USING DISTANCE AND ZIP CODE CENSUS INFORMATION FOR NONRESPONSE CORRECTION IN THE ANALYSIS OF MAIL SURVEY DATA

by

Trudy Ann Cameron, W. Douglass Shaw,
Shannon R. Ragland, Sally Keefe,
and John M. (Mac) Callaway*

April, 1996

UCLA Department of Economics
Working Paper # 751

Respectively: Department of Economics, University of California, Los Angeles, CA; Department of Applied Economics and Statistics, University of Nevada, Reno, NV; Hagler Bailly, Inc., Boulder, CO; Hagler Bailly, Inc., Boulder, CO; and UNEP Collaborating Centre on Energy and the Environment, Risoe National Laboratory, Roskilde, Denmark. This data set employed in this paper was collected as part of a study managed by Hagler Bailly Consulting, Inc., under contract with the U.S. Army Corps of Engineers (contract # TCN 93357).

USING DISTANCE AND ZIP CODE INFORMATION FOR NONRESPONSE CORRECTION IN THE ANALYSIS OF MAIL SURVEY DATA

Abstract

Even when researchers using mail surveys go to great lengths to maximize response rates, there will always be some portion of the intended sample that does not respond, either at all, or sufficiently completely to allow inclusion in the estimating sample. This paper examines nonresponse and its apparent consequences for a survey of water-based recreational participation conducted in the Northwest US. We describe how zip codes can be used in combination with special software to determine distances from each address in our intended sample of 7034 households to each of the recreational sites featured in our survey. Zip codes also allow us to merge our intended sample with 1990 Census data aggregated to the zip code level. We demonstrate how to model the survey response/nonresponse decision explicitly, and show that statistical corrections for nonresponse can have a potentially important effect on the apparent inferences from our models. We strongly advocate, based on these results, that any researcher using a mail survey can and should explore analogous response/nonresponse models and corrections before making any claims as to the robustness of empirical results to non-random sample selection.

USING DISTANCE AND ZIP CODE INFORMATION FOR NONRESPONSE CORRECTION IN THE ANALYSIS OF MAIL SURVEY DATA

1. Introduction

The purpose of this paper is to demonstrate the opportunities for, and utility of, explicit modelling of survey response/nonresponse. A good understanding of the relationship between survey response propensities and observable relationships among the subsample of respondents can help inform researchers and policy makers about the likely nature of nonresponse biases. Mail surveys, in particular, have long been a popular method for gathering research information. They continue to be employed in a wide variety of disciplines where household decisions, preferences, or behaviors need to be quantified. A perennial concern with mail surveys is the maximization of response rates (Dillman, 1978). However, even with aggressive campaigns of follow-up reminder postcards, nominal payments to respondents, and replacement mailings, there almost always remains a persistent nonresponse group.

The issues discussed in this paper are relevant to a very broad community of investigators who rely on data gathered from mail surveys, but the discussion here will be cast in terms of an example where a mail survey has been used to collect demand information concerning non-market environmental goods (economic models used with these types of data have included travel cost models, random utility models, and contingent valuation or behavior models).

Section 2 of this paper reviews a selection of findings concerning survey response/nonresponse that have appeared in the broader marketing and social science literature, as well as a small number of studies focussing on this issue within the boundaries of environmental economics. Section 3 covers the manner in which these earlier insights are reflected in the modelling of nonresponse in our specific illustrative environmental valuation context. Section 4 outlines a rudimentary model of water-based recreational trip-taking, to be estimated using the sample of respondents to a mail survey--with and without corrections for selectivity. Section 5 formally lays out the log-likelihood function that would ideally be maximized in order to simultaneously identify the parameters of both the response/nonresponse model and the trips model, plus the correlation between the error terms in these two submodels. We also explain how best to proceed if this function proves impossible to optimize for a given data set, as is the case here.

Section 6 provides a discussion of the empirical findings in our specific example, focussing upon some possible consequences of failure to correct for nonresponse. Section 7 concludes with comments on the apparent implications of our simple illustration for more-formal survey-based estimates of demand and welfare in other contexts.

2. Review of the Relevant Literature on Response/Nonresponse

Much of the general survey research literature on nonresponse has been devoted to studies of the relative effectiveness of different techniques that might be used during survey design and administration to maximize response rates. Fox et al. (1988) describe a meta-analysis of some of these techniques as they apply to mail survey responses. The classes of factors they consider are under the control of the researcher and include several related to the content of the cover letter, the amount and type of incentives offered, the form of respondent contact and follow-ups, the type of postage used for outgoing and return mail, and the topic, length, color, complexity and format of the questionnaire itself.

Among environmental economists, Dillman (1978) has been the standard handbook on design methods to maximize response for mail and telephone surveys.[1] Maximizing response rates is extremely important, but we are concerned here with the task of correcting demand and welfare models for any non-response that remains. To do this, we must acknowledge that nonresponse results from the decision-making process of survey recipients. Heterogeneity in sociodemographic characteristics across the intended sample may account for a significant portion of the systematic differences in response rates.

McDaniel et al. (1987) provide an excellent summary of survey research that focusses on the demographic characteristics of non-respondents. They also cite research that investigates psychographic or behavioral differences between respondents and non-respondents. Paraphrasing their summary, non-respondents to surveys tend to be less educated, of lower socioeconomic class, white, of foreign-born parentage, older, married, residents in urban areas, and living in the Northeast U.S. Psychologically and behaviorally, nonrespondents also tend to be less emotionally stable, less effective as employees, less gregarious, lower in sense-of-leadership, less widely read, less proficient in writing ability, low on order and dependency but high on aggression, dominance, autonomy and intraception, and less responsible, less tolerant, and less intellectual in personality characteristics.

McDaniel et al. (1987) make the point that this assortment of apparent tendencies from individual studies of nonresponse may not be generalizable to other studies. Their results do strongly support the common contention that the "salience" of the survey topic to the survey recipient can have a substantial bearing on the probability that the survey will be completed and returned. This conforms with an earlier meta-analysis by Heberlein and Baumgartner (1978).

Regional differences in populations may have an effect on response rates. (See, for example, Jobber and Saunders (1988), Jobber et al. (1991), Ayal and Hornick, 1986, and evidence of Canadian-U.S. differences in Goyder (1985).)

Previous Solutions

Mitchell and Carson (1989, pp. 267-282) review the problem of non-response as it affects contingent valuation surveys. They review econometric methods for sample selection bias correction but conclude that "Unfortunately, these methods may be of limited use in contingent valuation studies when little or no information is available on factors affecting the probability of responding to the survey. ... We know of no CV study that has attempted to use these techniques to correct for sample selection bias."

The most common strategy for addressing non-response in environmental valuation surveys is to provide marginal means for a limited set of sociodemographic variables (e.g. income, age), calculated for the respondent sample and for the population it is intended to represent. If these means are similar, little more is said. The problem is that even though respondents and non-respondents may appear similar on a selection of observable sociodemographic attributes, there may be important unobservable forms of heterogeneity that affect both response propensities and demand for the environmental good. For example, respondents to a recreational fishing survey may tend to be more-avid anglers than nonrespondents, and avidity may not be measured.

Most researchers have treated survey non-response as a problem that has no easy solution. Whitehead (1991) asserts that correction for self-selection bias requires information about nonrespondents, obtained either through screener surveys or follow-up surveys. Edwards and Anderson (1987) emphasize that "from a practical standpoint the test for selection bias resulting from nonrespondents' self-censorship" requires that one "interview a high percentage of nonrespondents." They note that "This need presents a substantial, technical challenge for contingent valuation studies." In contrast, the present paper offers a tractable general strategy for modelling and correcting for nonresponse.

There are only a very few cases in the existing literature on environmental valuation via survey-based methods where researchers have attempted to control for nonresponse bias. Edwards and Anderson (1987) limit their empirical analysis to cases of questionnaire item nonresponse, rather than complete unit nonresponse. In particular, they find that omission of observations due to protest bids or zero willingness-to-pay does not appear to produce any additional selectivity bias. Aggressive nonresponse conversion efforts allowed them to achieve an eventually very high response rate, but no data were available on non-respondents who remained.

In two other cases, the task of nonresponse evaluation has been facilitated because the researchers have access to supplementary databases where other *individual-specific* information can be linked to each targeted potential respondent. (See Whitehead et al. (1993), Englin et al. (1996), and Fisher (1996).

This paper differs from previous and concurrent efforts to explore nonresponse bias in the environmental economics literature in that it illustrates a technique that can be applied with any mail survey conducted in the United States.[2]

3. Modelling Propensity to Respond to Our Specific Mail Survey

The mail survey data we will use for our illustration comes from a four-version survey of water-based recreational participation (at lakes, reservoirs, and rivers) within the Columbia River system in the U.S. states of Washington, Oregon, Idaho and Western Montana, plus the southern portions of the Canadian provinces of British Columbia and Alberta. The larger study is discussed in detail in Callaway et al. (1995) and portions are also summarized in Cameron et al. (1996).

A copy of one of the four different versions of our survey instrument was sent to each of

7034 addresses. We are interested, first of all, in modelling the propensity for each copy of the survey instrument to be returned. Of these targeted households, 2513 returned surveys that were sufficiently complete for their data to be included in our demand analyses, for an overall response rate of 35.73%.[3]

Another task we face is to control for the fundamentally different expected response rates in different sampling strata. It is not always feasible to rely upon a strictly representative sample from the general population in modeling the demand for environmental amenities. In many cases, it is expedient to combine a basic general population samples with other convenience samples. In our illustration, for example, we over-sample people who live in close proximity to the environmental good to be valued. We also include people who are known to be users of the resources in question, by intercepting them on site. We also incorporate a subsample of potential foreign users, drawn from major urban areas of the nearest cross-border regions. Dummy variables identifying our four auxiliary strata groups are also included in our response/nonresponse model.

For this survey, Dillman's prescriptions were followed as closely as possible, in order to maximize overall response rates. There are some key features of the survey design, however. First, each of the four different versions focusses on a separate geographic region. The visual aids (photographs, both actual and computer-modified) that accompanied each questionnaire were different. While effort was made to make the written portions of the survey as comparable as possible, their different regional emphases may have resulted in differing appeal or salience to different types of respondents. In our model, response decisions are allowed to vary systematically according to survey version.

The zip code information is the key to generating variables that may go part way towards capturing the salience of the survey topic as well as the demographic or socioeconomic characteristics of the potential respondent's neighborhood. In designing our response/nonresponse model, we need to keep in mind that for a mail survey in the US, often nothing is likely to be known about each member of the intended sample beyond their mailing address (including zip code) and what version of the survey they were sent. Even if a survey research firm protects respondent confidentiality by redacting the name and street address of each respondent, it is possible that zip codes can be retained, since they rarely would allow unique identification of any respondent.

We thus attempt to capture salience of the survey topic to each respondent using an array of proxy variables. Even from the cover of the survey--the first thing a respondent would see--our survey topic is easily construed to be water-based recreation. Each recipient's actual and potential experience with water based recreation could be expected to influence response propensity. One way to attempt to capture this potential experience is by using distances between the recipient's home and each of the bodies of water featured in their particular version of the questionnaire. First, we employ the individual distances between the recipient's origin zip code and each of the three or four specific "Federal Projects" along the Columbia River system which are singled out on his or her particular version of the questionnaire. These projects are either reservoirs behind hydroelectric dams or run-of-river stretches below these dams. The identities of these waters differs by survey version, so we interact each of these distances with dummy variables for each version. We have also calculated the distances from each respondent's zip code to each of the five nearest "other" fresh waters in the target region that were not included among the set of Columbia River projects. The average of these five distances is used to represent the accessibility of other nearby water recreation opportunities. This average is also interacted with survey version dummy variables, since the set from which these "other" waters is drawn differs across survey versions.[4]

Our distance data were calculated using the ZIPFIP computer program (Hellerstein et al., 1993). Given origin and destination zip codes, this software allows the user to generate approximate road distances between the centers of these zip codes. These distances are constructed from "great circle" distances, modified by a factor (unique to each state) that converts these distances into average road distances in a manner that controls for differing densities of roadways in each state. We identified the zip codes containing (or nearest to) each of the water bodies in our study region. In combination with the zip code for each potential respondent's address, we merge these distances with the response/nonresponse data.

The zip codes for the intended sample are also the key that allows us to merge the data set with a wide array of variables from the 1990 Census (available from the STF3 data tapes). All variables are descriptors of the zip code area, rather than the individual, but to the extent that the geographical areas covered by zip codes are relatively homogeneous, some portion of the heterogeneity in these characteristics across survey recipients can be captured by these aggregate data. The Census data provide zip code populations as well as counts of persons in each of a variety of categories. Appendix Table A-1 gives the Census-based variables we have considered, with details concerning how these were calculated from the constituent variables (using the variable names from the Census tapes). Census variables other than these ones may be important predictors of response/nonresponse in other applications.

A portion of our intended sample also consisted of Canadians. The general nonresponse research cited in the previous section certainly suggests that response rates should be allowed to vary systematically by country. If there is enough heterogeneity between different subregions of the Northwest U.S.--along the dimensions suggested for Canadians--there is good reason to allow for possible regional variation in response rates there based on similar arguments. Within the U.S., however, rather than using regional jurisdictional dummy variables such as states or counties, we elect to rely directly upon sociodemographic variations. These are the factors that such dummy variables would presumably be capturing.

It is also potentially important to control for any variations in survey format across the sample. Many environmental valuation surveys rely upon contingent scenarios (either contingent valuation or contingent behavior). These scenarios differ across versions of the survey instrument and these differences could conceivably influence response rates. In the wake of the debate about "embedding," other types of environmental valuation surveys have been designed to assess the effectiveness of different amounts of context for the valuation exercise. The level of descriptive detail for each scenario involved may differ a little or a lot across the individuals who make up the intended sample. Alternately, the nature or scope of the good to be valued may differ across survey version. These variations in the survey instrument may themselves lead to differential response rates, and this possibility has not generally been pursued in the literature on non-market resource valuation.

4. Modelling Water-Based Recreational Trips

The first stage in our analysis (Stage A) is a discrete choice (probit) model of the response/nonresponse decisions among the intended sample. The second stage, conditional on response, is a model for the number of water-based recreational trips taken by respondents. We demonstrate the empirical importance of controlling for nonresponse in two different types of trip demand specifications (Stage B1 and B2). In Stage B1, we model the individual's total demand for trips to *any* water in the region featured on the questionnaire, regardless of specific destination. (Since the set of relevant waters varies across versions, we estimate these second-stage models separately for each version.) We construct a rough proxy for "accessibility" of water recreation opportunities by calculating the average distance from the respondent's home zip code to the nearest five waters (be they federal projects or "other" waters). Ex ante, one would expect that the less accessible these recreation opportunities--i.e. the greater this average distance--the fewer trips an individual will take.

In an alternative specification, Stage B2, the dependent variable is trips to *one* particular water in the choice set on a particular survey version. This allows us to attain crude estimates of the apparent own- and cross-price effects.[5] We illustrate the potential consequences of nonresponse bias in these disaggregate specifications using one individual water from each of two versions of the survey.

Our analysis is intended to illustrate the that nonresponse selectivity effects are potentially a very important consideration, regardless of the demand specification employed.

5. Full Information Maximum Likelihood Estimation and a Compromise

1347 of our 2513 respondents took positive numbers of trips and these trip-takers averaged 12.4 trips apiece, with a standard deviation of 16.4 trips. A continuous distribution is assumed to be an adequate approximation to the conditional distribution of trip-taking propensities, suggesting a Tobit-type model for trips, in order to accommodate the sizeable observed frequency of zero trips.

Let $y_i^* = x_i'\beta + \epsilon_i$ be the latent propensity to return a completed response to the questionnaire that was mailed to household i. Since y_i^* is unobservable, the response/nonresponse outcome associated with each mailing is evaluated in terms of the associated observable variable $y_i = 1$ if the questionnaire is returned completed and $y_i = 0$ if the questionnaire is either not returned, or is returned insufficiently complete to be included in the analysis.[6] The vector of variables x_i includes Census zip code characteristics, variables that capture the differences among survey versions, the different sample strata, and the distance variables that partially proxy for the probable salience of the survey topic to targeted households.

Let $q_i^* = z_i'\gamma + \nu_i$ be each respondent's propensity to take water-based recreation trips to water bodies in the geographical area stipulated in each version of the questionnaire. If $q_i^* > 0$, then observed water recreation trips $q_i = q_i^*$. If we have $q_i^* \leq 0$, then observed trips will be $q_i = 0$.

We assume y_i^* is distributed $N(0,1)$, with variance normalized to unity because the binary nature of y_i will not allow us to discern the scale of y_i^*. Let q_i^* be distributed $N(0,\sigma^2)$, since the observable portion of q_i^* does allow the scale to be identified. We wish to jointly model both the individual's decision about whether to respond to the questionnaire, and, conditional on response, the number of trips taken. This is a Tobit model with a sample selection correction, ideally estimated by Full Information Maximum Likelihood (FIML). For a description, see Greene (1995, p. 624), summarized here in Appendix 1.

A full-information maximum likelihood (FIML) Tobit model with sample selection is available in the LIMDEP econometric software package as a one-line command. We have also experimented extensively with programming and estimating the FIML log-likelihood directly using the GQOPT general nonlinear function optimization package (Goldfeld and Quandt, 1995). While, in principle, this log-likelihood is valid, it is notoriously difficult to optimize, even starting from the consistent parameter estimates produced by two-stage models. We have found that even very trivial specifications, with only one or two regressors, fail to converge successfully.

Given the frustrations of FIML estimation, we opt to rely upon consistent (though not efficient) two-stage estimates in the tradition of Heckman (1979). Note that the variance-covariance matrix obtained for the second-stage Tobit model is not valid. We correct it using the method of Murphy and Topel (1985). See Appendix 2.

Another minor inconvenience is that a consistent estimate of the error variance for the Tobit latent variable is necessary before a point estimate of the predicted number of trips can be recovered from the second-stage Tobit model. The necessary calculations require the point estimates of β from the first-stage probit model. (See Appendix 3.)

The potential consequences of ignoring the problem of nonresponse to mail surveys for environmental valuation are too important to forestall examination of the issue until FIML estimation of models in this genre can be rendered generally tractable. Thus we proceed below with corrected two-stage estimation methods.

6. Results

The list in Appendix Table A-1 is an inventory of all of the Census variables which were examined in preliminary models. Only those variables that were persistently statistically significant determinants of response rates across a variety of exploratory specifications are included in the models to follow. In other applications, different variables may prove important.

Table 1 provides descriptive statistics for the entire intended sample of 7034 addresses. This is the universe of addresses to which questionnaires were mailed. The variables (described briefly in the body or footnotes to the table) either describe the type of subsample or are obtained by utilizing zip codes to calculate distances or to merge with the available Census data.

Table 2 gives the results for a pooled-data probit model that uses all 7034 addresses in the intended sample and attempts to explain provision of a usable response as a function of everything known about the zip code of the target household, the type of subsample it belonged to, and the version of the survey it received. We find that the "Phase 1" and "known-user"

TABLE 1

Descriptive Statistics[a], Response/Nonresponse Sample (n = 7034)
(means; standard deviations for non-binary variables in parentheses)

VERSION	1 (n = 1428)	2 (n = 1433)	3 (n = 2095)	4 (n = 2078)
OVERALL RESPONSE	0.3573	= 1 if response sufficiently complete (0 otherwise)		
VERSION RESPONSE	0.3480	0.3517	0.3518	0.3730
HAV-DIST	0.9414	= 1 if distance data available for all specified waters		
HAV-OTH	0.9663	= 1 if distance data available, 5 nearest "other" waters		
HAV-CENSUS	0.9555	= 1 if 1990 Census data available for zip code		
POP-2	0.4362	= 1 if adjacent county sample (0 otherwise)		
POP-3	0.05530	= 1 if "Phase 1" sample (0 otherwise)		
POP-4	0.08189	= 1 if "known user" sample (0 otherwise)		
POP-5	0.02132	= 1 if Canadian sample (0 otherwise)		
VERSION	-	0.2037	0.2978	0.2954
DIST-1	0.2386 (1.061)	0.3616 (0.9550)	0.8134 (1.481)	0.5058 (0.9349)
DIST-2	0.3467 (1.200)	0.3334 (0.9204)	0.4664 (0.8796	0.7981 (1.437))
DIST-3	0.2702 (1.088)	0.2892 (0.8043)	0.7835 (1.393)	0.6676 (1.233)
DIST-4	0.2579 (1.062)	0.4300 (1.210)	-	0.7302 (1.344)
DIST-OTHER	0.2090 (1.048)	0.1979 (0.6640)	0.3039 (0.6683)	0.3881 (0.7705)
PLANGIS	0.004839 (0.006962)	= Proportion of zip code population language-isolated		
PPUBINC	0.02456 (0.01493)	= Proportion of zip code population on public assistance		
PURBAN	0.5041 (0.3978)	= Proportion of zip code population in urban area		
PAGOCC	0.02690 (0.03265)	= Proportion of zip code population in ag., fishing, or forestry-related occupations		
PSSINC	0.1023 (0.04508)	= Proportion of zip code population on social security		

[a] Identity of the specific waters to which distances are measured differs across survey versions:
Version 1: DIST-1 = Hungry Horse Reservoir, DIST-2 = Lake Pend Oreille, DIST-3 = Lake Koocanusa, DIST-4 = Kootenai River, DIST-OTHER = other waters in the Version 1 region;
Version 2: DIST-1 = Dworshak Lake, DIST-2 = Clearwater River, DIST-3 = Lower Granite Lake, DIST-4 = Lake Pend Oreille, DIST-OTHER = other waters in the Version 2 region;
Version 3: DIST-1 = Lake Roosevelt, DIST-2 = Lake Umatilla, DIST-3 = Lower Granite Lake, DIST-OTHER = other waters in the Version 3 region;
Version 4: DIST-1 = Lake Roosevelt, DIST-2 = Dworshak Lake, DIST-3 = Lower Granite Lake, DIST-4 = Lake Pend Oreille, DIST-OTHER = other waters in the Version 4 region.

TABLE 2

Stage A: Probit Model for Survey Response/Nonresponse (n = 7034)
(point estimates; asymptotic t-ratios in parentheses; ** = 5%, * = 10% level)

VARIABLE	VERSION 1[a]	VERSION 2	VERSION 3	VERSION 4
CONSTANT	-5.4090 (-0.101)	-	-	-
HAV-DIST	-0.94358 (-8.038)**	-	-	-
HAV-OTH	5.8271 (0.109)	-	-	-
HAV-CENSUS	-0.041423 (-0.255)	-	-	-
POP-2	0.051603 (0.935)	-	-	-
POP-3	0.19816 (2.766)**	-	-	-
POP-4	0.45169 (5.787)**	-	-	-
POP-5	0.056020 (0.001)	-	-	-
VERSION DUMMY	-	-0.011991 (-0.133)	0.33752 (2.073)**	0.10617 (0.956)
DIST-1	0.029962 (0.251)	-0.012718 (-0.061)	-0.10331 (-1.242)	-0.13793 (-1.620)
DIST-2	0.00064882 (0.014)	-1.4318 (-2.676)**	-0.14959 (-1.464)	0.33916 (1.353)
DIST-3	0.24446 (0.898)	1.5193 (3.020)**	-0.017056 (-0.302)	-0.31424 (-1.164)
DIST-4	-0.22852 (-1.190)	0.15348 (3.289)**	-	0.16944 (1.325)
DIST-OTHER	-0.089274 (-2.334)**	-0.11004 (-1.284)	0.30916 (1.727)*	-0.27913 (-1.515)
PLANGIS	-5.9098 (-2.056)**	-	-	-
PPUBINC	-4.2524 (-3.395)**	-	-	-
PURBAN	0.14932 (2.449)**	-	-	-
PAGOCC	0.73584 (1.042)	-	-	-
PSSINC	1.0420 (2.392)**	-	-	-
Log $\mathcal{L}$	-4341.5			

[a] All parameters estimated in a single model. Version dummy variables are interacted with all distance variables because waters corresponding to each "numbered distance" differ across versions.

samples were statistically significantly more likely to respond. For known users, the subject matter of the questionnaire is undeniably salient, so this is not surprising. In contrast to expectations, the Canadian sample (population 5) does not appear to be statistically less likely to respond. However, this subsample is very small, and while comparable distance data were calculated by hand, no Census data were available for this sample, so the HAV-CENSUS indicator variable is highly correlated with membership in the Canadian subsample. Thus our finding may not be conclusive.

Version 3 of the questionnaire appears to have produced systematically larger response rates. Distances to three of the specific waters described in the questionnaire significantly influenced response rates for recipients of version 2 of the questionnaire. The effect was negative for one water, and positive for two others.[7] Distances to the nearest "other" waters appears to matter only for versions 1 and 3 of the questionnaire.

Among the Census variables examined, the most robustly individually statistically significant variables were those intended to capture language isolation, proportion on public assistance income, proportion urban, and proportion on social security income. Language isolation decreases response probabilities, as does a greater neighborhood prevalence of public assistance income. Recipients in more urbanized areas are more likely to respond, as are those from areas with higher level of employment in agriculture, fisheries or forestry, although the last is not statistically significant. Response propensity is also significantly higher, the greater the proportion of the neighborhood receiving social security income.[8]

Table 3 gives descriptive statistics for the sample of usable responses from each of two versions of the survey. For these subsamples, we have actual respondent-specific individual sociodemographic information, which certainly involves less measurement error than the zip code proportions used for all observations in the first stage response/nonresponse model. The

TABLE 3

Descriptive Statistics, Respondents Only
(means; standard deviations for non-binary variables in parentheses)

VARIABLE	VERSION 1 (n = 497)	VERSION 3 (n = 737)	VERSION 4 (n = 775)	DESCRIPTION
TOTAL-TRIPS	10.12 (17.43)	3.550 (7.532)	5.737 (12.80)	number of water-based recreation trips in 1993
SINGLE-SITE TRIPS	0.2676 (1.330)	-	0.2787 (1.485)	Version 1; trips to Lake Koocanusa Version 4; trips to Lake Pend Oreille
HAV-DIS	0.8350 (0.3715)	-	0.9458 (0.2265)	distance data available to all relevant waters
AVG-DIST	0.07074 (0.1242)	0.1037 (0.1105)	0.1199 (0.08057)	Average distance to nearest five waters of any description
HAV-INC	0.8330 (0.3734)	0.8521 (0.3552)	0.8490 (0.3582)	= 1 if respondent provided income data (0 otherwise)
INC	0.02945 (0.02885)	0.03193 (0.02860)	0.03331 (0.02969)	Respondent average monthly income in $ 100,000
HAV-AGE	0.9678 (0.1767)	0.9647 (0.1846)	0.9523 (0.2134)	= 1 if respondent provided age data (0 otherwise)
AGE	5.161 (1.734)	5.089 (1.821)	4.937 (1.878)	Respondent age (in tens of years)
FISH-LICENSE	0.6016 (0.4901)	0.4301 (0.4954)	0.4052 (0.4912)	= 1 if respondent holds a current fishing license (0 otherwise)
OWN-BOAT	0.5171 (0.5002)	0.3636 (0.4814)	0.3858 (0.4871)	= 1 if respondent owns a boat (0 otherwise)
IMR	0.9747 (0.2985)	1.025 (0.1218)	0.9672 (0.1812)	= fitted inverse Mill's ratio from first-stage probit model

individual "home zip code to each water" distances are the same values that were computed for the entire intended sample of 7034, so they are reused to calculate the accessibility variables used here. A Heckman-type two-stage method is used below, involving the additional explanatory variable constructed from the first stage: IMR. This is the standard inverse Mill's ratio for sample selectivity correction.

Table 4 contains results for the second stage Tobit regression models, for total trips to any water in a given region. We report the nonresponse-corrected and uncorrected estimates only for versions 1 and 3 of our survey, as nonresponse bias appears not to be statistically significant for versions 2 and 4 under this type of specification. Focussing first on the corrected models (the first column of results in each pair), note that there are many individually statistically significant parameter estimates.

Recall that the distances in this model have been simplified to measure an index of average distances to the nearest five waters in the region featured on the questionnaire (be it a focus water for that version, or any other water). The coefficient on this distance index is very strongly statistically significant. The incremental effect of variations in income is positive in both cases, but not significant. Age has a negative effect on trips in both versions. Not surprisingly, the (potentially somewhat endogenous) variables indicating possession of a fishing license or a boat are strongly significantly correlated with the latent number of trips underlying this type of Tobit specification.[9]

The coefficient on the IMR variable provides insights into the nonresponse bias. In models with an OLS second stage, the coefficient on the IMR is typically interpreted as the product of the correlation between the error in the response model and the (latent) trips model (ρ), and the error standard deviation in the latent trips model (σ). If ρ is zero, $\rho\sigma$ will be zero. The coefficient on IMR is statistically significant for survey versions 1 and 3, and in each case,

TABLE 4

Stage B1: TOBIT PARAMETER ESTIMATES: With and Without Nonresponse Selectivity Correction
(point estimates; Murphy-Topel corrected asymptotic standard errors in parentheses)
Dependent Variable: Total Number of Trips to *Any* Water in Region

	VERSION 1 (n = 497)		VERSION 3 (n = 737)	
VARIABLE	Corrected	Uncorrected	Corrected	Uncorrected
CONSTANT	9.2763 (1.284)	-0.93730 (-0.144)	12.695 (2.456)**	-7.0195 (-2.232)**
HAV-DIS	0.53492 (0.091)	-16.707 (-6.776)**	-	-
AVG-DIST	-38.778 (-4.116)**	-43.334 (-4.751)**	-26.359 (-5.735)**	-20.916 (-4.647)**
HAV-INC	0.75177 (0.260)	0.66804 (0.230)	1.5582 (0.955)	1.1215 (0.686)
INC	11.290 (0.321)	-1.0807 (-0.031)	6.2325 (0.325)	6.8328 (0.354)
HAV-AGE	32.242 (4.463)**	33.566 (4.663)**	8.2059 (2.386)**	9.0126 (2.611)**
AGE	-3.7163 (-5.700)**	-3.7889 (-5.779)**	-1.3805 (-4.198)**	-1.4887 (-4.494)**
FISH-LICENSE	7.6388 (3.657)**	8.0040 (3.816)**	3.8794 (3.602)**	4.0515 (3.730)**
OWN-BOAT	7.1948 (3.558)**	7.6478 (3.768)**	6.7424 (6.191)**	7.3264 (6.696)**
IMR	-24.809 (-3.235)**	-	-18.579 (-4.673)**	-
SIGMA	18.840 (25.559)**	18.995 (25.513)**	11.240 (24.704)**	11.388 (24.631)**
Log $\mathcal{L}$	-1566.8	-1572.1	-1532.7	-1543.7
Fitted trips	23.91	12.47	13.05	6.275

is negative and rather large. This suggests a substantial negative correlation between the response/nonresponse decision and the trip-taking behavior for these two subsamples.[10]

The second column in each pair in Table 4 shows the results of an analogous model estimated without benefit of control for non-random nonresponse via the IMR term. The coefficient on the IMR term reveals that failure to control for nonresponse will lead to underestimates of the number of trips. For all versions of the survey where the selectivity effects are significant, our models suggest that unobserved factors which make targeted households *less likely* to respond to our questionnaire than our response model predicts also make them *more likely* to take water-based recreational trips than our trips models would predict. While it is sometimes difficult to label these unobserved factors, a reasonable speculation would be that these factors include tastes for outdoor activity, general levels of physical health and energy, and/or family composition, among others. One interpretation is that people who are busy engaging in activities related to outdoor freshwater-based recreation are too busy to waste time responding to our questionnaires. Alternatively, infrequent participants and non-users of these waters may have found it much easier to fill in our questionnaire. Rather than remembering numbers of trips to each site in different months, these people would simply have to fill in a lot of zeros.[11] It is worth noting that the finding of a negative error correlation is at odds with the common assumption that households with higher participation in an activity should be more likely to respond to surveys about that activity.

Setting to zero the correlation between the errors in the response/nonresponse model and the trips model is equivalent to eliminating the inverse Mill's ratio term from the calculation of fitted trips in the second-stage Tobit model. As expected, given the strongly significant negative coefficients on this term for versions 1 and 3, removal of sample selectivity results in the implication that a truly random sample from the population would have predicted much higher numbers of trips on average, and therefore greater aggregate utilization of the resource than implied by any demand model ignoring the selectivity problem. The differences are summarized in the last row of Table 4.

We should also note that failure to control for nonresponse can also distort the coefficient on the distance variable. For version 1, the corrected coefficient is -39, while the uncorrected one is -43, an overestimate of responsiveness. For version 3, the corrected estimate is -26, whereas the uncorrected one is -20, an underestimate of responsiveness. Clearly there is no generalizable bias. The AVG-DIST variable plays the role that the price variable would play in more-sophisticated consumer demand specifications. The estimated price coefficient is typically a key ingredient in consumer's surplus calculations. This offers some evidence that any eventual welfare estimates could potentially be distorted in a nontrivial fashion by failure to account for nonresponse. The distance (price) variable is probably most affected because distances are embodied in the IMR nonresponse correction terms. Few of the other slope parameters in this model appear to be seriously distorted by failure to correct for selectivity.

In our second type of specification, we explore a more-standard demand specification using our illustrative sample. These models concern demand for trips to a single site, and include travel costs for that site and other individual sites mentioned in that version (as well as "other" waters). Due in part to the smaller numbers of trips to individual waters, the two-stage Tobit specification did not converge for all individual waters.[12] For version 1, we illustrate with a demand model for Lake Koocanusa (W3) and for version 4, we provide estimates for demand for Lake Pend Oreille (W4).

We can now interpret the round-trip travel cost variables as prices in these simple demand models. The own-water price effects are negative and significant in the corrected specifications (with statistically significant IMR terms), suggesting downward-sloping demand curves. There is also evidence from version 1 that Hungry Horse Reservoir (W1) is viewed by recreationists as a substitute for Lake Koocanusa. Hungry Horse is often an overflow recreation site for Koocanusa. For version 4, it appears that Lake Roosevelt (W1) is a substitute for Lake Pend Oreille. Camping is the most important activity at Lake Roosevelt, as at Pend Oreille, and the large population centers in the region are located between these two sites, so it is reasonable that they might be viewed as substitutes. In no case is the average distance to the nearest five "other" waters influential.

The effect of holding a fishing license seems to matter for trips to Lake Koocanusa, but not for Lake Pend Oreille. Fishing is the most popular activity at Koocanusa, and most fishing there is done from boats. Boat ownership is significant for Koocanusa, but is not important in explaining trips to Pend Oreille. At Pend Oreille, camping and picnicking are the most important activities, so fishing licenses and boat ownership may well have not much of an effect on demand for this water.

Potential distortions to the demand relationship because of nonresponse bias remain our primary consideration in these examples, however. In both of the corrected models in Table 5, the inverse Mills ratio term, IMR, is strongly statistically different from zero and negative. This again suggests that unobserved heterogeneity that makes recipients more likely to respond to the survey also makes them less likely to take trips to each of these waters.

Comparison of the corrected and uncorrected models in Table 5 reveals the implications of failing to control for nonresponse. The magnitude of the own-price effect for Lake Koocanusa (in Version 1) is distorted slightly upwards, while that for Pend Oreille (in Version 4), is distorted substantially upwards (from -40 to -57). If the negative slope for quantity as a function of price is too great, the demand curve as usually depicted will be too flat. Welfare estimates such as consumer surplus, based on projection of the estimated demand curve up to

TABLE 5

Stage B2: TOBIT PARAMETER ESTIMATES: With and Without Nonresponse Selectivity Correction
(point estimates; Murphy-Topel corrected asymptotic standard errors in parentheses)
Dependent Variable = Number of Trips to **Specific Single Water** in Region

VERSION	1 (n = 497) Lake Koocanusa (W3)		4 (n = 775) Lake Pend Oreille (W4)	
VARIABLE[a]	Corrected	Uncorrected	Corrected	Uncorrected
CONSTANT	7.7610 (1.541)	-0.36457 (-0.086)	-2.0284 (-0.007)	-20.373 (-0.064)
HAV-COST	-9.3358 (-2.471)**	-8.7272 (-2.258)**	6.7455 (0.022)	11.155 (0.035)
RTC-W1	28.348 (2.855)**	35.548 (3.541)**	29.685 (3.191)**	43.012 (4.716)**
RTC-W2	23.307 (1.524)	26.335 (1.666)*	34.774 (0.938)	40.528 (1.234)
RTC-W3	**-58.351** **(-2.858)****	**-61.149** **(-2.965)****	-23.500 (-0.640)	-27.670 (-0.856)
RTC-W4	0.19742 (0.008)	-5.4258 (-0.212)	**-40.798** **(-4.781)****	**-57.414** **(-6.418)****
DIST-OTHER	-0.50069 (-1.509)	-0.51961 (-1.526)	-0.64864 (-0.275)	-0.99794 (-0.455)
HAV-INC	-0.20109 (-0.116)	0.21138 (0.121)	-1.5609 (-0.654)	-0.54145 (-0.235)
INC	24.844 (1.094)	18.336 (0.792)	-56.832 (-1.287)	-65.482 (-1.535)
HAV-AGE	3.6719 (1.104)	3.6488 (1.075)	3.7734 (1.020)	3.4295 (0.963)
AGE	-0.73097 (-1.989)**	-0.81161 (-2.118)**	-0.62715 (-1.292)	-0.70838 (-1.504)
FISH-LICENSE	-0.34727 (-0.306)	0.20197 (0.177)	1.0845 (0.794)	0.68323 (0.502)
OWN-BOAT	3.1394 (2.587)**	3.5133 (2.813)**	1.8448 (1.304)	2.9300 (2.061)**
IMR	-7.0387 (-2.804)**	-	-16.170 (-3.906)**	-
SIGMA	5.4358 (8.275)**	5.6438 (8.220)**	6.2956 (8.039)**	6.5700 (7.962)**
Max Log $\mathcal{L}$	-216.66	-220.75	-186.91	-196.64

[a]Round-trip costs ($'00) for different versions: Version 1: RTC-W1 = Hungry Horse Reservoir, RTC-W2 = Lake Pend Oreille, RTC-W3 = Lake Koocanusa, RTC-W4 = Kootenai River; Version 4: RTC-W1 = Lake Roosevelt, RTC-W2 = Dworshak Lake, RTC-W3 = Lower Granite Lake, RTC-W4 = Lake Pend Oreille.

the choke price, will imply too low a choke price if selectivity is not recognized in the estimation process. Resulting consumer's surplus estimates will then be too small (at least in these specifications--recall that the biases were mixed in Table 4 for total trips).

The apparent substitutability of Koocanusa for Hungry Horse and Pend Oreille for Roosevelt is also exaggerated if selectivity effects are ignored. Existence of satisfactory substitutes lessens the impact of compromises in the quality or availability of a particular water. The false impression that good substitutes exist could lead to undervaluing of social losses due to damage or reduced access to any of these waters.

7. Conclusions and Suggestions for Subsequent Mail-Survey-Based Research

We have demonstrated that non-random nonresponse to a mail survey has the potential to cause substantial distortion in empirical estimates of subsequent econometric models. Our available illustrative sample of data is far from ideal for truly detailed utility-theoretic demand modelling of environmental values. Nevertheless, the persistent appearance of nontrivial biases in key parameter estimates, even in a selection of simplistic demand models, certainly leads one to suspect that analogous biases would be possible in more sophisticated demand and/or utility models. This inference can readily be extended beyond the boundaries of environmental valuation to all types of other studies using mail survey data.

The main contribution of this paper is its demonstration that reliance on little more than the zip code information available for each household in the target sample allows one to reconstruct a selection of variables that can potentially be used in a response/nonresponse discrete choice model. Since our surveys ask respondents to consider environmental goods at specific identifiable geographical locations, distance is likely to be related to the salience of the good. We also rely on the 1990 Census, aggregated to the level of zip codes, to provide crude measures of the sociodemographic characteristic of each potential respondent's neighborhood. We can also control for membership in different types of subsamples. In our example, many of these Census variables are shown to make a statistically significant contribution to explaining a potential respondent's propensity to complete and return our questionnaire. Similar geographic or sociodemographic considerations or convenience samples will be present in many other types of surveys and the implications of our example extend to any study using these data.

It would have been advantageous to have access to additional variables for this study. In some cases of environmental valuation, for example, it may be possible to solicit from each state information on the numbers of fishing licenses per zip code and/or the number of licensed boat trailers per zip code (for example).[13] One must assume that the direct mail advertising industry also knows a lot about the preferences of US residents by zip code. While such data are unlikely to be free, it may be possible to acquire data on the number of subscribers to certain publications by zip code, or membership in certain organizations. A selection of such zip code frequency variables could paint an even more informative picture of probable survey topic salience.

In order to focus attention on the problem of nonresponse bias in survey research concerning the demand for environmental goods, our first featured examples employ a very simplified model of demand for a aggregates of similar environmental goods. Clear nonresponse biases can show up here. Next, we resort to demand models for individual localized goods. This second class of models is most theoretically satisfying, and provides evidence that the basic implications of demand theory are met. However, the difficulties we experienced in getting these models to converge leads us to offer them as supplementary evidence, rather than to present them as our main results.

Implementing a model of response/nonresponse requires only that sufficient geographic information be retained for the entire intended sample. Researchers must also have access to recent Census data at a corresponding level of aggregation, as well as relevant distance-calculating software.

What is our recommendation? Any researcher using mail survey data should be strongly encouraged to *plan for*, and then to undertake, explicit modelling of response/nonresponse to his or her survey instrument in a manner analogous to that presented here. This is especially important if one expects considerable heterogeneity in the sociodemographic characteristics of potential respondents, or if geographical proximity to the place(s) or object(s) featured in the subject matter of the survey varies substantially across potential respondents. It is also important if there are different versions of the survey, or if portions of the working sample consist of non-random convenience samples appended to a base sample that is reasonably representative. The key insight is that *without* formal nonresponse modelling and correction, the default presumption must be that substantial nonresponse biases could easily be present in any statistical work conducted using only a sample of mail survey *respondents*. These biases can distort not only estimates of the level of demand in the population, but also estimates of the degree of substitutability among goods and overall welfare calculations.

REFERENCES

Ayal, I. and J. Hornik (1986) "Foreign Source Effects and Response Behavior in Cross-National Mail Surveys," *International Journal of Research in Marketing*, 3, pp. 157-67.

Callaway, John M., Shannon Ragland, Sally Keefe, Trudy Ann Cameron, W. Douglass Shaw (1995) *Columbia River System Operation Review Recreation Impacts: Demand Model and Simulation Results*, report prepared for U.S. Army Corps of Engineers, Portland Oregon, by Hagler Bailly Consulting, Inc., Boulder, CO.

Cameron, T.A., W.D. Shaw, S.R. Ragland, J.M. Callaway, and S. Keefe (1996) "Using Actual and Contingent Behavior Data with Differing Levels of Time Aggregation to Model Recreation Demand," *Journal of Agricultural and Resource Economics* (forthcoming)

Dillman, D.A. (1978) *Mail and Telephone Surveys: The Total Design Method*, New York: Wiley.

Dillman, D.A. J.J. Dillman, and C.J. Makela (1984) "The Importance of Adhering to Details of the Total Design Method (TDM) for Mail Surveys," in D.C Lockhart (ed.) *Making Effective Use of Mailed Questionnaires*. New Directions for Program Evaluation, no. 21. San Francisco: Jossey-Bass.

Edwards, S.F. and G.D. Anderson (1987) "Overlooked Biases in Contingent Valuation Surveys: Some Considerations," *Land Economics*, 63(2), pp. 168-178.

Englin, J., J.S. Shonkwiler, and W.D. Shaw (1996, in progress) "Count Models with Self-Selectivity Corrections: An Application to Recreational Demand Modelling," manuscript, Department of Applied Economics and Statistics, University of Nevada, Reno, NV 89557.

Fisher, Mark R. (1996) "Estimating the Effect of Nonresponse Bias on Angler Surveys," *Transactions of the American Fisheries Society*, 125, 118-126.

Fox, R.J., M.R. Crask, and J. Kim, "Mail Survey Response Rate: A Meta-Analysis of Selected Techniques for Inducing Response," *Public Opinion Quarterly*, 52, pp. 467-491.

Goyder, J. (1985) "Nonresponse on Surveys: A Canada-United States Comparison," *Canadian Journal of Sociology*, 10(3), pp. 231-251.

Greene, W.H. (1993) *Econometric Analysis*, New York: Macmillan.

Greene, W.H. (1995) *LIMDEP: User's Manual and Reference Guide, Version 7.0*. Bellport, NY: Econometric Software, Inc.

Heberlein, T., and R. Baumgartner (1978), "Factors Affecting Response Rates to Mailed Questionnaires: A Quantitative Analysis of the Published Literature," *American Sociological Review*, 43, pp. 447-462.

Heckman, J. (1979) "Sample Selection Bias as a Specification Error," *Econometrica* 47, pp. 153-161.

Hellerstein, D., D. Woo, D. McCollum, and D. Donnelly (1993) *ZIPFIP: A ZIP and FIPS Database Users Manual*, Economic Research Service (ERA-NASS), 340 Victory Drive, Herndon, VA 22070 (1-800-999-6779)

Jaffe, E.D. (1982) "The Efficacy of Mail Surveys in Developing Countries--The Case of Israel," *European Research*, 10, pp. 102-104.

Jobber, D., H. Mizra, and K.H. Wee (1991) "Incentives and Response Rates to Cross-National Business Surveys: A Logit Model Analysis," *Journal of International Business Studies*, pp. 711-721.

Jobber, D., and J. Saunders (1988) "An Experimental Investigation into Cross-National Mail Survey Response Rates," *Journal of International Business Studies*, pp. 483-489.

McDaniel, S.W., C.S. Madden, and P. Verille, (1987) "Do Topic Differences Affect Survey Non-Response?" *Journal of the Market Research Society*, 29 (1), pp. 55-66.

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

Murphy, K.M. and R.H. Topel (1985) "Estimation and Inference in Two-Step Econometric Models," *Journal of Business and Economic Statistics* 3 (4), 370-379.

Whitehead, J.C. (1991) "Environmental Interest Group Behavior and Self-Selection Bias in Contingent Valuation Mail Surveys," *Growth and Change*, 22(1), pp. 10-21.

Whitehead, J.C., P.A. Groothuis, and G.C. Blomquist (1993) "Testing for Non-Response and Sample Selection Bias in Contingent Valuation; Analysis of a Combination Phone/Mail Survey," *Economics Letters* **41**, pp. 215-220.

Wiseman, F. and M. Billington (1984) "Comment on a Standard Definition of Response Rates," *Journal of Marketing Research*, 21, pp. 336-338.

TABLE A-1

Candidate Census Variables for Response/Nonresponse Probit Submodel

ACRONYM	CONSTRUCTION FROM STANDARD CENSUS STF3 VARIABLES	INTERPRETATION
PERSONS	P1_1	Population of zip code area
PURBAN	P6_1/P1_1	Proportion urban
PWHITE	P8_1/P1_1	Proportion White
PBLACK	P8_2/P1_1	Proportion Black
PAMIN	P8_3/P1_1	Proportion Native American
PASIAN	P8_4/P1_1	Proportion Asian
POTHER	P8_5/P1_1	Proportion other ethnicity
PLANGIS	(P29_2+P29_4+P29_6)/P1_1	Proportion language-isolated
PLTERM	P43_1/P1_1	Proportion long-term resident (same dwelling in 1985)
PCOLL	(P60_6+P60_7)/P1_1	Proportion college-educated
PAGIND	P77_1/P1_1	Proportion in agriculture, fishing or forestry industries
PAGOCC	P78_9/P1_1	Proportion in agriculture, fishing or forestry occupations
PSSINC	P94_1/P1_1	Proportion on social security income
PPUBINC	P95_1/P1_1	Proportion on public assistance income
PRETINC	P96_1/P1_1	Proportion with retirement income
INCM	P80A_1/1000	Median household income ($'000)
RENT	H43A_1/1000	Median rental rate ($'000)
VALUE	H61A_1/1000	Median house value ($'000)

Appendix 1

Tobit Model with Sample Selection

A little intuition will help with the development of the appropriate log-likelihood function. The domain of the joint density function can be partitioned into three distinct regions. The first region is characterized by $y_i^* > 0$ and $q_i > 0$ (respondents with nonzero observed trips). The second region has $y_i^* > 0$ and $q_i = 0$ (respondents with zero trips). The third region has $y_i^* < 0$ and thus q_i unknown (the nonrespondents).

For observations in the first region, the joint density can be conveniently expressed as the marginal density of q_i (observed) times the conditional density of y_i^* given the value of q_i. The random variable q_i is $N(z_i'\gamma, \sigma^2)$ and $f(y_i^* | q_i)$ is also normal with mean $x_i'\beta + \rho[(q_i - z_i'\gamma)/\sigma]$ and variance $(1 - \rho^2)$, since the variance of y_i^* is normalized to unity. The term for the marginal distribution of q_i will look like the ordinary maximum likelihood regression formula. The term for the conditional distribution will look like the term for a conventional MLE probit model for the positive domain of the latent variable. For this region, then, the contribution of one observation to the log-likelihood function is:

$$\log \mathcal{L}_{1i} = \{ -.5 \log(2\pi) - \log \sigma - .5 \, [\, (q_i - z_i'\gamma)^2/\sigma^2 \,] \} + \log[1 - \Phi(R_i)]$$

where $R_i = - \{ x_i'\beta + \rho[(q_i - z_i'\gamma)/\sigma] \} / (1 - \rho^2)^{.5}$.

For the second region, we assume that all values of $q_i^* < 0$ are manifested in the observed data as $q_i = 0$. Here, we must use the appropriate cumulative density associated with the bivariate normal distribution. If $\Phi_2(a,b,\rho)$ denotes the cumulative standard bivariate normal density function evaluated up to limits a and b, the log-likelihood terms for observations in this second region are given by:

$$\log \mathcal{L}_{2i} = \log[\, \Phi_2 (\, x_i'\beta, \, -z_i'\gamma/\sigma, \, -\rho \,)$$

For the third region of the domain of the joint density, all that is known is that $y_i^* < 0$, so we use the simple marginal distribution of y_i^*, employing a term like the one that applies to the negative domain of a conventional probit log-likelihood:

$$\log \mathcal{L}_{3i} = \log \Phi \, [\, -x_i'\beta \,].$$

Putting all three of these terms together, the full log-likelihood objective function can be expressed as:

$$\max_{\beta,\gamma,\sigma,\rho} \log \mathcal{L} = \Sigma_{y_i=1, q_i>0} \log \mathcal{L}_{1i} + \Sigma_{y_i=1, q_i=0} \log \mathcal{L}_{2i} + \Sigma_{y_i=0} \log \mathcal{L}_{3i}$$

Appendix 2

Murphy-Topel Corrected Second-Stage Variance-Covariance Matrix

Since FIML estimates and the desirable variance-covariance matrix cannot be attained in this application, we adopt the framework of Murphy and Topel (1985) in order to correct the second-stage variance-covariance matrix in our two-stage Tobit model. Let $N = nr + nn$ be the total number of observations in the target sample, with nr being the number of respondents and nn the number of nonrespondents. Let $I_i = 1$ if $q_i > 0$, and $I_i = 0$ if $q_i = 0$. The two separate log-likelihood functions employed in the two-stage method are:

$$\log\mathcal{L}_1 = \Sigma_N\, y_i \log \Phi(x_i'\beta) + (1 - y_i) \log [1 - \Phi(x_i'\beta)], \text{ and}$$

$$\log\mathcal{L}_2 = \Sigma_{nr} (-I_i/2)[\log(2\pi) + \log\sigma^2 + ((q_i - z_i'\gamma - \gamma_\lambda\lambda_i)/\sigma)^2] +$$

$$(1 - I_i) \log[1 - \Phi((z_i'\gamma + \gamma_\lambda\lambda_i)/\sigma)],$$

where $\lambda = \phi(x_i'\beta)/\Phi(x_i'\beta)$. If we now define $\theta = (\gamma', \gamma_\lambda, \sigma)'$, Murphy and Topel (1985) demonstrate that the correction formula for the second stage standard error estimates involves four matrices:

$$R_1 = -E[(\partial\log\mathcal{L}_1/\partial\beta)(\partial\log\mathcal{L}_1/\partial\beta'] = -E[\partial^2\log\mathcal{L}_1/\partial\beta\partial\beta']$$

$$R_2 = -E[(\partial\log\mathcal{L}_2/\partial\theta)(\partial\log\mathcal{L}_2/\partial\theta'] = -E[\partial^2\log\mathcal{L}_2/\partial\theta\partial\theta']$$

$$R_3 = -E[(\partial\log\mathcal{L}_2/\partial\beta)(\partial\log\mathcal{L}_2/\partial\theta'] = -E[\partial^2\log\mathcal{L}_2/\partial\beta\partial\theta']$$

$$R_4 = -E[(\partial\log\mathcal{L}_1/\partial\beta)(\partial\log\mathcal{L}_2/\partial\theta']$$

The matrices R_1 and R_2 can be replaced by the inverses of the uncorrected estimators for the asymptotic covariance matrices for the first stage probit and the second stage Tobit coefficients, respectively. Matrices R_3 and R_4 must be specially constructed. Based on the two-stage log-likelihood expressions, we have the vectors of derivatives:

$$\partial\log\mathcal{L}_1/\partial\beta = \Sigma_N \{ y_i\lambda_i - (1-y_i)\, \phi(x_i'\beta)/(1 - \Phi(x_i'\beta)) \}\, x_i$$

$$\partial\log\mathcal{L}_2/\partial\gamma = \Sigma_{nr} (1/\sigma) [I_i ((q_i-z_i'\gamma-\gamma_\lambda\lambda_i)/\sigma) - (1-I_i)\, \phi((z_i'\gamma+\gamma_\lambda\lambda_i)/\sigma)/(1 - \Phi((z_i'\gamma+\gamma_\lambda\lambda_i)/\sigma))]\, z_i$$

$$\partial\log\mathcal{L}_2/\partial\gamma_\lambda = \Sigma_{nr} (1/\sigma) [I_i ((q_i-z_i'\gamma-\gamma_\lambda\lambda_i)/\sigma) - (1-I_i)\, \phi((z_i'\gamma+\gamma_\lambda\lambda_i)/\sigma)/(1 - \Phi((z_i'\gamma+\gamma_\lambda\lambda_i)/\sigma))]\, \lambda_i$$

$$\partial\log\mathcal{L}_2/\partial\beta = \Sigma_{nr} \{ [I_i ((q_i-z_i'\gamma-\gamma_\lambda\lambda_i)/\sigma) - (1-I_i)\, \phi((z_i'\gamma+\gamma_\lambda\lambda_i)/\sigma)/(1 - \Phi((z_i'\gamma+\gamma_\lambda\lambda_i)/\sigma))] *$$

$$(\gamma_\lambda/\sigma) [(x_i'\beta)\, \lambda_i - \lambda_i^2] \}\, x_i.$$

and the scalar derivative:

$$\partial\log\mathcal{L}_2/\partial\sigma = \Sigma_{nr} (1/\sigma) \{I_i [((q_i-z_i'\gamma-\gamma_\lambda\lambda_i)/\sigma)^2 - 1] +$$

$$(1-I_i) ((z_i'\gamma+\gamma_\lambda\lambda_i)/\sigma)\ \phi((z_i'\gamma+\gamma_\lambda\lambda_i)/\sigma)/(1 - \Phi((z_i'\gamma+\gamma_\lambda\lambda_i)/\sigma))$$

The complete vector $\partial\log\mathcal{L}_2/\partial\theta$ is constructed from $((\partial\log\mathcal{L}_2/\partial\gamma)',(\partial\log\mathcal{L}_2/\partial\gamma_\lambda)',\partial\log\mathcal{L}_2/\partial\sigma)')$. The outer products of the appropriate vectors of derivatives are used to calculate R_3 and R_4.

Once the component R matrices have been calculated, the corrected variance-covariance matrix for the parameter vector of the second-stage tobit model will be $\sqrt{n}(\theta\text{-}\theta) \sim N(0,\Sigma)$, where

$$\Sigma = R_2^{-1} + R_2^{-1}[R_3'R_1^{-1}R_3 - R_4'R_1^{-1}R_3 - R_3R_1^{-1}R_4]\, R_2^{-1}$$

Appendix 3

Consistent Estimation of Tobit Conditional Error Variance

As in Greene (1993, p. 711), define:

$$\alpha_y = -x_i'\beta$$

$$\lambda_i = \phi(\alpha_y)/\Phi(\alpha_y)$$

$$\delta_i = \lambda_i [\lambda_i + x_i'\beta]$$

For each observation, i, the true error variance would be $\sigma_i^2 = \sigma_\epsilon^2(1 - \rho^2\delta_i)$. The average variance for the sample errors would converge in the limit to:

$$\text{plim } (1/n)\ \sigma_i^2 = \sigma_\epsilon^2(1 - \rho^2\delta^*)$$

where δ^* is the mean of the δ_i values. The maximum likelihood second stage Tobit algorithm provides an estimate, σ_0^2, of the desired quantity plim $(1/n)\sigma_i^2$.

Another necessary component is provided by the square of the coefficient on the inverse Mill's ratio term, λ_i. Let this coefficient be denoted γ_λ. We can use the result that plim $\gamma_\lambda^2 = \rho^2\sigma_\epsilon^2$. The first-stage probit model provides individual estimates of δ_i and plim $(1/n)\Sigma_i\delta_i = \delta^*$. Finally, we can generate a consistent estimator for the desired σ_ϵ^2 using the formula:

$$\sigma_\epsilon^2 = \sigma_0^2 + \delta^*\ \gamma_\lambda^2.$$

Tobit models produce estimated parameters, which, when employed in linear combination with the explanatory variables, produce fitted values of the Tobit "index." This index is the conditional expected value of the latent q_i^* variable. Fitted values of $q_i^* < 0$ are interpreted as zero fitted values of the observable trips variable, q_i. To determine the expected number of trips for a given vector of explanatory variables, the index must therefore be manipulated somewhat further. The expected number of trips is given by the probability of positive trips times the expected number of trips, conditional on trips being positive. This conditional probability depends upon the estimate of the error variance. One must be careful to use σ_ϵ^2 rather than the value of σ_0^2 produced automatically by the second-stage ordinary Tobit estimator.

Recall that the expected value of a standard normal random variable truncated below at c is given by $\phi(c)/[1 - \Phi(c)]$. Thus the implied conditional density function for individual i, truncated below at zero, will have a mean of

$$(z_i'\gamma+\gamma_\lambda\lambda_i) + \sigma_\epsilon\ [\ \phi((z_i'\gamma+\gamma_\lambda\lambda_i)/\sigma_\epsilon)/\Phi((z_i'\gamma+\gamma_\lambda\lambda_i)/\sigma_\epsilon)\ ].$$

The fitted $E[q_i]$ will be this value multiplied by the fitted probability of positive trips for this individual: $\Phi((z_i'\gamma+\gamma_\lambda\lambda_i)/\sigma_\epsilon)$. To simulate circumstances with no non-response bias, we can set $\gamma_\lambda\lambda_i = 0$ for all observations.

ENDNOTES

1. Some subsequent research concerning the prescriptions in Dillman (1978) is described in Dillman et al. (1984). Twenty-nine different elements of the "total design method" were either adhered-to or not for samples taken from eleven different states in the U.S. and the consequences for response-rates evaluated.

2. For a number of other countries, analogous methods are potentially feasible, depending on the availability of similar types of Census and distance data.

3. Wiseman and Billington (1984) address the issue of standardizing the definition of "response rate" in applied statistics. In the present study, the response rate is defined as the number of usable returned questionnaires divided by the total number of questionnaires mailed out. No adjustments are included for "returned undeliverable" or other exclusions that are occasionally allowed before making this computation.

4. Saltwater recreation opportunities may be viewed by some households as substitutes for the freshwater recreation opportunities they are being asked to consider. Saltwater resources were not considered in this survey.

5. Trips to any one water mentioned in any one survey version are much more sparse than total trips to all waters for that version, so not all submodels converge.

6. In a more-elaborate model, the specification could distinguish between complete non-response and unusable responses. However, this would require a trivariate joint density for FIML estimation.

7. Since distances to waters may be negatively correlated with distances to major urban areas, which have not been controlled for in our models, these results may be open to different interpretations.

8. Varying degrees of multicollinearity among some of the Census variables exist, but the purpose of the first-stage response/nonresponse model is to predict response probabilities, so this problem is not too troubling. The important result is that there is significant systematic variation in response probabilities.

9. Keeping in mind that the aggregate total of all types of water-based recreation is being modelled in this illustrative application (sight-seeing, camping, picnicking, etc., not just fishing trips, for example), the fishing license and boat-ownership dummy variables are less likely to be completely jointly determined with the dependent TRIPS variable.

10. The implied point value of ρ exceeds one in absolute value. In finite samples, and without parameter restrictions, this is possible.

11. We owe this eminently sensible explanation to Michael Hanemann.

12. Englin et al. (1996) consider Poisson-based selectivity models.

13. We use individual data on fishing licenses and boat ownership available for the respondent sample in the demand portion of our two-stage modelling exercise, but analogous zip code level variables could also contribute substantially to capturing the salience of water-recreation issues to the overall target population.