Cost-Benefit Analysis of Environmental Quality Improvement Projects: Uncertain Benefits of Willingness to Pay from Referendum Contingent Valuation

by

Diego J. Rodriguez

Thesis submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

MASTER OF ARTS

in

Economics

APPROVED: ____
Roger N. Waud, Chairman

______________________ Thomas Lutton

______________________ Nancy A. Wentzler

January 2000

Keywords: Cost-benefit analysis, contingent valuation, uncertainty, benefit estimation, natural resources economics.

Cost-Benefit Analysis of Environmental Quality Improvement Projects: Uncertain Benefits of Willingness to Pay from Referendum Contingent Valuation

Diego J. Rodriguez

Abstract

The use of contingent valuation (CV) methods to estimate benefits has become increasingly common in project analysis. Ever since the NOAA Blue Ribbon Panel Report in 1993 (NOAA, 1993) recommended the
use of the referendum form of CV, it seems to have become the method of choice in practical settings.

Referendum-type questions are thought to be easier to answer than the open-ended variety. But there is a downside: econometric techniques must be applied to the referendum data in order to infer the mean or median willingness to pay (WTP) of the sample and, thus, of the population of potential beneficiaries.

This is not, however, just a technical point. Its implications are demonstrated with data obtained from a referendum CV study done for a proposed sewer and wastewater treatment project designed to improve water quality in the Tietê River flowing through the city of São Paulo, Brazil. The results show that:

A factor of 4 separates lowest from highest central tendency estimates of WTP, ignoring one implausible outlier that is 14 times larger than the largest of the other figures.

This variation is ample enough to make a difference in the cost-benefit analysis results for the project under conservative assumptions.

Analysts that use referendum CV data must be sensitive to the problems they buy into, and decide how to deal with the resulting benefits uncertainty in their project analysis. If the principal use of CV survey data is to produce a mean or median estimate of WTP for Cost-Benefit analysis rather than to test for the factors influencing referendum choice responses and, by implication, WTP, nonparametric approaches have the advantage of simplicity over parametric approaches.

Acknowledgments

I wish to thank Thesis Committee Professors Roger Waud, Nancy Wentzler and Thomas Lutton for their guidance and support during the research process. Special thanks to Thomas Lutton for his continuous support throughout the preparation of the thesis.

A special mention goes to Dr. William J. Vaughan, Senior Economist at the Inter-American Development Bank; my friend, my mentor. Without his support this degree would have been difficult to fulfill.

To my family, my wife Marisil and my son Sebastian
In memory of my father

Outline

Section I: Introduction

Cost-benefit (CB) analysis of proposed projects is an inherently uncertain enterprise because it involves the future, which we can never know. Costs and project performance can be different from our expectations. The economy in which the project is embedded may change, and the tastes, incomes and preferences of the population affected by the project may change as well in ways hard to predict. When the proposed project involves environmental public goods, such as improved water quality, another widely recognized source of uncertainty is the behavior of the natural system involved. Completing this familiar list is uncertainty about project benefits, the issue of concern in this paper.

An increasingly respectable and common way of estimating the benefits of environmental projects is to use the so-called contingent valuation (CV) method, which involves directly asking people about their willingness to pay (WTP) for the environmental effects to be provided by the project. Two broad alternative ways of asking the valuation question are available: “open-ended,” in which the respondent can name any amount s/he wishes when asked some version of “What are you WTP?”; and “dichotomous choice” (referendum or yes/no) in which the respondent is asked: “Are you willing to pay (at least) B$ (per period)?”

There is a vast and rapidly growing literature on these methods, especially the problems that arise in creating successful survey instruments, obtaining satisfactory response rates, and interpreting responses. See, for example, the relatively early seminal work by Mitchell and Carson (1989); and the exchange in *Economic Perspectives* occasioned by the huge damage-estimation efforts done by both sides in the Exxon Valdez case, Portney *et al*. (1994).

One ambitious attempt to assess the usefulness of the method and to suggest ground rules for future applications was made by the National Oceanic and Atmospheric Agency (NOAA) in the U.S. in the early 1990s. This effort involved a panel of distinguished economists, and the panel’s report (NOAA 1993) has been very influential since its publication.

In particular, and most relevant to this paper, the panel recommended that CV studies be done using yes/no referendum format questions. This recommendation has been adopted by many practitioners who deal with real-world program or project evaluation. For example, at the Inter-American Development Bank (IDB) contingent valuation has become the method of choice for estimating the benefits of investment projects whose primary objective is improving water quality. Over the past decade the Bank approved 18 projects with sewer provision and/or wastewater treatment components, and 13 of them employed CB analysis whose benefits came at least in part from CV estimates. Most of the stated preference CV surveys used the referendum format (Ardila *et al*. 1998).

The referendum CV approach opens up a new and substantial source of uncertainty in benefit estimation. That source is the choice of econometric technique and subsequent calculation rules used to translate yes/no responses into mean or median WTP numbers. In project analysis this source of uncertainty is easily overlooked; almost none of the projects reviewed by Ardila *et al*. (1998) addressed it. Moreover, most analyses appear to have used an estimation formula that understates benefits.

This paper demonstrates the range alternative central tendency measures for WTP produced under alternative parametric and nonparametric[1] approaches using data gathered from a recent referendum CV survey that was conducted in Brazil to analyze a large, multi-phase water quality improvement project. It explains why one of the most commonly used measures, the unrestricted mean of the conditional inverse distribution function of WTP, may be less desirable and more computationally intensive than simpler alternatives like the nonparametric mean of the marginal inverse distribution function.

The paper is organized as follows. First, the surrounding issues in project analysis of environmental investments at the Inter-American Development Bank (IDB), including the IDB requirements for project economic analysis and a review of current practice are presented. Following, a brief introduction to the CV method is presented with a utility theoretic motivation for the method. The probability models used in the study are presented. Moving closer to reality, alternative central tendency measures are proposed and illustrated using referendum contingent valuation survey data collected for the case study presented. The water quality impacts of the case study project that respondents were asked to value in a referendum CV survey are briefly described. Then, eleven different versions of a central tendency measure of per household benefits from the project data are produced using methods suggested in the literature. Six come from the economist's customary route of econometrically estimating a binary choice model relating the respondent's

[1] In the context of this paper, nonparametric means "distribution-free"; that is, the distribution function of the random variable producing the data need not be specified.

acceptance probability to the bid offered and socioeconomic characteristics, using a Logit specification of the inverse distribution function.[2] The rest are alternatives which either involve nonparametric measures that can be easily obtained without econometrics from the pooled data (i.e. the marginal rather than conditional distribution), or a more complex method that imposes lower and upper bounds on median WTP in econometric estimation. Finally, the effect that uncertainty about "actual" WTP has on the discounted net benefits of the case study project is explored.

[2] The emphasis throughout the paper is on function evaluation to extract a measure of central tendency, not on the prior steps of CV survey design or choice model estimation. Instead, the survey is taken as a given. The specifications of the functional form and arguments in the econometric choice model follow the selections made by the Brazilian consultant who initially analyzed the data.

Section II: Issues in Project Analysis of Environmental Investments at the Inter-American Development Bank (IDB)

Bank Requirements for Project Economic Analysis

The IDB's Operational Policy of 1980 requires, when feasible, a cost-benefit analysis of project feasibility, where all inputs and outputs must be valued in opportunity cost or economic efficiency terms. Also, it states that feasible projects must have an economic internal rate of return of at least 12 percent, or a positive net present value using a 12 percent discount rate, and also concedes the use of cost-effectiveness analysis where the value of benefits cannot be reliably estimated (Ardila, Quiroga, and Vaughan 1998).

Review of Actual Practice at the IDB

When economists analyze investment projects, and when benefits can be measured monetarily, Cost-Benefit (CB) Analysis is the method of preference. When benefits cannot be measured accurately, Cost-Effectiveness (CE) Analysis is used to compare costs of alternatives of reaching a specific target and identify the option that costs the least (Vaughan and Ardila 1993). Although CE analysis is a acceptable alternative, economists tend to favor CBA for analyzing standard investment projects.

The methods and techniques used at the IDB to estimate benefits for environmental quality improvement projects have varied. In the 1980s, very few projects in environmental quality improvement or natural resources management were approved, with the exception of a handful of projects in urban sanitation and urban housing. When benefits were estimated, the most accepted method was the hedonic markets methods[3]. Economists were unfamiliar with the use of contingent valuation and in rare occasions used travel cost methods (Vaughan and Ardila 1993).

---

[3]The hedonic markets methods use regression analysis of the relation between property values and the attributes of the property, the local environment, and the neighborhood to calculate the marginal value of a change in these attributes (Vaughan and Ardila 1993).

In the 1990s the situation changed dramatically. The number of projects for environmental quality improvement or natural resources management increased substantially; technical expertise became more versed in contingent valuation methods, gradually replacing the use of the hedonic market and travel cost methods. Part of the reason for this shift in methods was attributed to the tendency to overestimate benefits with the manner in which the hedonic method was applied in the IDB where economists never tried to estimate the marginal value functions for attributes (Vaughan and Ardila 1993). In lieu, the hedonic price function was used to produce an upper bound estimate of benefits of an attributed change (Vaughan and Ardila 1993). Additionally, hedonic market methods are data intensive; the lack of appropriate data sets is a major issue in the developing countries. In projects addressing air pollution, water pollution, and solid waste disposal, sources of waste discharge should be inventoried. Proper measurements of ambient conditions need to be made in order to determine how often air and water quality standards are not being met. Then, the effects of the proposed investments on ambient conditions should be established which means determining the changes in ambient quality indicators across time and space that the project will bring about.

In one ex-post evaluation exercise, the IDB compared hedonic and contingent valuation estimates of benefits of supplying alternative housing solutions to people living in precarious conditions (see Table 1). The results were consistent with the belief that hedonic method provided an upper bound estimate of benefits.

Table 1. Comparison of Hedonic and Contingent Valuation Methods (CVM)

		Hedonic	**CVM**
Basic 2 story home 90 m^2 living space	Net Present Value Benefit (pesos)	2,129,037	1,769,500
Minimum 1 story home 30 m^2 living space	Net Present Value Benefit (pesos)	909,255	1,039,780
Self constructed cement home 72 m^2 living space	Net Present Value Benefit (pesos)	1,505,844	1,205,150
Overall Project Rate of Return	Percent	5.2	-5.2

Source: Vaughan and Ardila 1993.

In 1998, a desk review of 27[4] projects approved since 1989 that had ambient water quality improvement as an objective was undertaken. The projects were identified by searching the IDB's project data base, by interviewing staff involved in the preparation of the projects, and by reviewing project documents.

The study concluded that the preferred benefit estimation approach for sewer and treatment components was contingent valuation; 14 out of 18 projects. Each one of the projects from the selected sample, presented the expected willingness to pay estimates as a single number and consequently, a single net present value for the entire operation. Nowadays, CV is widely used and accepted at the IDB to estimate the benefits of environmental quality improvement operations.

[4] 18 were specific investment loans and 9 were multiple works loans. Global multiple works involve numerous investments in several cities and only a representative sample of investments is analyzed a priori. Global multiple works documents do not contain many useful information and were, therefore, omitted from the study.

Section III: Direct Benefit Revelation Theory: Non-Marketed Environmental Benefits

Valuing Public Goods: The Hypothetical (or Contingent) Market Experimental Technique

The Contingent Valuation Method (CVM) has been used since the early 1960s. Robert Davis used questionnaire/interviews of 121 hunters and recreationists in the Maine area to estimate the benefits of outdoor recreation (Mitchell and Carson 1989). In the late 1960s, Ronald Ridker used the CVM in Philadelphia and Syracuse to estimate air pollution benefits. In the 1970s, CVM was used to value various recreational amenities. Ciccheti and Smith used it to estimate willingness-to-pay to reduce congestion in a hiking area; Arthur Darling used the CVM to value amenities of three urban parks in California; Alan Randall used CVM to estimate air visibility benefits in the Four Corners area in the southwest; Acton applied the method to valuing programs which reduced the risk of dying from a heart attack (Mitchell and Carson 1989). By the late 1970s, CVM became a recommended method for determining project benefits for empirical resources such as environmental amenities. In 1979, the Water Resources Council recommended CVM in its "Principles and Standards for Water and Related Land Resources Planning". In early 1980s, the U.S. Army Corps of Engineers began to use CVM to measure project benefits (Mitchell and Carson 1989) and CVM was also recognized under the Comprehensive Environmental Response, Compensation, and Liability Act of 1980 (Mitchell and Carson 1989).

CVM is a stated preference technique used to measure willingness to pay for certain nonmarket commodities. Since it may include nonuse values, it may be one’s only alternative method for estimating use and nonuse values, especially for goods not traded in markets. Furthermore, the CVM is the only technique that measures Hicksian surplus directly, without requiring additional manipulation.

Under CVM a hypothetical scenario is created and individuals are surveyed through phone, mail or personal interview methods. To get a monetary measure of a scenario involving welfare change, individuals are asked the amount they would be willing to pay for the good (the welfare improvement) in question.

CVM assumes that individuals respond the same way to a hypothetical situation as they do to a real scenario. If so, their willingness to pay will be a realistic money measure of the individuals' worth or utility gained or lost from changes in the availability or quality of non-marketed environmental goods. CVM is been widely used in environmental and natural resource economics and it is the most popular method for applied research (Hazilla 1997).

Contingent valuation approaches to project benefit estimation necessarily involve surveying samples of the population of interest. If the sample is representative of the population, the sample mean of willingness to pay per capita (or per household) can simply be attributed to everyone in the beneficiary population of size N, so total project benefits are obtained as the product of N and per capita WTP.

There are three basic survey designs: (a) open-ended surveys where individuals are asked to state a willingness to pay; (b) closed-end referendums where individuals are presented with a bid and respond yes/no binary decision (single-response referendum); and (c) closed-end double referendums where individuals are presented with a sequence of two payments to obtain binary decisions (double-response referendum).

In the early years of CV, the method of payment elicitation was open ended. People were asked to reveal the specific monetary amount they would be willing to sacrifice for the provision of a non-marketed good such as an improvement in ambient environmental quality. Obtaining a measure of central tendency from this kind of data was as simple as calculating the mean or median of the WTP values provided by the survey respondents. The econometric analysis involved was minimal, usually being confined to plausibility checks undertaken by split sample comparisons or by regressing the payment amounts on income and other socioeconomic variables to see if the signs on the parameter estimates in the relationship were consistent with prior expectations (e.g. WTP increasing with income).

Issues with Referendum Models

All of this changed with the advent of the referendum format, which only asks if the respondent would or would not be willing to pay a specific pre-selected amount. Under this format it is not possible to know the true WTP of any individual directly. Because those who answer in the affirmative might actually be willing to pay even more, and those who answer in the negative might be willing to pay something less, econometric techniques have to be brought to bear to somehow interpolate and infer an expected value or other central tendency measure from the dichotomous choice information. Simplicity of data analysis was sacrificed in the referendum method in order to construct what many felt was a more realistic choice game.

In consequence, the notion that contingent valuation experiments of the referendum type can reveal a unique number which accurately and unambiguously represents individual willingness to pay for water quality improvement is unrealistic. Rather, there are several possible numbers, each dependent upon the way the initial survey was designed and administered and the way the resulting raw data was passed through the summarizing econometric sieve and reconstituted in the form of a central tendency measure. In short, such estimates are always uncertain when we acknowledge the existence of many routes that potentially can be taken to get at them and the several decision alternatives present at each step along the way. This is not a counsel of doom, or a suggestion that CB analysis based on referendum CV not be undertaken. But it is a fact that any benefit estimate to a greater or lesser degree is always a product of the analyst's protocol and judgement, something respectable analysts recognize and communicate to the users of their results.

The potential for a negative estimate of the expected value of willingness to pay is only a special example of a more general issue with referendum CV, which is that the willingness to pay value extracted from the data can be heavily influenced by the methodological approach taken.

There are basically two routes to analyzing referendum data. The one most frequently pursued by project economists involves several steps, beginning with the specification and statistical estimation of one or more probability models of individual choice, employing prior assumptions about the form of the inverse distribution, and the covariates belonging in the distribution which serve to change its location and shape across respondents. This is followed by the evaluation of conditional mean or median formulas derived from the choice model, which depend on its estimated parameters. After calculating individual-specific means or medians, averages are taken over the entire sample to produce global central tendency measures. A less frequently traveled but much easier route ignores covariates and does not specify any particular inverse distribution. Instead it uses all the data in pooled form (i.e. the marginal distribution) to produce nonparametric measures of central tendency. Given its prominence, the next two sections concentrate on the parametric route, followed by a discussion of nonparametric options.

The parametric route can quickly become quite complex, producing a wide array of central tendency estimates. It is not uncommon to find instances where predicted WTP can vary from low to high by a factor of two, five or ten with the same data, depending on the analyst's choice of density function, the specification of the functional form of the indirect utility index and its arguments, and whether a mean, a truncated mean, or a median is used. In short, with referendum data there are a host of possible measures of central tendency of willingness to pay. Gauged by their frequency of use by practitioners, all of them might seem equally legitimate, but this is not a useful criterion. For instance, the untruncated mean extracted from Logit estimation of a random utility model (see Table 2 below) has been one of the most popular measures used in IDB project analysis and in the literature more generally, even though it is potentially vulnerable to the problem of negative WTP.

The Models

In order to estimate WTP we have to estimate the probability of accepting or rejecting the offered price as a function of the price itself and some socioeconomic variables that shift the indirect utility function ($\triangle h$ below). We assume that the probability function follows a logistic distribution. Therefore, the method of maximum likelihood[5] via a binary choice Logit model will be developed.

The Logit Model

Consider the following logistic representation of accepting the obligation to pay a price for an environmental quality improvement:

[5]The method of maximum likelihood involve estimating those parameter values that most likely yield the observed data set. We begin by defining a likelihood function, which is the product of the PDF evaluated at each of the observation values (Morgan and Henrion 1992). By taking the first derivative of the log likelihood function with respect to each one of the parameters and setting them equal to zero we obtain the values of those parameters that maximize the function.

(1) $$P_k = E(Y = 1|X_k) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+.....+\beta_k X_k)}}$$

where Y=1 if a "yes" response and 0 otherwise and X_k = socioeconomic variables.

Rewritten for simplification:

(2) $$P_k = \frac{1}{1+e^{-G_k}}$$

where $G_k = \beta_0 + \beta_1 X_1 + ... + \beta_k X_k$

Equation (2) is known as the **logistic distribution function**. We can see that as G_k ranges from - ∞ to + ∞, P_k ranges between 0 and 1. Furthermore, we can also observe that P_k is nonlinearly related to G_k.

Now, P_k is the probability of accepting the price (equation (2), so (1 - P_k) is the probability of rejecting the price.

That is:

(3) $$1 - P_k = \frac{1}{1+e^{G_k}}$$

If the data are grouped (i.e. instead of observing 0, 1 for each observation we observe the aggregate percent acceptance at each price:

(4) $$\frac{P_k}{1 - P_k} = \frac{1 + e^{G_k}}{1 + e^{-G_k}} = e^{G_k}$$

The expression on the left is simply the odds ratio in favor of accepting the price. We can take the natural log of the expression and get:

(5) $$L_i = \ln\left(\frac{P_k}{1 - P_k}\right) = G_k = \beta_0 + \beta_1 X_1 + \ldots\ldots + \beta_k X_k$$

From this expression we can see that L_i, the log of the odds ratio is linear in X and linear in the parameters. **L** is called the **logit model** for grouped data (Gujarati 1995). Since our data are not grouped, we estimate (2) by maximum likelihood using the LIMDEP econometrics package.

Using the survey findings, an econometric model will be developed to estimate the likelihood that an individual will or will not agree to pay the stated amount to secure the benefits of the program. Based on the estimates, an average willingness-to-pay (WTP) per month per family will be calculated.

The Probit Model

As we've seen, the Logit model uses the cumulative logistic function. In some instances, the normal cumulative distribution function has been found useful. The estimating model emerging from this normal function is the Probit Model.

Consider the same function as used in the preceding example:

$G_k = \beta_0 + \beta_1 X_1 + \ldots.. + \beta_k X_k$

Given the assumption of normality and following same notation as above, we have the Probit model:

(6) $$P_k = E(Y=1|X_k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{G_k} e^{-\frac{t^2}{2}}\, dt$$

where t is a standardized normal random variable, i.e., $t \sim N(0,1)$. In this probability model, the probability P_k that Y=1 lies between zero and one, since it is the probability that the standardized normal random variable t is less or equal to $G_k = \beta_0 + \beta_1 X_1 + ... + \beta_k X_k$. As the utility index G_k increases from $-\infty$ to $+\infty$, the probability P_k that Y=1 increases monotonically (Griffiths *et al*. 1993). Since P_k represents the probability of a "yes" response, it is measured by the area of the standard normal curve from $-\infty$ to G^k.

Logit or Probit?

Both models are very close to each other except at the tails, with the logistic (logit model) having slightly flatter tails. Therefore, the choice of one model over the other is a matter of mathematical convenience and availability of computer programs.

Specification Issues

In order to estimate a Logit or Probit model, several specifications have to be made:

1) Specifications of the arguments of the indirect utility functions. In other words, the explanatory variables that belong in the model must be chosen from the survey.
2) Specifications of the functional form of the indirect utility function (linear, power, etc.).
3) Specifications of the cumulative density function (linear, logistic).
4) Specification of the limits of integration of the function of interest to derive the expected value of the willingness-to-pay.

Utility-Theoretic Motivation

Consider an individual who must decide whether to answer yes or no to the following: *Would you vote for a program to increase environmental quality from q^0 to q^1 if it would decrease your annual income by $B?*

Let the indirect utility function be $u(Y,q,X)$ where **Y** represents income, **X** is a vector of individual characteristics and the vector of market prices **P** is omitted since prices are assumed to be constant.

The individual responds yes if:

(7) $$u(Y\text{-}B, q^1, X) - u(Y, q^0, X) \geq 0$$

and no otherwise.

Let $h(\cdot)$ be the observable component of utility. Here h represents and indirect utility function which in statistical estimation is often called the index function or utility index, denoted as the summed product of the parameter estimates and the explanatory variables, $X\beta$ (Greene 1990, p. 673). The probability of a "yes" response is given by:

(8) $$P_1 = P[h(Y - B, q^1, X) + \epsilon_1 > h(Y, q^0, X) + \epsilon_0]$$

Where ϵ_i(i= 0,1) are independent, identically distributed random variables with zero means and the error term represents influences on utility not observed by the analyst, or just random error in the choice process itself. Assuming the error difference follows a Logistic distribution, the probability of a "yes" response can be expressed as an estimable random utility (difference) model, or RUM:

(9) $$P_1 = e^{\Delta h} / (1+ e^{\Delta h}) = (1 + e^{-\Delta h})^{-1}$$

Where $\Delta h = h^1 - h^0$ and h^0 represents the initial indirect utility function and h^1 is the indirect utility function reflecting the decrease in Y by B and the increase in environmental quality from q^0 to q^1. The linear utility difference index Δh in the "no income effects" RUM is usually specified as a function of the bid level, B, and a set of socioeconomic variables, S, including a constant term $(\alpha_1\text{-}\alpha_0)$ but not including income as an argument (i.e.$\Delta h=(\alpha_1\text{-}\alpha_0)+ \beta B +\zeta S$). This most basic of specifications imposes the assumption of a constant marginal utility of income, which simplifies recovery of an expected value for WTP.

By reversing the sign on the probability difference, we get the expression for the probability of rejecting the offer:

(10) $$P_0 = (1 + e^{\Delta h})^{-1}$$

We define the Willingness-to-pay for q^1 (WTP) by the amount of money that must be taken away from the individual enjoying an improved amenity level, q^1, that leaves he/she as well off as the initial amenity and income situation.

(11) $$u(Y\text{-}WTP, q^1) = u(Y, q^0)$$

and

(12) $$h(Y - WTP, q^1) + \epsilon_1 - \epsilon_0 = h(Y, q^0)$$

Because of the term $\epsilon_1 - \epsilon_0$, WTP is a random variable. Then, the probability of accepting the offer is also the probability that WTP $\geq$ B, and the probability of rejecting the offer is the probability that WTP < B. This is a cumulative distribution function and can be denoted as F(WTP). As pointed out by Hanemann (1984), the truncated expected value of the random variable (WTP) can be found from the cumulative density function as follows:

(13) $$E[WTP] = \int_0^\infty [1 - F(WTP)] dWTP$$

Here, the integration is only over positive values of WTP, because if there is utility improvement, WTP theoretically cannot be negative (although it can depend on who you ask and how the question is phrased). Similarly, the untruncated expected value of the random variable (WTP) can be found from the cumulative density function:

(14) $$E[WTP] = \int_0^\infty [1 - F(WTP)] dWTP - \int_{-\infty}^0 [F(WTP)] dWTP$$

The latter, treating the negative domain of WTP as admissible, will generally be less than or equal to the truncated WTP represented by the first term in the above expression (Johansson *et al.* 1989).

For the Logit probability model, Hanemann (1984) and Ardila (1993) provide the WTP formulas shown in Table 2 for the unrestricted expected value, the median, and the truncated expected value that restricts WTP to be positive.

The α term in the table is shorthand for an augmented intercept absorbing the estimated constant and the socioeconomic variable influences on $\triangle h$ (α below equals $(\alpha_1-\alpha_0)+\zeta S$). The letter C in the table is shorthand for the central tendency measure of WTP, following the notation of Hanneman (1984), the original source. In models with several explanatory variables, the parameter α can be replaced by an augmented intercept, using the coefficient estimates evaluated at the means of the independent variables, except of course, the bid price, β.[6]

Table 2. Formulae for Central Tendencies from the Probability Model

Description	Symbol	Equation
Mean, E(WTP), $-\infty <$ WTP $< \infty$	C+	α/β
Median WTP	C*	α/β
Truncated Mean, E(WTP), $0 <$ WTP $< \infty$	C’	$\ln(1+\exp(\alpha))/\beta$
Truncated Mean, E(WTP), $0<$WTP$<B_{max}$ where B_{max} is the maximum bid	$C^{\sim}$	$1/\beta \ln[(1+\exp(\alpha))/(1+\exp(\alpha-\beta B_{max}))]$
Truncated Mean, Log Transform, $E(\exp^{\ln(WTP)})$, $-\infty < \ln$WTP $<\infty$ (utility difference logit, log of bid, 0 Lower Limit, No Upper Limit)	$C^{+}_{\ln}$	$\exp(-\alpha/\beta)\ [(\pi/\beta)/(\sin(\pi/\beta))]$ (Only applies if $0 <1/\beta<1$, otherwise numerical approximation required)
Truncated Mean, Log Transform, $E(\exp^{\ln(WTP)})$, $-\infty < \ln$WTP $<\ln$ Income (utility difference logit, log of bid, 0 Lower Limit, Income Upper Limit)	$C^{\sim}_{\ln}$	No Analytic Expression–Requires Numerical Approximation

[6] The augmented intercept, α, referred to in Table 2 is simply the original intercept (for purposes of this note call it β_0) plus the rest of the i=1...n-1 parameter estimates other than the bid parameter estimate multiplied by the respective sample means of the explanatory variables $\overline{X}_i$. The β attached to bid in Table 2 is, in this notation, equivalent to β_n.

Table 2. Formulae for Central Tendencies from the Probability Model		
Description	**Symbol**	**Equation**
Truncated Median, Log Transform	C^*_{ln}	$\exp(\alpha/\beta)$

Model Assumptions

Given the theoretical underpinnings of the conventional random utility model (RUM) sketched above, it is necessary to recognize that when the RUM is specified as a Logit model with a linear utility difference index specification, a fundamental contradiction arises because the Logit potentially allows predicted willingness to pay to fall between minus and plus infinity, admitting the possibility of negative values. Negative WTP should be ruled out for well conceived environmental improvements, as should expected payments exceeding actual income.[7] The expedients for guaranteeing satisfaction of one or both of these limits by evaluating the linear utility index model estimated with Logit or Probit from zero bid to either plus infinity or income (truncated means), or by forcing the estimated density to lie in the positive region by using the logarithm of bid rather than the untransformed bid in estimation, leave a great deal to be desired. They are just ad-hoc fixes to the conventional random utility model's fundamental specification error of an unrestricted error term.

Although it was originally discussed in the late 1980s (Johansson, Kriström and Mäler, 1989; Hannemann 1989) the issue has recently been brought more fully to light by Haab and McConnell (May 1998). The latter suggest employing a beta distribution for the density of willingness to pay to consistently hold WTP between zero and some upper bound such as income. In an unpublished study Haab and McConnell (August 1997, January 1999) have proposed an alternative way to achieve a similar restriction by bounded Probit (or Logit) estimation. Because this method is much simpler to implement than the beta, it is applied to the Tietê project

[7] This is strictly true only if the answer supplied reflects an understanding that payments for the good offered are to be taken out of current income without drawing down savings or liquidating other forms of wealth. It is unlikely that low income survey respondents (who usually dominate CV surveys taken in developing countries), would either have assets to pledge or be willing to pledge them in excess of current income when valuing a non-unique environmental good like water quality improvement. However, the preservation of unique natural assets may evoke contributions in excess of income, especially among the upper strata, and especially if the question is posed as a one-time payment rather than a series of payments strung out over several years.

referendum survey data, where it produces reasonable estimates for the median, but curious estimates for the mean.

Section IV: The Case Study Context: A Water Quality Improvement Program in Brazil

The Tietê River Pollution Problem: Description of the Project

The state of Sao Paulo in Brazil occupies 240,000 km^2 (2.9% of Brazil's area) and has a population of 33 million. Its industrial park is one of the most important in Latin America, generating almost 30% of Brazil's Gross Domestic Product. The Sao Paulo Metropolitan Area (SPMA) occupies 8,000 km^2 and has a population of 16 million (11.2% of the country's total).

The Tietê River originates just 95 km east of the city, picks up its pollution load upon passing through it, and flows for another 1095 km before joining the Parana River. The majority of the municipalities in the area are located in the watershed of the Tietê River (upper Tietê) and its main tributaries: the Pinheiros, Tamanduatei, and Juqueri. The industrial center of Cubatao, which generates the highest atmospheric and water pollution in Brazil, is located in Sao Paulo State.

The parts of the Tietê River and its tributaries flowing through the SPMA are the most polluted bodies of water in the State. The Tietê enters the metropolitan area with acceptable water quality characteristics but in Guarulhos, at the confluence of the Jacu river, it becomes anaerobic (see Map 1 below). From the Jacu downstream the large volume of untreated domestic and industrial waste dumped into the relatively small volume of river flow has made the river an open sewer that supports no aquatic life, smells most of the year, and is used only as a sewer canal for more than 80 kilometers.

The city of São Paulo has developed around the Tietê in a way that adjusts for the river's extreme pollution. On either side of the river, the Paulistas have built large expressways, which impede access. Land adjacent to the expressways is used predominantly for industry or commercial storage and wholesale activities. Land use has adjusted, but the problem remains. Surveys indicate that people who drive the expressways and work in the areas are aware of the stench of the river. Sections of the expressways frequently flood in rainy season exposing people to health risks. The water is too contaminated even for industrial use.

The Edgard de Souza dam, located about 20 kilometers downstream from the confluence of the Tietê and the Pinheiros rivers, permits the diversion of the Tietê into the Pinheiros. Between 1930 and 1991 the elevating plants of Traição and Pedreira pumped Tietê/Pinheiros water to the Billings Reservoir where it was used by the 887 MW Henry Borden hydroelectric plant. Because of the increasing deterioration of the quality of the Tietê, the 1989 Constitution of the State of São Paulo required state and municipal authorities "to take effective measures" to stop the pumping of waste waters, and other polluting substances to Billings reservoir. The pollution is costing the power generating company, EMAE, about US$66 million in foregone revenue.

The tributaries of the Tietê in the metropolitan area and the Tiête itself receive waste well beyond the river's natural processing capacity. At present the organic load is predominantly from households (360 tons per day, 80% of the total). Surface runoff accounts for 62 tons per day and industry contributes 30 tons per day. The problem is severe all year long and becomes critical in the dry season.

Domestic Contamination

The Tietê River and its principal tributaries are fed by a number of smaller tributaries in the SPMA. These tributaries lie in the service areas of four sanitation companies. Three of these companies (Saneamento Basico do São Paulo--SABESP, the municipality of Guarulhos, and ABC are upstream from the critical areas of the river affected by the project), the fourth, the municipality of Osasco, is downstream.

Table 3 presents the data on the percentage of houses in each of these four areas with water and sewerage services supplied by public sanitation companies. Table 3 shows that SABESP treats more of the sewage it collects (61%) than any of the other three companies operating in the SPMA. ABC has higher coverage with sewerage services (85% vs. 79%) but a much lower percentage treated 48%. Guarulhos has the worst indicators with 70% sewerage coverage and no treatment at all. Thus it is hardly surprising that the Tietê becomes anaerobic when it flow through Guarulhos.

Table 3. Percentage of Households in the SPMA Served by Publicly Supplied Water and Sewerage Service in 1998			
Company	**Water Service Coverage**	**Sewer Service Coverage**	**Sewage Treated**
SABESP	98%	79%	61%
Guarulhos	95%	70%	0 %
ABC	100%	85%	48%
Osasco	98%	70%	24%

The SPMA still has 89 collectors serving sewer networks that discharge untreated sewerage directly into the tributaries of the Tietê. In addition, there are a large --but unknown-- number of households with sewers connected to the storm sewer system also discharging raw sewerage in the Tietê River and its tributaries.

Surface Runoff

A second source of contamination comes from material washed by rainfall from the streets and land into the storm sewer system or directly into the streams and rivers. Much of this contamination comes from households that are not connected to the sanitary sewer system. Contamination also comes from solid wastes that have been thrown in streets, gullies, or streams and from organic matter (such as fallen leaves) that is washed to the storm sewers or local streams.

The households not connected to the sewer system have a separate and important problem of contamination of the local environment. These conditions lead to localized health problems.

Industrial Contamination

During the first stage of the Tietê project (see below) completed in 1998 the Companhia de Tecnologia de Saneamento Ambiental (CETESB) focused on controlling the discharges of the 1,250 most important polluters. It succeeded in reducing the organic discharges of industry by 59% and the inorganic discharges by 74%. Many of these industries now pre-treat their effluents before they are discharged to the Tietê, but these pre-treated effluents still go into the river system. In addition, there are an additional 3,486 industries with significant pollution potential that are discharging untreated wastes into the river. Thus, each day roughly 150 tons of biochemical oxygen demand (BOD) and 1.5 tons of inorganic load per day are still being discharged by industry into the Tietê.

Pollution Control in the Multistage Program

The proposed project for cleaning up the Tietê River involves extension of sewers to currently unsewered households (and businesses) and the provision of wastewater treatment plants at the discharge ends of those sewers. The major objective is the removal of oxygen-demanding organic materials (measured as BOD) and safe disposal of sewage sludge.

The problem of contamination of the Tietê is enormous. The solution is expensive and will take many years to achieve. SABESP and the Bank agreed to divide the project in three stages, taking the technical and financial resources of SABESP into consideration. The stages are not independent; all are needed to attain any benefits.

Stage I (1993-1998)

The main objectives of the first stage, which has been completed, were to: (i) enhance the quality of life for the population of the SPMA; (ii) improve health and environmental conditions in the area; (iii) reduce the pollution of the Tietê River and its main tributaries; (iv) study the use of the water resources and formulate subsequent stages of the project; (v) strengthen the legal and institutional structure of the state of Sao Paulo for control of industrial waste; and (vi) train technical and administrative staff to operate and maintain the wastewater treatment plants.

In addition to the wastewater treatment plants of direct concern here, the project also provided for sewer construction. The treatment component involved the construction of two new wastewater treatment plants and the expansion of an existing plant; increasing the proportion of wastewater treated from 19% in 1992 to 45% by 1998. Specifically, the works were:

❶ Sao Miguel Plant: construction of the first module using the activated sludge treatment process with digestion by anaerobic bacteria. This plant will treat 40% of the flow from industries located in the area and serve a population of approximately 720,000.

❷ Parque Novo Mundo Plant: construction of an initial module using activated sludge treatment. The sludge produced will be chemically stabilized and primary sedimentation omitted. The plant will serve a population of 1.2 million and it will treat 14% of the flow from industries located in the area.

❸ Barueri Plant: expansion of the number of secondary sedimentation units. The plant will serve an additional 1.2 million persons and 14% of the total flow will be from industries located in the area.

Stage I of the project removes about 25% of organic material of domestic and industrial origin discharged into the Tietê River, and similar amounts of other pollutants such as inorganic material, toxic compounds, and fecal coliforms. BOD5 concentrations in the most critical (worst) reach should fall from a "without project" level of 86 mg/l to 40 mg/l. However, despite the BOD reductions, dissolved oxygen (DO) recovery is limited, since absolute BOD levels are still too high (well over the 5 mg/l of BOD defining a "clean" river). Increases in DO between 0.5 to 1.0 mg/l would only occur just before and after the long anaerobic stretch, which Stage I shrinks from 100 km to 75 km. Odor reduction is the major beneficial water quality effect of Stage I, but it still leaves DO at levels that are too low to support aquatic life.

Cost-Benefit (CB) analysis was only undertaken for the sewer connection component (including costs for sewers but not treatment plants), presumably because the benefits of Stage I alone were negligible. To choose the treatment plant capacities, locations and construction timing, a Regional Least-Cost Mixed Integer Programming model was used to minimize the sum of treatment plant investment, operation and maintenance costs, allowing construction to begin in either of two time periods subject to plant flow capacity constraints.

Stages II & III (1999-2008)

The main objectives of these subsequent stages is to continue supporting the State of Sao Paulo in its efforts to improve the ambient environmental quality of the Tietê Basin and use the State's water resources efficiently. The water-quality improvement component will include additional collection of wastewater and extension of sewers to currently unsewered households and businesses, along with some treatment plant capacity expansion. Works will be prioritized based on the results of a water quality model developed in Stage I. Interceptors will be built along the margins of the Rio Pinheiros. The improvement in water quality is expected to increase the use of water for hydroelectric generation at the Henry Borden power plant.

Cost Effectiveness Analysis

To select the collection networks that most effectively reduced contamination, SABESP ranked 91 networks by the ratio of investment and operating costs to load collected. The ratios ranged from R$0.12 to R$7.34 per cubic meter per year. Sixty-nine of these systems had ratios under R$1.00. Twenty-six of the 69 (29%) belong to Guarulhos and Osasco and have not been included in the second stage project investments. Twelve of these account for a significant amount of waste. This is a potential problem since Guarulhos and Osasco are not included in this project.

SABESP also analyzed 89 collection systems not connected to interceptors and ranked them by the ratio of investment and operating cost to organic load diverted. The ratios ranged from R$0.08 to R$19.83 per cubic meter per year. Fifty-six of these have ratios less than R$ 1.00 and eighteen of these are in Guarulhos and Osasco, and have been left for the third stage. Again, this is a potential problem for achieving the improvement of water quality. SABESP analyzed the impact of these works and the proposed treatment plants using a water quality simulation model.

Modeling the Impact of Contamination and Its Reduction on Water Quality

To simulate the impact of various discharges on water quality in the Tietê Basin, SABESP has adapted the QUAL2E Stream Water Quality Model provided by the US Environmental Protection Agency. The model is deterministic and relatively simple. Hydrological variations are determined outside the quality model and the quality is calculated on the basis of a particular river flow and the contaminant loads that enter. The model separately accounts for contaminant loads coming in above the SPMA, point discharges of industrial and sewerage outfalls, contaminant loads from tributaries, non-point discharges from surface runoff, and river reflows from underground lenses.

The model considers the principal mechanisms of transport, advection and dispersion in the direction of the flow of the river (but not horizontally across the river). It accommodates 15 water quality indicators, the most important of which--from the point of view of the economic analysis--is dissolved oxygen. The basic technical relations that determine the level of dissolved oxygen are: atmospheric reaeration (which is calculated as a function of the velocity and depth of the river), plant evapo-transpiration, benthonic and biochemical nitrification, and temperature.

The model divides the rivers into sections. Each section is characterized in hydrological terms by volume of flow entering the section, entering or departing lateral flows, and flows exiting the section. The amount of any particular quality characteristic being traced can be described in terms of advective or dispersive transport and can be increased or decreased along subsections by external inflows or biochemical processes.

For quality calculations, SABESP uses the average flow and the "minimum" flow. SABESP has 60 years of data on daily hydrological flows[8]. The "average flow" is the mathematical average of daily flows throughout the year. The actual flow is less than the average flow 60% of the time. SABESP defines "minimum flow" as a flow that will be exceeded 90% of the time or, in more intuitive terms, the river flow will exceed this volume 329 days a year and have a flow of smaller volume 36 days a year.

---

[8] This data was collected on natural river flows between 1900 and 1960. After 1960, the river was regulated and the data series discontinued.

Table 4. Quality of Water in the Tiete Basin at Minimum Flow after the Conclusion of Each of the Project Stages						
	Operation to Carry Away Wastes			**Operation to Carry Away Wastes and Generate**		
Point at Which Quality is Measured	Dissolved Oxygen mg/l	Biological Oxygen Demand mg/l	Fecal Coliforms no/100 ml	Dissolved Oxygen Mg/l	Biological Oxygen Demand mg/l	Fecal Coliforms no/100 ml
Tietê confluence Tamanduatei						
1998	0.00	33.49	850,200	0.00	33.49	850,200
2003	0.00	23.19	547,800	0.00	23.19	547,800
2010	1.46	13.64	233,300	1.46	13.64	233,300
Tietê confluence Pinheiros						
1998	0.00	15.22	150,200	0.00	28.89	777,700
2003	1.33	3.73	20,400	0.43	22.27	563,600
2010	1.14	3.70	22,000	1.98	12.55	246,900
Pinheiros Pumping Station						
1998	0.00	18.47	156,500	0.00	32.22	587,300
2003	1.99	7.21	10,500	0.55	16.66	292,800
2010	2.07	7.20	10,900	2.18	11.62	148,500
Edgar Souza						
1998	0.29	29.16	664,000	0.98	31.95	651,200
2003	1.34	22.88	505,400	2.95	26.10	460,500
2010	2.48	12.65	216,000	4.01	13.09	174,900
Pirapora						
1998	2.24	20.60	8,400	4.35	14.27	2,200
2003	2.60	17.59	10,200	4.50	13.41	2,400
2010	3.03	10.97	3,700	4.27	8.35	800

The Ponte Nova and Edgard de Souza Dams can control the flow of water in the Tietê River system. This effects volume and velocity and therefore quality. If the river is operated exclusively for carrying wastes away from the SPMA, all rivers run in their natural direction. If, however, water is to be diverted to Billings Reservoir, the gates at the Edgard de Souza Dam are partially closed to raise the level of water and cause the Pinheiros River to flow in the reverse direction. SABESP modeled two different operating regimes that are relevant to the economic analysis: (1) operation exclusively to carry away wastes, and (2) a joint operation in which 60% of the water goes to Billings and 40% continues downstream for other uses.[9]

[9] This 60-40 division of water is an arbitrary suggestion of Hidroplan, a consulting firm that developed the master plan for State water use. This division of water does not imply optimal operation and appears to reflect a judgement of what might be politically feasible. Before the Constitutional restriction, water volumes were divided 50-50 between the Tietê and Billings.

Table 4 shows the quality of water at "minimum flow" on various segments of the river system at the end of each of the "stages". Water quality will be better than the level shown 90% of the time (329 days of the year). The results indicate that, by the end of the second stage in 2003, dissolved oxygen will exceed the critical level of 0.5 mg/l from the confluence of the Pinheiros downstream (with the exception of the confluence itself, which does not quite reach 0.5 mg/l if the project is operated for hydroelectric generation). By the completion of the third stage in 2010, there will be significant levels of dissolved oxygen in all segments of the Tietê and Pinheiros whether the system is operated exclusively for carrying wastes or for combined waste disposal and generation of electricity.

Overview of the Economic Analysis

Households with well-maintained individual systems and proper systems of disposition of sludge from septic tanks have little adverse impact on the environment. The waste water of less adequate individual systems (which runs into the street or storm sewer system) may degrade naturally before entering the river, and therefore not cause a significant pollution problem. On the other hand, properly collected household waste channeled directly to the principal tributaries by collectors or storm sewers can have severe specific "point" impacts if the waste water does not have prior treatment. Changes in the number of connections have a relatively small impact on the quality of the Tietê and its major tributaries. Direct dumping of untreated sewage (the case of Guarulhos and Osasco) has a significant impact.

Because connection to a collection system may bring substantial benefits unrelated to improving the quality of the Tietê and its major tributaries, the economic analysis discussed here deals exclusively with clean-up of the river and does not discuss a separate CB analysis done for the provision of household sewer connections. The two are only closely related if one considers the clean-up project as the cost of mitigation of the connection project. In this case, the need for mitigation is debatable. The receptor body is dead; direct dumping will not make it deader [10].

---

[10] This poses a paradox for economic analysis. Connection programs usually generate large benefits sufficient to cover the cost of mitigation (i.e. treatment). However, in the initial stages of sewage collection projects, the discharge of wastes directly into receptor bodies may cause no significant deterioration and are not economically justified. Expansions of sewer collection may finally start to degrade the water, but the surplus of the marginal population (often the poorest) may not be sufficient to cover the cost of cleaning-up everything. The willingness to pay for clean-up of the whole population may not be sufficient to justify the clean-up project.

The analysis that follows develops two economic calculations for the river clean-up project; one for the whole project including Stages I, II, and III and one for the incremental project involving Stages II and III, which have yet to be built. As noted above, Stage I was originally analyzed and approved on the basis of minimum cost, without looking at the entire project, which was presumed to be economically worthwhile. The reprise NPV calculation for the whole project is presented for reasons of transparency, not because of its relevance to the decision as to whether to complete Stages II and III. It demonstrates the well-known weakness of using cost-effectiveness analysis to justify a project that is assumed to be socially beneficial, when in reality it is not.

The other calculation is for the economic return on investments of the second and third stages taken together. This calculation is relevant because the second stage alone is not sufficient to bring any lasting improvement of the quality of the Tietê. The costs and benefits of the first stage are not relevant to the investment decision on Stages II and III in combination because the Stage I costs have already been incurred and cannot be recovered (sunk cost). Therefore the decision to continue depends only on avoidable costs and attainable benefits.

Project Costs and Shadow Pricing Adjustments

SABESP separated the costs to clean-up the Tietê and its tributaries (Table 5) from the sub-programs to connect new users. Table 5 reports capital and operating costs for the pollution control program before the application of shadow price factors. Included are crude estimates of the costs that CETESB and industries will incur to control industrial pollution and monitor performance.[11]

---

[11] There was little information to develop these costs. A World Bank study for control of industrial pollution for 22 industries in the Baixada Santista, gave a range of US$ 13,000 to US$ 6,452,000 per industry with an average of US$ 717,000. It was not possible to extrapolate from this study without knowing the composition of industries that were regulated in the first stage or that will be regulated in the second and third. CETESB estimated that the cost of compliance for the first stage would be US$ 500 million (US$ 400,000 per industry) before the first stage of the Tietê project was implemented. In retrospect, it estimates a cost of US$ 200 million (US$ 171,000 per industry). There is little empirical basis for either of CETESB's estimates. The analysis here uses a range of costs from US$ 171,000 to US$ 342,000 per industry for the 1,168 industries controlled in the first stage and the 350 controlled in the second. It also assumes that the operating cost will be ten per cent per year of the capital cost. The environmental specialist estimated that the monitoring cost for CETESB will be US$ 1.44 million per year for the 1,168 industries covered in the first stage and US$ 0.55 million for the 350 industries in the second stage. These estimates were based on estimates of personnel time needed to check the reports supplied by industries. Operating

The cost of all works and operations relevant to project benefits was subdivided into four categories: traded goods, non-traded goods, skilled labor and unskilled labor. The economic analysis assumes all contingency costs are for tradable items and that all industry investment and operating pollution control programs are for tradable goods. The other costs were adjusted to economic opportunity costs using research done for a prior project, PASS/BID (BR-0269). This study estimated conversion factors[12] for skilled labor of 0.79 and a conversion factor for unskilled labor of 0.48. The PASS study used the reciprocal of the weighted average tariff to estimate a standard conversion factor of 0.91 for nontraded goods but it did not take into account the impact of high interest rates (tight monetary policy) in maintaining the level of the exchange rate. Therefore, this analysis uses a conversion factor of 0.75 non-tradeables.

Table 5. Project (Stages I, II and III) Capital and Operating Costs for Wastewater Treatment
(Undiscounted Thousand 1998 Reals)

Time Period and Stage (Year 1=1992)	Total Investment Costs	Total Operating Costs	Total Costs
BEGIN STAGE I Construction 1	18,304	0	18,304
2	33,559	0	33,559
3	333,948	0	333,948
4	60,864	30,000	90,864
5	72,549	30,001	102,550
END STAGE I Construction 6	87,370	30,001	117,371
STAGE I SUB-TOTAL	**606,594**	**90,002**	**696,596**
BEGIN STAGE II Construction 7	42,333	65,716	108,049
8	54,339	66,124	120,463
9	97,432	66,727	164,160
10	200,685	67,845	268,529
END STAGE II Construction 11	72,633	79,190	151,822
STAGE II SUB-TOTAL	**467,421**	**345,602**	**813,023**
BEGIN STAGE III Construction 12	42,761	104,349	147,110
13	38,090	104,864	142,954
14	60,182	105,100	165,282
15	60,333	105,392	165,725
16	59,667	105,627	165,294
17	59,666	105,916	165,582
END STAGE III Construction 18	57,563	106,100	163,662
STAGE III SUB-TOTAL	**378,262**	**737,347**	**1,115,610**

costs increase over time as the project works are used to capacity.

[12]The conversion factors are used to convert market prices reflecting imperfections to their social value.

Table 5. Project (Stages I, II and III) Capital and Operating Costs for Wastewater Treatment *(Undiscounted Thousand 1998 Reals)*			
Time Period and Stage (Year 1=1992)	**Total Investment Costs**	**Total Operating Costs**	**Total Costs**
19	0	135,288	135,288
20	0	137,415	137,415
21	0	137,592	137,592
22	0	137,769	137,769
23	0	137,947	137,947
24	0	138,036	138,036
25	0	138,127	138,127
26	0	138,216	138,216
27	0	138,306	138,306
28	0	138,306	138,306
29	0	138,306	138,306
END ANALYSIS PERIOD 30	0	138,306	138,306
GRAND TOTAL	**1,452,277**	**1,172,952**	**2,625,228**

Note: Costs exclude household sewer connections, the collection system, and the cost of collectors sufficient to carry untreated effluent to the nearest dumping point in the river. These costs were balanced against local household sewering benefits in a separate CB exercise not reported here. The costs above are related to pollution control and including interceptors, treatment and the industrial environmental cleanup program.

Project Benefits

There are two principal benefits from Stages II and III of the Tietê project: (1) increased welfare of residents based on reduction of odors and aesthetic blight, and (2) increased hydroelectric power generation. The benefits from Stage I are assumed to be negligible. Each of the benefit categories is discussed in turn. [13]

Benefit One: Reduction of Odors and Improvement of River Water Quality

To determine the benefits from reducing odors and permitting aquatic life, SABESP contracted a contingent valuation study of improvements in water quality. The questionnaire was developed by Robert Mitchell, an internationally known expert in the design of contingent valuation questionnaires. It was tested with two

[13] In addition, there are other benefits that have not been quantified which include (a) increased recreation benefits at Pirapora do Bom Jesus and downstream, (b) the retardation of saline intrusion in the Cubatão River in the Baixada Santista, and (c) the provision of a more economic source of potable water for the Baixada.

focus groups with people of different educational and income levels[14], in-depth interviews, and a pilot survey of 150 households in ten municipalities and neighborhoods of the SPMA.

The focus group sessions revealed a clear concern about governance and the judiciousness with which government spends. Many stated that the government collected sufficient money and that an additional contribution should not be needed. The two together led to a number of protest responses: refusal to pay. Those that distinguished SABESP from the government, had more confidence in SABESP. Many of those who stated that they were willing to pay clearly associated the river clean-up with better health. Health aspects were not, however, emphasized in the focus groups or in the questionnaire, because the relationship between the river clean-up and health is much less direct than the relationship between health and sewer systems that remove waste water from areas where people live.

To deal with the issues detected in the focus groups, the willingness to pay question specifically emphasized that: (1) money would be collected only during the 10 years of construction of the second and third stages, (2) the money would be used exclusively for the clean-up project, and (3) SABESP would annually present information on television and on radio about how the money was being used. To reflect the quality that would actually result from the proposed works, the questionnaire used maps to show what parts of the river would improve. It indicated that the greatest improvement that could be expected was that the water quality would permit boating and the existence of fish in some segments. It emphasized that it would not be safe to swim in any of the rivers. A translation of the core valuation question and the maps illustrating project impact appear below.

The questionnaire used the referendum design which asked the respondent whether he would be willing to pay one of five monthly prices (R$0.50, 2.00, 5.00, 12.00, 20.00) until the year 2010 when the project would be completed. These prices were found to cover the relevant range of price acceptability found in the focus groups and were assigned to an equal number of questionnaires, i.e. each price appeared on one fifth of the questionnaires.

Although the questionnaire was superior to those generally used, it does not describe the improvement in the quality of the river as precisely as it might have. The willingness to pay question is not precise about what "average quality" of water means. Because water quality depends on river flow and flow varies, quality

[14] More detail on the composition of the focus groups and the concepts studied appear in "Relatorio Preliminar de Projeto, Anexo, Relatorio da Realização de Pesquisa de Disposição a Pagar," June 1998.

will vary. For economic valuation, it is necessary to have a specific meaning to average quality. This analysis assumes that "average" means that the water at "minimum flow" either has no odor or has no odor and supports aquatic life. This implies that the respondent accepts the fact that 36 days a year (10% of the time) the water will have some odor and may not support aquatic life. It should be noted that the days in which the river does not comply will be roughly continuous during the dry season. The concentration of the timing will make failure to provide "average quality" more obvious. Table 4 above shows the predicted quality of critical segments of the river at the end of each stage of the project. The model projects that the critical variable dissolved oxygen will be above 0.5 mg/l or 0.2 mg/l in the segments asked about in the contingent valuation study.

To apply the questionnaire, SABESP drew a sample based on IBGE's 1996 survey of households in the SPMA. The strategy was to represent the population of São Paulo in terms of those factors that are likely to determine willingness to pay. In theory such factors would include income, degree to which the household is affected by the river's odors, environmental awareness, and education. Of these, the census has information only on income and education and these are highly correlated.

According to the Census, the average income of SPMA is R$ 828/month with a standard deviation of R$702. Using a 95% confidence interval and a 10% sampling error, SABESP calculated that it needed a sample of 276 homes.[15]

The Contingent Valuation Survey

---

[15] The necessary sample size was initially calculated by SABESP based on the amount of tolerable error in the sample estimate of mean income rather than mean WTP (which was unknown), using a standard statistical formula (e.g. Paffenberger and Patterson 1987, p. 391). The result, 276 households, was more than doubled when the survey was actually applied, presumably because the available budget permitted a larger sample and hence more precise results. The sample size determination formula is:

$$n = [z_{\alpha/2}\ \sigma/E]^2 = 276$$

where: n = desired sample size
z = the 95% confidence interval statistic (1.96) at significance level α = 5%, 2 sided test.
σ = standard deviation of income (R$702).
E = acceptable error in sample average estimate of population mean ($R 82.8) obtained as one-tenth of census estimate of average household income of $R828 (i.e. a 10% error).

Note that the variable of interest is household willingness to pay (WTP), not income, so the formula only holds if the mean and standard deviation of WTP bear a fixed proportional relationship to the mean and standard deviation of income.

The following excerpts from the CV questionnaire show how the valuation question, which followed questions on household characteristics, was structured:

Look at ***Map 1 (map available from author)****. The triangles and circles depict SABESP's five water treatment plants. The larger the size of the symbol the larger the quantity of wastewater treated. The two plants represented by the triangle have been operational for some time, treating 20% of SPMA wastewater.*

In 1993, SABESP initiated works for Stage I of the River Tietê decontamination program. Three new plants (depicted by the circles) are planned to be operational by the year 1998. With these new stations, 40% of the industrial and domestic load will be treated. Consequently, water quality of the Tietê River and its tributaries will improve. Still, 60% of the domestic and industrial load will reach the rivers untreated.

Even with three new treatment plants operational by 1998 water quality of the Rio Pinheiros will continue to be poor. The sections of the rivers in grey depict an acceptable level of water quality mainly due to the elimination of odors; still, no aquatic life is supported. On the other hand, the river sections delineated in white support some aquatic life and boating is permitted.

SABESP has a project to continue the decontamination of the River Tietê . Under the new project, more treatment plants will be built and an expansion of the existing treatment plants is foreseen. If the project is pursued, in 10 years 95% of pollutants will be treated, improving water quality of the rivers. ***Map 2*** *(map available from author) depicts the improvement in water quality during the next 10 years.*

As shown in the map, in the next five years, the Rio Pinheiros will show a considerable improvement in water quality. On the other hand, water quality in the River Tietê and Tamanduateí will not improve. By 2008, at the conclusion of the proposed project, all of the rivers will have an acceptable or good water quality level.

The costs involved in such a project are high and there are not enough financial resources.

What would you prefer:

Pay R$(bid amounts: 0.5, 2, 5, 12, and 20) rendered as an increase in you monthly water utility bill for the next 10 years for an improvement in water quality as depicted in Map 2 or not pay and the project will not be executed leaving water quality of the rivers of Sao Paulo at the current levels?

The actual sample size was 600. Based on data of the 1991 Census, this number of survey households would give a sampling error of 6.8% To increase the efficiency of the survey, SABESP stratified the sample by sub-regions. The central sub-region (metropolitan São Paulo) and the northeast (Guarulhos) were pre-selected. The other three were selected at random.

To extract an average measure of WTP from the referendum CV data, Logit probability models were fit, using as explanatory variables the bid (Valor), the age of the respondent (Idade), an indicator of household status/wealth (P118), and the contiguity of the area of residence to the Tietê and its major tributaries (Bairro). For the linear - in - bid models a large portion of the predicted cumulative distribution function lies in the negative quadrant. This implies that a large percentage of the population has a negative willingness to pay which obviously is not consistent with economic theory.

The next section depicts all of the central tendency measures of WTP calculated from the Tietê referendum survey data.

Benefit Two: Additional Hydroelectric Generation

Until 1992, half of the flow of the Tietê was pumped to Billings Reservoir to generate electricity at the Henry Borden power plant. But "Transitory Provisions" of the State's 1989 Constitution prevent pumping waste water to Billings. With the treatment plants of Stages II and III, the possibility exists of having water good enough to pump, and there are potential benefits from additional hydroelectric generation. It is not certain, however, that pumping will be allowed. Billings Reservoir is a Class II water body. By regulation, water pumped to a Class II body can not degrade the quality to less than 5 mg/l of dissolved oxygen, more than 5 mg/l of biological oxygen demand, or more that 4,000 fecal coliforms per 100 ml. At the point in the Billings Reservoir where the flow diverted from the Tietê/Pinheiros would be injected, the receiving water is already out of compliance. Thus, in principle, the only water that could be pumped there would be distilled water.

The water at the Pinheiros pumping station will not attain Class II quality in either stages II or III. Billings, however, is a very large reservoir. Its present quality ranges from Class IV (where Pinheiros/Tietê water would be injected), to Class I in the area where water is released for potable uses to the Santos region. It would be acceptable from the technical point of view to reclassify sections of Billings to reflect present reality and the fact that the reservoir functions as a natural treatment plant. If the part of Billings where the Tietê water would be pumped were reclassified as Class IV, Tietê water could be pumped to Billings. It is highly uncertain, however, whether Billings will be reclassified, if such pumping will be allowed, when it will be allowed, and how much will be allowed.

The economic analysis explores a number of hypotheses. The analysis assigns a 50 percent probability to the most adverse case; the law never permits diversion of water from the Tietê to Billings so no hydroelectric benefits are generated. The analysis assumes that there is 5% probability that pumping will be possible the year after the second stage is completed, a 10% probability in the fourth year after Stage II is completed, a 17.5% probability the year after Stage III is completed and a 17.5% chance four years after Stage III is completed.

If pumping is allowed, the benefits from using Tietê/Pinheiros water will depend on the amount of energy and the time of day when it will be generated. To estimate the amount of additional energy that might be generated, historical data was obtained on the amount of water processed by the Henry Borden Power Station before and after the restriction imposed on pumping from the Tietê/Pinheiros. The average difference is equivalent to a continuous flow of 67.0 m^3/s (Table 6 below). This flow resulted from a 50-50 division of the Tietê's flows. A master plan for the water resources of the State of Sao Paulo suggests that a 60-40 division might be possible. If so, it might be possible to pump the equivalent of a continuous 80 m^3/s to Billings.

Table 6: Henry Borden Power Plant Energy Production and Water Use Before and After Restriction on Pumping from the Tietê to Billings

Year	Energy Produced (MWh)	Water Used (M^3/S)
1985	3,702,424.5	75.4
1986	4,244,978.9	86.4
1987	5,056,923.8	104.5
1988	4,816,698.2	99.1
1989	5,230,506.9	108.1
1990	3,606,258.4	74.2
1991	4,798,196.6	98.1
Average 1985-1991	**4,493,283.9**	**92.3**
1992 [a/]	2,811,472.1	57.2
1993	1,579,454.0	32.1
1994	694,913.7	14.1
1995	1,255,767.6	26.2
1996	1,535,861.1	31.0
1997	1,131,306.8	23.3
Average 1993-1997	**1,239,460.6**	**25.3**
DIFFERENCE OF AVE.	**3,253,823.3**	**67.0**

a/ In 1992, the restrictions on pumping were imposed.

The incremental energy generated is the difference between the energy that can be generated at Henry Borden with the pumped water less (1) the energy that could be generated with the water on the ten downstream plants on the Tietê, and (2) the energy used in pumping. Henry Borden has a production capacity of 5.654 $MW/m^3/s$.

The consumption of energy to pump a cubic meter to Billings is 0.34 m^3/sec. Thus, Henry Borden's net gain from receiving a cubic meter per second is 5.34 $MW/m^3/s$. This net gain is offset by the losses of the hydroelectric plants on the lower Tietê. Table 7 shows that the production generated by a cubic meter passing through all the plants is 2.1336 $MW/m^3/s$. Thus, the net national gain from transferring a cubic meter from the Tietê to Billings is 3.206 $MW/m^3/s$ (5.3400-2.1336). This converts to 27,084.56 Mwh of additional energy per cubic meter per year. For an incremental flow between 67 and 80 cubic meters, the incremental energy is in the range of 1,881,665 Mwh to 2,246,764 Mwh per year.

Table 7: Power Stations on the Tietê Downstream from Edgard Souza

Power Plant	Production ($MW/m^3/s$)
Rasgao	0.1754
Porto Góes	0.1887
Barra Bonita	0.1727
A. Souza Lima	0.1881
Ibitinga	0.1872
Promissao	0.2057
Nova Avanhandava	0.2605
Ilha Solteira	0.3902
Jupiá	0.1982
Porto Primavera	0.1669
Total	**2.1336**

Because Billings is an enormous reservoir with inter-annual storage, it is possible to produce most of the incremental energy at peak. The plants on the lower Tietê have enough storage capacity to guarantee peak operations with or without the diversion to Billings. The decrease in power on the Tietê will be power off-peak. It is not certain, however, that Borden would be allowed to use all the Tietê water during peak hours, because the power company may be ordered to increase the constant release (mostly off-peak) to the Cubatao. The value of high voltage energy during peak demand is R$37.33 and R$42.69 per Mwh, depending on whether it is wet or dry season, and the value off peak is R$25.67 and R$30.20 per Mwh, again depending on the season. The economic analysis uses a range of $R30.20 to R$42.69 in its calculations. These values are based on the long run average incremental cost of supply at high voltage.

Section V: Central Tendency Measures

An issue is which measure of central tendency to use, once having estimated some probability- of-bid-acceptance model from referendum data. Again, the debate goes back at least ten years. Hanemann (1989) and Haab and McConnell (1997, August 1997, July 1998, January 1999) argue for the median of individual WTP because in probability models it is less sensitive to distributional misspecification and estimation method. Hanemann (1989) also points out that the median is a more equitable social choice rule for aggregation of willingness to pay across the population for a cost-benefit test than the mean or the mode.

Sometimes the discrepancies among the alternative central tendency measures can be large enough to confound a project acceptance or rejection decision using CB criteria — the project passes the test using some subset of central tendency measures and fails it using others. Put simply, the unbounded expected value measure obtained by using a linear utility index in estimation of a probability model is not generally satisfactory and may understate benefits. But, when distributional asymmetry is introduced to correct for this by either truncating the range of expected value function evaluation or by introducing non-linearity in the utility index, the mean individual WTP extracted from referendum models no longer equals the median and will usually exceed it. In this case using the median as a benefit measure means that project acceptance will not be as strongly influenced by a few extreme observations lying in the tails of the (asymmetric) WTP distribution as it would be using the mean. Experienced analysts know that to get the highest benefits possible and unabashedly seek project acceptance under an NPV or EIRR criterion, the mean of an asymmetric distribution can be used, but its median will provide a more cautious, conservative lower bound on project payoff. It seems reasonable to recommend at least taking a look at the latter, or reporting both mean and median.

Parametric Choice Models

To demonstrate, the standard central tendency measures described above and depicted in Table 2 above were obtained by applying a Logit choice model to the 600 survey sample observations, coding the dependent variable as 1 if the offer was accepted, and 0 if not. Simple linear and log bid specifications of the utility index were used.[16]

Probability Model Estimation

The independent variables in the statistical Logit model included the bid value (Valor), the age of the respondent (Idade), and a household wealth/social status indicator (P118). A dummy variable was included to distinguish between residents who live close to the river (Bairro) (184 households), and are significantly more affected by its pollution, than households not residing in close proximity.

The specification of the model is:

(15) $$\text{Prob(Yes)} = \alpha_0 + \beta_1\text{Bairro}+\beta_2\text{P118}+\beta_3\text{Idade}+\beta_4\text{Valor}+\text{error}$$

for the linear utility difference logit and:

(16) $$\text{Prob(Yes)} = \alpha_0 + \beta_1\text{Bairro}+\beta_2\text{P118}+\beta_3\text{Idade}+\beta_4\text{LNValor}+\text{error}$$

for the utility difference Logit with log of bid[17].

The estimation results appear in Table 8.

---

[16] Note that the dummy variable specification shifts the function but imposes the restriction that households living near or far from the river share the same regime with respect to the other parameters. The log bid model's expected value could not be evaluated using an analytic formula because its parameters fell outside the limits of the formula's applicability (Hanemann 1984, p. 337). Numerical approximation was used to compute the means of the log bid model (see Annex 1).

[17] Note that the only difference between equation 15 and 16 is that in the latter one, we are taking the natural log of the Bid variable (Valor). By taking the natural log of the bid, we restrain the model into the positive quadrant.

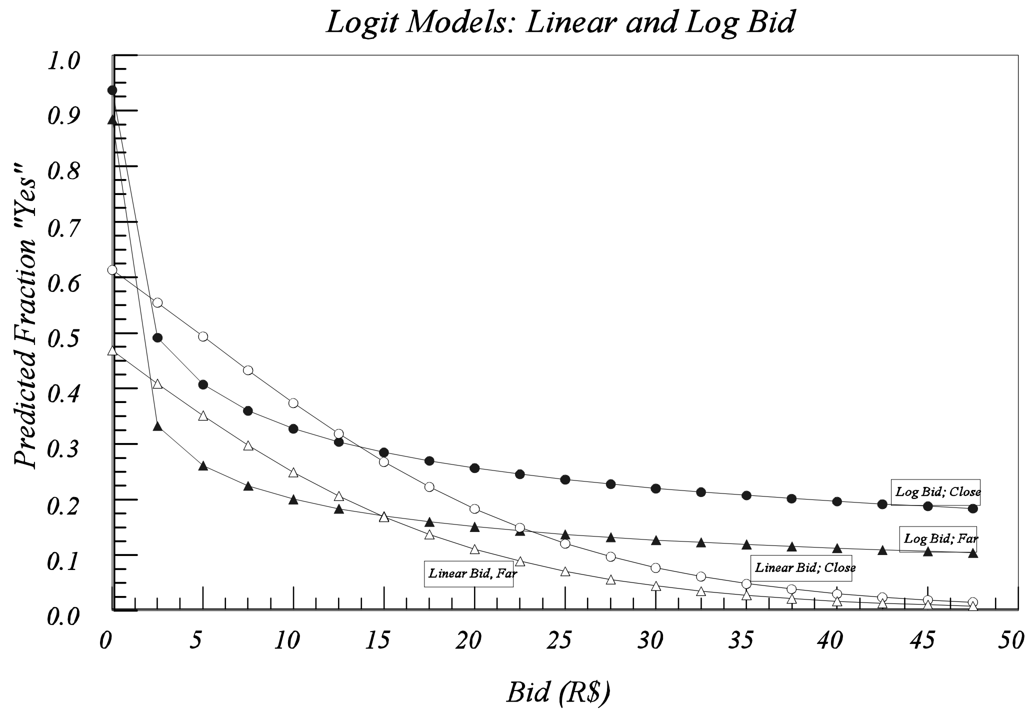
Table 8. Logit Model Parameter Estimates and Variable Means					
Variable	**Linear Bid Model Coefficient** (*t stat.*)	**Log Bid Model Coefficient** (*t stat.*)	**Means of Variables**		
			Full Sample	**Close Sub-Sample**	**Far Sub-Sample**
Constant	0.7769 *(2.38)*	0.7608 *(2.30)*	...	...	...
Close to River (1 if Yes, 0 Else)	0.6551 *(3.29)*	0.6629 *(3.33)*	0.3066	1	0
Status (1 if Upper, 0 Else)	0.8357 *(2.92)*	0.7968 *(2.78)*	0.11	0.1467	0.0938
Age of Household Head (Years)	-0.0221 *(-3.20)*	-0.0227 *(-3.27)*	45.88	49.38	44.34
Bid (R$/Household/Month)	-0.0978 *(-6.78)*	...	7.9	7.99	7.86
Log of Bid (ln R$/Household/Month	...	-0.4945 *(-6.99)*	1.42	1.43	1.41

Note: For the linear bid index model, Unrestricted Log Likelihood=-350.00, Restricted Log Likelihood (intercept only) = -389.08, Chi-squared statistic = 78.15, significant at >1% level, and Pseudo R^2 = 0.10. For the log bid index model, Unrestricted Log Likelihood=-350.65, Restricted Log Likelihood (intercept only) =-389.08, Chi-squared statistic = 76.79, significant at >1% level, and Pseudo R^2 = 0.098.

All parameter estimates are significant at better than the 5% level, and most have signs that are consistent with prior expectations. Households close to the river are more likely to be willing to pay than more distant households, as are wealthier households.

Predictions of the acceptance rates across bid levels for both models, evaluated at their respective sub-sample means, are displayed in Figure 1. Notice that the logarithmic specification confines all of the distribution function to the positive bid quadrant, while the linear specification potentially extends to the left of zero, even though this region is omitted from the figure. The thicker tails of the log bid models suggest arithmetic means that should exceed the arithmetic means of the linear bid models. However, we used geometric means for the log bid models, which explains why they fall below the arithmetic means of the linear bid models in Table 9 below.

Figure 1



Central Tendency Measures from Parametric Methods: Results

Applying the expected value and median formulas produces the WTP estimates in Table 9 for the untruncated mean, the mean truncated at zero but untruncated from above, the truncated mean confined between zero and the maximim bid (20 reals), and the median.[18]

[18] The unit of currency used throughout is the Brazilian real (reais), denoted as R$. The rate of exchange in March 1998 was 1.14 reals per U.S. dollar. All estimates presented were produced by evaluating the relevant formulas at the means of the explanatory variables rather than calculating individual-specific values and averaging them over the sample to obtain a grand mean.

Table 9. Parametric Central Tendency Estimates			
Central Tendency Measure		Household Willingness to Pay per Month (*1998 Reals)*	
		Close to River	Far from River
Median = Untruncated Mean, E(WTP), - ∞ < WTP < ∞ (utility difference logit, linear in bid)	C+ C*	4.74	-1.27
Truncated Mean, E(WTP), 0 < WTP < ∞ (utility difference logit, linear in bid)	C'	9.73	6.47
Truncated Mean, E(WTP), 0<WTP<B_{max} (utility difference logit, linear in bid)	*C~*	7.65	5.27
Truncated Mean, Log Transform, E($exp^{ln(WTP)}$), - ∞ < lnWTP <∞	C^{+}_{ln}	5.01	1.68
Truncated Mean, Log Transform, E($exp^{ln(B)}$), - ∞ < lnWTP <ln Income	$C^{\sim}_{ln}$	3.71	1.39
Truncated Median, Log Transform	C^{*}_{ln}	2.36	0.71

Note: The augmented intercepts are 0.4633 for Close and -0.1246 for Far in the linear model. For both cases, β, the marginal utility of income estimate, is 0.09778 (after multiplying by -1 to make it positive). In the log of bid model the augmented intercepts are 0.4245 for Close and -0.1665 for Far. For both cases, β on the natural log of bid is 0.4945 (after multiplying by -1 to make it positive). Geometric means were calculated for the log transform models by taking the antilog of the mean log bid found by numerical approximation.

These results pose two dilemmas. First, the unrestricted mean WTP for households living far from the river is negative. Second, there is a large disparity between the several alternative truncated means using either a linear or log bid specification.

If project justification (rather than analysis) is the goal, it might be tempting to use the truncated mean that gives the highest benefit. Few would ever detect this sleight of hand. However, an honest project appraisal would admit that things are not quite so simple. Hanneman (1989) indicates that the measure C' unambiguously overstates the true mean in situations where the augmented intercept is greater than zero (i.e. when the probability of acceptance at a zero bid is greater than 0.5).

Also, it is inconsistent to use an untruncated distributional assumption for estimation and a truncated rule like C' for function evaluation. In other words, an inconsistency arises because in estimation of a Logit model with a linear utility index difference the domain of the fitted cumulative density is theoretically allowed to include all the real numbers even though the random variable is known a-priori to exclude negative values. Then, in function evaluation, a "correction" like C' or $C^{\sim}$ is made ex-post by using only that portion of the fitted distribution lying in the positive probability/bid quadrant to compute the expected value integral.

Analysis Options: Non-Parametric Methods

Haab and McConnell offer two simple yet effective alternatives for estimating WTP that overcome the necessity of arbitrarily truncating WTP at zero or some upper bound (or both) in discrete choice referendum models. The first route is a "distribution-free" nonparametric technique for getting lower-bound estimates of the mean and median (McConnell 1995; Haab and McConnell 1997). The other involves a reformulation of the Probit or Logit model that automatically guarantees that median WTP will be greater than a lower bound of zero but never be greater than income (Haab and McConnell, August 1998, January 1999). At a minimum, it is probably a good idea to calculate a nonparametric[19] estimate of the mean and median before getting too deeply involved in estimation of WTP, just to have a benchmark.

The Turnbull Nonparametric Technique

Consider a stylized contingent valuation question. Respondents are asked: "Would you be willing to pay an amount b_j?" The b_j are indexed j = 0,1 ... M+1 and $b_j > b_k$ for $j > k$, and $b_0 = 0$. Let p_j be the probability that the respondent's WTP is in the bid interval b_{j-1} to b_j. This can be written:[20]

(17) $$p_j = P(b_{j-1} < w \leq b_j) \text{ for } j = 1, ..., M+1..$$

Alternatively, the cumulative distribution function (CDF) is written:

[19] In the context of this paper, nonparametric means "distribution-free"; that is, the distribution function of the random variable producing the data need not be specified.

[20] This section is an abridged version of the presentation in McConnell (1995). A complete treatment is available in Haab and McConnell (1997).

(18) $F_j = P(w \leq b_j)$ for j = 1, ..., M+1, where $F_{M+1} = 1$.

For reasons already discussed, one aims to have b_{M+1} high enough that $F_{M+1} = 1$. That is, b_{M+1} is effectively infinite in the problem setting. Then

(19) $$p_j = F_j - F_{j-1}$$

and $F_0 \equiv 0$. The Turnbull can be estimated by treating either the F_j, j = 1 →M or p_j, j = 1 →M as parameters.

The p's can be estimated quite simply. Let N_j represent the number of "no" responses registered in each bid group j. If $[N_j /(N_j + Y_j)] > [N_{j-1}/(N_{j-1} + Y_{j-1})]$ for all j between one and M, then $p_j = [N_j /(N_j + Y_j)] - [N_{j-1}/(N_{j-1} + Y_{j-1})]$. The probability $N_j/(Y_j + N_j)$ represents the proportion of respondents who say 'no' to b_j. As such, it is a natural estimator of F_j.[21]

Hence, the estimator of p_j could be written:

(19) $$p_j = F_j - F_{j-1}, \ where \ F_j = \frac{N_j}{N_j + Y_j}$$

Expected willingness to pay can be written as:

(20) $$E(WTP) = \int_0^{\infty} WTP \ dF(WTP) = \sum_{J=1}^{M+1} \int_{b_{j-1}}^{b_j} WTP \ dF(WTP)$$

Replacing willingness to pay by the lower bound of each interval produces a lower bound estimate of the expected value of willingness to pay:

(21) $$(LB_{WTP}) = 0 \cdot P(0 \leq w < b_1) + b_1 P(b_1 \leq w < b_2) + ... + b_m \ P(b_m \leq w < b_{m+1}) = \sum^{M+1} b_{j-1} p_j$$

[21] The estimate of F_j assumes the proportion of no responses increases as the bid increases across all bid classes. If not, McConnell and Haab (1997) show how to join bid groups to achieve monotonically increasing proportions. This was not necessary with the Tietê survey data, except for the first two bid groups in the far- from-river sub-sample.

where p_{M+1} = 1-F_M. The variance of the lower bound mean is:

(22) $$V(\sum^{M+1} p_j \ b_{j-1}) = \sum^{M+1} b_{j-1}^2 (V(F_j) + V(F_{j-1})) - 2\sum^{M} b_j b_{j-1} V(F_j)$$

where the variance of each proportion $V(F_j)$ is equal to $F_j(1-F_j) / (N_j + Y_j)$.

This too can be calculated rather easily from a simple table of proportions of yes's or no's and the total number of respondents in each grouping. The results of applying these formulas are displayed in Tables 10 and 11, which also provide a linear interpolation for the median.

Notice in the tables that b_M is the highest bid actually offered respondents and is the lower bound of the final interval running from b_M to infinity. In the expected value formula in Eq. (22), b_M is used with no attempt to guess at an appropriate value to apply to the portions of the two sub-samples who had WTPs greater than R$20 (24 and 11 percent respectively). This is what produces the lower bound label and distinguishes the Turnbull approach from Kriström's method discussed next.

Table 10. Turnbull Lower Bound Mean and Median Estimates: Close to River Sub-Sample

Bid Group j	Bid ($/month)	Bid Range	Total # of "No" Answers N_j	Total # of Obs. $TOTAL_j$	CDF=F_j= N_j /$TOTAL_j$	PDF=Pj= F(j)-F(j-1)	Lower Bound Estimate of E(WTP)
0	0.50	0-0.50	10	37	0.270	0.270	0.00
1	2.00	0.50 - 2.00	13	33	0.394	0.124	0.06
2	5.00	2.00 - 5.00	26	41	0.634	0.240	0.48
3	12.00	5.0 - 12.00	26	35	0.743	0.109	0.54
4	20.00	12.00 -20.00	29	38	0.763	0.020	0.24
5	>20.0				1.000	0.237	4.74
		Totals :	104	184		1	
						E(WTP):	R$6.07
						Variance E(WTP)	R$0.64
						Std. Error (Dev) E(WTP)	R$0.80
						Median WTP	R$3.33

Note:

The median bid was found by linear interpolation between the bids attached to the cumulative frequencies (CDF values) above and below 50%. That is, Med=B_l + k(i) where B_l is the lower (left) boundary of the class containing the median ($2.00), i is the class interval ($3.00) and k approximates where the 50% point lies inside the CDF values at the lower and upper boundaries ((0.5-0.394)/(0.634-0.394)). So, $3.33=$2.00 + 0.44*$3.00.

Table 11. Turnbull Lower Bound Mean and Median Estimates: Far from River Sub-Sample							
Bid Group j	**Bid ($/month)**	**Bid Range**	**Total # of "No" Answers N_j**	**Total # of Obs. $TOTAL_j$**	**CDF=F_j= N_j /$TOTAL_j$**	**PDF=Pj= F(j)-F(j-1)**	**Lower Bound Estimate of E(WTP)**
0	2.00	0.00 - 2.00	93	170	0.547	0.547	0
1	5.00	2.00 - 5.00	57	79	0.722	0.174	0.35
2	12.00	5.0 - 12.00	62	85	0.729	0.008	0.04
3	20.00	12.00 -20.00	73	82	0.89	0.161	1.93
4	>20.0				1	0.11	2.2
		Totals :	285	416		1	
						E(WTP:	R$4.51
						Variance E(WTP)	R$0.22
						Std. Error (Dev) E(WTP)	R$0.47
						Median WTP	R$1.83

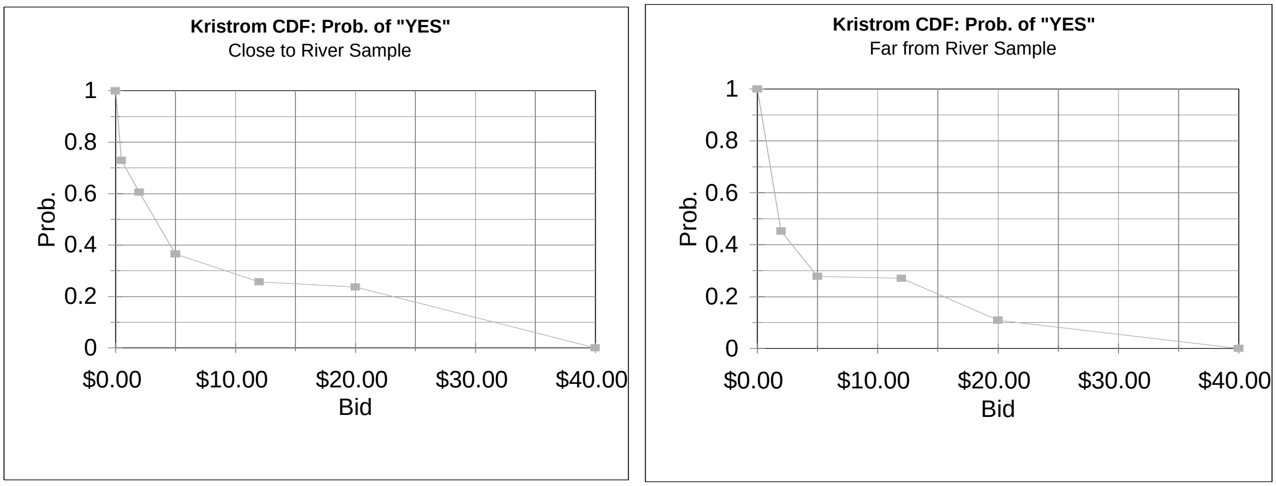
Note:

The median bid was found by linear interpolation between the bids attached to the cumulative frequencies (CDF values) above and below 50%. That is, Med=B_l + k(i) where B_l is the lower (left) boundary of the class containing the median ($0.00), i is the class interval ($2.00) and k approximates where the 50% point lies inside the CDF values at the lower and upper boundaries (0.5/0.547). So, $1.83=$0.00 + 0.914*$2.00.

<u>Kriström's Nonparametric Mean</u>

Kriström's (1990) nonparametric method is even easier to calculate and understand than the Turnbull. In words, all one does is array the frequency of affirmative responses in each bid class in monotonically descending order with ascending bids, connect the points by linear interpolation, and approximate the integral under the resultant empirical cumulative density to get the mean (see Annex 1). The figures below show the approximate empirical distributions. Average income in the close-to-river sample is 30 percent higher than the far-from-river average, which probably causes the corresponding density to be more stretched out toward the higher bid levels.

Figure 2. Nonparametric Inverse Cumulative Distributions



Unlike the Turnbull, the bid that drives the probability of acceptance to zero must be specified by the analyst if the survey does not reveal it, so Kriström's mean depends in part on this arbitrary value. To construct the empirical cumulative densities pictured above, a conservative upper limit of R$40 for b_{M+1} was assumed, which is approximately three percent of average household income (see Ardila *et al.* 1998). Tables 12 and 13 show the calculation steps.

The influence of the final interval between the last posited bid and the assumed bid driving acceptance to zero is evident from the entries in the penultimate row and last column of the tables, just above their shaded "Average WTP" cells. In the close-to-river case, this value accounts for nearly seventy-five percent of the overall mean value, and in the far-from-river-case, forty-five percent of the mean value is due to the last interval. If the upper limit driving the acceptance rate to zero were set to R$30 rather than R$40, the close and far means would fall by about 50¢ and 75¢, respectively, illustrating their sensitivity to this assumption. The nonparametric estimates of location would probably be better had the sample included more bid intervals spanning a wider bid range.

Table 12. Kriström Nonparametric Mean: Close to River Sample								
Bid Group j	**Bid (R$/month)**	**Bid Range**	**Bid Mid-Point**	**Total # of "Yes" Answers (Y_j)**	**Total # of Obs. Total j**	**$1-F_j$ = Y_j/Total j**	**P_j = $[1-F_{j-1}]-[1-F_j)]$**	**Kriström Estimate of WTP**
na	0.00	0	0	na	na	1	na	0.00
0	0.50	0-0.5	0.25	27	37	0.7297	0.2703	0.07
1	2.00	0.5-2.0	1.25	20	33	0.6061	0.1237	0.15
2	5.00	2.0-5.0	3.5	15	41	0.3659	0.2402	0.84
3	12.00	5.0-12.0	8.5	9	35	0.2571	0.1087	0.92
4	20.00	12.0-20.0	16	9	38	0.2368	0.0203	0.32
5	40.00	20-40	30	0	0	0.0000	0.2368	7.11
Note: The median bid was found by linear interpolation between the actually offered bids (not mid-points) attached to the cumulative frequencies (CDF values) above and below 50% acceptance. That is, Med=B_u - k*i where B_u is the bid in the first class containing more than 50% of "yes" observations ($5.00), i is the interval between adjacent bids bordering the median ($3.00) and k approximates where the 50% point lies ((0.6061-0.50)/(0.6061-0.3659)). So, $3.67 = $5.00 - 0.44*$3.00.							**Average WTP**:	R$9.42
							Median WTP:	R$3.67

Table 13. Kriström Nonparametric Mean: Far from River Sample								
Bid Group j	**Bid (R$/month)**	**Bid Range**	**Bid Mid-Point**	**Total # of "Yes" Answers (Y_j)**	**Total # of Obs. Total j**	**$1-F_j$ = Y_j/Total j**	**P_j = $[1-F_{j-1}]-[1-F_j)]$**	**Kriström Estimate of WTP**
na	0.00	0	0	na	na	1	na	0.00
0	2.00	0.0-2.0	1.25	77	170	0.4529	0.5471	0.55
1	5.00	2.0-5.0	3.5	22	79	0.2785	0.1745	0.61
2	12.00	5.0-12.0	8.5	23	85	0.2706	0.0079	0.07
3	20.00	12.0-20.0	16	9	82	0.1098	0.1608	2.57
4	40.00	20-40	30	0	0	0.0000	0.1098	3.29
Note: The median bid was found by linear interpolation between the actually offered bids (not mid-points) attached to the cumulative frequencies (CDF values) above and below 50% acceptance. That is, Med=B_u - k*i where B_u is the bid in the first class containing more than 50% of "yes" observations ($2.00), i is the interval between adjacent bids bordering the median ($2.00) and k approximates where the 50% point lies in the interval (([1-.4529]-.50)/(1.00-0.4529)). So, $1.83 = $2.00 - 0.086*$2.00.							**Average WTP**:	R$7.09
							Median WTP:	R$1.83

The Bounded Probit or Logit of Haab and McConnell

Rather than starting from a RUM model specification, Haab and McConnell (August 1997, July 1998, January 1999) start at the other end with an expression for WTP that represents the amount of income the individual is willing to pay, expressed as the product of income and a proportion of income lying between zero and one. Somewhat analogous to the conventional RUM, the proportion is estimated as a function of the bid amount and other socioeconomic variables but the bid-related variable disappears when predicting the median proportion.[22]

While this approach makes no claim to being consistent with any theoretical indirect utility function, it solves the practical problem of finding a non-zero WTP that at the same time will not exceed income. Haab and McConnell suppose that WTP lies between zero and some upper bound, A, such that:

(23) $$Median(WTP_i) = \frac{A_i}{1+e^{-X_i\beta-\varepsilon_i}} = p(\varepsilon_i)A_i$$

where $p(\epsilon_i) = 1/(1+e^{-X(i)\beta-\epsilon(i)})$ falls in the (0,1) interval, $\epsilon_i \sim N(0, \sigma^2)$, $X_i\ \beta$ is the inner product of the J covariates ($X_i = X_{i1}...X_{iJ}$) and a vector of coefficients β and A_i is a known constant for individual i, such as income, which is assumed to be a reasonable upper bound on willingness to pay. When A_i is interpreted as income, equation (23) shows that WTP goes to zero for very large negative errors or $X_i\beta$ and to income with very large positive errors or $X_i\beta$.

If the ith respondent is asked "Would you pay 'B_i' for a proposed water quality improvement?" the probability of a no response is the probability that willingness to pay would be less than B_i. Haab and McConnell write this as:

(24) $$P(WTP_i < B_i) = P\left(\frac{A_i}{1+e^{-X_i\beta-\varepsilon_i}} < B_i\right) = P\left(\frac{\varepsilon_i}{\sigma} < \frac{-\ln\left(\frac{A_i-B_i}{B_i}\right) - X_i\beta}{\sigma}\right)$$

When ϵ_i is distributed N(0,1), the last expression on the right hand side is the contribution to the likelihood function for a standard probit model, where the probability of a 'no' response is modeled with the covariates X_i and ln [$(A_i-B_i)/B_i$]. Similarly, the probability of a 'yes' response becomes:

---

[22] The balance of this section is drawn directly from parts of Haab and McConnell's papers.

(25) $$P(WTP < B_i) = P\left(\frac{\varepsilon_i}{\sigma} < \frac{\ln\left(\frac{A_i - B_i}{B_i}\right) + X_i\beta}{\sigma}\right)$$

Combining (24) and (25) results in a standard probit model with X_i (including a constant) and ln $[(A_i-B_i)/B_i]$ as covariates. The estimated coefficient on X_i will be an estimate of β/σ and the estimated coefficient for ln $[(A_i-B_i)/B_i]$ will be an estimate of 1/σ. The unscaled βs can be recovered by dividing the estimates of β/σ by the estimated parameter 1/σ attached to the constructed variable ln $[(A_i-B_i)/B_i]$. The median WTP for each individual is then obtained by setting ϵ_i in (23) to zero because that is the value that splits the symmetric error distribution in half.

(26) $$Median(WTP_i) = \frac{A_i}{1 + e^{-X_i\beta}} = p(\varepsilon_i)\,A_i$$

Application of the Bounded Probit estimator to the Tietê data leads to the median calculations demonstrated in Tables 14 and 15, using individual household income for the upper limit.[23]

The first two columns of each table refer to estimation of a Probit probability model for each of the two sub-samples (Close, Far) where the dependent variable is 0 if the respondent rejected the survey offer (a "no") and 1 if it was accepted (a "yes"). The Probit parameter estimates are reported in the third column. In general (Maddala 1983, p. 23) they are measurable and estimable only up to a scalar (1/σ) but the model specification in this particular case provides an independent estimate of that scalar (see the Btrans variable row in the tables) that allows unscaled parameter estimates to be recovered. They are reported in the fourth column. The summed product of the untransformed parameters and the explanatory variables gives an estimate of the average value of the index function Xβ. Inserting that index function value in Eq. 26's expression $1/(1+e^{(-X\beta)})$ produces a median estimate of the fraction of income that would be offered to get the water quality improvements provided by the project (0.0021 for beneficiaries close to the river and 0.0005 for those living farther away). Multiplying the fraction by average income ("A" in Eq. 26) produces a

---

[23] At the request of a referee, similar mean and median calculations were done based on estimation of a Bounded Probit model imposing an upper limit on WTP at 20% of household income. Bounds much less than 20% could not be imposed using the full sample since in some cases the bid offered was around 18% of income, so going below that would involve a negative sign on the variable (A-B)/B, which has no logarithm. For the medians, not much was gained or lost by imposing the limit. The bounded median under a 20% of income constraint was $3.25 for the close-to-river group and $0.60 for those far away.

Bounded Probit estimate of Median (WTP). The results of this exercise are reported in the summary table in the next section where all the WTP estimates are collected.

There is no closed form analytical solution for the expected value of WTP in the bounded probit or logit formulation, so it must be found by numerical integration (Haab and McConnell, August 1997). The general form of expected willingness to pay is given by:

(27) $$E(WTP_i) = \int_{-\infty}^{\infty} WTP\,(X_i\,\beta, \epsilon)\, f(\epsilon)\, d\epsilon$$

The integral in (27) can be approximated by:

(28) $$E(WTP) \approx \sum_{k=1}^{n} (1/\sigma)\, \phi\left[\frac{\epsilon_k}{\sigma}\right] WTP(X_i\beta, \epsilon_k)(\epsilon_k - \epsilon_{k-1})$$

where $\phi(\bullet)$ is the standard normal pdf, ϵ_k are points on the distributional support of ϵ and n is large enough so that the approximation is smooth. We used 5000 points to apply (28), approximating $\phi(\bullet)$ by successive differences in the standard normal CDF, $\Phi(\bullet)$, a technique explained in Annex 1.[24] Table 16 immediately following the median calculations illustrates selected portions of the 5000 evaluation points used to get a numerical approximation to the Bounded Probit mean for the close-to-river group. Similar calculations (not shown) were done for the far-from-river group.

[24] Haab and McConnell (1997) provide a quick numerical approximation technique based on a few point estimates of the pdf that dispenses with setting up a large number of points, n, but is less smooth and hence less accurate than (28). Although we applied it to test whether our more exact approximation worked, it is not discussed here because the shortcut can be fairly imprecise if the range in the standard normal deviate, ε/σ, and the number of evaluation points are not properly chosen.

Table 14. Bounded Probit Median: Close to River Sub-Sample					
Limit=100% of Income (Mean Income = $1,524.39 Reals/Household/Month)					
Variable	Variable Definition	Original Probit Parameter Estimates[2] (β/σ)	Unscaled Parameter Estimates[3] (β)	Variable Means (X)	Variable Means *Unscaled Parameters (Xβ)
Constant		-1.3089	-5.5886 *	1	-5.5886
Status	1 if Upper;0 Else	0.2715	1.1592	0.147	0.1704
Age	Age of Household Head, Years	-0.0108	-0.0459 *	49.38	-2.2677
Btrans[1]	ln ((Income-Bid)/Bid)	0.2342	0.2342 *	5.324	n.a.
Barrio	1 if Close to River; 0 Else	0.3569	1.5237 *	1	1.0000
				Xβ=Column Sum	-6.1622
				Fraction of Income =1/(1+exp(-Xβ)	0.0021
				Median =Share*Income	R$3.21

Notes:

1 This is the bounding variable whose parameter estimate, 1/σ, is used to unscale the rest of the βs.
2 A * denotes significance at the 1% level or better.
3 Original parameter estimates divided by 1/σ, the parameter attached to Btrans.

Table 15. Bounded Probit Median: Far from River Sub-Sample					
Limit=100% of Income (Mean Income = $1,148.97 Reals/Household/Month)					
Variable	Variable Definition	Original Probit Parameter Estimates[2] (β/σ)	Unscaled Parameter Estimates[3] (β)	Variable Means (X)	Variable Means *Unscaled Parameters (Xβ)
Constant		-1.3089 *	-5.5886	1	-5.5886
Status	1 if Upper;0 Else	0.2715	1.1592	0.094	0.1090
Age	Age of Household Head, Years	-0.0108 *	-0.0459	44.34	-2.0363
Btrans[1]	ln ((Income-Bid)/Bid)	0.2342 *	n.a	5.324	n.a.
Barrio	1 if Close to River; 0 Else	0.3569 *	1.5237	0	0.0000
				Xβ=Column Sum	-7.5159
				Fraction of Income =1/(1+exp(-Xβ)	0.0005
				Median =Share*Income	R$0.63

Notes:

1 This is the bounding variable whose parameter estimate, 1/σ, is used to unscale the rest of the βs.
2 A * denotes significance at the 1% level or better.
3 Original parameter estimates divided by 1/σ, the parameter attached to Btrans.

Table 16. Numerical Approximation of Bounded Probit Mean WTP: Close to River Sub-Sample								
Location of Median	Step #	Standard Normal Deviate (ε/σ)	Error (ε)	Cumulative Normal Density, CDF Φ(ε/σ)	Approximate pdf f(ε/σ)≈ΔΦ(ε/σ)	-Xβ-ε	WTP Ratio, R 1/(1+e(-XB-ε))	Product of WTP Ratio*pdf ΔΦ(ε/σ)*R
	1	-6.0000	-25.618	9.8659e-10				
	2	-5.9976	-25.608	1.0013e-09	1.4691e-11	31.7702	0.0000e+00	2.3412e-25
	•	•	•	•	•	•	•	•
	417	-5.0014	-21.354	2.8458e-07	3.5227e-09	27.5167	1.1211e-12	3.9495e-21
	418	-4.9990	-21.344	2.8814e-07	3.5653e-09	27.5064	1.1327e-12	4.0384e-21
	•	•	•	•	•	•	•	•
	834	-4.0004	-17.081	3.1618e-05	3.1921e-07	23.2427	8.0504e-11	2.5698e-17
	835	-3.9980	-17.070	3.1940e-05	3.2229e-07	23.2325	8.1334e-11	2.6213e-17
	•	•	•	•	•	•	•	•
	1250	-3.0018	-12.825	1.3419e-03	1.0543e-05	18.9868	5.6770e-09	5.9854e-14
	1251	-2.9994	-12.807	1.3526e-03	1.0619e-05	18.9687	5.7807e-09	6.1388e-14
	•	•	•	•	•	•	•	•
	1667	-2.0008	-8.543	0.022707	1.2909e-04	14.7050	4.1086e-07	5.3036e-11
	1668	-1.9984	-8.533	0.022837	1.2971e-04	14.6948	4.1509e-07	5.3841e-11
	•	•	•	•	•	•	•	•
	2083	-1.0022	-4.279	0.158123	5.7887e-04	10.4413	2.9201e-05	1.6903e-08
	2084	-0.9998	-4.269	0.158704	5.8026e-04	10.4310	2.9501e-05	1.7119e-08
	•	•	•	•	•	•	•	•
Median R ->	2500	-0.0012	-0.005	0.499521	9.5765e-04	6.1673	0.00209	2.0038e-06
	2501	0.0012	0.005	0.500479	9.5765e-04	6.1571	0.00211	2.0244e-06
	•	•	•	•	•	•	•	•
	2917	0.9998	4.269	0.841296	5.8166e-04	1.8934	0.13086	7.6117e-05
	2918	1.0022	4.279	0.841877	5.8026e-04	1.8831	0.13203	7.6614e-05
	•	•	•	•	•	•	•	•
	3333	1.9984	8.533	0.977163	1.3033e-04	-2.3704	0.91454	1.1919e-04
	3334	2.0008	8.543	0.977293	1.2971e-04	-2.3806	0.91534	1.1873e-04
		•	•	•	•	•	•	•
	3750	2.9994	12.807	0.998647	1.0696e-05	-6.6443	0.99870	5.5230e-05
	3751	3.0018	12.817	0.998658	1.0619e-05	-6.6546	0.99871	1.0606e-05
	•	•	•	•	•	•	•	•
	4166	3.9980	17.070	9.9997e-01	3.2540e-07	-10.9081	9.9998e-01	3.2539e-07
	4167	4.0004	17.081	9.9997e-01	3.2229e-07	-10.9183	9.9998e-01	3.2228e-07
	4583	4.9990	21.344	1.0000e+00	3.6083e-09	-15.1820	1.0000e+00	3.6083e-09
	4584	5.0014	21.354	1.0000e+00	3.5653e-09	-15.1923	1.0000e+00	3.5653e-09
	•	•	•	•	•	•	•	•
	4999	5.9976	25.608	1.0000e+00	1.4904e-11	-19.4458	1.0000e+00	1.4904e-11
	5000	6.0000	25.618	1.0000e+00	1.4691e-11	-19.4560	1.0000e+00	1.4691e-11
				Grand Totals:	1.0000e+00			0.0919
	E(WTP) = ∑ΔΦ(ε/σ) * R * Income = 0.0919 * R$1524:							**R$140.10**

Note: A • indicates intervening calculations that are not shown. Givens for the approximation are 5000 evaluation points; an ε/σ range from -6 to +6; a step size (Δ(ε/σ) of 0.0024, a standard deviation (σ) of 4.2697 (Table 14, reciprocal of the Btrans parameter), and an index value (-Xβ) of 6.1622 (Table14).

The Bounded Probit mean results (R$140.10 and R$60.46 for households close to and far from the river, respectively) are completely inconsistent with all that has come before, being more than a factor of ten greater than the highest of all of the preceding estimates, and 45 and 100 times larger than their respective close and far-from-river sub-sample Bounded Probit medians.[25]

[25] While this phenomenon might be an artifact of one or more mistakes in setting up the approximation, we were able to replicate all of the examples given in Haab and McConnell (1997) successfully. In addition, in the example in Table 7 of their paper, the Bounded Probit mean exceeds the median by a factor of 38, which is similar to what happens with the Tietê data. Reference to Table11 shows that the median ratio is properly located, but the distribution is heavily skewed. Imposing a bound on median WTP at 20% of income brought the near and far means down to $50.05 and $25.54 which are only slightly more plausible. Some doubt about the usefulness of the Bounded Probit mean (but not the median) in CB analysis is probably warranted.

Section VI: Central Tendency Measures Results and its Effect on the Project's Net Present Value

Uncertainty about WTP need not translate into uncertainty about a project's net benefits. If the analysis decision is unambiguous because a project's net present value (NPV) is either consistently positive or consistently negative across the plausible range of possible WTP estimates, then any one of them will suffice. But it is impossible to establish in advance, without actually doing the exercise, that a given project analysis decision will be impervious to variations in the central tendency measure of WTP, making the choice of measure a matter of indifference. On the contrary, uncertainty about benefits is a project-specific issue, and has to be handled on a case-by-case basis. Robustness is likely to be the exception rather than the rule.

The second and third columns in Table 17 depicted below collect all of the central tendency measures calculated from the Tietê project's WTP survey data. Sorting them from high to low in the near-to-river category confirms rather dramatically the introductory warning that a wide range of plausible estimates can be extracted from referendum data. Even disregarding the bounded Probit mean, the highest near-to-river WTP exceeds the lowest by a factor of four, and the factor is ten for the far-from river estimates.

The next four columns of the table show, in deterministic sensitivity fashion, the effect that using each of the alternative WTP measures would have on the economic feasibility of the project at issue, expressed in terms of net present value (NPV) using a twelve percent interest rate. In general, under optimistic assumptions about execution timing and the earliest possible manifestation of energy benefits[26] (the "best case" scenario), the project decision is not severely affected by the wide variety of per household benefit measures available to appraise it in this particular case. Under the most optimistic of assumptions the project as a whole (Stages I, II and III) is not viable except under the Bounded Probit mean benefit measure, and barely under Kriström's nonparametric mean, while the incremental project (Stages II and III) that treats Stage I costs as sunk is economically justified for all but the lowest WTP measure. Said otherwise, if the

[26] These benefits arise from resuming the use of water from the Tietê for hydro-electric generation after transfer to a different sub-basin. This use had been suspended because the low quality of the Tietê water was degrading the reservoir into which the Tietê was diverted.

initial conditions were set optimistically and the problem posed to different analysts each using a different WTP measure, the final conclusion would be near unanimous and unaffected by the measure chosen.

The apparent absence of a grey area or zone of ambiguity in the incremental project appraisal decision vanishes when the initial conditions are set less favorably (the "worst case" scenario in the table). While the project as a whole gets even worse and is consistently rejected, the once favorable decision on the configuration of Stages II and III becomes cloudier if the execution period is extended over fifteen years rather than completed in ten and if energy benefits do not materialize at all. Then, the final column of the table shows that the incremental project only looks economically feasible for six of the measures, mostly means, and is infeasible (negative NPV) for the other five, which are mainly medians of one sort or another. This result demonstrates another remark made early-on about the implications of using the mean rather than the median — the former will generally produce a more favorable outcome with WTP distributions that are skewed to the right.

Table 17. Cost-Benefit Comparisons

Central Tendency Measure	WTP per Household per Month (*1998 Reals*)		Net Present Value (Million Reals)			
			Scenario & Project Stages			
	Close	Far	Best Case I, II & III	Best Case II&III	Worst Case I, II & III	Worst Case II & III
Bounded Probit Mean, Limit =100% of Income	140.10	60.46	11,220	11,953	6,235	6,965
Truncated Mean, E(C), 0 < C < ∞ (utility difference logit, linear in bid, C)	9.73	6.47	(6)[c]	726	(394)	334
Kriström's Nonparametric Mean	9.42	7.09	0.1	733	(390)	338
Truncated Mean, E(C), $0<C<B_{max}$ (utility difference logit, linear in bid)	7.65	5.27	(200)	532	(509)	220
Turnbull Nonparametric Lower Bound Mean	6.07	4.51	(341)	391	(592)	136
Truncated Mean, Log Transform, ∞UL (utility difference logit, log of bid)	5.01	1.68	(538)	194	(709)	19

Central Tendency Measure	WTP per Household per Month (*1998 Reals*)		Net Present Value (Million Reals)			
			Scenario & Project Stages			
	Close	Far	Best Case I, II & III	Best Case II&III	Worst Case I, II & III	Worst Case II & III
Untruncated Mean, E(C)= Median, - ∞ < C < ∞ (utility difference logit, linear in bid, C)	4.74	–1.27	(631)[b]	101	(765)	(35)
Truncated Mean, Log Transform, Income UL (utility difference logit, log of bid)	3.71	1.39	(639)	93	(769)	(39)
Nonparametric Median (Linear Interpolation)	3.33	1.83	(645)	87	(772)	(43)
Bounded Probit Median, Limit =100% of Income	3.21	0.63	(707)	25	(809)	(79)
Truncated Median, Log Transform (utility difference logit, log of bid)	2.36	0.71	(761)	(28)	(841)	(111)

Notes:

a. The "Best Case" sets the construction period to 5 years each for Stages II and III, and has energy benefits on line in the first year after Stage II is built. The "Worst Case" sets the execution period to 10 years for Stage II and 5 years for Stage III, and assumes no energy benefits come on line over a 30-year horizon.
b. Far-from-river WTP arbitrarily set to zero to compute NPV
c. Amounts in parenthesis denote negative amounts.

However, the median measure only indicates the price at which a project proposal would be accepted by a majority vote under a one-person, one-vote rule. If the project's NPV is negative using the median, that does not necessarily imply it is not worth doing from a social welfare standpoint. Aggregating up using the mean to get total benefits is more consistent with standard cost-benefit practice where the "votes" are in monetary units, and outliers with high willingness to pay count in the calculation of the ability of the winners to compensate the losers and still come out ahead (McFadden and Leonard 1993, p. 193).

There is no golden rule for resolving ambiguities about project approval brought on by uncertainty about the central tendency measure of willingness to pay except, perhaps, to be aware of this source of uncertainty and to explicitly acknowledge it rather than ignore or conceal it. At a minimum, a search for the existence of a grey area should be conducted. If the project is either economically unjustified using the highest of all legitimate benefit measures or justified using the lowest among the candidates, all the better because benefits uncertainty is demonstrably not an issue.

If, on the other hand, the project acceptance decision is reversed somewhere along the spectrum of possible measures, there are several simple decision rules that could be applied, including picking the greatest WTP to push the project ahead and avoid controversy, choosing a measure somewhere in the middle of the range to impart some balance to the final recommendation, or taking a conservative posture by selecting a measure at the low end. A more sophisticated approach would be to fold all of the empirical distributions of the expected value measures together, either with equal probability of drawing from each (akin to picking something in the middle) or with unequal weights reflecting the analyst's judgement or confidence.[27] Finally, one could try to argue for a specific choice on theoretical or econometric grounds, although abstruse technical explanations are unlikely to be popular with decision makers who are ultimately responsible for financing multi-million dollar projects.

Looking at the preceding table, it would be prudent to discard the Bounded Probit and the Untruncated Rum means–the former is ridiculously high and the latter is theoretically inconsistent and ridiculously negative. The choice between means and medians is philosophical; choosing a mean is consistent with standard aggregation practice in CB. Eschewing the medians and moving on to the remaining means, the Kriström nonparametric mean is too heavily influenced by tail value assumptions to be reliable in this case. After this process of elimination, the remaining means are all legitimate contenders. If one had to chose a single measure, a reasonable choice would be the Turnbull expected value because it is a conservative lower bound measure that in this case falls in the middle of the pack.

[27] Ardila (1993) and Hazilla (forthcoming) show how empirical distributions of mean WTP can be generated, given knowledge of the variances and covariances of the statistically estimated parameter estimates that appear in the E(WTP) formulas.

Section VII: Conclusions and Recommendations

We have to begin with a simple question: How valuable is economic valuation? In this case that million dollar question has a half billion dollar answer. A cost benefit analysis for the full project was not performed at the onset. If it had been, the project may not have ever been undertaken and the borrower could have avoided a potentially large social loss. Even in the face of severe uncertainty about benefits, in hindsight its costs appear too great to make it economically attractive: it's NPV is unambiguously negative under almost all conceivable circumstances.

However, the prospect of a major water-borne disease outbreak if nothing were done to correct the pollution situation was not reflected in the original analysis or explicitly emphasized in the CV exercise done to estimate benefits for Stages II and III, assuming Stage I benefits were nil. The likelihood of such an event is not known, nor is the extent to which CV respondents considered health issues in formulating their answers. Finally there may have been non-economic reasons for trying to reduce pollution in this major metropolitan area which influenced the decision to go ahead. The project was approved around the time of the heavily publicized United Nations Conference on Environment and Development held in Rio de Janeiro, Brazil in 1992, an event which raised expectations and encouraged countries to pursue sustainable development following the principles of *Agenda 21*.

Whatever the pressures surrounding the project were, cost-effectiveness analysis was used to justify going ahead with the first stage of investment, on the unproven assumption that the entire project was economically viable. This decision locked the borrower into contractual obligations for operation of the first stage facilities, repayment of the first stage loan, and continuance of the pollution control program. Now, paradoxically, if the entire effort cannot be abandoned the next-best thing to do is to go ahead and complete the program.

We suggest that when millions of dollars are at stake in water pollution control program investments, cost-benefit analysis is definitely worth the effort, even though the benefits of water quality improvement are hard to pin down.

The use of contingent valuation (CV) methods to estimate benefits has become increasingly common in project analysis. Ever since the NOAA Blue Ribbon Panel Report in 1993 (NOAA, 1993) recommended the use of the referendum form of CV, it seems to have become the method of choice in practical settings.

Referendum-type questions are thought to be easier to answer than the open-ended variety. But there is a downside: econometric techniques must be applied to the referendum data in order to infer the mean or median willingness to pay (WTP) of the sample and, thus, of the population of potential beneficiaries.

This is not, however, just a technical point. Its implications are demonstrated with data obtained from a referendum CV study done for a proposed sewer and wastewater treatment project designed to improve water quality in the Tietê River flowing through the city of São Paulo, Brazil. The results show that:

- A factor of 4 separates lowest from highest central tendency estimates of WTP, ignoring one implausible outlier that is 14 times larger than the largest of the other figures.

- This variation is ample enough to make a difference in the cost-benefit analysis results for the project under conservative assumptions.

Analysts that use referendum CV data must be sensitive to the problems they buy into, and decide how to deal with the resulting benefits uncertainty in their project analysis[28]. If the principal use of CV survey data is to produce a mean or median estimate of WTP for Cost-Benefit analysis rather than to test for the factors influencing referendum choice responses and, by implication, WTP, nonparametric approaches have the advantage of simplicity over parametric approaches.

No code of silence has been broken here by revealing the uncertainty inherent in referendum CV estimates of WTP — the academic literature, particularly of late, has covered the issue in some depth and many

[28] An important issue that cannot be omitted but that it could well be the area of further research is that of the costs of achieving accuracy when working with any of the methods described in this paper. In the case of CVM, the standard error of the expected value of the mean reflect the accuracy of the method. A large variance would signify not so robust results. In order to achieve greater accuracy, the standard error needs to be reduced and this entails conducting additional interviews in order to increase the sample size, hence, decreasing the variance. Our experience with the project presented in this paper is that in Brazil, the marginal cost of an additional interview is approximately $100. Therefore, the cost of decreasing the variance can be large, not justifying the additional investment. The optimization problem becomes then on estimating the number of additional interviews that would minimize the opportunity loss of the investment.

experienced project analysts are probably well aware of it. Yet that literature is at times inaccessible and hard to understand, and no synthesis exists emphasizing the implications of using these several CV measures in investment project appraisal. Therefore, the main purpose of this paper has been to explain, in simple terms using worked examples, the nature of the problem and the solutions available to everyday practitioners.

What practical recommendations can be made? The most obvious would seem to be:

- Do an open-ended survey at the pre-test stage to get an idea of the bid range to use in a full-blown referendum survey and produce a tentative benchmark WTP from the open-ended data to compare against.

- Design the referendum to cover the bid range so nonparametric means and medians can be computed reliably. Make sure the sample is representative of the population, and does not involve oversampling of selected socioeconomic groups or geographical areas. Monitor the survey results, perhaps executing it in phases, so adjustments in the bid range can be made if coverage deficiencies become apparent.

- Run a battery of central tendency measures, definitely including a nonparametric measure and perhaps including the bounded Probit median, rather than arbitrarily picking one or two of the more familiar parametric measures.

- Explore the influence of the several WTP measures on the cost-benefit analysis outcome, looking for the existence or absence of the uncertain grey area.

- Reach a reasoned final recommendation about project feasibility based on the above, and be able to explain it.

In sum, before becoming completely and inextricably caught up in the fine points of econometric estimation of parametric choice models it is worth pausing to consider the options available and the point of the exercise. If the primary goal is to explain and understand respondent behavior, verify whether CV survey responses are consistent with economic theory, or estimate WTP for a population other than the one sampled, parametric choice models must be estimated. If all one needs is a benefit measure for CB analysis, on the other hand, nonparametric estimates of WTP may have the edge. McFadden and Leonard (1993, pp. 167-168) summarize the advantages and disadvantages of each route:

> *...direct approaches to valuing a resource do not require any parameterization of preferences or the distribution of tastes, and do not require that WTP be related to any consumer characteristics such as age or income, because the final impact of these variations is taken care of by random sampling from the population...The advantages of parametric methods are that they make it relatively easy to impose preference axioms, pool data across experiments, and extrapolate the calculations of value to different populations than the sampled population. Their primary limitation is that, if the parameterization is not flexible enough to describe behavior, then the misspecification will usually cause the mean WTP calculated from the estimated model to be a biased estimate of true WTP*.

References

Ajzen, Icek, and George L. Peterson.1988. "Contingent Value Measurement: The Price of Everything and the Value of Nothing?" in *Amenity Resource Valuation: Integrating Economics with other Disciplines*. George L. Peterson, B.L. Driver, and Robin Gregory, eds. State College: Ventura Publishing, Inc.

Adamowicz, Wiktor et al. February 1998. "Sated Preference Approaches for Measuring Passive Use Values: Choice Experiments and Contingent Valuation." *American Journal of Agricultural Economics* 80, pp. 64-75.

Alberini, Anna, Barbara Kanninen, and Richard Carson. August 1997. "Modeling Response Incentive Effects in Dichotomous Choice Contingent Valuation Data." *Land Economics* 73(3), pp. 309-24.

Amemiya, T. December 1981. "Qualitative Response Models: A Survey." *Journal of Economic Literature* Vol. XIX, pp. 1483-1536.

Anderson, Joch R. and Enrico Porceddu. December 1995. "Agricultural Economics and the Environment." *Rivista di Economia Agraria* nn. 3-4, pp. 447-453.

Ardila, Sergio, Ricardo Quiroga, and William J. Vaughan. 1998. "A Review of the Use of Contingent Valuation Methods in Project Analysis at the Inter-American Development Bank." Unpublished paper to be presented at the National Science Foundation Workshop on Alternatives to Traditional CVM in Environmental Valuation, Vanderbilt University, Nashville Tennessee, October 15-16, 1998.

Ardila Sergio. Diciembre 1993. "Guía para la utilizacíon de modelos econométricos en aplicaciones del método de valoración contingente". División de Protección del Medio Ambiente Documento de Trabajo ENP101. Washington, D.C.: Banco Interamericano de Desarollo.

Bishop, Richard C., Patricia A. Champ, and Daniel J. Mullarkey. 1995. "Contingent Valuation." in *The Handbook of Environmental Economics*. Daniel W. Bromley, ed. Blackwell Handbooks in Economics.

Blamey, R. K., J. W. Bennett, and M. D. Morrison. February 1999. "Yea-Saying in Contingent Valuation Surveys." *Land Economics* 754 (1): 126-141.

Brajer, Victor and Ronald G. Cummings. 1988. "Integrating Social Science Research in the Contingent Valuation Method: An Overview." in *Amenity Resource Valuation: Integrating Economics with other Disciplines*. George L. Peterson, B.L. Driver, and Robin Gregory, eds. State College: Ventura Publishing, Inc.

Boardman, Anthony E., David H. Greenberg, Aidan R. Vining, and David L. Weimer. 1996. *Cost-Benefit Analysis: Concepts and Practice*. Upper Saddle River: Prentice Hall, Inc.

Boman, Mattias, Göran Bostedt, and Bengt Kriström. May 1999. "Obtaining Welfare Bounds in Discrete-Response Valuation Studies: A Non-Parametric Approach." *Land Economics* 75 (2): 284-294.

Boyle, Kevin J. et. al. February 1998. "Bid Design and Yea Saying in Single-Bounded, Dichotomous-Choice Questions." *Land Economics* 74(1), pp. 49-64.

Boyle, Kevin J., F. Reed Johnson, and Daniel W. Mc.Collum. 1997. "Anchoring and Adjustment in Single-Bounded, Contingent Valuation Questions." *American Journal of Agricultural Economics* 79, pp. 1495-1500.

Breen, Richard. 1996. Regression Models: Censored, Sample Selected, or Truncated Data. Series Number 07-111. Series: Quantitative Applications in the Social Sciences. Thousand Oaks, California: Sage Publications, Inc.

Brent, Robert J. 1996. *Applied Cost-Benefit Analysis*. Brookfield: Edward Elgar Publishing Company.

Carson, Richard. 1997. "Contingent Valuation: Theoretical Advances and Empirical Tests since the NOAA Panel." *American Journal of Agricultural Economics* 79, pp. 1501-1507.

Carson, Richard *et al*. 1996. "Was the NOAA Panel Correct about Contingent Valuation?" Discussion Paper 96-20. Washington, D.C.: Resources for the Future.

Carson, Richard *et al*. 1995. "Referendum Design and Contingent Valuation: The NOAA Panel's No-Vote Recommendation." Discussion Paper 96-05. Washington, D.C.: Resources for the Future.

Carson, Richard. 1991. "Constructed Markets." Chapter 5 in *Measuring the Demand for Environmental Quality* by John B. Braden and Charles D. Kolstad (Editors). Elsevier Science Publishers B.V. (North Holland).

Carlson, Gerald A., David Zilberman, and John A. Miranowski. 1993. *Agricultural and Environmental Resource Economics*. New York: Oxford University Press.

Chen, Heng Z., and Stephen R. Cosslett. August 1998. "Environmental Quality Preference and Benefit Estimation in Multinomial Probit Models: A Simulation Approach." *American Journal of Agricultural Economics* 80, pp. 512-520.

Cooper, Joseph C. February 1994. "A Comparison of Approaches to Calculating Confidence Intervals for Benefit Measures from Dichotomous Choice Contingent Valuation Surveys." *Land Economics* 70 (1): 111-122.

Cornes, Richard, and Todd Sandler. 1986. *The Theory of Externalities, Public Goods and Club Goods*. New York: Cambridge University Press.

Creel, Michael. 1998. "A Note on Consistent Estimation of Mean WTP Using a Misspecified Logit Contingent Valuation Model." *Journal of Environmental Economics and Management* 35, pp. 277-284.

Darling, Arthur C. Principal Economist. Region 1 Environment and Natural Resources Division, Inter-American Development Bank. Personal Communication, Various Dates.

Diaconis, Persi and Bradley Efron. May 1983. "Computer-Intensive Methods in Statistics." *Scientific American*. pp. 116-130.

Diamond, Peter A. and Jerry A. Hausman. 1994. "Contingent Valuation: Is Some Number Better than No Number?" *Journal of Economic Perspectives* Volume 8, Number 4, pp. 45-64.

Ducci, Jorge. Noviembre 1988. "Metolodía de Cuantificación de Beneficios: Uruguay, Sanemiento Ambiental de Montevideo". Anexo Técnico. Unpublished Background Paper. Washington, D.C.: Inter-American Development Bank.

Fischhoff, Baruch. 1988. "Specifying Value Measurements." in *Amenity Resource Valuation: Integrating Economics with other Disciplines*. George L. Peterson, B.L. Driver, and Robin Gregory, eds. State College: Ventura Publishing, Inc.

Fomby, Thomas B., R. Carter Hill and Stanley R. Johnson. 1984. *Advanced Econometric Methods*. New York: Springer-Verlag.

Freeman, III, A.M. 1993. *The Measurement of Environmental and Resource Values: Theory and Methods*. Washington, D.C.: Resources for the Future.

Graham-Tomasi, Theodore. 1995. "Quasi-Option Value." in *The Handbook of Environmental Economics*. Daniel W. Bromley, ed. Blackwell Handbooks in Economics.

Greene, W. H. 1997. *Econometric Analysis*. New Jersey: Prentice Hall.

Greene, W. H. 1995. *LIMDEP. Version 7.0. User's Manual.* Australia: Econometric Software, Inc.

Greene, W. H. 1990. *Econometric Analysis.* New Jersey: Prentice Hall.

Gregory, Robin, and Richard C. Bishop. 1988. "Willingness to Pay or Compensated Demanded: Issues in Applied Resource Valuation." in *Amenity Resource Valuation: Integrating Economics with other Disciplines*. George L. Peterson, B.L. Driver, and Robin Gregory, eds. State College: Ventura Publishing, Inc.

Griffiths, William E., Carter Hill, R. and Judge, George C. 1993. *Learning and Practicing Econometrics*. New York: John Wiley & Sons, Inc.

Gujarati, Damodar N. 1995. *Basic Econometrics*. New York: McGraw-Hill, Inc.

Haab, T.C. and McConnell, K.E. July 1998. "Simple Bounds for Willingness to Pay Using a Probit or Logit Model." Unpublished.

Haab, T.C. and Kenneth E. McConnell. May 1998. "Referendum Models and Economic Values: Theoretical, Intuitive, and Practical Bounds on Willingness to Pay". *Land Economics* Vol. 74, No. 2, pp. 216-19.

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

Haab, T.C. and Kenneth E. McConnell. August 1997. "A Simple Method for Bounding Willingness to Pay Using a Probit or Logit Model. Unpublished.

Halvorsen, Bente, and Kjartan Soelensminde. May 1998. "Differences between Willingness-to-Pay Estimates from Open-Ended and Discrete-Choice Contingent Valuation Methods: The Effects of Heteroscedasticity." *Land Economics* 74(2), pp. 262-282.

Hanemann, W. Michael and Barbara Kanninen. March 1996. "The Statistical Analysis of Discrete-Response CV Data." To appear in *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Methods in the US, EC, and Developing Countries*. Ian Bateman and Ken G. Willis, Editors. Oxford: Oxford University Press.

Hanemann, W. Michael. 1994. "Valuing the Environment Through Contingent Valuation." *Journal of Economic Perspectives* Volume 8, Number 4, pp. 19-43.

Hanemann, W. Michael, J. Loomis and B. Kanninen. 1991. "Statistical Efficiency of Double-Bounded Dichotomous Choice Contingent Valuation". *American Journal of Agricultural Economics* Vol. 71, No. 4. pp 1255-63.

Hanemann, W. Michael. August 1984. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses." *American Journal of Agricultural Economics* pp. 332-340.

Harberger, Arnold. 1996. "Reflections on Social Project Evaluation." in *The Economic Evaluation of Projects: Papers for a Curriculum Development Workshop*. David G. Davies, ed. Washington, D.C.: The World Bank.

Harberger, Arnold. 1996. "Project Evaluation for the Next Decade." in *The Economic Evaluation of Projects: Papers for a Curriculum Development Workshop*. David G. Davies, ed. Washington, D.C.: The World Bank.

Harris, Charles C., Howard E. A. Tinsley, and Dennis M. Donnelly. 1988. "Research Methods for Public Amenity Resource Valuation: Issues and Recommendations." in *Amenity Resource Valuation: Integrating Economics with other Disciplines*. George L. Peterson, B.L. Driver, and Robin Gregory, eds. State College: Ventura Publishing, Inc.

Harrison, Glenn W. and James Lesley. 1996. "Must Contingent Valuation Surveys Cost So Much?" *Journal of Environmental Economics and Management* 31, 79-95.

Harrison, Glenn W. 1992. "Valuing Public Goods with the Contingent Valuation Method: A Critique of Kahneman and Knetsch." *Journal of Environmental Economics and Management* 23, pp. 248-257.

Hausman, Jerry A., Gregory K. Leonard, and Daniel McFadden. "Assessing Use Value Losses Caused by Natural Resource Injury." in *Contingent Valuation: A Critical Assessment*. J. A. Hausman (Editor). Elsevier Science Publishers B.V.

Hazilla, Michael. Forthcoming 1999. "Econometric Methods for Measurement and Valuation of Environmental Quality". Technical Paper, Environment Division, Sustainable Development Department. Washington D.C.: Inter-American Development Bank.

Hazilla, Michael. June 1999. "An Introduction to Operational Contingent Valuation Methods." Paper presented at the World Bank Teaching Clinic on Contingent Valuation held 2 June 1999. Washington, D.C.

Herriges, Joseph A., and Catherine L. Kling. August 1997. "The Performance of Nested Logit Models When Welfare Estimation is the Goal." *American Journal of Agricultural Economics* 79, pp. 792-802.

Hoehn, John P. and Alan Randall. June 1989. "Too Many Proposals Pass the Benefit Cost Test." *The American Economic Review* Vol. 79 No. 3. pp. 544-551.

Hoehn, John P., and Cindy Sorg Swanson.1988. "Toward a Satisfactory Model of Contingent Valuation Behavior in a Policy Valuation Context" in *Amenity Resource Valuation: Integrating Economics with other Disciplines*. George L. Peterson, B.L. Driver, and Robin Gregory, eds. State College: Ventura Publishing, Inc.

Hoehn, John P., and Alan Randall. 1987. "A Satisfactory Benefit Cost Indicator from Contingent Valuation." *Journal of Environmental Economics and Management* 14, pp. 226-247.

Hoevenagel, Ruud. 1994. "An Assessment of the Contingent Valuation Method." in *Valuing the Environment: Methodological and Measurement Issues*. Netherlands: Kluwer Academic Publishers.

Huang, Ju-Chin, Douglas W. Nychka, and V. Kerry Smith. December 21, 1995. "Nonparametric Discrete Choice Methods for Measuring Economic Values." Unpublished.

Hulett, David T. 1999. "Project Cost Risk Analysis Using Crystal Ball." Hulett & Associates.

Inter-American Development Bank. July 1998. "1997 Annual Report on Projects in Execution." Internal document.

Inter-American Development Bank. July 1997. "Annual Report on Portfolio Management for 1996." Internal document.

Inter-American Development Bank. 1980. "PR-302: Operational Policy." Internal document.

Jenkins, Glenn P. May 1997. "Project Analysis and the World Bank." *American Journal of Agricultural Economics* Vol. 87 No. 2, pp. 38-42.

Johansson, Per-Olov. 1993. *Cost-Benefit Analysis of Environmental Change*. New York: Cambridge University Press.

Johansson, Per-Olov. 1991. *An Introduction to Modern Welfare Economics*. New York: Cambridge University Press.

Johansson, Per-Olov, Bengt Kriström, and Karl Goran Mäler. 1989. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Response Data: Comment." *American Journal of Agricultural Economics.* Vol. 71. Pp. 1054-1056.

Kennedy, Peter. 1992. *A Guide to Econometrics*. Cambridge: The MIT Press.

Kmenta, Jan. 1971. *Elements of Econometrics*. New York: Macmillan Publishing Co., Inc.

Kopp, Raymond J., Alan Krupnick, and Michael Toman. 1997. "Cost-Benefit Analysis and Regulatory Reform: An Assessment of the Science and the Art." Discussion Paper 97-19. Washington, D.C.: Resources for the Future.

Kravchuck, Robert S. Winter 1989-90. "A Footnote to Cost-Benefit Analysis Applied Under Conditions of Radical Ignorance." *Policy Studies Journal* Vol. 18, No. 2, pp. 326-341.

Kriström, Bengt. 1990. "A Nonparametric Approach to the Estimation of Welfare Evaluations in Discrete Response Valuation Studies." *Land Economics*. Vol. 66, No. 2, (May) pp. 135-139.

Langford, Ian H. et. al. February 1998. "Improved Estimation of Willingness to Pay in Dichotomous Choice Contingent Valuation Studies." *Land Economics* 74(1), pp. 65-75.

Lazo, Jeffrey K., Gary H. McClelland, and William D. Schulze. August 1997. "Economic Theory and Psychology of Non-Use Values." *Land Economics* 73 (3), pp. 358-371.

Liao, Tim Futing. 1994. *Interpreting Probability Models: Logit, Probit, and Other Generalized Linear Models*. Series Number 07-101. Series: Quantitative Applications in the Social Sciences. Thousand Oaks, California: Sage Publications, Inc.

Londero, Elio, Editor. 1992. *Precios de cuenta: Principios, metodología y estudios de caso*. Washington, D.C.: Banco Interamericano de Desarrollo.

Mackenzie, John. August 1993. "A Comparison of Contingent Preference Models." *American Journal of Agricultural Economics* 75 pp. 593-603.

Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Econometric Society Monographs No. 3. New York: Cambridge University Press.

Mariño, Manuel and John Boland. 1998. "Deciding Where, When and How Much to Invest in Wastewater Treatment: An Integrated Approach." A paper prepared for the Water and Sanitation Division. Transport, Water and Urban Development Department. World Bank.

Markandya, A. 1991. "The Economic Appraisal of Projects: The Environmental Dimension." Project Analysis Department. Washington, D.C.: The Inter-American Development Bank.

McClave, James T., Frank H. Dietrich, II, and Terry Sincich. 1997. *Statistics*. Upper Saddle River: Prentice Hall, Inc.

McConnell, K.E. 1997. "Does Altruism Undermine Existence Value?" *Journal of Environmental Economics and Management* 32, pp. 22-37.

McConnell, K.E. 1995. "Issues in Estimating Benefits with Non-Market Methods". Office of the Chief Economist. Working Paper Series 308. Inter-American Development Bank. Washington, D.C.

McConnell, K.E. and Ducci Jorge H. 1989. "Valuing Environmental Quality in Developing Countries: Two Case Studies". Unpublished paper prepared for AERE Session on Contingent Valuation Surveys in Developing Countries. December 29, 1989. Atlanta, Georgia.

McConnell, K.E. 1988. "Introducing Referendum Models". Unpublished paper prepared for IDB Workshop on Valuation Techniques in Project Analysis. November 21, 1988. Washington, D.C.

McConnell, K.E.1990. "Models for Referendum Data: The Structure of Discrete Choice Models for Contingent Valuation." *Journal of Environmental Economics and Management* 18, pp. 19-34.

McFadden, Daniel. November 1994. "Contingent Valuation and Social Choice." *American Journal of Agricultural Economics* 76, pp. 689-708.

McFadden, Daniel and Gregory K. Leonard. 1993. "Issues in the Contingent Valuation of Environmental Goods: Methodologies for Data Collection and Analysis." In *Contingent Valuation: A Critical Assessment*. J. Hausmann (ed.). Amsterdam: Elsevier. Pp. 165-215.

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

Mitchell, Robert C. 1988. "Evaluating the Validity of Contingent Valuation Studies." in *Amenity Resource Valuation: Integrating Economics with other Disciplines*. George L. Peterson, B.L. Driver, and Robin Gregory, eds. State College: Ventura Publishing, Inc.

Mitchell, Robert C. April 1987. "How Far Along the Learning Curve is Contingent Valuation?". Discussion Paper QE 87-07. Washington, D.C.: Resources for the Future.

Nardini, Andrea. September 1997. "A Proposal for Integrating Environmental Impact Assessment, Cost-Benefit Analysis and Multicriteria Analysis in Decision-Making." *Project Appraisal* Volume 12, Number 3, pp. 173-184.

Newbold, Paul. 1994. *Statistics for Business & Economics*. Fourth Edition. Upper Saddle River: Prentice Hall, Inc.

Niklitschek, Mario and Javier León. 1996. "Combining Intended Demand and Yes/No Responses in the Estimation of Contingent Valuation Models." *Jounrnal of Environmental Economics and Management* 31, 387-402.

NOAA, National Oceanic and Atmospheric Administration. U.S. Department of Commerce. 1993. "Natural Resource Damage Assessments Under the Oil Pollution Act of 1990." *Federal Register*, 58(10), (January 15), pp. 4601-4614.

Novara, Juan. 1990. Internal Memorandum: Request for Suggestions for the Review of the Economic Analysis of the Bank Operations. Inter-American Development Bank.

Ongley, Edwin D. and William G. Booty. 1999. "Pollution Remediation Planning in Developing Countries: Conventional Modelling versus Knowledge-Based Prediction." *Water International*. (in press).

Ongley, Edwin D. 1998. "Modernization of water quality programmes in developing countries: Issues of relevancy and cost efficiency. *Water Quality International* Sept/Oct. 1998: 37-42.

Ott, Wayne R. 1995. *Environmental Statistics and Data Analysis*. Boca Raton: CRC Press, Inc.

Paffenberger, Roger C. and James H. Patterson. 1987. *Statistical Methods for Business and Economics*. 3rd Edition. Homewood: Richard D. Irwin.

Pearce, David W., (ed). 1995. *The MIT Dictionary of Modern Economics.* Fourth Edition. Cambridge: The MIT Press.

Perman, Roger, Yue Ma and James McGilvray. 1996. *Natural Resource & Environmental Economics*. New York: Longman Limited.

Portney, Paul R. 1994. "The Contingent Valuation Debate: Why Economists Should Care." *Journal of Economic Perspectives* Volume 8, Number 4, pp. 3-17.

Quiggin, John. February 1998. "Individual and Household Willingness to Pay for Public Goods." *American Journal of Agricultural Economics* 80, pp. 58-63.

Randall, Alan. 1997. "The NOAA Panel Report: A New Beginning or the End of an Era?" *American Journal of Agricultural Economics* 79, pp. 1489-1494.

Randall, Alan.1988. "A Camel is a Camel is...." in *Amenity Resource Valuation: Integrating Economics with other Disciplines*. George L. Peterson, B.L. Driver, and Robin Gregory, eds. State College: Ventura Publishing, Inc.

Ready, Richard C. 1995. "Environmental Valuation under Uncertainty." in *The Handbook of Environmental Economics*. Daniel W. Bromley, ed. Blackwell Handbooks in Economics.

Ready, Richard C. and Dayuan Hu. November 1995. "Statistical Approaches to the Fat Tail Problem for Dichotomous Choice Contingent Valuation." *Land Economics* 71 (4): 491-99.

Rodriguez, Diego J. August 1998. "Assessing the Benefits of Groundwater: A Cost-Benefit Analysis Approach." Unpublished.

Rodriguez, Diego J. May 1998. "Estimating Willingness-To-Pay from Referendum Models: Some Empirical Issues." Unpublished Environment Division Background Paper. Washington, D.C.: Inter-American Development Bank.

Russell, Clifford S., Christopher D. Clark, William J. Vaughan, Arthur Darling, and Diego J. Rodriguez. Forthcoming 1999. "The Economic Analysis of Water Quality Improvement Investments: A Good Practice Paper". Technical Paper, Environment Division, Sustainable Development Department. Washington D.C.: Inter-American Development Bank.

Sang, Heng-Kang. 1988. *Project Evaluation: Techniques and Practices for Developing Countries*. New York: Wilson Press.

Savedoff, William D. 1994. "A Review of Indicators Used in Economic Analysis of Projects: Strengths, Weaknesses, and Cautionary Tales." Working Paper Series No. 107-94. Project Advisory Unit, Project Analysis Department. Washington, D.C.: Inter-American Development Bank.

Schroeder, Larry D., David L. Sjoquist, and Paula E. Stephan. 1986. Understanding Regression Analysis: An Introductory Guide. Series Number 07-057. Series: Quantitative Applications in the Social Sciences. Thousand Oaks, California: Sage Publications, Inc.

Schroeder, Herbert W., and John F. Dwyer. 1988. "Overview: Gains, Losses, and Contingent Valuation." in *Amenity Resource Valuation: Integrating Economics with other Disciplines*. George L. Peterson, B.L. Driver, and Robin Gregory, eds. State College: Ventura Publishing, Inc.

Shavell, Steven. 1993. "Contingent Valuation of the Nonuse Value of Natural Resources: Implications for Public Policy and the Liability System." in *Contingent Valuation: A Critical Assessment*. J. A. Hausman (Editor). Elsevier Science Publishers B. V.

Smith, V. Kerry and George Van Houtven. April 1998. "Non-Market Valuation and the Household." Unpublished.

Smith, V. Kerry. November 1997. "Time and the Valuation of Environmental Resources." Discussion Paper 98-07. Washington, D.C.: Resources for the Future.

Smith, V. Kerry. August 1996. "Pricing What is Priceless: A Status Report on Non-Market Valuation of Environmental Resources." Unpublished.

Smith, V. Kerry and Laura L. Osborne. 1996. "Do Contingent Valuation Estimates Pass a "Scope" Test? A Meta-Analysis." *Journal of Environmental Economics and Management* 31, 287-301.

Smith, V. Kerry.1996. "Pricing What is Priceless: A Status Report on Non-Market Valuation of Environmental Resources." Unpublished version.

Smith, V. Kerry. Winter 1994. "Lightning Rods, Dart Boards, and Contingent Valuation." *Natural Resources Journal*. Vol. 34, pp. 121-152.

Smith, V. Kerry and William H. Desvousges. 1986. *Measuring Water Quality Benefits*. Boston: Kluwer-Nijhoff.

Squire, Lyn and Herman G. van der Tak. 1975. *Economic Analysis of Projects*. Baltimore: The Johns Hopkins University Press.

Tietenberg, Tom. 1996. *Environmental and Natural Resource Economics*. New York: HarperCollins Publishers, Inc.

Tryon, Joseph L. and Forrest E. Cookson. 1966. "A Critical Survey of Project Planning." Working Paper M-8625. Center for Development Planning. Washington, D.C.: National Planning Association.

U.S. Environmental Protection Agency. 1994. "Comments on Proposed NOAA/DOI Regulations on Natural Resource Damage Assessment." Washington, D.C.: U.S. Environmental Protection Agency.

Varian, Hal R. 1992. *Microeconomic Analysis*. Third Edition. New York: W. W. Norton & Company, Inc.

Vaughan, William J. Senior Economist. Environment Division, Social Programs and Sustainable Development Department, Inter-American Development Bank. Personal Communication, Various Dates.

Vaughan, William J.and Arne Paulson. April 1997. "Eligibility Delays and Portfolio Performance: A Special Study for the 1996 Annual Report on Porftolio Managment." Unpublished.

Vaughan, William J. January 1994. "Sample Size for the Ex-Ante Economic Evaluation of Multiple Works Programs". Project Analysis Department, Environment Protection Division Working Paper ENP103. Washington, D.C.: Inter-American Development Bank.

Vaughan, William J. and Sergio Ardila. December 1993. "Economic Analysis of the Environmental Aspects of Investment Projects". Environment Protection Division Working Paper ENP100. Washington, D.C.: Inter-American Development Bank.

Vaughan, William J., Jaime Millan, Arthur Darling, Jeffrey Poyo and Arne Paulson. December 1993. "Economic Analysis and Project Quality". Working Paper No. 102-93. Project Advisory Office, Project Analysis Department. Washington, D.C.: Inter-American Development Bank.

Ward, William A. and Barry J. Deren with Emmanuel H. D'Silva. 1991. "The Economics of Project Analysis: A Practitioner's Guide." EDI Technical Materials. Economic Development Institute o The World Bank. Washington, D.C.: The World Bank.

Annex 1: Numerical Integration Formulas for the Mean

The mean E(x) of a continuous random variable x with a cumulative distribution function F(x)[29] and probability density function f(x) –which is the first derivative of F(x) w.r.t. x–is given by:

$$(1)\ E(x) = \int_{-\infty}^{+\infty} x\, f(x)\, dx$$

The problem is to use a discrete approximation to (1) above to compute:

$$(2)\ E(x) \approx \sum_{x} x\, f(x)$$

where the range of x is approximately minus to plus infinity for the untruncated mean and zero to some upper limit $x_{\max}$ for the truncated mean.

The fundamental theorem of the calculus tells us that the area under a curve f(x) between the limits x_1 and x_2 is (i) the sum of a number of infinitesimally small subdivisions in x of length n; (ii) the definite integral of f(x) between the limits; or the difference between the integral F(x) evaluated at x_1 and x_2:

$$(3)\ \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i)\, \Delta x_i = \int_{x_1}^{x_2} f(x)\, dx = F(x_2) - F(x_1)$$

We know the value of F(x) for any bid x from the bid group proportions. Therefore, we can split the x range into "small" intervals and sum the means from each small interval to get the grand mean. That is, the contribution to the overall mean from the approximate mean *within* any bid group interval is the product of some x within the interval (i.e. the lower limit, x_1, the upper limit, x_2, or some arbitrary value of value of x in between which Kriström's method sets at the group mid-point) times the probability that x lies between

[29] To obtain the mean from the survival function, 1-F(x), t he same reasoning developed below also applies.

x_1 and x_2:

$$(4)\ E(x) \text{ in interval } x_2 - x_1 = \int_{x_1}^{x_2} x\, f(x)\, dx = x[F(x_2) - F(x_1)] \text{ for } (x_1 \leq x \leq x_2):$$

Generalizing, then, the grand mean is the sum of the interval sub-means. That is, symbolically, using the lower limit of each interval for each x_i and repeatedly applying (4) above:

$$(5)\ E(x) \approx x_1[F(x_2) - F(x_1)] + x_2[F(x_3) - F(x_2)] + x_3[F(x_4) - F(x_3)] \ldots + \quad x_{n-1}[F(x_n) - F(x_{n-1})]$$

> where x_1= a large negative number for the unrestricted mean or 0 for the truncated mean and x_n equals a large positive number for the unrestricted mean and the truncated mean bounded at zero but unbounded from above, or x_{max} when bounding from above at average income or some fraction thereof.

In addition, the density (a.k.a. pdf and f(x)), at some point in any interval given ascending values for x (i.e. $x_1 < x_2 < x_3 < \ldots x_n$) is approximated by — and proportional to — the difference between adjacent CDF values (Freund and Walpole, Theorem 3.3, p. 80), where the factor of proportionality is the sum of f(x) over the sampled points to normalize to one (Pollard 1977):

$$(6)\ f(x_i)\ (1/\sum f(x)) \approx [F(x_i) - F(x_{i-1})]$$

The above relationships can be used to compute the mean by numerical integration for any of the formulas in Table 1, even without access to specialized software. While admittedly crude, with a sufficient number of points it is possible to come very close to the analytical results in a simple spreadsheet setup by computing the sum of the products of the interval mid-points (or lower bounds) times the difference in adjacent CDF values, ΔF(x). Equivalently, f(x) values can be multiplied by the successive values of x and summed, but the result has to be divided by the normalizing factor ∑f(x) to get the mean.

Annex 2: A Few Observations of the Sample Data

Bairro	P118	DAP	VALOR	RENDA	IDADE
1	0	0	2	241	72
1	0	1	1	1801	41
1	0	0	5	120	77
1	0	0	12	901	44
1	0	0	5	400	68
1	0	0	5	1801	53
1	0	0	1	481	59
1	0	0	12	481	57
1	0	1	5	901	43
1	0	1	20	900	44
2	0	0	5	1200	43
2	0	0	2	1700	44
2	0	0	12	2500	51
2	0	0	5	500	66
2	0	1	1	241	46
2	0	0	5	2500	39
2	0	0	12	1000	34
2	0	1	2	1500	45
2	0	0	20	1500	53
2	0	0	20	2000	45
3	0	0	12	500	37
3	0	0	1	2300	44
3	0	0	2	1000	37
3	0	1	20	2000	47
3	1	0	1	4000	25
3	0	0	5	1400	42
3	0	0	12	900	29
3	0	0	20	1100	46
3	0	0	2	1000	36
3	0	0	1	1500	36
4	0	0	20	1200	39
4	1	0	5	3000	55
4	1	0	12	4000	56
4	0	1	1	950	22
4	0	0	12	1200	36
4	0	1	2	1500	36
4	0	0	12	901	56

Annex 3: LIMDEP Code

```
READ;NAME;FILE= e:\diego\econ\thesis\tiete\TIETE.WK1;FORMAT=WKS$
SAMPLE;1-600$
NameLIST;MOD1=Bairro, P118, DAP, IDADE, VALOR$
RECODE;Bairro;1,4,5,6,9,10,12,13,20,27,28,32,33,34,38,45,49,51,53=1;*=0$
DSTAT;RHS=MOD1;OUTPUT=2$
SAMPLE;ALL$
CREATE; IF(VALOR=.50)BID1=1;(ELSE)BID1=0$
CREATE; IF(VALOR=2)BID2=1;(ELSE)BID2=0$
CREATE; IF(VALOR=5)BID3=1;(ELSE)BID3=0$
CREATE; IF(VALOR=12)BID4=1;(ELSE)BID4=0$
CREATE; IF(VALOR=20)BID5=1;(ELSE)BID5=0$
?DSTAT AND COUNT BY BID RANGE FOR BAIRRO=0 (FAR FROM RIVER)
REJECT; BAIRRO$
REJECT; BID2$
REJECT; BID3$
REJECT; BID4$
REJECT; BID5$
DSTAT;RHS=BID1$
DSTAT;RHS=DAP$
SAMPLE; ALL$
REJECT; BAIRRO$
REJECT; BID1$
REJECT; BID3$
REJECT; BID4$
REJECT; BID5$
DSTAT;RHS=BID2$
DSTAT;RHS=DAP$
SAMPLE; ALL$
REJECT; BAIRRO$
REJECT; BID1$
REJECT; BID2$
REJECT; BID4$
REJECT; BID5$
DSTAT;RHS=BID3$
DSTAT;RHS=DAP$
SAMPLE; ALL$
REJECT; BAIRRO$
REJECT; BID1$
REJECT; BID2$
REJECT; BID3$
REJECT; BID5$
DSTAT;RHS=BID4$
DSTAT;RHS=DAP$
SAMPLE; ALL$
```

```
REJECT; BAIRRO$
REJECT; BID1$
REJECT; BID2$
REJECT; BID3$
REJECT; BID4$
DSTAT;RHS=BID5$
DSTAT;RHS=DAP$
?DSTAT AND COUNT FOR BAIRRO=1 (NEAR THE RIVER)
SAMPLE; ALL$
REJECT; BAIRRO#1$
REJECT; BID2$
REJECT; BID3$
REJECT; BID4$
REJECT; BID5$
DSTAT;RHS=BID1$
DSTAT;RHS=DAP$
SAMPLE; ALL$
REJECT; BAIRRO#1$
REJECT; BID1$
REJECT; BID3$
REJECT; BID4$
REJECT; BID5$
DSTAT;RHS=BID2$
DSTAT;RHS=DAP$
SAMPLE; ALL$
REJECT; BAIRRO#1$
REJECT; BID1$
REJECT; BID2$
REJECT; BID4$
REJECT; BID5$
DSTAT;RHS=BID3$
DSTAT;RHS=DAP$
SAMPLE; ALL$
REJECT; BAIRRO#1$
REJECT; BID1$
REJECT; BID2$
REJECT; BID3$
REJECT; BID5$
DSTAT;RHS=BID4$
DSTAT;RHS=DAP$
SAMPLE; ALL$
REJECT; BAIRRO#1$
REJECT; BID1$
REJECT; BID2$
REJECT; BID3$
REJECT; BID4$
DSTAT;RHS=BID5$
DSTAT;RHS=DAP$
```

?MODEL 1: LOGIT LINEAR IN BID (VALOR)

```
SAMPLE; ALL$
LOGIT;Lhs=DAP;Rhs=ONE,BAIRRO,P118,IDADE,VALOR$
DSTAT;RHS=ONE, BAIRRO, P118, IDADE, VALOR$
```

?MODEL 2: PROBIT LINEAR IN BID (VALOR)

```
PROBIT;Lhs=DAP;Rhs=ONE,BAIRRO,P118,IDADE,VALOR$
DSTAT; RHS=ONE,BAIRRO,P118,IDADE,VALOR$
?CREATION OF EXTRA VARIABLE FOR MCCONNELL BOUNDED LOGIT ESTIMATION
CREATE; BTRANS=LOG((RENDA-VALOR)/VALOR)$
SAMPLE; ALL$
```

?MODEL 3: BOUNDED LOGIT

```
LOGIT;Lhs=DAP;Rhs=ONE,BAIRRO,P118,IDADE,BTRANS$
DSTAT; RHS=ONE,BAIRRO,P118,IDADE,BTRANS$
```

?MODEL 4: BOUNDED PROBIT

```
PROBIT;Lhs=DAP;Rhs=ONE,BAIRRO,P118,IDADE,BTRANS$
DSTAT; RHS=ONE,BAIRRO,P118,IDADE,BTRANS$
```

?MODEL 5: BOUNDED PROBIT INCORPORATING RENDA AS EXPLANATORY VARIABLE

```
SAMPLE; ALL$
PROBIT; Lhs=DAP; Rhs=ONE,BAIRRO,P118,IDADE,BTRANS,RENDA$
DSTAT; Rhs=ONE,BAIRRO,P118,IDADE,BTRANS,RENDA$
```

?MODEL 6: LOGIT - LOG OF BID

```
CREATE; LNBID=LOG(VALOR)$
SAMPLE; ALL$
LOGIT;Lhs=DAP;RhS=ONE,BAIRRO,P118,IDADE,LNBID$
DSTAT; Rhs=ONE,BAIRRO,P118,IDADE,LNBID$
```

?MODEL 7: PROBIT - LOG OF BID

```
SAMPLE; ALL$
PROBIT;Lhs=DAP;RhS=ONE,BAIRRO,P118,IDADE,LNBID$
DSTAT; Rhs=ONE,BAIRRO,P118,IDADE,LNBID$
```

Annex 4: LIMDEP Output Results

```
--> RESET
--> READ;NAME;FILE= e:\diego\econ\thesis\tiete\TIETE.WK1;FORMAT=WKS$
--> SAMPLE;1-600$
--> NameLIST;MOD1=Bairro, P118, DAP, IDADE, VALOR$
--> RECODE;Bairro;1,4,5,6,9,10,12,13,20,27,28,32,33,34,38,45,49,51,53=1;
    *=0$
--> DSTAT;RHS=MOD1;OUTPUT=2$
                             Descriptive Statistics
               All results based on nonmissing observations.
Variable        Mean         Std.Dev.        Minimum         Maximum      Cases
-------------------------------------------------------------------------------
BAIRRO    .306666667      .461494512      .000000000      1.00000000        600
P118      .110000000      .313150825      .000000000      1.00000000        600
DAP       .351666667      .477888954      .000000000      1.00000000        600
IDADE     45.8883333      13.8308116      18.0000000      84.0000000        600
VALOR     7.90000000      7.23375504      .500000000      20.0000000        600

 Matrix COR.MAT. has  5 rows and  5 columns.
         BAIRRO        P118          DAP           IDADE         VALOR

       +-----------------------------------------------------------------------
BAIRRO  |  .1000000D+01  .7809078D-01  .1157662D+00  .1680597D+00  .7951333D-02
P118    |  .7809078D-01  .1000000D+01  .9805765D-01  .8725510D-01  .4945137D-01
DAP     |  .1157662D+00  .9805765D-01  .1000000D+01 -.1079623D+00 -.2829475D+00
IDADE   |  .1680597D+00  .8725510D-01 -.1079623D+00  .1000000D+01  .4024417D-01
VALOR   |  .7951333D-02  .4945137D-01 -.2829475D+00  .4024417D-01  .1000000D+01
--> SAMPLE; ALL$
--> LOGIT;Lhs=DAP;Rhs=ONE,BAIRRO,P118,IDADE,VALOR$

            +------------------------------------------------+
            | Multinomial logit model                        |
            | There are  2 outcomes for LH variable DAP      |
            | These are the OLS start values based on the    |
            | binary variables for each outcome Y(i) = j.    |
            | Coefficients for LHS=0 outcome are set to 0.0  |
            +------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Characteristics in numerator of Prob[Y = 1]
 Constant  .6431154899      .66086141E-01    9.731   .0000
 BAIRRO    .1354127874      .40470099E-01    3.346   .0008  .30666667
 P118      .1726147737      .59081041E-01    2.922   .0035  .11000000
 IDADE    -.4435241881E-02  .13522723E-02   -3.280   .0010  45.888333
 VALOR    -.1878951628E-01  .25446196E-02   -7.384   .0000  7.9000000

Normal exit from iterations. Exit status=0.


              +--------------------------------------------+
              | Multinomial Logit Model                    |
              | Maximum Likelihood Estimates               |
              | Dependent variable                DAP      |
              | Weighting variable                ONE      |
              | Number of observations            600      |
              | Iterations completed                5      |
              | Log likelihood function     -350.0071      |
```

```
               | Restricted log likelihood     -389.0834     |
               | Chi-squared                    78.15258     |
               | Degrees of freedom                    4     |
               | Significance level             .0000000     |
               +---------------------------------------------+
+---------+-------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+-------------+----------------+--------+---------+----------+
          Characteristics in numerator of Prob[Y = 1]
 Constant  .7769122002      .32805802        2.368   .0179
 BAIRRO    .6551473302      .19911602        3.290   .0010  .30666667
 P118      .8357227963      .28946674        2.887   .0039  .11000000
 IDADE    -.2210396196E-01  .69025015E-02   -3.202   .0014  45.888333
 VALOR    -.9775946274E-01  .14412255E-01   -6.783   .0000  7.9000000

Frequencies of actual & predicted outcomes

Predicted outcome has maximum probability.

            Predicted
------  ----------  +  -----
Actual      0    1  |  Total
------  ----------  +  -----
  0       344   45  |    389
  1       141   70  |    211
------  ----------  +  -----
Total     485  115  |    600
--> DSTAT;RHS=ONE, BAIRRO, P118, IDADE, VALOR$
                             Descriptive Statistics
               All results based on nonmissing observations.
Variable        Mean          Std.Dev.        Minimum         Maximum      Cases
-----------------------------------------------------------------------------
BAIRRO    .306666667      .461494512      .000000000      1.00000000        600
P118      .110000000      .313150825      .000000000      1.00000000        600
IDADE     45.8883333      13.8308116      18.0000000      84.0000000        600
VALOR     7.90000000      7.23375504      .500000000      20.0000000        600

--> PROBIT;Lhs=DAP;Rhs=ONE,BAIRRO,P118,IDADE,VALOR$
```

```
+-----------------------------------------------------------------------+
| Dependent variable is binary, y=0 or y not equal 0                    |
| Ordinary    least squares regression    Weighting variable = none     |
| Dep. var. = DAP      Mean=   .3516666667    , S.D.=   .4778889536     |
| Model size: Observations =     600, Parameters =   5, Deg.Fr.=    595 |
| Residuals:  Sum of squares= 120.3055666    , Std.Dev.=         .44966 |
| Fit:        R-squared=  .120563, Adjusted R-squared =          .11465 |
| Model test: F[  4,    595] =   20.39,    Prob value =          .00000 |
| Diagnostic: Log-L =   -369.2947, Restricted(b=0) Log-L =    -407.8366 |
|             LogAmemiyaPrCrt.=   -1.590, Akaike Info. Crt.=      1.248 |
+-----------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant  .6431154899      .66086141E-01    9.731   .0000
 BAIRRO    .1354127874      .40470099E-01    3.346   .0008  .30666667
 P118      .1726147737      .59081041E-01    2.922   .0035  .11000000
 IDADE    -.4435241881E-02  .13522723E-02   -3.280   .0010  45.888333
 VALOR    -.1878951628E-01  .25446196E-02   -7.384   .0000  7.9000000
```

Normal exit from iterations. Exit status=0.

```
              +---------------------------------------------+
              | Binomial Probit Model                       |
              | Maximum Likelihood Estimates                |
              | Dependent variable                 DAP      |
              | Weighting variable                 ONE      |
              | Number of observations             600      |
              | Iterations completed                 5      |
              | Log likelihood function      -350.2144      |
              | Restricted log likelihood    -389.0834      |
              | Chi-squared                   77.73793      |
              | Degrees of freedom                   4      |
              | Significance level            .0000000      |
              +---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Index function for probability
 Constant  .4609386122      .19854088        2.322   .0203
 BAIRRO    .3981059009      .12010968        3.315   .0009  .30666667
 P118      .4998137959      .17483091        2.859   .0043  .11000000
 IDADE    -.1331684141E-01  .41367778E-02   -3.219   .0013  45.888333
 VALOR    -.5777231876E-01  .82078888E-02   -7.039   .0000  7.9000000
```

Frequencies of actual & predicted outcomes

Predicted outcome has maximum probability.

```
            Predicted
------  ----------  +  -----
Actual      0    1  |  Total
------  ----------  +  -----
  0       344   45  |    389
  1       141   70  |    211
------  ----------  +  -----
Total     485  115  |    600

--> DSTAT; RHS=ONE,BAIRRO,P118,IDADE,VALOR$

                             Descriptive Statistics
               All results based on nonmissing observations.
Variable        Mean          Std.Dev.        Minimum         Maximum      Cases
==============================================================================
BAIRRO    .306666667      .461494512      .000000000      1.00000000         600
P118      .110000000      .313150825      .000000000      1.00000000         600
IDADE     45.8883333      13.8308116      18.0000000      84.0000000         600
VALOR     7.90000000      7.23375504      .500000000      20.0000000         600

--> CREATE; BTRANS=LOG((RENDA-VALOR)/VALOR)$
--> SAMPLE; ALL$
--> LOGIT;Lhs=DAP;Rhs=ONE,BAIRRO,P118,IDADE,BTRANS$


            +-------------------------------------------------+
            | Multinomial logit model                         |
            | There are  2 outcomes for LH variable DAP       |
            | These are the OLS start values based on the     |
            | binary variables for each outcome Y(i) = j.     |
            | Coefficients for LHS=0 outcome are set to 0.0   |
            +-------------------------------------------------+

+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Characteristics in numerator of Prob[Y = 1]
 Constant  .2878710711E-01  .93781252E-01      .307   .7589
 BAIRRO    .1225804927      .40660733E-01     3.015   .0026  .30666667
 P118      .9307960006E-01  .59826264E-01     1.556   .1197  .11000000
 IDADE    -.3423558858E-02  .13701142E-02    -2.499   .0125  45.888333
 BTRANS    .8116640161E-01  .11496711E-01     7.060   .0000  5.3242580

Normal exit from iterations. Exit status=0.

              +---------------------------------------------+
              | Multinomial Logit Model                     |
              | Maximum Likelihood Estimates                |
              | Dependent variable                DAP       |
              | Weighting variable                ONE       |
              | Number of observations            600       |
              | Iterations completed                5       |
              | Log likelihood function     -353.3194       |
              | Restricted log likelihood   -389.0834       |
              | Chi-squared                  71.52789       |
              | Degrees of freedom                  4       |
              | Significance level           .0000000       |
              +---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
```

```
          Characteristics in numerator of Prob[Y = 1]
 Constant -2.179939869      .47236955       -4.615   .0000
 BAIRRO    .5815176444      .19764249        2.942   .0033  .30666667
 P118      .4434263855      .28539552        1.554   .1202  .11000000
 IDADE    -.1773196887E-01  .69715254E-02   -2.543   .0110  45.888333
 BTRANS    .3904748561      .59575263E-01    6.554   .0000  5.3242580

Frequencies of actual & predicted outcomes

Predicted outcome has maximum probability.

            Predicted
------  ----------  +  -----
Actual      0    1  |  Total
------  ----------  +  -----
  0       346   43  |    389
  1       143   68  |    211
------  ----------  +  -----
Total     489  111  |    600
--> DSTAT; RHS=ONE,BAIRRO,P118,IDADE,BTRANS$
                             Descriptive Statistics
               All results based on nonmissing observations.
Variable       Mean         Std.Dev.        Minimum         Maximum      Cases
------------------------------------------------------------------------------
BAIRRO    .306666667     .461494512      .000000000      1.00000000        600
P118      .110000000     .313150825      .000000000      1.00000000        600
IDADE     45.8883333     13.8308116      18.0000000      84.0000000        600
BTRANS    5.32425799     1.63481165      1.60943791      9.54674118        600

--> PROBIT;Lhs=DAP;Rhs=ONE,BAIRRO,P118,IDADE,BTRANS$

+-----------------------------------------------------------------------+
| Dependent variable is binary, y=0 or y not equal 0                    |
| Ordinary    least squares regression    Weighting variable = none     |
| Dep. var. = DAP      Mean=   .3516666667    , S.D.=    .4778889536    |
| Model size: Observations =     600, Parameters =   5, Deg.Fr.=    595 |
| Residuals:  Sum of squares= 121.1788024    , Std.Dev.=         .45129 |
| Fit:        R-squared=  .114179, Adjusted R-squared =          .10822 |
| Model test: F[  4,    595] =   19.17,    Prob value =          .00000 |
| Diagnostic: Log-L =   -371.4644, Restricted(b=0) Log-L =    -407.8366 |
|             LogAmemiyaPrCrt.=   -1.583, Akaike Info. Crt.=      1.255 |
+-----------------------------------------------------------------------+


+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant  .2878710711E-01  .93781252E-01     .307   .7589
 BAIRRO    .1225804927      .40660733E-01    3.015   .0026  .30666667
 P118      .9307960006E-01  .59826264E-01    1.556   .1197  .11000000
 IDADE    -.3423558858E-02  .13701142E-02   -2.499   .0125  45.888333
 BTRANS    .8116640161E-01  .11496711E-01    7.060   .0000  5.3242580

Normal exit from iterations. Exit status=0.

              +---------------------------------------------+
              | Binomial Probit Model                       |
              | Maximum Likelihood Estimates                |
              | Dependent variable                DAP       |
              | Weighting variable                ONE       |
              | Number of observations            600       |
              | Iterations completed                5       |
```

```
               | Log likelihood function       -353.4328     |
               | Restricted log likelihood     -389.0834     |
               | Chi-squared                    71.30103     |
               | Degrees of freedom                    4     |
               | Significance level             .0000000     |
               +---------------------------------------------+

+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+

          Index function for probability
 Constant -1.308901389      .27976749       -4.679   .0000
 BAIRRO    .3568740472      .11991763        2.976   .0029  .30666667
 P118      .2715018344      .17399950        1.560   .1187  .11000000
 IDADE    -.1075599556E-01  .41579951E-02   -2.587   .0097  45.888333
 BTRANS    .2342110463      .34858330E-01    6.719   .0000  5.3242580

Frequencies of actual & predicted outcomes
Predicted outcome has maximum probability.


            Predicted
------  ----------  +  -----
Actual      0    1  |  Total
------  ----------  +  -----
  0       345   44  |    389
  1       144   67  |    211
------  ----------  +  -----
Total     489  111  |    600

--> DSTAT; RHS=ONE,BAIRRO,P118,IDADE,BTRANS$

                             Descriptive Statistics
               All results based on nonmissing observations.
Variable        Mean         Std.Dev.        Minimum         Maximum      Cases
-------------------------------------------------------------------------------
BAIRRO    .306666667      .461494512      .000000000      1.00000000        600
P118      .110000000      .313150825      .000000000      1.00000000        600
IDADE     45.8883333      13.8308116      18.0000000      84.0000000        600
BTRANS    5.32425799      1.63481165      1.60943791      9.54674118        600

--> SAMPLE; ALL$
--> PROBIT; Lhs=DAP; Rhs=ONE,BAIRRO,P118,IDADE,BTRANS,RENDA$

+-----------------------------------------------------------------------+
| Dependent variable is binary, y=0 or y not equal 0                    |
| Ordinary    least squares regression    Weighting variable = none     |
| Dep. var. = DAP      Mean=   .3516666667    , S.D.=   .4778889536     |
| Model size: Observations =     600, Parameters =   6, Deg.Fr.=    594 |
| Residuals:  Sum of squares= 120.7822422    , Std.Dev.=         .45093 |
| Fit:        R-squared=  .117078, Adjusted R-squared =          .10965 |
| Model test: F[  5,    594] =   15.75,    Prob value =          .00000 |
| Diagnostic: Log-L =   -370.4810, Restricted(b=0) Log-L =    -407.8366 |
|             LogAmemiyaPrCrt.=   -1.583, Akaike Info. Crt.=      1.255 |
+-----------------------------------------------------------------------+

+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant  .1506108542E-01  .94220506E-01     .160   .8730
 BAIRRO    .1300683355      .40980579E-01    3.174   .0015  .30666667
```

```
 P118       .1217191411      .63198466E-01    1.926   .0541  .11000000
 IDADE    -.3584378694E-02  .13738562E-02   -2.609   .0091  45.888333
 BTRANS    .8994840778E-01  .13096142E-01    6.868   .0000  5.3242580
 RENDA    -.2460143653E-04  .17616281E-04   -1.397   .1626  1264.0933

Normal exit from iterations. Exit status=0.

              +---------------------------------------------+
              | Binomial Probit Model                       |
              | Maximum Likelihood Estimates                |
              | Dependent variable                 DAP      |
              | Weighting variable                 ONE      |
              | Number of observations             600      |
              | Iterations completed                 5      |
              | Log likelihood function      -352.4272      |
              | Restricted log likelihood    -389.0834      |
              | Chi-squared                   73.31239      |
              | Degrees of freedom                   5      |
              | Significance level            .0000000      |
              +---------------------------------------------+


+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Index function for probability
 Constant -1.350778813      .28115935       -4.804   .0000
 BAIRRO    .3801977864      .12115352        3.138   .0017  .30666667
 P118      .3566690662      .18408042        1.938   .0527  .11000000
 IDADE    -.1130777971E-01  .41787489E-02   -2.706   .0068  45.888333
 BTRANS    .2604147781      .39504108E-01    6.592   .0000  5.3242580
 RENDA    -.7175350370E-04  .50228352E-04   -1.429   .1531  1264.0933

Frequencies of actual & predicted outcomes

Predicted outcome has maximum probability.

            Predicted
------  ----------  +  -----
Actual     0    1   |  Total
------  ----------  +  -----
  0       348   41  |   389
  1       143   68  |   211
------  ----------  +  -----
Total     491  109  |   600


--> DSTAT; Rhs=ONE,BAIRRO,P118,IDADE,BTRANS,RENDA$


                            Descriptive Statistics
               All results based on nonmissing observations.
Variable       Mean          Std.Dev.         Minimum          Maximum       Cases
-------------------------------------------------------------------------------
BAIRRO    .306666667      .461494512      .000000000      1.00000000        600
P118      .110000000      .313150825      .000000000      1.00000000        600
IDADE     45.8883333      13.8308116      18.0000000      84.0000000        600
BTRANS    5.32425799      1.63481165      1.60943791      9.54674118        600
RENDA     1264.09333      1291.32590      120.000000      15000.0000        600
```

```
--> CREATE; LNBID=LOG(VALOR)$
--> SAMPLE; ALL$
--> LOGIT;Lhs=DAP;RhS=ONE,BAIRRO,P118,IDADE,LNBID$

            +-------------------------------------------------+
            | Multinomial logit model                         |
            | There are  2 outcomes for LH variable DAP       |
            | These are the OLS start values based on the     |
            | binary variables for each outcome Y(i) = j.     |
            | Coefficients for LHS=0 outcome are set to 0.0   |
            +-------------------------------------------------+


+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Characteristics in numerator of Prob[Y = 1]
 Constant  .6454784693      .65975391E-01    9.784   .0000
 BAIRRO    .1361385842      .40402095E-01    3.370   .0008  .30666667
 P118      .1657001508      .58945122E-01    2.811   .0049  .11000000
 IDADE    -.4463183732E-02  .13498453E-02   -3.306   .0009  45.888333
 LNBID    -.1050622763      .13951340E-01   -7.531   .0000  1.4180154

Normal exit from iterations. Exit status=0.

              +----------------------------------------------+
              | Multinomial Logit Model                      |
              | Maximum Likelihood Estimates                 |
              | Dependent variable                 DAP       |
              | Weighting variable                 ONE       |
              | Number of observations             600       |
              | Iterations completed                 5       |
              | Log likelihood function      -350.6859       |
              | Restricted log likelihood    -389.0834       |
              | Chi-squared                   76.79498       |
              | Degrees of freedom                   4       |
              | Significance level            .0000000       |
              +----------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Characteristics in numerator of Prob[Y = 1]
 Constant  .7608212397      .33069388        2.301   .0214
 BAIRRO    .6628953053      .19896660        3.332   .0009  .30666667
 P118      .7968283208      .28637660        2.782   .0054  .11000000
 IDADE    -.2269546651E-01  .69428735E-02   -3.269   .0011  45.888333
 LNBID    -.4945452768      .70747029E-01   -6.990   .0000  1.4180154

Frequencies of actual & predicted outcomes

Predicted outcome has maximum probability.

            Predicted
------  ----------  +  -----
Actual      0    1  |  Total
------  ----------  +  -----
  0       342   47  |    389
  1       141   70  |    211
------  ----------  +  -----
Total     483  117  |    600

--> DSTAT; Rhs=ONE,BAIRRO,P118,IDADE,LNBID$
```

```
                         Descriptive Statistics
               All results based on nonmissing observations.
Variable        Mean          Std.Dev.        Minimum          Maximum      Cases
-------------------------------------------------------------------------------
BAIRRO    .306666667      .461494512       .000000000      1.00000000         600
P118      .110000000      .313150825       .000000000      1.00000000         600
IDADE     45.8883333      13.8308116       18.0000000      84.0000000         600
LNBID     1.41801537      1.31612814      -.693147181      2.99573227         600

--> SAMPLE; ALL$
--> PROBIT;Lhs=DAP;RhS=ONE,BAIRRO,P118,IDADE,LNBID$

+-----------------------------------------------------------------------+
| Dependent variable is binary, y=0 or y not equal 0                    |
| Ordinary    least squares regression    Weighting variable = none     |
| Dep. var. = DAP      Mean=   .3516666667    , S.D.=   .4778889536     |
| Model size: Observations =     600, Parameters =   5, Deg.Fr.=    595 |
| Residuals:  Sum of squares= 119.9019277    , Std.Dev.=         .44891 |
| Fit:        R-squared=  .123513, Adjusted R-squared =          .11762 |
| Model test: F[  4,    595] =   20.96,    Prob value =          .00000 |
| Diagnostic: Log-L =   -368.2865, Restricted(b=0) Log-L =    -407.8366 |
|             LogAmemiyaPrCrt.=   -1.594, Akaike Info. Crt.=      1.244 |
+-----------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant  .6454784693      .65975391E-01    9.784   .0000
 BAIRRO    .1361385842      .40402095E-01    3.370   .0008  .30666667
 P118      .1657001508      .58945122E-01    2.811   .0049  .11000000
 IDADE    -.4463183732E-02  .13498453E-02   -3.306   .0009  45.888333
 LNBID    -.1050622763      .13951340E-01   -7.531   .0000  1.4180154

Normal exit from iterations. Exit status=0.

              +---------------------------------------------+
              | Binomial Probit Model                       |
              | Maximum Likelihood Estimates                |
              | Dependent variable                 DAP      |
              | Weighting variable                 ONE      |
              | Number of observations             600      |
              | Iterations completed                 5      |
              | Log likelihood function      -350.5878      |
              | Restricted log likelihood    -389.0834      |
              | Chi-squared                   76.99104      |
              | Degrees of freedom                   4      |
              | Significance level            .0000000      |
              +---------------------------------------------+

+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Index function for probability
 Constant  .4573279453      .19867019        2.302   .0213
 BAIRRO    .4016311053      .12022914        3.341   .0008  .30666667
 P118      .4824650934      .17386014        2.775   .0055  .11000000
 IDADE    -.1363441327E-01  .41402963E-02   -3.293   .0010  45.888333
```

```
 LNBID    -.3013339001      .42215180E-01   -7.138   .0000  1.4180154
```

Frequencies of actual & predicted outcomes

Predicted outcome has maximum probability.

```
            Predicted
------  ----------  +  -----
Actual      0    1  |  Total
------  ----------  +  -----
  0       342   47  |    389
  1       141   70  |    211
------  ----------  +  -----
Total     483  117  |    600
--> DSTAT; Rhs=ONE,BAIRRO,P118,IDADE,LNBID$
```

Descriptive Statistics

All results based on nonmissing observations.

Variable	Mean	Std.Dev.	Minimum	Maximum	Cases
BAIRRO	.306666667	.461494512	.000000000	1.00000000	600
P118	.110000000	.313150825	.000000000	1.00000000	600
IDADE	45.8883333	13.8308116	18.0000000	84.0000000	600
LNBID	1.41801537	1.31612814	-.693147181	2.99573227	600

Annex 5 – Estimates WTP - Parametric Methods - Spreadsheet Calculations

Estimates of WTP - Utility Difference Logit - Linear in Bid

CLOSE TO RIVER SUB SAMPLE

Variable	Mean	Coefficient	Mean * Coefficient
Constant	1.0000	0.7769	0.7769
BAIRRO=Close to River 1=Yes, 0 Else	1.0000	0.6551	0.6551
P118=Status	0.1467	0.8357	0.1226
IDADE= Age of Household Head	49.3800	-0.0221	-1.0913

α = Augmented intercept= **0.4633**

β = **marginal utlity of income=VALOR=Bid after multiplying by -1 to make it positive** 0.0978

ANALYTICAL FORMULAS:

1. Untruncated Mean=Median, ∞<WTP<+∞ : C+ = C*	- α/β =	**4.74**
2. Truncated Mean, 0<WTP<+∞: C'	**ln(1+exp(α))/β =**	**9.73**
3. Truncated Mean, 0<WTP<B_{max} : $C^{\sim}$	**1/β ln[(1+exp(α))/((1+exp(α–βB_{max}))]=**	**7.65**

FAR FROM RIVER SUB SAMPLE

Variable	Mean	Coefficient	Mean * Coefficient
Constant	1.0000	0.7769	0.7769
BAIRRO=Close to River 1=Yes, 0 Else	0.0000	0.6551	0.0000
P118=Status	0.0938	0.8357	0.0784
IDADE= Age of Household Head	44.3400	-0.0221	-0.9799

α = Augmented intercept= **-0.1246**

β = **marginal utlity of income=VALOR=Bid after multiplying by -1 to make it positive** 0.0978

ANALYTICAL FORMULAS:

1. Untruncated Mean=Median, ∞<WTP<∞ : C+ = C*	- α/β =	**-1.27**
2. Truncated Mean, 0<WTP<+∞: C'	**ln(1+exp(α))/β =**	**6.47**
3. Truncated Mean, 0<WTP<B_{max} : C~	**1/β ln[(1+exp(α))/((1+exp(α–βB_{max}))]=**	**5.27**

Estimates of WTP - Utility Difference Logit - Log in Bid

CLOSE TO RIVER SUB SAMPLE

Variable	Mean	Coefficient	Mean * Coefficient
Constant	1.0000	0.7608	0.7608
BAIRRO=Close to River 1=Yes, 0 Else	1.0000	0.6628	0.6628
P118=Status	0.1467	0.7968	0.1169
IDADE= Age of Household Head	49.3800	-0.0226	-1.1160
		α = Augmented intercept=	**0.4245**
β = natural log of bid =LNBID=Log of Bid after multiplying by -1 to make it positive		0.4945	

<u>exp (-α/β)[(π/β)/(sin(π/β)]: Geometric Means were calculated for the log transform models 1 and 2 by taking the antilog of the mean log bid found by numerical integration</u>

1. Truncated Mean, Log Tansform, E(exp $^{\ln(WTP)}$), - ∞ lnWTP < +∞ (utility difference logit, log of bid, 0 Lower Limit, No Upper Limit: $C^{+\ln}$

2. Truncated Mean, Log Tansform, E(exp $^{\ln(WTP)}$), - ∞ lnWTP < ln Income (utility difference logit, log of bid, 0 Lower Limit, Income Upper Limit: $C^{\sim\ln}$

Truncated Median, Log Transform: exp (α/β) $C^*_{\ln}$ **2.36**

FAR FROM RIVER SUB SAMPLE

Variable	Mean	Coefficient	Mean * Coefficient
Constant	1.0000	0.7608	0.7608
BAIRRO=Close to River 1=Yes, 0 Else	0.0000	0.6628	0.0000
P118=Status	0.0938	0.7968	0.0747
IDADE= Age of Household Head	44.3400	-0.0226	-1.0021
		α = Augmented intercept=	**-0.1665**
β = natural log of bid =LNBID=Log of Bid after multiplying by -1 to make it positive		0.4945	

<u>exp (-α/β)[(π/β)/(sin(π/β)]: Geometric Means were calculated for the log transform models 1 and 2 by taking the antilog of the mean log bid found by numerical integration</u>

1. Truncated Mean, Log Tansform, E(exp $^{\ln(WTP)}$), - ∞ lnWTP < +∞ (utility difference logit, log of bid, 0 Lower Limit, No Upper Limit: $C^{+\ln}$

2. Truncated Mean, Log Tansform, E(exp $^{\ln(WTP)}$), - ∞ lnWTP < ln Income (utility difference logit, log of bid, 0 Lower Limit, Income Upper Limit: $C^{\sim\ln}$

Truncated Median, Log Transform: exp (α/β) **0.71**

C^*_{ln}

Annex 6: Numerical Integration Spreadsheet Calculation for the Log Index Functions Evaluated up to Average Income ($1524) and Plus Infinity: Close to River

		MEAN OF NO	MEAN OF YES	OVERALL MEAN
AUGMENTED INTERCEPT= Alpha	**0.4622**	n.a.	3.71	**3.71 U.L. Inc.**
AVERAGE SLOPE=Beta	**-0.49453**		5.01	**5.01 U.L.+ Inf.**

where CDF Prob “Yes”= exp((alpha+beta*lnBid)/1+exp(alpha+beta*lnBid))

and CDF Prob “No” = 1 - CDF Prob “Yes”

PDF=CDF Prob “Yes” sub i - CDF Prob “Yes” sub i+1

Mean is lnBid*PDF at each bid level

		CDF	CDF	PDF
		Prob "Yes"	**Prob "No"**	**Prob "Yes"**
Bid	**ln Bid**	**F(xB)**	**G(xB)**	**f(xB)=g(xB)**
0.00000	(18.42068)	0.99993	0.00007	
0.01000	(4.60517)	0.93932	0.06068	0.06061
0.02000	(3.91202)	0.91658	0.08342	0.02274
0.03000	(3.50656)	0.89992	0.10008	0.01667
0.04000	(3.21888)	0.88635	0.11365	0.01356
0.05000	(2.99573)	0.87476	0.12524	0.01160
0.06000	(2.81341)	0.86454	0.13546	0.01022
0.07000	(2.65926)	0.85536	0.14464	0.00918
0.08000	(2.52573)	0.84700	0.15300	0.00836
0.09000	(2.40795)	0.83930	0.16070	0.00770
0.10000	(2.30259)	0.83214	0.16786	0.00715
0.11000	(2.20727)	0.82546	0.17454	0.00669
0.12000	(2.12026)	0.81917	0.18083	0.00629
0.13000	(2.04022)	0.81323	0.18677	0.00594
0.14000	(1.96611)	0.80760	0.19240	0.00563
0.15000	(1.89712)	0.80224	0.19776	0.00536
0.16000	(1.83258)	0.79713	0.20287	0.00511
0.17000	(1.77196)	0.79224	0.20776	0.00489
0.18000	(1.71480)	0.78755	0.21245	0.00469
0.19000	(1.66073)	0.78304	0.21696	0.00451
0.20000	(1.60944)	0.77870	0.22130	0.00434
•	•	•	•	•
•	•	•	•	•
61589	11.02824	0.00675	0.99325	0.000050000
62589	11.04434	0.00670	0.99330	0.000050000
63589	11.06020	0.00664	0.99336	0.000050000
64589	11.07580	0.00659	0.99341	0.000050000
65589	11.09116	0.00654	0.99346	0.000050000
66589	11.10629	0.00650	0.99350	0.000050000
67589	11.12120	0.00645	0.99355	0.000047394

Annex 7: Numerical Integration Spreadsheet Calculation for the Log Index Functions Evaluated up to Average Income ($1148) and Plus Infinity: Far from River Sub Sample

		MEAN OF NO	MEAN OF YES	OVERALL MEAN
AVERAGE INTERCEPT= Alpha	**-0.1665**	n.a.	1.39	**1.39 U.L. Inc.**
AVERAGE SLOPE= Beta	**-0.49453**		1.68	**1.68 U.L.+Inf.**

Bid	**ln Bid**	**CDF Prob "Yes" F(xB)**	**CDF Prob "No" G(xB)**	**PDF Prob "Yes" f(xB)=g(xB)**
1E-08	(18.42068)	0.99987	0.00013	
0.01	(4.60517)	0.89196	0.10804	0.10791
0.02	(3.91202)	0.85422	0.14578	0.03773
0.03	(3.50656)	0.82744	0.17256	0.02678
0.04	(3.21888)	0.80617	0.19383	0.02127
0.05	(2.99573)	0.78835	0.21165	0.01783
0.06	(2.81341)	0.77291	0.22709	0.01544
0.07	(2.65926)	0.75925	0.24075	0.01366
0.08	(2.52573)	0.74698	0.25302	0.01228
0.09	(2.40795)	0.73581	0.26419	0.01117
0.1	(2.30259)	0.72556	0.27444	0.01025
0.11	(2.20727)	0.71607	0.28393	0.00948
0.12	(2.12026)	0.70724	0.29276	0.00883
0.13	(2.04022)	0.69898	0.30102	0.00826
0.14	(1.96611)	0.69121	0.30879	0.00777
0.15	(1.89712)	0.68388	0.31612	0.00733
0.16	(1.83258)	0.67694	0.32306	0.00694
0.17	(1.77196)	0.67035	0.32965	0.00659
0.18	(1.71480)	0.66408	0.33592	0.00628
0.19	(1.66073)	0.65809	0.34191	0.00599
0.2	(1.60944)	0.65236	0.34764	0.00573
0.21	(1.56065)	0.64686	0.35314	0.00549
0.22	(1.51413)	0.64159	0.35841	0.00527
0.23	(1.46968)	0.63652	0.36348	0.00507
0.24	(1.42712)	0.63164	0.36836	0.00488
0.25	(1.38629)	0.62693	0.37307	0.00471
0.26	(1.34707)	0.62238	0.37762	0.00455
0.27	(1.30933)	0.61799	0.38201	0.00440
0.28	(1.27297)	0.61373	0.38627	0.00425
0.29	(1.23787)	0.60961	0.39039	0.00412
0.3	(1.20397)	0.60561	0.39439	0.00400
0.31	(1.17118)	0.60173	0.39827	0.00388
●	●	●	●	●
●	●	●	●	●
61589	11.02824	0.00361	0.99639	0.0000292
62589	11.04434	0.00358	0.99642	0.0000285
63589	11.06020	0.00355	0.99645	0.0000279

Vita

Diego J. Rodriguez was born on February 22, 1967 in Montevideo, Uruguay. He graduated in 1995 with a Bachelors of Science in Economics from University of Maryland. In 1990, Mr. Rodriguez was employed as a Technical Assistant, Inter-American Development Bank (IDB). At the time of submitting his thesis, Mr. Rodriguez was an Economist in the Environment Division of the Inter-American Development Bank where he performs economic analysis of IDB-funded development projects in Latin America and the Caribbean. He has been working in areas such as integrated water resources management; coastal zone management; environmental information systems; commercialization of hydro-meteorological services; socioeconomic analysis of natural disasters; economic analysis of climatic forecasting; remote sensing in coastal areas; fisheries; and dams and environment.