How Much Does the Far Future Matter? A Hierarchical Bayesian Analysis of the Public's Willingness to Mitigate Ecological Impacts of Climate Change

David F. Layton* University of Washington

Richard A. Levine† University of California, Davis

June 4, 2002

Abstract

How much does the far future matter? This question lies at the heart of many important environmental policy issues such as global climate change, biodiversity loss, and the disposal of radioactive waste. While philosophers, experts, and others offer their viewpoints on this deep question, the solution to many environmental problems lies in the willingness of the public to bear significant costs now in order to make the far future a better place. Short of national plebiscites, the only way to assess the public's willingness to mitigate impacts in the far future is to ask them. Using a unique set of survey data in which respondents were provided with sets of scenarios describing different amounts of forest loss due to climate change, along with associated mitigation methods and costs, we can infer their willingness to bear additional costs to mitigate future ecological impacts of climate change. The survey also varied the timing of the impacts which allows us to assess how the willingness to mitigate depends upon the timing of the impacts. The responses to the survey questions are a consequence of latent utilities with complex ordinal structures which result in non-rectangular probabilities. While the non-rectangular probabilities complicate standard maximum likelihood based approaches, we show how the non-rectangular probabilities fit neatly into a hierarchical Bayesian model. We show how to fit these models using the Gibbs sampler, overcoming problems in parameter identification to improve mixing of the induced Markov chain. The results indicate that the public's willingness to incur additional costs to mitigate ecological impacts of climate change is an increasing non-linear function of the magnitude of the impact, and that they discount future impacts at about 1% per year.

Keywords: non-market and environmental valuation, discount rate, stated preference, willingness to pay, forest loss, Gibbs sampler, Markov chain Monte Carlo, non-rectangular probabilities, identifiable parameters

*Daniel J. Evans School of Public Affairs; email:dflayton@u.washington.edu

†Department of Statistics; email: ralevine@ucdavis.edu

This work is supported by NSF grants DMS-9978321 ES and SCREMS DMS-0079430

1 Introduction

The debate over what actions should be undertaken in order to mitigate impacts of global climate change in the far future is complicated both by scientific uncertainty regarding the underlying mechanisms of climate change and legitimate debate over how society *should* weigh the costs of mitigation today versus the costs of impacts which will not occur for many decades or centuries. While many experts have strong beliefs about how society should weigh the costs of mitigation versus the costs of impacts imposed on future generations, there seems to be little understanding of how the public *actually* weighs them. If we acknowledge that in democratic societies it is the public that is alive today that will ultimately determine how much mitigation actually takes place, understanding the public's preferences regarding climate change mitigation becomes crucial. This paper analyzes two critical aspects of the public's willingness to mitigate climate change impacts. First we examine the costs the public perceives due to ecosystem shifts that could result from climate change, and then we examine how the timing of the impacts affects the public's willingness to mitigate.

The issue of ecosystem shifts is important because much of the debate over climate change mitigation takes place in the context of optimal investment. Economists point out that resources spent on mitigation reduce investment in other economically productive areas. Even if society considers only actions that would improve the well-being of future generations, the question becomes whether future generations would be better off with less climate change and lower economic growth, or by not mitigating as much and more economic growth? These economically based questions raise important issues, but do little to address the losses due to ecosystem shifts. Technological innovation may alleviate certain impacts upon humans due to climate change, but it cannot replace ecosystems. If we value ecosystems in their own right, perhaps simply because they exist, then the only way to protect them is to mitigate the impacts of climate change. Without knowing how much we value ecosystems, picking the right amount of mitigation is impossible.

The timing of climate change impacts is of enormous significance in any analysis of the benefits and costs of mitigation measures. The model used by nearly all economists to compare different streams of costs (or benefits) realized through time is based upon exponential discounting. In this process, costs (or benefits) in the future are compared to today's costs (or benefits) by first re-weighting them by an exponential factor, modeled as $\exp(-\rho t)$ where ρ is called the discount rate, and t indicates how many years into the future the cost (or benefit) will be realized. The discount rate, ρ, is quite possibly the single most important parameter in comparing the costs and benefits of mitigation measures that will occur over long periods of time. When t is large, say on the order of 75 or 100 years, small changes in this parameter can have an enormous impact on a final assessment of whether a certain mitigation action is, on net, beneficial or not. Remembering that the underlying science regarding climate change is uncertain enough so that it is easy to imagine updating forecasts so that impacts might move 50 years nearer or farther into the future, one can see that getting the discount rate right is critically important.

The economics profession is in almost complete agreement about discounting the future. In a survey of economists by Weitzman (2001), more than 98% said that the future costs and benefits of climate change should be discounted at some positive rate. But the agreement ends there. While the economics profession is nearly unanimous that the future should be discounted at a positive rate, the question of *what* rate causes considerable difference in expert opinion. Weitzman (2001, p. 260) states that "the most critical single problem with discounting future benefits and costs is that no consensus exists today, or, for that matter, has ever existed, about what rate of interest to use." (For exhaustive treatments of the issues see Arrow et al., 1996, Nordhaus, 1994, or Portney and Weyant, 1999.) All previous analyses of climate change have been based on the range of expert opinions described above, or opinions of other non-economist experts. In this paper we provide, to our knowledge, the first estimate of the public's discount rate for the far future impacts of climate change.

There is at least one good reason why relatively little is known about the public's willingness to bear the costs of climate change mitigation – there is simply no way to observe people making trade-offs that would indicate their preferences. Unlike housing markets where we can observe people trading-off living rooms for bedrooms through their choice of houses, or in financial markets where we can observe interest rates for periods of generally up to 30 years in length, there is no market place for climate change mitigation. When one considers that many impacts will not occur for a hundred years or more, it is clear that there is simply no guidance from past human behavior about the types of trade-offs people will be willing to make. Short of national plebiscites, the only way to assess the public's willingness to mitigate impacts in the far future is to ask them. So we use data collected from an innovative survey that used the Stated Preference (SP) technique to infer the public's willingness to make these trade-offs. The SP approach presents respondents with sets of alternatives composed of underlying attributes and asks them to state their preferences over the set of alternatives. By varying the attributes of the alternatives, much like changing the prices and the number and types of rooms in a set of houses, we can infer the value of each underlying attribute.

To address the two questions above (how does the public value ecosystem changes, and how does the timing of ecosystem impacts affect their values) we use a unique survey of residents in the Denver, Colorado area. The survey presented respondents with various amounts of forest loss along the Front Range of the Rocky Mountains due to climate change. By varying the amounts of forest loss, costs of mitigation, type of mitigation measures used, and the timing of the impacts, we can derive the Willingness to Pay (WTP) to prevent different degrees of forest loss. WTP is the economic measure of willingness to mitigate. This definition acknowledges that most people want or desire to make things better, but action requires more than desires or preferences; it requires a willingness to bear costs. By analyzing how the WTP changes as a function of the timing of the impacts we can derive implicit discount rates. Discount rates are the economic measure of how timing affects values.

Typically in SP surveys, the respondent is asked to choose their most preferred alternative from a set. Then standard discrete choice models can be applied. The data set we analyze here collected additional data in the form of each respondent's least preferred alternative. This additional information is important as it will allow us to draw much more reliable inferences regarding the public's preferences, and the implicit discount rate, but unfortunately complicates the analysis as it results in non-rectangular probabilities. For this, and other reasons detailed below, we analyze the data via a hierarchical Bayesian modeling approach fit by Markov chain Monte Carlo.

The next section describes the survey data and economic model in detail. Section 3 lays out the latent variable model and the implementation of the Gibbs sampler. The results are provided in Section 4, with discussion and interpretation presented in Section 5.

2 Survey data and economic model

As the climate warms, significant shifts in temperature isotherms are predicted to occur. These would be equivalent to poleward geographic shifts on the order of hundreds of kilometers, or altitude shifts on the order of hundreds of meters. This is predicted to cause noticeable shifts in biomes across mid-latitude regions of the world including the United States [VEMAP (1995), IPCC (1998)]. The survey describes one possible ecosystem shift: forests along the Colorado Front Range of the Rocky mountains. Along the Colorado Front Range there is a sharp boundary between the grasslands of the plains to the east of the mountains and forests that begin in the foothills. If global climate changes, the forests of the Front Range may retreat up the mountains and be replaced by encroaching grasslands. We will describe the retreat of forests as "forest loss". Forest loss along the Front Range of the Rocky Mountains was chosen because the existing abrupt change between the grassland and forests systems makes it much easier for people to visualize slowly shifting ecosystems due to climate change.

In Section 2.1, we describe the stated preference (SP) survey data collected. In Section 2.2, we discuss difficulties in analyzing the survey data with discrete choice models, standard in the economics literature.

2.1 Stated Preference survey

The SP survey was in booklet form and first provided relevant background information on climate change and the greenhouse effect; the latest predictions for global temperature rise and sea level rise; that regional impacts are much harder to predict than global averages; that one impact of global climate change in Colorado may be the retreat of forests along the Front Range of the Rocky mountains. Given the background information, the SP scenario is developed which is the context in which the respondents will actually make trade-offs that reveal their preferences. The scenario is that one outcome of climate change may be that forests retreat from their current elevation boundary of about 5400 feet (essentially at the bottom of the foothills), and are replaced by grasslands. Four levels of forest impacts due to climate change were utilized. If forest loss is prevented, then forests will still begin at about 5400 feet with a 0 foot loss in elevation. If climate change moves forests to 6000 feet of elevation there is a 600 foot loss of forest, at 6660 feet there is a 1200 foot loss, and at 7900 feet there is a 2500 foot loss. Each forest loss level was described in words and by computer-altered photographs linked to actual elevation levels.

Two potential means for slowing or stopping the retreat of forests are described. One attacks the problem locally through forestry practices that help the trees to reproduce themselves and other practices that discourage the encroachment of the prairie as the climate changes. This approach does nothing to remedy climate change and its global impacts, but reduces the local impacts on forests. The other approach is to attack climate change at its root cause by reducing carbon emissions through international agreements. These agreements have associated costs due to higher fuel prices, product prices, and so on. International agreements have enormously greater mitigation potential than forestry practices which just deal with local forests, but of course international agreements will cost much more. This is what generates the trade-offs: mitigating most or all of the *global* impacts of climate change is very costly, while more targeted *local* intervention is much cheaper.

The Stated Preference (SP) approach takes the basic scenario as above, and presents respondents with alternatives which are composed of underlying attributes. Each alternative in the climate change-forest loss scenario consists of a view of what the future might look like described in terms of four basic attributes: the amount of forest loss (none, 600 foot loss, 1200 foot loss, 2500 foot loss), the type of local forest loss mitigation employed (none, forestry practices used), the type of global climate change mitigation employed (none, limited agreements, vigorous agreements), and costs to households today ($ per month for the next 20 years). One last aspect of the scenario is important: how soon or far-off in the future the bulk of the impacts will occur is expected to be critically important, so two levels of timing were utilized consisting of impacts in 60 or 150 years. This results in a total of five attributes.

"Menus" consisting of five or six alternatives were created. See Figure 1 for a sample menu. The respondent was asked to state which alternative in the menu was their most preferred and which was their least preferred after carefully considering all of the aspects of the alternatives. This exercise was conducted four times with four different menus for each respondent. Each menu contained the following combination of abatement and forestry (one combination in each alternative): no action (no forestry or abatement), forestry alone, limited abatement alone, combined forestry and limited abatement, and vigorous abatement alone for a total of five alternatives. (There were also some six alternative menus which were the same as the five programs menus with the addition of a program that utilized vigorous abatement plus forestry.) By varying the costs of the alternatives, the impacts on forests, and the local forestry and global mitigation approaches, we can un-bundle the values people hold for preventing forest loss from all of the other impacts of climate change.

Twenty-eight menus were created which all had the basic structure described above in terms of abatement and forestry, but across menus the costs of the alternatives varied as did the impacts on forest loss. So for a given combination of abatement and forestry, different respondents were given different costs and forest loss. Each of the 28 menus represents a different set of trade-offs. These 28 menus were placed into 7 different survey versions consisting of four menus each (each menu appeared in only 1 survey version). Each respondent received one survey version. The time horizon did not vary within a survey version (each respondent saw only one time horizon). The sample was split, with one half receiving a time horizon of 60 years, and the other half receiving a time horizon of 150 years. The seven survey versions (of 28 menus) were replicated for both time horizons. That is, one respondent would see a menu embodying a particular set of trade-offs with the impacts occurring in 60 years, and another respondent would see the same menu and trade-offs with the exception that the impacts would occur in 150 years. This design allows us to directly compare the results across the two horizons to infer the public's discount rate. Table 1 summarizes the experimental design.

The survey sample consists of 373 randomly chosen Denver, Colorado area residents. The data set consists of 752 observations (188 people times a pair of most and least preferred choices from each of 4 menus) for the 60 year horizon and 740 observations (185 people times 4 pairs of most and least preferred choices each) for the 150 year horizon. This survey was subjected to many levels of peer review and pre-testing on its content, text, graphics, elicitation format, and over-all design. See Layton and Brown (2000) for a more complete discussion of the survey format and pre-testing.

The goal of the data analysis will be to characterize the public's willingness to incur costs over the next 20 years to provide mitigation of impacts that would not occur until 60 or 150 years. Layton and Brown (2000) utilized this data set but analyzed the most preferred choices only (not the least preferred choices), using a random parameters model estimated by maximum likelihood. They found little evidence that the public's WTP was a function of the time horizon. We expect that by adding in the least preferred choices, which essentially doubles the size of the data, we will be able to estimate a more flexible model while still providing enough explanatory power to draw firm collusions regarding the public's WTP for mitigation and how the WTP changes as a function of the timing of the impacts.

2.2 The economic model

Our estimates of WTP will be derived from an economic model of choice based on utility theory. Economic utility theory begins with the assumption that consumers can order a set of bundles or alternatives according to the amount of utility (or desirability) they receive from each alternative. Utility is itself unobservable, but choices can be observed. Importantly, utility theory is ordinal and not cardinal meaning that it is only utility differences and not utility levels that determines choice. The extension of basic utility theory to the statistical analysis of choice data is accomplished via the Random Utility Model (RUM) (see McFadden, 1981). In the discrete choice case of J distinct alternatives, we assume that each individual evaluates their utility and chooses their most preferred alternative. By allowing utility to depend upon error terms, choice becomes random

$$U_{ij} = V_{ij} + \epsilon_{ij} \tag{1}$$

where i indexes person, j indexes alternatives, V_{ij} is the deterministic portion of utility which will be modeled as a function of explanatory variables and parameters to be estimated, ϵ_{ij} is the error term, and U_{ij} is the total utility the person receives from the alternative.

Maximization of random utility indicates that U_{ij} is chosen if $U_{ij} > U_{ik}$ for all k not equal to j. Without loss of generality, placing the chosen alternative U_{ij} first, the probability, P_{ij}, of choice equals

$$P_{ij} = \int_{-\infty}^{\infty} \int_{-\infty}^{U_{ij}} \cdots \int_{-\infty}^{U_{ij}} f(U)\, dU \tag{2}$$

where $f(U)$ is the density function of the J-vector of utilities induced by the errors ϵ_{ij}.

The assumption of Type I Extreme Value errors leads to the commonly used Multinomial or Conditional Logit Model of McFadden (1974). Conditional Logit models are restrictive and so with the advent of high speed computation and simulation techniques, Multinomial Probit (MNP) models estimated within the frequentist approach have become increasingly popular. To implement the frequentist approach in problems with more than a few alternatives, the probability in (2) (or some function related to it) must be simulated. Three simulation based approaches have been suggested: the method of simulated moments of McFadden (1989), simulated maximum likelihood as in Borsch-Supan and Hajivassiliou (1993), and simulation of the score functions by Hajivassiliou and McFadden (1998). The probability in (2) is rectangular, and most research has centered on developing simulators useful for integrals of the form in (2) for their use in one of the simulation based approaches (Hajivassiliou and Ruud, 1994 and Hajivassiliou, McFadden and Ruud, 1996). To illustrate the rectangular nature of the probability in (2), define Z to be the $J-1$ vector of utility differences, differenced with respect to the most preferred alternative, $U_{ik}-U_{ij}$ for all k not equal to j. Next define its density function to be $g(Z)$. Then (2) can be written as

$$P_{ij} = \int_{-\infty}^{0} \cdots \int_{-\infty}^{0} g(Z)\, dZ. \tag{3}$$

Revealed preference data based on purchases or other decisions made by consumers typically result in rectangular probabilities. Survey based data that takes the form of choices obviously results in rectangular probabilities, and so do rankings, but other seemingly minor deviations from choice or ranking result in non-rectangular probabilities. For instance, asking respondents to group alternatives into say three groups of "acceptable", "unacceptable", or "neutral" would not, and neither does asking respondents to provide their most and least preferred responses. For the most and least preferred (or "best" and "worst") data, place the utility of the most preferred alternative first and the least preferred alternative last. The resulting J-dimensional integral yields the probability

$$P_{ibw} = \int_{-\infty}^{\infty} \int_{-\infty}^{U_b} \cdots \int_{-\infty}^{U_b} \int_{-\infty}^{\min\{\text{all } U_l\}} f(U)\, dU \tag{4}$$

where b indexes the best or most preferred, l all of the "middle" alternatives, and w the worst or least preferred. This integral is not rectangular.

Layton and Lee (1998) show that non-rectangular probabilities associated with any type of preference ordering can be decomposed into a sum of rectangular probabilities. In the case of most and least preferred alternatives in (4), this sum increases in a factorial manner as the number of alternatives increases. In many other contexts, such as grouping alternatives into "acceptable", "unacceptable" or "neutral", the actual sets of rectangular probabilities will not only increase factorially as the number of alternatives increases, but the form of the sum will vary substantially across individuals, making a brute force approach computationally expensive. With enough alternatives, it may become impossible to compute the likelihood for a given individual even once. While different methods of asking respondents to evaluate alternatives may result in complicated probabilities, in many cases they are worth the extra effort. They may provide more information about preferences reducing the sample size needed to draw reliable conclusions, which is crucial given the high costs of administering time consuming valuation surveys. Different response formats may arise naturally in the context of the survey information, resource use, or due to future uses in policy evaluation.

Utility theory begins with the assumption that respondents order the alternatives according to relative utilities. No matter how responses are elicited, in each pair of utilities, U_{ij} is either greater than, less than, or indeterminate relative to U_{ik}. The amount and type of information retrieved will depend upon the response format, but this basic pairwise relationship holds for all response formats. Put another way, conditional on each U_{ik}, U_{ij} will either be untruncated, truncated from above, or truncated from below. This suggests that a Bayesian approach based on Gibbs sampling is a natural method for estimating MNP style models in the presence of complex response probabilities. McCulloch and Rossi (1994) first used Gibbs sampling to conduct Bayesian inference in the MNP model. With only minor modifications, their approach can be used to estimate models with *any* kind of response pattern. Complex response patterns merely change the truncation structure, not the fact that ones needs to draw a univariate normal subject to truncation. So in the context of a multivariate normal RUM, the Gibbs sampler results in a completely general approach for handling both rectangular and non-rectangular probabilities.

It is well known that the Gibbs sampler may be prohibitively slow and model parameterization can have a major impact on the rate of convergence or mixing of the induced Markov chain. In the next section, we lay out the Bayesian model for our problem and show how to implement a fast mixing Gibbs sampler to fit the model.

3 Hierarchical Bayesian model

3.1 Utilities as latent variables

Recall that our goal is to estimate WTP, which is accomplished via the Random Utility Model (1). We relate the individual random utilities to the price, forest loss, and abatement strategy attributes addressed in the survey. In our problem, we do not observe the utilities for each alternative by individual, but instead record the most and least preferred choices for each individual. Importantly, we do not observe the relative ranking of the middle alternatives, and thus can not differentiate between the utility of the other alternatives with the given observations. Since the observed choices are a function of the unobserved utilities, we shall use a *latent variable model* with latent data being the utilities implied from the observed choice data.

Let U_{ij} denote the utilities of alternative j for individual i. We denote the J-vectors of utilities over all J alternatives for individual i by $\mathbf{U}_i$. We relate these utilities to the covariates of interest through the linear mixed model (see Searle, Casella, and McCulloch, 1992, Chapter 9)

$$\begin{aligned} \mathbf{U}_i &= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{R}_i\boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_i \\ \boldsymbol{\epsilon}_i &\sim N_J(\mathbf{0},\ \boldsymbol{\Sigma}) \end{aligned} \tag{5}$$

where $\mathbf{X}_i$ denotes a $J \times K$ design matrix over the "fixed" effects, $\mathbf{R}_i$ denotes a $J \times L$ design matrix over the "random" effects, $\boldsymbol{\beta}$ denotes a K-vector of regression coefficients, $\boldsymbol{\gamma}_i$ denotes an L-vector of subject effects with respect to each of the L elements of $\mathbf{R}_i$, $\mathbf{0}$ is a J-vector of zero means, and $\boldsymbol{\Sigma}$ is the $J \times J$ dispersion of the normal distribution. In some, but not all cases, L will equal K.

Again, in our problem we do not observe the utilities, U_{ij}, for each alternative by individual, but instead record the most and least preferred choices of the J alternatives for each individual. We denote these choices by Y_{ij} for each individual by choice and the J-vectors $\mathbf{Y}_i$ for choices of individual i over all J alternatives. The specification of the observed data $\mathbf{Y}_i$ is flexible, merely a task of record keeping. For convenience, we denote the most preferred of the J alternatives considered by J, the least preferred alternative by 1, and all other alternatives in between by the value $\lfloor \frac{J+1}{2} \rfloor$ where $\lfloor \cdot \rfloor$ denotes the greatest integer function. Analogous, but different specifications do not affect the inferences drawn from the model.

We approach modeling from a Bayesian viewpoint, specifying the unknown parameters $(\boldsymbol{\beta}, \boldsymbol{\gamma}_i, \boldsymbol{\Sigma})$ through a hierarchy of prior distributions. In particular, we assume

$$\begin{aligned} \boldsymbol{\beta} &\sim N_K(\bar{\boldsymbol{\beta}},\ \mathbf{A}) \\ \boldsymbol{\gamma}_i &\sim N_L(\mathbf{0},\ \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}) \\ \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} &\sim Wishart(\nu_{\boldsymbol{\gamma}},\ \mathbf{V}_{\boldsymbol{\gamma}}) \end{aligned} \tag{6}$$

where $\bar{\boldsymbol{\beta}}$ is a known K-vector mean and $\mathbf{A}$ a known $K \times K$ dispersion matrix for the normal prior on $\boldsymbol{\beta}$ and $\nu_{\boldsymbol{\gamma}}$ and $\mathbf{V}_{\boldsymbol{\gamma}}$ are the known parameters in the Wishart prior on $\mathbf{H} = \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1}$ with density

$$\pi(H \mid \nu_{\boldsymbol{\gamma}}, \mathbf{V}_{\boldsymbol{\gamma}}) \propto |H|^{(\nu_{\boldsymbol{\gamma}} - L - 1)/2} \exp\left[tr(-0.5\mathbf{H}\mathbf{V}_{\boldsymbol{\gamma}})\right].$$

We finally presume $\boldsymbol{\Sigma} = \mathbf{I}$, a $J \times J$ identity matrix, a typical assumption in the econometric valuation literature (see Layton and Brown, 2000).

The parameters in the hierarchical mixed model (5) and (6) have intuitive interpretation in the context of our problem. The parameters $\boldsymbol{\beta}$ represent the average effects of each attribute on the utilities $\mathbf{U}_i$. This average effect is over all individuals $i = 1, \ldots, n$. The subject effects $\boldsymbol{\gamma}_i$ represent the variability in the mean effects $\boldsymbol{\beta}$, though differing over each individual. Thus the effects for each attribute on the utilities of a given individual i may be denoted by

$$\begin{aligned} \mathbf{U}_i &= \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\tau}_i \\ \boldsymbol{\tau}_i = \mathbf{R}_i \boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_i &\sim N_J(\mathbf{0},\ \boldsymbol{\Sigma} + \mathbf{R}_i \boldsymbol{\Sigma}_{\boldsymbol{\gamma}} \mathbf{R}_i^T) \end{aligned}$$

or

$$\mathbf{U}_i \sim N(\mathbf{X}_i \boldsymbol{\beta},\ \boldsymbol{\Sigma} + \mathbf{R}_i \boldsymbol{\Sigma}_{\boldsymbol{\gamma}} \mathbf{R}_i^T) \tag{7}$$

by combining the random effects and the error terms. The model specifications (5) and (7) not only shows how we, in effect, partition subject variation into iid noise, $\boldsymbol{\epsilon}_i$, and the subject effects $\boldsymbol{\gamma}_i$, but also suggests a fast implementation of the Gibbs sampler discussed in Section 3.2.

In our problem, we assume the effect of price on utilities, β_1, is the same across individuals but the effect of the other attributes on utilities varies across individuals so that $\boldsymbol{\gamma}_i$ is a $(K-1)$-vector and the design matrix $\mathbf{R}_i$ is a $J \times (K-1)$ subset of $\mathbf{X}_i$. This assumption is common in the econometric literature (see Layton and Brown, 2000), aids in parameter identification in the model (McCulloch and Rossi, 1994), and is useful for computing WTP in Section 4.4. The supposition $\boldsymbol{\Sigma} = \mathbf{I}$ forces all unknown correlations to be modeled through the subject effect dispersion $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$. This assumption is standard in econometric Bayesian analyses (see for example Albert and Chib, 1993 and Chib and Greenberg, 1996) to force model identifiability.

The prior specifications of a normal distribution on the unknown mean parameter and Wishart distribution on the unknown dispersion is standard in the Bayesian literature (Bernardo and Smith, 1995). In particular, they ensure proper posterior distributions and relatively straightforward application of the Gibbs sampler for drawing posterior inferences from $\pi(\boldsymbol{\beta} \mid \mathbf{Y})$ (see Section 3.2). Furthermore, these distributions allow for diffuse prior specification when the practitioner has no strong prior beliefs in the parameters. In fact, for diffuse priors on $\boldsymbol{\beta}$, choice of $\bar{\boldsymbol{\beta}}$ and $\mathbf{A}$ is not critical (McCulloch and Rossi, 1994). The Wishart parameters may be specified from previous analyses with sample size $\nu_{\boldsymbol{\gamma}}$ as we do in Section 4.

We may choose to place further prior distributions on the prior parameters $\bar{\boldsymbol{\beta}}, \mathbf{A}, \nu_{\boldsymbol{\gamma}}$, and $\mathbf{V}_{\boldsymbol{\gamma}}$.

However, the additional level greatly increases the complexity in fitting the model without much gain in inferential power.

3.2 Implementing the Gibbs sampler: identifiable parameters

Analytical and numerical calculation of the posterior distribution from the model (5) and (6) is infeasible due to the latent variable structure. However, the prior specifications allow for application of the Gibbs sampler for Monte Carlo posterior inferences. The Gibbs sampler generates a sample $\boldsymbol{\beta}^{(1)}, \ldots, \boldsymbol{\beta}^{(T)}$ from the posterior distribution $\pi(\boldsymbol{\beta} \mid \mathbf{Y})$ without having to directly sample from this analytically unspecified distribution. In particular, the sampler iteratively generates variates from the full conditional distributions $p(\boldsymbol{\beta} \mid \mathbf{U},\ \boldsymbol{\gamma},\ \mathbf{H},\ \mathbf{Y})$, $p(\boldsymbol{\gamma} \mid \mathbf{U},\ \boldsymbol{\beta},\ \mathbf{H},\ \mathbf{Y})$, $p(\mathbf{H} \mid \boldsymbol{\beta},\ \mathbf{U},\ \boldsymbol{\gamma},\ \mathbf{Y})$, and $p(\mathbf{U} \mid \boldsymbol{\beta},\ \mathbf{H},\ \boldsymbol{\gamma},\ \mathbf{Y})$ in such a way that the samples induce a Markov chain with stationary distribution $\pi(\boldsymbol{\beta},\ \mathbf{H} \mid \mathbf{Y})$. Here we are denoting the observed and latent data and subject effect variables over all subjects, the $(n \cdot J)$- vectors $\mathbf{Y}$, $\mathbf{U}$, and $\boldsymbol{\gamma}$, by stacking the variables $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$, $\mathbf{U}_1, \ldots, \mathbf{U}_n$, and $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_n$ respectively. If the conditional distributions are available to sample, the scheme circumvents the difficult problem of direct Monte Carlo inference from the posterior distribution of interest.

In the latent variable model, once we condition on the utilities $\mathbf{U}$ and either the subject effects $\boldsymbol{\gamma}$ or the subject effect inverted dispersion $\mathbf{H}$, the model reduces to a standard linear model with normal errors. Thus the full conditional distributions are straightforward to specify and consequently to sample. This Gibbs sampler implementation though is known to mix slowly; that is, convergence to the stationary distribution may require a prohibitively large number of iterations. See Zeger and Karim (1991), Gelfand, Sahu, and Carlin (1995), Chib, Greenberg, and Winkelmann (1998), and Chib and Carlin (1999) for further discussion and examples of this difficulty. The model (5) and (6), in particular, has difficulty in identifying parameters on a number of fronts.

The multinomial probit model has a fundamental identification problem. The model is identified only in utility differences due to its ordinal structure. Differencing with respect to the most preferred alternative resolves this. Henceforth we work with the differenced model without making it notationally explicit. Additionally the covariance matrix for the alternatives specific errors should be specified in the differenced, not un-differenced, model (see Bunch, 1991). This presents no difficulties in our implementation given our focus on the hierarchical structure. The more difficult problem is that the random effects and utilities are not easily separated or identified. The cause of slow mixing in the Gibbs sampler is a high correlation between the random effects $\boldsymbol{\gamma}$ and the utilities $\mathbf{U}$. As a consequence, the sampler does not traverse the utility space very well and we cannot easily identify the random effects apart from the other parameters in the model. This problem is further compounded when the number of choices per respondent is small as it is in our application. In this section, we discuss methods for overcoming these identifiability problems and in the process speeding up the Gibbs sampler.

Recall the model specification (7) in which we combine the random effects and the error terms. This formulation suggests that we may generate utilities $\mathbf{U}$ from a distribution dependent on the random effects $\boldsymbol{\gamma}$ through only the dispersion $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$. Furthermore, since the updates of the fixed effects $\boldsymbol{\beta}$ is a regression on the utilities with known residual error, this Gibbs step is also independent of the random effects $\boldsymbol{\gamma}$. Intuitively, once we are given the error dispersion $\boldsymbol{\Sigma} + \mathbf{R}_i\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\mathbf{R}_i^T$, knowledge of the actual $\boldsymbol{\gamma}$ values is redundant when updating the vectors $\mathbf{U}$ and $\boldsymbol{\beta}$ (thus the problem in identifying these parameters).

The conditional distributions required by this alternative implementation of the Gibbs sampler may be derived as follows. Let $\boldsymbol{\Omega}_i = \mathbf{R}_i\mathbf{H}^{-1}\mathbf{R}_i + \mathbf{I}$. We then have

$$
\begin{aligned}
\boldsymbol{\beta} \mid \mathbf{U}, \mathbf{H}, \mathbf{Y} &\sim N_K(\hat{\boldsymbol{\beta}},\ \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}) \\
\boldsymbol{\gamma}_i \mid \mathbf{U}, \boldsymbol{\beta}, \mathbf{H}, \mathbf{Y} &\sim N_K(\hat{\gamma}_i,\ \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}_i}) \\
\mathbf{H} \mid \mathbf{U}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{Y} &\sim Wishart(\nu_{\boldsymbol{\gamma}} + n, \mathbf{V}_{\boldsymbol{\gamma}} + \sum_{i=1}^{n}(\gamma_i - \bar{\gamma})(\gamma_i - \bar{\gamma})^T) \\
\mathbf{U}_{ij} \mid \mathbf{U}_{i;-j}, \boldsymbol{\beta}, \mathbf{H}, \mathbf{Y} &\sim Truncated\ Normal(m_{ij},\ \eta_{ij}^2;\ \tau_{ij}^+,\ \tau_{ij}^-)
\end{aligned} \tag{8}
$$

where $i = 1, \ldots, n$, $j = 1, \ldots, J$, $U_{i;-j}$ is the $(J-1)$-vector of $\mathbf{U}_i$ with the jth element removed, $\bar{\gamma}$ is the sample mean of the $\boldsymbol{\gamma}_i$, T denotes vector transpose, and

$$
\begin{aligned}
\hat{\beta} &= \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{n} (\mathbf{X}_i^T \boldsymbol{\Omega}_i^{-1} \mathbf{U}_i) + \mathbf{A}^{-1}\bar{\beta} \right\} \\
\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}} &= \left(\sum_{i=1}^{n} \mathbf{X}_i^T \boldsymbol{\Omega}_i^{-1} \mathbf{X}_i + \mathbf{A}^{-1} \right)^{-1} \\
\hat{\gamma}_i &= \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}_i} \left\{ \mathbf{R}_i^T (\mathbf{U}_i - \mathbf{X}_i \boldsymbol{\beta}) \right\} \\
\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}_i} &= \left(\mathbf{H} + \mathbf{R}_i^T \mathbf{R}_i \right)^{-1} \\
m_{ij} &= \mathbf{x}_{ij}\boldsymbol{\beta} + \mathbf{F}^T(\mathbf{U}_{i;-j} - \mathbf{X}_{i;-j}\boldsymbol{\beta})
\end{aligned}
$$

$$\eta_{ij}^2 = \omega_i(j;j) - \boldsymbol{\omega}_i(j;-j)\mathbf{F}$$
$$\mathbf{F} = \boldsymbol{\Omega}_i^{-1}(-j;-j)\, \boldsymbol{\omega}_i(-j;j)$$

Here $\mathbf{x}_{ij}$ is the jth row of $\mathbf{X}_i$, $\mathbf{X}_{i;-j}$ denotes $\mathbf{X}_i$ with the jth row removed, $\omega_i(j;j)$ is the (j,j) element of $\boldsymbol{\Omega}_i$, $\boldsymbol{\omega}_i(j;-j)$ and $\boldsymbol{\omega}(-j;j)$ are the jth row and column of $\boldsymbol{\Omega}_i$ with the jth element removed, and $\boldsymbol{\Omega}_i^{-1}(-j;-j)$ is $\boldsymbol{\Omega}_i$ with the jth row and column removed. The conditional means and dispersions above follow from general normal linear model theory.

The truncated normal distribution in (8) has mean m_{ij}, variance η_{ij}^2, and upper and lower truncation points τ_{ij}^+ and τ_{ij}^- respectively. The truncation points allow us to account for the ordinal content of the most/least preferred choice data in the model fit. Specifically, the truncations are functions of the choices, bounding each utility according to the utilities from the next largest and smallest choices and utilities. For the most (least) preferred alternative, the upper (lower) truncation point will be infinity (negative infinity). Therefore, the truncated normal distributions for these utilities are truncated on one side, not two. Note that any ordinal structure, not just most and least preferred choices, can be handled by suitably defining the truncation points.

The Gibbs sampler successively samples from each of the $(n \cdot J + n + 2)$ conditional distributions listed in (8). The Wishart random variate generations use the standard Monte Carlo scheme of Johnson (1987, pp. 203-204). Samples from the truncated normal distributions are generated using a combination of the accept-reject routine of Robert (1995) and the "throw away" method. The method of Robert (1995) has the most robust acceptance rate against unusually small (large) upper (lower) truncation points; that is the acceptance rates are reasonably large across the range of truncation values. We thus use this method when the upper (lower) truncation points are larger (smaller) than the mean m_{ij}. In other circumstances, we simply simulate from a univariate normal distribution until the generated numbers are between $[\tau_{ij}^-,\ \tau_{ij}^+]$.

We update each utility U_{ij} separately during each iteration of the Gibbs sampler. Thus, the utility simulation is in itself a Gibbs sampler. The structure of this scheme arises from the multivariate truncated normal distribution $p(\mathbf{U}_i \mid \boldsymbol{\beta},\ \boldsymbol{\gamma},\ \mathbf{H},\ \mathbf{Y})$. We may directly sample from this J-dimensional multivariate conditional distribution. However, successively sampling from the much easier *univariate* truncated normal conditional on U_{ij}, as in (8), is more computationally efficient. The overall algorithm is thus a Gibbs sampler embedded within a Gibbs sampler requiring updates of $(n \cdot J + n + 2)$ conditional distributions; a large number of random variate generations each step, but less costly than sampling all U_{ij} at once.

The Bayesian hierarchical model may be intuitively linked to the scheme for fitting preference data of Layton and Lee (1998). The Layton and Lee scheme permutes through all possible utilities given the stated/observed preferences. The Gibbs sampler also traverses the space of underlying utilities $\mathbf{U}$, but using the probabilistic structure of the utilities given the choice data. By defining the direction of the Gibbs chain through the utility space with the distribution of the latent variable, the Gibbs sampler more efficiently traverses this space than the algorithm of Layton and Lee (1998).

Note that the inverted dispersion matrix $\mathbf{H}$ is generated from a distribution dependent on $\boldsymbol{\gamma}$. Intuitively, we update $\boldsymbol{\gamma}$ at iteration t through a regression on the residuals $\mathbf{U} - \mathbf{X}\boldsymbol{\beta}$ with known error $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{(t-1)}$ from the previous iteration. We then use these $\boldsymbol{\gamma}^{(t)}$ to estimate $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$. Consequently, we may update the utilities $\mathbf{U}$ and the effects $\boldsymbol{\beta}$ independently of $\boldsymbol{\gamma}$, using only this dispersion $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$.

This result was also realized by Chib and Carlin (1999) from the perspective of *marginalizing* the distribution of $\mathbf{U}$ over the random effects $\boldsymbol{\gamma}$. The Gibbs sampler based on the conditional distributions (8) is similar to the naive Gibbs sampler in which we do not marginalize over the random effects, differing only in the updates of the variables $\mathbf{U}$ and $\boldsymbol{\beta}$. However, the improved mixing of the chain is substantial. In our problem as analyzed in Section 4, the Gibbs sampler above requires a burn-in on the order of 10^2 iterations (though in Section 4 we apply a conservative burn-in on the order of 10^3). The Gibbs sampler dependent on the full conditional distributions with respect to each parameter requires burn-in on the order of 10^5. The reason for this substantial improvement is that we have few observations per person and $\mathbf{R}_i$ is a subset of $\mathbf{X}_i$ in our problem. Consequently, it is difficult to identify the individual random effects $\boldsymbol{\gamma}$ from the other parameters in this model. Marginalizing over the $\boldsymbol{\gamma}$ parameters is thus essential for good mixing.

4 Results

4.1 Model specifics

We fit the survey data discussed in Section 2 using the Bayesian model of Section 3 to study an individual's willingness to pay to alleviate the effects of climate change on forest loss. Recall that in our survey, the number of alternatives J differs for each respondent in that menus consisted of five or six alternatives. The Gibbs sampler of Section 3.2, however, considers conditionals for each individual separately. Therefore, the number of alternatives is free to change over the subjects. Further, the model is identified in differences only. We denote the differenced number of alternatives for each subject by J_i, being either four or five in this data. Following Layton and Brown (2000), we consider the linear relationship between utility and seven attributes: a continuous price variable, discrete forest loss variables at three levels (loss of 600 feet, 1200 feet, and 2500 feet) with respect to the 5400 feet baseline, and discrete mitigation strategy variables at three levels (forestry, limited abatement, vigorous abatement) with respect to no action. Therefore, the model fits a $J_i \times K$ design matrix $\mathbf{X}_{ij}$ and K-vector regression coefficient $\boldsymbol{\beta}$ where $K = 7$. As noted earlier, we assume the subject effect on price has zero variance. Thus, L equals six and the model fits only six random coefficients $\boldsymbol{\gamma}_i$ for each individual.

We estimate separate models for each of the two time horizons (60 years and 150 years). The prior parameters for $\boldsymbol{\beta}$ in (6) are set to the estimates obtained of Layton and Brown (2000) on the most preferred program data. In particular,

$$\bar{\boldsymbol{\beta}} = (-0.1185,\ -11.1871,\ -5.1586,\ -1.9483,\ 1.3719,\ 3.2839,\ 1.5303)$$

and the prior dispersion on the effects $\boldsymbol{\gamma}_i$ is taken as

$$\mathbf{V}_{\boldsymbol{\gamma}}^{-1} = \begin{bmatrix} 77.78 & 42.42 & 24.13 & 0 & 0 & 0 \\ 42.42 & 25.50 & 15.16 & 0 & 0 & 0 \\ 24.13 & 15.16 & 10.49 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.60 & 0 & 0 \\ 0 & 0 & 0 & 0 & 12.63 & 18.57 \\ 0 & 0 & 0 & 0 & 18.57 & 47.06 \end{bmatrix}$$

for the 60 year horizon data and

$$\bar{\boldsymbol{\beta}} = (-0.1467,\ -14.4531,\ -5.7531,\ -1.7072,\ 1.5508,\ 4.7503,\ 3.1237)$$

and the prior dispersion on the effects $\boldsymbol{\gamma}_i$ is taken as

$$\mathbf{V}_{\boldsymbol{\gamma}}^{-1} = \begin{bmatrix} 168.72 & 74.47 & 40.05 & 0 & 0 & 0 \\ 74.47 & 47.62 & 22.98 & 0 & 0 & 0 \\ 40.05 & 22.98 & 13.97 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.34 & 0 & 0 \\ 0 & 0 & 0 & 0 & 26.03 & 45.99 \\ 0 & 0 & 0 & 0 & 45.99 & 86.04 \end{bmatrix}$$

for the 150 year horizon data. The zeros in the dispersion matrix are due to the more restrictive model estimated by Layton and Brown (2000).

We assume the prior dispersion on the effects $\boldsymbol{\beta}$ is a diagonal matrix with variances 1000 and we set $\nu = K - 1 = 6$. These values provide a diffuse prior on both $\boldsymbol{\beta}$ and $\mathbf{H}$. We initialize the Gibbs sampler with $\mathbf{H} = \mathbf{V}_{\boldsymbol{\gamma}}$, $\boldsymbol{\gamma}_i$ as a six vector of zeros for all respondents, and $\mathbf{U}_i$ as $(4 \cdot J_i)$-vectors of zeroes (for four menus).

4.2 Model comparisons

Our model imposes little a priori structure on how the survey attributes affect utility and hence WTP, and the model allows for a rich correlation structure. This approach minimizes the impact of possible specification errors, but it comes at a high computational cost. So we first compare fitting more parsimonious models to the survey data. Simpler models can be formulated by utilizing fewer random effects and/or imposing a linear structure on some of the variables. We chose these alternative models based on their relative ease of estimation and their economic reasonableness. The models are described below.

- Model I ("our model"): The model proposed in Section 3.1 and 4.1 has seven fixed effect and 21 covariance parameters.

- Model II ("no random effects"): The simplest model with respect to covariance parameters has seven fixed effect parameters and no covariance parameters.

- Model III ("regression on forest loss"): We make utility (and hence WTP) a linear function of forest loss by collapsing the four levels of forest loss into one variable measuring footage lost, yielding five fixed effect parameters. We consider forest loss as continuous via a Bayesian random coefficient model with one random effect.

- Model IV ("regression on forest loss with random effects on abatement strategies"): We extend Model III by adding random effects on the abatement strategy, yielding five fixed effect and ten covariance parameters.

We use the approach of Chib (1995), extending the algorithm of Yu and Chan (2001), for computing marginal likelihoods to compare our model with the three competing, restricted models. Table 2 shows that our model performed best, displaying a larger marginal likelihood and larger Bayes factor with respect to each of the other three models under the 60 year and 150 year horizon data. We thus draw posterior inferences and conclusions from our more flexible model (model I) fit to the survey data.

4.3 Posterior inferences

Table 3 presents posterior point and interval estimates and standard errors for each of the parameters $\boldsymbol{\beta}$ under study. The inferences are based on 5000 Gibbs samples after a burn-in of 5000 samples. MCMC diagnostic routines and plots show that the Gibbs sampler converges rapidly, suggesting a sufficient (and very conservative) burn-in period after 5000 iterations. These diagnostics are not shown here for brevity. Table 4 display posterior point and interval estimates of the dispersion parameters in $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$. Note that all covariance parameters are significantly positive. We do not have strong a priori inclinations as to the correlation between forestry and the abatement strategies. However, the positive correlations between all other variables is economically reasonable.

For both time horizons, price clearly has a strong effect on the choice of most and least preferred alternatives. The magnitude of the forest loss has an economically intuitive impact as larger amounts of forest loss lead to greater utility losses. The dispersion matrices for both time horizons exhibit the same essential structure in terms of the relative magnitudes of parameters within each matrix, and their signs. The two models can not be directly compared in terms of the magnitudes of the $\boldsymbol{\beta}$ as discrete choice models are identified only up to a scale factor. For this reason, and more importantly for economic reasons, we compare the models in terms of WTP.

4.4 Willingness to pay results

Using the parameter estimates, WTP_i is computed as the dollar amount that would leave respondent i with equal utility before and after a change in one or more of the attributes. To compute WTP_i we imagine one "status quo" or before change alternative and define its utility to person i as U_{iB}, where "B" stands for before. Next we imagine offering the person a new or "after" the change alternative and label its utility as U_{iA}. Both U_{iB} and U_{iA} depend on the attributes, estimated coefficients, and errors as in (5). At this point it is useful to distinguish between the price attribute, which is modeled as a fixed effect through the coefficient β_1, and the other $K-1$ attributes $(2:K)$. We rewrite (5) for the before and after utilities as

$$\mathbf{U}_{iA} \quad = \quad \mathbf{X}_{iA,1}\boldsymbol{\beta}_1 + \mathbf{X}_{iA,2:K}\boldsymbol{\beta}_{2:K} + \mathbf{R}_{iA}\boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_{iA}$$

$$\mathbf{U}_{iB} \quad = \quad \mathbf{X}_{iB,1}\boldsymbol{\beta}_1 + \mathbf{X}_{iB,2:K}\boldsymbol{\beta}_{2:K} + \mathbf{R}_{iB}\boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_{iB}.$$

WTP is implicitly defined for any person as the amount of money, above and beyond any other costs, that the person would be willing to pay to have the "after" alternative as opposed to the "before" alternative.

$$\mathbf{X}_{iB,1}\boldsymbol{\beta_1} + \mathbf{X}_{iB,2:k}\boldsymbol{\beta_{2:k}} + \mathbf{R}_{iB}\boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_{iB} \quad = \quad (\mathbf{X}_{iA,1} + WTP_i)\boldsymbol{\beta_1} + \mathbf{X}_{iA,2:k}\boldsymbol{\beta_{2:k}} + \mathbf{R}_{iA}\boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_{iA}. \tag{9}$$

Given the random utility model, WTP is a random variable, and expected WTP is typically used (which following convention we will denote simply as WTP as opposed to explicitly acknowledging the expectation). To find the WTP, note that each attribute in (9) is separable in utility from the others, so we can compute the expected WTP for a one unit change in each attribute separately. Next, for the policy options we consider and given that costs enter the model linearly, we can normalize the program costs to zero so that we will find the WTP for a change in the state of the world. Then using the independence of γ_i and ϵ_i, and that price is modeled as a fixed effect through the coefficient β_1, the expected WTP for a unit change in each attribute $2 : K$ takes a particularly simple form by solving for WTP_i. Note that the difference of the error terms have an expected value of zero as do the random effects, so

$$WTP_k = -\beta_k/\beta_1. \tag{10}$$

Using the 5000 draws from the posterior distribution of the parameter estimates, we compute 5000 draws from the posterior distribution of WTP for each of the three forest loss levels, across both time horizons using (10). If for a given draw the coefficients on the forest loss and price parameters are both negative, the WTP derived from (10) will be negative, indicating that individuals experience utility losses due to forest losses, or that they would be willing to pay a positive amount to avoid forest losses. In Table 5 we present point and interval estimates for the WTP to avoid the three forest loss levels. For both time horizons, the WTP for the 2500 foot loss is significant. In WTP per foot of forest loss, the incremental WTP appears to increase non-linearly with the severity of forest loss.

Notably, the three WTP's for the 150 year horizons are lower than the three WTP's for the 60 year model. It certainly looks like respondents are discounting the impacts that are much farther into the future. The draws for the poster distribution of WTP for the two time horizons make it easy to compute the implicit discount rate used by the sample of respondents. Labeling the WTP for an impact occurring in 60 or 150 years as WTP60 and WTP150 respectively (and suppressing an index for each of the three levels of impacts), we note that under the assumption of discounting, the WTP60 and WTP150 are each related to the WTP to prevent an immediate impact, WTP0, as

$$WTP60 = \exp(-60\rho) \cdot WTP0 \text{ and } WTP150 = \exp(-150\rho) \cdot WTP0.$$

This holds for each level of impact (e.g. comparing a 600 foot forest loss occurring in 60 versus a 600 foot forest loss occurring in 150 years, but not comparing a 600 foot forest loss in 60 years versus a 1200 foot forest loss in 150 years). We can take the ratio of WTP60 and WTP150 and solve for the discount rate

$$\rho = \{\ln(WTP60/WTP150)\}/90. \tag{11}$$

The draws from the posterior of WTP60 and WTP150 are independent, and so we sample pairs for WTP60 and WTP150 for all three levels of impact and compute ρ as in (11). Although the discount rate should be positive, random draws will of course sometimes result in a negative rate. Any such draws are treated as zero in the analysis which follows. The alternative approach of attempting to place priors on the models such that the implied discount rate is always positive does not appear feasible.

Weitzman (2001) surveyed 2,160 economists and asked them “Taking all relevant considerations into account, what real interest rate do you think should be used to discount over time the (expected) benefits and (expected) costs of projects being proposed to mitigate possible effects of global climate change?” This is exactly what our two sub-samples were implicitly asked to do for forest loss in Colorado. Weitzman (2001) provides the entire distribution of responses rounded to the nearest whole percent, so for comparison purposes we tabulate our results in the same fashion (we omit the 20 out of the 2160 responses in Weitzman’s table that were less than 0 or greater than 15.5% to make the table easier to read. For this reason column one does not sum to 100%). Table 6 provides the percentage of discount rate draws in each discount rate interval, and a number of other measures we discuss below.

Before discussing the results, it is important to consider what the nature of interesting hypotheses regarding the discount rate are. First, point null hypotheses regarding the discount rate are not interesting. Policy analysis will not turn on a question such as “is the discount rate equal to 3%”.

Economic simulation models of climate change such as Nordhaus (1994) use at least a "high" and "low" estimate of the discount rate to test sensitivity of the results. Interesting and policy useful hypotheses must be interval in nature. As Levine and Casella (1996) argue that the posterior odds are a more useful measure than the Bayes factor for interval null hypotheses, we present posterior odds for each 1% interval. Table 6 presents the posterior and prior odds in favor of the discount rate interval with respect to all other intervals based on the Gibbs sample and the Weitzman (2001) data respectively. Examination of the posterior odds indicates that the majority of the probability mass falls in the 0.5% to 1.5% range. We conclude that the best estimate of the discount rate is about 1%.

We next compare our results to the collective beliefs of the economics profession. In columns 2, 5, and 6 of Table 6, we reproduce the aggregated responses from Weitzman (2001), the odds for the Weitzman (2001) data, and a "Bayes factor" style normalization of our posterior odds using the Weitzman (2001) odds as a "prior". The Bayes factor in Table 6 is not a true Bayes factor as we do not use our own prior odds. In fact, it would be difficult to construct a computationally tractable joint model by placing a prior over the implicit discount rate. The "Bayes factors" we present in Table 6 are meant to be illustrative of the following thought experiment: imagine that one takes Weitzman's (2001) survey of economists and conducts a new experiment (ours) and attaches 50% weight to each, then what conclusion do we draw. First, the odds for the economics profession appear to be much more diffuse than in our model, with most of the probability falling in the 0.5% - 5.5% range as opposed to 0% - 2.5% in this study. The "Bayes factor" suggests that in our hypothetical experiment the weight of evidence has shifted sharply downwards towards less than 1.5%.

5 Discussion

5.1 Generality of the modeling strategy

The hierarchical Bayesian model of Section 3 provides for a tractable modeling strategy for analyzing complex ordinal data. The approach is quite general, not restricted to rectangular probabilities or specific alternative structures in the survey. Such a modeling approach is thus applicable to many practical problems in which the surveyor elicits some kind of discrete response with respect to well defined alternatives. We have seen such surveys concerning quality of life among the permanently disabled, customer satisfaction surveys, new product development, quality assurance in cancer detection, recitivism and parolee program evaluations, election poll surveys, resource management, to name a few, which pervade sociological, agricultural, economic, marketing, transportation research, and medical sciences.

Of particular note is the flexibility in the Bayesian hierarchical model. We are not restricted to conjugate prior distributions. We may explicitly model the willingness to pay and force truncated (greater or less than zero) or skewed distributions on these parameters. For example, we may assume specific structure on a random price coefficient. A natural candidate would be an inverted gamma distribution. Furthermore, we may lift the supposition of iid errors in (5) to allow for more general error dispersions. Hastings algorithms provide for feasible implementations of MCMC sampling towards fitting such complex models. We note however, that careful study of parameter identification may be necessary to ensure a reasonably mixing Markov chain sampler, as discussed in Section 3.2.

5.2 Conclusion

The research has sought to address two important questions relating to current climate change policy debates. The first concerns the willingness of the public to bear costs today in order to prevent future ecosystem impacts due to climate change. Second, we have provided the first estimates to our knowledge of the public's discount rate for mitigating climate change impacts. The hierarchical Bayesian model and the Gibbs sampler allowed us to formulate a rich model for a high dimensional problem, while easily handling non-rectangular probabilities that can prove difficult in simulated maximum likelihood approaches. Our model comparisons demonstrated that the additional complexity of our model was well worth the extra effort. In terms of implementation, we utilized a novel blocking strategy for the Gibbs sampler. We found that this simple modification of the naive sampler resulted in a very substantial improvement in convergence, improving burn-in by three orders of magnitude.

The public's WTP to prevent ecosystem impacts is insignificantly different from zero for "small" impacts, but increases sharply as the impacts approach complete ecosystem change. Some have criticized environmental valuation surveys for providing significant values for any level of impact, no matter how small, and for generating values that appear insensitive to the magnitude of the impacts. This is clearly not the case here: scope matters, the public is willing to spend a great deal to prevent complete ecosystem change, but small to moderate changes do not appear to be very costly to the public. This strikes us as entirely realistic.

To our knowledge, this study provides the first estimate of the public's discount rate for climate change. Previously economists have chosen, and argued over, the appropriate discount rate through a combination of economic theory and empirical work focusing on the rates of return to capital throughout the economy, future economic growth, the marginal utility of consumption, and a myriad of other important economic variables (Arrow et al., 1996; Nordhaus, 1994). We do not wish to dispute the importance of this work, but do wish to point out that this approach is complicated by the need for forecasting into the far future. Our approach has the appeal of policy simplicity. Ultimately it is the public that will bear the cost of climate change mitigation, why not ascertain what discount rate the public uses? Our estimates conclude that the public's discount rate is about 1%. Interestingly, and probably not surprisingly to non-economists, the public's estimated discount rate is lower than that suggested by the economics profession. This means that the costs of future climate change impacts loom larger for the public than for the economics profession, and the public would find in its best interest some programs economists would reject. Further, the posterior odds indicate a relatively tight interval for the estimated discount rate - much tighter than for the economist data. This relatively precise estimate for the discount rate is due to the quality of the survey data, but also due to our ability to model the most and least preferred choices and the associated non-rectangular probabilities via the Gibbs sampler.

References

[1] Albert, J. H. and Chib, S. (1993). Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association* 88, 669-679.

[2] Arrow, K. J., Cline, W. R., Maler, K-G., Munasinghe, R., Squitieri, R., and Stiglitz, J. E. (1996). Intertemporal Equity, Discounting, and Economic Efficiency. In *Climate Change 1995: Economic and Social Dimensions of Climate Change*, contribution of Working Group III to the Second Assessment Report of the Intergovernmental Panel on Climate Change. Bruce, J. P., Lee, H., and Haites, E. F. (eds.). Cambridge University Press, Cambridge, U.K.

[3] Bernardo, J. and Smith, A. F. M. (1995). *Bayesian Theory.* Wiley, New York.

[4] Borsch-Supan, A. and Hajivassiliou, V.A. (1993). Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models. *Journal of Econometrics* 58, 347-68.

[5] Bunch, D.S. (1991). Estimability in the Multinomial Probit Model. *Transportation Research B* 25, 1-12.

[6] Chib, S. (1995). Marginal Likelihood from the Gibbs Output. *Journal of the American Statistical Association* 90, 1313-1321.

[7] Chib, S. and Carlin, B. P. (1999). On MCMC Sampling in Hierarchical Longitudinal Models. *Statistics and Computing* 9, 17-26.

[8] Chib, S. and Greenberg, E. (1996). Markov Chain Monte Carlo Simulation Methods in Econometrics. *Econometric Theory* 12, 409-431.

[9] Chib, S., Greenberg, E., and Winkelmann, R. (1998). Posterior Simulation and Bayes Factors in Panel Count Data Models. *Journal of Econometrics* 86, 33-54.

[10] Gelfand, A. E., Sahu, S. K., and Carlin, B. P. (1995). Efficient Parametrisations for Normal Linear Mixed Models. *Biometrika* 82, 479-488.

[11] Hajivassiliou, V.A. and McFadden, D.L. (1998). The Method of Simulated Scores for the Estimation of LDV Models. *Econometrica* 66, 863-96.

[12] Hajivassiliou, V.A., McFadden, D.L., and Ruud, P. (1996). Simulation of Multivariate Normal Rectangle Probabilities and Their Derivatives: Theoretical and Computational Results. *Journal of Econometrics* 72, 85-134.

[13] Hajivassiliou, V. A. and Ruud, P. A. (1994). Classical Estimation Methods for LDV Models Using Simulation. In R. Engle and D. McFadden (eds.) *Handbook of Econometrics*, Volume 4. Elsevier, North-Holland, 2383-2441.

[14] Intergovernmental Panel on Climate Change (1998). *The Regional Impacts of Climate Change: An Assessment of Vulnerability.* Cambridge University Press, Cambridge, U.K.

[15] Johnson, M. (1987). *Multivariate Statistical Simulation.* Wiley, New York.

[16] Layton, D. F. and Brown, G. (2000). Heterogeneous Preferences Regarding Global Climate Change. *The Review of Economics and Statistics* 82(4), 616-624.

[17] Layton, D. F. and Lee, S. T. (1998). From Ratings to Rankings: The Econometric Analysis of Stated Preference Ratings Data. Presented at the World Congress of Environmental and Resource Economists, Venice Italy, June 1998.

[18] Levine, R. A. and Casella, G. (1996). Convergence of Posterior Odds. *Journal of Statistical Planning and Inference* 55, 331-344.

[19] McCulloch, R. and Rossi, R. E. (1994). An Exact Likelihood Analysis of the Multinomial Probit Model. *Journal of Econometrics* 64, 207-240.

[20] McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka (ed.), *Frontiers in Econometrics*, New York, Academic Press.

[21] McFadden, D. (1981). Econometric Models of Probabilistic Choice. In C.F. Manski and D. McFadden (eds.) *Structural Analysis of Discrete Choice Data with Econometric Applications*. MIT Press, Cambridge MA.

[22] McFadden, D. (1989). A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration. *Econometrica* 57, 995-1026.

[23] Nordhaus, W.D. (1994). *Managing the Global Commons*. MIT Press. Cambridge, Massachusetts.

[24] Portney, P.R. and Weyant, J.P. (eds.) (1999). *Discounting and Intergenerational Equity*. Resources for the Future, Washington, D.C.

[25] Robert, C. P. (1995). Simulation of Truncated Normal Variables. *Statistics and Computing* 5, 121-125.

[26] Searle, S. R., Casella, G., and McCulloch, C. E. (1992). *Variance Components*. Wiley, New York.

[27] VEMAP Members (1995). Vegetation/Ecosystem Modeling and Analysis Project (VEMAP): Comparing Biogeography and Biogeochemistry Models in a Continental-Scale Study of Terrestrial Ecosystem Responses to Climate Change and CO2 Doubling. *Global Biogeochemical Cycles*, 9(4), 407-437.

[28] Weitzman, M.L. (2001). Gamma Discounting. *American Economic Review* 91, 260-271.

[29] Yu, P. L. H. and Chan, L. K. Y. (2001). Bayesian Analysis of Wandering Vector Models for Displaying Ranking Data. *Statistics Sinica* 11, 445-461.

[30] Zeger, S. L. and Karim, R. (1991). Generalized Linear Models with Random Effects: A Gibbs Sampling Approach. *Journal of the American Statistical Association* 86, 79-86.

Table 1: Survey Experimental Design: The SP survey experimental design presented to each respondent four different menus of alternatives from which they chose their most and least preferred alternative. Each alternative consisted of five attributes present at different levels. The attributes were: Forest Loss, Forestry, Abatement, Program Cost, and Time Horizon. These five attributes could vary across alternatives within a menu, across the four menus presented to a respondent, and/or across respondents.

Attribute	*Levels used*	*Varies over alternatives*	*Varies over menus*	*Varies over individuals*
Forest Loss (measured in elevation)	Four levels: 0 ft., 600 ft., 1200 ft., 2500 ft.	Yes	Yes	Yes
Forestry	Two levels: Utilized or Not utilized	Yes	No	No
Abatement	Three levels: None, Limited, Vigorous	Yes	No	No
Program Cost (in $'s per month)	Fifteen levels: $0, $2.5, $10, $12.5, $20, $22.5, $30, $32.5, $40, $42.5 $50, $60, $62.5, $70, $80	Yes	Yes	Yes
Time Horizon	Two levels: 60 years or 150 years	No	No	Yes

Table 2: Log-marginal likelihood of the four models under consideration.

Model	Log-marginal likelihood 60 year horizon	Log-marginal likelihood 150 year horizon
Model I	-676.5	-596.7
Model II	-769.7	-819.6
Model III	-1,957.4	-1,930.7
Model IV	-995.4	-927.8

Table 3: Posterior point and interval estimates for the seven regression coefficients $\boldsymbol{\beta}$ under the best and least preferred alternatives data.

	60 year horizon			*150 year horizon*		
Parameters	Estimate	Std. error	95% credible set	Estimate	Std. error	95% credible set
Price	-0.02	0.01	(-0.04, -0.01)	-0.03	0.01	(-0.05, -0.02)
2500 ft. loss	-3.50	0.70	(-4.89, -2.11)	-2.66	0.72	(-4.07, -1.51)
1200 ft. loss	-0.90	0.45	(-1.80, -0.01)	-0.55	0.44	(-1.41, 0.29)
600 ft. loss	-0.17	0.28	(-0.73, 0.38)	0.10	0.29	(-0.47, 0.66)
Forestry	0.31	0.29	(-0.24, 0.90)	0.67	0.30	(0.06, 1.27)
Lim. Abate	1.09	0.39	(0.34, 1.85)	1.44	0.40	(0.66, 2.21)
Vig. Abate	0.77	0.75	(-0.68, 2.25)	1.55	0.79	(0.00, 3.06)

Table 4: Posterior point and interval estimates for the 21 dispersion parameters in $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ under the best and least preferred alternatives data. The subscripts, one through six, corresponds to the order of the latter six variables in Table 3.

	60 year horizon			*150 year horizon*		
Parameters	Estimate	Std. error	95% credible set	Estimate	Std. error	95% credible set
σ_{11}	27.73	5.05	(19.15, 38.90)	34.94	6.48	(24.07, 49.27)
σ_{12}	17.13	3.19	(11.76, 24.26)	20.23	3.91	(13.68, 28.81)
σ_{13}	9.77	1.92	(6.44, 14.10)	11.12	2.30	(7.24, 16.31)
σ_{14}	9.72	1.89	(6.50, 13.94)	12.63	2.54	(8.42, 18.34)
σ_{15}	9.68	2.15	(6.02, 14.41)	13.03	2.84	(8.20, 19.34)
σ_{16}	20.18	4.25	(13.02, 29.50)	26.51	5.72	(16.95, 39.20)
σ_{22}	11.47	2.12	(7.92, 16.27)	13.07	2.52	(8.83, 18.53)
σ_{23}	6.43	1.25	(4.30, 9.19)	6.91	1.43	(4.52, 10.11)
σ_{24}	5.83	1.17	(3.83, 8.48)	6.90	1.47	(4.37, 10.21)
σ_{25}	5.50	1.30	(3.29, 8.37)	6.98	1.63	(4.23, 10.58)
σ_{26}	12.06	2.61	(7.59, 17.68)	14.57	3.29	(8.97, 21.72)
σ_{33}	4.97	0.87	(3.50, 6.85)	5.26	0.98	(3.63, 7.47)
σ_{34}	3.15	0.72	(1.91, 4.75)	3.74	0.92	(2.20, 5.82)
σ_{35}	2.67	0.82	(1.23, 4.45)	3.71	1.06	(1.90, 6.08)
σ_{36}	6.10	1.62	(3.27, 9.57)	8.13	2.14	(4.55, 12.97)
σ_{44}	4.53	0.84	(3.09, 6.38)	6.44	1.24	(4.38, 9.20)
σ_{45}	4.69	0.96	(3.03, 6.84)	7.05	1.42	(4.66, 10.17)
σ_{46}	9.56	1.92	(6.31, 13.88)	14.12	2.88	(9.30, 20.49)
σ_{55}	6.91	1.28	(4.73, 9.81)	9.45	1.79	(6.46, 13.43)
σ_{56}	12.88	2.44	(8.73, 18.37)	18.04	3.52	(12.13, 25.92)
σ_{66}	27.16	4.99	(18.73, 38.19)	37.10	7.25	(24.95, 52.95)

Table 5: Posterior willingness to pay in dollars with interval estimates for the three forest loss levels under the best and least preferred alternatives data.

	60 year horizon		*150 year horizon*	
Parameters	WTP	95% credible set	WTP	95% credible set
2500 ft. loss	182.07	(71.23, 459.35)	87.74	(35.00, 176.34)
1200 ft. loss	46.91	(-0.46, 142.18)	18.25	(-9.53, 54.17)
600 ft. loss	8.89	(-20.54, 45.39)	-3.15	(-22.91, 16.76)

Table 6: Discount rate results from the data collected by Weitzman (2001) and the hierarchical Bayesian model fit to our survey data. Columns 2 and 3 present the percentage of discount rate (ρ) draws in the given discount rate intervals as based on the data collected by Weitzman (2001) and the Gibbs sample respectively. Columns 4 and 5 present the posterior and prior odds in favor of the discount rate interval with respect to all other intervals based on the Gibbs sample and the Weitzman (2001) data respectively. Column 6 presents the Bayes factor being the ratio of the posterior odds to the prior odds.

ρ interval in percent	Weitzman	Gibbs	posterior odds	prior odds	Bayes factor
0 - 0.5	2.1	36.4	0.573	0.022	26.33
0.5 - 1.5	10.9	50.8	1.034	0.123	8.43
1.5 - 2.5	21.0	11.1	0.125	0.266	0.47
2.5 - 3.5	19.8	1.2	0.013	0.246	0.05
3.5 - 4.5	16.8	0.2	0.003	0.201	0.01
4.5 - 5.5	10.5	0.1	0.001	0.117	0.01
5.5 - 6.5	6.3	0.0	0.000	0.067	0.00
6.5 - 7.5	3.3	0.0	0.000	0.034	0.01
7.5 - 8.5	2.0	0.0	0.000	0.021	0.00
8.5 - 9.5	1.3	0.0	0.000	0.013	0.00
9.5 - 10.5	2.0	0.0	0.000	0.021	0.00
10.5 - 11.5	0.7	0.0	0.000	0.007	0.00
11.5 - 12.5	1.2	0.0	0.000	0.012	0.00
12.5 - 13.5	0.6	0.0	0.000	0.006	0.00
13.5 - 14.5	0.2	0.0	0.000	0.002	0.00
14.5 - 15.5	0.4	0.0	0.000	0.004	0.00