Blending Supplementary Surveys with Citizen Science Data to Estimate Representative Use Values for Non-market Environmental Goods: An eBird example

Sonja Kolstoe and Trudy Ann Cameron

This Version: December 1, 2019
Check for Newer Version Online
Preliminary; not to be quoted without permission

Abstract

In this study, we address how to render the benefits estimates from citizen science data scalable to the general population from a random utility maximization (RUM) model of destination site choice. We conducted a special-purpose supplementary survey of eBird members to collect additional data above what is normally collected by eBird, to include their engagement with eBird. We also collected data from a Qualtrics Omnibus (qBus) survey of the general population with engagement in citizen science projects and bird trip behavior. This additional information allows us to develop population weights to correct for the systematic sampling problem when these citizen science data are used by themselves. We use these two data sources to generate de-meaned response propensities to allow all key parameters in the destination site choice model to reflect the average propensity in the general population. We also generate weights based on predicted engagement intensity with eBird as part of our sample selection correction to make it representative of the distribution of engagement propensities in the general population of the U.S. Both of these sample selection corrections are necessary to make citizen science data useful for policy makers who need population-level benefits of recreational demand of non-market environmental goods and services for comprehensive benefit-cost analyses to overcome the sample of convenience problem in citizen science data. We also relax an assumption made in previous work that the marginal benefit of an additional bird species is the same across all species. In our current analysis, we allow the marginal utility of species richness (a biodiversity measure) to be different across four different categories of bird species: water birds, birds of prey, perching birds and game birds.

*Steve Kelling at the Cornell University Laboratory of Ornithology generously provided us with access to the eBird data. We are grateful for comments from participants at the 2019 Southern Economic Association conference as well as feedback on earlier versions of this paper from participants at the 2018 Society for Benefit-Cost Analysis conference, 2018 World Congress for Environmental and Resource Economists, Clarkson University Biology Seminar, California State University - Chico Seminar, DePauw University Seminar and University of Delaware Seminar. This work has been supported in part by the endowment accompanying the Raymond F. Mikesell Chair in Environmental and Resource Economics at the University of Oregon. All errors are the responsibility of the authors.

1 Introduction

Volunteered geographic information (VGI) gathered by citizen science projects and social-media platforms are continually capturing "passively generated" data on human behavior and their use of non-market environmental goods and services. These data may be exploited to model recreation demand using destination site choice models based on travel-cost information that reveal people's willingness-to-pay (WTP) for improvements in environmental assets (Kolstoe and Cameron, 2017; Kolstoe et al., 2018). These are "passively generated" data on human behavior and their use of non-market environmental goods and services, meaning data on this behavior now exists where it previously did only in a limited capacity or not at all. Plus, some projects now involve numbers of observations that qualify as "big data."

One way in which economists and other social scientists can help policy-makers take advantage of the power of citizen science data is to demonstrate the ways in which these data can be brought to bear on questions concerning the non-market valuation of environmental goods, especially for use in benefit-cost analysis of environmental policy. In recognition of the vast potential of citizen science data to inform policy-making, the U.S. Environmental Protection Agency (EPA), via its National Advisory Council for Environmental Policy and Technology (NACEPT), issued a report on 13 December 2016 stating that the EPA needs to "embrace citizen science as a core tenet of environmental protection." However, prior to this data being used to inform policy-making through the estimation of benefits, the question of how to correct for the sample of convenience problem must be addressed.

In this paper, we endeavor to answer two specific research questions. The first question concerns how the value of bird biodiversity (to eBirders) varies across different categories of bird species (e.g. water birds, birds of prey, perching birds and game birds), as well as by birder type (e.g. for casual birders versus "listers"). The second question concerns what additional information is necessary to render benefits estimates from citizen science data scalable to the general population. We use a different and newer sample from the eBird project, supplemented with additional information about each eBirder from a supplementary survey of these same people, to build on the research reported in Kolstoe and Cameron (2017) and Kolstoe et al. (2018).

Only a few published papers have used citizen science data, or geocoded information from social-media platforms to estimate willingness-to-pay (WTP) use values for the represented populations. A paper by Kolstoe and Cameron (2017) uses off-the-shelf citizen science data from eBird, a large year-round citizen science project run by the Cornell Lab of Ornithology to estimate eBirders WTP for bird biodiversity as evidenced by the destination choices in their reported birding trips. Kolstoe et al. (2018) extends the earlier paper to look at predicted welfare impacts of urbanization and climate change using forecasts from Homer et al. (2015) and Langham et al. (2015). A paper by Keeler et al. (2015) uses geo-coded photography data from the social-media platform, Flickr, to estimate members WTP for lake water clarity. Both of these papers demonstrate the potential to use a sample of convenience to infer non-market values of an environmental good, but these values are not necessarily scalable to the general population because the estimating samples are not representative. Both of these papers include caveats that the estimated benefits apply only to the sample and not to the general population, but population-level benefits are necessary for comprehensive benefit-cost analyses for policy purposes.

We specifically address how to employ supplementary survey data to help overcome the "sample of convenience" problem by gathering supplemental data on the sample of convenience of interest (eBird members) as well as in a sample from the general population (Qualtrics Omnibus), building on the work in Cameron and Kolstoe (2019). Our auxiliary survey data was gathered specifically to permit the construction of population weights to correct for the systematic samples to be expected in data contributed by volunteers (eBird members), who (in this case) are more likely than average to be avid birders (Green, 2007). We also use a representative sample from the Qualtrics Omnibus (qBus) survey of the general population of the U.S. where we contributed two questions, one on their level of engagement with the eBird citizen-science project and another on their bird-watching away from home (i.e. more than a mile) behavior in the past year.[1] Respondents are invited to indicate their level of engagement with the eBird project in six ordered and mutually exclusive categories: whether they are (1) unfamiliar with the project; (2) have heard of the project, but are not a member; are a member, but (3) report their bird sightings only rarely; (4) report sighting less than half of the time; (5) report more than half the time; or whether they (6) report virtually all of their sightings. Our second question asked about the number of days per year, if any, they traveled more than one mile from home to see birds. This allows us to calculate population weights to correct for over- or under-representation at the four different engagement levels represented in the eBird sample based on the eBird engagement-level probabilities for both the qBus sample and the separate sample of eBird participants.

We also work to incorporate work from Cameron and Kolstoe (2019) who build on sample selection techniques from Cameron and DeShazo (2013), but focus in this paper specifically on those that are suitable for use in choice models. We use the predicted relative probabilities of eBird engagement levels in the general-population qBus sample and the selected eBird citizen-science sample to build weights for each observation in the eBird citizen science sample that more accurately reflect the relative frequencies in the general population of each of four different eBird participation levels. We allow these probability-based weights to vary systematically with each individual's sociodemographic characteristics, rather than be calculated simply from the observed relative frequencies at different participation levels in each sample, conditional on being a member of eBird from either sample. This allows

[1]Respondents were asked about their level of engagement with eBird as well as thirteen other birding-related citizen-science projects. The bird-watching away from home definition comes from the U.S. Fish and Wildlife quinquennial survey.

for the possibility that the qBus sample represents a different mix of sociodemographic characteristics than the eBird sample (which it does).

Section 2 provides a brief overview of the relevant literatures: recreation demand models, sample selection and VGI data uses. Section 3 outlines the data sets used in this analysis. Section 4 outlines our empirical approach, including the sample selection correction model for participation, the random utility model and its estimation. We also describe how to construct and use population weights based on the U.S. Fish & Wildlife quinquennial survey of Fishing, Hunting, and Wildlife-Associated Recreation, as well as a special-purpose question posed to 2,000 nationwide general-population respondents as part of the Qualtrics Omnibus survey to construct weights correcting for participation intensity with citizen science projects. Section 5 discusses our estimation results. Section 6 discusses the willingness-to-pay (benefits) estimates inferred from the parameter estimates for our preferred specification employing our auxiliary data, and Section 7 concludes.

2 Relevant Literature

2.1 Recreation Demand Models

Recreational demand models originated with the single-site travel cost method (TCM) to destination site choice as methods evolved to be able to handle choices among multiple sites, based on site attributes, using the random utility maximization (RUM) model framework.[2] The conditional logit estimator used with RUM models has long been a workhorse in the

[2]McFadden (1974) first proposed the random utility maximization (RUM) model as an approach that would make it possible to infer utility-theoretic population demands for a new public transportation system from consumer choices among a variety of existing transportation modes. Bockstael et al. (1989) applied McFadden's model to environmental valuation and it is now the basis for the predominant approach to valuation of environmental attributes, at least in a context where "use" demand can be observed and demand is expressed through a willingness to incur travel costs to gain access to an environmental good; it exploits travel costs and destination attributes, along with the socioeconomic characteristics of recreationists to explain destination choices.

literature; advances in computer processing power and econometrics have led to the development of logit models which relax the restrictive property known as the independence of irrelevant alternatives (IIA) of the conditional logit (Phaneuf and Requate, 2017, Chapter 17). The IIA assumption assumes that the unobserved components of utility (for a given choice occasion) are uncorrelated across alternatives, which may not be true. Train (2009, Chapter 6) points to this assumption as being unrealistic, given the strong possibility that unobserved site attributes are correlated within subsets of alternatives. Concern about the problem of unobserved heterogeneity, and a desire for a more-flexible RUM model, led to the development of the mixed logit, which allows for random parameters, based on some parametric distribution of the random parameter(s) (e.g normal or log-normal). This approach has desirable characteristics when modeling repeated choices to include allowing for unobserved heterogeneity and a more-flexible pattern of substitution between alternatives and for correlations across choices, by estimating a joint distribution for a subset of random parameters in the model (Train, 2009, Chapter 6).

2.1.1 Heterogeneity in Preferences

To accommodate preference heterogeneity, we follow the first two recommendations of Moeltner and Von Haefen (2011), who point out that there are basically three ways to do this: (1) including observable socio-demographic variables; (2) the use of so-called "mixed" logit models with one or more random parameters; and (3) the use of latent class models.[3] The first method is perhaps the most straightforward when the mix of sociodemographic characteristics in the sample can be mapped to the corresponding (joint) distribution of these characteristics in the population. This information permits the researcher to generate exogenous weights based on the relative frequencies of these observable characteristics in the population as opposed to the sample. The first two methods are the most commonly seen

[3]Due to convergence problems with a large number of parameters we choose to not use latent class models for this analysis.

in the literature, and in particular the second option because the mixed logit model also allows for implicit but very flexible substitution patterns between sites. This is particularly advantageous in the "repeated RUM" model with panel data when there is information for multiple choice occasions for each individual. In additional to allowing for heterogeneity across consumers in the sizes of the marginal utilities derived from specific attributes, the mixed logit allows for correlations across consumers in the marginal utility derived from different attributes. It may be expected that consumers' preferences for one attribute at a site can be related to their preferences for another attribute at the site (as well as over time, in the case of the repeated RUM). For example, see Revelt and Train (1998); Train (1998). Finally, McFadden and Train (2000) show that any choice model can be approximated by a mixed logit.

2.1.2 Willingness To Pay Space

The standard approach to estimating discrete choice models to obtain willingness to pay (WTP) measures for welfare or benefit-cost analysis can be somewhat problematic when the model is estimated in preference space and then solved to obtain the corresponding WTP function. Calculation of the WTP function involves ratios of asymptotically normally distributed preference parameters estimated by maximum likelihood. Using the Delta, Fieller or Krinsky-Robb approaches may lead to large positive WTP measures if the price coefficient is small (and large negative values if the price coefficient is unconstrained and passes slightly into the negative domain). WTP estimates by these methods can have large variances, whereas more-reasonable WTP estimates tend to be found when models are estimated directly in WTP space. Both Train and Weeks (2005) and Scarpa et al. (2008) note that the marginal utility of net income, as captured by the price coefficient in these models, is unlikely to be constant across people. To ensure that a random price parameter is negative, researchers may parameterize the coefficient on price as $\beta = -exp(\beta^*)$ so that the estimated parameter, β^*, has an unbounded normal distribution but the effective coefficient $-\beta$, while random, remains strictly negative. However, the model can be re-parameterized and estimated directly in WTP space. In that case, the ratios of the marginal utility of each attribute to the marginal utility of net income can be estimated directly as redefined single parameters, rather than estimating the two components of the ratio separately and having to calculate corresponding values for these ratios. In WTP space, the price coefficient as well as other key parameters can be random (and, if appropriate, correlated). This specification allows for scale-heterogeneity in addition to the other desirable features of the mixed logit, which include its ability to relax the independence of irrelevant alternatives (IIA) assumption to capture unobserved heterogeneity and to allow for more-flexible substitution patterns across alternatives (see Train (1998); Scarpa et al. (2008); Thiene and Scarpa (2009); Hess and Train (2017)).

2.2 Sample Selection

To overcome the problem of citizen-science samples of human participants being "samples of convenience" requires addressing the sample-selection problem which is not well developed for choice models as Yuan et al. (2015) and Johnston and Abdulrahman (2017) point out. A reduced-form approach adjusting for response propensity is the typical form seen in choice models (see Cameron and DeShazo (2013); Johnston and Abdulrahman (2017); Kolstoe and Cameron (2017)). Cameron and Kolstoe (2019) point out while an inverse Mills Ratio is commonly used to correct for systematic selection in models, it is not appropriate when the outcome model in the second stage is a logit specification. This is because the bivariate normality assumption for the errors in the two equations (selection and outcome) does not hold and thus the Heckman logic motivating the inclusion of something like an inverse Mills Ratio is only appropriate when the selection propensity and the outcome variable are jointly normally distributed.

2.3 Data from VGI Projects

Citizen science data (e.g. eBird), as well as the voluminous quantities of data that can be harvested from geotagged sources (e.g. Flickr, Instagram and Twitter) can provide researchers with more and different information than they have ever enjoyed before. The eBird citizen science dataset contains information contributed by bird-watchers who are project "members." The early data were sparse, but improved as eBird transformed from being a citizen science project focused on collecting data for just scientists to use, into a collaborative effort with the exchange of information between citizen scientists and researchers (Sullivan et al., 2014). The success of eBird is attributed to its role as a project that engages the data-gathers (the citizen scientists) with the researchers and the research products being generated from the data. The different eBird participants, be they data-gathers or data-users, are all stakeholders in the success of the eBird project. This level of two-way engagement appears to be an important factor contributing to the sustained success of this particular citizen science project (Sullivan et al., 2014; Venturelli et al., 2017).

3 Empirical Strategy

3.1 The Mixed Logit Model & WTP Space

The destination choice models in this paper use a mixed logit RUM framework in WTP space, following Scarpa et al. (2008). In the context of this model of birders' choices among publicly accessible birding "hot spots," birder i derives utility U_{jt}^{i} from a birdwatching trip to site j (among J alternatives) on choice occasion t within the sample time period.[4] We are using diary data which yields an unbalanced panel dataset, so the total number of trips, T^{i},

[4]Per eBird, a "hotspot" is a good public birding location that has been identified by eBird members and by which aggregate results about birds is available through eBird. See `http://help.ebird.org/customer/portal/articles/1900690-all-about-hotspots?b_id=1928` for additional information

varies across individuals. The birder's utility function has a systematic component, V^i_{jt}, that depends (linearly, for convenience) on income net of the cost of travel to that site: $(Y^i - C^i_{jt})$. The marginal utility of net income (i.e. the marginal utility of other consumption) is given by the coefficient α. Utility depends on expected bird biodiversity at the destination, by species categories, denoted as the vector X^i_{jt}. Utility also depends on other observable site attributes (included in vector A_{jt}). We allow for preferences for bird biodiversity to vary systematically with the seasons and to trend across calendar years, with these time indicators included as the vector T_t. Preferences for bird biodiversity are also allowed to vary systematically with observable sociodemographic characteristics. These birder-specific characteristics are included in vector Z^i.

The utility function is thus specified as:

$$U^i_{jt} = \alpha^i(Y^i - C^i_{jt}) + (\theta^i_0 + \theta_1 T_t + \theta_2 Z^i)' X^i_{jt} + \theta_3' A_{jt} + \epsilon^i_{jt} \tag{1}$$

where the marginal utility of net income, α^i, and the vector of marginal utilities of species richness by category of species, θ^i_0, are allowed to vary randomly over eBirders, but we assume that the vectors of parameters θ_1, θ_2 and θ_3 are fixed. The error term ϵ^i_{jt} is assumed to be approximated by a Gumbel distribution.

We can rewrite equation (1) by redefining the coefficients to account for the transformation and this will allow us to then estimate the model directly in WTP-space (Train and Weeks, 2005). The utility model must be reparameterized so that the distribution of WTP is directly estimated by the model, and thus already divided through by the price parameter, $wtp^R_i = -\theta^i_0/\alpha_i$ and $wtp^F = -\theta/\alpha_i$ (for $\theta = \theta_1, \theta_2, \theta_3$) for the random and fixed parameters respectively. By allowing α^i to be a random parameter this allow for scale heterogeneity in the model, but since not all parameters in the model are not random and correlated, it does does not all for all sources of correlation (Hess and Train, 2017). We assume the negative of the "price" coefficient, α_i (the marginal utility of net income) has a log-normal distribution to force the random travel cost coefficient to be bounded away from zero (Hess and Train, 2017).

$$U^i_{jt} = \alpha^i(Y^i - C^i_{jt}) + \alpha_i((\frac{\theta^i_0}{\alpha_i}) + (\frac{\theta_1}{\alpha_i})T_t + (\frac{\theta_2}{\alpha_i})Z^i)'X^i_{jt} + \alpha^i(\frac{\theta_3}{\alpha^i})'A_{jt} + \epsilon^i_{jt} \tag{2}$$

This can be rewritten as:

$$U^i_{jt} = \alpha^i(Y^i - C^i_{jt}) + \alpha_i(wtp^i_0 + wtp_1 T_t + wtp_2 Z^i)'X^i_{jt} + \alpha^i(wtp_3)'A_{jt} + \epsilon^i_{jt} \tag{3}$$

Note this conversion into "utility in WTP space" only works when assuming that net income (income minus travel costs) enter linearly and additively separably into the utility function.

The choice probabilities for a mixed logit are based on birder i choosing alternative j on choice occasion t. To simplify the representation of equation (2), in fashion similar to Thiene and Scarpa (2009), let utility be written as $U^i_{jt} = V^i_{jt}(\omega^i, \kappa)$ where $\omega^i = (\alpha^i, \theta^i_0)$, the random parameters within the model and $\kappa = (\theta_1, \theta_2, \theta_3)$, the fixed parameters in the model. Then following the usual intuition behind discrete choice models that birder i chooses destination j in period t because $U^i_{jt} > U^i_{kt}$, $\forall\ j \neq k$. Let y_{it} denote the birder's selection on choice occasion t over their Y^i choice occasions and $y^i = (y^i_1, y^i_2, \ldots, y^i_{Y^i})$ be their sequence of choices. The conventional logit probabilities evaluated at parameters (ω^i, κ) for birder i is:

$$L^i_{jt} = \sum_{t=1}^{t=T^i} \frac{e^{V^i_{jt}(\omega^i,\kappa)}}{\sum_j e^{V^i_{jt}(\omega^i,\kappa)}} \tag{4}$$

On any given choice occasion, then, the mixed logit choice probabilities are given by the integral of $L(y^i|\omega^i)$ over the distribution of the random parameters ω^i:

$$P^i_{jt}(y^i) = \int L^i_{jt}(y^i|\omega^i, \kappa) f(\omega^i|\kappa) d\omega^i \tag{5}$$

where $f(\omega^i|\kappa)$ in equation (5) is the density function for the random parameters ω^i, given the fixed parameters κ.

3.2 Sample Selection Corrections

This prototype analysis employs a sample of convenience and thus we need to control for selection into the sample: responding to the survey and taking a trip they reported to eBird. For this analysis, travel costs are calculated only for eBird members who responded to the auxiliary survey so we can geocode their home address (their trip origin). This restriction limits us to approximately 10 percent of total eBird membership, so some effort to assess and correct for any systematic sample selection is required.[5] We use the sample selection method employed in Cameron and Kolstoe (2019) which builds on the sample selection method employed in Kolstoe and Cameron (2017), as well as look at the sample selection stemming from the choice to report a trip to eBird by using both demeaned propensities to be in the estimating sample as well as engagement-intensity weights.[6] The auxiliary data on the general population came from two survey questions we asked as part of Qualtrics Omnibus survey in April and May of 2018. Our two questions, as shown to the respondents are reproduced in Figures 4a and 4b. We used these data to construct weights that reflect the relative frequency and scope, in the general population, of engagement with eBird (as

[5]There are 7,469 user who submitted reports to eBird for the 2013-2015 time frame in the states of Oregon and Washington.

[6]In the presence of limited data, the best recourse was to model who selected to report home address information versus who did not. This method is based on an ad hoc method proposed by Cameron and DeShazo (2013) that employs sociodemographic data at the county level for the "likely" home address of each birder, based on the spatial center of gravity of their trips

well as other major citizen science projects, though these others are not used in this paper) related to wild birds, and people's corresponding levels of participation in bird watching. These are engagement-intensity weights for the eBird sample. Greater detail about how these weights are constructed can be found in Cameron and Kolstoe (2019).

3.2.1 De-meaned response propensity

By estimating the engagement propensity, the estimated propensity can be demeaned relative to the average engagement propensity for the general population of interest to be sampled. In a choice model, these can be interacted with the key parameters of interest so that the baseline parameter than reflects the parameter for the average engagement propensity in the general population of interest. Plus, after the model has been estimated, the demeaned response propensity can be counterfactually set to zero.

3.2.2 Intensity Weights

We use the intensity weights described in Cameron and Kolstoe (2019) to capture both the extensive and intensive margin of how people participate with eBird. We choose to use intensity weights rather than the conventional response/nonresponse correction because we understand that our sample of eBird members, the respondents to our 2015 survey, may not be representative of the distribution of engagement propensities in the general population of the U.S. Qualtrics qBus respondents were asked to indicate their level of engagement with the eBird project in six mutually exclusive ordered categories: whether they are (1) unfamiliar with the project; (2) have heard of the project but are not a member; are a member but (3) report their bird sightings only rarely, (4) report sighting less than half of the time, (5) report more than half the time, or whether they (6) report virtually all of their sightings. We estimate an ordered probit model using the qBus sample and respondent's engagement with eBird, for all six possible bins and calculate a set of six fitted probabilities for each bin, conditional on the Z_i vector for each individual. Then we assign weights to each response in the eBird dataset that serve to scale the fitted probability of an individual being in their observed engagement-intensity bin, but allow multiple factors to affect expected levels of engagement intensity for each eBird respondent. We weight each observation in the eBird sample so that the weights sum to the sample size for the eBird sample. For additional detail about how these are constructed, see Cameron and Kolstoe (2019).

3.3 Variable Selection - LASSO & Elastic Net

In the era of "big data" new tools have been developed to help deal with modeling issues arising from the high-dimensional problem when there are many potential control covariates (Belloni et al., 2014) Chernozhukov et al. (2015). For example, the least absolute shrinkage and selection operator (LASSO) developed by Tibshirani (1996) is now used by researchers to select the set of control variables as it seeks to find a vector of coefficient estimates to minimize the value of λ (the stop value). Additional LASSO methods have proposed to include cross-validation, double-selection and adaptive lasso which all seek to address weaknesses of LASSO such as small mistakes about which covariates to include and exclude from the model. Zou and Hastie (2005) proposed elastic net is an extension of LASSO with a penalty term derived from a combination of the absolute-value penalty and the squared penalty, used by LASSO and ridge regression respectively. Elastic net may be a preferable method to use if the proposed variables may be highly correlated such as in the case of preferences about birds.

3.4 Estimating TWTP

Estimating the model "in WTP space" yields direct estimates of the mean marginal WTP (MWTP) associated with each destination attribute, which can then be used to calculate the mean total willingness to pay (TWTP) for a single trip to a birding destination with specified attributes, in a specific season and year, for a birder with specific characteristics. If we want further to incorporate the distributions of the coefficients which were allowed to be random across the population of birders, a subsequent step is necessary. Assume that $\gamma \sim N(b, \Omega)$ represents the vector of coefficients on the expected species richness variables by category. These coefficients are specifed as being random and correlated across the population of birders. Then we can use the vector of means for these parameters and the symmetric matrix of parameter covariances and make draws from this distribution, based on a vector of independent standard normal deviates, z, using $\gamma = b + Lz$ where L is the lower-triangular Cholesky factorization of Ω (such that $LL' = \Omega$). Including the seasonal and individual-characteristic systematic shifters, the MWTP for species richness for each category of bird species, c, now becomes:

$$MWTP_c = \gamma_c + \beta_{1c} T_t + \beta_{2c} Z^i \tag{6}$$

Thus the TWTP for one trip to a given site is:

$$TWTP = \left[\sum_{c=1}^{C} \left(\gamma_c + \beta_{1c} T_t + \beta_{2c} Z^i \right) \right]' X_{jt} + \beta_3' A_{jt} \tag{7}$$

4 Data

Our recreational site-choice data for birdwatchers are drawn from eBird, an extensive citizen science project managed by the Cornell Ornithology Lab and the National Audubon Society and funded since 2002 through several grants from the National Science Foundation.[7] The eBird dataset contains information contributed by birdwatchers who are project "members."

[7] The eBird project is online at `www.ebird.org`.

Early membership was relatively low, expanded greatly since 2009, especially with the availability of the smart-phone app and quality controls for the contributed data on bird sightings have been implemented. The available information from eBird includes the trip entries of individual birdwatchers, where it is possible to connect the trip origin (i.e. the member's home address from their member profile) and the recorded destination for each trip. We use trips taken by Oregon and Washington eBirders to destinations in Oregon and Washington states to build on the work in Kolstoe and Cameron (2017).[8] Summary statistics concerning site and trip attributes, and the characteristics of eBird members taking these trips, are provided in Tables 1 and 2.

4.1 eBird Individual Data and the Auxiliary Supplementary Survey Data

The individual characteristics in the analysis reported in Kolstoe and Cameron (2017) were limited to the profile information eBird collected from its members when they registered for the project. To collect additional information deemed valuable for demand modeling, we conducted an auxiliary survey of eBirders in the states of Oregon and Washington during the winter of 2015-2016. This auxiliary survey included standard demographic questions (to permit matching to the relative population frequencies from the American Community Survey of the U.S. Census). Also included were questions about eBirders' recreational behavior (similar to those asked by the U.S. Fish and Wildlife as part of their quinquennial survey), and questions about their personal birding and trip preferences (the survey instrument is

[8] Boundary-related problems are also minimized for this pair of states, because the Canadian border serves to limit trips to the north, while the Pacific Ocean constrains trips to the west. The relatively low populations along the eastern boundaries of these states, or along the Oregon-California border, likewise minimize the truncation of trips. This is one reason we combine the two states. The Portland, Oregon and Vancouver, Washington metropolitan area spans the border between Washington and Oregon, which would complicate any analysis of either of these two states considered separately.

available upon request from the authors).[9] This analysis takes advantage of about nine times as much trip data as our earlier study, and augments the simple eBirder membership profile data with additional information (see Kolstoe and Cameron (2017)).[10] Summary statistics about the respondents are provided in Table A3.

Travel costs. Distances and travel times for our study are calculated for the "best route" based on the algorithm coded as ORSMtime.ado in STATA and ggmap in R.[11] We do not model reported bird sightings that involve a travel distance of less than one mile, so utility from backyard birds or other very local bird populations does not enter into our analysis.[12] Thus we have no revealed-preference measures of WTP for backyard birds, even though such sightings undoubtedly contribute substantially to the aggregate net social welfare associated with avian biodiversity.

The opportunity cost of time is always an important consideration in the construction of the travel cost variable for a site-choice model. The basic eBird data do not include individual-specific income or wage information. However, we collected data on household income brackets in our auxiliary survey. We use the midpoint of each income bracket as an approximation for household income. We convert this annual income into an approximate hourly wage. We then count the value of travel time at one-third of this wage, a common approximation in the literature.[13]

[9] The survey had 1,277 respondents, of which 974 completed income questions and 531 of which reported bird sightings from a birding trip more than one mile away from home during the sample period of 2013-2015. Not all of these trips satisfy the necessary inclusion criteria for the estimating sample. These criteria pare down the sample to 321.

[10] The eBird data collects limited information from its members when they first sign-up for the project (e.g. home address, age-range, education, gender) and thus the additional data available on each member is sparse).

[11] The OSRMtime.ado relies on OpenStreet map to query "best route" data. As a small handful of routes were not found using OSRMtime, ggmap in R which uses Google Maps API as the back end was used to fill-in the missing routes.

[12] This metric conforms with how the U.S. Fish and Wildlife define a trip away from home Carver (2013). We do have a stated preference survey underway that will allow us to get at the value of backyard birding in future work.

[13] Larson and Lew (2014) and Fezzi et al. (2014) test this assumption allowing for a noisy-wage fraction and find that this assumption is reasonable.

4.2 Expected Number of Bird Species

The eBird data is rich with reports of actual bird sightings on each trip. However, one must be careful to build a measure of *ex ante* expected bird sighting to avoid endogeneity problems associated with trip-specific observations. Destinations for each trip are chosen in advance of knowing which species will actually be seen. Each eBird trip record includes information about which bird species and how many of each are observed during each outing. Not all sites are equally visited, or visited in every year, so to fill in the gaps in the eBird data, we take advantage of a second external data set, this one from BirdLife International, via Ridgely et al. (2011). This dataset allows us to incorporate seasonal variations in species richness into our calculations of *ex ante* "expected sightings." The BirdLife dataset provides geographic references for bird ranges, their presence (i.e. their likelihood of being seen), their origin (e.g., native or introduced) and seasonality (e.g., resident, breeding, non-breeding or passage). The BirdLife data are particularly important when no eBird visits are recorded in the same month of the prior year for a particular birding hot-spot destination.[14] Our RUM models require a conformable set of attributes for all sites that comprise an individual's potential choice set, even when no eBird member visited that site during the time period before which forms the basis of the construction of the expected species measure. We also break out our "expected sighting" attribute into "expected sighting by type" for four categories of bird species—water birds, birds of prey, perching birds and game birds.[15]

[14]In the future, we plan to alter the expected species variable to be constructed as a moving average of sightings in the area to better reflect how quickly eBirders can access the data submitted by other members (once it has been verified).

[15]We distinguish these four types based on the way that birds are grouped for the Pocket Naturalist Guide. Species not featured in *Oregon Birds: An Introduction to Familiar Species* are grouped using class and genus according to the Clements Checklist, version 6.8.

4.3 Other destination attributes.

Destination attributes that vary across sites can be included as variables that shift the overall *level* of total willingness to pay for a trip to a specific destination, independent of the specific levels of biodiversity in categories of bird species at that destination. These attributes include indicators for site management regimes that may be related to the overall level of biodiversity, indicators for the expected presence of an endangered bird species (state and federal listings), an indicator for whether the site lies in an urban area, land cover at a destination and indicators for the type of ecosystem at the destination. We use data from the U.S. Geological Survey—namely the Protected Area Database of the U.S.[16] These data categorize the ecological management regime for each hotspot location, such as who owns the land, who manages the land, whether and how it is managed for biodiversity, and the spatial extent of the protected area. These datasets were created to facilitate landscape analyses. Our omitted GAP category is "no known ecological management of the site."[17] We use data from the U.S. Census to indicate whether a site is within an urban area (based on the 2010 U.S. Census). We include data from the USGS's 2011 National Land Cover Data which gives us land cover information on a 30-meter by 30-meter resolution level. We also use the U.S. Environmental Protection Agency (EPA) ecoregion dataset (level III) to capture broader abiotic and biotic differences across the various hotspots (for twelve regions within the two-state area for this study).

4.4 Collective prior behavior

Our current "expected congestion/popularity" site attribute is based on the share of total eBird member visits to the site in question, in the same month of the previous year. In

[16] Specifically we use the USGS Gap Analysis Program (GAP) Version 1.4).

[17] GAP Status Code Definitions can be found in the USGS GAP-PAD-US Standards and Methods for State Data Stewards, 5 May 2016)

the future, we plan to comparing the results from this version of the variable to results that we obtain when we measure congestion as the spatial densities of visits per unit area per month, rather than as shares of total eBird visits, thus exploring an approach used by Bujosa et al. (2015). Regardless of whether the variable is constructed as a share of total visits, or as an average monthly density of visitors per unit of site area, this congestion measure is observationally equivalent to a measure of the popularity of the site.[18] The intent of a congestion/popularity measure is to serve as a proxy for the number of other visitors the individual might expect to encounter during a visit to a particular birding hotspot. Congestion/popularity may be undesirable from the individual's perspective if it implies crowding for a rivalrous quasi-public good. But some degree of congestion/popularity may be desirable if the environmental good in question is essentially non-rival and there are pleasant social interactions among like-minded participants at most destinations.[19]

5 Results - Preliminary and Do Not Cite

The main results are featured in Tables 3 and 4. The results shown are based on 8,800 trips taken by 309 birders who responded to the auxiliary survey. Our preferred model is featured in column 4 in Table 4 (full results in Table A2 in the Appendix).

Consideration Sets We limit eBirders choice set to all possible sites within 60-miles of their home address. Due to computational limits, we are currently only using 10 percent of the non-chosen alternatives from the full set of alternatives; the number of sites in a consideration set ranges from 2 to 89, with the average being 44. Note the results presented here are preliminary. Early versions of this paper used the population weights constructed with the

[18] A measure of relative congestion, consisting of visits to a given site as a share of all visits to all sites, has been used previously in the literature (Murdock (2006); Phaneuf et al. (2009); Timmins and Murdock (2007)).

[19] While our congestion/popularity measure is not completely exogenous, it is at least predetermined. Most other studies that attempt to control for expected congestion/popularity do not enjoy our luxury of *prior* years of trip data for each individual.

2016 U.S. FWS data and lasso was not decide which variables to include as additional controls and are available in the appendix. Feedback from audience members about the large number of control variables and collinearity between the controls led us to use LASSO and elastic net to see what additional interaction terms to include in the model to capture observed heterogeneity.

5.1 Mixed Logit in WTP Space

The main results are featured in Tables 3 and 4. The full results are in the appendix in Tables A1 and A2. The key parameters of interests, the expected number of bird species by type, are shown in both the tables and are estimated under different models. The first set of results are estimated using a conditional logit and the second set of results are estimated using a mixed logit in WTP-space. These results estimated using the conditional logit shows how the parameters estimates are affected by using different weights to correct for the sample of convenience problem. The results estimated using the mixed logit in WTP-space show how the parameter estimates and thus WTP estimates change as you included different interactions with the key parameters of interest, the expected number of bird species by type. These results are based on 8,647 trips taken by 306 birders who responded to the auxiliary survey. Earlier results using exogenous weights from the 2016 U.S. FWS data are featured in the appendix.

Consideration Sets We limit eBirders choice set to all possible sites within 60-miles of their home address. Due to computational limits, we are currently only using 10 percent of the non-chosen alternatives from the full set of alternatives; the number of sites in a consideration set ranges from 2 to 89, with the average being 44. Note the results presented here are preliminary. Early versions of this paper used the population weights constructed with the 2016 U.S. FWS data and lasso was not decide which variables to include as additional controls and are available in the appendix. Feedback from audience members about the large number of control variables and collinearity between the controls led us to use LASSO and its related tools to select for which control variables to include in the model to reduce the dimensionality and risk of having overparameterized the model.

Marginal Utilities: Net Income Across all models in Tables 3 and 4 the marginal utility of other consumption (the negative α coefficient on the travel cost variable, linear in this specification) is statistically significant and negative, the expected sign. The parameter is robust across all of our specifications.

Marginal Utilities of Expected Number by Type In Tables 3 and 4, the main variable of interest in the model is β_1^i and X_{jt} which contain a vector of expected species by type (water birds, birds of prey, perching birds and game birds). Our preferred specification is shown in column 4 in Table 4 as we also capture seasonality effects by including time fixed effects. In the future we will explore how to incorporate time fixed effects to reduce the number of additional parameters to include in the model as controls. We find that the marginal utility of expected species by type is heterogeneous. Part of the heterogeneity is captured through interaction terms with observable characteristics, temporal fixed effects, and random parameters which are allowed to be correlated (the results for the latter two are shown in Table A2). The baseline parameters for the expected number of water birds, birds of prey, and game birds are each respectively positive and significant, except perching birds which is negative, but statistically insignificant in all four of the models shown in Table 4. whereas the parameter for the expected number of perching birds is negative and statistically significant across all four specifications.

Systematic Variation in the Marginal Utilities of Expected Number by Type: Water Birds The sources of observable heterogeneity in the margin utility of an additional water bird species does not come from differences in income (not statistically significant as shown in Table A2, but does comes from seasonal differences. The temporal differences in marginal utility for an additional water bird species is higher for August and the fall months of September and October, but negative for the month of April and the linear time trend.

Systematic Variation in the Marginal Utilities of Expected Number by Type: Birds of Prey The sources of observable heterogeneity in the margin utility of an additional birds of prey species does not come from differences in income (not statistically significant as shown in Table A2, but does comes from seasonal differences. The results suggest that birders have a lower marginal utility and thereby WTP of seeing an additional bird of prey in July and August. The linear time trend is positive and statistically significant.

Systematic Variation in the Marginal Utilities of Expected Number by Type: Perching Birds The MWTP for an additional perching bird species appears to be zero based on the lack of statistical significance. This may be tied to the fact people tend to see these types of birds in and around their home.

Systematic Variation in the Marginal Utilities of Expected Number by Type: Game Birds The sources of observable heterogeneity in the margin utility of an additional game bird species does not come from differences in income (not statistically significant as shown in Table A2, but does comes from seasonal differences. The temporal differences in marginal utility for an additional game bird species is lower in July and December.

Other Site Attributes

A number of other site attributes bear statistically significant marginal-utility coefficients in Table A2. These include indicators for different management regimes for biodiversity, ecoregions, land cover classes, urban hotspots, an indicator for expected presence of an endangered bird species along with our continuous "congestion/popularity" measure for each hotspot.As in Kolstoe and Cameron (2017) and Kolstoe et al. (2018), the estimated marginal utility from visits to National Wildlife Refuges which are specifically managed for bird biodiversity yields the highest to eBirders. The marginal WTP for a visit to a National Wildlife Refuge is statistically significant, implying there is a higher value (utility) for a birder for specifically visiting a site managed for bird biodiversity, above and beyond managing a site managed for biodiversity (this parameter is additive to the parameter for National Parks, etc.). This may not be surprising in light of when individuals are traveling to see birds, they are primarily traveling to see waterfowl which can be found at National Wildlife Refuges (Wilson, 2010). The marginal utility of a visit to a site such as National Parks (e.g. a site with GAP 1 or 2 status) or National Forests (e.g. a site with GAP 3 status) sites are positive, but are not statistically significant.

We also find a persistently negative coefficient on the indicator for an urban destination (activated for hotspots which lie within an Urban Area as defined by the 2010 Census), as was the case in Kolstoe and Cameron (2017). This suggests there are latent attributes of sites in urban areas that confers disutility to non-urban sites, independent of species richness.

To control destination ecoregions and land cover, to pre-empt any omitted variable bias from these two sources in the estimated coefficients on the expected number of bird species. These two sets of controls account for any utility that is accounted for directly from either the type of ecoregion or land cover at a site. The baseline ecoregion for the trips is the Puget Lowland. The indicators for the Coast Range, Columbia Plateau, North Rockies, Willamette Valley, North Cascades, Klamath Mountains, and the Northern Basin Rage are positive and statistically significant, indicating a higher value (utility) than to a site in the Puget Lowland. However the indicators for the Blue Mountains is negative and statistically significant, whereas the indicator for the Cascades is negative and statistically insignificant. The land cover class indicators for barren land, forest, planted and wetlands are positive and statistically significant whereas the indicators for water, shrub/scrub and herbaceous land cover are not statistically significant. As in Kolstoe and Cameron (2017) the congestion/popularity variable is proxied using all eBird members' reported visits to *this* hotspot in the same month last year, as a proportion of all eBird members' visits to *any* hotspot during that same time frame. The parameters suggest that the marginal WTP of anticipated congestion/popularity is a good thing, and only is in the range where it diminishes for one-percent of the site alternatives. Only a few sites in our sample receive this share of eBird visits in the previous year, thus over most of the range of the data, this variable is positive.[20]

6 Discussion

The results suggest there is significant preference heterogeneity for birds, and most notably by bird species type by eBirders. While we would have like to explore the heterogeneity across groups further with the latent class models for the models by species by type, it is evident that controlling for differences in income and lister status are critical. The LASSO operator predicts it is important to control for differences in income, but not other controls. Additional LASSO operators beyond the cross-validation LASSO operator will be used in the future to see as to what control variables to include in the model to capture preference heterogeneity. The results suggest there is significant preference heterogeneity for birds, and most notably by bird species type by eBirders, and more generally based on the results with

7 Still To Do

- Explore how to capture seasonality with a minimal number of fixed effects (currently includes 48 additional variables)
- Explore what variables LASSO and elastic net select when you require the seasonal fixed effects to be included.

[20]Linearity in this variable is strongly rejected by a likelihood ratio test.

- Construct the TWTP estimates by illustrative site characteristics for preferred specification

8 Caveats and Directions for Future Research

Citizen science data offers researchers the possibility of working with rich data sets, however, the caveat remains that these are samples of conveniences. Our work here is to provide a flexible model to estimate the value of bird biodiversity and best approximate the general population of birders' value of different bird species types. To date this work has focused solely on birders who travel away from home (more than a mile) to see birds. This does not address the question of the value of bird biodiversity to birders who see them around their home. Per the U.S. FWS, most birders see birds around their homes, and the vast majority of these individuals, feed the wild birds around their home (U.S. FWS Report, 2018). In our other work, we are seeking to address this issue.

9 Tables

Table 1: Descriptive Statistics across All Alternatives, Featured Variables, Oregon and Washington States[a]

Variable	Mean	Std. Dev	Min	Max
Roundtrip, 1/3 wage	27.74	27.92764	1.555612	166.8201
Bird species type				
Water Birds (WT)	12.10	2.10	1	20.84
Birds of Prey (PR)	13.52	2.16	1	20.5
Perching Birds (PH)	24.43	3.64	1	35.06
Game Birds (GM)	12.37	4.85	0	20.63
Destination type				
1(National Wildlife Refuge)	0.100			
1(National Parks, etc.)	0.198			
1(National Forests, etc.)	0.408			
1(Urban Area)	0.728			
†$Congestion/Popularity_{jt}$	.0017	.003	0	.0227142
Seasonal and trend variables				
$1(January)$	.102			
$1(February)$	.092			
$1(March)$	.091			
$1(April)$	.106			
$1(May)$	.101			
$1(June)$	.072			
$1(July)$	.069			
$1(August)$	.068			
$1(September)$	.072			
$1(October)$	.0813			
$1(November)$	.085			
$1(December)$	.062			
$t13$ ($t13 = year - 2013$)	1.05			
Ecoregion indicators				
$1(Coast\ Range)_j$	0.069			
$1(Columbia\ Plateau)_j$	0.056			
$1(Blue\ Mountains)_j$	0.041			
$1(North\ Rockies)_j$	0.033			
$1(PugetLowlands)_j$	0.465			
$1(Willamette\ Valley)_j$	0.218			
$1(Cascades)_j$	0.0058			
$1(North\ Cascades)_j$	0.010			
$1(Klamath\ Mtns,\ Coast\ Range)_j$	0.043			
$1(North\ Basin)_j$	0.0005			
Land cover indicators				
$1(LC\ Developed)$	0.37			
$1(LC\ Water/Perennial\ Snow\ \&\ Ice)$	0.051			
$1(LC\ Barren\ Land)$	0.080			
$1(LC\ Forest)$	0.210			
$1(LC\ Shrub/Scrub)$	0.061			
$1(LC\ Herbaceous)$	0.011			
$1(LC\ Planted)$	0.084			
$1(LC\ Wetlands)$	0.120			

NOTES: [a]60-miles maximum travel distance for considered alternative hotspots, 2013-2015 trips; statistics for other control variables used in our models have been relegated to the Appendix.
† Share of all eBird trips, same month, last year, to site j

Table 2: Descriptive Statistics across All Alternatives,
Featured Birder Characteristics, Oregon and Washington States

eBirders in Sample: Sociodemographics	Mean	Min	Max
Females	.517	-	-
Listers	.561	-	-
Avg. Age	57.26	14	78
Diff. from Sample Avg. Age	-0.55	-44.38	19.62
Working	.433	-	-
Avg. Income (in $10,000)	9.484	1.8	22.5
Diff. from Sample Avg. Income (in $10,000)	1.80	-6.81	13.89

Table 3: Comparison of Weights - Elastic Net Specification, Key Parameters

	(1) Fitted Propensity	(2) FWS weights	(3) Intensity weights
Roundtrip TC	-0.0630***	-0.0656***	-0.0649***
	(0.00106)	(0.00161)	(0.00152)
E(# water bird species) (WT)	0.00717	0.0280*	-0.0176
	(0.0149)	(0.0158)	(0.0154)
E(# birds of prey species) (PR)	0.0607***	0.0661***	0.0907***
	(0.0157)	(0.0158)	(0.0168)
E(# perching bird species) (PH)	-0.0301***	-0.0335***	-0.0228***
	(0.00840)	(0.00824)	(0.00774)
E(# game bird species) (GM)	0.00146	-0.00964	-0.0106**
	(0.00575)	(0.00602)	(0.00539)
Ecoregions & NLCD?	Yes	Yes	Yes
Observations	298,766	298,231	293,263
Log Likelihood	-19158.63	-344848.34	-26363.68
AIC	38389.27	689768.68	52799.35

Note: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Estimated in STATA 16 using CLOGIT.

Table 4: Comparison of Specification, Key Parameters with Intensity Weights

	(1)	(2)	(3)	(4)
	No Interactions	Income	Elastic Net	Income + Time
Roundtrip TC	-2.325***	-2.331***	-2.314***	-2.327***
	(0.0701)	(0.0678)	(0.0945)	(0.0782)
E(# water bird species)	2.006***	1.949***	2.180**	2.756***
	(0.523)	(0.546)	(0.911)	(0.921)
E(# birds of prey species)	2.376***	2.387***	2.674*	2.275**
	(0.770)	(0.819)	(1.541)	(1.090)
E(# perching bird species)	-0.862	-0.853	-0.377	-1.081
	(0.637)	(0.531)	(0.974)	(0.850)
E(# game bird species)	2.041***	2.088***	2.269***	2.261***
	(0.489)	(0.569)	(0.668)	(0.491)
Site Controls, Ecoregion, NLCD?	Yes	Yes	Yes	Yes
Time Fixed Effects?	No	No	No	Yes
Log Likelihood	-23229.81	-23223.65	-23224.82	-23120.05
AIC	46553.63	46551.29	46551.64	46440.10

Note: N = 293,263. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Estimated in STATA 16 using MIXLOGITWTP.

10 Figures

Figure 1: Map of the home addresses of the survey respondents where dots represent home address location

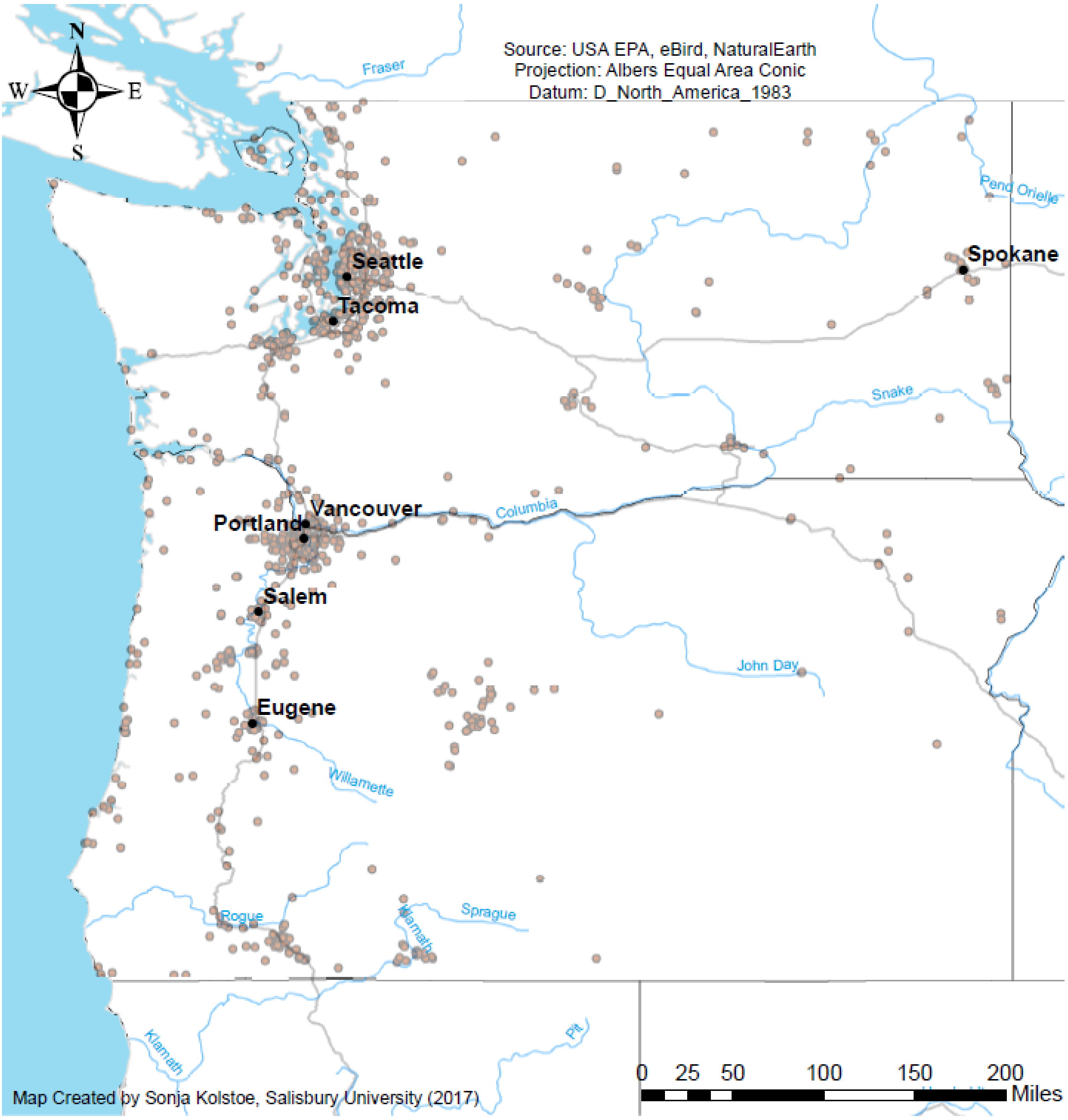


Figure 2: Map of the eBird Hotspots

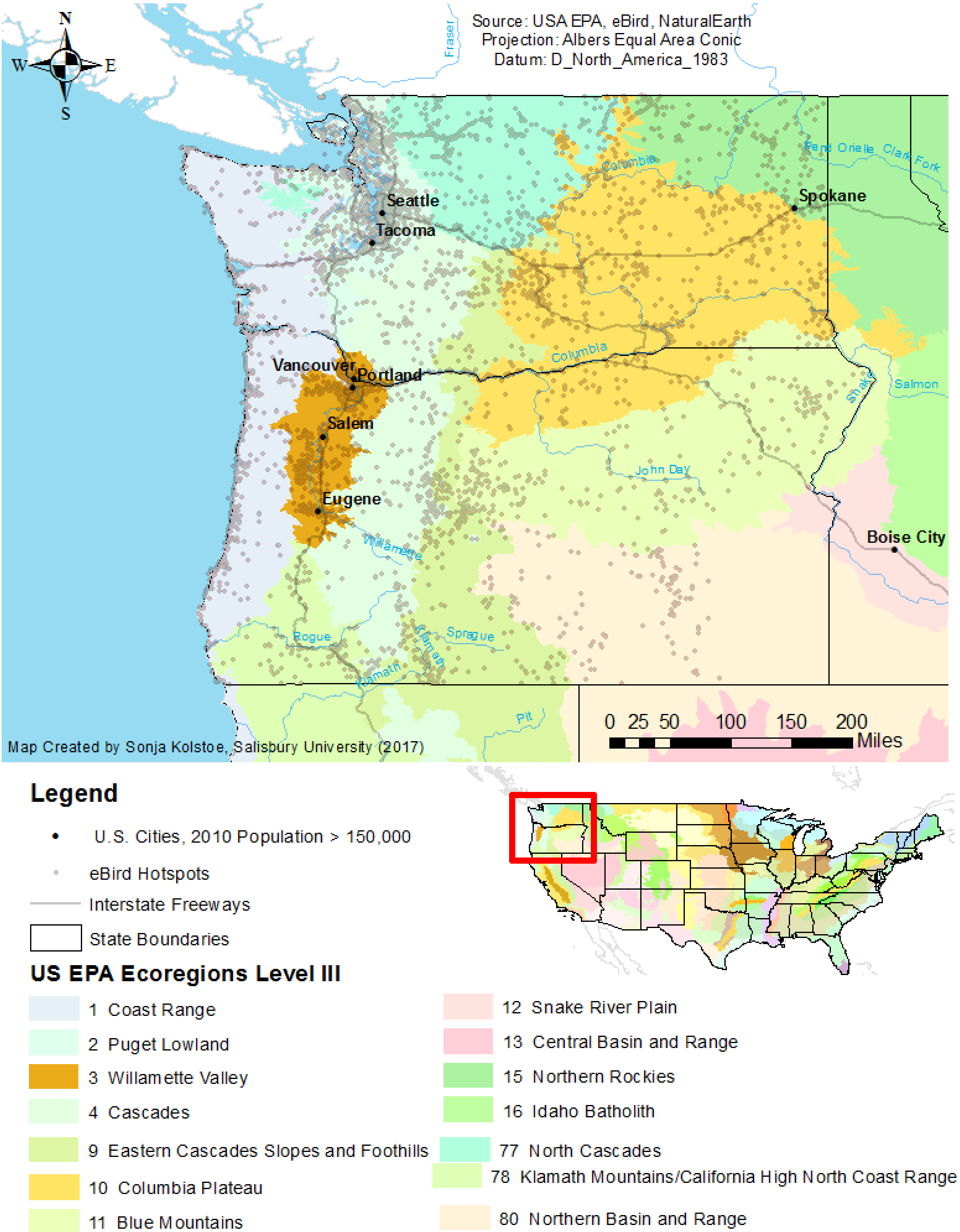
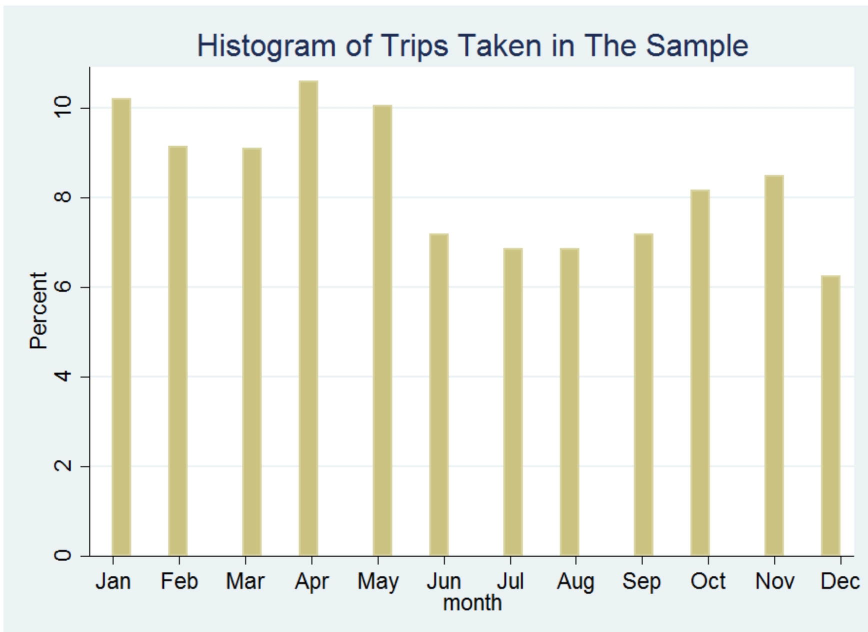


Figure 3: Histogram of Trips

(a) By Month



(b) By Year

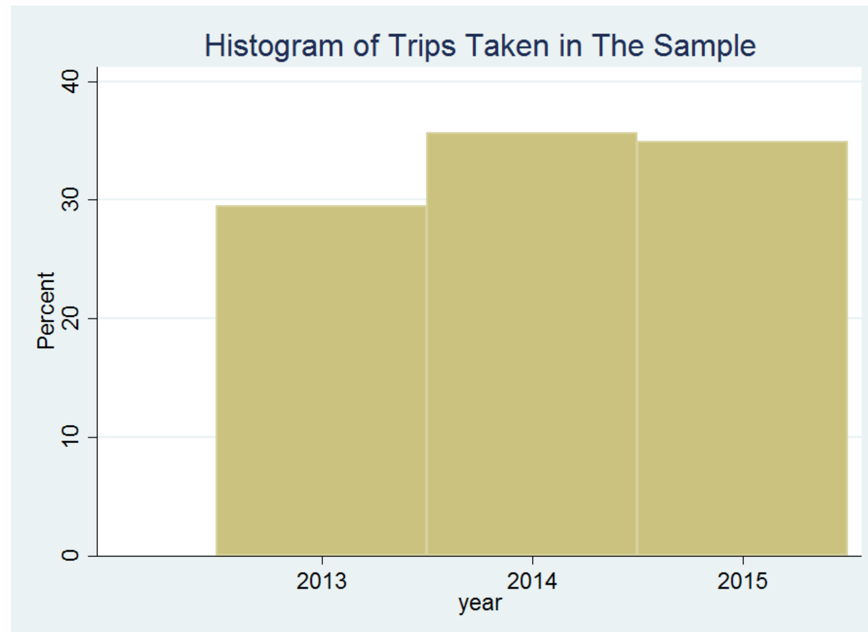
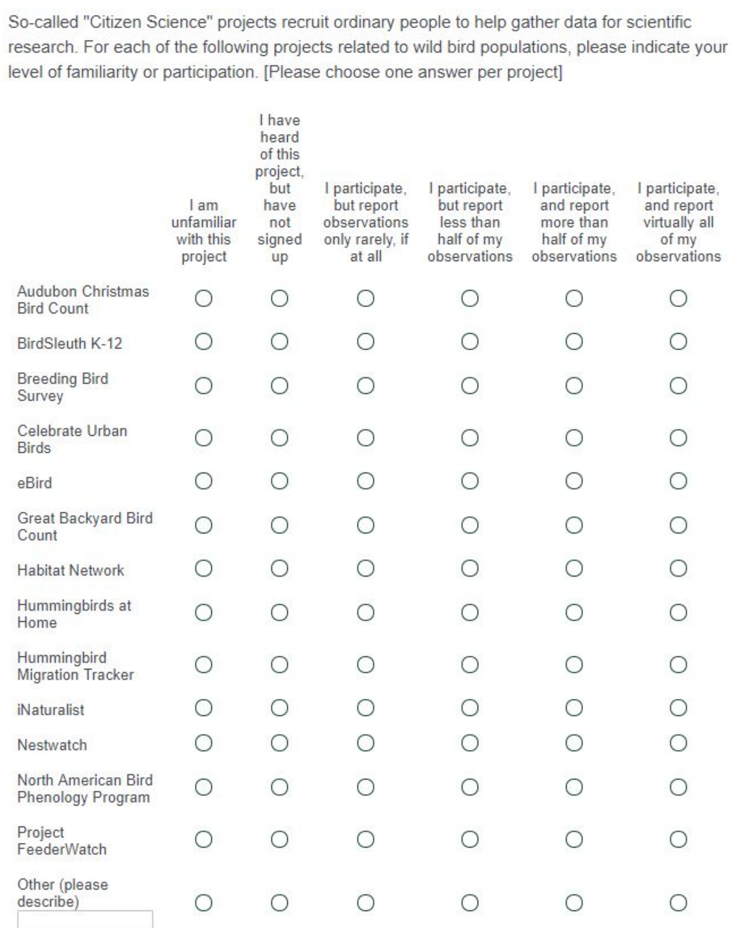


Figure 4: Qualtric Questions

(a) Bird Citizen Science Project



(b) Level of Engagement in Birding

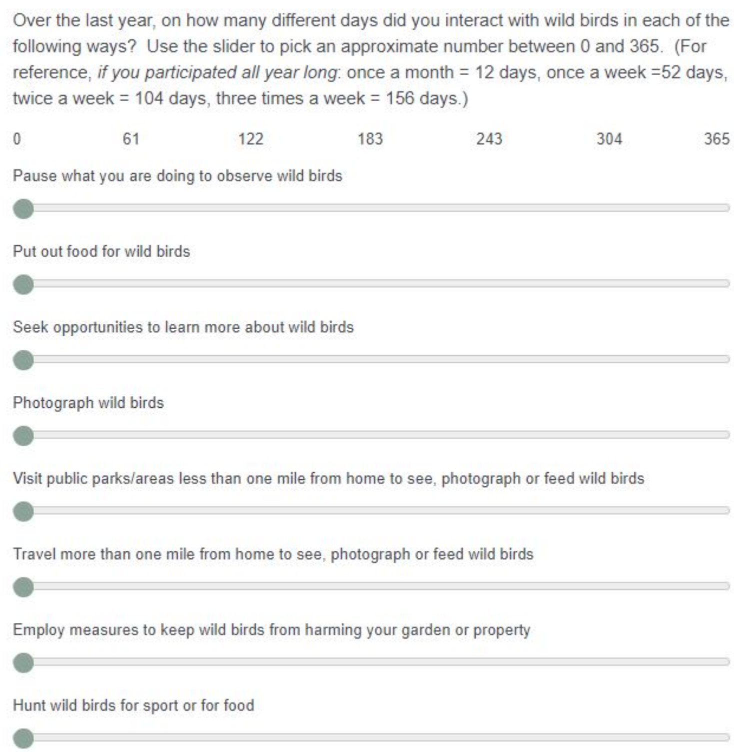


Figure 5: Coefficient Selected by Cross-Validation Elastic Net & LASSO

(a) Elastic Net

	active
TC3_IWsiteIRS_dzg_oprobit	x
TC3_IWsiteIRS	x
WT	x
WT_dzg_oprobit	x
PR	x
PR_dzg_oprobit	x
PH	x
PH_dzg_oprobit	x
GM	x
GM_dzg_oprobit	x
LC_water	x
LC_barren	x
LC_forest	x
LC_shrub	x
LC_herb	x
LC_planted	x
LC_wetlands	x
EcoReg_CoastRange	x
EcoReg_ColumbiaPlat	x
EcoReg_BlueMtns	x
EcoReg_N_Rockies	x
EcoReg_WillametteValley	x
EcoReg_Cascades	x
EcoReg_N_Cascades	x
EcoReg_KlamMtns_CA_NCoastRange	x
EcoReg_N_Basin_Range	x
PR_dev_mean_inc0000	x
PH_dev_mean_inc0000	x
GM_dev_mean_inc0000	x

Legend:
b - base level
e - empty cell
o - omitted
x - estimated

(b) LASSO

	cvlasso_type
TC3_IWsiteIRS_dzg_oprobit	x
TC3_IWsiteIRS	x
WT	x
WT_dzg_oprobit	x
PR	x
PR_dzg_oprobit	x
PH	x
PH_dzg_oprobit	x
GM	x
GM_dzg_oprobit	x
LC_water	x
LC_barren	x
LC_forest	x
LC_shrub	x
LC_herb	x
LC_planted	x
LC_wetlands	x
EcoReg_CoastRange	x
EcoReg_ColumbiaPlat	x
EcoReg_BlueMtns	x
EcoReg_N_Rockies	x
EcoReg_WillametteValley	x
EcoReg_Cascades	x
EcoReg_N_Cascades	x
EcoReg_KlamMtns_CA_NCoastRange	x
EcoReg_N_Basin_Range	x
GM_LC_forest	x
PR_dev_mean_inc0000	x
PH_dev_mean_inc0000	x
GM_dev_mean_inc0000	x

Legend:
b - base level
e - empty cell
o - omitted
x - estimated

References

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, 28(2):29–50.

Bockstael, N. E., McConnell, K. E., and Strand, I. E. (1989). A random utility model for sportfishing: some preliminary results for florida. *Marine Resource Economics*, 6(3):245–260.

Bujosa, A., Riera, A., Hicks, R. L., and McConnell, K. E. (2015). Densities rather than shares: improving the measurement of congestion in recreation demand models. *Environmental and Resource Economics*, 61(2):127–140.

Cameron, T. A. and DeShazo, J. (2013). Demand for health risk reductions. *Journal of Environmental Economics and Management*, 65(1):87–109.

Cameron, T. A. and Kolstoe, S. (2019). Hurry Up and Weight (etc.): Auxiliary Population Samples to Correct for Systematic Selection in Citizen Science Data. Working paper.

Carver, E. (2013). Birding in the United States: a demographic and economic analysis addendum to the 2011 national survey of fishing, hunting, and wildlife-associated recreation. Technical report.

Chernozhukov, V., Hansen, C., and Spindler, M. (2015). Valid post-selection and post-regularization inference: An elementary, general approach. *Annual Review of Economics*, 7(1):649–688.

Fezzi, C., Bateman, I. J., and Ferrini, S. (2014). Using revealed preferences to estimate the Value of Travel Time to recreation sites. *Journal of Environmental Economics and Management*, 67(1):58–70.

Green, W. H. (2007). *Econometric Analysis (7 th)*. Upper Saddle River, NJ: Prentice Hall.

Hess, S. and Train, K. (2017). Correlation and scale in mixed logit models. *Journal of choice modelling*, 23:1–8.

Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N., Wickham, J., and Megown, K. (2015). Completion of the 2011 national land cover database for the conterminous united states-representing a decade of land cover change information. *Photogrammetric Engineering and Remote Sensing*, 81(5):345–354.

Johnston, R. J. and Abdulrahman, A. S. (2017). Systematic non-response in discrete choice experiments: implications for the valuation of climate risk reductions. *Journal of Environmental Economics and Policy*, 6(3):246–267.

Keeler, B. L., Wood, S. A., Polasky, S., Kling, C., Filstrup, C. T., and Downing, J. A. (2015). Recreational demand for clean water: evidence from geotagged photographs by visitors to lakes. *Frontiers in Ecology and the Environment*, 13(2):76–81.

Kolstoe, S. and Cameron, T. A. (2017). The Non-market Value of Birding Sites and the Marginal Value of Additional Species: Biodiversity in a Random Utility Model of Site Choice by eBird Members. *Ecological Economics*, 137:1–12.

Kolstoe, S., Cameron, T. A., and Wilsey, C. (2018). Climate, land cover, and bird populations: Differential impacts on the future welfare of birders across the pacific northwest. *Agricultural and Resource Economics Review*, 47(2):272–310.

Langham, G. M., Schuetz, J. G., Distler, T., Soykan, C. U., and Wilsey, C. (2015). Conservation status of North American birds in the face of future climate change. *PLoS One*, 10(9):e0135350.

Larson, D. M. and Lew, D. K. (2014). The opportunity cost of travel time as a noisy wage fraction. *American Journal of Agricultural Economics*, 96(2):420–437.

McFadden, D. (1974). The measurement of urban travel demand. *Journal of Public Economics*, 3(4):303–328.

McFadden, D. and Train, K. E. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15(5):447–470.

Moeltner, K. and Von Haefen, R. (2011). Microeconometric strategies for dealing with unobservables and endogenous variables in recreation demand models. *Annual Review of Resource Economics*, 3(1):375–396.

Murdock, J. (2006). Handling unobserved site characteristics in random utility models of recreation demand. *Journal of Environmental Economics and Management*, 51(1):1–25.

Phaneuf, D. J., Carbone, J. C., and Herriges, J. A. (2009). Non-price equilibria for non-marketed goods. *Journal of Environmental Economics and Management*, 57(1):45–64.

Phaneuf, D. J. and Requate, T. (2017). *A Course in Environmental Economics: Theory, Policy, and Practice*. Cambridge University Press.

Revelt, D. and Train, K. (1998). Mixed logit with repeated choices: Households' choices of appliance efficiency level. *Review of Economics and Statistics*, 80(4):647–657.

Ridgely, R., Allnutt, T., Brooks, T., McNicol, D., Mehlman, D., Young, B., Zook, J., and International, B. (2011). *Digital Distribution Maps of the Birds of the Western Hemisphere, version 4.0*. BirdLife International, Cambridge, UK and NatureServe, Arlington, USA.

Scarpa, R., Thiene, M., and Train, K. (2008). Utility in Willingness to Pay Space: a Tool to Address Confounding Random Scale Effects in Destination Choice to the Alps. *American Journal of Agricultural Economics*, 90(4):994–1010.

Sullivan, B. L., Aycrigg, J. L., Barry, J. H., Bonney, R. E., Bruns, N., Cooper, C. B., Damoulas, T., Dhondt, A. A., Dietterich, T., Farnsworth, A., et al. (2014). The eBird Enterprise: an Integrated Approach to Development and Application of Citizen Science. *Biological Conservation*, 169:31–40.

Thiene, M. and Scarpa, R. (2009). Deriving and testing efficient estimates of wtp distributions in destination choice models. *Environmental and Resource Economics*, 44(3):379–395. Copyright - Springer Science+Business Media B.V. 2009; Document feature - ; Tables; Equations; Graphs; Last updated - 2014-08-23.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Timmins, C. and Murdock, J. (2007). A revealed preference approach to the measurement of congestion in travel cost models. *Journal of Environmental Economics and Management*, 53(2):230–249.

Train, K. (1998). Recreation demand models with taste differences over people. *Land Economics*, 74(2):230–239.

Train, K. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press, second edition.

Train, K. and Weeks, M. (2005). Discrete choice models in preference space and willingness-to-pay space. In *Applications of simulation methods in environmental and resource economics*, pages 1–16. Springer.

Venturelli, P. A., Hyder, K., and Skov, C. (2017). Angler apps as a source of recreational fisheries data: opportunities, challenges and proposed standards. *Fish and Fisheries*, 18(3):578–595.

Wilson, R. M. (2010). *Seeking refuge: Birds and landscapes of the Pacific Flyway*. University of Washington Press.

Yuan, Y., Boyle, K. J., and You, W. (2015). Sample selection, individual heterogeneity, and regional heterogeneity in valuing farmland conservation easements. *Land Economics*, 91(4):627–649.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B*, 67(2):301–320.

Online Appendix to Accompany:

Appendix: Blending Supplementary Surveys with Citizen Science Data to Estimate Representative Use Values for Non-market Environmental Goods: An eBird example

Table A1: Comparison of Weights - Elastic Net Specification, Key Parameters

	(1) Fitted Propensity	(2) FWS weights	(3) Intensity weights
Roundtrip TC	-0.0630***	-0.0656***	-0.0649***
	(0.00106)	(0.00161)	(0.00152)
E(# water bird species) (WT)	0.00717	0.0280*	-0.0176
	(0.0149)	(0.0158)	(0.0154)
E(# birds of prey species) (PR)	0.0607***	0.0661***	0.0907***
	(0.0157)	(0.0158)	(0.0168)
E(# perching bird species) (PH)	-0.0301***	-0.0335***	-0.0228***
	(0.00840)	(0.00824)	(0.00774)
E(# game bird species) (GM)	0.00146	-0.00964	-0.0106**
	(0.00575)	(0.00602)	(0.00539)
PR × demean income $10,000	0.000192	0.00165	-0.00248
	(0.00195)	(0.00218)	(0.00191)
PH × demean income $10,000	-0.00320***	-0.00417***	-0.00300***
	(0.00101)	(0.00104)	(0.000943)
GM × demean income $10,000	0.00500***	0.00665***	0.00580***
	(0.000627)	(0.000684)	(0.000641)
1(GAP status 1 or 2)	-0.0409	-0.0388	-0.105**
	(0.0446)	(0.0515)	(0.0459)
1(GAP status 3)	0.0281	0.0359	-0.0795**
	(0.0322)	(0.0351)	(0.0333)
1(GAP status 1 or 2) × 1(NWR)	0.622***	0.541***	0.737***
	(0.0634)	(0.0747)	(0.0733)
1(US Census 2010 Urbanized Area)	-0.438***	-0.441***	-0.399***
	(0.0422)	(0.0480)	(0.0450)
(mean) CongestionPMY	646.3***	658.2***	645.7***
	(12.82)	(16.09)	(15.04)
CongestionPMY2	-26380.3***	-27982.9***	-27260.0***
	(966.2)	(1366.7)	(1342.3)
1(Endangered)	-0.687	-0.0438	-0.649

Continued on next page

	(0.868)	(0.737)	(1.249)
1(Coast Range)	-0.572***	-0.489***	-0.558***
	(0.151)	(0.176)	(0.169)
1(Columbia Plateau)	-1.471***	-0.604*	-1.310***
	(0.159)	(0.330)	(0.262)
1(Blue Mountains)	-0.510***	-0.527***	-0.674***
	(0.150)	(0.146)	(0.149)
1(Northern Rockies)	-0.322	0.669*	-0.479
	(0.298)	(0.407)	(0.352)
1(Willamette Valley)	0.568***	0.549***	0.706***
	(0.177)	(0.187)	(0.187)
1(Cascades)	-0.275	-0.133	-0.261
	(0.169)	(0.178)	(0.176)
1(North Cascades)	-0.910***	-0.409	-0.794***
	(0.190)	(0.331)	(0.307)
1(Klamath Mountains/California High North Coast Range)	0.877***	1.225***	0.789***
	(0.246)	(0.277)	(0.291)
1(Northern Basin and Range)	1.221**	1.229**	1.074**
	(0.494)	(0.530)	(0.506)
1(LC water)	-0.0779	-0.146**	-0.0121
	(0.0549)	(0.0618)	(0.0571)
1(LC barren)	0.593***	0.222***	0.692***
	(0.0503)	(0.0587)	(0.0530)
1(LC forest)	0.775***	0.826***	0.906***
	(0.0386)	(0.0369)	(0.0364)
1(LC shrub)	0.335***	0.351***	0.148**
	(0.0635)	(0.0704)	(0.0693)
1(LC herbaceous)	-0.140	-0.196*	-0.162
	(0.105)	(0.110)	(0.110)
1(LC planted)	0.395***	0.469***	0.475***
	(0.0481)	(0.0575)	(0.0510)
1(LC wetlands)	0.644***	0.629***	0.608***
	(0.0475)	(0.0522)	(0.0517)
WT × Fit Propens. oprobit $j\hat{\gamma}$ for eBird-qBus mean $i\hat{\gamma}$	0.0254	-0.00632	0.0451**
	(0.0203)	(0.0222)	(0.0209)
PR × Fit Propens. oprobit $j\hat{\gamma}$ for eBird-qBus mean $i\hat{\gamma}$	0.0211	0.0333	0.00417
	(0.0212)	(0.0216)	(0.0233)
PH × Fit Propens. oprobit $j\hat{\gamma}$ for eBird-qBus mean $i\hat{\gamma}$	0.00933	0.0179	0.00636

Continued on next page

	(0.0120)	(0.0118)	(0.0117)
GM × Fit Propens. oprobit $j\hat{\gamma}$ for eBird-qBus mean $i\hat{\gamma}$	-0.0231***	-0.0438***	-0.0211***
	(0.00751)	(0.00797)	(0.00731)
Roundtrip TC × Fit Propens. oprobit $j\hat{\gamma}$ for eBird-qBus mean $i\hat{\gamma}$	0.0180***	0.0190***	0.0186***
	(0.00126)	(0.00187)	(0.00176)
Ecoregions & NLCD?	Yes	Yes	Yes
Observations	298,766	298,231	293,263
Log Likelihood	-19158.63	-344848.34	-26363.68
AIC	38389.27	689768.68	52799.35
BIC	38771.14	690150.49	53180.55

Note: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Estimated in STATA 16 using CLOGIT.

Table A2: Comparison of Specification, Key Parameters with Intensity Weights

	(1) No Interactions	(2) Income	(3) Elastic Net	(4) Income + Time
Roundtrip TC	-0.0630***	-0.0656***	-0.0649***	
	(0.00106)	(0.00161)	(0.00152)	
E(# water bird species) (WT)	0.00717	0.0280*	-0.0176	
	(0.0149)	(0.0158)	(0.0154)	
E(# birds of prey species) (PR)	0.0607***	0.0661***	0.0907***	
	(0.0157)	(0.0158)	(0.0168)	
E(# perching bird species) (PH)	-0.0301***	-0.0335***	-0.0228***	
	(0.00840)	(0.00824)	(0.00774)	
E(# game bird species) (GM)	0.00146	-0.00964	-0.0106**	
	(0.00575)	(0.00602)	(0.00539)	
WT $\times$demeaned income ($10,000)		0.0367		0.0347
		(0.0451)		(0.0501)
PR $\times$demeaned income ($10,000)		0.0415	-0.00190	0.0181
		(0.0730)	(0.0732)	(0.0733)
PH $\times$demeaned income ($10,000)		0.00113	-0.0364	0.0136
		(0.0516)	(0.0379)	(0.0567)
GM $\times$demeaned income ($10,000)		0.00787	0.00783	0.00611
		(0.0146)	(0.0237)	(0.0167)
1(GAP Status 1 or 2)	0.0753	0.0638	0.138	-0.0501
	(1.594)	(1.390)	(2.363)	(1.496)
1(GAP Status 3)	0.209	0.289	0.174	0.193
	(1.383)	(1.394)	(1.445)	(1.430)
1(GAP Status 1 or 2) $\times$ 1(NWR)	9.750***	9.831***	9.591***	10.01***
	(2.455)	(2.440)	(2.675)	(2.376)

Continued on next page

	(1) No Interactions	(2) Income	(3) Elastic Net	(4) Income + Time
1(U.S. Census Urban Area)	-5.001***	-4.985***	-5.126**	-4.930***
	(1.778)	(1.775)	(2.126)	(1.774)
CongestionPMY	5836.3***	5879.5***	5770.0***	5871.6***
	(625.6)	(622.0)	(1058.0)	(655.9)
CongestionPMY2	-242843.0***	-244051.8***	-239524.4***	-249186.5***
	(29268.6)	(29299.4)	(47820.0)	(30286.7)
1(Endangered)		1.019	-1.356	-1.154
		(8.571)	(9.785)	(9.890)
WT ×1(Feb)				0.552
				(0.624)
WT ×1(Mar)				0.538
				(0.966)
WT ×1(Apr)				-1.457**
				(0.679)
WT ×1(May)				-1.104
				(0.721)
WT ×1(Jun)				-0.911
				(0.740)
WT ×1(Jul)				0.716
				(0.667)
WT ×1(Aug)				1.552*
				(0.799)
WT ×1(Sep)				1.972*
				(1.105)
WT ×1(Oct)				1.845**
				(0.850)

Continued on next page

	(1) No Interactions	(2) Income	(3) Elastic Net	(4) Income + Time
WT $\times 1(\text{Nov})$				1.464
				(0.912)
WT $\times 1(\text{Dec})$				-0.740
				(0.855)
WT $\times$t13 (t13=0 in 2013)				-0.950**
				(0.418)
PR $\times 1(\text{Feb})$				-0.644
				(0.826)
PR $\times 1(\text{Mar})$				0.344
				(0.899)
PR $\times 1(\text{Apr})$				0.0937
				(0.848)
PR $\times 1(\text{May})$				-0.451
				(0.749)
PR $\times 1(\text{Jun})$				-0.865
				(0.703)
PR $\times 1(\text{Jul})$				-1.446*
				(0.859)
PR $\times 1(\text{Aug})$				-1.705*
				(0.905)
PR $\times 1(\text{Sep})$				-0.906
				(0.888)
PR $\times 1(\text{Oct})$				-0.429
				(1.018)
PR $\times 1(\text{Nov})$				-0.717
				(0.595)

Continued on next page

	(1) No Interactions	(2) Income	(3) Elastic Net	(4) Income + Time
PR ×1(Dec)				-0.108
				(0.702)
PR ×t13 (t13=0 in 2013)				0.667*
				(0.383)
PH ×1(Feb)				-0.548
				(0.385)
PH ×1(Mar)				0.343
				(0.568)
PH ×1(Apr)				0.243
				(0.419)
PH ×1(May)				0.528
				(0.421)
PH ×1(Jun)				0.558
				(0.532)
PH ×1(Jul)				0.548
				(0.581)
PH ×1(Aug)				0.361
				(0.648)
PH ×1(Sep)				-0.639
				(0.515)
PH ×1(Oct)				-0.119
				(0.584)
PH ×1(Nov)				-0.0364
				(0.489)
PH ×1(Dec)				0.0515
				(0.501)

Continued on next page

	(1) No Interactions	(2) Income	(3) Elastic Net	(4) Income + Time
PH $\times$t13 (t13=0 in 2013)				-0.0215
				(0.252)
GM $\times$1(Feb)				0.163
				(0.161)
GM $\times$1(Mar)				-0.151
				(0.220)
GM $\times$1(Apr)				-0.0145
				(0.230)
GM $\times$1(May)				0.247
				(0.250)
GM $\times$ 1(Jun)				-0.100
				(0.261)
GM $\times$ 1(Jul)				-0.368**
				(0.183)
GM $\times$ 1(Aug)				0.0255
				(0.195)
GM $\times$ 1(Sep)				-0.211
				(0.292)
GM $\times$ 1(Oct)				-0.230
				(0.219)
GM $\times$ 1(Nov)				-0.488
				(0.368)
GM $\times$ 1(Dec)				-0.599***
				(0.214)
GM $\times$ t13 (t13=0 in 2013)				0.0493
				(0.0936)

Continued on next page

	(1) No Interactions	(2) Income	(3) Elastic Net	(4) Income + Time
TC3 × demeaned fitted propensity	-0.299***	-0.296***	-0.306***	-0.298***
	(0.0628)	(0.0646)	(0.0740)	(0.0683)
WT × demeaned fitted propensity	0.640	0.617	0.396	0.705
	(0.558)	(0.546)	(0.916)	(0.518)
PR × demeaned fitted propensity	-1.738	-1.789*	-2.044	-1.882
	(1.133)	(0.982)	(1.997)	(1.360)
PH × demeaned fitted propensity	1.758***	1.790***	1.249	1.864***
	(0.602)	(0.547)	(0.811)	(0.694)
GM × demeaned fitted propensity	-1.671***	-1.737***	-1.813***	-1.763***
	(0.406)	(0.534)	(0.493)	(0.461)
1(Eco Reg Coast Range)	12.57**	12.45***	12.72***	11.99***
	(5.082)	(4.364)	(4.680)	(4.552)
1(Eco Reg Columbia Plat)	18.16	17.84	19.19	17.26
	(19.54)	(18.85)	(17.46)	(24.49)
1(Eco Reg Blue Mtns)	-9.480***	-9.251***	-9.598***	-9.498***
	(3.449)	(2.808)	(3.711)	(2.846)
1(Eco Reg N Rockies)	15.66	15.26	16.38	15.01
	(19.53)	(18.87)	(15.64)	(24.14)
1(Eco Reg Willamette Valley)	28.33***	28.43***	28.75***	27.94***
	(7.270)	(6.900)	(7.028)	(6.892)
1(Eco Reg Cascades)	3.836	3.941	4.136	3.791
	(5.607)	(5.380)	(4.941)	(5.483)
1(Eco Reg N Cascades)	13.12	12.57	13.36	13.22
	(10.27)	(9.649)	(10.62)	(9.941)
1(Eco Reg Klam Mtns CA N Coast Range)	41.10***	41.38***	48.67***	41.48***
	(8.070)	(7.543)	(7.915)	(7.522)

Continued on next page

	(1) No Interactions	(2) Income	(3) Elastic Net	(4) Income + Time
1(Eco Reg N Basin Range)	2.783	3.061	2.765	2.901
	(7.092)	(6.772)	(7.621)	(7.001)
1(LC water)	1.623	1.594	1.593	1.431
	(2.224)	(2.246)	(2.254)	(2.284)
1(LC barren)	9.711**	9.879***	9.830**	9.662**
	(4.029)	(3.784)	(3.920)	(3.938)
1(LC forest)	6.096***	6.087***	6.162***	6.116***
	(1.966)	(1.938)	(2.054)	(1.905)
1(LC shrub)	-2.830	-2.856	-2.808	-2.717
	(3.573)	(3.451)	(8.435)	(3.321)
1(LC herbaceous)	-1.550	-1.487	-1.810	-1.473
	(3.259)	(3.129)	(3.241)	(3.216)
1(LC planted)	5.097**	5.062**	4.969**	5.174**
	(2.044)	(2.064)	(2.308)	(2.109)
1(LC wetlands)	5.763*	5.796*	5.683*	5.861*
	(2.991)	(2.963)	(3.060)	(3.065)
var(WT)	-5.426***	-5.542***	-5.428***	-5.212***
	(0.670)	(0.664)	(0.948)	(0.741)
cov(WT,PR)	-2.038*	-2.070*	-1.872	-1.879*
	(1.062)	(1.089)	(2.040)	(1.107)
cov(WT,PH)	-0.167	-0.175	-0.00363	-0.125
	(0.770)	(0.636)	(0.965)	(0.619)
cov(WT, GM)	0.831***	0.846***	0.979	0.912***
	(0.250)	(0.243)	(0.673)	(0.224)

Continued on next page

	(1) No Interactions	(2) Income	(3) Elastic Net	(4) Income + Time
cov(WT, TC3)	0.691***	0.693***	0.690***	0.685***
	(0.0654)	(0.0576)	(0.0807)	(0.0660)
cov(PR)	5.880***	5.795***	5.889***	5.748***
	(1.227)	(1.126)	(1.497)	(1.117)
cov(PR,PH)	-3.928***	-4.012***	-3.834***	-4.048***
	(0.816)	(0.814)	(0.863)	(0.835)
cov(PR,GM)	-0.311	-0.281	-0.330	-0.301
	(0.404)	(0.565)	(0.748)	(0.408)
cov(PR,TC3)	-0.788***	-0.787***	-0.775***	-0.787***
	(0.0467)	(0.0483)	(0.0620)	(0.0530)
var(PH)	-0.208	-0.197	-0.0310	-0.238
	(0.334)	(0.252)	(0.648)	(0.267)
cov(PH,GM)	0.771***	0.761***	0.804***	0.811***
	(0.126)	(0.128)	(0.229)	(0.138)
cov(PH,TC3)	-0.152***	-0.153***	-0.160***	-0.147***
	(0.0465)	(0.0498)	(0.0469)	(0.0502)
var(GM)	1.505***	1.543***	1.639***	1.628***
	(0.230)	(0.229)	(0.510)	(0.241)
cov(GM,TC3)	0.171***	0.170***	0.163***	0.173***

Continued on next page

	(1) No Interactions	(2) Income	(3) Elastic Net	(4) Income + Time
	(0.0320)	(0.0362)	(0.0483)	(0.0318)
var(TC3)	0.432***	0.426***	0.432***	0.424***
	(0.0752)	(0.0595)	(0.0985)	(0.0640)
Observations	293263	293263	293263	293263
Log Likelihood	-23229.81	-23223.65	-23224.82	-23120.05
AIC	46553.63	46551.29	46551.64	46440.10
BIC	47051.30	47101.91	47091.67	47498.98
Weighted?				

Note: N = 293,263. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Estimated in STATA 16 using MIXLOGITWTP.

A Results with weights from 2016 FWS data without LASSO selection of variables

The summary statistics from the 2015 eBird auxiliary survey relative to the 2016 US FWS survey for birders who took trip(s) away from home is featured in Table A3.

The main results are featured in Tables A4, A5, A6 and A7. The results shown are based on 9,287 trips taken by 321 birders who responded to the auxiliary survey. Our preferred model is featured in column 6 of each of these tables which incorporates the exogenous weights from the 2016 U.S. FWS data.

Consideration Sets We limit eBirders choice set to all possible sites within 60-miles of their home address. Due to computational limits, we are currently only using 10 percent of the non-chosen alternatives from the full set of alternatives; the number of sites in a consideration set ranges from 2 to 89, with the average being 44. Note the results presented here are preliminary. Early versions of this paper used the population weights constructed with the 2016 U.S. FWS data and lasso was not decide which variables to include as additional controls and are available in the appendix. Feedback from audience members about the large number of control variables and collinearity between the controls led us to use LASSO and its related tools to select for which control variables to include in the model to reduce the dimensionality and risk of having overparameterized the model.

A.1 Mixed Logit in WTP Space

Marginal Utilities: Net Income Across all models in Tables A4, A5, A6 and A7, the marginal WTP of other consumption (the negative α coefficient on the travel cost variable, linear in this specification) is statistically significant and negative, the expected sign. The parameter is robust across all of our specifications.

Marginal WTP of Expected Number by Type and Correlations In Table A4, the main variable of interest in the model is β_1^i and X_{jt} which contain a vector of expected species by type (water birds, birds of prey, perching birds and game birds). In our preferred specification, shown in column 6 in Tables A4 and A5 , we find that the marginal WTP of expected species by type is heterogeneous. Part of the heterogeneity is captured through interaction terms with observable characteristics and temporal fixed effects. Considerable unobservable heterogeneity remains as evidence by the statistical significance of the random parameters and the covariance-variance matrix for these parameters. The baseline parameters for the expected number of water birds and birds of prey are each respectively positive and significant, game birds is positive, but statistically insignificant, whereas the parameter for the expected number of perching birds is negative. The variance parameters for each type is statistically significant and only the covariance term that is not statistically significant is between the expected number of perching birds and travel cost.

The negative and statistically significant covariance terms between the expected number of water birds and birds of prey, game birds and travel cost as well as between the expected number of game birds and perching birds suggesting lower levels of utility along the distribution on one of the attributes whereas all the other statistically significant covariance terms are positive.

Systematic Variation in the Marginal WTP of Expected Number by Type: Water Birds The sources of observable heterogeneity in the marginal WTP of an additional water bird species comes from observable differences in birder characteristics and time. The results suggest birders have a higher marginal WTP of seeing an additional water bird species if they are both female and a lister. The temporal differences in marginal WTP for an additional water bird species is higher for March, August and the fall months of September, October and November, but negative for the months of April, May, June and the linear time trend.

Systematic Variation in the Marginal WTP of Expected Number by Type: Birds of Prey

The sources of observable heterogeneity in the marginal WTP of an additional birds of prey species comes from observable differences in birder characteristics and time. The results suggest that birders have a lower marginal WTP of seeing an additional bird of prey if they are from a higher income bracket (than the sample average), and/or are older (than the sample average), and/or are female or a lister. However, female listers marginal WTP is higher for an additional bird of prey species than a male lister. The temporal differences in marginal WTP for an additional birds of prey bird species is higher in March, and the linear time trend parameter is positive and statistically significant.

Systematic Variation in the Marginal WTP of Expected Number by Type: Perching Birds The sources of observable heterogeneity in the marginal WTP of an additional perching bird species comes from observable differences in birder characteristics and time. The results suggest birders have a higher marginal WTP of seeing an additional perching bird if they are from a higher income bracket (than the sample average), and/or are a lister. The temporal differences in marginal WTP for an additional birds of prey bird species is lower in September.

Systematic Variation in the Marginal WTP of Expected Number by Type: Game Birds The sources of observable heterogeneity in the marginal WTP of an additional game bird species comes from observable differences in birder characteristics and time. The results suggest birders have a higher marginal WTP of seeing an additional game bird species if they are older, and/or female, and/or are a lister, and lower a lower marginal WTP if they are from a higher income bracket (than the sample average). The marginal WTP of a female lister is lower for an additional game bird species than a male lister. The temporal differences in marginal WTP for an additional game bird species is lower in July, November, December, and for the linear time trend.

Other Site Attributes A number of other site attributes bear statistically significant marginal-utility coefficients in Table A6 and A7. These include indicators for different management regimes for biodiversity, ecoregions, land cover classes, urban hotspots, along with our continuous "congestion/popularity" measure for each hotspot.[21]

[21] We originally did include a variable to capture the expected presence of an endangered bird species,

As in Kolstoe and Cameron (2017), the estimated marginal WTP from visits to National Wildlife Refuges which are specifically managed for bird biodiversity yields the highest to eBirders. The marginal WTP for a visit to a National Wildlife Refuge is statistically significant, implying there is a higher value (utility) for a birder for specifically visiting a site managed for bird biodiversity, above and beyond managing a site managed for biodiversity (this parameter is additive to the parameter for National Parks, etc.). This may not be surprising in light of when individuals are traveling to see birds, they are primarily traveling to see waterfowl which can be found at National Wildlife Refuges (Wilson, 2010). The marginal WTP of a visit to a site such as National Parks (e.g. a site with GAP 1 or 2 status) or National Forests (e.g. a site with GAP 3 status) sites are positive, but are not statistically significant.

We also find a persistently negative coefficient on the indicator for an urban destination (activated for hotspots which lie within an Urban Area as defined by the 2010 Census), as was the case in Kolstoe and Cameron (2017). This suggests there are latent attributes of sites in urban areas that confers disutility to non-urban sites, independent of species richness.

To control destination ecoregions and land cover, to pre-empt any omitted variable bias from these two sources in the estimated coefficients on the expected number of bird species. These two sets of controls account for any utility that is accounted for directly from either the type of ecoregion or land cover at a site. The baseline ecoregion for the trips is the Puget Lowland. The indicators for the Coast Range, Columbia Plateau, North Rockies, Willamette Valley, North Cascades, Klamath Mountains, and the Northern Basin Rage are positive and statistically significant, indicating a higher value (utility) than to a site in the Puget Lowland. However the indicators for the Blue Mountains is negative and statistically significant, whereas the indicator for the Cascades is negative and statistically insignificant. The land cover class indicators for barren land, forest, planted and wetlands are positive and statistically significant whereas the indicators for water, shrub/scrub and herbaceous land cover are not statistically significant.

As in Kolstoe and Cameron (2017) the congestion/popularity variable is proxied using all eBird members' reported visits to *this* hotspot in the same month last year, as a proportion of all eBird members' visits to *any* hotspot during that same time frame. The parameters suggest that the marginal WTP of anticipated congestion/popularity is a good thing, and only is in the range where it diminishes for one-percent of the site alternatives. Only a few sites in our sample receive this share of eBird visits in the previous year, thus over most of the range of the data, this variable is positive.[22]

however, it was not statistically significant in any of the models, nor was the expected presence of threatened bird species so we have since omitted them from the model.

[22]Linearity in this variable is strongly rejected by a likelihood ratio test.

B Additional Tables from Earlier Versions of the Paper

Table A3: Selected Summary Statistics from the 2015 eBird Auxiliary Survey Relative to the 2016 USFWS Survey for Birders Who Took Trip(s) Away from Home

Survey Sample Statistics	% New Survey	% USFW 2016
Gender		
Male	43	55
Female	57	45
Age		
16 to 24	1	5
25 to 34	6	13.9
35 to 44	9	16.8
45 to 54	15	17.8
55 to 64	31	17.8
65 to 74	31	20.8
75 +	2.8	7.9
Income		
Less than $20,000	3.34	0
$20,000 to $24,999	3.02	4
$25,000 to $29,999	3.18	0
$30,000 to $49,999	12.71	5
$50,000 to $74,999	18.34	6.9
$75,000 to $99,999	13.53	28.7
$100,000 to $149,999	15.15	45.6
$150,000 or more	9.86	11.9

Table A4: Progression of Models, Pooled Oregon and Washington Sample; Select Key Coefficients Results – 60-Mile Choice Set

	No Interactions		Ecological Economics Specification		+ Gender, Age, Lister Status	
	No Weights	2016 Weights	No Weights	2016 Weights	No Weights	2016 Weights
Mean Coefficients	(1)	(2)	(3)	(4)	(5)	(6)
Expected # water bird species (WT)	0.878***	1.630***	3.241**	3.173***	2.980***	3.403***
	(0.202)	(0.449)	(1.267)	(0.996)	(0.771)	(0.749)
Expected # bird of prey species (PR)	-0.565***	2.990***	3.342**	2.380**	-1.356*	4.351***
	(0.216)	(0.654)	(1.395)	(1.036)	(0.741)	(0.747)
Expected # perching bird species (PH)	0.0398	-0.802***	-0.312	-1.223*	-1.868***	-2.347***
	(0.137)	(0.242)	(0.509)	(0.726)	(0.402)	(0.436)
Expected # game bird species (GM)	0.930***	-0.786**	-0.711**	0.914**	2.228***	0.621
	(0.132)	(0.340)	(0.310)	(0.451)	(0.330)	(0.384)
Var(WT)	5.930***	0.655	1.260	6.456***	5.686***	6.591***
	(0.260)	(1.446)	(1.158)	(1.561)	(0.267)	(0.302)
Cov(PR,WT)	-3.722***	-6.883***	-7.252***	-2.343***	-3.786***	-2.687***
	(0.274)	(1.709)	(1.195)	(0.862)	(0.357)	(0.354)
Cov(PH,WT)	-0.205	2.539***	2.281***	2.203**	0.0286	1.884***
	(0.141)	(0.741)	(0.838)	(1.007)	(0.163)	(0.183)
Cov(GM,WT)	1.636***	6.455***	7.950***	-0.530*	2.289***	-0.806***
	(0.184)	(1.990)	(1.425)	(0.319)	(0.140)	(0.175)
Cov(TC3, WT)	0.422***	0.806***	0.835***	-0.168***	0.391***	-0.0828***
	(0.0206)	(0.122)	(0.199)	(0.0406)	(0.0158)	(0.0184)
Var(PR)	3.913***	3.026**	2.385**	5.150***	5.350***	7.013***
	(0.321)	(1.354)	(1.184)	(0.707)	(0.260)	(0.340)
Cov(PH,PR)	-1.376***	-3.632***	-3.943***	2.068***	-1.777***	2.640***
	(0.132)	(0.628)	(0.594)	(0.364)	(0.133)	(0.209)
Cov(GM,PR)	2.211***	2.901***	3.560***	0.867**	1.708***	1.114***
	(0.149)	(0.465)	(0.421)	(0.405)	(0.138)	(0.233)
Cov(TC3,PR)	0.239***	0.402***	0.411***	0.397***	0.0968***	0.493***

Table A4 Continued

	No Interactions		Ecological Economics Specification		+ Gender, Age, Lister Status	
	No Weights	2016 Weights	No Weights	2016 Weights	No Weights	2016 Weights
Mean Coefficients	(1)	(2)	(3)	(4)	(5)	(6)
	(0.0200)	(0.0728)	(0.0712)	(0.0486)	(0.0184)	(0.0219)
Var(PH)	0.385**	1.566***	1.665***	1.969***	0.745***	1.688***
	(0.166)	(0.321)	(0.291)	(0.373)	(0.166)	(0.233)
Cov(GM,PH)	3.894***	0.651**	0.637***	-4.049***	4.398***	-6.133***
	(0.230)	(0.299)	(0.202)	(0.735)	(0.199)	(0.286)
Cov(TC3,PH)	-0.498***	-0.408***	-0.433***	0.118**	-0.700***	-0.0252
	(0.0355)	(0.0502)	(0.0393)	(0.0536)	(0.0240)	(0.0208)
Var(GM)	2.921***	2.168***	2.724***	0.572***	3.793***	0.958***
	(0.255)	(0.430)	(0.407)	(0.163)	(0.164)	(0.108)
Cov(TC3, GM)	0.564***	0.135***	0.147**	0.293***	0.396***	0.258***
	(0.0541)	(0.0487)	(0.0710)	(0.0423)	(0.0238)	(0.0155)
Var(TC3)	0.116***	0.420***	0.409***	0.728***	0.0867***	0.667***
	(0.0246)	(0.0375)	(0.0680)	(0.0978)	(0.0163)	(0.0210)
Sample Selection?	No	No	No	No	No	No
Time fixed effects?	No	No	Yes	Yes	Yes	Yes
Ecoregion?	Yes	Yes	Yes	Yes	Yes	Yes
Total Alternatives	308,922	308,387	308,387	308,387	308,620	308,620
Log Likelihood	-17421.11	-352125.71	-347230.04	-311840.57	-17240.16	-17228.46
AIC	34926.21	704335.41	694648.07	623869.14	34700.32	34676.92
BIC	35373.13	704782.25	695648.15	624869.22	35870.70	35847.30

Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
† Share of all eBird trips, same month, last year, to site j

NOTES: Estimates estimated via STATA *mixlogitwtp.ado*. These results use 500 Halton draws for the mixed logit in WTP model simulations. Baseline coefficient represents the marginal utility for an eBirder who is visiting a rural site that is not managed for biodiversity in the Puget Lowland in January of 2013. Models are the results for choice sets within a 60-mile drive from a member's home.

Table A5: Progression of Models, Pooled Oregon and Washington Sample; By Interaction Terms Results – 60-Mile Choice Set

Select Coefficients

	No Interactions		Ecological Economics Specification		+ Gender, Age, Lister Status	
	No Weights	2016 Weights	No Weights	2016 Weights	No Weights	2016 Weights
Mean Coefficients	(1)	(2)	(3)	(4)	(5)	(6)
WT × diff. mean ($10,000)			0.0214	-0.0121	-0.0425	0.0257
			(0.0809)	(0.0878)	(0.0376)	(0.0375)
WT× diff. mean age					-0.0207*	-0.00757
					(0.0126)	(0.0139)
WT× 1($female$)					0.129	0.573
					(0.855)	(0.810)
WT× 1($lister$)					-1.357**	-0.845
					(0.588)	(0.567)
WT× 1($female$)× 1($lister$)					-0.603	2.952***
					(0.953)	(0.936)
PR × diff. mean income ($10,000)			-0.0666	-0.359***	-0.179***	-0.400***
			(0.112)	(0.127)	(0.0516)	(0.0573)
PR× diff. mean age					0.0278**	-0.0370**
					(0.0138)	(0.0151)
PR× 1($female$)					-1.109	-2.801***
					(0.828)	(0.764)
PR× 1($lister$)					1.878***	-1.483**
					(0.668)	(0.610)
PR× 1($female$)× 1($lister$)					-0.209	2.617***
					(0.999)	(0.954)
PH × diff. mean income ($10,000)			-0.109***	0.248**	0.00855	0.206***
			(0.0304)	(0.107)	(0.0245)	(0.0301)
PH× diff. mean age					0.00944	0.00408
					(0.00752)	(0.00823)
PH× 1($female$)					-0.979**	-0.497

Table A5 Continued

Mean Coefficients	No Interactions		Ecological Economics Specification		+ Gender, Age, Lister Status	
	No Weights	2016 Weights	No Weights	2016 Weights	No Weights	2016 Weights
	(1)	(2)	(3)	(4)	(5)	(6)
					(0.486)	(0.475)
PH× 1($lister$)					1.603***	1.867***
					(0.309)	(0.344)
PH× 1($female$)× 1($lister$)					0.882	-0.885
					(0.562)	(0.564)
GM× diff. mean income ($10,000)			0.129***	-0.0646	0.00663	-0.116***
			(0.0231)	(0.0505)	(0.0153)	(0.0226)
GM× diff. mean age					-0.0222***	0.00975*
					(0.00488)	(0.00534)
GM× 1($female$)					1.081**	2.860***
					(0.450)	(0.457)
GM× 1($lister$)					-0.369	0.856***
					(0.248)	(0.303)
GM× 1($female$)× 1($lister$)					0.168	-3.131***
					(0.472)	(0.533)
WT× 1($February$)			-0.641	0.904	0.658	1.174
			(2.177)	(0.876)	(0.696)	(0.726)
WT× 1($March$)			0.744	1.400	0.684	1.478*
			(1.728)	(1.188)	(0.787)	(0.777)
WT× 1($April$)			-2.836***	-2.513**	-1.803**	-1.664**
			(1.085)	(1.038)	(0.717)	(0.735)
WT× 1(May)			-1.483	-2.243***	-1.622**	-1.430*
			(1.307)	(0.857)	(0.730)	(0.736)
WT× 1($June$)			-2.101**	-1.645**	-1.451*	-1.561*
			(1.010)	(0.812)	(0.767)	(0.821)
WT× 1($July$)			2.587	1.043	0.446	0.471
			(2.105)	(0.829)	(0.847)	(0.827)

Table A5 Continued

	No Interactions		Ecological Economics Specification		+ Gender, Age, Lister Status	
	No Weights	2016 Weights	No Weights	2016 Weights	No Weights	2016 Weights
Mean Coefficients	(1)	(2)	(3)	(4)	(5)	(6)
WT× 1(*August*)			0.517	2.038*	2.507***	2.061**
			(2.572)	(1.129)	(0.898)	(0.980)
WT× 1(*September*)			-0.210	1.101	2.228***	1.957**
			(2.651)	(1.283)	(0.848)	(0.864)
WT× 1(*October*)			1.960	1.599	1.725**	2.194***
			(1.293)	(1.081)	(0.785)	(0.822)
WT× 1(*November*)			2.886***	1.731*	2.156***	2.107***
			(0.918)	(0.979)	(0.808)	(0.775)
WT× 1(*December*)			-0.922	-0.966	-1.554**	-0.979
			(1.719)	(1.004)	(0.744)	(0.818)
WT× $t13$			-1.044	-1.041**	-1.245***	-1.327***
			(0.755)	(0.447)	(0.246)	(0.252)
PR× 1(*February*)			0.242	-0.989	-0.716	-0.809
			(1.816)	(0.663)	(0.766)	(0.795)
PR× 1(*March*)			0.572	0.974	1.140	1.494*
			(1.004)	(0.696)	(0.929)	(0.904)
PR× 1(*April*)			1.012	0.494	0.301	0.313
			(1.300)	(0.890)	(0.788)	(0.789)
PR× 1(*May*)			-2.032*	0.818	0.329	0.722
			(1.045)	(0.886)	(0.796)	(0.780)
PR× 1(*June*)			-0.419	-0.748	-1.026	-0.541
			(0.956)	(0.666)	(0.800)	(0.839)
PR× 1(*July*)			-0.440	-1.104	-1.023	-0.728
			(1.215)	(0.739)	(0.843)	(0.845)
PR× 1(*August*)			-1.387	-1.648	-2.107***	-1.328
			(1.755)	(1.006)	(0.807)	(0.853)
PR× 1(*September*)			-0.154	-0.521	-0.804	-0.365

Table A5 Continued

	No Interactions		Ecological Economics Specification		+ Gender, Age, Lister Status	
	No Weights	2016 Weights	No Weights	2016 Weights	No Weights	2016 Weights
Mean Coefficients	(1)	(2)	(3)	(4)	(5)	(6)
			(1.875)	(0.960)	(0.775)	(0.786)
PR$\times$ 1($October$)			-0.0727	-0.485	-0.250	-0.150
			(1.552)	(0.826)	(0.802)	(0.833)
PR$\times$ 1($November$)			-0.703	-0.247	-0.581	0.303
			(1.219)	(0.798)	(0.840)	(0.809)
PR$\times$ 1($December$)			-0.102	0.544	0.109	0.595
			(1.315)	(0.852)	(0.827)	(0.863)
PR$\times$ $t13$			0.124	0.575	0.162	0.415*
			(0.833)	(0.408)	(0.242)	(0.241)
PH$\times$ 1($February$)			-0.0234	-0.127	-0.491	-0.563
			(0.497)	(0.411)	(0.435)	(0.443)
PH$\times$ 1($March$)			0.590	0.965	0.420	0.400
			(0.510)	(0.673)	(0.553)	(0.564)
PH$\times$ 1($April$)			-0.193	0.266	0.408	0.206
			(0.595)	(0.469)	(0.417)	(0.442)
PH$\times$ 1(May)			0.736	0.559	0.708*	0.475
			(0.784)	(0.527)	(0.420)	(0.446)
PH$\times$ 1($June$)			0.615	0.734*	0.850*	0.502
			(0.430)	(0.427)	(0.441)	(0.491)
PH$\times$ 1($July$)			-0.863	0.430	0.824*	0.494
			(0.786)	(0.627)	(0.476)	(0.491)
PH$\times$ 1($August$)			0.0340	0.344	1.104**	0.117
			(0.933)	(0.778)	(0.503)	(0.547)
PH$\times$ 1($September$)			-1.002	-0.543	-0.694*	-0.955**
			(0.826)	(0.668)	(0.418)	(0.459)
PH$\times$ 1($Octber$)			-1.558**	-0.204	0.0250	-0.247
			(0.710)	(0.624)	(0.435)	(0.479)

Table A5 Continued

	No Interactions		Ecological Economics Specification		+ Gender, Age, Lister Status	
	No Weights	2016 Weights	No Weights	2016 Weights	No Weights	2016 Weights
Mean Coefficients	(1)	(2)	(3)	(4)	(5)	(6)
PH× 1(*November*)			-1.256***	-0.357	-0.392	-0.546
			(0.486)	(0.589)	(0.449)	(0.447)
PH× 1(*December*)			-0.656	-0.424	0.372	-0.138
			(0.498)	(0.551)	(0.427)	(0.486)
PH× $t13$			-0.0831	-0.155	0.181	0.0493
			(0.245)	(0.304)	(0.135)	(0.142)
GM× 1(*February*)			0.132	0.0783	0.146	0.117
			(0.314)	(0.228)	(0.302)	(0.324)
GM× 1(*March*)			-0.693**	-0.551*	-0.217	-0.270
			(0.287)	(0.282)	(0.288)	(0.312)
GM× 1(*April*)			-0.125	0.0537	-0.0919	0.178
			(0.290)	(0.244)	(0.296)	(0.325)
GM× 1(*May*)			-0.0837	0.0624	0.373	0.504
			(0.365)	(0.422)	(0.303)	(0.330)
GM× 1(*June*)			-0.169	0.0582	0.0144	0.0442
			(0.418)	(0.309)	(0.361)	(0.384)
GM× 1(*July*)			-0.279	-0.632**	-0.413	-0.624*
			(0.178)	(0.289)	(0.323)	(0.360)
GM× 1(*August*)			0.165	-0.300	-0.153	-0.197
			(0.450)	(0.433)	(0.310)	(0.346)
GM× 1(*September*)			0.222	-0.156	-0.197	-0.254
			(0.362)	(0.295)	(0.329)	(0.372)
GM× 1(*Octber*)			-0.0325	-0.346	-0.305	-0.368
			(0.298)	(0.363)	(0.281)	(0.328)
GM× 1(*November*)			0.244	-0.504	-0.599**	-0.785**
			(0.409)	(0.527)	(0.274)	(0.316)
GM× 1(*December*)			-0.459**	-0.743***	-0.715**	-0.926***

Table A5 Continued

Mean Coefficients	No Interactions No Weights (1)	 2016 Weights (2)	Ecological Economics Specification No Weights (3)	 2016 Weights (4)	+ Gender, Age, Lister Status No Weights (5)	 2016 Weights (6)
			(0.207)	(0.247)	(0.309)	(0.348)
GM× $t13$			0.149*	-0.0986	-0.0421	-0.177*
			(0.0884)	(0.131)	(0.0922)	(0.104)
Sample Selection?	No	No	No	No	No	No
Time fixed effects?	No	No	Yes	Yes	Yes	Yes
Ecoregion indicators?	Yes	Yes	Yes	Yes	Yes	Yes
Total Alternatives	308,922	308,387	308,387	308,387	308,620	308,620
Log Likelihood	-17421.11	-352125.71	-347230.04	-311840.57	-17240.16	-17228.46
AIC	34926.21	704335.41	694648.07	623869.14	34700.32	34676.92
BIC	35373.13	704782.25	695648.15	624869.22	35870.70	35847.30

Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
† Share of all eBird trips, same month, last year, to site j

NOTES: Estimates estimated via STATA *mixlogitwtp.ado*. These results use 500 Halton draws for the mixed logit in WTP model simulations. Baseline coefficient represents the marginal utility for an eBirder who is visiting a rural site that is not managed for biodiversity in the Puget Lowland in January of 2013. Models are the results for choice sets within a 60-mile drive from a member's home.

Table A6: Progression of Models, Pooled Oregon and Washington Sample; Select Key Site Attribute Coefficients Results – 60-Mile Choice Set

	No Interactions		Ecological Economics Specification		+ Gender, Age, Lister Status	
	No Weights	2016 Weights	No Weights	2016 Weights	No Weights	2016 Weights
	(1)	(2)	(3)	(4)	(5)	(6)
1(National Wildlife Refuge)	10.48***	1.487	1.918	6.191*	9.954***	10.72***
	(1.042)	(7.128)	(7.059)	(3.411)	(1.019)	(1.068)
1(National Parks, etc.)	-0.0768	1.644	1.707	1.207	0.918	0.605
	(0.738)	(2.500)	(2.749)	(2.347)	(0.753)	(0.779)
1(National Forests, etc.)	0.0519	-0.923	-0.929	0.755	0.531	0.551
	(0.499)	(1.584)	(1.528)	(1.608)	(0.518)	(0.531)
1(Urban Area)	-7.913***	-5.243***	-5.032***	-5.380**	-6.971***	-5.865***
	(0.769)	(1.581)	(1.579)	(2.531)	(0.782)	(0.792)
†$Congestion/Popularity_{jt}$	7936.6***	7479.5***	7459.8***	7865.1***	7962.5***	8042.5***
	(253.1)	(1032.1)	(923.8)	(941.6)	(252.1)	(256.1)
$(Congestion/Popularity_{jt})^2$	-312837.3***	-303912.9***	-305342.5***	-336081.2***	-316396.7***	-330533.7***
	(16051.4)	(60670.3)	(52251.1)	(46911.0)	(16616.4)	(16625.7)
Sample Selection?	No	No	No	No	No	No
Time fixed effects?	No	No	Yes	Yes	Yes	Yes
Ecoregion indicators?	Yes	Yes	Yes	Yes	Yes	Yes
Total Alternatives	308,922	308,387	308,387	308,387	308,620	308,620
Log Likelihood	-17421.11	-352125.71	-347230.04	-311840.57	-17240.16	-17228.46
AIC	34926.21	704335.41	694648.07	623869.14	34700.32	34676.92
BIC	35373.13	704782.25	695648.15	624869.22	35870.70	35847.30

Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
† Share of all eBird trips, same month, last year, to site j

NOTES: Estimates estimated via STATA *mixlogitwtp.ado*. These results use 500 Halton draws for the mixed logit in WTP model simulations. Baseline coefficient represents the marginal utility for an eBirder who is visiting a rural site that is not managed for biodiversity in the Puget Lowland in January of 2013. Models are the results for choice sets within a 60-mile drive from a member's home.

Table A7: Progression of Models, Pooled Oregon and Washington Sample; Select Coefficients By Interaction Terms Results – 60-Mile Choice Set

	No Interactions		Ecological Economics Specification		+ Gender, Age, Lister Status	
	No Weights	2016 Weights	No Weights	2016 Weights	No Weights	2016 Weights
Mean Coefficients	(1)	(2)	(3)	(4)	(5)	(6)
$1(Coast\ Range)_j$	9.201***	26.82***	27.84***	14.50**	10.63***	14.18***
	(2.515)	(3.406)	(4.517)	(5.829)	(2.440)	(2.490)
$1(Columbia\ Plateau)_j$	5.021	16.03	16.98	23.03	8.336*	14.69***
	(4.134)	(12.39)	(11.68)	(102.9)	(4.408)	(4.992)
$1(Blue\ Mountains)_j$	-2.744	-0.355	-0.421	-4.492	-6.212***	-3.562*
	(2.085)	(4.373)	(3.891)	(3.408)	(2.088)	(2.119)
$1(North\ Rockies)_j$	6.414	14.40	15.16	23.53	11.73**	15.60**
	(6.049)	(13.83)	(13.20)	(101.7)	(5.290)	(6.310)
$1(Willamette\ Valley)_j$	17.70***	41.35***	41.49***	19.70***	21.73***	27.52***
	(2.823)	(8.012)	(7.172)	(7.496)	(2.951)	(3.236)
$1(Cascades)_j$	1.255	-0.104	-0.0132	-0.834	2.030	-2.454
	(3.401)	(6.575)	(7.153)	(6.907)	(3.120)	(3.676)
$1(North\ Cascades)_j$	8.796***	12.03**	13.51**	12.67	11.92***	16.64***
	(2.966)	(6.012)	(5.381)	(9.384)	(3.009)	(3.538)
$1(Klamath\ Mtns,\ Coast\ Range)_j$	32.94***	52.62***	58.42***	39.38***	32.73***	43.14***
	(4.688)	(6.184)	(7.108)	(9.770)	(4.183)	(5.118)
$1(N.\ Basin\ Range)$	20.26**	15.46*	15.98*	17.40	16.52**	15.42**
	(7.966)	(8.880)	(8.960)	(12.51)	(7.105)	(7.102)
$1(LC\ Water/Perennial\ Snow\ \&\ Ice)$	0.762	1.450	1.213	-0.165	0.642	0.488
	(0.818)	(2.336)	(2.278)	(2.992)	(0.842)	(0.910)
$1(LC\ Barren\ Land)$	10.88***	6.159	5.766	4.579	10.60***	9.800***
	(0.833)	(4.349)	(4.766)	(3.436)	(0.839)	(0.886)
$1(LC\ Forest)$	7.806***	11.24***	11.00***	6.542***	7.032***	7.058***
	(0.677)	(2.333)	(2.118)	(2.428)	(0.644)	(0.661)
$1(LC\ Shrub/Scrub)$	-0.283	-0.149	0.143	-1.239	-1.328	0.844
	(1.001)	(3.975)	(3.985)	(3.607)	(0.993)	(1.092)

Table A7 Continued

	No Interactions		Ecological Economics Specification		+ Gender, Age, Lister Status	
	No Weights	2016 Weights	No Weights	2016 Weights	No Weights	2016 Weights
Mean Coefficients	(1)	(2)	(3)	(4)	(5)	(6)
1($LC\ Herbaceous$)	-0.0118	-1.170	-0.932	-1.342	0.105	-0.462
	(1.694)	(3.343)	(3.349)	(3.642)	(1.602)	(1.669)
1($LC\ Planted$)	4.917***	6.869***	6.923**	6.559***	4.981***	5.734***
	(0.750)	(2.498)	(2.699)	(2.258)	(0.796)	(0.800)
1($LC\ Wetlands$)	9.156***	7.838***	7.119**	9.420***	8.960***	10.49***
	(0.702)	(2.989)	(3.069)	(3.432)	(0.718)	(0.737)
Sample Selection?	No	No	No	No	No	No
Time fixed effects?	No	No	Yes	Yes	Yes	Yes
Ecoregion?	Yes	Yes	Yes	Yes	Yes	Yes
Total Alternatives	308,922	308,387	308,387	308,387	308,620	308,620
Log Likelihood	-17421.11	-352125.71	-347230.04	-311840.57	-17240.16	-17228.46
AIC	34926.21	704335.41	694648.07	623869.14	34700.32	34676.92
BIC	35373.13	704782.25	695648.15	624869.22	35870.70	35847.30

Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
† Share of all eBird trips, same month, last year, to site j

NOTES: Estimates estimated via STATA *mixlogitwtp.ado*. These results use 500 Halton draws for the mixed logit in WTP model simulations. Baseline coefficient represents the marginal utility for an eBirder who is visiting a rural site that is not managed for biodiversity in the Puget Lowland in January of 2013. Models are the results for choice sets within a 60-mile drive from a member's home.

Table A8: Variations in the value of a birding trip based on the expected number of bird species by type (mean of continuous variables, for January 2013, unmanaged site, non-urban developed destination in the Puget Lowlands). Estimates incorporate the 2016 FWS weights.

Simulation		$ Total WTP for trip
Minimum	(1 water bird, 1 bird of prey, 1 perching bird, 0 game birds)	$1.78 (.75, 2.81)
10th percentile	(10 water birds, 12 birds of prey, 22 perching birds, 0.14 game birds)	$111.66 (105.33, 117.99)
25th percentile	(11 water birds, 12.84 birds of prey, 24 perching birds, 12 game birds)	$111.37 (104.4, 118.34)
50th percentile	(12.03 water birds, 13.07 birds of prey, 25 perching birds, 14 game birds)	$118.87 (111.42, 126.32)
75th percentile	(13.36 water birds, 15 birds of prey, 26 perching birds, 14 game birds)	$105.49 (97.38, 113.59)
90th percentile	(14.29 water birds, 15.11 birds of prey, 27.39 perching birds, 16.02 game birds)	$126.19 (117.63, 134.76)
Maximum	(20.84 water birds, 20.5 birds of prey, 35 perching birds, 20.63 game birds)	$170.07 (158.27, 181.87)

NOTE: The values in parenthesis represent the values within one standard deviation and include the correlations between the random variables

Table A9: Systematic seasonal variations in the value of a birding trip (calculated at mean species richness and mean congestion level, for January 2013, unmanaged site, non-urban developed destination in the Puget Lowlands). Estimates incorporate the 2016 FWS weights.

Simulation	Total WTP for a trip	MWTP per WT	MWTP per PR	MWTP per PH	MWTP per GM
B. By month (T_t variable)					
(At mean $E[S]$, means of cont. variables, January, 2013, not managed, rural, Puget Lowlands)					
January	$120.3	$3.45	$3.63	$-1.98	$.41
	(113.01, 127.59)	(3.08, 3.82)	(3.59, 3.68)	(-2.05, -1.91)	(.4, .42)
February	$-165.44	$4.62	$2.82	$-2.54	$.53
	(-172.73, -158.15)	(4.25, 4.99)	(2.78, 2.87)	(-2.61, -2.47)	(.52, .54)
March	$771.67	$4.93	$5.13	$-1.58	$.14
	(764.38, 778.96)	(4.56, 5.3)	(5.08, 5.17)	(-1.65, -1.51)	(.13, .15)
April	$86.01	$1.79	$3.94	$-1.77	$.59
	(78.72, 93.3)	(1.41, 2.16)	(3.9, 3.99)	(-1.84, -1.7)	(.58, .6)
May	$390.36	$2.02	$4.35	$-1.5	$.92
	(383.07, 397.65)	(1.65, 2.39)	(4.31, 4.4)	(-1.57, -1.43)	(.91, .92)
June	$115.62	$1.89	$3.09	$-1.47	$.46
	(108.32, 122.91)	(1.52, 2.26)	(3.05, 3.13)	(-1.54, -1.4)	(.45, .46)
July	$270.89	$3.92	$2.9	$-1.48	$-.21
	(263.6, 278.18)	(3.55, 4.29)	(2.86, 2.95)	(-1.55, -1.41)	(-.22, -.2)
August	$230	$5.51	$2.3	$-1.86	$.22
	(222.7, 237.29)	(5.14, 5.88)	(2.26, 2.35)	(-1.93, -1.79)	(.21, .22)
September	$120.3	$3.45	$3.63	$-1.98	$.41
	(113.01, 127.59)	(3.08, 3.82)	(3.59, 3.68)	(-2.05, -1.91)	(.4, .42)
October	$204.99	$5.64	$3.48	$-2.22	$.04
	(197.7, 212.29)	(5.27, 6.02)	(3.44, 3.53)	(-2.29, -2.15)	(.04, .05)
November	$33.41	$5.56	$3.93	$-2.52	$-.37
	(26.12, 40.7)	(5.18, 5.93)	(3.89, 3.98)	(-2.59, -2.45)	(-.38, -.37)
December	$-128.09	$2.47	$4.23	$-2.12	$-.51
	(-135.38, -120.79)	(2.1, 2.84)	(4.18, 4.27)	(-2.19, -2.04)	(-.52, -.51)
C. By year (T_t variable)					
(At mean $E[S]$, mean continuous variables, mean congestion, January, not managed, rural, developed, Puget Lowlands)					
2013	$120.3	$3.45	$3.63	$-1.98	$.41
	(113.01, 127.59)	(3.08, 3.82)	(3.59, 3.68)	(-2.05, -1.91)	(.4, .42)
2014	$8.27	$2.12	$4.05	$-1.93	$.23
	(.98, 15.56)	(1.75, 2.49)	(4, 4.09)	(-2, -1.86)	(.23, .24)
2015	$-103.76	$.79	$4.46	$-1.88	$.06
	(-111.05, -96.47)	(.42, 1.17)	(4.42, 4.51)	(-1.95, -1.81)	(.05, .07)

NOTE: The values in parenthesis represent the values within one standard deviation and include the correlations between the random variables

Table A10: Systematic variation by sociodemographics in the value of a birding trip (calculated at mean species richness and mean congestion level, for January 2013, unmanaged site, non-urban developed destination in the Puget Lowlands). Estimates incorporate the 2016 FWS weights.

Simulation	Total WTP for a trip	MWTP per WT	MWTP per PR	MWTP per PH	MWTP per GM
D. By gender					
(At means of cont. variables, January, 2013, not managed, rural, Puget Lowlands)					
Male, not a Lister(baseline)	$120.3	$3.45	$3.63	$-1.98	$.41
	(113.01, 127.59)	(3.08, 3.82)	(3.59, 3.68)	(-2.05, -1.91)	(.4, .42)
Female	$120.79	$3.45	$3.63	$-1.98	$.41
	(113.5, 128.08)	(3.08, 3.82)	(3.59, 3.68)	(-2.05, -1.91)	(.4, .42)
E. By lister status					
(At means of cont. variables, January, 2013, not managed, rural, Puget Lowlands)					
Lister	$130.1	$3.45	$3.63	$-1.98	$.41
	(122.81, 137.39)	(3.08, 3.82)	(3.59, 3.68)	(-2.05, -1.91)	(.4, .42)
Female Lister	$127.36	$3.45	$3.63	$-1.98	$.41
	(120.07, 134.65)	(3.08, 3.82)	(3.59, 3.68)	(-2.05, -1.91)	(.4, .42)
F. By difference in income (Z^i variables)					
(At mean $E[S]$, mean of other continuous variables, mean congestion, January, 2013, rural, developed, Puget Lowlands)					
Min of difference from sample mean income ($0000)=-6.812	$-249.21	$3.23	$7.07	$-3.75	$1.41
	(-256.5, -241.92)	(2.86, 3.6)	(7.03, 7.12)	(-3.82, -3.68)	(1.4, 1.42)
10th %ile of difference from sample mean income ($0000)=-4.6121	$-154.82	$3.28	$6.2	$-3.3	$1.16
	(-162.11, -147.53)	(2.91, 3.66)	(6.15, 6.24)	(-3.37, -3.23)	(1.15, 1.16)
25th %ile of difference from sample mean income ($0000)=-2.3621	$-58.27	$3.34	$5.3	$-2.83	$.89
	(-65.57, -50.98)	(2.97, 3.71)	(5.25, 5.34)	(-2.9, -2.76)	(.89, .9)
50th %ile of difference from sample mean income ($0000)=-2.3621	$-58.27	$3.34	$5.3	$-2.83	$.89
	(-65.57, -50.98)	(2.97, 3.71)	(5.25, 5.34)	(-2.9, -2.76)	(.89, .9)
Mean difference from sample mean income ($0000)=.985421	$85.36	$3.43	$3.96	$-2.14	$.51
	(78.07, 92.65)	(3.06, 3.8)	(3.91, 4)	(-2.21, -2.07)	(.5, .51)
90th %ile of difference from sample mean income ($0000)=5.1379	$263.53	$3.54	$2.3	$-1.29	$.02
	(256.24, 270.82)	(3.16, 3.91)	(2.25, 2.34)	(-1.36, -1.22)	(.02, .03)
Max difference from sample mean income ($0000)=10.1379	$478.07	$3.66	$.3	$-.26	$-.56
	(470.78, 485.36)	(3.29, 4.04)	(.25, .34)	(-.33, -.19)	(-.56, -.55)
G. By difference in age (Z^i variables)					
(At mean $E[S]$, mean of other continuous variables, mean congestion, January, 2013, rural, developed, Puget Lowlands)					
-44 year difference from sample mean age=-44.378689	$120.3	$3.45	$3.63	$-1.98	$.41
	(113.01, 127.59)	(3.08, 3.82)	(3.59, 3.68)	(-2.05, -1.91)	(.4, .42)
-38 year difference from sample mean age=-38.378689	$120.3	$3.45	$3.63	$-1.98	$.41
	(113.01, 127.59)	(3.08, 3.82)	(3.59, 3.68)	(-2.05, -1.91)	(.4, .42)
-28 year difference from sample mean age=-28.378689	$120.3	$3.45	$3.63	$-1.98	$.41
	(113.01, 127.59)	(3.08, 3.82)	(3.59, 3.68)	(-2.05, -1.91)	(.4, .42)
-18 year difference from sample mean age=-18.378689	$120.3	$3.45	$3.63	$-1.98	$.41
	(113.01, 127.59)	(3.08, 3.82)	(3.59, 3.68)	(-2.05, -1.91)	(.4, .42)
-8 year difference from sample mean age=-8.378689	$120.3	$3.45	$3.63	$-1.98	$.41
	(113.01, 127.59)	(3.08, 3.82)	(3.59, 3.68)	(-2.05, -1.91)	(.4, .42)
1.6 year difference from sample mean age=1.6213112	$120.3	$3.45	$3.63	$-1.98	$.41
	(113.01, 127.59)	(3.08, 3.82)	(3.59, 3.68)	(-2.05, -1.91)	(.4, .42)
11.6 year difference from sample mean age=11.6213112	$120.3	$3.45	$3.63	$-1.98	$.41
	(113.01, 127.59)	(3.08, 3.82)	(3.59, 3.68)	(-2.05, -1.91)	(.4, .42)
19.6 year difference from sample mean age=19.6213112	$120.3	$3.45	$3.63	$-1.98	$.41
	(113.01, 127.59)	(3.08, 3.82)	(3.59, 3.68)	(-2.05, -1.91)	(.4, .42)

NOTE: The values in parenthesis represent the values within one standard deviation and include the correlations between the random variables

Table A11: Variations in the value of a birding trip based on observable site attributes (mean of continuous variables, for January 2013, unmanaged site, non-urban developed destination in the Puget Lowlands). Estimates incorporate the 2016 FWS weights.

Simulation	\$ Total WTP for trip
H. By management regime (A_{jt} variables)	
(At mean $E[S]$, mean continuous variables, mean congestion, January, 2013, rural, developed, Puget Lowlands)	
National Wildlife Refuges	\$131.62 (124.33, 138.91)
National Parks, etc.	\$120.9 (113.61, 128.19)
National Forests, etc.	\$120.85 (113.56, 128.14)
Not managed (repeat)	\$120.3 (113.01, 127.59)
I. By urban/rural (a A_{jt} variable)	
(At mean $E[S]$, mean continuous variables, mean congestion, January, 2013, rural, developed, Puget Lowlands)	
Urban	\$120.3 (113.01, 127.59)
Rural	\$126.16 (118.87, 133.45)
J. By congestion/popularity measure (A_{jt} variables)	
(At mean $E[S]$, mean continuous variables, mean congestion, January, 2013, rural, developed, Puget Lowlands)	
10th %ile of eBird congestion=0	\$117.76 (110.47, 125.05)
Mean eBird congestion=.00032	\$120.3 (113.01, 127.59)
75th %ile of eBird congestion=.000221	\$119.52 (112.23, 126.81)
90th %ile of eBird congestion=.000768	\$123.74 (116.45, 131.03)
Max eBird congestion=.02441176	\$117.11 (109.82, 124.41)

NOTE: The values in parenthesis represent the values within one standard deviation and include the correlations between the random variables

Table A12: Variations in the value of a birding trip based on ecoregion (mean of continuous variables, for January 2013, unmanaged site, non-urban developed destination in the Puget Lowlands). Estimates incorporate the 2016 FWS weights.

Simulation	$ Total WTP for trip
K. By Ecoregion (A_{jt} variables)	
(At mean $E[S]$, mean continuous variables, mean congestion, January, 2013, rural, developed, Puget Lowlands)	
Blue Mountains	$116.74 (109.45, 124.03)
Cascades	$117.84 (110.55, 125.14)
Coast Range	$134.48 (127.18, 141.77)
Columbia Plateau	$134.99 (127.7, 142.28)
Eastern Cascades Slopes and Foothills	$120.3 (113.01, 127.59)
Klamath Mtns and CA High N. Coast Range	$163.44 (156.15, 170.73)
North Cascades	$136.94 (129.65, 144.23)
Northern Basin and Range	$135.72 (128.42, 143.01)
Northern Rockies	$135.9 (128.61, 143.19)
Puget Lowlands	$120.3 (113.01, 127.59)
Willamette Valley	$147.81 (140.52, 155.11)

NOTE: The values in parenthesis represent the values within one standard deviation and include the correlations between the random variables

Table A13: Variations in the value of a birding trip based on ecoregion (mean of continuous variables, for January 2013, unmanaged site, non-urban developed destination in the Puget Lowlands). Estimates incorporate the 2016 FWS weights.

Simulation	$ Total WTP for trip
L. By land cover type	
(At means of cont. variables, January, 2013, not managed, rural, Puget Lowlands)	
Developed (baseline)	$120.3 (113.01, 127.59)
Water	$120.79 (113.5, 128.08)
Barren land	$130.1 (122.81, 137.39)
Forest	$127.36 (120.07, 134.65)
Shrubland	$120.3 (113.01, 127.59)
Herbaceous	$119.84 (112.55, 127.13)
Planted/cultivated	$126.03 (118.74, 133.32)
Wetlands	$130.78 (123.49, 138.08)

NOTE: The values in parenthesis represent the values within one standard deviation and include the correlations between the random variables