_k(q^k,\ y-A_k)/\tau_k} \Big/ \sum_{l\,=\,k}^{K} e^{\bar{v}_l(q^l,\ y-A_l)/\tau_k} \right] .$$

They find that the τ_k's significantly increase with k, though not monotically. Ben-Akiva et al. (1992) find the same. But, they also find that, while their first two rankings can be combined together, their third is still not consistent with the other two rankings; even after re-scaling, the rankings still violate the IIA property assumed by (86). When Layton (1995) analyzed partial rankings data, he found that the rank-ordered heteroscedastic logit model was inappropriate; he concluded that the variation in parameter estimates for different rankings was caused not by heteroscedasticity across rankings but rather by a violation of IIA. To handle this, he employed a rank-ordered heteroscedastic multivariate normal model (79), derived from a random coefficient version of (12). He found this fitted the data well and eliminated the inconsistency among the rankings.

While more experience is needed, the rank-ordered heteroscedastic multivariate normal model seems promising since it provides considerable flexibility and permits one to model a variety of response phenomena. For example, the variances of the u_k's could be made a function of the degree to which the item is similar in its characteristics to other items in the set, as measured by some distance metric. One hypothesis is that greater similarity among items makes it harder to rank them. An alternative hypothesis is that greater *divergence* makes it harder to rank items.[17] One might also be able to obtain from respondents an assessment of how hard they found the rankings, and then parametrize the variances of the u_k's as a function of task difficulty.[18]

In addition to changing the stochastic specification of RUM preferences, another way to improve the results from ranking experiments is to modify the ranking task so as to lighten the cognitive burden on respondents. Since respondent fatigue could be a problem, it may be desirable to limit the number of items in the set to be considered. It may also be useful to mix ranking with choice, in order to limit the

---

[15] It should be noted that (86) does not appear to be strictly consistent with the RUM framework. If the u_k's are extreme value variates with a common scale parameter, as noted above, (85) is equivalent to (78). But, if they have *different* scale parameters, one no longer obtains the logistic for the distribution of their differences. We are not aware of a stochastic specification of the u_k's that would make (86) equivalent to (78). In effect, (86) implies a separate set of preferences for each ranking activity.

[16] The heteroscedastic logit model is also used for choices where data from separate choice tasks are to be combined in a single estimation; the choice probabilities for each task are given by (82b), with the τ_k's varying across the data sets. For examples and references, see the chapter by Adamowicz et al.

[17] The results in Mazzotta and Opaluch (1995) provide some support for the second hypothesis; these are consistent with observing a higher variance when items differ on more attributes.

[18] For example, an interesting model by Swait and Adamowicz (1996) makes the variances a function of an entropy measure of the differences in attribute vectors.

number of ranking evaluations. Suppose the survey includes the status quo along with other programs. One might ask respondents to separate the programs into those that are better than the status quo, and therefore worth paying for, versus those that are not, and then have them rank the former but not the latter. Ranking items which one does not want may be a futile task (Mitchell and Carson, 1989). Suppose the individual considers that items 1,..,I are an improvement over the status quo, while items I+1,.,,K are not, and she ranks the former in descending order; with this mixture of choice and ranking, the response probability is the following blend of (78) and (80) :

Pr {item 1 ranked 1st, item 2 ranked 2nd, .., item I ranked Ith, and items I+1,..,K deemed unacceptable} =

$$\text{(87)} \qquad \Pr\{v_1(q^1,y-A_1,\varepsilon_1) \geq v_2(q^2,y-A_2,\varepsilon_2) \geq \cdots \geq v_I(q^I,y-A_I,\varepsilon_I) \geq v_0(q^0,y,\varepsilon_0), \text{ and } v_0(q^0,y,\varepsilon_0) \geq v_k(q^k,y-A_k,\varepsilon_k),\ k = I+1,..,K\}.$$

With the additively random formulation of (79) and (81) the response probability formula becomes:

Pr {item 1 ranked 1st, item 2 ranked 2nd, .., item I ranked Ith, and items I+1,..,K deemed unacceptable} =

$$\text{(87a)} \qquad \int_{-\infty}^{\bar\delta_1} \cdots \int_{-\infty}^{\bar\delta_I} \int_{-\infty}^{\delta_{I+1}} \cdots \int_{-\infty}^{\delta_K} f_{\bar\eta\eta}(\bar\eta_1, \ldots, \bar\eta_I, \eta_{I+1}, \ldots \eta_K)\, d\bar\eta_1 \ldots d\bar\eta_I\, d\eta_{I+1} \ldots d\eta_K \ ,$$

where $\bar\delta_k \equiv \bar v_k(q^k,y-A_k) - \bar v_{k+1}(q^{k+1},y-A_{k+1})$, $k = 1,..,I-1$, $\bar\delta_I \equiv \bar v_I(q^I,y-A_I) - \bar v_0(q^0,y-A_0)$, $\delta_k \equiv \bar v_k(q^k,y-A_k) - \bar v_0(q^0,y-A_0)$, $k = I+1,..,K$, $\bar\eta_k \equiv \varepsilon_{k+1} - \varepsilon_k$, $k = 1,..,I-1$, $\bar\eta_I \equiv \varepsilon_I - \varepsilon_0$, $\eta_k \equiv \varepsilon_k - \varepsilon_0$, $k = I+1,..,K$. While an extreme value distribution for the ε_k's yields a closed form expression for (87a), a multivariate normal distribution seems preferable for the reasons given above.

We end this discussion of contingent ranking and choice with some comments on (1) welfare evaluation with these preference elicitation methods, and (2) their relation to conjoint analysis. Contingent choice and ranking differ from referendum CV in two ways. Instead of valuing a single item, as in referendum CV, the goal with choice and ranking is to estimate a utility function, or a system of utility functions, that can be used to value *many* items, perhaps including items not covered in the CV survey whose value can be extrapolated from the utility model fitted to the survey responses. Second, contingent choice and ranking break the direct tie between response probabilities and welfare measures that is present in referendum (binary choice) CV. The response probability for ranking or multivariate choice is constructed from a multivariate utility distribution rather than a multivariate WTP distribution (i.e., from the joint distribution of the u_k's rather than the C_k's). Formulation and estimation of monetary measures of value proceeds separately from, and subsequent to, the modelling of survey responses.

It should also be noted that, in a multivariate setting, there is an important distinction between *conditional* and *unconditional* welfare evaluation. The welfare measures used with referendum CV, as discussed in Section 2, are conditional since they measure the WTP or WTA to have one specific q^k versus another; the C_k in (5) is constructed so as to equate the conditional utility u_k with the conditional utility u_0. Multivariate choice, by contrast, generally takes an unconditional approach to welfare evaluation and focuses on the WTP or WTA for a change in the attributes of one or more of the items unconditional on the individual's selecting those items. It values a change in the entire vector of q^k's, from, say, $Q^0 \equiv (q^{10},..,q^{k0},..,q^{K0})$ to $Q^1 \equiv (q^{11},..,q^{k1},..,q^{K1})$ and employs the unconditional indirect utility function, $v(Q,A,y,\varepsilon) \equiv \max\{v_1(q^1,y-A_1,\varepsilon_1),..,v_K(q^K,y-A_K,\varepsilon_K)\}$; the welfare measure is the monetary compensation that equates the post-change unconditional utility with the pre-change unconditional utility. This turns out to be somewhat like taking conditional welfare measures and then weighing them by the probability that the item involved is the one selected.[19]

The term *conjoint analysis* was coined by Green and Srinivasan (1978) to cover a variety of preference elicitation techniques used in psychometrics and market research.[20] These include forms of ranking and choice, as well as *rating*, where respondents are asked to rate a set of items on a scale from, say, 1 to 10 or 1 to 100.[21] A distinctive feature is that respondents typically perform multiple tasks, and a separate preference model is often fitted for each respondent.[22] The items to which subjects respond are selected by the researcher to span some attribute space, often with special designs such as fractionated sampling (Green, 1974; Louviere, 1988). Data collected through one form of elicitation may be converted into another equivalent form for the purposes of analysis. For example, subjects may be asked to rate a large set of items but, for model fitting, their ratings are converted into implied rankings or choices. Or, subjects may be asked to make a sequence of choices among pairs of items, but their pairwise choices are then converted into an implied ranking over the entire set of items. The statistical models used to analyze conjoint responses are typically not designed to satisfy the economic restrictions discussed in Section 2.3. They tend to be either linear models of the form $u_k = \Sigma_r \beta_r x_{kr} + \varepsilon_k$, where x_{kr} is the level of the r^{th} attribute in item k, which is essentially equivalent to (12), or *ideal point* models where $u_k = - \Sigma_r \beta_r(x_{kr} - x^*_{kr})^2 + \varepsilon_k$, where x^*_{kr} is the individual's ideal level for the r^{th} attribute.[23] Aside from the multiple responses per subject and the form of the statistical model, to the extent that conjoint data are analyzed as ranking or choices, there is no essential difference with the contingent ranking or choice techniques described above.

The older tradition in conjoint analysis is to take the ratings as a cardinal measure of preference and fit a utility function directly to them, $u_k = v_k(q^k,y-A_k,\varepsilon_k)$, using ordinary least squares regression, linear

---

[19] For a formal development of unconditional welfare measurement in multivariate RUM choice models, see Hanemann (1983a, 1985b).

[20] For summaries of what is now a vast literature, see Green and Srinivasan (1990) or Carroll and Green (1995). The review by Louviere (1994) emphasizes choice-based conjoint analysis, which we have called contingent choice.

[21] Sometimes, the rating scale is cast in terms likelihood of purchase -- e.g., 1 means "I would never buy this," while 100 means "I would definitely buy it." In that case, it may be possible to apply to the conjoint ratings the statistical models of preference uncertainty discussed at the end of Section 4.3.

[22] Some form of cluster analysis may be used to group subjects into market segments for the purposes of demand analysis and projection.

[23] The x^*_{kr}'s may be elicited directly as part of the survey, or they may be estimated from subjects' responses. Another model commonly used in conjoint analysis, besides linear and ideal point utility models, is the *part-worth function*; this posits an additively separable utility function $u_k = \Sigma_r \psi_r(x_{kr}) + \varepsilon_k$ where $\psi_r(\cdot)$ is the sub-utility (part-worth) function for the r^{th} attribute. In practice, $\psi_r(x_{kr})$ is usually estimated for only a few selected levels of x_{kr}, with the value for other levels of x_{kr} obtained by extrapolation. This corresponds to estimating the step function hedonic model (71) without cross-product terms, and then making a piecewise linear approximation to an additively separable function.

programming or some other method of estimation.[24] [25] As cardinal measures of utility, ratings have the advantage of providing more information since they convey intensity of preference, unlike the ordinal measures of preference associated with ranking or choice. However, the extra information comes at some cost. Pooling responses across individuals presumes inter-personal comparability of utility, the practical feasibility of which has been debated in economics since the 1930's.[26] For example, a rating of "5" may mean different things to different people. Certainly, it often happens that different subjects center their ratings on different parts of the scale. Fitting a separate response model for each subject can be viewed as one way of circumventing inter-personal comparisons; but the problem reappears as soon as one extrapolates from individual subjects to market segments. A recent CV application of conjoint analysis by Roe et al. (forthcoming), which pooled ratings across individuals, attempted to control for differences in centering by having one scenario (the status quo) that was rated by everybody and then estimating utility difference functions where the differences in ratings between the common scenario and each other scenario were regressed on the differences in attributes between the scenarios. This appeared to reduce the noise in the data and it produced the most significant coefficients.[27] But, even if it corrected for differences in centering, it does not preclude other factors, e.g., differences in range or variance, that could invalidate inter-personal utility comparisons.

---

[24] CV applications of traditional conjoint analysis, where the fitted utility model is used to compute monetary welfare measures for changes in attributes, appear to have started with Mackenzie (1993).

[25] Ordinary least squares and linear programming are known as *decompositional* methods of estimation, since they fit the utility function in one step. *Compositional* methods of conjoint analysis, by contrast, fit the function in stages, first estimating part-worth functions for individual attributes (for example, by asking subjects to scale alternative levels of each attribute taken separately) and then estimating weights for combining the part-worth functions into an overall utility index. The compositional approach is also used in *multiattribute utility assessment* (MAUA) (Keeney and Raiffa, 1976; von Winterfeldt and Edwards, 1986). As with conjoint analysis, the general practice in MAUA is to estimate a separate utility function for each subject. The multivariate utility models used in MAUA generally allow for some interaction among the part-worth functions; the particular form is chosen by testing whether the subject's preferences satisfy certain axioms. To elicit the part-worth functions, MAUA generally uses pairwise choices or *matching* (essentially equivalent to open-ended CV, in which subjects are asked to identify the level of an attribute required to adjust the utility of one item so that it exactly matches that of another); for model fitting, calibration is generally used rather than statistical estimation. MAUA typically involves more intensive data collection from subjects than conjoint analysis. How this affects the results in prediction of behavior or assessment of value is still an open question. Gregory et al. (1993) advocate switching from conventional CV to MAUA because they consider the compositional approach to preference elicitation more reliable. In our view, the case for a compositional approach is more normative than positive: this may be how individuals *should* make judgments, but we are not sure it is how they *do* make judgments.

[26] Approaches which model behavior in terms of ordinal preferences, such as referendum CV, contingent ranking and choice, or demand function estimation, avoid inter-personal comparisons of utility because they allow a separate monotone transformation of the utility index for each individual.

[27] Faced with the same problem, Mackenzie (1993) simply included the subject's mean rating over all items as an explanatory variable in each rating regression; first differencing is a better solution.

Aside from the problem of inter-personal comparisons, one may doubt whether subjects can always measure their preferences with the precision implied by a cardinal rating. To ease the burden on respondents, some conjoint researchers employ scales with only a few integer ratings; for example, both Mackenzie and Roe et al. use a scale from 1 ("very undesirable") to 10 ("very desirable"). This may be more convenient, but it also can turn the ratings into an *ordinal* measure of preference; it is not clear whether, on this scale, subjects treat an "8" as twice a "4", or a "4" as twice a "2". With this scale it may be more appropriate to apply *ordered* probit or logit models, due originally to Aitchison and Silvey (1957) and Ashford (1959) and popularized in the social sciences by Zavoina and McElvey (1975), which treat the dependent variable as an ordering and permit one to test whether it is measured on an ordinal or cardinal scale. These models postulate that, while the observed dependent variable is an ordinally-ranked categorical variable, it is derived from a latent continuous variable reflecting the respondent's underlying sentiment. Let the stated rating of the k^{th} item by R_k; in this case $R_k = 1, 2, 3,.., 9$ or 10. In addition to the underlying conditional indirect utility function $u_k = v_k(q^k,y\text{-}A_k,\varepsilon_k)$, the model involves a set of thresholds $\bar{r}_1,..,\bar{r}_9$, which can be treated as constants or as functions of covariates. The relation mapping the subject's underlying utility for the item into her stated rating is given by: [28] [29]

$$
\begin{array}{llll}
R_k = 1 & \text{if} & v_k(q^k,y\text{-}A_k,\varepsilon_k) \le \bar{r}_1 \\
R_k = 2 & \text{if} & \bar{r}_1 \le v_k(q^k,y\text{-}A_k,\varepsilon_k) \le \bar{r}_2 \\
R_k = 3 & \text{if} & \bar{r}_2 \le v_k(q^k,y\text{-}A_k,\varepsilon_k) \le \bar{r}_3 \\
\cdot & & \cdot \\
\cdot & & \cdot \\
R_k = 9 & \text{if} & \bar{r}_8 \le v_k(q^k,y\text{-}A_k,\varepsilon_k) \le \bar{r}_9 \\
R_k = 10 & \text{if} & v_k(q^k,y\text{-}A_k,\varepsilon_k) \ge \bar{r}_9.
\end{array}
\tag{88}
$$

If ε_k enters $v_k(q^k,y\text{-}A_k,\varepsilon_k)$ additively and is normally distributed, this is an ordered probit model; if it is logistic, this is an ordered logit model. The coefficients to be estimated include the parameters of $v_k(\cdot)$, the parameters of the joint distribution of the ε_k's, and the $\bar{r}_k$'s (possibly subject to a normalization restriction);[30] estimation is by maximum likelihood or, if there are grouped data, by minimum chi-squared. If the $\bar{r}_k$'s are *equally spaced*, i.e., $\bar{r}_2 \equiv \bar{r}_1 + \Delta$, $\bar{r}_3 \equiv \bar{r}_1 + 2\Delta$, .. , $\bar{r}_9 \equiv \bar{r}_1 + 8\Delta$, the R_k's are an

---

[28] In their analysis, Roe et al. use a double-hurdle tobit model for regressions of ratings or ratings differences. This gives R_k a censored normal distribution with spikes at "1" and "10"; within the interval (1,10), it makes R_k a continuous variable, not an integer.

[29] The model could be adjusted for differences among subjects in the centering of ratings by compressing and relabelling some of the ratings categories. Suppose some subjects' ratings fall in the interval (1,5) while others' fall in the interval (5,9); one could, for example, convert the ratings to a common, five-point scale, making "5" the highest rating and relabelling the second group's ratings so that they range from "1" to "5" instead of "5" to "9."

[30] To the extent that the $\bar{r}_k$'s are taken to be the same across subjects, this makes the u_k's cardinal utility measures and implies inter-personal comparability of utility.

approximately cardinal measure of preference. This can be tested as a restriction on the estimated $\bar{r}_k$'s.

In summary, while there is a close relationship at the conceptual level between rating, ranking, and choice, there are important practical differences in terms of the cognitive burden on respondents and the implications for inter-personal comparability of utility. When comparisons have been made among alternative modes of preference expression, the general experience is that inconsistencies appear: different modes imply different preferences even though, in theory, a single utility function should suffice to generate all the various types of response. The differences in welfare measures estimated from the fitted utility functions can be quite substantial (Mackenzie; Roe et al.). The evidence suggests that this task-dependence is a general phenomenon, and is not confined to non-use values, environmental goods, or contingent valuation. For researchers, it has two implications. First, it creates a challenge to develop explicit models of task-dependence in the stochastic and/or deterministic components of the response model. Second, it requires the researcher to exercise judgment in deciding which set of results is the most appropriate to her research goal.

4.1.4 Random Coefficient Models

So far we have discussed how the deterministic or stochastic components of RUM models can be made to reflect variation in the item being valued. In this section we touch on aspects of model specification dealing with variation *among respondents*. This can arise, for example, from socio-economic differences, or from heterogeneity in preferences for programs.

From a statistical perspective, the CV evaluations of programs can be viewed as panel data -- we pool observations on individuals over programs, rather than over time. To illustrate this, we employ the linear model (12), adding subscript i to denote the i^{th} individual. Some or all of the program attributes mentioned in the preceding paragraph may vary across individuals; therefore, let x_{ikr} denote the r^{th} attribute of program k for the i^{th} individual. In addition, x_{is} denotes the s^{th} attribute or socio-demographic characteristic of the i^{th} individual. Thus, we write (12) as $u_{ik} = \alpha_{ik} + \beta_y y_i + \varepsilon_{ik}$ where $\alpha_{ik} \equiv$ *intercept* $+ \Sigma_r \beta_r x_{ikr} + \Sigma_s \beta_s x_{is}$; different models correspond to different treatments of the intercept and different stochastic specifications for ε_{ik}. In panel data terminology, many of the models discussed above would be categorized as *one-way fixed effects* models. This applies, for example, to the model in (72) with $\lambda = 1$,

$$u_{ik} = \alpha_k + \Sigma_r \beta_r x_{ikr} + \Sigma_s \beta_s x_{is} + \beta y_i + \varepsilon_{ik} \tag{89}$$

where the ε_{ik}'s are all iid with mean zero and variance σ_ε^2 and the α_k's are fixed parameters to be estimated from the data along with σ_ε^2 and the various β's. This formulation is appropriate if preferences for different programs are viewed as parametric shifts of a utility function. An alternative is the *one-way random effects* model, which treats the program-specific terms as varying randomly across programs; this would apply if one saw the sampled programs as drawn from a larger set of programs. This model takes the form

$$u_{ik} = \alpha + \Sigma_r \beta_r x_{ikr} + \Sigma_s \beta_s x_{is} + \beta y_i + \upsilon_k + \omega_{ik} \tag{90}$$

where υ_k is a stochastic component which is constant across respondents and represents unmeasured factors specific to the k^{th} program. We assume that $E\{\upsilon_k\} = E\{\omega_{ik}\} = 0$, $E\{\upsilon_k{}^2\} = \sigma_\upsilon^2$, $E\{\omega_{ik}{}^2\} = \sigma_\omega^2$, $E\{\upsilon_k \upsilon_l\} = 0$ if $k \neq l$, $E\{\omega_{ik}\omega_{jl}\} = 0$ if $i \neq j$ or $k \neq l$, and $E\{\upsilon_k \omega_{il}\} = 0$ for all i,k and l; σ_υ^2 and σ_ω^2 are to be estimated from the data along with α and the various β's. In addition to (89) and (90), *two-way* fixed- and random-effects models are possible, with both program-specific and individual-specific parameters or variance components. The two-way fixed-effect model is[31]

$$u_{ik} = \gamma_i + \alpha_k + \Sigma_r\, \beta_r x_{ikr} + \Sigma_s\, \beta_s x_{is} + \beta y_i + \varepsilon_{ik} \tag{89'}$$

where the γ_i's and α_k's are fixed parameters and, as before, the ε_{ik}'s are iid with mean zero and variance σ^2_ε. The two-way random effect model is

$$u_{ik} = \alpha + \Sigma_r\, \beta_r x_{ikr} + \Sigma_s\, \beta_s x_{is} + \beta y_i + \nu_i + \upsilon_k + \omega_{ik} \tag{90'}$$

where, in addition to the above specification of the variances and covariances of the υ_k's and ω_{ik}'s, one has $E\{\nu_i\} = 0$, $E\{\nu_i^2\} = \sigma^2_\nu$, $E\{\nu_i\nu_j\} = 0$ if $i \neq j$, and $E\{\nu_i\upsilon_k\} = E\{\nu_i\omega_{jk}\} = 0$ for all i,j and k. The two-way fixed-effect formulation is appropriate if individual differences in preferences for a program are viewed as parametric shifts of the utility function for that program; the random-effects formulation is appropriate if one sees the respondents as drawn from a larger population with randomly varying preferences. It is also possible to have a *mixed* two-way model combining, say, fixed-effects over programs with random effects over individuals:

$$u_{ik} = \alpha_k + \Sigma_r\, \beta_r x_{ikr} + \Sigma_s\, \beta_s x_{is} + \beta y_i + \nu_i + \omega_{ik} \tag{91}$$

where $E\{\nu_i^2\} = \sigma^2_\nu$, $E\{\omega_{ik}^2\} = \sigma^2_\omega$, and $E\{\nu_i\omega_{jk}\} = 0$ for all i,j and k.

The choice between fixed- and random-effects formulations has been the subject of much debate. This originally focused on conventional linear regression, but has recently been extended to quantal response models. From the point of view of estimation, an important difference is that, while the random-effects approach adds only two parameters, σ^2_υ and σ^2_ν, the fixed-effects approach adds N γ_i's and K α_k's. This is computationally burdensome when K or, more likely, N is large. In linear models, the ordinary least squares estimate of the γ_i's and α_k's is unbiased. In quantal response models estimated by maximum likelihood, there are only asymptotic properties. In both cases, the estimates of γ_i's and α_k's are consistent only as N and K approach infinity. In practice, K will be small with panel data, making the estimate of the γ_i's inconsistent. This is the *incidental parameters* problem (Neyman and Scott, 1948): an increase in N provides no extra information about γ_i, but just raises the number of γ_i's. For large-sample properties, any estimate of the fixed-effect γ_i's is unreliable when K is finite, even if N tends to infinity.

In linear models, the inability to obtain consistent estimates of the γ_i's does not affect estimation of the β's; fixed- and random-effects estimators of the β's are unbiased and consistent as long as N tends to infinity. In quantal response models, unlike linear models, the maximum likelihood estimators of parameters are not independent of each other (Hsiao, 1986). In fixed-effect quantal models with finite K, the inconsistency of γ_i is transmitted to the estimators of the α_k's and β's. Hence, even if N tends to infinity, the maximum likelihood estimate of the β's remains inconsistent. Generalizing Neyman and Scott (1948), Anderson (1970) showed a way to obtain consistent estimates of structural parameters by maximizing a conditional likelihood function conditioned on a minimum sufficient statistic for the incidental parameters, if one exists that is independent of the structural parameters. Chamberlain (1980) applied this to obtain a computationally convenient and consistent estimate of the β's in fixed-effects logit

[31] We noted above the common practice in conjoint analysis of eliciting multiple responses from subjects and then fitting a separate response probability model for each subject. That is equivalent to a fixed-effects formulation, with individual-specific intercepts and slope coefficients.

models. For fixed-effects probit models, however, there does not appear to exist a consistent estimator of the β's.

In random-effects quantal models, maximum likelihood estimation does provide consistent estimation of the β's. However, the estimation is computationally burdensome because of the complex nature of the likelihood function in this case. For simplicity, we focus on the mixed model (91). Even when the ω_{ik}'s are independently distributed over i and k in (91), $E\{u_{ik}u_{il}\} = \sigma^2_\nu \neq 0$, and the joint log-likelihood of a sample of responses $(y_{11},..,y_{it},..,y_{NT})$, where y_{it} is the i^{th} individual's response to the t^{th} discrete-response CV question, can no longer be written as the sum of the marginal log-likelihoods of the y_{it}'s, as in (47). Let $P(y_{it} \mid \nu; X_{it}, \theta)$ denote the probability formula for the i^{th} individual's response to the t^{th} CV question conditional on the random effect ν_i, given the vector of individual-specific and item-specific attributes, X_{it}, where θ is the vector of parameters not associated with the random effects (i.e. the β's and σ^2_ω). Let the density of ν be $g_\nu(\nu;\sigma^2_\nu)$. With this random-effects formulation, the log-likelihood function takes the form

$$(92) \qquad L(\theta,\sigma^2_\nu \mid y,X) \;=\; \Sigma_i \;\ln \int \Big[\, \Pi_t \; P(y_{it} \mid \nu; X_{it},\theta) \,\Big] \; g_\nu(\nu;\sigma^2_\nu)\; d\nu \;\; .$$

By contrast, if there is no heterogeneity among individuals (the ω_{ik} are iid and the individual-specific effects ν_i do not exist), the log-likelihood takes the far simpler form

$$(93) \qquad L(\theta \mid y,X) \;=\; \Sigma_i\; \Sigma_t \ln P(y_{it} \mid X_{it}, \theta).$$

Maximizing the random-effects log-likelihood, (92), is computationally challenging since, in general, there are integrals in each of the response probabilities $P(y_{it} \mid \nu; X_{it}, \theta)$. The first application to random-effects probit was by Heckman and Willis (1975); a faster computational algorithm was developed by Butler and Moffitt (1982). There are also several alternatives to maximum likelihood estimation. Chamberlain (1984) proposes a minimum distance estimator, and Avery, Hansen and Hotz (1983) describe a related generalized method of moments estimator. The biometrics literature tends to emphasize quasi-likelihood and similar techniques (Longford, 1993).[32]

In both linear and nonlinear models, empirically it can make a substantial difference to the estimates of the β's whether one adopts a fixed- or random-effects formulation. Hence, the choice of model is contentious. Mundlak (1978) suggests two criteria. One is the type of inference desired: the effects can always be considered random but, when using the fixed-effects model, the researcher makes inferences *conditional on the effects in the observed sample*. When the random-effects model is used, the researcher makes specific assumptions about the distribution of the random effects, and this permits *unconditional* or *marginal* inferences. The other is robustness to specification error. If the distributional assumptions about the error components are correct, the random-effects estimator is more efficient. But, if the distributional assumptions are incorrect, the random-effects estimator is inconsistent. Because the fixed-effects approach does not make any specific distributional assumption about the error components

[32] Another possibility, mentioned by Maddala (1986), is to simply ignore the correlation among responses and maximize (93) instead of (92); this yields estimates of the β's which are consistent but inefficient. However, the conventional formula for the variance-covariance matrix of these estimates is incorrect (Robinson, 1982). In a Monte Carlo study, Guilkey and Murphy (1993) find that, despite the mis-specification of the error structure, standard probit with a corrected covariance matrix often works relatively well.

of the model, it is more robust. For example, Mundlak argued that the conventional random-effects formulation assumes no correlation between the individual-specific effect, ν_i, and the individual-specific covariates, x_{ik}, whereas there is likely to be some correlation in practice. As an alternative, Chamberlain (1980) suggested estimating random-effects models with a linear dependence between the x_{it} and ν_i. Hsiao (1986) provides details, along with a description of specification tests for the independence of the ν_i's. In addition, he discusses specification tests for the random-effects model (92) against the no-correlation formulation (93). Lechner (1995) and Hamerle and Ronning (1995) are also useful on these topics.

We have focused so far on fixed- versus random-effects with respect to the *intercept* of the utility function; the same types of formulation can also be applied to the *slope coefficients*. For example, compared to (11), the Box-Cox model (10) is a one-way fixed-effect slope-coefficient model, since the income coefficient, β, varies with the item (program) being valued -- in the set-up of (89), where $\lambda = 1$, we write this as

$$u_{ik} = \alpha_k + \Sigma_r \beta_r x_{ikr} + \Sigma_s \beta_s x_{is} + \beta_k y_i + \varepsilon_{ik}. \tag{94}$$

The corresponding one-way, random-effect formulation is

$$u_{ik} = \alpha_k + \Sigma_r \beta_r x_{ikr} + \Sigma_s \beta_s x_{is} + (\beta + \upsilon_k) y_i + \omega_{ik}. \tag{95}$$

where $E\{\upsilon_k\} = 0$, $E\{\upsilon_k^2\} = \sigma_\upsilon^2$, $E\{\upsilon_k \upsilon_l\} = 0$ if $k \neq l$, and $E\{\upsilon_k \omega_{il}\} = 0$ for all i,k and l. This can also be written

$$u_{ik} = \alpha_k + \Sigma_r \beta_r x_{ikr} + \Sigma_s \beta_s x_{is} + \beta y_i + \zeta_{ik} \tag{95'}$$

where $\zeta_{ik} \equiv \upsilon_k y_i + \omega_{ik}$; because of the dependence on y_i, the error term is now heteroscedastic. Such models, also known as *random coefficient* models, were first introduced by Swamy (1970); Hausman and Wise (1978) give a probit application. A two-way, slope coefficient model is

$$u_{ik} = \alpha_k + \Sigma_r \beta_r x_{ikr} + \Sigma_s \beta_s x_{is} + (\beta + \nu_i + \upsilon_k) y_i + \varepsilon_{ik}, \tag{96}$$

where the individual-specific component, ν_i, and the program-specific component, υ_k, of the income coefficient are either fixed parameters to be estimated (the fixed-effect model) or random variables with some covariance structure to be estimated (the random-effect model). Estimation of these models raises essentially the same issues as varying intercept models.

The random-effect formulation can be extended to more general error component structures. In presenting (90), for example, we assumed that $\sigma_\upsilon^2 = E\{u_{ik} u_{jk}\}$ is the same for every pair of individuals and does not vary with k, while $\sigma_\nu^2 = E\{u_{ik} u_{il}\}$ is the same for every pair of programs and does not vary with i. Random coefficient models such as (95) are one way to introduce some heteroscedasticity and relax these restrictions. Another way is to specify directly a more complex covariance matrix for the error components; an example, mentioned in the previous section, is parameterizing the covariance terms in a multivariate normal distribution for contingent ranking as functions of similarity indices. Another extension is to have multiple levels of nesting for the variance components. For example, instead of the one-way individual effects assumed so far, one might identify a hierarchy of groups such that individuals in the same group are more similar than individuals in different groups.[33] The random terms in u_{ik} have both individual and group components. Individual components are all independent. Group components are independent between groups but perfectly correlated within groups. Some groups might be more homogeneous than other groups, which means that the variance of the group components can differ. Such *hierarchical* or *multi-level* models can be rich enough to provide a good fit to complex survey data, while still being parsimonious with the number of parameters. They can be estimated by the EM procedure (Raudenbush and Bryk, 1986), quasi-likelihood (Longford, 1988), iterated generalized least squares (Goldstein, 1991), or two-stage probit (Borjas and Sueyoshi, 1994). Chapter 12 provides a CV application.

In closing, we note that (92) is an example of *compounding*, where heterogeneity among individuals leads to the averaging or integration over a response probability formula. In this case, the compounding is with respect to a variance component, but one can also have compounding with respect to a location parameter; a classic example is the generation of the beta-binomial distribution by compounding a binomial distribution with a beta distribution for the binomial parameter. Another example is *aggregation* of quantal responses. This can occur when there is aggregate rather than individual data on CV responses, or when there is data on individual responses but no observations for one or more variables believed to affect the individual responses. Let X be some such variables; the aim is to estimate the structural coefficients associated with X despite the fact that X is not observed. For this purpose, we assume that the values which X can take follow some density, $f_X(X)$. For example, X could be the respondent's income, which is unobserved in the data, and $f_X(\cdot)$ is the income distribution that applies to the sample of respondents. Given the i^{th} observed response, y_i, the probability of this response can be written

$$P(y_i) = \int P(y_i|X)\ f_X(X)\ dX. \tag{97}$$

The functional form of $f_X(\cdot)$ is assumed to be known by the researcher. If its parameters are known (e.g. it is log-normally distributed with mean and variance known from some external source), one uses the aggregate response probability formula in (97) to estimate the parameters of $P(y_i|X)$. If the parameters of $f_X(\cdot)$ are not known, they are estimated along with those of $P(y_i|X)$ by maximizing the likelihood function based on (97).

4.2 Double-Bounded CV

The double-bounded version of discrete response CV, proposed by Hanemann (1985a) and Carson (1985), follows up on the initial question with a second question, again involving a specific dollar cost to which the respondent can respond with a "yes" or a "no." Let A denote the amount of the first bid. The amount presented in the second bid depends on the response to first bid; if the individual answered "no" to A, the second bid is some lower amount $A_d < A$, while if she answered "yes" it is some higher amount $A_u > A$. Thus, there are four possible response sequences: (a) both answers are yes; (b) both answers are no; (c) a yes followed by a no; and (d) a no followed by a yes. Following the logic of (6b), for any given underlying WTP distribution $G_C(\cdot)$, the probability of the responses is given by

$$\Pr\{\text{yes/yes}\} \equiv P^{yy} = 1 - G_C(A_u)$$

[33] In educational statistics, where this originated, one might have information about individual students, class-level information, and school-level information.

$$\Pr\{\text{no/no}\} \equiv P^{nn} = G_C(A_d) \tag{98}$$

$$\Pr\{\text{yes/no}\} \equiv P^{yn} = G_C(A_u) - G_C(A)$$

$$\Pr\{\text{no/yes}\} \equiv P^{ny} = G_C(A) - G_C(A_d).$$

Any of the WTP distributions in Appendix Tables 2-4 can be used to construct these response probability formulas. Given these, the log-likelihood function for the double-bounded model is

$$\ln L = \sum_{i=1}^{n} [\ I_{yy} \ln\ P_i^{yy} + I_{yn} \ln\ P_i^{yn} + I_{ny} \ln\ P_i^{ny} + I_{nn} \ln\ P_i^{nn}\ ] \tag{99}$$

where I^{xz} is an indicator function that equals one when the two responses are xz, and zero otherwise. Estimation is by maximum likelihood or one of the other techniques described in Section 3.4.[34]

Using maximum likelihood, Hanemann, Loomis and Kanninen (1991) compared this double-bounded model with a single-bounded model estimated from the responses to the initial bid, A, using the logit model (15). They found that the double-bounded model provided a substantial gain in precision for the variance-covariance matrix of the coefficient estimates, leading to much tighter confidence intervals for the estimate of median WTP. Similar results have been obtained by other researchers who have compared the double- and single-bounded approach. Hanemann, Loomis and Kanninen also found that the double-bounded data yielded a lower point estimate of median WTP than the single-bounded data. The same result has been found in many other applications. Both the gain in efficiency and the direction of change in the estimate of median WTP can be explained by a poor selection of the initial bid. When setting their bids, Hanemann, Loomis and Kanninen had very limited pre-test results and their guess at the point value of median WTP turned out to be far too low. With the hindsight of the field survey results, an optimal design would have used a higher initial bid, A. The double-bounded format corrected for this on the second bid—the follow-up bid A_u came closer to where an optimal design would have placed the bid, and this helped considerably to pin down the estimate of median WTP. In effect, A_u provides insurance against too low a choice of A, and A_d provides insurance against too high a choice of A.

Table 4, which parallels Table 1 for the single-bounded logit model (15), presents calculations of the bias and asymptotic variance of the maximum likelihood estimator for the double-bounded logit model. In the first three examples, the initial bid amount is the median value and the follow-up bid amounts are placed at increasing distances from the median value.[35] As in the single-bounded case, as the bid amounts are placed further away from the middle of the distribution, the biases and variances increase. The fourth example places the initial bid amounts on one side of the median value—at the 62nd, 82nd and 92nd percentiles of the distribution—with follow-up bid amounts placed $100 on either side of these initial bids.

---

[34] Alberini (1995c) calls the normal version of (99) *interval-data probit*. As noted in the introduction, this is formally identical to the likelihood function for failure time data with censoring, where the analyst knows upper and lower bounds but not the exact time of failure. Kalbfleisch and Prentice (1980) is a standard reference on such models; Lancaster (1990) covers economic applications to unemployment etc.

[35] In general, a one-point design will not identify the two parameters of the model. But with the double-bounded model, two pieces of information are elicited from each respondent, yielding enough information to estimate both parameters.

In this case, most of the biases and the asymptotic variance are greater than any of the cases where the initial bid is the median value. But the biases and variance are far less than in the analogous single-bounded model. This is due to the adaptive nature of the double-bounded model. Even though the initial bid design is poor, the follow-up bids provide a second chance to recoup the situation.

With the double-bounded model, experimental design involves the specification of three bid amounts per observation. Although each respondent is offered only two bids, an initial and a follow-up, three bids enter the log-likelihood function because *a priori* it is not known which follow-up bid amount will be offered. Optimal designs for the double-bounded logit and probit models are presented in Table 5. In each case, the initial bid amount is the median value, and follow-up bid amounts are symmetrically placed around the median. The results in Table 5 have interesting implications for the importance of "bounding" individuals' WTP values. We can derive the probability of receiving a "yes" response to the follow-up bid *conditional* on the probability of receiving a "yes" response to the initial bid. If the initial bid is the median value, then the probability of receiving a "yes" response to the initial bid is exactly equal to .5 as is the probability of receiving a "no" response. Given the initial response, we can derive the conditional probability for the response to the follow-up bid by dividing the unconditional c.d.f. of the second bid value by .5. We obtain the result that the D-optimal bid design for the logit model "bounds" slightly over 75% of the respondents' WTP values, while that for the C-optimal bid design "bounds" exactly half. A general rule of thumb might be that it is best to bound between 50% and 75% of observations. Note that this implies that the follow-up bids should not be so high or so low that all observations are bounded.

After the second valuation question, one could follow up with a third or fourth closed-ended valuation question, again using with a higher bid amount if the individual responded"yes" to the previous bid and a lower bid amount if the response was "no." The additional questions should give sharper bounds on the individual's WTP. An empirical application of the triple-bounded approach is described in Bateman, Langford and Rasbash (this volume). However, in a Monte Carlo simulation analysis, Cooper and Hanemann (1995) found that, compared to the double-bounded approach, adding a third round of closed-ended questions produced a relatively small gain in efficiency. Most of the statistical benefit from the extra information was reaped in going from the single- to the double-bounded format.

Although there is a gain in efficiency, there can also be some bias in going from a single- to a double-bounded format because there is evidence that some of the responses to the second bid are inconsistent with the responses to the first bid. We have assumed so far that, for a given individual, the *same* utility model and stochastic process generate the answers to both valuation questions. If so, this implies certain relations between the response probabilities for first and second bids which can be used to construct nonparametric tests of consistency among the responses. Suppose, for example, that there are two separate overlapping sets of bids: {$A = \$10$, $A_u = \$20$, $A_d = \$5$) and ($A = \$20$, $A_u = \$40$, $A_d = \$10$). The first set starts with \$10, and then either goes up to \$20 or down to \$5; the second starts with \$20, and then goes up to \$40 or down to \$10. Individuals would be assigned one or the other bid structures at random. It is a standard result in probability that:

$$\text{(100)} \qquad \Pr\{\text{"yes" to \$20}\} \ = \ \Pr\{\text{"yes" to \$20 | "yes" to \$10}\} \cdot \Pr\{\text{"yes" to \$10}\}.$$

The LHS of (100) can be estimated by taking the respondents who received the second set of bids and calculating the proportion who said yes to the first bid in that set. The RHS of (100) can be estimated by taking the respondents who received the first set of bids and multiplying the proportion who said yes to the first bid in that set by the proportion of those respondents who subsequently said yes to the second bid in that set. If the responses to the first and second bids come from the same stochastic generating process, the estimates of the LHS and RHS of (100) should be the same, except for sampling variation. An alternative versions of this test is to compare two estimates of the probability Pr {\$10 ≤ WTP ≤ \$20}. One estimate takes the respondents who received the first bid structure and calculates the proportion of yes/no responses. The other estimate takes those who received the second bid structure and calculates the proportion of no/yes responses. Again, these two proportions should be the same, except for sampling variation. The first test was used by Hanemann (1991), and the second by McFadden and Leonard (1993); in both cases, the test failed. Similar failures of consistency tests have been observed in other applications of double-bounded CV.[36] [37]

How should one deal with the inconsistency? Several approaches are found in the literature. One approach is to continue pooling the responses to the first and second bids despite the inconsistency; while there may be some bias in the coefficient estimates, there also can be a gain in efficiency which offsets this (Alberini, 1995c). A second approach, due originally to Cameron and Quiggin (1994), is to treat the responses to the two bids as though they were valuations of separate items, using one or another of the models described in Section 4.1.[38] In Cameron and Quiggin's general model, there are separate WTP distributions for the first and second responses, both based on the linear utility function (12), leading to marginal cdf's $G_k(C_k)$, k = 1, 2, with marginal means $E\{C_k\} \equiv \mu_k = \alpha_k/\beta_k$ and variances $var\{C_k\} \equiv \sigma_k^2 = 1/\beta_k$. In terms of the underlying utility functions, this is a one-way fixed-effects formulation with separate response-specific intercepts, α_k, and slope coefficients, β_k. In addition, there is a non-zero correlation between the first and second responses, represented by a correlation coefficient ρ. Thus, the two WTP distributions are neither identical nor independent. Let $g_{12}(C_1, C_2; \rho)$ be the resulting bivariate WTP density. Instead of (98), the double-bounded response probabilities now take a form analogous to (73):

$$
\begin{aligned}
P^{yy} &= \int_{A}^{C_1^{\max}} \int_{A_U}^{C_2^{\max}} g_{12}(C_1, C_2; \rho)\, dC_1\, dC_2 \\
P^{yn} &= \int_{A}^{C_1^{\max}} \int_{C_2^{\min}}^{A_U} g_{12}(C_1, C_2; \rho)\, dC_1\, dC_2 \qquad (101) \\
P^{ny} &= \int_{C_1^{\min}}^{A} \int_{A_L}^{C_2^{\max}} g_{12}(C_1, C_2; \rho)\, dC_1\, dC_2 \\
P^{nn} &= \int_{C_1^{\min}}^{A} \int_{C_2^{\min}}^{A_L} g_{12}(C_1, C_2; \rho)\, dC_1\, dC_2 \quad .
\end{aligned}
$$

[36] There are also various parametric tests of consistency. For example, one can analyze the two sets of responses separately, and then test whether they come from the same distribution.

[37] The results obtained by Bateman, Langford and Rabash (this volume) suggest that there may be an additional inconsistency in going from the double- to the triple-bounded format. Almost all the existing literature has focused on consistency between the single- and double-bounded formats.

[38] This is analogous to the practice in contingent ranking, discussed in footnote 82, of dealing with inconsistencies in rankings by assuming a separate set of preferences for each ranking.

In their empirical application, Cameron and Quiggin make $g_{12}(C_1,C_2;\rho)$ a bivariate normal, so that (101) becomes a correlated, bivariate version of the probit model (14).[39]

Cameron and Quiggin also consider formulations that lie somewhere between (98) and (101). They find empirically that they cannot reject the hypotheses that $\mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2$ (i.e, $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$). If it were also the case that $\rho = 1$, then (101) would reduce to (98), where $G_C(\cdot)$ is the common univariate WTP distribution that generates both responses. With their data, however, Cameron and Quiggin find they *can* reject the hypothesis that $\rho = 1$; ρ is significantly positive but less than unity.[40] Thus, the two WTP distributions are identical, not independent, but not perfectly correlated either. In terms of the underlying utility functions, this is equivalent to a one-way random-effects formulation for (12) with common (fixed) slope coefficient, β, and random intercepts, α, that are correlated draws from a common distribution. Let $G(\cdot,\cdot;\rho)$ be a correlated, bivariate WTP distribution with common marginal cdf's given by $G_C(\cdot)$; in Cameron and Quiggin's empirical application these are bivariate and univariate normal, respectively. Instead of (101), the double-bounded response probabilities now take the form[41]

$$
\begin{aligned}
P^{yy} &= 1 - G_C(A) - G_C(A_u) + G(A,A_u;\rho) \\
P^{yn} &= G_C(A_u) - G(A,A_u;\rho) \\
P^{ny} &= G_C(A) - G(A,A_d;\rho) \\
P^{nn} &= G(A,A_d;\rho).
\end{aligned}
\tag{102}
$$

Alberini, Kanninen and Carson (1994a) develop a two-way random-effect version of this model, with individual-specific as well as response-specific effects.

A third approach postulates that the inconsistency arises as some form of response effect, and models this effect explicitly. This can be done in various ways, depending on the type of response effect believed to occur. For example, many CV studies implement the double-bounded approach in a manner that makes the second valuation question come as something of a surprise. After having been assured

[39] Cameron and Quiggin consider that negative as well as positive preferences are possible, and they model this along the lines illustrated in the lower panel of Figure 3, rather than through a mixture model as in the upper panel of Figure 3. They set $C^{min} = -\infty$ and $C^{max} = \infty$, but their WTP distribution needs to be modified by censoring or truncation for the reasons mentioned in Section 2.3.

[40] Ward (1995) notes that these hypotheses tests may not be robust if the true WTP distribution is non-normal. Even if the distribution were normal, Alberini (1995c) points out that, since testing for $\rho = 1$ involves a hypothesis at the boundary of the parameter space, the distribution of the test statistic is non-standard and the size of the test statistic is likely to diverge from its nominal level, making the test results in Cameron and Quiggin somewhat unreliable.

[41] Alberini (1995c) conducts a Monte Carlo analysis of the bivariate normal version of (102) and the normal version of the double-bounded model (98), simulating the error that arises when one model is erroneously fitted to data actually generated by the other. She finds that the double-bounded estimates of mean/median WTP can be surprisingly robust to low values of ρ; even if the double-bounded model is technically misspecified, it is often superior to the bivariate model in terms of mean square error.

during the presentation of the first bid that the item will cost \$A, the respondent is suddenly informed that the cost could instead be \$$A_u$ (or \$$A_d$). She may not react to this new information in a neutral manner. If she had said "yes" to the initial bid, she might see the increased second bid as a crude attempt at bargaining, which she could well resent (Altaf and DeShazo, 1994). Or, she might view it as implying a possible cost overrun, which she could similarly resent. In both cases, she might then be inclined to say "no" to the second bid, not because she doesn't like the item but because she objects to the way in which it is now being offered to her. If so, this would cause a greater preponderance of yes/no responses than would be anticipated on the basis of the "yes" response to the first question alone. In a similar vein, if she had said "no" to the initial bid, she might see the reduced second bid as an attempt to pressure her to change her mind, which she could resent, or she might view it as an indication that she is now being offered an inferior and cheaper version of the commodity. If so, this would cause a greater preponderance of no/no responses than would be anticipated on the basis of the "no" response to the first question alone. Some researchers have suggested a different scenario. Under this scenario, respondents who say "yes" initially feel called upon to assert and expand their commitment in the face of the "challenge" implied by the raised follow-up bid; hence, there could be a greater preponderance of yes/yes responses than would be anticipated on the basis of the "yes" response to the first question. Conversely, those who say "no" initially may feel guilty and decide that they should at least agree to the second bid, causing a greater preponderance of no/yes responses than would be anticipated on the basis of the "no" response to the first question. In one scenario, the second bid creates resentment; in the other, acquiescence.

Both of these scenarios can lead to the Cameron-Quiggin model (101), where there are two WTP distributions because the second bid induces a shift in the location parameter of the WTP distribution generating the second response. In terms of the of the Cameron-Quiggin model, these scenarios explain why μ_2 might be different from μ_1. In the resentment scenario, the second bid causes $\mu_2 < \mu_1$; in the acquiescence scenario, it causes $\mu_2 > \mu_1$. However, they can also be modelled directly as a single WTP distribution combined with a response effect. For the resentment scenario, one extends (98) by introducing a background disposition to say "no" that applies to the second but not the first valuation question. This changes the response probabilities from P^{yy}, P^{yn}, P^{ny}, P^{nn} as given in (98) to

$$\begin{aligned} \bar{P}^{yy} &= (1 - \theta_u)\, P^{yy} \\ \bar{P}^{yn} &= \theta_u\, P^{y} + (1 - \theta_u)\, P^{yn} \\ \bar{P}^{ny} &= (1 - \theta_d)\, P^{ny} \\ \bar{P}^{nn} &= \theta_d\, P^{n} + (1 - \theta_d)\, P^{nn} \end{aligned} \tag{103}$$

where θ_u is the background probability of saying "no" to the second bid given that the respondent said "yes" initially, θ_d is the background probability of saying "no" to the second bid given that the respondent said "no" initially, $P^y = 1 - G_C(A)$ is the probability of saying "yes" initially, and $P^n = G_C(A)$ is the probability of saying "no" initially. Under the resentment hypothesis, $\theta_u > 0$ and $\theta_d > 0$.[42] The acquiescence hypothesis is treated similarly by introducing into (98) a background disposition to say "yes", which leads to a modified set of response probabilities analogous to (103).

[42] We leave open the possibility that $\theta_u \neq \theta_d$. Both of these background probabilities could be parametrized as functions of covariates. Kanninen (1995) develops a model similar to (103) but with a background probability of saying "yes/yes."

Instead of modelling the second response in terms of resentment or acquiescence, Herriges and Shogren (1996) approach it in terms of anchoring. They postulate that the respondent's revises her valuation of the item after the first bid, forming some weighted average of the bid and her original WTP. In linear models like (12) they use an arithmetic average while, in the nonlinear model (19), they assume a geometric average; for simplicity, we focus here on the linear model. If C_1 is the individual's WTP for the item on receiving the first bid, $G_C(\cdot)$ is the cdf of C_1, A is the first bid, and C_2 is her WTP after the first bid, Herriges and Shogren postulate that

(104) $$C_2 = (1 - \gamma)C_1 + \gamma A.$$

for some $\gamma \in [0, 1]$; if $\gamma = 0$, there is no anchoring. Given (104), the double-bounded response probability formulas take the following form instead of (98)[43]

(105)
$$\begin{aligned} P^{yy} &= 1 - G_C[(A_u - \gamma A)/(1-\gamma)]) \\ P^{nn} &= G_C[(A_d - \gamma A)/(1-\gamma)] \\ P^{yn} &= G_C[(A_u - \gamma A)/(1-\gamma)] - G_C(A) \\ P^{ny} &= G_C(A) - G_C[(A_d - \gamma A)/(1-\gamma)]. \end{aligned}$$

In their empirical application, Herriges and Shogren find no evidence of anchoring in the valuation of an improvement in water quality by local residents, but definite evidence of anchoring for visiting recreationists.

Alternatively, instead of operating on the location parameter, Alberini, Kanninen and Carson (1994a) suggest that the response effect might shift the dispersion of the WTP distribution. They propose a model where, while $\mu \equiv E\{C_k\}$ is the same for both questions, $\sigma_k^2 \equiv \mathrm{var}\{C_k\}$ varies as a function of the difference between the bid presented to the respondent in the k^{th} question and the central tendency of her WTP for the item. Their formula allows for differences in the strength of this effect among questions

(106) $$\sigma_k^2 = \exp\,[\phi + \psi_k \cdot (\mu - A_k)^2], \qquad k = 1, 2$$

where A_k is the bid used in the k^{th} question.[44] Different signs for ψ_k correspond to different models of cognitive behavior. If $\psi_k < 0$, this is model where, the closer the individual is to indifference (i.e., what she has to pay for the item nearly exhausts what it is worth to her), the less certain she is about how to

[43] These are slightly different from the formulas published by Herriges and Shogren.

[44] In this case, some care must be exercised in translating back from the parameters of the WTP distribution to those of the underlying utility function. In cognitive heteroscedasticity models like (106), it would be natural to identify *two separate* components of var{C}: one (denoted $\tilde{\sigma}^2$) arises from the variation in preferences implied by the RUM formulation, while the other reflects the respondent's cognitive burden in answering the CV question. In terms of the linear model (12), for example, where $\mu = a/\beta$, the variance associated with RUM preferences is $\tilde{\sigma}^2 = 1/\beta$. The ϕ in (106) is intended to reflect $\tilde{\sigma}^2$; the formulation there implies a multiplicative relation between the two components of variance. It might instead be more natural to have an additive formulation: $\sigma_k^2 = \tilde{\sigma}^2 + \exp[\psi_k \cdot (\mu - A_k)^2]$.

answer the CV question. If $\psi_k > 0$, this is a model where the individual is less certain of her response when the bid is more surprising or implausible, in the sense that the costs strikes her as unusually high or unusually low. The model is thus a particular case of the Cameron-Quiggin model (101) where $\rho = 1$, $\mu_1 = \mu_2$, and σ_1^2 and σ_2^2 are given by (106). When Alberini, Kanninen and Carson apply this model to three different double-bounded data sets, they find no evidence of heteroscedasticity in two cases but significant evidence in the third case, with ψ positive and decreasing across questions ($\psi_1 > \psi_2 > 0$).

Several conclusions can be drawn from the literature on the double-bounded CV format. While, a variety of response-effect models exist for which there is some empirical support, most analyses apply a particular model to the data at hand rather than pitting many of them against one another. Thus, while it is clear that response effects can occur with the double-bounded format, there is less certainty about their specific form. Moreover, this is something that may vary with the circumstances of the interview. Whether or not the second question connotes bargaining or induces surprise, resentment or guilt, and whether or not the first question is seen as providing a hint as to how subjects should value the item, could depend on just what is said in the interview and how it is conducted.

For example, it might be possible to eliminate the surprise effect of the second valuation question by arranging for the interviewer to state, well before the first valuation question, that the cost of the item is expected to fall within a range from A_l to A_h. In the first valuation question, the interviewer selects one end of that range at random and asks whether the subject is willing to pay this amount. Suppose the interviewer selects A_l; if the subject says yes, the interviewer follows up with a second question about paying A_h, but he does *not* follow up with a second question if the subject says no to A_l. Conversely, if the interviewer starts with A_h for the first bid and the subject says no, he follows up with a second question about paying A_l, but does *not* follow up with a second question if the subject says yes to A_h. On average, the interviewer follows up with a second bid only half as often as with the conventional double-bounded format. Therefore, Cooper and Hanemann (1995) call this the "one-and-a-half-bounded" format. Through a simulation analysis they find that it provides parameter estimates much closer in efficiency to those associated with the double-bounded than the single-bounded format. Thus, it may offer most of the statistical advantages of the double-bounded format without the response effects.

Pending testing of the one-and-a-half-bounded format, we would recommend using the double-bounded format when collecting CV data because of the extra information it provides.[45] To be sure, statistical information could be maximized by just asking an open-ended WTP question. But, this ignores considerations of cognitive capacity. In our experience, even for market commodities, people often cannot simply state their maximum WTP off the top of their head; they go through a thought process of considering particular amounts and deciding whether or not they would be willing to pay those amounts. Thus, asking people directly for their WTP, even for many market commodities, is asking for more precision than many people can offer. The closed-ended format comes closer to how they think and what they can answer. The double-bounded approach stays within this constraint while providing more information. Our own experience has been that, even if it produces some bias compared to the single-bounded approach, the bias is in a conservative direction and is outweighed by the gain in efficiency.

4.3 Extending the Response Probability Model

---

[45] We would *not* recommend the triple-bounded format, because it appears to offer only a small gain in efficiency while having a greater chance of inducing response effects.

Most of the response probability models considered so far—including the logistic, normal, log-logistic, lognormal and Weibull—involve two-parameter probability distributions, with one location parameter and one scale parameter.[46] Adding more location or scale parameters can give greater flexibility, which allows a more sensitive modelling of the data, especially in the tails of the distribution.[47] This is important if the goal is to predict tail probabilities or to measure quantities, such as mean WTP, which are particularly sensitive to them.

There are several reason why a more flexible response probability model could be appropriate. First, even at the level of the individual respondent, the WTP distribution could possess a more complex shape than that yielded by conventional, two-parameter models. Second, the researcher may want to draw a distinction between the individual's underlying preferences and the way these are expressed through the survey medium; the additional parameters could be used to model response effects or the individual's uncertainty about her preferences. Third, even if the WTP distribution has a simple shape at the individual level, heterogeneity among respondents due to variation in preferences or unmeasured covariates may produce a more complex shape in the aggregate, population-level WTP distribution. Rather than explicitly compounding a two-parameter model as in (97), it may be convenient to use a three- or four-parameter distribution to approximate the aggregate response probability model.

The overdispersion model in Section 3.4. can be regarded as an instance where one adds a parameter to get a more realistic shape for a WTP distribution. Another example is the model with a spike at zero in (33). Besides this, the statistical literature contains a number of multi-parameter distributions which generalize the logistic, including Aranda-Ordaz's (1981) standardized asymmetric distribution[48]

(107) $$F_\eta(\eta) = 1 - [1 + \lambda e^{\eta}]^{1/\lambda},$$

which yields the logistic cdf when $\lambda = 1$ and the extreme value cdf when $\lambda = 0$, and the cubic-logistic of Morgan (1985) with parameters μ, $\gamma > 0$ and $\delta > 0$

(108) $$F_\eta(\eta) = \left[1 + e^{-\gamma(\eta-\mu) - \delta(\eta-\mu)^3}\right]^{-1} .$$

There are also Prentice's (1976) generalization of probit and logit, Stukel's (1988) generalized logistic model, and several other distributions reviewed in El-Saidi, Singh and Moussa (1992). The mixing approach to generalizing the logistic is discussed by Follmann and Lambert (1989), who apply a nonparametric mixture model with cdf

(109) $$F_\eta(\eta) = \sum_{m=1}^{M} \gamma_m \left[1 + e^{-\theta_m - \eta}\right]^{-1} ,$$

where M (the number of atoms in the mixture), the α_m's (coefficients unique to each atom), and the γ_m's

[46] There are additional coefficients in the model when other covariates are used along with the bid variable, but these all form part of the location and/or shape parameters.

[47] While adding parameters reduces the bias estimates of welfare measures and other quantities of interest, it also increases the variance because of the extra variation associated with the estimates of the additional parameters (Taylor, 1988).

[48] This was applied to discrete-response CV data by Johansson and Kristrom (1988).

(non-negative weights that sum to unity and constitute the nonparametric cdf of the logistic parameter θ) are all to be estimated from the data. Follmann and Lambert show that, in the case of a single covariate (e.g., the CV bid) which takes k distinct values, there is a restriction that $M \leq [(k+2)^{0.5} - 1]$ for parameter identifiability.[49]

The double-bounded model with a background disposition to answer "yes" or "no," as in (103), is an example of a model that introduces an additional parameter in order to represent a response effect in a CV survey. The single-bounded version was introduced by Hanemann (1983b); instead of (6b), the response probability is given by[50]

$$\text{(110)} \qquad \Pr\{\text{response is "yes"}\} = \theta + (1-\theta)\,[1 - G_C(A)]$$

where $\theta \in [0,1]$ is the background probability of saying "yes," which could be taken as a constant or made a function of covariates other than the bid, A; this model was derived from the concept of a background mortality rate in biometrics (Hasselblad et al., 1980). Kanninen (1995) notes an alternative interpretation based on an asymmetric version of Copas' (1988) transposition model, (62): if some proportion, θ, of the "no" responses is wrongly recorded as a "yes," but there is no error in recording the "yes" responses, this generates the response probability in (110).

Another model that uses a parameter to represent response effects is the model of "thick" indifference curves where a parameter $\delta \geq 0$ determines a zone of indifference or ambivalence. The idea that consumers might be insensitive to small differences in utility, thus creating a threshold which must be exceeded in order for them to express a preference for one alternative over another, goes back to Georgescu-Roegen (1936, 1958) and Luce (1956). It was first linked to the concept of a RUM model by Quandt (1956), and was formally incorporated into binary choice models by Deacon and Shapiro (1975) and Krishnan (1977). The first application to discrete-response CV was by Svento (1993).[51] In the single-bounded context, the respondent answers yes only if $v(q^1, y - A, \varepsilon) \geq v(q^0, y, \varepsilon) + \delta$; she answers no only if $v(q^1, y - A, \varepsilon) \leq v(q^0, y, \varepsilon) - \delta$; otherwise, if $[v(q^1, y - A, \varepsilon) - v(q^0, y, \varepsilon)] \leq |\delta|$, she answers "don't know" or abstains from making a choice. Recasting the indifference zone into monetary units via a parameter Δ, instead of the utility units measured by δ, one obtains a trichotomous choice model where, instead of (6b),

$$\text{(111a)} \qquad \Pr\{\text{response is "yes"}\} = 1 - G_C(A + \Delta)$$

$$\text{(111b)} \qquad \Pr\{\text{response is "no"}\} = G_C(A - \Delta)$$

$$\text{(111c)} \qquad \Pr\{\text{response is "don't know"}\} = G_C(A + \Delta) - G_C(A - \Delta).$$

---

[49] If one uses these distributions, it will still be necessary to apply censoring or truncation to ensure that the resulting WTP distribution satisfies the economic restriction that $0 \leq C \leq y$.

[50] This should be distinguished from the spike model, (33), where the response probability takes the form: Pr {response is "yes"} = $\theta[1-G_C(A)]$, where $\theta \equiv 1- \gamma$.

[51] Opaluch and Segerson (1989) propose that thick indifference curves can explain CV responses without formally modelling it in this way; Mazzotta and Opaluch (1995) model ambivalence as an increase in the variance of the random utility associated with the item being valued.

To estimate the model -- and to identify Δ -- one needs explicit data on abstentions or "don't knows."[52] Svento (1993) used the linear probit and logit models (14) and (15) in (111), and parametrized Δ on attributes of the respondent, finding that it increased slightly with income. Svento and Mantymaa (1994) tested for separate (asymmetric) Δ's in (111a) and (111b) and found that this actually provided a better fit than the symmetric formulation.[53]

The heteroscedastic model in (106) is an example of an approach where one represents response effects in terms of increased dispersion in preferences. In general with this approach, in addition to the random terms ε and η used above to represent variation in preferences among individuals in a population, one adds a new random term υ which represents the individual's uncertainty about her own preferences. As in (106), the extent of this uncertainty may vary with the circumstances of the CV survey. We now distinguish between C, as defined in (5), which is the individual's underlying preference for the item in terms of her true WTP, and $\tilde{C}$, which is how she perceives or expresses her WTP in the survey context. We assume that

(112a) $$\tilde{C} \equiv C + \upsilon$$

where υ is a random variable with a mean of zero (in the absence of incentives for strategic behavior) and some cdf $F_\upsilon(\cdot)$. It is the cdf of $\tilde{C}$, $\tilde{G}_C(\cdot)$, rather than $G_C(\cdot)$ which governs the responses to the CV survey. Thus, instead of (6b), the probability that the respondent answers yes in the survey is

(112b) $$\Pr\{\text{response is "yes"}\} = 1 - \tilde{G}_C(A).$$

Whether or not $F_\upsilon(\cdot)$ and $G_C(\cdot)$ can both be recovered depends on the stochastic specification and the form of the data.

Li and Mattsson (1995), who first introduced this model, used the linear model (12) for C with η a standard normal, which generates the probit model (14), combined with a normal distribution for υ having a mean of zero and a variance of σ_υ^2 . Thus, $\tilde{C} = C + \upsilon \equiv (\alpha/\beta) + \omega$ is normally distributed with a mean of $E\{\tilde{C}\} = E\{C\} \equiv \mu = \alpha/\beta$,[54] and a variance of $\sigma_{\tilde{C}}^2 \equiv \text{var}\{\tilde{C}\} = \sigma_\omega^2 = \sigma^2 + \sigma_\upsilon^2$, where $\sigma^2 \equiv \text{var}\{C\} = 1/\beta^2$. With this specification, σ^2 and σ_υ^2 cannot be separately identified from data on the discrete CV responses alone. However, Li and Mattsson had conducted a survey in which, after presenting a closed-ended, single-bounded valuation question, they asked "How certain were you of your answer to the

---

[52] While it is always good practice to record a response of "don't know" when this is spontaneously volunteered by a respondent, there is some question about the necessity of automatically providing an explicit "don't know" option, as suggested by the NOAA Panel (Arrow et al., 1993). A split-sample experiment by Carson et al. (1995) found that, if one treats "don't know" as "no" in analyzing CV responses, offering the "don't know" option does *not* change the breakdown of yes/no responses or alter the estimate of WTP.

[53] The trichotomous choice model in (111) and the asymmetric model of Svento and Mantymaa can also be viewed as versions of the ordered categorical dependent variable model in (88).

[54] This is also the median of C and $\tilde{C}$. For the purpose of exposition, we ignore the restriction that $0 \leq C \leq y$.

previous question?" Respondents were asked to record their reply on a scale from 0% to 100%.[55] Li and Mattsson interpreted the answer to the confidence question as information about $F_\upsilon(\cdot)$ that would suffice to recover the unknown parameters.

Suppose the individual said yes to a bid of A and then said 30% in response to the confidence question. The probability of this joint answer can be factored into

(113) $\Pr\{\text{says "yes" to A and 30\% confident}\} = \Pr\{\text{says 30\%} \mid \text{"yes" to A}\}\cdot\Pr\{\text{says "yes" to A}\}$,

where the second probability on the RHS corresponds to (112b). To obtain the first term on the RHS, Li and Mattsson assume that, if the subject says she is 30% confident of her response, she means "exactly 30%" in the sense that

$$
\begin{aligned}
\text{(114a)} \qquad 0.3 &= \Pr\{\tilde{C} \geq A \mid C\} \\
&= \Pr\{C + \upsilon \geq A\} \\
&= 1 - F_\upsilon(A - C).
\end{aligned}
$$

Hence, the first term on the RHS of (113) is given by

$$
\begin{aligned}
\Pr\{\text{says 30\%} \mid \text{"yes" to A}\} &= \Pr\{0.3 = 1 - F_\upsilon(A - C)] \mid \text{"yes" to A}\} \\
\text{(114b)} \qquad &= \Pr\{C = A - F_\upsilon^{-1}(0.7) \mid \tilde{C} \geq A\} \\
&= \Pr\{C = A - F_\upsilon^{-1}(0.7) \mid C + \upsilon \geq A\}
\end{aligned}
$$

where $F_\upsilon^{-1}(\cdot)$ is the inverse of $F_\upsilon(\cdot)$. Rather than estimating the joint probability on the LHS of (113) directly, Li and Mattsson estimate it in two stages. First, they estimate (112b) from the CV responses alone, which produces an estimate of (α/β) and $\sigma_{\tilde{C}}^2$. Taking these as given, they estimate (114b) from the answers to the confidence question, which produces an estimate of σ_υ^2. The estimate of σ^2 is recovered by subtracting the estimate of σ_υ^2 from that of $\sigma_{\tilde{C}}^2$. Li and Mattsson find that this provides more efficient coefficient estimates than the conventional approach based on (6b).

It would be still more efficient to estimate the LHS of (113) in one step directly. When the answer to the confidence question is interpreted as "exactly 30%", as in (114a), one obtains the following direct expression for (113):

$$\text{(114c)} \qquad \Pr\{\text{says "yes" to A and 30\% confident}\} = \Pr\{\eta = A - (\alpha/\beta) - F_\upsilon^{-1}(0.7)\}.$$

---

[55] Qualitative follow-up questions about how confident one is in one’s answer, or how strongly one supports the program, are not uncommon in CV surveys. For example, Ready et al. (1995) offered subjects six responses in a single-bounded survey: "definitely yes", "probably yes", "maybe yes", "maybe no", "probably no", and "definitely no". To use Li and Mattsson’s approach based on respondents’ preference uncertainty, one has to convert the qualitative response into some sort of quantitative probability level or range. Instead, Ready et al. treat this as a case of varying intensity of preference and model it on the lines of an ordered categorical dependent variable model.

A more interesting model is obtained when one changes the interpretation or format of the confidence question. Two other formulations are possible. One is that, when the respondent says she is 30% confident, she means "*at least* 30%" in the sense that

(115a) $$0.3 \leq \Pr\{\tilde{C} \geq A \mid C\} = \Pr\{C + \upsilon \geq A\}.$$

In this case, instead of (114c), the joint probability of the responses to the CV question and the confidence question takes the form

(115b) $$\Pr\{\text{says "yes" to A and 30\% confident}\} = \int_{A-\mu-F_\upsilon^{-1}(0.7)}^{\eta^{max}} [1 - F_\upsilon(A-\mu-\eta)]\, g(\eta)\, d\eta$$

where, for the sake of generality, we write $C = \mu + \eta$, where η is assumed to be independent of υ with an upper support of η^{max} and a density $g(\eta)$.[56] The other formulation is that, when the respondent says she is 30% confident, this is taken to imply a particular *probability range*, e.g., from 25% to 35%, so that

(116a) $$0.25 \leq \Pr\{\tilde{C} \geq A \mid C\} = \Pr\{C + \upsilon \geq A\} \leq 0.35.$$

The joint probability of the responses to the CV question and the confidence question then becomes

(116b) $$\Pr\{\text{says "yes" to A and 25-35\% confident}\} = \int_{A-\mu-F_\upsilon^{-1}(0.75)}^{A-\mu-F_\upsilon^{-1}(0.65)} [1 - F_\upsilon(A-\mu-\eta)]\, g(\eta)\, d\eta \quad .$$

Hanemann, Kriström and Li (1995) develop an alternative specification of preference uncertainty which uses (112a) but assumes that, conditional on C, the uncertainty term υ is *uniformly* distributed over an interval [C-h,C+h] for some parameter $h \geq 0$. This creates a probabilistic zone of indifference with width $\Delta = 2h$. It can be shown that, with this stochastic specification, $\tilde{G}_C(\cdot)$ takes the form

(117) $$\tilde{G}_C(x) = (1/2h)\int_{x-h}^{x+h} G_C(x)\, dx \quad .$$

Thus, for example, when $G_C(\cdot)$ is the standard logistic, (15), one obtains

(118) $$\tilde{G}_C(x) = (2h\beta)^{-1} \ln[(1 + e^{\alpha + \beta(x+h)})/(1 + e^{\alpha + \beta(x-h)})].$$

This cdf can be regarded as a multi-parameter generalization of the logistic that is an alternative to those in (107) - (109). Like them, it can be estimated from the CV responses alone, using (112b), without requiring a question about confidence in one's answer. Li, Lofgren and Hanemann (1995) apply a version of this model to Bishop and Heberlein's (1979) data on WTA responses in real and hypothetical markets for goose hunting permits, based on a multiplicative formulation, $\tilde{C} = Ce^{\upsilon}$ rather than the linear formulation in (112a) and using the Bishop-Heberlein model utility model, (19a), for $G_C(\cdot)$. They find a substantial and interesting difference in the value of h for the two sets of data: h is not significantly different from zero in the real WTA experiment, but it is significantly positive in the hypothetical experiment and its inclusion brings the estimated welfare measure from the hypothetical experiment much closer to the value of the welfare measure estimated from the real data. Given the difference in the

[56] For example, when the probit model (14) is modified to ensure that $C \leq y$, η^{max} is given by (24).

temporal resolution of choice, it is not surprising that responses in hypothetical markets exhibit substantially more preference uncertainty than those in actual markets.

In this section we have reviewed recent efforts to generalize the response probability distribution so as to provide a more flexible and realistic framework for modelling CV responses. When tested against CV data, each of the models has been found to have some validity in the sense of providing a better fit than the conventional logit and probit models in the earlier literature. This creates quite a strong case for considering a multi-parameter generalization of the conventional models. However, it still remains to be seen which is the *more appropriate* generalization. The next step in CV research is to test these generalized models against one another in order to gain a better of understanding of their relative merits.[57]

4.4 Nonparametric and Semiparametric Models

4.4.1 Single-Bounded Data

We first consider nonparametric and semiparametric estimation for single-bounded CV data, and then address double-bounded data in section 4.4.2. Our starting point is the log-likelihood function (48), which we repeat here for convenience

$$\ln L = \sum_{i=1}^{N} y_i \ln P_i + (1-y_i) \ln (1-P_i) \quad , \tag{119}$$

where $P_i \equiv \Pr\{i^{th}$ observation is a "yes"$\}$. Focusing for the moment on the bid, A_i, and ignoring all other covariates, following (1) - (3) and (6) we can write

$$\begin{aligned} P_i \;&\equiv\; H(A_i) \;=\; 1 - G_C(A_i) \\ &=\; 1 - F_\eta[\Delta\bar{v}(A_i)], \end{aligned} \tag{120}$$

where the second line uses the utility difference formulation associated with an additively random utility model, as noted in footnote 12. In a parametric statistical model one adopts a parametric cdf (or survivor function) for $H(\cdot)$ in (120). In a parametric RUM model, one adopts a parametric specification for the WTP distribution, $G_C(\cdot)$, or parametric specifications for the indirect utility function, $v(q,y,\varepsilon)$, and the cdf of ε. A fully nonparametric approach estimates $H(A)$ -- or $G_C(A)$ -- nonparametrically, without specifying a parametric formula. A semiparametric approach estimates either $F_\eta(\cdot)$ or $\Delta\bar{v}(\cdot)$ in the second line of (120) -- but not both -- nonparametrically. We start with the nonparametric approach and then consider various semiparametric approaches.

With binary response data, nonparametric estimation involves estimation of an unknown function of a variable, $H(A)$, whose value is observed at a discrete set of points A_j, j = 1,..,J, (we assume these are arrayed from low to high). The function can be estimated only at those points, not at other points which are unobserved. Economic theory restricts $H(A)$ to be monotonically non-increasing in A (see footnote 8). Ayer et al. (1955) developed a nonparametric maximum likelihood estimator for $H(A)$ by treating P_j =

[57] Also, bid design for generalized parametric response models is an important area for future research.

$H(A_j)$, j = 1,..,J, as a set of unknowns to be estimated, subject to the constraints that $P_j \geq P_{j+1}$—this is the ultimate extension of the idea of adding extra parameters to better model the observed data. Let $\hat{P}_j$ be the observed sample proportion of respondents who said yes to the j^{th} bid. Ayer et al. proved that, if the $\hat{P}_j$'s form a non-increasing sequence, they are the nonparametric maximum likelihood estimator of H(A). Otherwise, if they are increasing in some region (i.e., $\hat{P}_j < \hat{P}_{j+1}$ for some j), one simply combines the responses for adjacent bids until the revised sequence of sample proportions *is* monotonically non-increasing. Ayer et al. proved that this "pool-adjacent-violators" algorithm generates a consistent estimate of the true H(A).[58] This estimator was first applied to single-bounded CV data by Kristrom (1990).

An alternative to the Ayer et al. nonparametric estimator of $H(\cdot)$ is smoothing of the empirical distribution of responses via a kernel function. This was first applied to binary response data in biometrics by Copas (1983), who saw it as being especially useful when there are so few observations at each bid (dose) that one cannot obtain a reliable estimate of the sample proportions of "yes" and "no" responses without resorting to some smoothing procedure. Copas describes a method of correcting for the bias due to smoothing, and he presents an estimate of the variance of the estimated response probability. Kernel estimation of dose response models, together with confidence intervals for estimated response probabilities, is also considered by Kappenman (1987) and Staniswalis and Cooper (1988). We are not aware of any CV applications as yet.

In the last decade or so, several approaches have been developed for modeling binary data in the semiparametric case where the response probability function $H(\cdot)$ can be decomposed, as in (2) or the second line of (120), into a cdf, $F_\eta(\cdot)$, which is of unknown form and what is called an *index function*, $T(\cdot)$ or $\Delta v(\cdot)$, which has a known functional form but with unknown coefficients.[59] This formulation allows for other covariates besides the bid. The index function is written $T(Z,\theta)$, where Z is a set of covariates (including A) and θ is a vector of coefficients to be estimated along with $F_\eta(\cdot)$. Some normalization is imposed on θ and it is usually assumed that $T(Z,\theta) = Z\theta$, which would correspond to a linear Box-Cox utility model (12') where $Z\theta = \alpha - \beta A$, but nonlinear formulations are also possible.

Coslett's (1983) semiparametric estimator extends the Ayer et al. estimator to the case where $H(A) = F_\eta(T(Z,\theta))$ using a two-part maximization of the likelihood function. Given the current estimate of θ, say $\hat{\theta}$, evaluate the index function for each observation. Let $\hat{T}_j$, j = 1,..,J denote the resulting distinct values of the index function, indexed so that they are in increasing order. Estimate $F_\eta(\cdot)$ nonparametrically as a set of constants corresponding to $\hat{P}_j = F_\eta(\hat{T}_j)$, using the pool-adjacent-violators algorithm. Then, concentrate the log-likelihood function and numerically maximize with respect to θ. This is difficult since the concentrated likelihood function varies in discrete steps over θ-space, which precludes conventional maximization techniques based on analytical or numerical derivatives. Instead, following Manski (1975), Coslett takes random sets of orthogonal directions in θ-space, searching for the maximum until a convergence criterion is satisfied. He established that this yields a consistent estimate of θ and $F_\eta(\cdot)$. Li (1996) applied Coslett's estimator to single-bounded CV data and found that, while computer intensive, it was certainly feasible. He compared it to conventional probit, based on (14), using Monte Carlo

---

[58] Applications of the Ayer et al. estimator are discussed in Barlow et al. (1972). Schmoyer (1984) extends their approach to nonparametric estimation where the unknown WTP distribution is assumed to be unimodal, so that H(A) is sigmoidal as well as non-increasing. Glasbey (1987) extends it to cases where the WTP distribution is assumed to be symmetric and/or bell-shaped.

[59] Horowitz (1993) has an excellent survey of this literature.

simulation and found that, although probit gave more efficient coefficient estimates when η actually was normally distributed, it produced seriously biased estimates when η was non-normal while the semiparametric estimates remained close to the true values and had a considerably lower mean squared error.

Because the concentrated likelihood in Coslett's approach is not a smooth function of θ, the asymptotic distribution of his estimator is not known; while it is consistent, there is no reason to believe that it is asymptotically normal or efficient. To circumvent this problem, Klein and Spady (1993) replace Coslett's concentrated likelihood function with a nonparametrically estimated function that locally approximates the true likelihood but is a smooth function of θ. They apply Bayes Rule to obtain

$$(121) \qquad P_j(\theta) = F_\eta(Z_j\theta) = \Pr\{\text{"yes"} \mid T_j = Z_j\theta\} =$$

$$\Pr\{T_j = Z_j\theta \mid \text{"yes"}\} \cdot \Pr\{\text{"yes"}\} \;/\; [\Pr\{T_j = Z_j\theta \mid \text{"yes"}\}\cdot\Pr\{\text{"yes"}\} + \Pr\{T_j = Z_j\theta \mid \text{"no"}\}\cdot\Pr\{\text{"no"}\}].$$

Although $F_\eta(\cdot)$ is unknown, they estimate the RHS of (121) by using the sample proportions of "yes" and "no" responses as estimates of Pr{"yes"} and Pr{"no"}, and nonparametric kernel estimates of the conditional densities $\Pr\{T_j = Z_j\theta \mid \text{"yes"}\}$ and $\Pr\{T_j = Z_j\theta \mid \text{"no"}\}$, which are smooth functions of θ. Because of this, the RHS of (121) can be maximized with respect to θ in a conventional manner. Klein and Spady establish that the resulting estimator of θ is $N^{1/2}$ consistent and asymptotically normal, and attains the asymptotic efficiency bound of Coslett (1987) if η and Z are independent. It has not yet been applied to CV data.

There are several other semiparametric estimators for the case where $F_\eta(\cdot)$ is of unknown form while the index function has a known form, including Han's (1987) maximum rank correlation estimator, the Powell et al. (1989) weighted average derivative estimator, Ichimura's (1993) semiparametric least squares estimator, Manski's (1975) maximum score estimator, and Horowitz's (1992) smoothed maximum score estimator. However, these provide estimates of the coefficients θ but not of $F_\eta(\cdot)$, which makes them less useful in a CV context where the interest lies not just in θ but also in welfare measures that depend on $F_\eta(\cdot)$ or its quantiles.

The other major approach to semiparametric estimation of a binary response model reverses this assumption and postulates a known, parametric cdf for $F_\eta(\cdot)$ combined with an unknown form for the index function which is approximated by some flexible form. For example, one might assume that $F_\eta(\cdot)$ is a logistic or normal cdf, while $\Delta\bar{v}(A)$ is a spline function of the bid and other covariates (O'Sullivan, Yandell and Raynor, 1986), a generalized additive function (Hastie and Tibshirani, 1987; Preisler, 1989), or Gallant's (1981) Fourier flexible form. Estimation is by maximum likelihood or, for grouped data, by some form of least squares; in the logistic case, the analog of minimum chi-squared estimates the RHS of (15') as a spline, generalized additive function, or Fourier function of A. The logistic-Fourier model has been applied to single-bounded CV data by Chen (1993) and Creel and Loomis (1995).

4.4.2 Double-Bounded Data

The log-likelihood function for double bounded data, (98)-(99), can be written equivalently in terms of the J distinct values used for the initial bid A and the follow-up bids A_u and A_d, augmented by two "artificial" bid values of $A_0 \equiv 0$ and $A_{J+1} \equiv \infty$, as

$$(122) \qquad \ln L = \sum_{j=1}^{J+1} N_j \ln \left[G_C(A_j) - G_C(A_{j-1}) \right] \equiv \sum_{j=1}^{J+1} N_j \ln \pi_j \ ,$$

where $G_C(\cdot)$ is the WTP distribution function and N_j is the number of respondents for whom $A_{j-1} < C \leq A_j$. The aim is to estimate $G_C(\cdot)$ nonparametrically. Define $S_j \equiv G_C(A_j)$ = Pr{says "no" to A_j}; then $S_j = \Sigma_{k=1}^{j} \pi_k$. Turnbull (1974, 1976) developed an analog of the Ayer et al. non-parametric estimator for likelihood functions like (122), treating the π_j's as a set of unknowns to be estimated, subject to the constraints that $\pi_j \geq 0$ (which makes the S_j's monotone non-decreasing) and $\Sigma\pi_j = 1$. Turnbull shows that the first order conditions for the likelihood maximization imply the pool-adjacent-violators principle, so that if the observed sample $\hat{S}_j$'s are decreasing in some region one combines adjacent bids until the sequence is non-decreasing.[60] This was first applied to double-bounded CV data by Carson and Steinberg (1990); recent applications include Carson et al. (1992), Carson, Wilks and Imber (1994), and Huhtala (1994). In principle, the Turnbull estimator can be extended to the semiparametric case where $G_C(A) = F_\eta(T(Z,\theta))$, where Z includes covariates besides A and $T(\cdot)$ has a known parametric form while $F_\eta(\cdot)$ is estimated nonparametrically, in the same way that Coslett (1983) and Klein and Spady (1993) extend the Ayer et al. estimator. This would be computationally challenging, however, and no applications are yet known. The alternative semiparametric formulation where $F_\eta(\cdot)$ is a known cdf and $T(\cdot)$ is estimated nonparametrically as some flexible function of covariates might not be as difficult to implement with double-bounded data.

4.4.3 Nonparametric Welfare Measurement

With nonparametric methods that estimate H(A) by smoothing, and semiparametric methods that assume a known $F_\eta(\cdot)$ such as the logistic-Fourier model, the estimated response probability distribution is a continuous function to which (36) and (37) can be applied directly to obtain an estimate of the median WTP, C*, or any other quantile of the WTP distribution. For the mean, C^+, one must resort to some form of numerical analysis to evaluate the area under the empirical response probability graph based on (45) or (46).[61] The situation is more complicated when the Ayer et al., Coslett or Klein and Spady estimators are used with single-bounded data, or the Turnbull estimator with double-bounded data. In those cases, the resulting estimate of the response probability function is a step function consisting of J (or fewer, if there is pooling) distinct estimates, $\hat{P}_1,..,\hat{P}_J$ which are monotonically decreasing and are located at the bid values, $A_1,..,A_J$, where $\hat{P}_j$ is the nonparametric estimate of $H(A_j)$ = Pr{"yes" to A_j}. An example of what the researcher might face is given in Figure 5(a), which graphs estimated $\hat{P}_j$'s against A_j's. Unless one of the $\hat{P}_j$'s just happens to take the value 0.5, the estimate of C* cannot be obtained without some form of interpolation; in Figure 5(a), for example, the point estimate of C* lies somewhere in the interval $[A_4, A_5]$.

---

[60] In (122), the observations on WTP are said to be interval censored for j = 1,..,J, and left (right) censored for j = 0 (j = J+1). The case where observations are right censored was treated by Kaplan and Meier (1958), who showed that what is known as the product limit estimate of the P_j's constitutes the nonparametric maximum likelihood estimator of $G_C(\cdot)$. Turnbull extended this to more complex models, including (122). Statistical packages for survival analysis generally include the Kaplan-Meier estimator and sometimes the Turnbull estimator.

[61] Staniswalis and Cooper (1988) show how to construct a confidence interval for C* when H(A) is estimated by smoothing. Otherwise, confidence intervals for these estimates of C* and C^+ can be obtained by bootstrapping.

Thus, C^* no longer has a computational advantage over the C^+, since both require numerical evaluation.

One way to proceed is via linear interpolation, which produces the trapezoidal approximation to the unknown H(A) shown in Figure 5(b). The resulting estimate of median WTP, $\hat{C}^*$, is marked on the horizontal axis.[62] We noted earlier that the estimate of C^+ is sensitive to the tails of the WTP distribution. In the nonparametric case, it is heavily influenced by what one assumes about the shape of the WTP distribution to the left of A_1 and to the right of A_J. Suppose one assumes that $P_0 \equiv H(A_0) = 1$, thus ruling out negative preferences. Suppose one also assumes that there is some $C^{max} \leq y$ such that $H(C^{max}) = 0$. The graph in Figure 5(b) applies the linear interpolation between A_0 and A_1 and between A_7 and C^{max}. In biometrics, the estimate of C^+ obtained by integrating the area under the linearly interpolated graph in Figure 5(b) is known as the Spearman-Karber estimate of mean WTP, $\hat{\mu}_{SK}$.[63]

An alternative approach to calculating C^+ from the nonparametrically estimated $\hat{P}_j$'s is shown in Figure 5(c). This graph differs from that in Figure 5(b) in three ways, all of which cause it to produce a more conservative estimate of C^+. First, in the main body of the graph it assumes that, over the interval $[A_j, A_{j+1}]$, $H(A) = \hat{P}_{j+1}$, $j = 1,..,J-1$. This is a conservative assumption since, with $H(\cdot)$ monotone nonincreasing, we know that $H(A) \geq H(A_{j+1})$ over the interval. Second, in the left tail, it assumes that $H(A) = \hat{P}_1$ for $0 < A \leq A_1$, which is conservative for the same reason; in effect, this introduces a spike of $\gamma = \hat{P}_1$ into the estimated WTP distribution at C = 0. Third, it censors the WTP distribution from above by imposing a spike at the highest observed bid, A_J. The estimate obtained by integrating the area under the graph in Figure 5(c) is known as the Kaplan-Meier-Turnbull estimate of mean WTP, $\hat{\mu}_{KMT}$.[64]

The formulas for these two estimators are

(123)
$$\hat{\mu}_{KMT} = \sum_{j=1}^{J} (\hat{P}_j - \hat{P}_{j+1}) A_j \equiv \sum_{j=1}^{J} \hat{\pi}_j A_j$$

(124)
$$\hat{\mu}_{KS} = \hat{\mu}_{KMT} + (\frac{1}{2}) \sum_{j=1}^{J} (\hat{P}_j - \hat{P}_{j+1})(A_{j+1} - A_j)$$
$$= \sum_{j=1}^{J} (\hat{P}_j - \hat{P}_{j+1})[(A_j + A_{j+1})/2]$$

where $\hat{P}_{J+1} \equiv 0$ and $A_{J+1} \equiv C^{max}$. The Spearman-Karber estimator $\hat{\mu}_{SK}$ has been shown to be a uniform minimum variance unbiased estimator and a nonparametric maximum likelihood estimator of the discretized mean of the WTP distribution (Miller, 1973; Church and Cobb, 1973). Morgan (1992) gives the formula for the variance of $\hat{\mu}_{SK}$, from which a confidence interval can be constructed. Hamilton et al. (1977) proposed that $\hat{\mu}_{KS}$ could be made more robust by first trimming the data, dropping, say, 5% or 10%

[62] This is how Kristrom (1990) estimated C^* when using the Ayer et al. estimator with CV data.

[63] In the CV literature, this is used by Kristrom (1990) and Duffield and Patterson (1991).

[64] CV applications include Carson et al. (1992), Carson, Wilks and Imber (1994) and Huhtala (1994).

of the observations corresponding to the lowest and highest bids. Morgan reviews the results of a number of simulation studies comparing alternative estimators and concludes that a moderately trimmed Spearman-Karber estimator is often a good general-purpose estimator of the mean of a quantal response distribution. In contrast, the Kaplan-Meier-Turnbull measure $\hat{\mu}_{KMT}$ is an estimator of a *lower bound* on the mean of the WTP distribution, C^+, since it is constructed from a lower bound on the graph of $H(A) = [1-G_C(A)]$. This makes it a conservative estimate if the mean is the chosen welfare measure. Carson et al. (1994) give the formula for the variance of $\hat{\mu}_{KMT}$, from which a confidence interval can be obtained.

Nonparametric or semiparametric estimation of the response probability distribution fundamentally changes the nature of bid design in CV experiments. With parametric response models, the probability distribution is given (assumed by the researcher) and has some small number of parameters which are to be estimated from the data. The bid design takes the distribution into account and makes an initial assumption about the parameter values; in those circumstances, a small number of well-placed design points suffices to provide good estimates of the parameters. Nonparametric or semiparametric response models eliminate the crutch of a predetermined distributional form, resorting instead to a greatly expanded number of parameters to represent the response distribution. Indeed, with the Ayer et al., Coslett, Klein and Spady or Turnbull estimators, the design points literally determine all that can be known about the response distribution, as Figure 5(a) illustrates. Even more than with parametric models, the optimal design depends crucially on what the researcher wants to estimate -- the mean, the median, or some other quantile of the WTP distribution.

If the goal is to estimate median WTP and it is possible to vary the bids between experiments, the classic sequential estimator of C* is the "up and down" estimator (Dixon and Mood, 1948; Dixon, 1965). In the single-bounded version, one starts with an initial bid, A, obtains some CV responses, computes the proportion of "yes" responses, $\hat{P}$, and then revises the bid if $\hat{P} \neq 0.5$; the bid is raised by some predetermined amount if $\hat{P} > 0.5$, and lowered if $\hat{P} < 0.5$, and the sequence is repeated until convergence. Modifications designed to make the estimate of C* more robust or efficient include incorporating an initial delay, corresponding to excluding initial trials (Brownlee et al, 1953), and allowing the step length to decrease as the iterations progress (Cochran and Davis, 1965). There is now an extensive literature on nonparametric sequential estimation of C*, which is summarized in Morgan (1992).

If sequential estimation is impossible and one is restricted to a single bid vector, Muller and Schmitt (1990) recommend identifying an interval in which C* is expected to fall and then using as many bid points as possible evenly dispersed over this interval, with only one or a few observations at each bid. McFadden (1994) gives similar advice for estimating C^+ -- he recommends a relatively large number of closely-spaced bids, spread widely over the interval $[C^{min}, C^{max}]$. There certainly is some logic to both recommendations. The fewer the design points, the cruder is the interpolation required to pin down C*. And, since C^+ depends on the shape of the WTP distribution over its entire length, maximizing the number of bid levels maximizes the number of points at which the WTP distribution is observed. Indeed, this would minimize the bias of estimated mean WTP. But, in our view the argument does not pay enough attention to the *variance* of the estimated response proportions. With a fixed total sample size, when there is an increase in the number of distinct bid levels, J, the number of observations at each bid declines, thereby raising the variance associated with the estimated $\hat{P}_j$. Moreover, the increase in var($\hat{P}_j$) raises the likelihood that some of the observed $\hat{P}_j$'s will violate the monotonicity requirement, thereby causing the pool-adjacent-violators algorithm to shrink the effective number of bid levels and dilute the information associated with the bids that were pooled. These considerations are not adequately addressed in either paper. Just how one should balance the benefits from having many distinct bid points against the benefits from having a reasonable number of observations at each point remains an open question that merits further research.

4.4.4 Concluding Observations on Nonparametric Estimation

There is currently much interest in nonparametric and semi-parametric approaches to the analysis of discrete-response CV data. Compared to parametric models, these are more robust against possible misspecification of the response probability distribution, and they offer the least restricted characterization of what the data have to say. Whether or not this makes much of a difference in practice depends partly on the goal of the analysis. The experience to date suggests that the estimates of median WTP obtained using nonparametric or semiparametric estimation do not differ greatly from those obtained with a parametric approach; however, the estimates of mean WTP can differ considerably. Moreover, the lack of structure implied by the nonparametric or semiparametric approach entails some costs as well as some benefits. Since there is no parametric specification of the underlying RUM model, it is not possible to extrapolate from the observed CV responses to the measurement of other items not directly covered in the survey. For example, one cannot derive an estimate of the WTA distribution from the fitted WTP distribution as indicated in footnote 33. The nonparametric and semiparametric estimators are often computationally demanding and, as noted above, some of them make it difficult to include covariates besides the bid. Moreover, in the absence of a parametric structure it may be difficult or impossible to model response effects using the approaches described in Section 4.3.

4.5 Analysis of Nonresponse Bias

There are two general classes of nonresponse: unit nonresponse where survey recipients may simply choose not to respond to the survey at all, and item nonresponse where they may choose to not answer certain questions, either out of protest, or because the questions are too difficult, time-consuming, or confusing. Bias occurs when nonresponses are not random over the sample population, but instead are correlated with WTP. Although a survey researcher cannot fully account for this bias without collecting responses from the nonrespondents, there are techniques available for addressing the component of nonresponse bias that can be linked to observable sociodemographic or attitudinal characteristics.

4.5.1 Unit Nonresponse

Let the total number of survey recipients be N. For the i^{th} recipient let R_i be an indicator which equals one if the recipient responds to the survey and zero if the recipient does not respond. As before, y_i is the discrete response to the WTP question, which is observed only if the individual responds, and P_i is the probability of a positive response. Thus, we can write the log-likelihood function as follows:

$$\ln\,L \;=\; \sum_{i=1}^{n} \; R_i \;[y_i \ln P_i \;+\; (1 - y_i)\ln(1 - P_i)] \qquad (125)$$

which is simply the sum of the individual log-likelihood functions for units that responded to the survey. The consistency of the maximum likelihood estimators is dependent on the response process which determines $R = (R_1,...R_i,...R_n)$. If R is randomly and independently distributed across survey recipients, and uncorrelated with factors related to WTP, then unit nonresponse merely enters as an additional level of sampling in the data collection process. That is, instead of the sample size being n, the number of survey recipients, it is m, the number of survey respondents. The response rate then, affects small sample properties but does not affect the maximum likelihood consistency and efficiency results. If however, the probability of responding to the survey varies systematically over the sample population, by demographic characteristics, attitudes, or directly by WTP, then maximization of (119) instead of (125) will result in inconsistent estimation, due to the heavier influence of certain groups with high response rates, relative to those with low response rates.

The best solution is to minimize nonresponse in the first place, by designing a data collection process which encourages high participation rates for all groups. This aside, the researcher must rely on what she knows about the characteristics of the population and the sample, and on what she knows or can assume about the response process, in order to infer something about the characteristics of the non-respondents. A key issue is whether, conditional on the observed covariates, members of the sample are missing from the set of respondents at random or not at random (Little and Rubin, 1987). If they are missing at random, there is an unconfounded probability process with $\Pr(R \mid Y) = \Pr(R)$, and $\Pr(R_i = 1 \mid Y) > 0$ for all i and all possible Y, where $Y = (y_1,...y_i,...y_n)$. Furthermore, assume that the response probabilities are all equal, positive, and independent, so that the response process is an independent Bernoulli sampling process with probability of response $\Phi > 0$. Given the number of survey respondents, m, and the total number of survey recipients, n, we have $\Pr(R_i = 1) = m/n$, or $E(R_i) = m/n$ for all i. Then, taking the expectation of (125), we have:

(126) $$E(\ln L) = (\frac{m}{n})\ E \sum_{i=1}^{n} [y_i \ln P_i + (1-y_i) \ln(1-P_i)]$$

which implies that maximum likelihood estimation is, on average, unaffected by nonresponses drawn from an unconfounded probability responding mechanism. The unit nonresponses represent an additional level of sampling which reduces the total sample size, but does not distort the maximum likelihood estimation in any systematic manner.

It is possible, however, that the probability of responding is correlated with demographics, and possibly other unobservable factors, that are themselves correlated with WTP responses. In that case, taking the expectation of (125) does not eliminate the variation in R across individuals. Thus, maximum likelihood estimation will be inconsistent unless the weights imposed by R are appropriately adjusted with a weighing scheme. We might therefore, assume that the response probabilities are functions of the demographics only, with demographic groups stratified by characteristics such as age, education and income. Within each demographic group, the responses are assumed to follow an unconfounded response probability process; between groups, the response process is *not* unconfounded. Under these assumptions, we can write the likelihood function as:

(127) $$\ln L = \sum_{h=1}^{H} \ln L_h$$

where h = 1,..,H is the index of the demographic group, n_h is the sample size of group h,

(128) $$\ln L_h = \sum_{i=1}^{n_h} R_{hi}\ [y_i \ln P_i + (1-y_i) \ln(1-P_i)],$$

and, within group h, we have $E(R_{hi}) = m_h/n_h$.

If we specify weight adjustments w_h so that the different groups carry the same weight as they do in the sample population, then we overcome the problem. In particular, we let

$$(129) \qquad w_h = \frac{n_h/n}{m_h/m}$$

where m_h/m is the response rate for group h in the sample, and n_h/n is the relative weight of group h in the sample population. The weighted likelihood function is:

$$(130) \qquad \ln L^W = \sum_{h=1}^{H} w_h \ln L_h$$

$$= \sum_{h=1}^{H} w_h \sum_{i=1}^{n_h} R_{hi} \, [y_i \ln P_i + (1-y_i) \ln (1-P_i)].$$

This weighted likelihood function has the same expected value as (126). Thus, the stratified weighing scheme effectively removes the unit nonresponse distortions in the sample, provided that the stratification fully accounts for the distorted response rates.

4.5.2 Item Nonresponse

If there are systematic differences between those who do and do not respond to the WTP question, there will be bias in the estimation of mean or median WTP. The only way to prevent this bias is to collect a complete sample at the start. Short of this, the most common practical remedy for item nonresponse is imputation. This is where values are imputed for the missing responses using existing information about the respondents and some prediction method based on the responses provided by the other survey respondents.

By substituting predicted values for the missing values, one creates a complete data set, but the data set lacks the variability found in a truly complete data set. The imputed values are typically predicted values based on some model derived from the existing data. But these models will not account for any differences between respondents and non-respondents. The researcher must therefore make assumptions, based on prior information, about the non-respondents' behavior; either they are assumed to be no different, or they are assumed to be different in some specific, quantifiable way. Either way, it is clear that the assumptions made by the researcher and the resulting predictive model used for imputation will not adequately substitute for the real data. Several alternative procedures have been used that provide a greater amount of variability in imputed responses. One alternative is to use a hot deck procedure, of which there are several varieties[65] where it is assumed that non-respondents and respondents within a classification group follow the same distribution, and the missing values are imputed by drawing randomly from the existing data within the corresponding classification groups and substituting the drawn responses for the missing values.

Another method which provides more variability of responses is the multiple imputation method (Rubin, 1987). Here, multiple data sets are created using different imputation algorithms, and estimation is performed separately, using each data set. The collection of resulting estimates should be stable in terms of their expected values, but should demonstrate adequate variability for inference purposes.

[65] See Madow, Olkin and Rubin (1983, Chapter 14) for a review of hot deck procedures.

As indicated, the imputation procedure chosen is based on the instincts of the survey researcher about the behavior of the non-respondents. The validity of the imputations is not verifiable unless actual responses are eventually obtained from the nonrespondents. The researcher must therefore be careful about making assumptions and using point estimates derived using the imputed values based on these assumptions. Alternatively, the researcher must be careful about making inferences using estimates derived without accounting for the missing responses, because these estimates might be systematically biased.

It is impossible to be certain about survey nonresponse biases. In order to adjust for biases in survey responses, the researcher must make assumptions about what types of biases exist and believe these biases can be addressed using other information, either theoretical or empirical. If assumptions are correct, then the resulting adjusted estimation is an improvement over unadjusted estimation. Unfortunately, in the case of nonresponse, the researcher is unable to verify assumptions. We recommend examining the sensitivity of WTP estimates to the assumptions made about nonrespondents. The greater the variation among estimates, the more cautious one must be about drawing inferences based upon any imputation scheme, including no imputation.

5. CONCLUSIONS

In this survey of the statistical issues in discrete-response CV, we have sought to emphasize three broad themes: the need for thoughtfulness in modelling the stochastic components of the response model; the importance of modelling the entire data generation mechanism, including the elicitation procedure as well as individual preferences; and the importance of experimental design.

There is some tendency in conventional demand analysis to treat the stochastic component as an afterthought when specifying an economic model, and to focus almost exclusively on the deterministic component. The stochastic component is assumed to arise outside the economic model and to have no real significance for the economic interpretation of the statistical results. In RUM models, including those used for discrete-response CV, the situation is different. The stochastic component is an essential part of the economic model; it interacts with the deterministic component and materially affects the model's implications for the prediction of behavior and the evaluation of welfare. Therefore, both components need to be taken seriously by the researcher. This is illustrated by our analysis in Section 2.3 of issues that arise in restricting the distribution of the stochastic components in order to ensure that the indirect utility function is monotone increasing in q. Another example is welfare measurement; with discrete response CV data one ends up estimating the entire WTP distribution and, as noted in Section 2.4, the welfare measures -- especially the mean -- can be highly sensitive to the specification of the response probability model. Until recently, the CV literature generally used two-parameter probability distributions, with a single location parameter (modelled as a function of covariates) and a single scale parameter. There is growing awareness that this is too restrictive; one often needs multi-parameter probability models that permit more flexibility in the shape of the WTP distribution. There are many possible ways to accomplish this, including inserting a spike at one or both tails of the WTP distribution, adding an overdispersion parameter, making the distribution heteroscedastic, or using models of thick indifference curves or preference uncertainty. One can also use various multi-parameter generalizations of the logistic or related distributions, or switch to a nonparametric approach. The next step in CV research will be comparative analysis, testing the alternative approaches against one another with a view to identifying which might work best in different circumstances.

Another recent development is the growing interest in alternative ways to elicit preferences, including discrete-response CV formats in which individuals can respond with varying degrees of confidence or certainty, and other procedures such as contingent choice, contingent ranking, scaling or other forms of conjoint analysis, and multiattribute utility assessment. These are seen as either alternatives to referendum CV or supplements that can be combined with it in a single interview. We certainly encourage methodological pluralism for the purpose of discovering which approaches work well with different items and different types of people. We have three general observations to offer.

First, it is important to decide whether the goal of the exercise is value a particular item or to estimate an entire valuation function, since that greatly influences the choice of elicitation procedure. There is a tradeoff since, although estimating an entire valuation function may more useful, it is also more costly. To value multiple items with the same precision as a single item, either one must increase the number of observation collected, or one must assume that preferences for the items are related to one another in a specific way through a single, underlying hedonic utility model. The hedonic approach makes it possible to value multiple items without more observations, but at the cost of imposing a maintained assumption about preferences which could be inaccurate or unduly restrictive.

Second, when using alternative elicitation formats one should not be surprised that different formats yield different estimates of preferences. In fact one should expect this, because the cognitive demands are not identical. People find choice different than ranking, ranking the top items different than ranking all items, rating different than ranking, matching different than choice, and open-ended CV different than closed-ended CV. Some critics of CV have asserted that one should expect the same results regardless of the elicitation procedure.[66] In our view, this is like asserting that words should have the same meaning regardless of context; it misunderstands the nature of both the organism and the process. At any rate, the evidence in market research as well as environmental valuation overwhelmingly rejects procedural or task invariance.[67] In consequence, the researcher should contemplate modelling the response process itself in combination with the underlying preferences -- in effect, the complete data generating process. This can be done through heteroscedastic models in which the variance of the individual's utility for an item is non-constant, reflecting features of the individual, the item, or the elicitation task; examples are the heteroscedastic models used in choice experiments, (82b), rankings, (86), and double-bounded CV, (106). Other ways to incorporate response effects or cognitive burdens in surveys include the models with background probability of saying "yes," (103) and (110), the model of thick indifference curves, (111), and the preference uncertainty model, (117).

Third, while we favor experimentation, we believe that researchers should resist the temptation of going too far in maximizing the amount of information collected per subject. One reason is cognitive limits. There is no free lunch in collecting information from subjects; wearying them or asking for more precision than they can supply becomes counterproductive. This is why we prefer closed- to open-ended

---

[66] "If people really have this willingness-to-pay, they should shine through. They should shine through these little variations" (Hausman 1993, p. 214).

[67] In addition to cognitive impacts, changing the wording of a question or the survey context may substantively affect the commodity that is being valued. Specifying the item less precisely may signify a wider class of environmental commodities; leaving the timing of commitment more flexible makes the item more economically valuable.

CV questions.[68] A second reason is the danger of an inadequately sized sample. Collecting more information from fewer subjects may reduce survey costs, but it is a fragile strategy because, while there is likely to be considerable heterogeneity in preferences between individuals, successive responses from the same subject are unlikely to capture the inter-individual variability. Indeed, successive responses are likely to be correlated because of fatigue or learning, because of anchoring phenomena, or because framing or other factors create individual-specific effects. Consequently, for the purpose of modelling preferences in a population, 100 subjects each valuing twenty items may provide much less information than 1,000 subjects each valuing two items.

In addition to experimentation with alternative elicitation procedures, there is much interest in comparing preferences elicited in surveys with those exhibited in actual (non-experimental) behavior. In transportation, where much of the research originated, the term *stated preference* is used to describe the elicitation of preferences through surveys or experiments, in contrast to *revealed preference* for the observation of actual behavior from which preferences are inferred.[69] It is useful to make a distinction between comparisons involving the *same* task performed in different settings, versus comparisons involving *different* tasks that are derived from the same underlying utility function. Examples of the former are comparing referendum CV responses with voting in actual referendum elections, comparing stated preference choices with actual market choices, or comparing self-predictions of demand behavior with realized demand behavior.[70] Examples of the latter are comparing CV with realized demand behavior, or comparing stated preference rankings with actual market choices.

The practice of comparing the results from CV with those from revealed preference goes back to the very first CV study, an open-ended survey of visitors to the Maine woods (Davis 1963); Knetsch and Davis (1966) compared the estimate of WTP with that from a travel cost study and found almost no difference. The first closed-ended CV study, by Bishop and Heberlein (1979), also contained a comparison with a travel cost study, and obtained similar results.[71] However, those were comparisons of reduced-form estimates of WTP; instead, Hanemann (1984a) argued for a structural approach in which one tests for

---

[68] Also, the open-ended format is significantly less incentive-compatible than the closed-ended format (Hoehn and Randall, 1987; Arrow et al., 1993), and subjects tend to respond to open-ended CV questions by estimating what the item might cost rather than their WTP (Schkade and Payne, 1994).

[69] The use of stated preference in transportation was the subject of special issues of the *Journal of Transport Economics and Policy* Volume XXII, No. 1 (January 1988) and *Transportation* Volume 21, No. 2 (May 1994); its use in market research was the subject of a special issue of the *Journal of Business Research*, Volume 24, No. 2 (March 1992). Morikawa (1989) is a seminal work in transportation; Hensher (1994) provides a summary of the transportation literature.

[70] Constructing a utility model for analyzing consumers' assessments of their response to possible price changes -- sometimes called *contingent behavior* -- and comparing or combining these with data on their realized demand behavior is discussed in Dickie, Fisher and Gerking (1987), Hanemann and Chapman (1988), Carson, Hanemann and Steinberg (1990), Cameron (1992), Chapman et al. (1993), Englin and Cameron (1996), and Loomis (1996).

[71] This is a fairly general finding in the literature. Carson et al. (1996) survey over 80 studies containing several hundred comparisons between estimates of WTP based on CV and revealed preference; the results are often fairly close and, overall, the CV estimates are slightly lower.

consistency across tasks by reference to a given, underlying utility function that is used to model the responses to each task. In a comparison of CV with travel cost, for example, the CV response is generated by an indirect utility function, as in (4) or (6a), while the recreation demand function is generated by applying Roy's identity to the indirect utility function:

$$x_i \;=\; h^i(\boldsymbol{p},\boldsymbol{q},y,s,\epsilon) \;=\; -\frac{\partial v(\boldsymbol{p},\boldsymbol{q},y,s,\epsilon)\,/\,\partial \boldsymbol{p_i}}{\partial v(\boldsymbol{p},\boldsymbol{q},y,s,\epsilon)\,/\,\partial y}\;. \tag{131}$$

Consequently, there is a mapping between the ordinary demand functions which underpin revealed preference analysis and the WTP function which underpins stated preference analysis; specifying a particular model for one implies a particular model for the other.[72] This structural approach has now become the standard for comparing or combining preferences elicited through alternative procedures.[73] There is one key limitation: a utility model of the form

$$u \;=\; T[v(p,q,y,s,\varepsilon),q] \tag{132}$$

yields the *same* ordinary demand functions for market goods, $h^i(p,q,y,s,\varepsilon)$, as the utility function $v(p,q,y,s,\varepsilon)$ because the ratio of price and income derivatives is the same, while producing *different* WTP functions, $C(p,q,y,s,\varepsilon)$. The revealed preference approach based on integrating an observed system of ordinary demand functions to obtain the underlying utility function is *incapable* of recovering transformations such as $T[\cdot,q]$ in (132). Therefore, it omits the portion of WTP which is associated with the second argument of the $T[\cdot,q]$.[74] This must be taken into consideration when comparing results from stated and revealed preference approaches.

However, while adopting a model-based approach, much of the current CV literature comparing stated and revealed preference still pays limited attention to the stochastic specification of the RUM model. Researchers typically assume that the stochastic structure is identical across elicitation tasks and across settings (e.g., hypothetical versus actual markets), and focus exclusively on consistency with respect to the deterministic components of the model. Transportation researchers, by contrast, have allowed for differences in variances when comparing or combining RUM preferences across tasks or settings, along the lines of the heteroscedastic logit model, (82b). In discrete choice experiments, they find that the variances typically *do* differ. If the heteroscedasticity is taken into account, the deterministic components of the RUM model often turn out to be consistent across tasks or settings, but not if the heteroscedasticity

---

[72] McConnell (1990) showed how this works, identifying the observable implications of an income-compensation function for ordinary demand functions; see also Whitehead (1995).

[73] Examples include Jakus and Smith (1992), Eom and Smith (1994), Jakus (1994), Chen (1994), and Dickie and Gerking (1996).

[74] McConnell (1983) and Hanemann (1988) equate the unrecovered component with what Krutilla (1967) had called *existence* value, also known as *nonuse* or *passive use* value. The failure to recover $T[\cdot,q]$ from observed ordinary demand functions arises from optimization at the margin by consumers, i.e., what Hanemann (1984b) calls their *continuous* choices. If they make discrete or mixed discrete/continuous choices, $T[\cdot,q]$ is in principle recoverable from revealed preference data.

is ignored.[75] The need to allow for heteroscedasticity could arise because of the cognitive differences mentioned above, differences in task performance (e.g., due to attentiveness or strategic behavior), differences in perception of the commodity (e.g., there may be a different degree of heterogeneity in perceptions of attributes for hypothetical versus actual choices, depending on how this is controlled in the stated preference exercise), or statistical factors such as differences in the degree of variation among the attributes in stated versus revealed preference. Several studies have found that the revealed preference data sometimes contain more noise and, without rescaling of variances, produce less reliable estimates of model parameters than the stated preference data (Morikawa, 1989; Adamowicz et al., 1994).

While heteroscedastic models have clearly proved their worth, we believe that it may be useful to consider a wider range of models. Examples include models such as (103) or (110) where the background probability of saying "yes" varies with the task or setting; the model in (111) where the width-of-indifference parameter varies with the context; or a model such as (117) where the extent of preference uncertainty varies with the context. For the purpose of comparing or combining stated and revealed preference data, it may pay to use specifications covering all facets of the data generation process that could vary with the task or the setting.

Since stated preference involves data collection in an experimental rather observational mode, experimental design is an important component of the research process. Experience has shown that careful attention to the statistical design of the CV study, no less than the questionnaire design and sampling plan, can have a large payoff in terms the quality and reliability of the data collected. But, experience also shows that achieving an effective design requires a clear sense of the goals of the research and the preference structure being modelled. Optimal bid designs are generally sensitive to what the researcher wishes to measure -- e.g., mean versus median WTP -- and to what he knows (or thinks he knows) about the structure of the response probability model.

Current practice in CV research draws on two distinct bodies of design literature. One is the literature on fractional factorial designs (Winer, 1971). This deals with the problem of estimating the effects of several factors on a dependent variable where there may be several possible levels of each factor and the factors can, in principle, vary independently. This creates a large number of potential combinations among the factors. The aim in fractional sampling is to reduce the number of sampling designs by varying several factors simultaneously, instead of experimenting with them one at a time. The typical situation where this arises is the estimation of the models discussed in Section 4.1 for valuing multiple items. The problem arises with linear regression as well as with nonlinear response probability models, and is essentially one of dimensionality reduction. Look-up tables can be consulted to find a parsimonious design with certain given properties regarding the interaction effects the researcher wishes to estimate (Green, 1974; Louviere, 1988). The other literature concerns optimal dose design in bioassay, which underpins our discussion of optimal bid design in Section 3.5. The context there is estimating the effect of a continuous exogenous variable (the bid) on the discrete CV response, where the response probability is a *nonlinear* function of the bid variable. In consequence, the asymptotic variance-covariance matrix for the parameters of the response probability distribution is itself a function of those parameters. The challenge is to choose design levels for the bid that will minimize the variance (or some related statistic) of the parameter estimates, or the welfare measure constructed from them.

---

[75] See the chapter by Adamowicz et al., Swait and Adamowicz (1996), or Louviere (1996), who contends that allowing variances to change has brought about a "rescaling revolution."

In fact, when one takes a hedonic approach to valuing multiple items, *both* problems tend to arise. On the one hand, CV responses depend on the price associated with each item, which in principle is a continuous variable that must be specified for the purpose of the experiment. On the other hand, preferences for the items depend on their characteristics, which are apt to be sufficiently multi-dimensional as to require some dimensionality reduction. We see combining the two traditions of experimental design in order to facilitate improved estimation of hedonic valuation models as an important area for future CV research.

One of our main themes has been the need to model the entire response process in CV, including the elicitation procedure as well as individual preferences. So far, most of the experimental design research in CV has not adopted this perspective, focusing on the modelling of preferences rather than elicitation procedures, and working off standard, simple response models such as the linear logit model (15), rather than the more complex models discussed in much of this chapter. We see bid design for models of the entire response process as another important area for future CV research.

We end with a word of caution. Good experimental design is rooted not just in statistics but also in survey research. This is because there may always be some limit to what information a researcher is able to convey to subjects, and what she can get them to believe about an item. Those limits constrain what experiments she can design. One usually discovers the limits the hard way, by testing the instrument using various cognitive survey techniques. In this chapter we have focused on what economic theory and statistics have to say about designing discrete-response CV experiments. The third leg on which CV stands is survey research. What makes CV such an interesting subject to study is the practical interplay between economics, statistics, and survey research.Appendix: Maximum Likelihood Algorithms

In general, optimization algorithms are iterative procedures that begin with a starting value for θ and update the estimate at each iteration using some rule of the form:

(A.1) $$\theta_{t+1} = \theta_t + \lambda_t \Delta_t$$

where t represents the iteration number, Δ represents the direction vector and λ represents step size. The various algorithms use different specifications for Δ and λ, and vary in their complexity and convergence properties. When the likelihood function is globally concave, as with the models discussed in this chapter, the Newton-Raphson Method is a popular, effective choice. It is based on the linear Taylor series expansion of the first-order condition around θ_0:

(A.2) $$\frac{\partial L(\theta)}{\partial \theta} = L_\theta = L_\theta^0 + H^0 (\theta - \theta_0) = 0$$

where the superscript 0 indicates that the function is evaluated at $\theta = \theta_0$, L_θ is the first derivative of the log-likelihood function with respect to the parameter vector, θ (also known as the score vector or gradient vector) and H is the Hessian matrix of second derivatives of the log-likelihood function:

(A.3) $$H = \frac{\partial^2 \ln L(\theta)}{\partial \theta \, \partial \theta'}$$

Solving for θ as a function of θ_0 gives:

(A.4) $$\theta = \theta_0 + (H^0)^{-1} L_\theta^0$$

Letting $\theta_t = \theta_0$ and $\theta_{t+1} = \theta$ gives the Newton-Raphson procedure:

(A.5) $$\theta_{t+1} = \theta_t + (H^t)^{-1} L_\theta^t$$

A disadvantage of this procedure is that it uses the second derivative matrix, H, which can be difficult to derive analytically and can easily be the source of errors.[76] It also does not always perform well when the current estimate is very far from the maximum. For the cases of logit and probit however, the Hessian matrix, H, is positive definite for all values of θ, so the algorithm should always converge.

Other methods are based on equation (A.5) but use different expressions for H. The method of scoring, for example, uses the negative of the Fisher information matrix for H:

(A.6) $$\theta_{t+1} = \theta_t + (-I^t)^{-1} L_\theta^t$$

The information matrix is the negative of the expected value of the Hessian matrix (- E[H^{-1}]) and must be derived analytically. In the case of the logit model, the Hessian is independent of the response vector so that E(H) = H, and the method of scoring is equivalent to the Newton-Raphson method.

The Berndt, Hall, Hall and Hausman (BHHH) method, uses the outer product of the gradient vector for H:

(A.7) $$\theta_{t+1} = \theta_t + (L_\theta^t \, L_\theta^{t\prime})^{-1} L_\theta^t$$

This choice has the advantage of being always positive semi-definite (for any functional form) and less burdensome to derive analytically because it does not involve the second derivative matrix. Furthermore, Berndt et al (1974) have shown that the outer product of the score vector is a consistent estimator for the information matrix.

Another popular algorithm is the Davidon-Fletcher-Powell (DFP) Method which is in the class of quasi-Newton methods. This method starts with an approximation of H (sometimes the identity matrix, sometimes the result of the BHHH method), and updates the estimate of H at each iteration using the previous estimate of H and an estimate of the change in the first derivative vector as a numerical substitute for the second derivative. In a sense, at each iteration, "information" is added to the estimate of H. The DFP method has excellent convergence properties but because of its iterative approach to estimating H, it cannot approximate the Hessian well when only a few iterations are made. Although this method generally takes more iterations to converge than the Newton-Raphson method, if numerical derivatives are being used, the DFP method can converge more quickly.

---

[76] Numerical derivatives are obtained by taking the difference between a function at two arbitrarily close points. They require twice as many function evaluations as analytical derivatives and can be computationally time-consuming.

Table 1: Analytical Bias and Asymptotic Variance of WTP - Single-Bounded Logit Model

(α = 2.5, β = -0.01, median WTP = $250, 250 observations)

Bid Design (percentile)	% Bias in α	% Bias in β	% Bias in Median WTP	Asymptotic Variance of Median WTP
$200, $300 (38%, 62%)	0.83%	-0.83%	0.00%	170.21
$100, $400 (18%, 82%)	1.14%	1.14%	0.00%	268.19
$5, $500 (8% , 92%)	1.90%	-1.90%	-0.03%	588.60
$5, $200, $300, $500 (8%, 38%, 62%, 92%)	1.94%	-1.95%	0.00%	258.85
$300, $400, $500 (62%,82%, 92%)	4.15%	-3.62%	0.51%	1779.19

Table 2: Some Optimal Designs for Single Bounded CV Data[a]

		Logit Models	Probit Models
Design Criterion	Criterion description	Single-bounded design (in percentiles)	Single-bounded design (in percentiles)
D-optimal	Maximize Determinant of Fisher Information matrix	Half of sample each at: B = 82.4 % B = 17.6 %	Half of sample each at: B = 87.25 % B = 12.75 %
C-optimal	Minimize variance of median WTP	B = 50.0 % (parameters unidentifiable)	B = 50.0 % (parameters unidentifiable)

[a] Results from Alberini (1995b), Minkin (1987), Wu (1988).

Table 3: Performance of Simulated Sequential Single Bounded C-optimal Logit CV Design[bc]

Number of survey iterations	Number of observations per iteration	Mean Squared Error of WTP relative to C-optimal design[d]
1	500	101.42
2	250	2.46
3	166	1.34
4	125	1.32
5	100	1.23

[b] These simulations are based on a logistic distribution with parameters α=2.5, β=-0.01; expected median WTP = 250, n = 500.

[c] Taken from Kanninen (1993a).

[d] Based on the specified parameters and sample size, a C-optimal design would produce a mean squared error estimate of $4/n\beta^2$ for the single-bounded model (Kanninen, 1993b).

Table 4: Analytical Bias and Asymptotic Variance of WTP - Double-Bounded Logit Model
α = 2.5, β = -0.01, median WTP = $250, number of observations =250

Bid Design	% Bias α	% Bias β	% Bias Median WTP	Asymptotic Variance of Median WTP
B = $250 (50%) Follow-up = B ± $100	0.53%	-0.53%	0.00%	128.15
B = $250 (50%) Follow-up = B ± $200	0.56%	-0.56%	0.00%	135.41
B = $250 (50%) Follow-up = B + $250 or B - $245	0.66%	-0.66%	0.00%	141.45
B = $300, $400, $500 Follow-up = B ± $100	0.90%	-0.81%	0.09%	217.11

Table 5: Some Optimal Designs for Double Bounded CV Data[e]

		Logit Models	Probit Models
Design Criterion	Criterion descriptio n	Double-bounded bid design (in percentile s)	Double-bounded bid design (in percentile s)
D-optimal	Maximize Determinan t of Fisher Informatio n matrix	B = 50.0 % B^{up} =87.9 % B^{dn} =12.1 %	B =50.0 % B^{up} = 91.7 % B^{dn} = 8.3 %
C-optimal	Minimize variance of median WTP	B = 50.0 % B^{up} = 75.0 % B^{dn} = 25.0 %	B = 50.0 % B^{up} = 83.7 % B^{dn} = 16.3 %

[e] Results from Alberini (1995b), Kanninen (1993b), Minkin (1987), Wu (1988).

REFERENCES

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

Adamowicz, W., P. C. Boxall, J. J. Louviere, J. Swait, and M. Williams (1996) "Stated Preference Methods for Valuing Environmental Amenities," chapter __ in this volume.

Adamowicz, W., J. Louviere and M. Williams (1994) "Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities," *Journal of Environmental Economics and Management*, 26, 271-292.

Aitchison J. and S.D. Silvey (1957) "The Generalization of Probit Analysis to the Case of Multiple Responses," *Biometrika*, 44, 131-140.

Alberini, A. (1995a) "Testing Willingness to Pay Models of Discrete Choice Contingent Valuation Data." *Land Economics*, 71:1, 83-95.

Alberini, A. (1995b) "Optimal Designs for Discrete Choice Contingent Valuation Surveys: Single-bound, Double-bound and Bivariate Models," *Journal of Environmental Economics and Management*, 28:3, May, 287-306.

Alberini, A. (1995c) "Efficiency vs Bias of Willingness-to-Pay Estimates: Bivariate and Interval-Data Models," *Journal of Environmental Economics and Management*, 29, 169-180.

Alberini, A. and R.T. Carson (1993) "Efficient Threshold Values for Binary Discrete Choice Contingent Valuation Surveys and Economic Experiments," Resources for the Future Discussion Paper, Quality of the Environment Division, Washington, D.C.

Alberini, A., B. Kanninen and R. Carson (1994) "Random-Effect Models of Willingness to Pay Using Discrete Response CV Survey Data," Resources for the Future Discussion Paper 94-34, Washington, D.C.

Alberini, A., B. Kanninen and R. Carson (1994) "Estimating Willingness to Pay with Discrete Choice Contingent Valuation Data Using Error Component Models," presented at the Allied Social Sciences Associations Annual Meeting, Boston MA, January.

Altaf, Mir Anjum, and J.R. DeShazo (1994) "Bid Elicitation in the Contingent Valuation Method: The Double Referendum Format and Induced Strategic Behavior," working paper, Harvard University, Arts and Sciences, Urban Planning Committee, March.

Amemiya, T. (1980) "The n^{-2}-Order Mean Squared Errors of the Maximum Likelihood and the Minimum Logit Chi-Square Estimates," *Annals of Statistics*, 8:3, 488-505.

Amemiya, Takeshi (1985) *Advanced Econometrics,* Harvard University Press, Cambridge.

Anderson, E.B. (1970) "Asymptotic Properties of Conditional Maximum Likelihood Estimators," *Journal of the Royal Statistical Society, Series B* 32, 283-301.

Anderson J.A. and Richardson (1979) "Logistic Discrimination and Bias Correction in Maximum Likelihood Estimation," *Technometrics*, 21, 71-78.

Aranda-Ordaz, F.J. (1981) "On Two Families of Transformations to Additivity for Binary Response Data," *Biometrika*, 68, 357-363.

Arnold, Barry C. (1992) "Multivariate Logistic Distributions" in *Handbook of the Logistic Distribution*, edited by N. Balakrishnan. New York: Marcel Dekker, Inc.

Arrow, K., R. Solow, P.R. Portney, E.E. Leamer, R. Radner and H Schuman (1993) "Report of the NOAA Panel on Contingent Valuation." *Federal Register*, 58:10, 4601-4614.

Ashford, J.R. (1959) "An Approach to the Analysis of Data for Semiquantal Responses in Biological Response," *Biometrics*, 15, 573-581.

Avery, R.B., L.P. Hansen, and V.J. Hotz (1983) "Multiperiod Probit Models and Orthogonality Condition Estimation," *International Economic Review*, 24, 21-35.

Ayer, M., B.D. Brunk, G.M Ewing, E. Silverman (1955) "An Empirical Distribution Function for Sampling with Incomplete Information," *Annals of Mathematical Statistics,* 26, 641-647.

Barbera, S., and P.K. Pattanaik (1986) "Falmagne and the Rationalizability of Stochastic Choices in Terms of Random Orderings," *Econometrica*, 54:3, 707-715.

Barlow, R.E., D.J. Bartholomew, J.M. Bremner, and H.D. Brunk (1972) *Statistical Inference Under Order Restrictions*, New York: John Wiley.

Bateman, Ian J., Ian H. Langford, and Jon Rasbash (1996) "Elicitation Effects in Contingent Valuation Studies," chapter 15, this volume.

Beggs, S., S. Cardell, and J. Hausman (1981) "Assessing the Potential Demand for Electric Cars," *Journal of Econometrics*, 16, 1-19.

Ben-Akiva, M. and S.R. Lerman (1985) *Discrete Choice Analysis: Theory and Applications to Travel Demand*, MIT Press, Cambridge.

Ben-Akiva, M. T. Morikawa, and F. Shiroishi (1992) "Analysis of the Reliability of Preference Ranking Data," *Journal Business Research*, 24, 149-164.

Berkson, J. (1953) "A Statistically Precise and Relatively Simple Method of Estimating the Bio-Assay with Quantal Response, Based on the Logistic Function." *Journal of the American Statistical Association*, 48, 565-599.

Berndt, E.R., B.H. Hall, R.E. Hall and J.A. Hausman (1974) "Estimation and Inference in Nonlinear Structural Models," *Annals of Economic and Social Measurement,* 3, 653-666.

Bishop, Richard C. and Thomas A. Heberlein (1979) "Measuring Values of Extra-Market Goods: Are Indirect Measures Biased?" *American Journal of Agricultural Economics* 61:5, 926-930.

Borjas, George J., and Glenn T. Sueyoshi (1994) "A Two-Stage Estimator for Probit Models with Structural Group Effects," *Journal of Econometrics*, 64, 165-182.

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

Brownlee, K.A., J.L. Hodges, and M. Rosenblatt (1953) "The Up-and-Down Method with Small Samples," *Journal of the American Statistical Association,* 48, 262-277.

Burbidge, John B., Lonnie Magee, and A. Leslie Robb (1988) "Alternative Transformations to Handle Extreme Values of the Dependent Variable," *Journal of the American Statistical Association*, 83:401 March, Theory and methods, 123-127.

Butler, J.S. and R. Moffit (1982) "A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial

Probit Model," *Econometrica*, 50, 761-764.

Cameron, Trudy A. (1992) "Combining Contingent Valuation and Travel Cost Data for the Valuation of Nonmarket Goods," *Land Economics*, 68, August, 302-317.

Cameron, T.A. (1988) "A New Paradigm for Valuing Non-Market Goods Using Referendum Data: maximum Likelihood Estimation by Censored Logistic Regression," *Journal of Environmental Economics and Management*, 15, 355-379.

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

Cameron, Trudy A. and Michelle D. James (1987) "Efficient Estimation Methods for 'Closed-Ended' Contingent Valuation Surveys." *Review of Economics and Statistics*, 69:2, 269-276.

Cameron, Trudy A. and John Quiggin (1994) "Estimation Using Contingent Valuation Data from a 'Dichotomous Choice with Follow-up' Questionnaire," *Journal of Environmental Economics and Management*, 27:3, 218-234.

Carroll, J. Douglas and Paul E. Green (1995) "Psychometric Methods in Marketing Research: Part 1, Conjoint Analysis," *Journal of Marketing Research*, 32, November, 385-391.

Carroll, R.J., D. Ruppert and L.A. Stefanski (1995) *Measurement Error in Nonlinear Models*, New York: Chapman and Hall.

Carroll, R.J., C.H. Spiegelman, K.K.G. Lan, K.T. Bailey and R.D. Abbott (1984) "On Errors-in Variables for Binary Regression Models," *Biometrika*, 71:1, 19-25.

Carson, Richard T. (1985) *Three Essays on Contingent Valuation (Welfare Economics, Non-Market Goods, Water Quality)*, Ph.D. dissertation, Department of Agricultural and Resource Economics, University of California, Berkeley.

Carson, Richard T., Nicholas E. Flores, and W. Michael Hanemann (1992) "On the Nature of Compensable Value in a Natural Resource Damage Assessment." Presented at the American Economic Association Annual Meeting, New Orleans, January 3, 1992.

Carson, Richard T., Nicholas E. Flores, Kerry M. Martin, and Jennifer L. Wright (1996) "Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-public Goods," *Land Economics*, 72, February, 80-99.

Carson, Richard T., W. Michael Hanemann, Raymond J. Kopp, Jon A. Krosnick, Robert C. Mitchell, Stanley Presser, Paul A. Ruud, and V. Kerry Smith, with Michael Conaway and Kerry Martin (1995) "Referendum Design and Contingent Valuation: The NOAA Panel's No-Vote Recommendation," Discussion Paper 96-05, Washington, D.C., Resources for the Future, November.

Carson, Richard T., Michael Hanemann, Raymond J. Kopp, Jon A. Krosnick, Robert C. Mitchell, Stanley Presser, Paul A. Ruud, and V. Kerry Smith (1994) *Prospective Interim Lost Use Value Due to DDT and PCB Contamination in the Southern California Bight*, A Report by Natural Resource Damage Assessment, Inc., La Jolla, California, September 30.

Carson, Richard T., Michael Hanemann, and Dan Steinberg (1990) "A Discrete Choice Contingent Valuation Estimate of the Value of Kenai King Salmon," *Journal of Behavioral Economics*, 19:1, 53-68.

Carson, Richard T., Robert C. Mitchell, W. Michael Hanemann, Raymond J. Kopp, Stanley Presser, and Paul A. Ruud (1992) *A Contingent Valuation Study of Lost Passive Use Values Resulting from the Exxon Valdez Oil Spill*. A report to the Attorney General of the state of Alaska, November 10.

Carson, Richard T. and Dan Steinberg (1990) "Experimental Design for Discrete Choice Voter Preference Surveys." in *1989 Proceeding of the Survey Methodology Section of the American Statistical Association.* Washington, DC: American Statistical Association.

Carson, Richard T., Leanne Wilks and David Imber (1994) "Valuing the Preservation of Australia's Kakadu Conservation Zone." *Oxford Economic Papers*, 46, 727-749.

Chaloner, K. (1986) "An Approach to Experimental Design for Generalized Models," in *Optimal Design and Analysis of Experiments*, eds. Y. Dodge, V. V. Fedorov, and H. P. Wynn; Elsevier Science Publishers.

Chamberlain, G. (1980) "Analysis of Covariance with Qualitative Data," *Review of Economic Studies*, 47, 225-238.

Chamberlain, G. (1984) "Panel data," in Z. Griliches and M. Intriligator, eds., *Handbook of Econometrics*, Amsterdam: North-Holland, 2:22.

Chapman, David, and W. Michael Hanemann (1993) "Correlated Discrete-Response Contingent Valuation" Department of Agricultural and Resource Economics, Working Paper, University of California. Berkeley, July.

Chapman, David, W. Michael Hanemann, and Barbara Kanninen (1993) "Non-Market Valuation Using Contingent Behavior: Model Specification and Consistency Tests." Department of Agricultural and Resource Economics, Working Paper, University of California. Berkeley. January.

Chapman, Randall G., and Richard Staelin (1982) "Exploiting Rank-Ordered Choice Set Data within the Stochastic Utility Model," *Journal of Marketing Research* 19, 288-301.

Chen, Heng Z. (1993) *Semiparametric and Semi-nonparametric Welfare Impact Analyses with Applications to Natural Resource Valuation*, Ph.D dissertation, Ohio State University.

Chen, Yu-Lan (1994) *Valuing Environmental Amenities with Revealed and Stated Preference Information: An Application to Gray Whales in California*, Ph.D. dissertation, Department of Agricultural Economics, University of California, Davis.

Chesher, Andrew, and Margaret Irish (1987) "Residual Analysis in the Grouped and Censored Normal Linear Model," *Journal of Econometrics*, 34, 33-61.

Church, J.D. and E.B. Cobb (1973) "On the equivalence of Spearman-Kärber and Maximum Likelihood Estimates of the Mean," *Journal of the American Statistical Association*, 68, 201-202.

Clark, C. (1961) "The Greatest of a Finite Set of Random Variables," *Operations Research*, 9, 145-162.

Cochran, W.G. and M. Davis (1965) "The Robbins-Monro Method for Estimating the Median Lethal Dose," *Journal of the Royal Statistical Society*, Ser. B, 27 28-44.

Collett, D. (1991) *Modelling Binary Data*, London: Chapman and Hall.

Cook, R. Dennis, Stanford Weisberg (1982), *Residual and Influence in Regression*, New York: Chapman and Hall.

Cook, R. Dennis, Stanford Weisberg (1994), *An Introduction to Regression Graphics*, New York: John Wiley and Sons.

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. (1994) "A Comparison of Approaches to Calculating Confidence Intervals for Benefit Measures from Dichotomous Choice Contingent Valuation Surveys," *Land Economics*, 70:1, 111-22.

Cooper, J., and W. Michael Hanemann (1995) "Referendum Contingent Valuation: How Many Bounds Are Enough?," USDA Economic Research Service, Food and Consumer Economics Division, working paper, May.

Cooper, J. and J. Loomis (1992) "Sensitivity of Willingness-to-Pay Estimates to Bid Design in Dichotomous Choice Contingent Valuation Models," *Land Economics*, 68:2 May, 211-224.

Cooper, J. and J. Loomis (1993) "Sensitivity of Willingness-to-Pay Estimates to Bid Design in Dichotomous Choice Contingent Valuation Models: Reply," *Land Economics*, 69:2 May, 203-208.

Copas, J.B. (1983) "Plotting *p* against *x*," *Applied Statistics* 32:1, 25-31.

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

Cosslett, S.R. (1983) "Distribution-Free Maximum Likelihood Estimator of the Binary Choice Model." *Econometrica* 51:3, 765-782.

Cosslett, Stephen R. (1987) "Efficiency Bounds for Distribution-Free Estimators of the Binary Choice and the Censored Regression Models," *Econometrica* 55, 559-586.

Cox, C. (1990) "Fieller's Theorem, The Likelihood and the Delta Method," *Biometrics*, 46, 709-718.

Cox, D.R. (1972) "The Analysis of Multivariate Binary Data," *Applied Statistics*, 21, 113-120.

Cox, D.R. and D.V. Hinkley (1974) *Theoretical Statistics,* London: Chapman and Hall.

Creel, Michael and John Loomis (1995) "Semi-Nonparametric Distribution-Free Dichotomous Choice Contingent Valuation," working paper, Department of Agricultural and Resource Economics, Colorado State University, Fort Collins, September.

Davidson, R. and J.G. MacKinnon (1981) "Several Tests for Model Specification in the Presence of Alternative Hypotheses," *Econometrica,* 49:3, 781-793.

Davidson, R. and J.G. MacKinnon (1984) "Convenient Specification Tests for Logit and Probit Models," *Journal of Econometrics,* 25, 241-262.

Davis, R. (1963) "Recreation Planning as an Economic Problem," *Natural Resources Journal*, 3, 239-249.

Deacon, R., and P. Shapiro (1975) "Private Preference for Collective Goods Revealed Through Voting on Referenda," *American Economic Review*, 65:5, 943-955.

Debreu, G. (1960) "Review of R.D. Luce Individual Choice Behavior." *American Economic Review*, 50, 186-188.

Desvousges, William H., F. Reed Johnson, Richard W. Dunford, Sara P. Hudson, K. Nicole Wilson, and Kevin J. Boyle (1993) *Measuring Nonuse Damages Using Contingent Valuation: An Experimental Evaluation of Accuracy* in *Contingent Valuation: A Critical Assessment*, J.A. Hausman, ed., New York: North-Holland, 91-164.

Dickie, M., A. Fisher and S. Gerking (1987) "Market Transactions and Hypothetical Demand: A Comparative Study," *Journal of the American Statistical Association*, 82, 67-75.

Dickie, Mark and Shelby Gerking (1996) "Defensive Action and Willingness to Pay for Reduced Health Risk: Inference from Actual and Contingent Behavior," presented at the 1996 AERE Workshop on Combining Stated and Revealed Preference, Tahoe City, June 2-4.

Dixon, W.J., (1965) "The Up and Down Method in Small Samples," *Journal of the American Statistical Association*, 60, 967-978.

Dixon, W.J. and A.M. Mood (1948) "A Method for Obtaining and Analyzing Sensitivity Data," *Journal of the American Statistical Association*, 43, March, 109-126.

Duffield, John W. (1993) "An Economic Analysis of Wolf Recovery in Yellowstone: Park Visitor Attitudes and Values," working paper, Department of Economics, University of Montana, March.

Duffield, J.W. and D.A. Patterson (1991) "Inference and Optimal Design for a Welfare Measure in Dichotomous Choice Contingent Valuation," *Land Economics*, 67, 225-239.

Efron, Bradley, and Robert J. Tibshirani (1993) *An Introduction to the Bootstrap*, New York: Chapman and Hall.

El-Saidi, Mohammed A., Karan P. Singh, and Effat A. Moussa (1992) "Symmetric Generalized Logistic Models for Quantal Assay Data: a Review," Proceedings of the American Statistical Association Statistical Computing Section Annual Meeting, 113-118.

Englin, Jeffrey and Trudy Ann Cameron (1996) "Augmenting Travel Cost Models with Contingent Behavior Data," *Environmental and Resource Economics*, 7, 133-147.

Eom, Young Sook and V. Kerry Smith (1994) "Calibrated Nonmarket Valuation," working paper, Department of Economics, North Carolina State University.

Fieller, E.C. (1954) "Some Problems in Interval Estimation," *Journal of the Royal Statistical Society*, Ser. B, 16, 175-185.

Finney, D.J. (1964) *Statistical Methods in Biological Assay* (2nd edition), London: Griffin.

Finney, D.J. (1978) *Statistical Methods in Biological Assays*, New York, NY: Oxford University Press.

Fitzmaurice, Garrett M., and Nan M. Laird (1993) "A Likelihood-Based Method for Analyzing Longitudinal Binary Responses," *Biometrika*, 80, 141-151.

Follmann, Dean A. and Diane Lambert (1989) "Generalizing Logistic Regression by Nonparametric Mixing," *Journal of the American Statistical Association*, 84:405, 295-300.

Freeman III, A. Myrick (1991) "Factorial Survey Methods and Willingness to Pay for Housing Characteristics: A Comment," *Journal of Environmental Economics and Management* 20, 92-96.

Gallant, A.R. (1981) "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form," *Journal of Econometrics*, 15, 211-245.

GAUSS 3.0 Applications: Maximum Likelihood (1992) Aptech Systems, Maple Valley, WA.

Georgescu-Roegen, N. (1936) "The Pure Theory of Consumer Behavior," *Quarterly Journal of Economics*, 570.

Georgescu-Roegen, N. (1958) "Threshold in Choice and the Theory of Demand," *Econometrica*, 26.

Gill, Philip E., Walter Murray, and Margaret H. Wright (1981) *Practical Optimization*, New York: Academic Press.

Glasbey, C.A. (1987) "Tolerance-Distribution-Free Analyses of Quantal Dose-Response Data," *Applied Statistics*, 36:3, 251-259.

Goldstein, Harvey (1991) "Nonlinear Multilevel Models, with an Application to Discrete Response Data," *Biometrika*, 78:1, 45-51.

Gourieroux, C., A. Monfort and A. Trognon (1983) "Testing Nested or Non-Nested Hypotheses," *Journal of Econometrics*, 21, 83-115.

Gourieroux, C., A. Monfort, E. Renault, and A. Trognon (1984) "Pseudo Maximum Likelihood Methods: Theory," *Econometrica*, 52:3 May, 681-700.

Green, Paul E. (1974) "On the Design of Choice Experiments Involving Multifactor Alternatives," *Journal of Consumer Research*, 1, September, 61-68.

Green, Paul E., and V. Srinivasan (1978) "Conjoint Analysis in Consumer Research: Issues and Outlook," *Journal of Consumer Research*, 5, 103-123.

Green, P.E., and V. Srinivasan (1990) "Conjoint Analysis in Marketing Research: New Developments and Directions," *Journal of Marketing*, 54:4, 3-19.

Greene, W.H. (1993) *Econometric Analysis*, New York: Macmillan.

Gregory, Robin, Sarah Lichtenstein and Paul Slovic (1993) "Valuing Environmental Resources: A Constructive Approach." *Journal of Risk and Uncertainty*, 7:177-197.

Griffiths, W.E., R. Carter Hill and P.J. Pope (1987) "Small Sample Properties of Probit Model Estimators," *Journal of the American Statistical Association*, 82, 929-37.

Guilkey, D.K. and J.L. Murphy (1993) "Estimation and Testing in the Random Effects Probit Model," *Journal of Econometrics*, 59, 301-317.

Hajivassiliou, V. (1993) "Simulation Estimation Methods for Limited Dependent Variable Models," in G.S. Maddala, C.R. Rao and H.D. Vinod, eds., *Handbook of Statistics (Econometrics)*, Vol. 11. Amsterdam: North-Holland, 519-543.

Hajivassiliou, Vassilis A. and Paul A. Ruud (1994) "Classical Estimation Methods for LDV Models Using Simulation," in R.F. Engle and D.L. McFadden eds. *Handbook of Econometrics*, Vol. 14, Elsevier Science B.V., Chapter 40, 2383-2441.

Hamerle, Alfred and Gerd Ronning (1995) "Panel Analysis for Qualitative Variables," in G. Arminger, C.C. Clogg, and M.E. Sobel, eds., *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, New York: Plenum Press, 401-451.

Hamilton, M.A., R.C. Russo, and R.V. Thurston (1977) "Trimmed Spearman-Karber Method for Estimating Median Lethal Concentrations in Toxicity Bioassay," *Environmental Science and Technology*, 11, 714-719.

Hampel, F.R. (1974) "The Influence Curve and Its Role in Robust Estimation," *Journal of the American Statistical Association* 62, 1179-1186.

Han, Aaron K. (1987) "Non-Parametric Analysis of a Generalized Regression Model: The Maximum Rank Correlation Estimator," *Journal of Econometrics* 35, 303-316.

Hanemann, W. Michael (1983a) "Marginal Welfare Measures for Discrete Choice Models." *Economics Letters*, 13, 129-136.

Hanemann, W. Michael (1983b) "Welfare Evaluations with Simulated and Hypothetical Market data: Bishop and Heberlein Revisited," Department of Agricultural and Resource Economics, Working paper 276, University of California, Berkeley, July.

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

Hanemann, W. Michael (1984b) "Discrete/Continuous Models of Consumer Demand." *Econometrica* 52, 541-562.

Hanemann, W. Michael (1985a) "Some Issues in Continuous- and Discrete-Response Contingent Valuation Studies." *Northeastern Journal of Agricultural Economics*, Vol. 14 (April), 5-13.

Hanemann, W. Michael (1985b) "Welfare Analysis with Discrete Choice Models." Department of Agricultural and Resource Economics, Working Paper, University of California. Berkeley, August.

Hanemann, W. Michael (1988) "Three Approaches to Defining 'Existence' or 'Non-Use' Value Under Certainty," Department of Agricultural and Resource Economics, Working paper 691, University of California, Berkeley.

Hanemann, W. Michael (1989) "Welfare Evaluations in Contingent Valuation Experiments with Discrete Response Data: Reply." *American Journal of Agricultural Economics*, Vol. 71 (November) 1057-1061.

Hanemann, W. Michael (1991) "Reviewer’s Comments" In David Imber, Gay Stevenson and Leanne Wilks, *A Contingent Valuation Survey of the Kakadu Conservation Zone.* Canberra, Australia: Resource Assessment Commission, RAC Research Paper No. 3, February, 187-192.

Hanemann, W. Michael (1996) "Theory Versus Data in the Contingent Valuation Debate," In *The Contingent Valuation of Environmental Resources: Methodological Issues and Research Needs*, edited by David J. Bjornstad and James R. Kahn, Cheltenhan: Edward Elgar Publishing.

Hanemann, W. Michael and David Chapman (1988) "Beyond Contingent Valuation: Deriving Environmental Benefits from Hypothetical Data." Department of Agricultural and Resource Economics, Working Paper, University of California. Berkeley, October.

Hanemann, W. Michael, Bengt Kriström, and Chuan-Zhong Li (1995) "Nonmarket Valuation under Preference Uncertainty: Econometric Models and Estimation," paper presented at the Annual Meeting of the European Association of Environmental & Resource Economists, Umea, Sweden, June 17-20.

Hanemann, W. Michael, John B. Loomis and Barbara J. Kanninen (1991) "Statistical Efficiency of Double-Bounded Dichotomous Choice Contingent Valuation." *American Journal of Agricultural Economics* 73:4, 1255-1263.

Hasselblad, Victor, Andrew G. Stead, and John P. Creason (1980) "Multiple Probit Analysis with a Non-zero Background," *Biometrics*, 36, December, 659-63.

Hastie, Trevor, and Robert Tibshirani (1987) "Non-Parametric Logistic and Proportional Odds Regression," *Applied Statistics*, 36:2, 260-276.

Hausman, J.A. ed., (1993) *Contingent Valuation: A Critical Assessment*. New York: North-Holland.

Hausman, J.A. and P.A. Ruud (1987) "Specifying and Testing Econometric Models for Rank-Ordered Data," *Journal of Econometrics* 34, 83-104.

Hausman, J.A. and D.A. Wise (1978) "Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica* 46:2, 403-426.

Heckman, J. (1978) "Dummy Exogenous Variables in a Simultaneous Equation System," *Econometrica* 46:931-59.

Heckman, J.J. and R. Willis (1975) "Estimation of a Stochastic Model of Reproduction: An Econometric Approach," in N. Terleckyi ed., *Household Production and Consumption*, New York: National Bureau of Economic Research, 99-138.

Hensher, David A. (1994) "Stated Preference Analysis of Travel Choices: The State of the Practice," *Transportation*, Volume 21, May.

Herriges, J.A. and J.F. Shogren (1996) "Starting Point Bias in Dichotomous Choice Valuation with Follow-up Questioning," *Journal of Environmental and Management*, 30(1): 112-131.

Hoaglin, David, Frederick Mosteller, John W. Tukey (1983) *Understanding Robust and Exploratory Data Analysis*, New York: John Wiley and Sons, Inc.

Hoehn, John P. (1991) "Valuing the Multidimensional Impacts of Environmental Policy; Theory and Methods," *American Journal of Agricultural Economics*, 73:2 May, 289-99.

Hoehn, John P., and John B. Loomis (1993) "Substitution Effects in the Valuation of Multiple Environmental Programs," *Journal of Environmental Economics and Management* 25:1 Part 1, July, 56-75.

Hoehn, John P., and Alan Randall (1987) "A Satisfactory Benefit Cost Indicator from Contingent Valuation," *Journal of Environmental Economics and Management*, 14, September, 226-247.

Horowitz, J. (1982) "An Evaluation of the Usefulness of Two Standard Goodness-of-Fit Indicators for Comparing Non-Nested Random Utility Models," *Trans. Research Record*, 874, 19-25.

Horowitz, J.L. (1983) "Statistical Comparison of Non-Nested Probabilistic Discrete Choice Models," *Transportation Science*, 17:3, 319-50.

Horowitz, J.L. (1992) "A Smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica*, 60:3 May, 505-531.

Horowitz, J.L. (1993) "Semiparametric and Nonparametric Estimation of Quantal Response Models," in *Handbook of Statistics, Vol 11*, G.S. Maddala, C.R. Rao and H.D. Vinod, eds., Elsevier Science Publishers, B.V.

Horowitz, J., Sparmonn, J. and C. Daganzo (1984) "An Investigation of the Accuracy of the Clark Approximation for the Multinomial Probit Model," *Transportation Science*, 16, 382-401.

Hsiao, C. (1986) *Analysis of Panel Data*, Cambridge: Cambridge University Press.

Huber, P.J. (1981) *Robust Statistics*, New York: John Wiley and Sons.

Huhtala, Anni H. (1994) *Is Environmental Guilt a Driving Force? An Economic Study on Recycling*, Ph.D. dissertation, Department of Agricultural and Resource Economics, University of California, Berkeley, December.

Hurwicz, L. and H. Uzawa (1971) "On the Integrability of Demand Functions." In J.S. Chipman, L. Hurwicz, M.K. Richter, and H.F. Sonnenschen, eds. *Preferences, Utility and Demand*, New York: Harcourt Brace, 114-148.

Ichimura, Hidehiko (1993) "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models," *Journal of Econometrics* 58, 71-120.

Jakus, Paul M. (1994) "Averting Behavior in the Presence of Public Spillovers: Household Control of Nuisance Pests," *Land Economics*, 70, August, 273-285.

Jakus, Paul M. and V. Kerry Smith (1992) "Measuring Use and Nonuse Values for Landscape Amenities: A Contingent Behavior Analysis of Gypsy Moth Control" presented at the American Economic Association Annual Meeting, New Orleans, January.

James, B.R. and James, K.L. (1983) "On the Influence Curve for Quantal Bioassay," *J. Statist. Pl. Inf.*, 8, 331-345.

Jennings, Dennis E. (1986) "Outlier and Residual Distribution in Logistic Regression," *Journal of the American Statistical Association*, 81:396 December, 987-990.

Johansson, P. and B. Kriström (1988) "Asymmetric and Symmetric Discrete Response Models in Contingent Valuation Experiments," Research Papers in Statistics, 7, Dept. of Statistics, University of Umeå.

Johnson, Norman L., and Samuel Kotz (1972) *Distributions in Statistics: Continuous Multivariate Distributions*. New York: John Wiley and Sons, Inc.

Judge G.G., R.C. Hill, W.E. Griffiths, H. Lütkepohl and L. Tsoung-Chao (1988) *Introduction to the Theory and Practice of Econometrics*, Second Edition, New York: John Wiley & Sons.

Kalbfleisch, J.D. and R.L. Prentice (1980) *The Statistical Analysis of Failure Time Data*, New York: John Wiley & Sons.

Kanninen, B.J. (1993a) "Design of Sequential Experiments for Contingent Valuation Studies," *Journal of Environmental Economics and Management*, 25, S-1-11.

Kanninen, B.J. (1993b) "Optimal Experimental Design for Double-Bounded Dichotomous Choice Contingent Valuation," *Land Economics*, 69:2, 138-146.

Kanninen, B.J. (1995) "Bias in Discrete Response Contingent Valuation," *Journal of Environmental Economics and Management*, 28:1 January, 114-125.

Kanninen, B.J. and B. Kriström (1993) "Sensitivity of Willingness-to-Pay Estimates to Bid Design in Dichotomous Choice Contingent Valuation Models: A Comment," *Land Economics*, 69:2, 199-202.

Kao, C and J.R. Schnell (1987) "Errors in Variables in the Multinomial Response Model," *Economics Letters*, 25, 249-254.

Kaplan, E.L. and P. Meier (1958) "Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, 53, 457-481.

Kappenman, Russell F. (1987) "Nonparametric Estimation of Dose-Response Curves with Application to ED50 Estimation," *J. Statist. Comput. Simul.*, 28, 1-13.

Keane, M. (1993) "Simulation Estimation Methods for Panel Data Limited Dependent Variable Models," in G.C. Maddala, C.R. Rao and H.D. Vinod, eds., *Handbook of Statistics (Econometrics)*, Vol. 11. Amsterdam: North-Holland.

Keeney, Ralph L. and Howard Raiffa (1976) *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, New York: John Wiley and Sons.

Kent, J.T. (1986) "The Underlying Structure of Nonnested Hypothesis Tests," *Biometrika*, 73:2, 333-343.

Klein, R.W. and R.H. Spady (1993) "An Efficient Semiparametric Estimator for Binary Response Models," *Econometrica*, 61:2, 387-421.

Knetch, J. L. and R. K. Davis (1966) "Comparisons of Methods for Recreation Evaluation" in Kneese A.V., and S. C. Smith, eds., *Water Research*, Baltimore: Resources for the Future Inc., Johns Hopkins Press, 125-142.

Krinsky, I. and A. Robb (1986) "Approximating the Statistical Properties of Elasticities," *Review of Economics and Statistics*, 68, 715-719.

Krishnan, K.S. (1977) "Incorporating Thresholds of Indifference in Probabilistic Choice Models," *Management Science*, 23, July, 1224-1233.

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

Kriström, Bengt, Hans Nyquist and W. Michael Hanemann (1992) "Dichotomous Choice Contingent Valuation Experiment: A Mixed Distribution Approach." Working Paper, Department of Economics, University of Umea, Sweden. November.

Krutilla, J. V. (1967) "Conservation Reconsidered," *American Economic Review*, 57:4, 777-786.

Künsch, Hans R., Leonard A. Stefanski, and Raymond J. Carroll (1989) "Conditionally Unbiased Bounded-Influence Estimation in General Regression Models, with Applications to Generalized Linear Models," *Journal of the American Statistical Association*, 84:406 June, 460-466.

Lancaster, Tony (1990) *The Econometric Analysis of Transition Data*, Cambridge, MA: Cambridge University Press.

Langford, I. (1994) "Using a Generalized Linear Mixed Model to Analyze Dichotomous Choice Contingent Valuation Data," *Land Economics*, 70, 507-514

Langford, Ian H. and Ian J. Bateman, "Multilevel Modelling and Contingent Valuation," chapter 12, this volume.

Layton, David Frost (1995) *Specifying and Testing Econometric Models for Stated Preference Surveys*, Ph.D. dissertation, Department of Economics, University of Washington, Seattle, June.

Lechner, Michael (1995) "Some Specification Tests for Probit Models Estimated on Panel Data," *Journal of Business and Economic Statistics*, 13:4 Oct., 475-488.

Lee, Lung-Fei (1982) "Specification Error in Multinomial Logit Models: Analysis of the Omitted Variable Bias," *Journal of Econometrics* 20, 197-209.

Lee, Lung-Fei (1983) "Generalized Econometric Models with Selectivity," *Econometrica* 55:2 March, 507-512.

Lerman, S.R. and C.F. Manski (1981) "The Use of Simulated Frequencies to Approximate Choice Probabilities," in Manski, C.F. and D. McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge, Mass: MIT Press, 305-319.

Li, Chuan-Zhong (1996) "Semiparametric Estimation of the Binary Choice Model for Contingent Valuation," *Land Economics*, forthcoming, November.

Li, Chuan-Zhong, Karl-Gustaf Löfgren, and W. Michael Hanemann (1995) "Real versus Hypothetical Willingness to Accept: The Bishop and Heberlein Model Revisited." Working Paper, Department of Economics, University of Umea, Sweden. September.

Li, Chuan-Zhong and Lief Mattsson (1995) "Discrete Choice Under Preference Uncertainty: An Improved Structural Model for Contingent Valuation." *Journal of Environmental Economics and Management*, 28:2 March, 256-269.

Liang, Kung-Yee, Scott L. Zeger, and Bahjat Qaqish (1992) "Multivariate Regression Analyses for Categorical Data." *Journal of the Royal Statistical Society*, Ser. B, 54:1, 3-40.

Little, I.M.D. (1957) *A Critique of Welfare Economics*, New York, NY: Oxford University Press.

Little, Roderick J.A. and Donald B. Rubin (1987) *Statistical Analysis with Missing Data*, New York: Wiley & Sons.

Longford, N.T. (1988) "A Quasi-likelihood Adaptation for Variance Component Analysis," in *Proc. Sect. Comp. Statist., Am. Statist. Assoc.*, 137-142.

Longford, N.T. (1993) *Random Coefficient Models*, Oxford: Clarendon Press.

Loomis, John B. (1996) "Panel Estimators Useful for Analyzing Combined Revealed and Stated Preference Dichotomous Choice Data: Application to Visit and Dichotomous Choice CVM Responses for Instream Flow," presented at the 1996 AERE Workshop on Combining Stated and Revealed Preference, Tahoe City, June 2-4.

Loomis, J. B., W. M. Hanemann, and T. C. Wegge. (1990) *Environmental Benefits Study of San Joaquin Valley's Fish and Wildlife Resources*. Sacramento, California: Jones & Stokes Associates, Inc. September.

Louviere, Jordan J. (1996) "Combining Revealed and Stated Preference Data: The Rescaling Revolution," presented at the 1996 AERE Workshop on Combining Stated and Revealed Preference, Tahoe City, June 2-4.

Louviere, Jordan J. (1994) "Conjoint Analysis," in R. Bagozzi ed. *Handbook of Marketing Research*, Oxford: Blackwell Publishers.

Louviere, Jordan J. (1988) *Analyzing Decision Making: Metric Conjoint Analysis*, Newbury Park: Sage Publications.

Luce, R.D. (1956) "Semiorders and a Theory of Utility Discrimination," *Econometrica*, 24, 178-191.

Luce, R.D. (1959) *Individual Choice Behavior: A Theoretical Analysis*, New York, NY: John Wiley and Sons.

Luce, R.D. and P. Suppes (1965) "Preference, Utility and Subjective Probability," in R.D. Luce and R.R. Bush and E. Galanter (eds.) *Handbook of Mathematical Psychology*, Vol. 3, New York: John Wiley and Sons, 249-410.

Mackenzie, John (1993) "A Comparison of Contingent Preference Models," *American Journal of Agricultural Economics*, 75, August, 593-603.

MacKinnon, J.G. (1983) "Model Specification Tests Against Non-Nested Alternatives," *Econometric Reviews*, 2:1, 85-110.

Maddala, G.S. (1983) *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge: Cambridge University Press.

Maddala, G.S. (1986) "Limited Dependent Variable Models Using Panel Data," *Journal of Human Resources*, 22, 307-338.

Madow, W.G., I. Olkin and D. Rubin, eds. (1983) *Incomplete Data in Sample Surveys, Volume 2, Theory and Bibliographies*, New York: Academic Press, Inc.

Manski, Charles F. (1975) "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics* 3, 205-228.

Mardia, K.V. (1970) *Families of Bivariate Distributions*, London: Griffin.

Mazzotta, Marisa J. and James J. Opaluch (1995) "Decision Making When Choices are Complex: A Test of Heiner's Hypothesis," *Land Economics*, 71:4, 500-515.

McConnell, K. E. (1990) "Models for Referendum Data: The Structure of Discrete Choice Models for Contingent Valuation," *Journal of Environmental Economics and Management*, 18, 19-35.

McConnell, K. E. (1983) "Existence and Bequest Values" in Rowe, R. and Chestnut, L. eds, *Managing Air Quality and Scenic Resources at National Parks and Wilderness Areas*, Boulder: Westview Press.

McCullagh, P., and J.A. Nelder (1989) *Generalized Linear Models*, 2nd ed., London: Chapman and Hall.

McFadden, D. (1974) "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, ed., *Frontiers in Econometrics*, New York: Academic Press, 105-142.

McFadden, D. (1976) "Quantal Choice Analysis: A Survey," *Annals of Economic and Social Measurement*, 5, 363-390.

McFadden, D. (1978) "Modelling the Choice of Residential Location," in A. Karlqvist, ed., *Spatial Interaction Theory and Residential Location*, Amsterdam: North-Holland, 75-96.

McFadden, D. (1987) "Regression-Based Specification Tests for the Multinomial Logit Model," *Journal of Econometrics*, 34, 63-82.

McFadden, D. (1989) "A Method of Simulated Moments for the Estimation: Discrete Response Models without Numerical Integration," *Econometrica* 57:5, 995-1026.

McFadden, Daniel (1994) "Contingent Valuation and Social Choice," *American Journal of Agricultural Economics*, 76, November, 689-708.

McFadden, Daniel, and Gregory Leonard (1993) "Issues in the Contingent Valuation of Environmental Goods: Methodologies for Data Collection and Analysis." In Hausman, J.A., ed. *Contingent Valuation: A Critical Assessment*. New York: North-Holland, 165-215.

Miller, R.G. (1973) "Nonparametric Estimators of the Mean Tolerance in Bioassay," *Biometrika*, 60, 535-542.

Miller, R.G. and J.W. Halpern (1980) "Robust estimators for quantal bioassay," *Biometrika*, 67, 103-110.

Minkin, S. (1987) "Optimal Design for Binary Data," *Journal of the American Statistical Association*, 82, 1098-1103.

Mitchell, Robert Cameron, and Richard T. Carson (1989) *Using Surveys to Value Public Goods: The Contingent Valuation Method,* Washington, D.C.: Resources for the Future.

Mitchell, Robert Cameron, and Richard T. Carson (1981) "An Experiment in Determining Willingness to Pay for National Water Quality Improvements," Report to the U.S. Environmental Protection Agency, Washington, D.C, June.

Morgan, B.J.T. (1985) "The Cubic Logistic Model for Quantal Assay Data," *Appl. Statist.* 34:2, 105-113.

Morgan, B.J.T. (1992) *Analysis of Quantal Response Data*, New York: Chapman and Hall.

Morikawa, T. (1989) *Incorporating Stated Preference Data in Travel Demand Analysis*, Ph.D. dissertation, Department of Civil Engineering, Massachusetts Institute of Technology.

Muller, H.G. and T. Schmitt (1990) "Choice of Number of Doses for Maximum Likelihood Estimation of the ED_{50} for Quantal Dose-Response Data," *Biometrics*, 46:1, 117-130.

Mundlak, Y. (1978) "On the Pooling of Cross-Section and Time-Series Data," *Econometrica,* 46, 69-86.

Nelder, J.A. and Wedderburn, R.W.M. (1972) "Generalized Linear Models," *Journal of the Royal Statistical Society*, A, 135, 370-84.

Nelson, Wayne (1982) *Applied Life Data Analysis*, New York: John Wiley & Sons.

Nerlove, M., and J. Press (1973) "Univariate and Multivariate Log-Linear and Logistic Models," RAND report R-1306-EDA/NIH.

Neyman, J. and E.L. Scott (1948) "Consistent Estimates Based on Partially Consistent Observations," *Econometrica* 16, 1-32.

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

Opaluch, James J., and Kathleen Segerson (1989) "Rational Roots of 'Irrational' Behavior: New Theories of Economic Decision-Making," *Northeastern Journal of Agricultural and Resource Economics* 18(Oct.), 81-95.

O'Sullivan, F., Yandell, B. and W. Raynor (1986) "Automatic Smoothing of Regression Functions in Generalized Linear Models," *Journal of the American Statistical Association*, 81, 96-103.

Pakes, A., and D. Pollard (1989) "Simulation and the Asymptotics of Optimization Estimators," *Econometrica,* 57:5, 1027-1057.

Park, T., J. Loomis and M. Creel (1991) "Confidence Intervals for Evaluating Benefits from Dichotomous Choice Contingent Valuation Studies," *Land Economics,* 67:1, 64-73.

Parzen, Emmanuel (1960) *Modern Probability Theory and its Applications*, New York: John Wiley & Sons.

Pesaran, M.H. and A.S. Deaton (1978) "Testing Non-Nested Nonlinear Regression Models," *Econometrica,* 46:3, 677-694.

Pierce, Donald A. and Daniel E. Schafer (1986) "Residuals in Generalized Linear Models," *Journal of the American Statistical Association*, 81:396, December, 977-986.

Poe, Gregory L., Eric K. Severance-Lossin, and Michael P. Welsh (1994) "Measuring the Difference (X–Y) of Simulated Distributions: A Convolutions Approach," *American Journal of Agricultural Economics*, 76, November, 904-915.

Powell, J.L., James H. Stock, and Thomas M. Stoker (1989) "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57:6, 1403-1430.

Pregibon, D. (1981) "Logistic Regression Diagnostics," *Annals of Statistics*, 9, 705-724.

Pregibon, D. (1982) "Resistant Fits for Some Commonly Used Logistic Models with Medical Application," *Biometrics*, 38, 485-498.

Pregibon, Daryl (1988) "Discussion of the Paper by Copas" contained in, "Binary Regression Models for Contaminated Data, by J.B. Copas" *Journal of Royal Statistical Society*, 50:2, 260-261.

Preisler, H.K. (1989) "Fitting Dose-Response Data with Non-Zero Background within Generalized Linear and Generalized Additive Models," *Computational Statistics and Data Analysis* 7, 279-290.

Prentice, R.L. (1976) "A Generalization of the Probit and Logit Methods for Dose Response Curves." *Biometrics* 32, 761-768.

Quandt, R. (1956) "A Probabilistic Theory of Consumer Behavior," *Quarterly Journal of Economics*, 70:4, 507-536.

Rae, Douglas, A. (1982) "Benefits of Visual Air Quality in Cincinnati," report to the Electric Power Research Institute by Charles River Associates, Boston.

Rae, Douglas, A. (1983) "The Value to Visitors of Improving Visibility at Mesa Verde and Great Smoky National Parks," in Robert D. Rowe and Lauraine G. Chestnut, eds., *Managing Air Quality and Scenic Resources at National Parks and Wilderness Areas*, Boulder: Westview Press.

Raudenbush, S.W., & A.S. Bryk (1986) "A Hierarchical Model for Studying School Effects," *Sociology of Education*, 59, 1-17.

Ready, Richard C., and Dayuan Hu (1995) "Statistical Approaches to the Fat Tail Problem for Dichotomous Choice Contingent Valuation," *Land Economics*, 71:4, 491-499.

Ready, Richard, John C. Whitehead, and Glenn C. Blomquist (1995) "Contingent Valuation When Respondents Are Ambivalent," *Journal of Environmental Economics and Management*, 29, 181-196.

Robinson, P.M. (1982) "On the Asymptotic Properties of Estimators of Models Containing Limited Dependent Variables," *Econometrica* 50, 27-41.

Roe, Brian, Kevin J. Boyle, Mario F. Teisl (forthcoming) "Using Conjoint Analysis to Derive Estimates of Compensating Variation," *Journal of Environmental Economics and Management*.

Rubin D. (1987) *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons, Inc.

Ruud, Paul A. (1983) "Sufficient Conditions for the Consistency of Maximum Likelihood Estimation Despite Misspecification of Distribution in Multinomial Discrete Choice Models," *Econometrica*, 51:1, January, 225-228.

Santner, Thomas J. and Diane E. Duffy (1989) *The Statistical Analysis of Discrete Data*. New York: Springer-Verlag.

Schkade, David A. and John W. Payne (1994) "How People Respond to Contingent Valuation Questions: A Verbal

Protocol Analysis of Willingness to Pay for Environmental regulation," *Journal of Environmental Economics and Management*, 26, January, 88-109.

Schmidt, Peter, and Ann Dryden Witte (1989) "Predicting Criminal Recidivism Using 'Split Population' Survival Time Models,' *Journal of Econometrics* 40, 141-159.

Schmidt, Peter, and Robert P. Strauss (1975) "Estimation of Models with Jointly Dependent Qualitative Variables: A Simultaneous Logit Approach," *Econometrica*, 43:4, July, 745-755.

Schmoyer, Richard L. (1984) "Sigmoidally Constrained Maximum Likelihood Estimation in Quantal Bioassay," *Journal of the American Statistical Association*, 79:386, 448-453.

Staniswalis, Joan G., and Vanessa Cooper (1988) "Kernel Estimates of Dose Response," *Biometrics* 44, 1103-1119.

Stavig, Gordon R., and Jean D. Gibbons (1977) "Comparing the Mean and Median as Measures of Centrality," *International Statistical Review*, 45, 63-70.

Steinberg, D. and P. Colla (1988) *Survival: A Supplementary Module for SYSTAT*, Evanston, IL, SYSTAT, Inc.

Strand, Jon (1981) "Valuation of Fresh Water as a Public Good in Norway," Mimeo, Department of Economics, University of Oslo, Norway.

Strauss, D. (1979) "Some Results on Random Utility Models," *Journal of Mathematical Psychology* 20:1, 35-52.

Stukel, Therese A. (1988) "Generalized Logistic Models." *Journal of the American Statistical Association* 83:402, 426-431.

Svento, Rauli (1993) "Some Notes on Trichotomous Choice Discrete Valuation," *Environmental and Resource Economics* 3, 533-543.

Svento, Rauli and Erkki Mäntymaa (1994) "A Generalized Test for Vagueness in Hypothetical Markets," Annual Meeting of the European Association of Environmental & Resource Economists, June.

Swait, Joffre, and Wiktor Adamowicz (1996) "The Effect of Choice Complexity on Random Utility Models: An Application to Combined Stated and Revealed Preference Models, or Tough Choices: Contribution or Confusion?" presented at the 1996 AERE Workshop on Combining Stated and Revealed Preference, Tahoe City, June 2-4.

Swamy, P.A.V.B. (1970) "Efficient Inference in a Random Coefficient Regression Model," *Econometrica* 38, 311-23.

Taylor, Jeremy M.G. (1988) "The Cost of Generalizing Logistic Regression," *Journal of the American Statistical Association,* Dec., 83:404, 1078-1083.

Tsutakawa, R.K. (1972) "Design of Experiment for Bioassay," *Journal of the American Statistical Association*, 67, 584-590.

Tsutakawa, R.K. (1980) "Selection of Dose Levels for Estimating a Percentage Point of a Logistic Quantal Response Curve," *Applied Statistics*, 29, 25-33.

Turnbull, B.W. (1974) "Nonparametric Estimation of a Survivorship Function with Doubly Censored Data," *Journal of the American Statistical Association*, 69:345.

Turnbull, B.W. (1976) "The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data," *Journal of the Royal Statistical Society,* 38, 290-295.

von Winterfeldt, Detlof, and Ward Edwards (1986) *Decision Analysis and Behavioral Research*, New York: Cambridge University Press.

Waldman, Donald M. (1983) "The Information Content of Data in a Random Utility Model," Working Paper, Department of Economics, University of North Carolina, December.

Ward, Michael (1995) *Nonparametric Methods in Economic Analysis*, Ph.D. dissertation, Department of Economics, University of Washington, Seattle.

Wedderburn, R.W.M. (1974) "Quasilikelihood Functions, Generalized Linear Models and the Gauss-Newton Method," *Biometrika*, 61, 439-447.

Welsh, Michael P. and Richard C. Bishop (1993) "Multiple Bounded Discrete Choice Models," in W-133 Benefits and Costs Transfer in Natural Resource Planning, Western Regional Research Publication compiled by J. Bergstrom, Dept. of Ag. and Applied Economics, University of Georgia, September.

Werner, Megan. (1994) *Public Choice Issues in Environmental Economics*, Ph.D dissertation, Department of Economics, University of California, San Diego, 1994.

White, Halbert (1994) *Estimation, Inference and Specification Analysis*, Cambridge, Mass: Cambridge University Press.

Whitehead, John C. (1995), "Willingness to Pay for Quality Improvements: Comparative Statics and Interpretation of Contingent Valuation Results," *Land Economics*, 71, May, 207-215.

Williams, D.A. (1982) "Extra-binomial variation in Logistic Linear Models," *Applied Statistics*, 31:2, 144-48.

Winer, B. J. (1971) *Statistical Principles in Experimental Design*, New York: John Wiley.

Wu, C.F.J. (1985) "Efficient Sequential Designs with Binary Data," *Journal of the American Statistical*

APPENDIX TABLE 1. SINGLE-BOUNDED DISCRETE RESPONSE PROBABILITY MODELS

(A) BOX-COX MODEL $C = y - \left[y^{\lambda} - \frac{\alpha}{b} - \frac{\eta}{b} \right]^{1/\lambda}$ $\quad a > 0 \quad b > 0$

(i) η Standard Normal

$$\boxed{\Pr\{\text{"}yes\text{"}\} \quad = \quad \Phi\left(\alpha - by^{\lambda} + b(y-A)^{\lambda}\right)}$$

(ii) η Standard Logistic

$$\boxed{\Pr\{\text{"}yes\text{"}\} \quad = \quad \left[1 + e^{by^{\lambda} - b(y-A)^{\lambda} - \alpha}\right]^{-1}}$$

(B) LINEAR MODEL $C = \frac{\alpha + \eta}{\beta}$

(i) η Standard Normal (Probit Model)

$$\boxed{\Pr\{\text{"}yes\text{"}\} \quad = \quad \Phi(\alpha - \beta A)}$$

(ii) η Standard Logistic (Logit Model)

$$\boxed{\Pr\{\text{"}yes\text{"}\} \quad = \quad \left[1 + e^{-\alpha + \beta A}\right]^{-1}}$$

(iii) η Standard Lognormal

$$\boxed{\Pr\{\text{"}yes\text{"}\} \quad = \quad \Phi\left[-\ln(-\alpha + \beta A)\right]}$$

(iv) η Standard Log-logistic

$$\boxed{\Pr\{\text{"}yes\text{"}\} \quad = \quad [1 - \alpha + \beta A]^{-1}}$$

(v) η Standard Weibull

$$\boxed{\Pr\{\text{"}yes\text{"}\} \quad = \quad e^{-(-\alpha + \beta A)^{c}} \qquad c > 0}$$

(C) LOG-MODEL $C = y\left[1 - e^{-\frac{(\alpha + \eta)}{\beta}}\right]$

(i) η Standard Normal

$$\boxed{\Pr\{\text{"}yes\text{"}\} \quad = \quad \Phi\left(\alpha + \beta \ln\left(1 - \frac{A}{y}\right)\right)}$$

(ii) η Standard Logistic

$$\boxed{\Pr\{\text{"}yes\text{"}\} \quad = \quad \left[1 + e^{-\alpha}\left(1 - \frac{A}{y}\right)^{-\beta}\right]^{-1}}$$

(D) MULTIPLICATIVE MODEL $C = e^{(\alpha + \varepsilon)/\beta}$

(i) ε Standard Normal (Lognormal Probability Model)

$$\boxed{\Pr\{\text{"}yes\text{"}\} \quad = \quad \Phi(\alpha - \beta \ln A)}$$

(ii) ε Standard Logistic (Log-Logistic Probability Model)

$$\boxed{\Pr\{\text{"}yes\text{"}\} \quad = \quad \frac{1}{1 + e^{-\alpha + \beta \ln A}}}$$

(iii) ε Standard Extreme Value (Weibull Probability Model)

$$\boxed{\Pr\{\text{"}yes\text{"}\} \quad = \quad 1 - \exp\left[-e^{\alpha - \beta \ln A}\right]}$$

APPENDIX TABLE 2. CANONICAL RESPONSE PROBABILITY MODELS SATISFYING C ε [0,y]

(A) MULTIPLICATIVE MODEL $C = e^{(\alpha+\varepsilon)/\beta}$

(1) TRUNCATION AT ε^{max}

(i) ε Standard Normal

$$\Pr\{\text{"}yes\text{"}\} \;=\; 1 - \frac{\Phi(-\alpha+\beta\ln A)}{\Phi(-\alpha+\beta\ln y)}$$

$$C^+ \;=\; y - \frac{1}{\Phi(-\alpha+\beta\ln y)}\int_0^y \Phi(-\alpha+\beta\ln A)\,dA$$

$$C^* \text{ solves: } \Phi\left(-\alpha+\beta\ln\left(C^*\right)\right) = \frac{1}{2}\Phi(-\alpha+\beta\ln y)$$

(ii) ε Standard Logistic

$$\Pr\{\text{"}yes\text{"}\} \;=\; 1 - \frac{1+e^{\alpha-\beta\ln y}}{1+e^{\alpha-\beta\ln A}}$$

$$C^+ \;=\; y - \left(1+e^{\alpha}y^{-\beta}\right)\int_0^y \left(1+e^{\alpha}A^{-\beta}\right)^{-1} dA$$

$$C^* \;=\; e^{\frac{\alpha-\ln\left(1+2e^{\alpha}y^{-\beta}\right)}{\beta}}$$

(iii) ε Standard extreme value

$$\Pr\{\text{"}yes\text{"}\} \;=\; 1 - \frac{\exp\left(-e^{\alpha-\beta\ln A}\right)}{\exp\left(-e^{\alpha-\beta\ln y}\right)}$$

$$C^+ \;=\; y - e^{e^{\alpha}y^{-\beta}}\int_0^y e^{-e^{\alpha}A^{-\beta}} dA$$

$$C^* \;=\; e^{\frac{\alpha-\ln\left(\ln(2)+e^{\alpha}y^{-\beta}\right)}{\beta}}$$

(2) SPIKE AT ε^{max}

(i) ε Normal

$$\Pr\{\text{"}yes\text{"}\} \;=\; \begin{cases} \Phi(\alpha-\beta\ln A) & if\ A<y \\ 0 & if\ A\ge y \end{cases}$$

$$C^+ = \int_0^y \Phi(\alpha-\beta A)\,dA$$

$$C^* \;=\; \min\left(y, e^{\alpha/\beta}\right)$$

(ii) ε Logistic

$$\Pr\{\text{"}yes\text{"}\} \;=\; \begin{cases} \left[1+e^{-\alpha+\beta\ln A}\right]^{-1} & if\ A<y \\ 0 & if\ A\ge y \end{cases}$$

$$C^+ \;=\; \int_0^y \left(1+e^{-\alpha+\beta\ln A}\right)^{-1} dA$$

$$C^* \;=\; \min\left(y, e^{\alpha/\beta}\right)$$

(iii) ε Extreme value

$$\Pr\{\text{"}yes\text{"}\} \;=\; \begin{cases} 1-\exp\left[-e^{\alpha-\beta\ln A}\right] & if\ A<y \\ 0 & if\ A\ge y \end{cases}$$

$$C^+ \;=\; y - \int_0^y \exp\left[-e^{\alpha-\beta\ln A}\right] dA$$

$$C^* \;=\; \min\left(y, e^{\alpha/\beta}(\ln 2)^{-1/\beta}\right)$$

(B) BOX-COX MODEL $C = y - \left[y^{\lambda} - \frac{\alpha}{b} - \frac{\eta}{b}\right]^{1/\lambda}$

(1) DOUBLE TRUNCATION

(i) η Beta

$$\Pr\{"yes"\} = \int_{by^{\lambda}-b(y-A)^{\lambda}-\alpha}^{by^{\lambda}-\alpha} \left[\frac{(\eta+\alpha)^{\alpha-1}(by^{\lambda}-\alpha-\eta)^{S-1}}{B(r,s)(by^{\lambda})^{r+S-1}}\right] d\eta$$

$$C^{+} = \int_0^y \left[\int_{by^{\lambda}-b(y-A)^{\lambda}}^{by^{\lambda}-\alpha} \left[\frac{(\eta+\alpha)^{\alpha-1}(by^{\lambda}-\alpha-\eta)^{S-1}}{B(r,s)(by^{\lambda})^{r+S-1}}\right] d\eta\right] dA$$

$$C^{*} \text{ solves: } \int_{by^{\lambda}-b(y-C^*)^{\lambda}-\alpha}^{by^{\lambda}-\alpha} \left[\frac{(\eta+\alpha)^{\alpha-1}(by^{\lambda}-\alpha-\eta)^{S-1}}{B(r,s)(by^{\lambda})^{r+S-1}}\right] d\eta = 0.5$$

(ii) η Standard Normal

$$\Pr\{"yes"\} = \frac{\Phi(-\alpha+by^{\lambda}) - \Phi(-\alpha+by^{\lambda}-b(y-A)^{\lambda})}{\Phi(-\alpha+by^{\lambda}) - \Phi(-\alpha)}$$

$$C^{+} = \frac{1}{\Phi(-\alpha+by^{\lambda}) - \Phi(-\alpha)}\left[y\Phi(-\alpha+by^{\lambda}) - \int_0^y \Phi(-\alpha+by^{\lambda}-b(y-A)^{\lambda})dA\right]$$

$$C^{*} \text{ solves: } \Phi(-\alpha+by^{\lambda}-b(y-C^{*})^{\lambda}) = \frac{\Phi(-\alpha+by^{\lambda})+\Phi(-\alpha)}{2}$$

(iii) η Standard Logistic

$$\Pr\{"yes"\} = \frac{[1+e^{\alpha-by^{\lambda}}]^{-1} - [1+e^{\alpha-by^{\lambda}+b(y-A)^{\lambda}}]^{-1}}{[1+e^{\alpha-by^{\lambda}}]^{-1} - [1+e^{\alpha}]^{-1}}$$

$$C^{+} = \frac{1}{[1+e^{\alpha-by^{\lambda}}]^{-1} - [1+e^{\alpha}]^{-1}}\left(y[1+e^{\alpha-by^{\lambda}}]^{-1} - \int_0^y [1+e^{\alpha-by^{\lambda}+b(y-A)^{\lambda}}]^{-1} dA\right)$$

$$C^{*} = y - \left[y^{\lambda} - \frac{\alpha}{b} + \frac{1}{b}\ln\left(\frac{2}{[1+e^{\alpha-by^{\lambda}}]^{-1} + [1+e^{\alpha}]^{-1}} - 1\right)\right]^{1/\lambda}$$

(2) SPIKE AT η^{max}, TRUNCATION AT η^{min}

(i) η Standard Normal

$$\Pr\{"yes"\} = \begin{cases} \dfrac{\Phi(\alpha-by^{\lambda}+b(y-A)^{\lambda})}{\Phi(\alpha)} & \text{if } A < y \\ 0 & \text{if } A \ge y \end{cases}$$

$$C^{+} = \frac{1}{\Phi(\alpha)}\int_0^y \Phi(\alpha-by^{\lambda}+b(y-A)^{\lambda})da$$

$$C^{*} \text{ solves: } \Phi(\alpha-by^{\lambda}+b(y-C^{*})^{\lambda}) = \frac{1}{2}\Phi(\alpha) \quad \text{if } \Phi(\alpha-by^{\lambda}) < \frac{1}{2}$$

$$C* = y, \quad \text{otherwise}$$

(ii) η Standard Logistic

$$\Pr\{"yes"\} = \begin{cases} \dfrac{1+e^{-\alpha}}{1+e^{-\alpha+by^{\lambda}-b(y-A)^{\lambda}}} & \text{if } A < y \\ 0 & A \ge y \end{cases}$$

$$C^{+} = (1+e^{-\alpha})\int_0^y \left(1+e^{-\alpha+by^{\lambda}-b(y-A)^{\lambda}}\right)^{-1} dA$$

$$C^{*} = \min\left(y, y - \left[y^{\lambda} - \frac{\alpha}{b} - \frac{1}{b}\ln(1+2e^{-\alpha})\right]^{1/\lambda}\right)$$

(C) LINEAR MODEL $C = \frac{\alpha + \eta}{\beta}$

(1) DOUBLE TRUNCATION

(i) η Standard Normal

$$\boxed{\Pr\{\text{"}yes\text{"}\} \quad = \quad \frac{\Phi(\alpha - \beta A) - \Phi(\alpha - \beta y)}{\Phi(\alpha) - \Phi(\alpha - \beta y)} \qquad if\ 0 \le A \le y}$$

$$\boxed{C^{+} \quad = \quad \frac{1}{\Phi(\alpha) - \Phi(\alpha - \beta y)}\left[\int_0^y \Phi(\alpha - \beta A)dA - y\Phi(\alpha - \beta y)\right]}$$

$$\boxed{C^{*} \quad \text{solves}: \quad \Phi\left(\alpha - \beta C^{*}\right) = \frac{1}{2}\left[\Phi(\alpha) + \Phi(\alpha + \beta y)\right]}$$

(ii) η Standard Logistic

$$\boxed{\Pr\{\text{"}yes\text{"}\} \quad = \frac{\left[1 + e^{\alpha - \beta y}\right]^{-1} - \left[1 + e^{\alpha - \beta A}\right]^{-1}}{\left[1 + e^{\alpha - \beta y}\right]^{-1} - \left[1 + e^{\alpha}\right]^{-1}} \quad if\, 0 \le A \le y}$$

$$\boxed{C^{+} \quad = \quad \frac{1}{\left[1 + e^{\alpha - \beta y}\right]^{-1} - \left[1 + e^{\alpha}\right]^{-1}}\left(y\left[1 + e^{\alpha - \beta y}\right]^{-1} - \int_0^y \left[1 + e^{\alpha - \beta A}\right]^{-1} dA\right)}$$

$$\boxed{C^{*} \quad = \quad \frac{\alpha}{\beta} - \frac{1}{\beta}\ln\left[\frac{2}{\left[1 + e^{\alpha - \beta y}\right]^{-1} + \left[1 + e^{\alpha}\right]^{-1}} - 1\right]}$$

(2) SPIKE AT η^{max}, TRUNCATION AT η^{min}

(i) η Standard Normal

$$\boxed{\Pr\{\text{"}yes\text{"}\} \quad = \quad \begin{cases} \dfrac{\Phi(\alpha - \beta A)}{\Phi(\alpha)} & if\ 0 \le A < y \\ 0 & if\ A \ge y \end{cases}}$$

$$\boxed{C^{+} \quad = \quad \frac{1}{\Phi(\alpha)}\int_0^y \Phi(\alpha - \beta A)da}$$

$$\boxed{C^{*} \quad \text{solves}: \quad \Phi\left(\alpha - \beta C^{*}\right) = \frac{1}{2}\Phi(\alpha)}$$

(ii) η Standard Logistic

$$\boxed{\Pr\{\text{"}yes\text{"}\} \quad = \quad \begin{cases} \dfrac{1 + e^{-\alpha}}{1 + e^{-\alpha + \beta A}} & if\ 0 \le A < y \\ 0 & if\ A \ge y \end{cases}}$$

$$\boxed{C^{+} \quad = \quad \left(1 + e^{-\alpha}\right)\int_0^y \left[1 + e^{-\alpha + \beta A}\right]^{-1} dA}$$

$$\boxed{C^{*} \quad = \quad \frac{\alpha}{\beta} + \frac{1}{\beta}\ln\left(1 + 2e^{-\alpha}\right)}$$

(D) LOG MODEL $C = y\left[1 - e^{-\left(\frac{\alpha+\eta}{\beta}\right)}\right]$

(1) TRUNCATION AT η^{min}

(i) η Standard Normal

$$\boxed{\Pr\{"yes"\} \quad = \quad \frac{\Phi\left(\alpha + \beta \ln\left(1 - \frac{A}{y}\right)\right)}{\Phi(\alpha)}}$$

$$\boxed{C^{+} \quad = \quad \frac{1}{\Phi(\alpha)} \int_0^y \Phi\left(\alpha + \beta \ln\left(1 - \frac{A}{y}\right)\right) dA}$$

$$\boxed{C^{*} \quad \text{solves:} \quad \Phi\left(\alpha + \beta \ln\left(1 - \frac{C^{*}}{y}\right)\right) = \frac{1}{2}\Phi(\alpha)}$$

(ii) η Standard Logistic

$$\boxed{\Pr\{"yes"\} \quad = \quad \frac{1 + e^{-\alpha}}{1 + e^{-\alpha}\left(1 - \frac{A}{y}\right)^{-\beta}}}$$

$$\boxed{C^{+} \quad = \quad \left(1 + e^{-\alpha}\right)\left(\int_0^y \left[1 + e^{-\alpha}\left(1 - \frac{A}{y}\right)^{-\beta}\right]^{-1} dA\right)}$$

$$\boxed{C^{*} \quad = \quad y\left[1 - \left(2 + e^{\alpha}\right)^{-1/\beta}\right]}$$

APPENDIX TABLE 3. RESPONSE PROBABILITY MODELS SATISFYING C ε [0,y] WITH A SPIKE AT ZERO

(1) MIXTURE MODELS

$\Pr\{"yes"\} = (1-\gamma)\cdot \Pr\{"yes"\}$ where $\Pr\{"yes"\}$ is any of the formulas in Table 2.

(2) CENSORED MODELS

(A) BOX-COX MODEL $C = y - \left[y^\lambda - \frac{\alpha}{b} - \frac{\eta}{b}\right]^{\frac{1}{\lambda}}$

TRUNCATION AT η^{max}, SPIKE AT η^{min}

(i) η Standard Normal

$$\Pr\{"yes"\} = \begin{cases} 1 & if\ A=0 \\ 1-\dfrac{\Phi\left[-\alpha+by^\lambda-b(y-A)^\lambda\right]}{\Phi\left(-\alpha+by^\lambda\right)} & if\ 0<A\le y \end{cases}$$

$$C^+ = y - \frac{1}{\Phi\left(-\alpha+by^\lambda\right)}\int_0^y \Phi\left(-\alpha+by^\lambda-b(y-A)^\lambda\right)dA$$

$$C^* \text{ solves: } 2\Phi\left(-\alpha+by^\lambda-b\left(y-C^*\right)^\lambda\right)=\Phi\left(-\alpha+by^\lambda\right) \quad if\ \ 2\Phi(-\alpha)<\Phi\left(-\alpha+by^\lambda\right)$$
$$C^* = 0 \quad otherwise$$

(ii) η Standard Logistic

$$\Pr\{"yes"\} = \begin{cases} 1 & if\ A=0 \\ 1-\dfrac{1+e^{\alpha-by^\lambda}}{1+e^{\alpha-by^\lambda+b(y-A)^\lambda}} & if\ 0<A\le y \end{cases}$$

$$C^+ = y-\left[1+e^{\alpha-by^\lambda}\right]\int_0^y\left[1+e^{\alpha-by^\lambda+b(y-A)^\lambda}\right]^{-1}dA$$

$$C^* = y-\left(\frac{\ln\left(1+e^{\alpha-by^\lambda}\right)}{b}-\frac{\alpha}{b}+y^\lambda\right)^{1/\lambda} \quad if\ \ e^{-\alpha}+2e^{-by^\lambda}<1$$
$$C^* = 0 \quad \text{otherwise}$$

SPIKE AT η^{max} AND η^{min}

(i) η Standard Normal

$$\Pr\{"yes"\} = \begin{cases} 1 & if\ A=0 \\ \Phi\left(\alpha-by^\lambda+b(y-A)^\lambda\right) & if\ 0<A<y \\ 0 & if\ A\ge y \end{cases}$$

$$C^+ = \int_0^y \Phi\left(\alpha-by^\lambda+b(y-A)^\lambda\right)da$$

$$C^* = \begin{cases} 0 & if\ \Phi(-\alpha)\ge 1/2 \\ y & if\ \Phi\left(-\alpha+by^\lambda\right)\le 1/2 \\ \text{solves: } \Phi\left(-\alpha+by^\lambda-b\left(y-C^*\right)^\lambda\right)=1/2 & otherwise \end{cases}$$

(ii) η Standard Logistic

$$\Pr\{"yes"\} = \begin{cases} 1 & if\ A=0 \\ \left[1+e^{-\alpha+by^\lambda-b(y-A)^\lambda}\right]^{-1} & if\ 0<A<y \\ 0 & if\ A\ge y \end{cases}$$

$$C^+ = \int_0^y\left[1+e^{-\alpha+by^\lambda-b(y-A)^\lambda}\right]^{-1}dA$$

$$C^* = \begin{cases} 0 & \alpha \le 0 \\ 1 & y^{\lambda} \le \alpha/\beta \\ y - \left(y^{\lambda} - \frac{\alpha}{b}\right)^{1/\lambda} & otherwise \end{cases}$$

(B) LINEAR MODEL $C = \dfrac{\alpha+\eta}{\beta}$

TRUNCATION AT η^{max}, SPIKE AT η^{min}

(i) η Standard Normal

$$\Pr\{"yes"\} = \begin{cases} 1 & if\ A = 0 \\ 1 - \dfrac{\Phi(-\alpha+\beta A)}{\Phi(-\alpha+\beta y)} & if\ 0 < A \le y \end{cases}$$

$$C^+ = y - \frac{1}{\Phi(-\alpha+\beta y)}\int_0^y \Phi(-\alpha+\beta A)dA$$

$$C^* \ \text{solves}: \ \Phi\left(-\alpha+\beta C^*\right) = \frac{\Phi(-\alpha+\beta y)}{2} \quad if\ 2\Phi(-\alpha) < \Phi(-\alpha+\beta y)$$

$$C^* = 0 \quad otherwise$$

(ii) η Standard Logistic

$$\Pr\{"yes"\} = \begin{cases} 1 & if\ A = 0 \\ 1 - \dfrac{1+e^{\alpha-\beta y}}{1+e^{\alpha-\beta A}} & if\ 0 < A \le y \end{cases}$$

$$C^+ = y - \left(1+e^{\alpha-\beta y}\right)\int_0^y \left[1+e^{\alpha-\beta A}\right]^{-1} dA$$

$$C^* = \frac{\alpha - \ln\left(2e^{\alpha-\beta y}+1\right)}{\beta} \quad if\ e^{-\alpha} + 2e^{-\beta y} < 1$$

$$C^* = 0 \quad otherwise$$

SPIKE AT η^{max} AND η^{min}

(i) η Standard Normal

$$\Pr\{"yes"\} = \begin{cases} 1 & if\ A = 0 \\ \Phi(\alpha - \beta A) & if\ 0 < A < y \\ 0 & if\ A \ge y \end{cases}$$

$$C^+ = \int_0^y \Phi(\alpha - \beta A)da$$

$$C^* = \begin{cases} 0 & if\ \Phi(\alpha) \le 1/2 \\ y & if\ \Phi(\alpha - \beta y) \ge 1/2 \\ \text{solves}: \Phi\left(\alpha - \beta C^*\right) = 1/2 & otherwise \end{cases}$$

(ii) η Standard Logistic

$$\Pr\{"yes"\} = \begin{cases} 1 & if\ A = 0 \\ \left[1+e^{-\alpha+\beta A}\right]^{-1} & if\ 0 < A < y \\ 0 & if\ A \ge y \end{cases}$$

$$C^+ = \int_0^y \left[1+e^{-\alpha+\beta A}\right]^{-1} dA$$

$$C^* = \begin{cases} 0 & \text{if } \alpha \le 0 \\ y & \text{if } y \le \alpha/\beta \\ \dfrac{\alpha}{\beta} & \text{otherwise} \end{cases}$$

$$C^* = \begin{cases} 0 & \text{if } \alpha \le 0 \\ y\left(1 - e^{-\alpha/\beta}\right) & \text{otherwise} \end{cases}$$

(C) LOG MODEL $C = y\left[1 - e^{-\left(\frac{\alpha+\eta}{\beta}\right)}\right]$

SPIKE AT η^{min}

(i) η Standard Normal

$$\Pr\{\text{"yes"}\} = \begin{cases} 1 & \text{if } A = 0 \\ \Phi\left(\alpha + \beta \ln\left(1 - \frac{A}{y}\right)\right) & \text{if } 0 < A \le y \end{cases}$$

$$C^+ = \int_0^y \Phi\left(\alpha + \beta \ln\left(1 - \frac{A}{y}\right)\right) dA$$

$$C^* = \begin{cases} 0 & \text{if } \Phi(\alpha) \le 1/2 \\ \text{solves: } \Phi\left(\alpha + \beta \ln\left(1 - \frac{C^*}{y}\right)\right) = 1/2 & \text{otherwise} \end{cases}$$

(ii) η Standard Logistic

$$\Pr\{\text{"yes"}\} = \begin{cases} 1 & \text{if } A = 0 \\ \left[1 + e^{-\alpha}\left(1 - \frac{A}{y}\right)^{-\beta}\right]^{-1} & \text{if } 0 < A \le y \end{cases}$$

$$C^+ = \int_0^y \left[1 + e^{-\alpha}\left(1 - \frac{A}{y}\right)^{-\beta}\right]^{-1} dA$$

APPENDIX TABLE 4. RESPONSE PROBABILITY MODELS SATISFYING C ε [-α,y]

(1) MIXTURE MODELS $\Pr\{"yes"\} = \gamma_1[1 - G\_(A)] + \gamma_2 \Pr\{"yes"\}$ where $\Pr\{"yes"\}$ is any of the formulas in Table 2 and G_(x) is a cdf for negative preferences as described in connection with equation (34).

(2) BOX-COX MODEL $C = y - \left[y^\lambda - \frac{\alpha}{b} - \frac{\eta}{b}\right]^{1/\lambda}$

TRUNCATION AT η^{max}

(i) η Standard Normal

$$\Pr\{"yes"\} \quad = \quad 1 - \frac{\Phi\left[-\alpha + by^\lambda - (b-A)^\lambda\right]}{\Phi\left(-\alpha + by^\lambda\right)}$$

(ii) η Standard Logistic

$$\Pr\{"yes"\} \quad = \quad 1 - \frac{\left[1 + e^{\alpha - by^\lambda + b(y-A)^\lambda}\right]^{-1}}{\left[1 + e^{\alpha - by^\lambda}\right]^{-1}}$$

SPIKE AT η^{max}

(i) η Standard Normal

$$\Pr\{"yes"\} = \begin{cases} \Phi\left[\alpha - by^\lambda + b(y-A)^\lambda\right] & if\ A < y \\ 0 & if\ A \geq y \end{cases}$$

(ii) η Standard Logistic

$$\Pr\{"yes"\} = \begin{cases} \left[1 + e^{-\alpha + by^\lambda - b(y-A)^\lambda}\right]^{-1} & if\ A < y \\ 0 & if\ A \geq y \end{cases}$$

(3) LOG MODEL $C \ = \ y\left[1 + e^{-\left(\frac{\alpha+\eta}{\beta}\right)}\right]$

(i) η Standard Normal

$$\Pr\{"yes"\} \quad = \quad \Phi\left(\alpha + \beta \ln\left(1 - \frac{A}{y}\right)\right)$$

(ii) η Standard Logistic

$$\Pr\{"yes"\} \quad = \quad \left[1 + e^{-\alpha}\left(1 - \frac{A}{y}\right)^{-\beta}\right]^{-1}$$

(4) LINEAR MODEL $C \ = \ \frac{\alpha + \eta}{\beta}$

DOUBLE TRUNCATION AT $\eta^{max} = -\alpha + \beta y$ AND $\eta^{min} = -\alpha - \beta y$

(i) η Standard Normal

$$\Pr\{"yes"\} \quad = \quad \frac{\Phi(-\alpha + \beta y) - \Phi(-\alpha + \beta A)}{\Phi(-\alpha + \beta y) - \Phi(-\alpha - \beta y)}$$

(i) η Standard Logistic

$$\Pr\{"yes"\} \quad = \quad \frac{\left[1 + e^{\alpha - \beta y}\right]^{-1} - \left[1 + e^{\alpha - \beta A}\right]^{-1}}{\left[1 + e^{\alpha - \beta y}\right]^{-1} - \left[1 + e^{\alpha + \beta y}\right]^{-1}}$$

DOUBLE SPIKE AT $\eta^{max} = -\alpha + \beta y$ AND $\eta^{min} = -\alpha - \beta y$

(i) η Standard Normal

$$\Pr\{"yes"\} = \begin{cases} 1 & if\ A \leq -y \\ \Phi(\alpha - \beta A) & if\ -y < A < y \\ 0 & if\ A \geq y \end{cases}$$

(ii) η Standard Logistic

$$\Pr\{"yes"\} = \begin{cases} 1 & \text{if } A \le -y \\ \left[1+e^{-\alpha+\beta A}\right]^{-1} & \text{if } -y < A < y \\ 0 & \text{if } A \ge y \end{cases}$$

TRUNCATION AT η^{max}, SPIKE AT η^{min}

(i) η Standard Normal

$$\Pr\{"yes"\} = \begin{cases} 1 & \text{if } A \le -y \\ 1-\dfrac{\Phi(-\alpha+\beta A)}{\Phi(-\alpha+\beta y)} & \text{if } -y < A \le y \end{cases}$$

(ii) η Standard Logistic

$$\Pr\{"yes"\} = \begin{cases} 1 & \text{if } A \le -y \\ 1-\dfrac{1+e^{\alpha-\beta y}}{1+e^{\alpha-\beta A}} & \text{if } -y < A \le y \end{cases}$$

SPIKE AT η^{max}, TRUNCATION AT η^{min}

(i) η Standard Normal

$$\Pr\{"yes"\} = \begin{cases} \dfrac{\Phi(\alpha-\beta A)}{\Phi(\alpha+\beta y)} & \text{if } A < y \\ 0 & \text{if } A \ge y \end{cases}$$

(ii) η Standard Logistic

$$\Pr\{"yes"\} = \begin{cases} \dfrac{1+e^{-\alpha-\beta y}}{1+e^{-\alpha+\beta A}} & \text{if } A < y \\ 0 & \text{if } A \ge y \end{cases}$$