Treatment of Incomplete and Missing Covariate

Information in a Bayesian Generalized Linear Model

of Marine Recreational Anglers' Choice of Fishing Site

Selected Paper for the

Annual Meeting of the American Agricultural Economics Association

Long Beach, CA

July 28 – 31, 2002

Dale K. Graybeal*

Economics Graduate Program

North Carolina State University

Raleigh, NC

* The author wishes to thank Brad Gentner of the National Marine Fisheries Service for his assistance in obtaining the data for this research. Also, thanks to Wally Thurman for his helpful comments and assistance.

Treatment of Incomplete and Missing Covariate Information in a Bayesian Generalized Linear Model of Marine Recreational Anglers' Choice of Fishing Site

ABSTRACT

Economic surveys often report income via a categorical variable, and income information is often missing altogether for a large fraction of the sample. The Bayesian inferential framework allows one to specify and estimate models for incomplete and missing covariate information. Here multiple choice models of recreational anglers' choice of fishing site are estimated and alternative specifications for incomplete and missing income data are compared.

1. Introduction

The random utility model has become a standard tool for estimating the value of non-market goods when consumers choose from among a discrete set of alternatives – see Sandefur *et al*. (1996) for an introductory review. Parson's online bibliography (Parson's , 2002) lists 121 citations that address the estimation of random utility models in the context of recreation demand analysis. Many of these articles use a random utility model to estimate the change in consumer welfare associated with a change in the quality of a natural resource.

The key to welfare analysis is the construction of a suitable proxy for the opportunity cost of choice alternatives. Travel cost is often considered a suitable measure of opportunity cost when the consumer chooses from among a set of recreation sites. Travel cost is rarely, if ever, directly observed and must be estimated on the basis of available information. The travel cost associated with a particular site consists of (i) direct travel costs such as gasoline expenditures, tolls, and depreciation of the automobile, and (ii) indirect travel costs such as the expenditure of time. Direct travel costs might vary from consumer to consumer, though one might be comfortable using some average cost of travel. The indirect travel cost – i.e., the opportunity cost of time, presents the difficult theoretical and empirical problem.

On the basis of microeconomic theory and the assumption that an interior solution to the labor-leisure problem exists, one could equate the marginal utility of time to the marginal utility of income multiplied by marginal income. In this case, formulation of the random utility model and thus estimation of the marginal utility of income requires that one estimate the consumer's marginal income. Marginal income, and indeed average income, are often unobserved and so one must estimate the value on the basis of observed variables. For example, Kaoru *et al*. (1995) used a hedonic wage model to predict wages for employed anglers and assigned the minimum wage to unemployed anglers.

Three issues arise here: (1) what is an appropriate and feasible model for the unobserved covariate, (2) should the model for the unobserved covariate be estimated as part of the random utility model, and (3) should uncertainty about the predicted values of the unobserved covariate be accounted for in the random utility model. These questions apply generally to problems involving measurement error, errors in variables, and models for missing observations on covariates. To see this, consider a Bayesian probability model for the simple linear regression with no intercept and unit variance:

$$p(Y_1,\ldots,Y_n \mid x_1,\ldots,x_n,\beta) \sim N\left(\begin{bmatrix}\beta x_1\\ \vdots \\ \beta x_n\end{bmatrix}, I\right).$$

Given prior probability for β and a sample, $(y_1,\ldots,y_n)'$, the problem can be expressed as that of finding the posterior distribution of the unobservable coefficient β, which is given by Bayes' rule as: $p(\beta|y,x) = p(y|\beta,x)p(\beta)/p(y)$. If $x=(x_1,\ldots,x_n)'$ is unobserved, then the posterior cannot be computed. However, suppose there exists another variable, $z=(z_1,\ldots,z_n)'$ that is observable, and that one has a probability model for x conditional on z, say $p(x|z)$. Then, one can compute the posterior probability for β conditional on the observed data. First note that Bayes' rule yields the joint posterior probability of all unobservable values as $p(\beta,x|y,z) = p(y|\beta,x)p(x|z)p(\beta)/p(y|z)$. The marginal posterior distribution for β can be found by integrating x out of the joint posterior. If only some elements of the vector x are unobserved, then z could include those values that are observed and perhaps other explanatory variables.

The approach taken in the previous paragraph is simply to condition all unobservable values on observable information, and then compute the marginal posterior distribution. Spiegelhalter *et al*. (1996, p. 49) demonstrate a possible deficiency with this approach in the context of a measurement error problem (also see MacMahon *et al*., 1990). Specifically, when the unobserved covariate is modeled jointly with the parameters of the linear regression, the predicted values of the unobserved covariate are pulled towards values that improve the fit of the regression equation. There is nothing inherently wrong with this effect. The posterior follows from Bayes' rule, and so it is a coherent result. However, the posterior probability of x is conditional on z and y, so a deficiency in the specification of the regression model will affect predictions of x and β. This result is related to Train and McFadden's (2000) criticism of Morey and Waldman's (2000) likelihood based approach to estimating unobservable expected catch-rates as part of a random utility model. They argue that Morey and Waldman's estimator is inconsistent when important site attributes are omitted from the model, while the standard approach of using the mean catch-rate remains consistent in this case.

The subjective Bayesian point of view is concerned with finite sample inference; and, the goal of Bayesian analysis is the posterior distribution of unobservable quantities of interest, which might include complex functions of the model parameters and explanatory variables. Furthermore, subjective Bayesian inference is based on the idea that the Bayes rule is the logical and coherent way to combine prior probability and observed data to form posterior inferences (Pelloni, 1996). Thus, the relevant question for a subjective Bayesian is: Which model best represents my prior beliefs, and do alternative assumptions affect substantive scientific inferences?

Here, Bayesian statistical methods and Markov chain Monte Carlo (MCMC) simulation are used to estimate alternative random utility models. The models differ in terms of the form of the model used to predict anglers' marginal income and in the treatment of measurement error in the random utility model. Two models for marginal income are considered. (i) In one model, marginal income is assumed to equal the angler's wage-rate, which is assumed proportional to angler income. (ii) In a more novel approach, a labor supply curve is specified for the angler and total income is equated to the wage multiplied by the quantity of labor supplied. This allows the wage-rate to be predicted given the angler's total income. When labor supply is concave in the wage-rate, the predicted wage will not be strictly proportional to income. Total income is unobserved in the data used for this study. Thus, in addition to the wage model, a parametric form is assumed for the distribution of income and the parameters of this distribution are estimated conditional on observed household income category (an ordered categorical variable). In a Bayesian model, this problem is handled in a straightforward fashion by supplementing the original model with additional conditional probabilities – just like the simple linear regression model above.

Hence, the random utility model is estimated conditional on predicted wages, which are random, and wages are predicted conditional on predicted angler incomes, which are also random variables. In other words, there is error associated with the covariates in the random utility model. Should the prediction of angler income and the prediction of angler wage be performed jointly with the estimation of the random utility model, or should point estimates be substituted for the unobserved covariates? As discussed above, there is no clear-cut answer to this question. However, Bayesian computational inference provides a convenient method for assessing how one's choice of model affects substantive scientific inferences. Here, compensating variation is the quantity of scientific interest and it is a function of the unobservable values in the model. As discussed in Jackman (2000), one advantage of Bayesian MCMC is that the posterior distribution of complex functions of model parameters can be recovered at little additional cost. As a result, MCMC simulation can be used to simulate the posterior distribution of compensating variation; one need not be satisfied with a point estimate for its expectation. Thus, one can evaluate how alternative models for angler wage and angler income affect the distribution of compensating variation. Furthermore, the posterior distribution of compensating variation accounts for all uncertainty that is included in the random utility model – uncertainty associated with unknown parameter values and uncertainty associated with unobserved covariates. In the model below, compensating variation is computed for the imposition of a flat fee for one of the choices. In this case, the absolute value of compensating variation is the minimum increase in wealth that would leave the angler as well-off as he was prior to the imposition of the fee.

2. Model Specification

2.1 Data

The random utility model described here is intended to model the choice of fishing location (county) made by marine anglers who fished by boat in coastal North Carolina during May and June, 1997. Data are from the economic supplement to the 1997 MRFSS, and only respondents to the follow-up telephone survey are considered. Only single-day trips are considered, so the travel distance is the roundtrip distance from the angler's home county to the county where they fished; the distance between the two county seats was used as a proxy for actual travel distance and was estimated using a AAA Road Atlas (American Automobile Association, 1993). There are $J=5$ sites and $I=131$ anglers with homes in one of 5 counties.

Site quality is related to the quality of fishing. In subsequent work, a model of each angler's expected catch will be incorporated in the model. For the present, the log of the number of intercept locations is used as a quality variable. Ben-Akiva and Lerman (1985) suggest that this measure accounts for the relative size of the county though it could also be influenced by the historical fishing quality in the county.

The sampling was stratified by state, fishing mode, and two-month wave. Samples were allocated within strata "…in proportion to average estimates of fishing pressure over the previous three survey years" (National Marine Fisheries Service, 2002, p.4). In other words, the data are not a random sample and the random utility model is estimated conditional on anglers' inclusion in the sample. For the purpose of illustration, suppose that there are only two sites described by angler-specific covariates, x_1 and x_2, and that the angler chooses site 1 if the utility $U = \theta(x_1 - x_2) + (e_1 - e_2)$ is positive, where e_1 and e_2 are independently and identically distributed random variables and let the cdf of $(e_1 - e_2)$ be F, assumed symmetric (e.g., normal cdf). Hence, $P(\text{choose site 1}) = P(e_1 - e_2 > -\theta(x_1 - x_2)) = 1-F(-\theta(x_1 - x_2))$. Now, if $x_1 = x_2$ for this particular angler, then $P(\text{choose site 1}) = 0.50$. But, this probability should be expressed conditional on the angler's inclusion in the sample. Specifically, if q_1 is the fraction of total samples collected at site 1, we should have $P(\text{choose site 1}) = q_1$ for the case described here. Another way to think of this is to suppose that you do not observe any explanatory variables and you want to estimate the probability that an arbitrarily chosen angler from the sample chooses site 1. If 2/3 of the samples were collected at site 1, then the probability that the arbitrarily chosen angler visited site 1 is 2/3. So, we might supplement the utility function with a site-specific constant that accounts for the sampling design: say, $U = \delta + \theta(x_1 - x_2) + (e_1 - e_2)$, such that $1-F(-\delta)= q_1$. Solving for δ yields: $\delta = F^{-1}(q_1)$, which implies that δ is increasing in q_1.

The problem is more complicated in the multinomial case. Furthermore, we cannot be certain that the telephone survey respondents are a random sample from the population of MRFSS respondents. Therefore, in order to capture the effect of the sampling design, the fraction of MRFSS samples collected from the site will be included in the random utility model as an explanatory variable.

The economic survey includes data on hourly wages and a categorical response for household income. Very few (21%) of the anglers provide information on hourly wages, but a somewhat higher response rate (58%) is observed for household income category. In addition, the employment status of the angler is recorded. Because a continuous income variable is desired, the observed categorical responses were fit to a log-normal distribution function. Let Z_i be the income category for observation $i = 1, \ldots, n$, where Z_i is a categorical variable, and let $L(Z_i)$ be the lower limit of the observed category and $U(Z_i)$ be the corresponding upper limit. If the underlying continuous latent variable has a cdf F with parameter θ, then the probability of observing $Z_i = z_i$ is $p(z_i|\theta) = F(U(Z_i)) - F(L(Z_i))$. This model was estimated using diffuse proper priors on the mean and variance with F specified as the normal cdf for log income. The posterior mean of the mean of log income was found to be 10.96 and the posterior mean of the variance of log income was found to be $(2.62)^{-1}$. This corresponds to an average household income of about $70,000 per year. Further results are deferred to section 3.3. A few observations of personal income were available. A simple regression supported the hypothesis that the individual angler's income is approximately half of household income.

2.2 Random Utility Model

The random utility model is based on the idea that there exists a latent value (utility) such that an angler chooses to visit the site with the highest expected utility. All anglers in the sample are assumed to have the same choice set and the specification of the utility function presumes that the angler's marginal income is equal to his opportunity cost of time. The form of the utility function for the i^{th} angler and the j^{th} site is assumed to be:

(1) $V_{i,j} = \alpha_1 \log r_j + \alpha_2 (n_j / n) + \beta(c + s^{-1} w_i) x_{i,j} + \varepsilon_{i,j}$,

where r_j is the number of intercept sites in county *j*, (n_j/n) is the fraction of MRFSS observations in county *j*, and $x_{i,j}$ is the round-trip travel distance to county *j* from angler *i*'s home county. Here, *c* is the direct travel cost in \$/mile, *s* is the speed of travel in miles/hour, and the travel distance is measured in miles. The opportunity cost of time is w_i and its form is discussed below. Unknown parameters in the model include α_1, α_2, *and* β; these are assigned proper but relatively diffuse prior probability centered about zero, though α_2 is restricted to be non-negative. The restriction is imposed because, *ceteris paribus*, the probability that a sampled angler chooses site *j* is increasing in the proportion of samples collected at site *j* – this is a result of the sampling design, as discussed in section 2.1. $\varepsilon_{i,j}$ is a constant that is known to the angler, but random from the researcher's perspective. Here, $\varepsilon_i = (\varepsilon_{i,1},\ldots, \varepsilon_{i,5})'$ is assumed to have a normal distribution with zero mean and a covariance matrix proportional to the identity matrix. This assumption about the covariance matrix is equivalent to the imposition of independence of irrelevant alternatives (IIA), which is appropriate when all choices are close substitutes. In later work, a heteroscedastic model will be considered that allows for variable substitution effects (Allenby and Ginter, 1995).

Two models for employed anglers' marginal income, *w*, are considered: (1) marginal income is assumed to equal the wage-rate, which is assumed proportional to annual income, and (2) an upward sloping labor supply exists such that the marginal income is less than proportional to annual income. The second model accounts for the individual's response to increased wages; as the wage increases, the opportunity cost of leisure increases and so the individual has an incentive to increase his supply of labor. Hence, higher incomes are the result of a higher wage and a higher number of work hours. Specifically, *log(income) = log(wage) + log(labor)*. Labor supply could become backward-bending at higher wage levels due to the income effect, but a model that includes that possibility will be reserved for future work. The first model is based on the assumption that *log(labor)* is constant – i.e., the labor supply curve is vertical and the wage is proportional to income. The second model is based on a labor supply model that allows for a positive elasticity of labor supply. The elasticity of labor supply is restricted to be less than 1.0 here so that the labor supply curve is concave in the wage-rate.

Let *m* be the angler's annual income, which is predicted conditional on observed categorical values for household income – see section 2.1. Then, the first model suggest that $E(log[w]|log[m]) = a_0 + log[m]$. For the second model, the angler's supply of labor, L, is assumed to be described by $E(log[L]|log[w]) = b_0 + b_1 log[w]$, where $0 < b_1 < 1.0$. Then, the angler's expected marginal income is $E(log[w]|log[m]) = c_0 + c_1 log[m]$, where $0.5 < c_1 < 1.0$. Unemployed anglers are assigned the minimum wage in both models.

Some additional assumptions are required in order to estimate the random utility model. First, missing household incomes are assumed to be missing at random. In other words, if angler *i* does not report income category then $log[m_i] \sim N(10.96, 2.62^{-1})$ is assumed. This is a strong assumption and alternatives must be considered at a future time. In addition, it is assumed that the relationship between observed hourly wages and income holds for anglers that do not report an hourly wage. Finally, independent and identical normal errors are assumed for the wage models.

3. Results

3.1 Overview

Three approaches to estimation of the random utility model with unobserved covariates are considered here. The three methods can be distinguished from one another by considering the general structure of the posterior distributions. Let *θ* denote the coefficients and *X* denote the observable covariates and fixed parameters in the utility function (1); let *γ* denote the parameters of the wage model; let *δ* denote the parameters of the income model; and let a hat (^) denote the posterior expectation of the value obtained prior to estimation of the random utility model. Then, the posterior distributions obtained fro the random utility model will have the form:

(2a) $Method\,1: \quad p(\theta, w, m \mid y, z, X, \hat{\gamma}, \hat{\delta})$;

(2b) $Method\,2: \quad p(\theta, w \mid y, \hat{m}, z, X, \hat{\gamma}, \hat{\delta})$;

(2c) $Method\,3: \quad p(\theta \mid y, \hat{w}, \hat{m}, z, X, \hat{\gamma}, \hat{\delta})$.

In all three methods, household income is predicted conditional on the posterior means of the parameters of the assumed distributional form (see section 2) and the observed income category. The parameters of the wage equation are estimated conditional on the sample of observed wages (n=28), income category, and the expected value of the parameters of the income distribution. Hence, some uncertainty is selectively ignored; this was done to ease the computational burden and because the income distribution and the wage model are thought to be rough approximations. In all three methods, if the wage rate is observed, it is its own prediction because $p(w_i|w_i)=1$.

The difference in the three methods is in the treatment of the predicted value of the wage in the random utility model. In the first method, the measurement error inherent in the observation of income is considered when forming the predictive distribution for the wage, and the uncertainty about the predicted wage is included in the random utility model. As a result, the posterior distribution for the random utility model is a joint distribution over the utility function coefficients, wage, and income. The predicted values for wage and income will be pulled towards values that improve the fit of the random utility model. The value of each angler's income is fixed at the mean of its income category in the second method; hence the variance of income does not enter into the predictive distribution for the wage. The predicted wage is still a random variable, however, and wage appears jointly in the posterior distribution of the random utility model. The third method omits all covariate error from the random utility model; utility function coefficients are estimated conditional on point estimates of the wage, which is predicted at the angler's mean income.

3.2 MCMC Simulation

MCMC is method for simulating draws from posterior distributions that have an inconvenient or intractable analytic form (Gelfand and Smith, 1990). An overview of Bayesian MCMC can be found in Gelman *et al*. (1995), Jackman (2000), Cappe and Robert (2000), and Gelfand (2000). The result of MCMC simulation is a sample from an asymptotic approximation of the posterior distribution. A model is defined by a sampling distribution for an observable variable, Y conditional on unobserved values, θ, and by a prior probability distribution for θ. The posterior distribution is found using Bayes' rule: $p(\theta \mid y) \propto p(y \mid \theta) p(\theta)$, which must be proper. MCMC simulation generates a sample $\{\theta^1, \ldots, \theta^L\}$ that is approximately a sample from the posterior distribution. Descriptive statistics (e.g., sample quintiles) for the sample can be used to describe the marginal posterior distribution of each element of θ. To ensure that the sample is a sample from the posterior and not a function of given starting values, it is common practice to generate a number of warm-ups before collecting values for inference. Often, sample values exhibit significant autocorrelation. If this happens, the sample moments and quintiles are not estimated with the precision implied by the precision calculated under the assumption of i.i.d draws. The problem of ensuring that a sufficient number of warm-ups have been generated and the problem of ensuring that a sufficient number of draws are available to precisely estimate sample quintiles is usually grouped under the heading of convergence diagnosis and several statistical tests are available for that purpose (Gelman *et al*., 1995 and Smith, 2002 provide references).

An advantage of MCMC simulation is that functions of θ, $g(\theta)$, can also be simulated. For each iteration of the sampler, g can be computed so that one obtains a sample, $\{g(\theta^1), \ldots, g(\theta^L)\}$, which describes the posterior distribution of *g*. Here, compensating variation is of interest; it can be computed for each iteration of the MCMC sampler and its posterior distribution described using descriptive statistics.

3.3 Predictive Model for Income and Wages

The predictive model for income was described in section 2.1. The log of income was assumed to follow a normal distribution, and the parameters of this distribution were estimated conditional on observed income categories. To check the fit of this model, the sample of categorical variables for income was replicated at each iteration of the MCMC simulation, and the proportion of the sample in each category was recorded – i.e., $p^l_k = \Pr(Z^l_{rep,i} = k)$ for each observation, *i*, each income category, *k*, and each iteration, *l*. In table 1, the posterior distribution of the predicted frequency of incomes that fall in each income category is described and compared to the empirical frequency. Note that all empirical frequencies fall within the 95% posterior interval for the predicted frequencies. This does not imply that the normal distribution is "correct" for log income, but it does indicate that the distribution of log income is well-approximated by this model.

Table 2 provides the marginal posterior distribution for the coefficients of the two wage models. Each of these models was estimated conditional on log income, which was treated as a random variable. Specifically, let w be the sample of observed wages, let m be the corresponding vector of incomes, let z be the observed income category, let γ be the unobserved parameters of the wage regression, and let θ^* be the posterior expectation of the parameters of the income distribution. Then the sampling distribution

for the wage models can be written as $p(w,m|z, \theta^*,\gamma) = p(w|m,\gamma)p(m|z, \theta^*)$. Hence, individual values of log income can be pulled towards favorable values, but the parameters of the income distribution are fixed *a priori*. The estimated labor elasticity in the second wage model can be recovered from γ_1 as
$elasticity = (1/\gamma_1)-1$. Here, the elasticity derived from the posterior expectation of γ_1 is approximately 0.667.

3.4 Random Utility Model

The random utility model was estimated six times. The three methods for accounting for uncertainty, (2a) – (2c), were used for each wage model. The variance was restricted to 1.0 for identification. The posterior estimates for the parameters of the random utility model, equation (1), are described in table 3. The prior probability for the parameters was specified using proper but diffuse distributions. Prior variances were increased by a factor of 10 with little effect on posterior values, so the priors are relatively diffuse relative to sample information. Figure 1 illustrates the prior distribution and the posterior distribution for the marginal utility of income for the random utility model. In addition to the posterior distribution, simulation was used to replicate utility function values and replicate site choices were derived. All models replicated the observed site choices with about 40% accuracy, on average. In terms of this measure, there was little difference in predictive performance between models.

The quantity of interest is compensating variation. Table 4 provides a comparison of alternative estimates of the average compensating variation assuming that a \$10 access fee is imposed for site number 2 (Dare County) – this is a hypothetical example constructed for the purpose of illustration. Here, for each iteration of the MCMC simulation, compensating variation is computed for each angler and is then averaged over anglers to obtain a measure of the average loss per angler – i.e., the average amount that would have to be paid to each angler to leave him as well off as before the imposition of the $10 fee for Dare county. The relatively small magnitude of compensating variation suggests that close substitutes are available for little additional cost.

4. Discussion

The purpose of this study is to evaluate how alternative wage models and alternative treatments of uncertainty about the value of unobserved covariates affect the posterior estimates of utility function parameters and compensating variation and that is the topic of the following discussion. More detailed and rigorous assessments of model adequacy and formal statistical model comparisons are reserved for future work.

The parameter of primary interest is β, the negative marginal utility of income. Here, the form of the wage model has a modest affect on the posterior distribution of β. The magnitude of the posterior median and the width of the 95% posterior interval are slightly smaller when the upward sloping labor supply function (model 2 in table 3) is used to predict the angler's wage-rate. When wages are predicted within the random utility model, the posterior estimate for β is slightly larger in magnitude than when uncertainty about the predicted wage is ignored; this is true for both model 1 and model 2. Hence, there is some evidence of attenuation of the parameter estimate. However, this effect is not very large for this particular situation, and the economic significance of the result is marginal.

The compensating variation per angler is also affected by both the choice of wage model and the treatment of prediction error. As with the parameter estimate, the magnitude of the effect is small. In table 4, one can see that the median loss per angler is about \$0.02 higher in model 2, though the 95th percentile is between \$0.40 and \$0.50 higher in model 2. Thus, the choice of wage model has a more significant effect on the tails of the posterior distribution of compensating variation. The higher magnitude of compensating variation for model 2 is most likely a direct result of the smaller magnitude of β that is obtained when the upward sloping labor supply is used to construct the prediction for wage because the inverse of β appears in the formula for compensating variation.

Given the small magnitude of the differences between alternative treatments of prediction error in the random utility model, it is apparent that inclusion of income and wage prediction as part of the estimation of the utility function parameters results in very little adjustment of predicted wages towards values that improve model fit. Thus, the concern expressed by Spiegelhalter *et al*. (1996, p. 49) and discussed above in the introduction does not appear to be a significant issue in this particular case. Predicted incomes are constrained by the observed income category and this fact might explain why the treatment of error in predicted income has little effect on posterior results. If the parameters of the prediction equation for wage were estimated along with the parameters of the utility function, greater adjustment of the wage model towards improved fit of the utility function would likely be observed.

5. Conclusions

Difficulties arising due to unobserved covariates can be handled in a straight-forward fashion in a Bayesian model, provided a suitable probability model that relates the unobserved value to observable explanatory variables can be constructed. Bayesian MCMC simulation provides a means of estimating complex models and simulating the posterior distribution of quantities of interest (e.g., compensating variation). This has the advantage that the effect of model assumptions on relevant scientific inferences can be evaluated.

In the discrete choice model estimated here, the choice of wage model and treatment of prediction error had little effect on posterior inferences. Morey and Waldman (2000) considered the estimation of catch-rates within the random utility framework, and Train and McFadden (2000) were skeptical about their results in the face of specification error. Here, the treatment of prediction error had little effect on finite sample inferences for compensating variation, and so the issue has little substantive economic significance. Future work is required to determine if the joint estimation of wage equation parameters and utility function parameters results in substantively different inferences.

The model developed here also accounted for the sampling design of the MRFSS. This is an issue that has not received much attention in the literature. To the extent that predicted fishing pressure is correlated with historic catch-rates, the inclusion of the latter accounts for the sample design. However, in this case, it is not clear that one can differentiate between the effect of catch-rate on choice probabilities and the effect of the sampling design.

6. References

Allenby, G. M. and J. L. Ginter (1995). The Effect of In-Store Displays and Feature Advertising on Consideration Sets. *International Journal of Research in Marketing*, 12, 67 – 80.

American Automobile Association (1993). *Road Atlas*.

Ben-Akiva, M. and S. R. Lerman (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press.

Cappe, O. and C. P. Robert (2000). Markov Chain Monte Carlo: 10 Years and Still Running. *Journal of the American Statistical Association*, 95, 1282-1286.

Gelfand, A. E. and A. F. M. Smith (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, 85, 398-409.

Gelfand, A. E. (2000). Gibbs Sampling. *Journal of the American Statistical Association*, 95, 1300-1304.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (1995). *Bayesian Data Analysis*. Chapman and Hall, New York.

Jackman, S. (2000). Estimation and Inference are Missing Data Problems: Unifying Social Science Statistics via Bayesian Simulation. *Political Analysis*, 8:4, 307 – 332.

Kaoru, Y., V. K. Smith, and J. L. Liu (1995). Using Random Utility Models to Value the Recreational Value of Estuarine Resources. *American Journal of Agricultural Economics*, 77, 141-151.

MacMahon, S., R. Peto, J. Cutler, R. Collins, P. Sorlie, J. Neaton, R. Abbott, J. Godwin, A. Dyer, and J. Stamler (1990). Blood Pressure, Stroke, and Coronary Heart Disease. Part 1, prolonged deficiencies in blood pressure: prospective observational studies corrected for the regression dilution bias. *The Lancet*, 335, 765-774.

Morey, E. R. and D. M. Walderman (1998). Measurement Error in Recreationa Demand Models: The Joint Estimation of Participation, Site Choice, and Site Characterisitcs, *Journal of Environmental Economics and Management*, 35, 262-276.

National Marine Fisheries Service (2002). *MRFSS Data Users Manual*. On-line at http://www.noaa.gov.

Parsons, G. (2002). *A Bibliography of Revealed Preference Random Utility Models in Recreation Demand*. On-line at http://www.ocean.udel.edu/level1/facultystaff/faculty/gparsons.

Pelloni, G. (1996). *De Finetti, Friedman, and the Methodology of Positive Economics*. Journal of Econometrics, 75, 33-50.

Sandefur, R. A., F. R. Johnson, and R. B. Fowler (1996). An Introduction to the Random Utility Model. TER General Working Paper No. G-9602, Triangle Economic Research, Durham, NC.

Smith, B. J. (2001*). Bayesian Output Analysis (BOA) Program Version 0.5.0 User Manual*. Department of Biostatistics, The University of Iowa College of Public Health.

Spiegelhalter, D., A. Thomas, N. Best, and W. Gilks (1996*). BUGS 0.5: Bayesian Inference Using Gibbs Sampling Manual (version ii)*. MRC Biostatistics Unit, Institute of public Health, Cambridge, U.K.

Spiegelhalter, D., A. Thomas, N. Best, and W. Gilks (1996*). WinBUGS 1.3:Software and Documentation*. MRC Biostatistics Unit, Institute of public Health, Cambridge, U.K.

Train, K. and D. McFadden (2000). Comment. Discussion of Morey and Walderman's "Measurement Error in Recreation Demand Models". *Journal of Environmental Economics and Management*, 40, 76-81.

Table 1. Empirical and Replicated Frequencies for Income Categories

		Posterior Interval		
Income Category	Obs. Freq.	5th %ile	Median	95th %ile
< $15,000	0.05	0.00	0.01	0.07
$15,001-$30,000	0.05	0.05	0.14	0.25
$30,001-$45,000	0.19	0.11	0.19	0.30
$45,001-$60,000	0.18	0.08	0.16	0.26
$60,001-$85,000	0.23	0.11	0.19	0.30
$85,001-$110,000	0.19	0.04	0.11	0.19
$110,001-$135,000	0.05	0.01	0.05	0.12
> $135,000	0.04	0.03	0.10	0.19

Table 2. Posterior Estimates for Wage Regressions

Wage Model 1:

parameter	5th %ile	Median	95th %ile
	95% Posterior Interval*		
γ_0	-8.58	-8.36	-8.14
γ_1	na	na	na
$1/\sigma^2$	2.02	3.91	6.68

Wage Model 2:

parameter	5th %ile	Median	95th %ile
	95% Posterior Interval*		
γ_0	-5.70	-3.54	-2.15
γ_1	0.43	0.56	0.76
$1/\sigma^2$	3.08	6.10	10.21

*Estimates based on 10,000 iterations after 10,000 warm-ups.

Table 3. Posterior Estimates of Utility Function Parameters

	Prior[**]	Posterior Interval[*] 5th %ile	Median	95th %ile
Model 1,Method 1:				
α1	$N(0,10^3)$	0.44	0.75	1.01
α2	$N(0,10^3)I(0,)$	0.01	0.17	0.47
β	$N(0,10^3)$	-1.59	-1.10	-0.65
Model 1,Method 2:				
α1	$N(0,10^3)$	0.42	0.75	1.01
α2	$N(0,10^3)I(0,)$	0.01	0.18	0.51
β	$N(0,10^3)$	-1.64	-1.15	-0.67
Model 1,Method 3:				
α1	$N(0,10^3)$	0.41	0.73	0.99
α2	$N(0,10^3)I(0,)$	0.01	0.19	0.52
β	$N(0,10^3)$	-1.74	-1.21	-0.72
Model 2,Method 1:				
α1	$N(0,10^3)$	0.42	0.72	0.98
α2	$N(0,10^3)I(0,)$	0.01	0.19	0.49
β	$N(0,10^3)$	-1.32	-0.91	-0.53
Model 2,Method 2:				
α1	$N(0,10^3)$	0.43	0.74	1.00
α2	$N(0,10^3)I(0,)$	0.01	0.17	0.47
β	$N(0,10^3)$	-1.34	-0.91	-0.53
Model 2,Method 3:				
α_1	$N(0,10^3)$	0.41	0.72	0.98
α_2	$N(0,10^3)I(0,)$	0.01	0.20	0.51
β	$N(0,10^3)$	-1.42	-0.98	-0.59

[*]Estimates based on 10,000 iterations after 10,000 warm-ups
[**]Increasing the prior variance by a factor of 10 had no substantive effect on posterior values. Hence the prior is deemed to be dominated by the sample information.

Table 4. Absolute Compensating Variation per Angler

	5th %ile	Median	95th %ile
Wage Model 1:			
method 1	0.00	0.11	2.46
method 2	0.00	0.12	2.38
method 3	0.00	0.12	2.25
Wage Model 2:			
method 1	0.00	0.14	2.99
method 2	0.00	0.14	2.99
method 3	0.00	0.14	2.75