How to Use Auxiliary General-Population Samples to Reduce Sample Selection Bias in Modeling: Geographic Ranges for eBirders versus All Birders

Trudy Ann Cameron, University of Oregon
Sonja Kolstoe, Salisbury University

March 4, 2020

JEL classification: C83, Q51, Q57, Q58

Keywords: citizen science, volunteered geographic information, sample selection, selection propensities, weighting, bird-watching, eBird, market extent

*Our thanks to Steve Kelling at the Cornell University Laboratory of Ornithology, who generously provided us with access to the eBird data. 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. We are grateful for comments from participants at the Salisbury University Brown Bag Seminar, the 2018 Southern Economic Association in Washington, D.C. and the USDA Workshop Applications and Potential of Ecosystem Services Valuation within USDA—Advancing the Science. Any remaining errors should be attributed to the authors.

How to Use Auxiliary General-Population Samples to Reduce Sample Selection Bias in Modeling: Geographic Ranges for eBirders versus All Birders

ABSTRACT

Volunteered geographic information can be contributed intentionally to citizen science (also known a community science) (CS) projects, or generated incidentally, via smartphone apps or social media. These data can provide a wealth of information about the natural world and associated human preferences or behaviors. However, citizen scientists self-select to participate in each CS project, so the problem of non-representative samples must be addressed before any attempt to scale CS members' preferences or behaviors to the general population. One birding-related CS project is eBird. We survey a sample of eBird members about their levels of engagement with the project as well as some other birding-related behaviors. Additionally, we conduct an independent survey of the general population, where we focus on respondents' propensities to participate in the eBird project, as well as their levels of engagement with the project. We develop an ordered-probit model to explain latent propensities to engage with eBird at different intensity levels, and then transfer the estimated parameters to our separate sample of eBird members, to (a) predict engagement-level propensities, (b) calculate predicted inverse Mills ratio terms, and (c) predict eBird engagement-level probabilities for both samples, which we use to construct population weights for each respondent in the eBird member survey sample. We employ these different methods in a model to explain the spatial extent-of-the-market for one-day bird-watching excursions, based on responses to a question posed only to our eBird member survey sample. An understanding of heterogeneous spatial market extents is important, for example, for destination choice models, where the researcher must define the relevant consideration sets for each person. Market extents are also important in calculating the spatial distribution of the human beneficiaries of conservation measures that affect bird populations.

1 Introduction

"Citizen science" or "community science" (CS) projects recruit volunteers from the general population to help scientists gather data about the natural world. For many CS projects, mobile apps or web portals enable people to store their observations, and to include related geographical information (either automatically or on a discretionary basis) as they contribute to the project.[1] CS databases, with their time stamps and volunteered geographic information (VGI) can sometimes provide a vast amount of individual-level panel data. In other contexts, of course, such data are already being used to track consumer behavior for market research (e.g. SmartGraph), but VGI can also be a resource for research concerning social benefits derived from non-market environmental goods and services.

Some valuable CS/VGI data, for environmental economics research, is related to bird-watching. About 45.1 million Americans observed birds, both around home and away from home, according to the 2016 Survey of Fishing, Hunting and Wildlife-Associated Recreation.[2] The eBird project, managed by the Cornell Lab of Ornithology, is the world's largest biodiversity-related CS undertaking. The eBird review for 2019 reports over 500,000 members collecting 737 million bird observations.[3] Users of eBird's mobile app contributed more than 6.2 million eBird checklists, worldwide, in 2019. Published examples of the use of eBird CS/VGI data to value environmental goods (specifically, avian biodiversity) include Kolstoe and Cameron (2017) and Kolstoe et al. (2018). Additional studies of the demand for non-market environmental goods have also employed, for example, data from social media

[1] CS projects have proliferated because of the growing ability of participants to contribute real-time field observations using convenient smart-phone applications. As of February 2020, there are now more than 2,000 active CS projects, according to the *Citizen Science Association* (see `CitizenScience.Org`), 448 are registered in the federal crowd-sourcing and CS registry (see `CitizenScience.Gov/Catalog/#`).

[2] Our own general-population survey asks about a wide variety of interactions with wild birds, including "pause what you are doing to observe wild birds." Fully 88% of our sample reports doing this at least once over the past year, suggesting that incidental interest in wild birds extends beyond just those people who consider themselves to be bird-watchers. See Appendix A, summarizing levels of activities related to wild birds.

[3] This review is available online at https://ebird.org/news/ebird-2019-year-in-review.

platforms such as Flickr.[4]

The observation and enjoyment of wild birds is thus a popular recreational pursuit that previously left little in the way of a "data footprint" compared to other recreational activities (such as those which require permits, licenses, or entrance fees). CS projects such as eBird, however, provide a wealth of observations about environmental public goods because they present unique opportunities to assemble vast quantities of data on the activities of project participants over time.

The widespread spatial and temporal availability of ecosystem-related CS data like eBird make these data potentially very useful for studies of how ecological changes affect welfare levels for people who value the non-market or recreational ecosystems services affected by these changes. For example, CS data have been used to document the fact that the degradation of ecosystems that constitute wild bird habitat has had a significant adverse effect on wild bird populations. Rosenberg et al. (2019) report an estimated net loss of about three billion birds since 1970—a 29% net loss in roughly five decades. Climate change, urbanization, and changes in land cover in coming decades are likely to produce even more losses, as well as changes in the geographical ranges of individual bird species. Efforts to monitor ongoing changes in wild bird populations (and benefit-cost analysis of alternative policies to protect or enhance these populations) require data on both (a) these bird populations themselves and (b) the preferences of the human population regarding alternative policies with respect to wild birds.

For benefit-cost analysis of environmental policies, it is helpful to know the total social benefits that society will derive. These need to be compared to the social (opportunity)

[4]To date, several different types of analyses have used CS data, including visitation patterns and valuation studies. Wood et al. (2013) first pointed to online social media applications as a possible source of "big data" concerning nature-based tourism and recreation. Visitation studies, such as those done by Sonter et al. (2016) and Wood et al. (2013), employ Flickr data as a source of information for nature-based tourism and recreation. A valuation study using the Flickr data is described in Keeler et al. (2015), who estimate the recreational demand for clean water.

costs that would be incurred with the policy. If a given policy passes a benefit-cost test in the aggregate, it is also helpful to know the incidence of net benefits from the policy, so that the distributional consequences of the policy (i.e. environmental equity) can be assessed. Inferences about total benefits require that estimates of preference parameters be derived for a representative sample from the general population.

Unfortunately, groups of citizen scientists are unlikely to be representative of the general population. Rather than being random samples, these groups must be classed as "samples of convenience."[5] Of course, non-representative voluntary surveys are often used to collect data, and a variety of different methods have been developed to correct for respondents' differing propensities to respond to the survey and therefore to be part of the estimating sample (either via the traditional method of Heckman (1979), a foundational paper for the least-squares context, now cited more than 10,000 times in Web of Science, or via alternative ad hoc approaches as in Cameron and DeShazo (2013) and Johnston and Abdulrahman (2017)). For data collected from CS participants, self-selection bias may arise from the potential correlation between the unobserved components of (a) their outcome variable of interest concerning the environmental good being studied in a CS sample, and (b) their propensities to engage with the CS project to different degrees.

In this paper, we address systematic sample selection that must be considered when harnessing CS/VGI data, to make any potential inferences based on these sample more useful for policy-makers. We propose three candidate strategies. The first strategy is to use, in tandem, a representative sample from a survey of the general population of the U.S. and a completely independent survey of eBird members in the Pacific Northwest. Both of our samples include information from respondents about the extent to which they participate

[5]Samples of convenience can still be immensely valuable if the research object is merely to establish that a relationship among variables *can* exist, but they do not permit the researcher to generalize statistical results from the sample to the overall population. Systematic sample selection has the potential to bias statistical inferences about the general population, so it can be very important to assess, and if necessary to correct for, systematic sample selection.

in the eBird project, so that we can distinguish not only the extensive margin (whether an individual participates in eBird), but how intensively they engage with this CS project. We use our two samples to construct estimated sampling weights that permit corrections for different relative frequencies of engagement with eBird at different levels of intensity, across our general population survey and our survey of eBird members.

Our second strategy, which is a more-structural approach, is to develop "inverse Mills ratio *functions*" that can be estimated using the general population sample and transferred to the eBird member survey sample. A third approach is to transfer "engagement propensity *functions*" from the general population sample to the eBird member survey sample (as a more ad hoc correction approach). Deviations of individual predicted engagement propensities from the sample mean can be allowed to shift the estimated parameters in the outcome equation of interest.

As our outcome equation, for this illustrative analysis, we use one question about birder behavior posed in our eBird member survey, concerning the respondent's personal geographical "extent of the market" for typical one-day birding excursions. The birding destinations within the respondent's market extents can be interpreted as the relevant "consideration set" for their destination choices. Previous research concerning recreational destination choices has tended to use a common market extent for all users, informed by the observed distribution of distances actually travelled across all trips in the data. Here, we seek to identify systematic variations in individual market extents across our sample of eBird members. The fitted market extent function may then be transferable to other samples of birders, if the results are corrected for sample selection bias in our sample of eBird members.

In Section 2 of this paper, we review existing approaches to sample selection correction in the related environmental economics literature. Section 3 reviews the conventional intuition and standard assumptions for two-step structural sample selection corrections with a binary probit selection equation (using inverse Mills ratios). We then outline the po-

tential to use a selection equation, estimated for one sample, to construct an inverse Mills ratio function that can then be transferred to an independent sample. We then generalize the usual binary-probit selection equation to a six-level ordered-probit selection equation as the basis for calculating inverse Mills ratio functions for "structural" selectivity corrections (to accommodate differing intensive margins of eBird participation in the CS sample and the general-population sample). We also consider an ad hoc correction procedure using demeaned response propensities calculated for the eBird sample based on the selection equation estimated for the general-population sample, and the calculation of sampling weights based on ordered-probit models fitted separately to the general-population sample and the eBird member survey sample. Section 4 discusses the estimated selection model, along with the estimated propensities, inverse Mills ratios, and weights. Section 5 discusses the estimated outcome model (to explain variations in the spatial market extent for birding excursions) estimated both naively and with our different types of selection correction strategies. Section 6 provides some further discussion and directions for future research, and Section 7 concludes.

2 Sample-Selection Bias and Correction in the Related Literature

Overcoming the problem of CS samples of human participants being "samples of convenience" requires addressing the sample-selection problem. To date, the literature has focused mostly on strategies to correct for survey "unit" non-responses, or the effects of additional exclusion restrictions based on missing data for key variables. Sample selection correction methods are familiar in the case of continuous outcome variables, as reviewed by Vella (1998) and Wooldridge (2002). But sample-selection correction methods for multiple discrete outcomes are not well developed. For discrete-choice outcome models, Johnston and Abdulrahman (2017) use an ad hoc approach that builds on earlier work by Cameron and DeShazo (2013)

to adjust for response propensity. Kolstoe and Cameron (2017) and Kolstoe et al. (2018) also use this approach, but employ the method to correct only for the individual's propensity to be in the estimating sample drawn from the population of eBird members, not the propensity to be an eBird member in the first place (see the Online Appendix from Kolstoe and Cameron (2017) for details).

Yuan et al. (2015) use a binary probit model to explain systematic selection into their estimating sample and compute an inverse Mills ratio (IMR) used as a regressor in their second-stage conditional logit choice model, to shift the coefficient on the status-quo alternative in their choice sets. However, the IMR transformation is appropriate only when the selection propensity and the (possibly transformed) outcome variable of interest are jointly normally distributed. The Heckman logic for using an IMR thus does not apply when the outcome model is a conditional logit specification. The individual-specific, rather than alternative-specific, IMR variable, in this case, is merely a monotonic transformation of the underlying fitted selection propensity index, and one cannot appeal to the usual bivariate normality assumption for the errors in the two equations to argue that the inclusion of this IMR variable in the outcome equation solves the problem of selection bias. Given that the IMR derived from the selection model is individual-specific but does not vary across alternatives, including it in the utility-difference "index" that underpins a conditional logit model requires that the IMR term be interacted with at least one regressor that actually *does* vary across alternatives. Yuan et al. (2015) choose to interact their fitted IMR term with the status-quo indicator variable, which is non-zero for only one alternative.

Table 1: Levels of participation intensity represented in each sample

CS Engagement Description	CS level	$CS6$ level	Observed for qBus general population sample	Observed for eBird citizen science sample
Does not know eBird	0	1	Y	N
Knows eBird, not a member	0	2	Y	N
Member, reports rarely	1	3	Y	Y
Member, reports $<$ half	1	4	Y	Y
Member, reports $>$ half	1	5	Y	Y
Member, reports almost all	1	6	Y	Y

3 Strategies for Dealing with Systematic Sample Selection in eBird data

Our "eBird member survey" sample is self-selected, consisting only of eBird members who chose to respond to our survey. These birders are also likely to participate in the eBird project with a different mix of levels of engagement than might be expected for members of the general population who participate in birding. Table 1 contrasts our binary indicator for eBird participation, CS, with our six ordered categories of engagement intensity, $CS6$, elicited in the qBus data and in our sample of eBird members.[6]

3.1 Review standard intuition based on binary selection

In framing the *intuition* for our models, we will start by aggregating our six possible engagement levels into just two: non-eBird members and eBird members. This allows us to develop our approach, initially, in terms of a more-familiar binary indicator for "selection into citizen science project participation." Let CS_i be 1 if the respondent is a member of eBird (reporting bird sightings at any frequency) and 0 if the respondent is not a member

[6]The Qualtrics Omnibus surveys have been discontinued, but there remains the equivalent (if somewhat more expensive) option of using the regular Qualtrics panel for a very short regular survey. Alternatively, any other general-population consumer panel could be used.

of eBird (regardless of whether they have heard of the eBird project).[7]

3.1.1 Binary selection and the general population sample

For the $i = 1, ..., N$ individuals in our general population (qBus) sample, we have observations for some people who are members of eBird and other observations for other people who are not. For everyone, we have conformable variables on sociodemographics and income, Z_i, that we will use to explain eBird participation or non-participation, where respondents $i = 1, ..., r$ participate in eBird and respondents $i = s, ..., N$ do not:

$$CS_i = \begin{bmatrix} 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, Z_i = \begin{bmatrix} Z_{11} & \dots & Z_{k1} \\ \vdots & & \vdots \\ Z_{1r} & \dots & Z_{kr} \\ Z_{1s} & \dots & Z_{ks} \\ \vdots & & \vdots \\ Z_{1N} & \dots & Z_{kN} \end{bmatrix}$$

For the qBus sample, we will model the latent propensity to be a member of eBird as $CS_i^* = Z_i\gamma + \eta_i$. Given only these data for the qBus sample, we cannot estimate a second model to explain the outcome variable of interest, $y_i = X_i\beta + \epsilon$, because there are no data for y_i in the qBus sample.

One might wonder whether a researcher could simply pose all the questions on our eBird member survey to a large sample of respondents from the general population. Such a survey would be possible, if we were interested only in the illustrative outcome variable explored in this study. However, such a survey could not take advantage of the wealth of other "passive" birding participation data also available for eBird members. It would be expensive (and less accurate) to attempt to elicit detailed retrospective birding participation information exclu-

[7] We will assume, in this proof-of-concept example, that respondents to the qBus questions are essentially a representative sample of the general population, and respondents to the analogous questions posed to our eBird member survey are essentially a representative sample of eBird members. Unlike our previous research with eBird data, the current analysis is not affected by the need for home address information to allow measurement of actual distances from each person's home to all relevant birding destinations.

sively from a Qualtrics sample. The identity protections in the Qualtrics panel preclude us from being able simply to link eBird members in the qBus sample to their eBird information via their membership number (if the respondent even knows their member number). However, *if, hypothetically,* we could somehow collect from each eBird member who turned up in our qBus sample all of the eBird data and responses to our separate survey of eBird members, we assume that we would specify both the selection equation and the equation to explain some outcome variable of interest, y, for the qBus data as:

$$\begin{aligned} CS_i^* &= Z_i\gamma + \eta_i \\ y_i &= X_i\beta + \epsilon_i \\ (\eta_i, \epsilon_i) &\sim BVN(0, 0, \sigma_\eta, \sigma_\epsilon, \rho) \end{aligned} \tag{1}$$

where the subscript i will be used to denote observations from the qBus sample.

3.1.2 Binary selection and the eBird CS sample

In contrast, for the $j = 1, ..., J$ observations from our sample of eBird members, we have Z_j variables that conform to the Z_i variables in the qBus sample, but we have no information for anyone for whom $CS_j = 0$ (i.e. everyone in this sample is a member of eBird). In this case, the selection process cannot be modeled using the eBird data alone because there is no variation in the selection outcome for this group. However, we have data on an outcome variable of interest for this sample, y_j (the individual's maximum one-way distance for a one-day birding trip), and regressors, X_j, to explain this outcome, where this information is not available for the qBus sample:

$$CS_j = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, Z_j = \begin{bmatrix} Z_{11} & \dots & Z_{k1} \\ \vdots & & \vdots \\ Z_{1J} & \dots & Z_{kJ} \end{bmatrix}, y_j = \begin{bmatrix} y_1 \\ \vdots \\ y_J \end{bmatrix}, X_j = \begin{bmatrix} X_{11} & \dots & X_{m1} \\ \vdots & & \vdots \\ Z_{1J} & \dots & Z_{mJ} \end{bmatrix}$$

For our eBird member survey sample, we will assume that there is the *identical* underlying relationship between CS and the Z variables. Our proposed selection correction method will be appropriate if the identical γ and β parameters *would* apply in this eBird sample (and the same $\sigma_\eta, \sigma_\epsilon, \rho$, as well). If the selection equation *could* be estimated for this sample, the relevant pair of equations would be:

$$\begin{aligned} CS_j^* &= Z_j\gamma + \eta_j \\ y_j &= X_j\beta + \epsilon_j \\ (\eta_j, \epsilon_j) &\sim BVN(0, 0, \sigma_\eta, \sigma_\epsilon, \rho) \end{aligned} \tag{2}$$

Of course, this joint model cannot be estimated because the observed CS_j variable is constant (at 1) in the data from our eBird member survey.[8]

3.1.3 Transferring a fitted selection equation

We do not have y_i and the X_i variables for the eBird members in the qBus general population sample. Instead, we have these variables only for an *independent* sample of eBird members. If we can assume that participation in eBird in the general-population qBus sample follows the same data-generating process as the one that determines participation in eBird among people in our eBird member survey sample, we can assume likewise that $\left(CS_j^*, y_j\right) \sim BVN\left(Z_j\gamma, X_j\beta, 1, \sigma_\epsilon, \rho\right)$ for $j = 1, ..., J$ for members of our eBird sample.[9]

The crux of this approach is that we transfer the $\hat{\gamma}^q$ estimates from the qBus sample

[8]For readers who may wish to review the conventional Heckman two-step sample-selection correction procedure, we provide a brief summary in Appendix B, available from the authors.

[9]Mechanically, it would be possible to pool our two samples and use the combined dataset to estimate one common selection equation. The advantage of using the qBus sample, alone, for the selection equation is that the qBus data represent a random sample from the general population. Pooling it with the eBird sample, however, produces a dataset that no longer represents the general population. One could, potentially, weight the eBird member survey sample according to the proportion of eBird members in the qBus sample, but this would still leave a sample for the selection equation that is not randomly selected from both eBird members and non-members.

to construct a fitted index, $Z_j\hat{\gamma^q}$, to control for the selection process for the eBird sample, even though we have no data from non-eBird members in the eBird member survey sample. The conditional expected value and variance for y_j will be calculated as follows, noting the j subscripts for the eBird data.

$$\begin{aligned} E\left[y_j|y_j\ observed\right] = E\left[y_j|CS_j^* > -Z_j\hat{\gamma^q}\right] &= X_j\beta + \rho\sigma_\epsilon\lambda\left(-Z_j\hat{\gamma^q}\right) \\ &= X_j\beta + \beta_\lambda\lambda\left(-Z_j\hat{\gamma^q}\right) \\ Var\left[y_j|y_j\ observed\right] = Var\left[y_j|CS_j^* > -Z_j\hat{\gamma^q}\right] &= \sigma_y^2\left[(1-\rho^2\delta\left(-Z_j\hat{\gamma^q}\right)\right] \end{aligned} \tag{3}$$

The inverse Mills ratio, $\lambda(-Z_j\hat{\gamma^q}))$ is equal to $\phi(-Z_j\hat{\gamma^q})/(1-\Phi(-Z_j\hat{\gamma^q})) = \phi(Z_j\hat{\gamma^q})/\Phi(Z_j\hat{\gamma^q})$, where $\phi(\cdot)$ is the standard normal probability density function (pdf) and $\Phi(\cdot)$ is the corresponding cumulative density function (cdf). The desired unconditional (i.e. non-systematically selected) expectation for y_j can be simulated, counterfactually, by setting $\rho = 0$, so that $E\left[y_j\right] = X_j\beta$.

It is not uncommon for different discrete-choice samples to have error distributions with different scales. The estimated coefficients in the selection model estimated using the qBus data are known only up to a scale factor. Each γ coefficient is implicitly γ^*/σ_η, where the σ_η applies to the qBus data. If the value of σ_η is larger or smaller for the eBird sample, employing the coefficients estimated on the qBus sample would lead to predicted engagement intensities in the eBird sample that are biased proportionately downward or upward, respectively. With the complex FIML estimation option, discussed later, it may be possible to estimate a scale differential for the eBird data. In this paper, however, we assume the qBus and eBird selection-equation error distributions are identical.[10]

[10]In Section 4.2.1, we assess whether there is compelling evidence that the corresponding engagement propensity equations for the qBus and the eBird samples have parameters that are proportional. We find that the ratio of the corresponding eBird to qBus slope coefficients averages 2.12, but has a standard deviation of 9.73, and the ratios vary from -11.6 to 39.7. Thus any evidence of proportionality is swamped by the differences between the model for the general population and the model for the selected sample of respondents to our eBird member survey.

3.2 Generalizing to a six-level ordered-probit selection model

We now change the model to acknowledge that participation or engagement in eBird is more complex than just a simple "extensive margin" (i.e. yes/no) decision. A standard *binary* probit selection model will fail to capture people's various different levels of engagement intensity. Our qBus questions elicit six levels of eBird engagement intensity, and our eBird member survey questions elicit four corresponding levels of eBird engagement intensity, conditional (obviously) on at least some level of participation in eBird. We now show how standard two-stage sample selection correction methods can be adapted to accommodate an ordered categorical selection variable with six levels (based on an analogous latent continuous "propensity to engage" with the eBird citizen science project). Compared to the standard binary selection equation, our generalization is built around a six-level ordered-probit model for the qBus general population sample and a four-level ordered-probit model for the eBird member survey.[11]

If the joint distribution of the error terms in the underlying (latent continuous) $CS6_i^*$ and y_i variables can be assumed still to be jointly normal, so that the formulas for the moments of a singly-truncated bivariate normal distribution still apply, the same calculated IMR selectivity-correction term based on the adjusted $\hat{\gamma^q}$ ordered-probit estimates from the qBus sample and the Z_j variables from the eBird sample can still be calculated as: $\phi(Z_j\hat{\gamma^q})/\Phi(Z_j\hat{\gamma^q})$. As in the case of a binary probit selection correction, this term can be

[11] An adjustment may be nececessary, to the intercept of the fitted propensity-to-engage with eBird, depending upon which software is used to estimate the ordered probit model. Comparability is necessary if one wishes to compare fitted "propensity" estimates from this ordered-probit selection model to those produced when a standard binary-probit model is used in first step of a typical two-step correction model. In Stata, for example, the *intercept* of the "index", $Z_i\gamma$, is normalized to zero and five distinct "cuts" (thresholds of the latent continuous propensity variable CS_i variable) are estimated. More generally, if there are K levels of an ordered probit model, Stata estimates K-1 "cuts" in the latent continuous propensity variable. To produce fitted propensity estimates that are comparable across the binary probit and ordered-probit versions of the selection equation, it is necessary normalize to zero not the *intercept* in the ordered probit model, but the *"cut" between the second and third categories* in the six-level ordered probit selection model.

appended to the list of regressors, X_j, in the outcome equation of interest.[12,13]

3.3 Ad hoc alternative: Interactions with demeaned propensities

In lieu of a Heckman-type selection-correction model, an ad hoc approach can sometimes be entertained. We might use the estimated engagement propensity model from our first-stage selection model to calculate fitted propensities to engage with eBird at each of six levels (in the qBus sample) or predicted propensities to engage with eBird at four levels (in the eBird member survey sample). For any individual in the eBird member survey sample with a given set of X_j variables, their "predicted engagement propensity" can then be used just like any other type of individual-speciic heterogeneity, such as indicators for gender, age, employment status, or educational attainment.

In a true random sample, every individual in the population of interest is equally likely to show up on the sample, which probably *not* the case with CS samples. To correct for this, the estimated engagement propensity can be demeaned relative to the average engagement propensity for the general population. This fitted demeaned engagement propensity variable can then be allowed to shift all of the key parameters in the outcome model of interest. After estimation of the outcome model, this demeaned response propensity can be counterfactually set to zero, effectively dropping all of the interaction terms in which it is involved. The resulting outcome equation, without these interaction terms, then applies (in principle) to the case where everyone in the estimating sample shares an engagement propensity equal to

[12]It will be appropriate, in future research, to graduate to a full-information maximum likelihood joint estimation of the selection equation and the outcome equation. The ordered-probit form for the selection model is atypical, so no packaged algorithms exist to permit FIML estimation of a "selection-on-ordered-probit" model. We note that there is a packaged algorithm for "ordered probit models with selection," but this is not what we need. That model has a conventional binary-probit selection model and an outcome equation that should be estimated as an ordered-probit model.

[13]In Appendix C, available from the authors, we provide a detailed discussion of the types of outcome models where it is appropriate to use an IMR to correct for sample selection. Here, we merely note that it is appropriate to entertain an IMR term appended to the outcome equation only if the latent continuous dependent variable selection equation and the observed (or latent) continuous dependent variable in the outcome equation can be modeled as involving error terms that are plausibly jointly normally distributed.

the average engagement propensity in the general population—namely, for a "representative" sample.

Ideally, one or more exogenous explanatory variables should be *included* in the $Z_i\gamma$ index that yields the fitted engagement propensities, but *excluded* from the $X_j\beta$ index that represents the conditional expected value of the outcome variable of interest. If there are too few variables in the selection model that do *not* also appear in the outcome equation, the fitted engagement propensity can be very close to a linear combination of the other variables used to explain the outcome variable of interest, and this multicollinearity can inflate the standard errors on the coefficient(s) of interest in the outcome equation and/or render those estimates insufficiently robust to minor differences in the specification of the joint model.[14]

3.4 Complementary method: Weights based on predicted engagement intensities

In this section, we focus on the actual eBird members in the qBus sample, comparing their different participation levels to those in the entire eBird sample. With our eBird member survey data, there is the added concern that the *intensity* with which these survey respondents engage with the eBird citizen science project may not be representative of the distribution of eBird engagement propensities in the general population of the U.S. To address this issue, we consider how to develop weights for each of the four levels of participation intensity among these eBird members. We base our weights on the *fitted probabilities* of a respondent being each of the four engagement intensity bins in each sample.

For the qBus sample, we estimate an ordered probit model for all six possible bins and calculate a set of fitted probabilities for each bin for each person, conditional on the Z_i vector for that person. Call these probabilities $\hat{p}_{ki}, k = 1, ..., 6$. We then make two cal-

[14] In two-step selection-correction models, of course, the non-linear transformation used to create the inverse Mills ratio variable breaks this collinearity, but it is considered poor form to identify the second-step parameters solely because of this transformation.

culations for each respondent in the eBird member survey sample. First, we use the fitted participation-intensity model, estimated using the qBus data, to predict (for the eBird member survey sample) the individual-specific set of six probabilities associated with each of the six participation-intense bins (even through nobody in the eBird sample is in non-participation bins 1 or 2). Call these fitted probabilities $\hat{p}^*_{kj}, k = 1, ..., 6$, Then, we use the eBird sample independently, with its four possible participation-intensity bins. We estimate a four-level ordered probit model using just the eBird member survey sample and calculate four fitted probabilities, which we will call $\hat{q}_{kj}, k = 3, ..., 6$.

The next step is to assign weights to each respondent in the eBird dataset. Theses weights serve to scale the fitted probability of an individual being in their observed engagement-intensity bin to match the fitted probability in the population (i.e. the qBus sample). First consider a counterfactual case where everyone in the qBus and eBird samples has been drawn from the general population and everyone had identical Z variables. Then we would expect, across the two samples, to have identical proportions of people in each of the six intensity bins. However, since nobody in the eBird sample is observed in bins 1 or 2, we must focus on the proportion of the participation-intensity distribution corresponding to eBird membership. For the qBus sample, we should consider the probabilities of being in bins 3 through 6 for the qBus sample, *conditional* on the probability of being in one of those four bins. Thus we define $\hat{p}_{kj} = \hat{p}^*_{kj} / \left(\hat{p}^*_{3j} + \hat{p}^*_{4j} + \hat{p}^*_{5j} + \hat{p}^*_{6j}\right)$.[15]

When we allow for potentially very different distributions of the explanatory variables Z_i and Z_j for the engagement-intensity model in the qBus and eBird samples, it is readily apparent that we should not use simply the differing observed *proportions* of people in each bin in the two samples to construct weights to be used in estimating the outcome model. Our preferred approach would be more akin to the common method of constructing *exogenous*

[15] An alternative would be to attempt to fit a four-level ordered probit for only the qBus respondents, but there are relatively few eBird members in the qBus sample.

weights based on age brackets or gender. We wish to allow multiple factors to affect *expected* levels of engagement intensity for each eBird respondent. Consequently, we weight each observation in the eBird sample by $\hat{p}_{kj}/\hat{q}_{kj}, k = 3, ..., 6$, normalized so that these weights sum to the sample size for the eBird sample. Use of these fitted probabilities recruits all of the factors that capture heterogeneity in response propensities (the Z_i and Z_j data) to build the empirical weights, rather than just 0/1 group membership indicators.

4 Selection Model: Implementation

4.1 Available variables

Our selection equations, either binary probit or ordered probit, require conformable measures of the Z_i and Z_j variables (i.e. these variables must be measured in the same way for the qBus and eBird member survey datasets). For the qBus data, unless one wishes to pay for additional questions, it is necessary to make do with the default set of sociodemographic and geographic characteristics that are available for all qBus panelists, as collected by Qualtrics. Thus we aggregate both the qBus and the eBird member survey Z variables to the same level, yielding conformable sets of indicator variables for the different levels of each of seven individual characteristics that can be allowed to influence either engagement or non-engagement in the eBird CS project (in the binary selection model) or the different intensities of engagement (in the more-general ordered-probit selection model).

The available variables for our selection model, conformably aggregated, are as follows:

- Age (7 brackets)
- Race (4 groups)
- Ethnicity (2 groups)
- Income (5 brackets)

- Geography (4 regions)
- Employment status (5 categories)
- Educational attainment (5 levels)

Across observations with no missing values, for the qBus data (N=4161) and for the eBird member survey data (J=1081), Table 2 summarizes the proportions of observations in each set of indicator variables. Note that respondents in the qBus sample have two more options than respondents in the eBird member survey sample. The qBus respondents can also choose the engagement categories "Unfamiliar with eBird CS project" or "Heard of eBird but not a member." As a consequence, it is not possible to compare directly the proportions in the other four eBird-member engagement-intensity categories across the qBus and eBird samples. If we calculate the conditional distribution solely for engagement levels 3 through 6 (where a qBus respondent is at least a member of eBird), then the pairs (proportion qBus, proportion eBird) for these four categories are (0.273,0.398), (0.252,0.275), (0.265,0.179), and (0.210, 0.146). While these relative proportions differ, it is also the case that the *types* of people who respond to the qBus survey may also differ from the types of people who are enrolled in eBird and responded to our survey of a random sample only from the eBird population.

As detailed in Table 2, there are a number of notable differences between our two sample. Our eBird member survey respondents are somewhat more likely to be female and to be older. They are also more likely to identify as White.[16] A considerably larger share of the eBird member survey sample, relative to the qBus sample, did not provide any income data (23.7 percent). Everyone in the eBird member survey sample is from the states of Washington and Oregon, whereas the qBus sample is nationwide. Compared to qBus respondents, about

[16]Indeed, the proportions of Black and Asian and Hispanic eBird respondents are all less than 1 percent. It may not be prudent to expect so few respondents in those three groups to be representative of the level of engagement for those groups, so we will not use the Race or Ethnicity indicators in our specifications.

Table 2: Descriptive statistics (proportions) for first-stage Bird engagement intensity Availability indicators are proportions of total sample; group shares are proportions of available data

	qBus	eBird member survey (prop. if known)
eBird engagement intensity:		
1=Unfamiliar with eBird CS project	0.802	–
2=Heard of eBird but not a member	0.083	–
3=eBird member, but report rarely	0.031	0.391
4=eBird member, report less than 1/2 of birds	0.029	0.280
5=eBird member, report more than 1/2 of birds	0.030	0.177
6=eBird member, report almost all birds	0.024	0.152
Explanatory variables (sets of indicators):		
• Travel 1 mile data available	0.442	0.995
Some 1+ mile trip for birds	0.846	0.939
• Gender data available	1.000	0.994
Gender = male	0.489	0.425
Gender = female	0.511	0.570
• Age data available	1.000	0.993
Age 24 years or less	0.125	0.018
Age 25 to 34 years	0.224	0.065
Age 35 to 44 years	0.196	0.089
Age 45 to 54 years	0.135	0.145
Age 55 to 64 years	0.175	0.309
Age 65 years and up	0.145	0.367
• Income data available	1.000	0.804
Income less than 25K	0.179	0.058
Income 25 K to 50 K	0.219	0.163
Income 50 K to 75 K	0.189	0.186
Income 75 K to 100 K	0.141	0.139
Income 100 K or more	0.272	0.258
• Region data available	1.000	1.000
Region: West	0.225	1.000
Region: Northeast	0.186	–
Region: Midwest	0.217	–
Region: South	0.372	–
• Employment status data available	1.000	0.849
Employment status: Full time	0.473	0.411
Employment status: Part time	0.132	0.068
Employment status: Looking for work	0.057	0.006
Employment status: Unemployed	0.145	0.056
Employment status: Retired	0.193	0.414
• Education data available	1.000	0.976
Education: High school	0.226	0.035
Education: Some college	0.356	0.154
Education: College grad	0.263	0.281
Education: Masters degree	0.118	0.387
Education: Doctoral degree	0.038	0.118
Observations	4161	1081

twice as many eBird member survey respondents are retired. Finally, the eBird member survey sample reports higher education attainment.

4.2 Estimation results

4.2.1 Ordered-probit qBus propensities to engage with eBird

The qBus sample has virtually complete data for its Z_i variables, but there are more missing values for the Z_j variables that our eBird member survey attempted to elicit from each respondent. We seek to transfer from the qBus sample to the eBird member survey sample the richest possible specification of the ordered-probit selection model, consistent with the data available for each respondent in the eBird member survey sample. Appendix D, available from the authors, explains in more detail our strategy for recruiting all available data for every eBird member when estimating ordered-probit models using the qBus sample that can be used to predict response propensities and inverse Mills ratio values for each respondent in the eBird member survey sample. To accommodate all of the patterns of missing values encountered in our eBird member survey data, we need to estimate 24 different ordered-probit specifications using the qBus data.

In Appendix E (available from the authors), includes eight analogous tables that contain the 24 different specifications of the ordered probit model, estimated using the qBus sample, to explain the *six* possible engagement levels that could be selected by the qBus respondents. In many empirical models, it is possible to use indicators for "variable value non-missing" in a specification where a given variable is unavailable for some fraction of respondents. However, because the qBus sociodemographics variables are essentially complete, it is not possible (in the qBus sample) to identify coefficients on such indicators for "variable value non-missing." Given that the estimates from the qBus dataset are to be transferred to the eBird data, where variable values are missing in many cases, it would be necessary to have estimates for

those data-availability indicators.[17]

Appendix F (likewise available from the authors) contains a second set of eight tables that show the corresponding set of 24 different specifications of the ordered probit model, estimated using the eBird member survey sample, to explain the *four* engagement levels that could be selected by our eBird respondents. These models are required for the construction of our weights. Recall that the default parameterization of the thresholds between intervals (e.g. in Stata) is not comparable between the six-interval qBus models and these four-interval eBird member survey models. E.g. for the qBus models, "cut5" is the threshold between engagement levels 5 and 6, whereas for the eBird member survey models, "cut3" is the corresponding threshold between these same two levels.

Table 3 presents results for the most complete specification that can be estimated using either the qBus or the eBird member survey samples. Note that the complete set of Z_j variables is available for only 725 of the 1081 eBird member survey respondents. At the other end of the spectrum of missing values in the eBird member survey sample, Table 4 shows a model that can be estimated for *every* respondent in the eBird member survey sample, even if they answered none of the questions needed for the Z_j vector. This model can use all 1081 eBird member survey respondents who answered the question about our outcome variable, but must employ fewer variables.

To predict participation intensities for respondents to our eBird member survey, we use the γ^q coefficients estimated from the qBus specification in the first column of estimates in Table 3 and Table 4 (and the rest of the 24 models). We use these parameter estimates to calculate predicted engagement-level probabilities, as well as predicted engagement inten-

[17] One could pretend that variable values were missing in the qBus data by randomly converting some of the observed values to missing. However, it is unlikely that variable values are missing at random in the eBird member survey data, so this approach may not be particularly appropriate. We definitely do not want to drop observations from the eBird member survey sample because the resulting subset of the data could be a biased sample with respect to eBird participation intensity. Our recourse is to retain every possible eBird respondent, but to employ different models to generate predicted values for participation intensity $Z_j\gamma^q$, according to whichever subset of the Z_j variables is available.

sities and inverse Mills ratio terms to be used for sample-selection corrections in outcome equations that use only the eBird member survey data. The second column of parameter estimates in each of these tables is estimated using the eBird member survey data alone. We need these models to calculate fitted engagement-level probabilities within that sample. To construct our weights to be applied to each observation in the eBird member survey sample, we require engagement-level probabilities for our eBird sample that are (a) predicted, based on estimates transferred from the qBus sample, and (b) fitted, based on estimates from the eBird sample alone.

In comparing corresponding model for each sample in Table 3, we note some common tendencies. If the individual reports having traveled more than one mile to see birds, they have a statistically significantly higher propensity to engage with eBird, as are male respondents. In the qBus sample, relative to the baseline propensity for the 45-to-64-year-old age group, younger age groups have higher propensities to engage with eBird, and older age groups have lower propensities. Income, assigned to five brackets, does not influence eBird participation intensities in either sample. However engagement intensities are statistically significantly higher in the Northeast region of the U.S. In the qBus sample, compared to those who are working full time, people who are unemployed, looking for work, or retired have lower propensities to engage with eBird (controlling for age group).

For education levels, the omitted category is people who are college graduates. In the qBus sample, compared to this group, people who have only "some college" have a lesser propensity to engage with eBird, but those with either a masters degree or a doctoral degree have higher propensities to engage with the project. In each of our samples, there are very few respondents whose educational attainment is only High School, so the statistical insignificance of the corresponding coefficient is unsurprising. The effect of having some college education is statistically significant at the 10% level or better and similar across samples, as is the effect of having a doctoral degree.

Table 3: Ordered-probit engagement-level models with maximum heterogeneity: six-level model for qBus sample and four-level model for the subset of 725 respondents to our eBird member survey sample with complete data for same specification

	Full qBus sample		eBird subsample (non-missing data)	
Engagement-level indicator				
Travel 1 mile data available	0.536***	(0.0486)	unident.[a]	
Some 1+ mile trip for birds	1.443***	(0.107)	1.747***	(0.306)
Gender = male	0.281***	(0.0484)	0.367***	(0.0867)
Age 24 years or less	0.776***	(0.0942)	0.678*	(0.403)
Age 25 to 34 years	0.738***	(0.0825)	0.189	(0.188)
Age 35 to 44 years	0.534***	(0.0843)	0.183	(0.168)
Age 55 to 64 years	-0.409***	(0.106)	-0.164	(0.136)
Age 65 years and up	-0.466***	(0.129)	-0.0397	(0.171)
Income less than 25K	-0.0238	(0.0820)	-0.111	(0.204)
Income 25 K to 50 K	-0.0160	(0.0744)	0.186	(0.128)
Income 75 K to 100 K	0.0856	(0.0818)	-0.209	(0.133)
Income 100 K or more	-0.000760	(0.0745)	-0.0303	(0.118)
Region: Northeast	0.255***	(0.0702)	unident.[b]	
Region: Midwest	-0.00564	(0.0717)	unident.[b]	
Region: South	0.0616	(0.0630)	unident.[b]	
Employment status: Part time	-0.0819	(0.0721)	-0.169	(0.161)
Employment status: Looking for work	-0.200**	(0.102)	-0.519	(0.465)
Employment status: Unemployed	-0.320***	(0.0779)	-0.109	(0.193)
Employment status: Retired	-0.171*	(0.103)	-0.318**	(0.138)
Education: High school	-0.0408	(0.0719)	0.163	(0.263)
Education: Some college	-0.180***	(0.0640)	-0.235*	(0.138)
Education: Masters degree	0.368***	(0.0780)	0.0675	(0.103)
Education: Doctoral degree	0.267**	(0.118)	0.284**	(0.144)
cut1	2.847***	(0.154)	1.240***	(0.333)
cut2	3.287***	(0.156)	2.135***	(0.336)
cut3	3.515***	(0.157)	2.742***	(0.338)
cut4	3.790***	(0.159)	n/a[c]	
cut5	4.238***	(0.164)	n/a[c]	
Observations	4161		725	
Max. log-likelihood	-2859.07		-902.54	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ [a]Too few zero cases to identify. [b]Var = 0 for all. [c]Only 4 levels.

Table 4: Ordered-probit engagement-level models ignoring heterogeneity in participation propensities: qBus sample and full eBird sample. Fitted values to be used only for eBird observations without needed regressors, other than implicit Region variable (= West for all eBird respondents)

	Full qBus sample		Full eBird sample	
Region: Northeast	0.215***	(0.0647)	unident.[a]	
Region: Midwest	-0.0754	(0.0656)	unident.[a]	
Region: South	-0.0453	(0.0576)	unident.[a]	
cut1	0.861***	(0.0456)	-0.276***	(0.0387)
cut2	1.217***	(0.0473)	0.444***	(0.0395)
cut3	1.399***	(0.0488)	1.029***	(0.0464)
cut4	1.620***	(0.0513)	n/a[b]	
cut5	1.995***	(0.0581)	n/a[b]	
Observations	4161		1081	
Max. log-likelihood	-3274.80		-1422.62	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ [a]Var = 0 for all. [b]Only 4 levels.

4.2.2 Transferring qBus selection model to eBird member survey sample

One innovation in this paper is the specification of an apparently new sample selection model where the selection equation is an ordered probit model. However, if a conventional selection equation is desired, the six categories in our ordered probit selection model can be aggregated into a simple binary indicator that is 0 for people who are not eBird members (engagement levels 1 and 2) and 1 for eBird members (engagement levels 3, 4, 5 and 6). We can estimate analogous propensity indexes, $Z_i\hat{\gamma}^q$, for either a six-level ordered-probit eBird engagement selection model or a binary probit eBird engagement selection model.[18]

For the qBus dataset, which includes both members and non-members of eBird, Figure 1 shows the joint distribution of the (adjusted) fitted propensity index, $Z_i\hat{\gamma}^q$, for our new ordered-probit selection equation as well as the fitted propensity index, $Z_i\hat{\gamma}^q$ from a conventional binary-probit selection model using the same qBus data. Our ordered-probit

[18]The only adjustment that is necessary for these propensity index measures to be comparable is that the threshold ("cut") between levels 2 and 3 in the ordered probit model must be normalized to zero (instead of Stata's default normalization that sets the intercept of the index to zero).

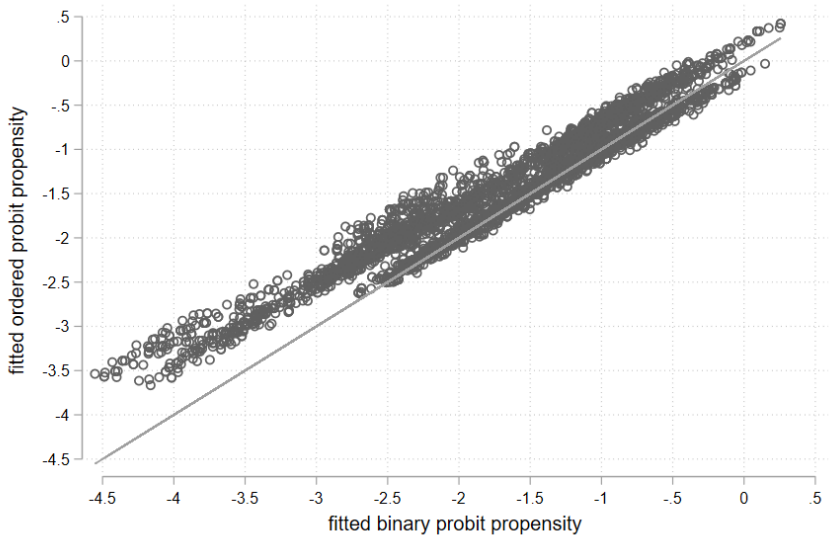

Figure 1: *qBus sample only:* compare within-sample engagement propensities calculated from ordered-probit parameters and within-sample engagement propensities calculated from binary-probit parameters

selection model recruits more information in its multiple categories among non-members and members of eBird in the qBus sample. The propensity index from our new ordered-probit selection model is somewhat higher than the index for the conventional binary model among people with low propensities to belong to eBird, but the upper part of the joint distribution coincides closely.

Our ordered-probit selection model (with the renormalization adjustment) thus tracks the conventional binary-probit selection specification closely when each is applied to the same sample of qBus respondents. Crucially, we assume that the same data-generating process would apply to the general population from which our survey sample of eBird members is drawn. Thus we assume that for each individual in our eBird member survey sample, we can transfer the relevant set of $\hat{\gamma}^q$ parameters estimated using the qBus data. The qBus selection model to be transferred needs to be estimated using the same set of non-missing regressors, so that we have exactly the necessary information to calculate a predicted propensity index,

$Z_j\hat{\gamma}^q$, that exploits as much information as we possess about that individual eBird member's sociodemographic characteristics.

Either none, some, or all of the selection-equation regressors may be missing for some eBird member survey respondents. Recall that an ordered-probit model estimated with *no* regressors will return just the thresholds of a univariate propensity distribution and will yield only as many different values of the probability in each interval as there are intervals (since there is no heterogeneity in these probabilities).[19]

Figure 2 displays histograms for the marginal distributions (i.e. the degree of heterogeneity) across respondents in the predicted probabilities of being at each of the four engagement levels (3, 4, 5 and 6), *conditional* on the individual being at observed to participate at one of these four levels (i.e. conditional on being a member of eBird). The left column of distribution in Figure 2 shows the *fitted* individual probabilities of being at each engagement level for the qBus sample. The right column of distributions in Figure 2 shows the *predicted* probabilities of being at each engagement level for the eBird member survey sample, calculated using the parameters of the relevant ordered probit model estimated using the qBus sample.

To reveal the differences between engagement propensities based on the binary probit model and the ordered-probit model, at the individual level for the qBus data only, Figure 1 plots the re-normalized (adjusted) ordered-probit engagement propensities against the binary-probit engagement propensities. The diagonal represents equal estimates. At lower levels of engagement propensity, the ordered probit model produces a fitted engagement propensity that is somewhat larger than the one produced by the binary probit selection

[19]The selection process for the eBird sample is the compound effect of selection into eBird and selection into our sample of survey respondents. To the extent that the sample from our eBird member survey does not represent the population of eBird members, there may be a second layer of selection to consider. We ignore that additional complexity in this paper, relying instead upon the six-level order-probit selection model to correct to some extent, for the different proportion of people with engagement levels 3 through 6 in the general population versus our eBird sample.

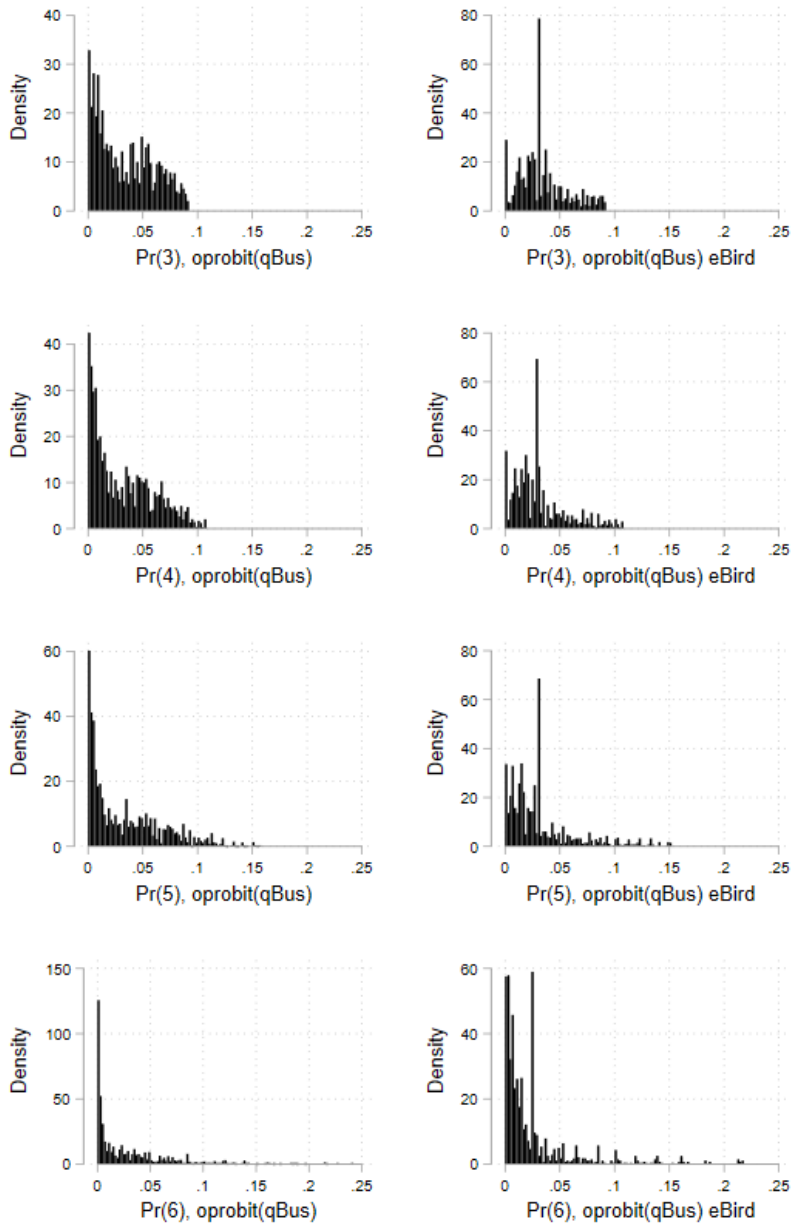

Figure 2: *Left column:* Predicted probabilities for engagement levels 3, 4, 5, and 6, conditional on being at one of these levels (i.e. on being an eBird member), estimated for qBus sample. *Right column:* Analogous probabilities *predicted* for eBird sample. eBird predictions use qBus specifications corresponding to available variables for each eBird individual; spikes correspond to respondents in the eBird sample who provided no sociodemographic variables.

model.

We can also consider the differences in the marginal distributions of the two calculated IMR terms, *for individuals in our eBird member survey sample*, based on our two different selection models estimated using the qBus data: (1) the binary-probit selection model and (2) the re-normalized (adjusted) ordered-probit selection model, Figure 3 shows these two IMR distributions for our eBird member survey respondents. The greater heaping in the predictioned IMR values based the probit model again reflects the lesser heterogeneity among calculated IMR values when an eBird member survey respondent has missing data for all sociodemographic questions used in the most general model.

4.2.3 Calculating engagement-intensity population weights for eBird member survey sample

Selection-correction models allow the analyst to accommodate the possibility that there is a correlation between the *unobserved error term* in the process that leads to an individual's presence in the estimating sample and the *unobserved error term* in the process that generates the outcome variable for that individual. However, it is also possible that the *observable heterogeneity* across the individuals who show up in the estimating sample could be different from the heterogeneity in the general population. In such a case, researchers often consider the use of exogenous weights. Weights are used to scale the relative frequency of people of different types in the estimating sample so that group proportions more closely match the group proportions in the population. With a fully representative estimating sample, each observation represents an equal number of people in the population as a whole, so that average preferences in the sample (for example) should be the same as average preferences in the population. If each observation in the sample represents a very different number of people in the population, then estimated average preferences are less likely to scale up to the general population.

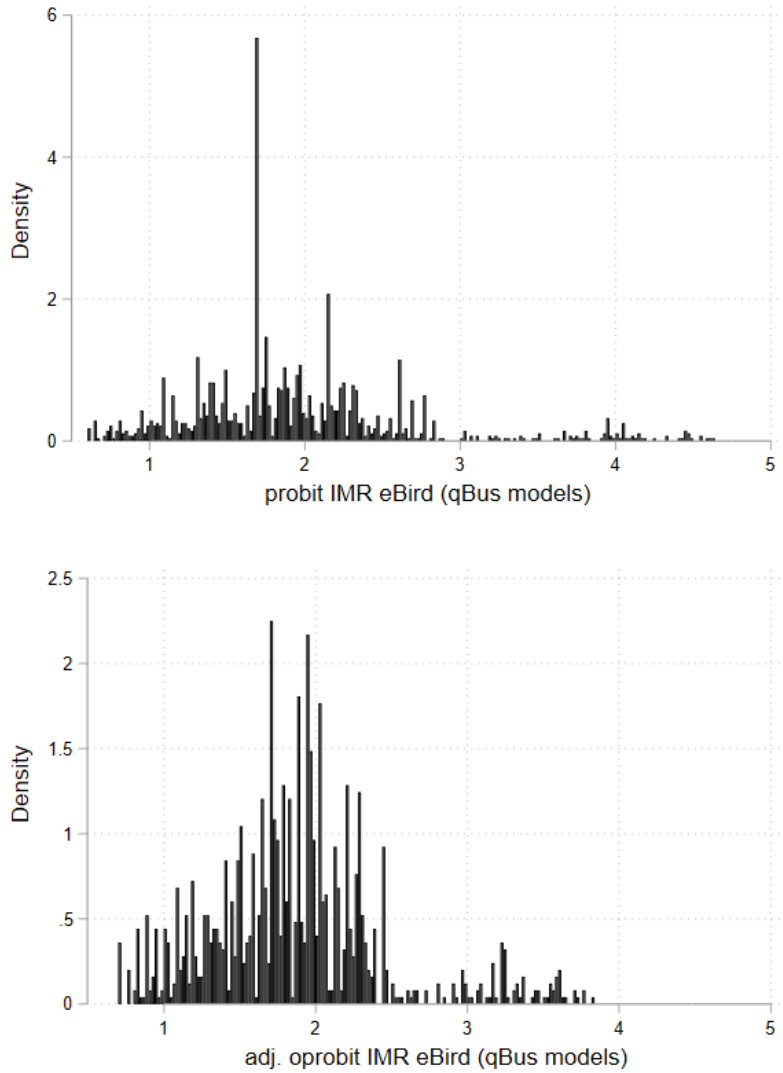

Figure 3: *eBird sample: predicted* distributions of binary-probit-based IMRs and ordered-probit-based IMRs, both calculated from qBus parameters for the most-detailed specification consistent with missing Z_j data in the eBird member survey sample. Prominent spikes in each case correspond to observations with few or no available Z_j variables.

We could use the observed and undifferentiated proportions at each engagement level in the two samples, (qBus, eBird) = (0.273,0.398), (0.252,0.275), (0.265,0.179), and (0.210, 0.146), as mentioned in Section 4.1. But these proportions do not allow for the posssibility of systematically different mixes of people in the two samples. Thus we explore how to compute weights based upon the winthin-sample predicted probabilties that each respondent participated in eBird at each of the four possible levels.

For our eBird data, we are concerned that (a) the relative *proportions* of respondents in our eBird sample who engage with the project at different levels might differ from (b) the corresponding proportions in the population of eBird members who turn up in a random sample from the general population (i.e. our qBus sample). We again transfer our qBus parameter estimates for the six-level ordered-probit models with the sociodemographic characteristics of each person in our eBird member survey sample to calculate the predicted individual conditional engagement-level *probabilities* for respondents in our eBird member survey sample, as are shown in the second column of histograms in Figure 2. But then we also use our eBird member survey sample, *independently*, to estimate four-level ordered-probit models for engagement levels 3, 4, 5 and 6, and calculate predicted probabilities for these four engagement levels based on those parameters. We treat these two sets of predicted probabilities as the "expected" probabilities and the "observed" probabilities in the eBird member survey sample.

We construct our weights for each observation in the eBird sample by considering the observed engagement level for that person. We then generate a weight that reflects (a) the out-of-sample *predicted* probability that a person with these same characteristics would be at that level of engagement in the general population (qBus) sample, in ratio to (b) the within-sample *fitted* probability, estimated using the eBird member survey data, that they are at their observed level of engagement. We scale these weights so that they sum to the sample size for the eBird member survey. Figure 4 shows the resulting distribution of weights

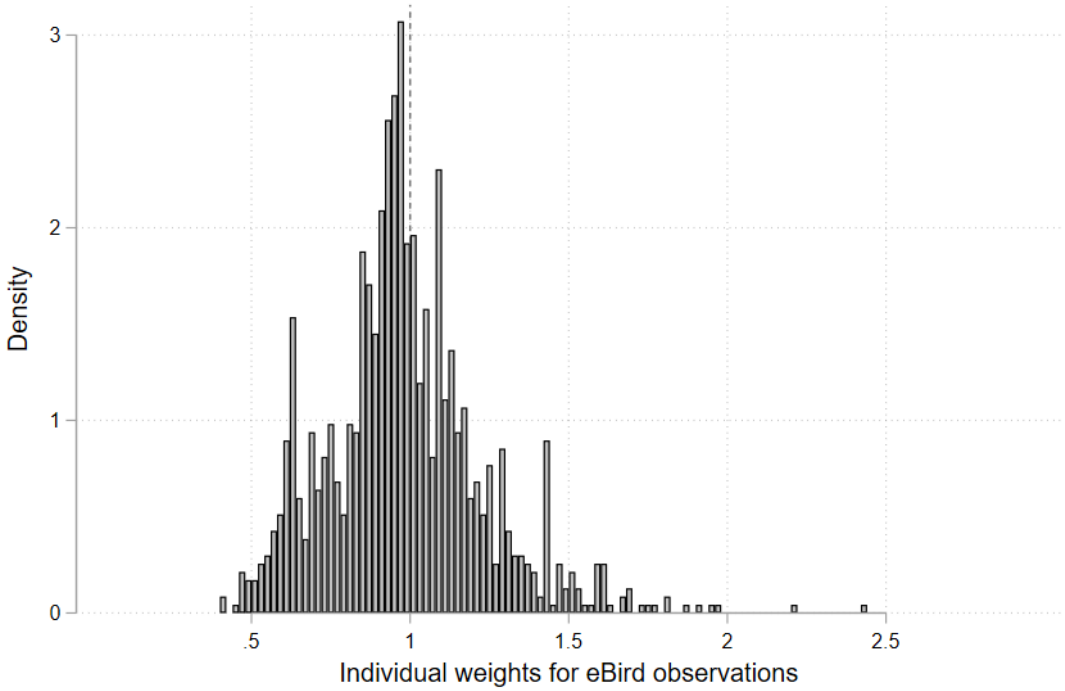

Figure 4: Distribution of weights across eBird member survey observations, where these weights serve to match engagement-intensity probabilities in the eBird member survey sample to engagement-intensity probabilities in the general-population qBus sample (six outlier weights, between 2.66 and 7.93, are not shown)

for use in estimation of the outcome model that uses only the eBird member survey data.

5 Outcome: Market Extent for Birding Excursions

5.1 Available variables for outcome model

This section illustrates the use of our predicted, rather than estimated, IMR terms in a model, to be estimated using only our eBird member survey sample, that explains the maximum distance that people state they would typically consider traveling for a one-day birding excursion. This survey response reveals the geographical "market extent" for birding excursions, for different types of people. The summary statistics for the eBird data used for these

Table 5: Descriptive statistics: Variables for outcome model, elicited from eBird member survey sample for n = 1,081 respondents who answered the question about maximum one-way distance for a birding day-trip

Variable	Mean	Std. Dev.	Min	Max
Dependent variable:				
Max One-Way Distances (miles)	83.28	58.10	10	225
Explanatory variables:				
1(Employed)	0.370			
1(Have Income Data)	0.80			
Income (in $10K, if known)	8.72	5.32	1.8	22.5
1(Female)	0.57			
1(Age<45)	0.170			
1(Age>64)	0.367			
1(Education: Some College or Less)	0.19			
1(Education: Some Graduate School or More)	0.51			
1(Not Interested in Perching Birds)	0.057			
1(Not Interested in Wading Birds)	0.0721			
1(Not Interested in Waterfowl)	0.079			
1(Not Interested in Other Game Birds)	0.111			
1(Not Interested in Birds of Prey)	0.072			
IMR - Binary Probit	2.20	0.67	0.831	4.66
IMR - Ordered Probit	0.33	0.24	0.016	1.46
Demeaned Engagement Propensity	0.450	0.606	-1.51	1.95

models are in Table 5.

The variables available to use as regressors in our market extent model are different than those used in either the binary-probit or ordered-probit models to explain levels of eBird engagement intensity. For our engagement intensity models, we were limited to variables that were available and could be measured conformably for both the qBus sample and the eBird member survey sample because we needed to perform a "model transfer." We have richer data from the eBird member survey that was not available in the qBus sample. For example, our eBird member survey elicits income in much finer brackets than we could use in the engagement intensity models, so we convert the income bracket data into an approximate continuous income variable. We also take advantage of eBird survey data concerning eBirders' interest in different species categories. For each of five categories of

bird species, between 6% and 11% of eBirders report that they have no interest in that category. The least popular category in our eBird member sample, for example, is "game birds other than waterfowl" (e.g. pheasants, turkeys, grouse or partridges).

The question of the effective geographical extent of the market has bearing on the potential "active use" versus "passive use" (option, bequest, or existence) values of environmental projects to protect or enhance local wild bird populations, or the welfare impacts of wholesale shifts in the geographic ranges of different bird species in response to climate change. (Birds are highly mobile and are likely to relocate more quickly than most bird-watchers, especially if climate change accelerates.) We thus adopt a strategy akin to that used in the transportation geography or marketing literature where we seek to understand the market range or extent, as defined by the maximum distance each consumer will travel to reach a service (Rodrigue et al., 2016, pg. 346). The definition of market extent in the non-market valuation literature often refers to the geographic area over which substantial numbers of non-zero WTP values are observed for the environmental resource. These WTP values can be aggregated over the population in this relevant market to estimate the overall value of the resource to that particular population (De Valck and Rolfe, 2018; Loomis, 1996).

5.2 Estimation results

Our dependent variable for these models, the maximum distance considered for a typical one-day birding trip, is elicited in "distance brackets" in our eBird survey. Given that these distances are strictly positive and the boundaries for these brackets are known, a reasonable estimation method assumes that the logarithm of the latent continuous dependent variable is conditionally normally distributed. An interval regression model can then be estimated by maximum likelihood methods.[20]

Model 1 in Table 6 is a naive specification to explain maximum willingness to travel to

[20]Stata's intreg estimator is available for such models.

see birds, with no corrections. The other columns show several alternative types of corrected models. Model 2 is an otherwise naive specification that employs only our constructed weights. Models 3 through 5 continue to employ these weights. Model 3 includes an IMR variable based on a conventional binary-probit selection equation, and Model 4 employs our novel ordered-probit selection equation. Finally, Model 5 shows the results from our ad hoc selection-correction strategy that interacts each main determinant of market extent with a demeaned predicted engagement propensity (based on our ordered-probit selection specification estimated on the qBus sample and transferred to the eBird member survey sample).[21]

[21]For Models 2 through 5 in Table 6, the weights are designed to correct for differences in observable determinants of engagement intensities among eBird members, to make this particular sample of eBird members more representative of engagement intensities among eBird members in the general population (i.e. in the qBus sample, in this case). The complete set of parameter estimates, including statistically insignificant coefficients on several interaction terms in Model 5, can be found in Appendix G, available from the authors.

Table 6: Market extent models with engagement-intensity weights and either sample selection corrections or interactions between all regressors and demeaned ordered-probit selection propensity. Statistically insignificant interactions are not displayed. Dependent variable: logarithm of maximum one-way distance willingly traveled on a typical birdwatching day-trip.

	(1) Naive	(2) Weights only	(3) Probit IMR	(4) Ordered probit IMR	(5) Demeaned propensity[a]
Explanatory variables:					
1(Employed)	-0.0400	-0.104	-0.140**	-0.134**	-0.211**
	(0.0693)	(0.0783)	(0.0674)	(0.0675)	(0.0894)
1(Have Self-Reported Income Bracket)	-0.420***	-0.456***	-0.370***	-0.420***	-0.539***
	(0.127)	(0.149)	(0.132)	(0.133)	(0.174)
ln(Income in 10Ks, If Reported)	0.210***	0.226***	0.173***	0.185***	0.230***
	(0.0532)	(0.0608)	(0.0538)	(0.0537)	(0.0811)
1(Female)	-0.126**	-0.0842	0.124*	0.103*	0.145
	(0.0587)	(0.0612)	(0.0648)	(0.0627)	(0.0973)
1(Age $<$ 45)	0.175**	0.205**	-0.292***	-0.158	0.371
	(0.0831)	(0.0867)	(0.110)	(0.0963)	(0.298)
1(Age $>$ 64)	-0.0407	-0.0587	0.315***	0.129*	0.124
	(0.0709)	(0.0768)	(0.0883)	(0.0728)	(0.0877)
1(Education - Some College or Less)	-0.0973	-0.0868	0.0527	0.0899	0.106
	(0.0842)	(0.0945)	(0.0927)	(0.0942)	(0.0977)
1(Education - Some Grad School or More)	0.0700	0.0481	-0.198**	-0.119*	-0.237**
	(0.0672)	(0.0677)	(0.0776)	(0.0712)	(0.105)
1(Not Interested in Perching Birds)	-0.514***	-0.521**	-0.428**	-0.471**	-0.402**
	(0.169)	(0.204)	(0.200)	(0.200)	(0.196)

Continued on next page

1(Not Interested in Wading Birds)	-0.253	-0.254	-0.0743	-0.0352	-0.111
	(0.180)	(0.188)	(0.197)	(0.200)	(0.208)
1(Not Interested in Waterfowl)	-0.216	-0.255	-0.0855	-0.0756	-0.178
	(0.166)	(0.178)	(0.186)	(0.188)	(0.198)
1(Not Interested in Other Game Birds)	-0.447***	-0.427***	-0.492***	-0.503***	-0.480***
	(0.117)	(0.122)	(0.124)	(0.126)	(0.138)
1(Not Interested in Birds of Prey)	-0.243	-0.170	-0.262	-0.244	-0.217
	(0.174)	(0.179)	(0.173)	(0.173)	(0.168)
Constant	4.123***	4.133***	5.429***	4.562***	3.898***
	(0.104)	(0.115)	(0.220)	(0.123)	(0.145)
Different inverse Mills ratio estimates, for Models (3) and (4):					
Binary Probit IMR			-0.620***		
			(0.0932)		
Ordered probit IMR				-1.439***	
				(0.213)	
Selected interactions with demeaned predicted engagement propensity, for Model (5):					
1(Age $<$ 45)					-0.629***
$\times$ Demeaned Engagement Propensity					(0.230)
Demeaned Engagement Propensity					0.514**
					(0.212)
$ln(\sigma)$					
Constant	-0.0879***	-0.0946***	-0.121***	-0.123***	-0.127***
	(0.0237)	(0.0270)	(0.0268)	(0.0269)	(0.0270)
Predicted maximum one-way distances for baseline individual:					
median (miles)	61.7	62.4	228.	95.8	49.3
mean (miles)	93.9	94.3	337.	142.	72.6
Observations	1081	1081	1081	1081	1081

Continued on next page

Log Likelihood	-2404.01	-2403.52	-2377.36	-2375.24	-2371.15
AIC	4838.02	4837.05	4786.72	4782.48	4800.29
BIC	4912.81	4911.83	4866.49	4862.25	4944.88
Weighted?	No	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
[a] Not shown for Model (5): the rest of the full set of interactions with demeaned predicted engagement propensity. None of these other interactions bear coefficients that are statistically significantly different from zero.

In the models reported in Table 6, our explanatory variables include whether the eBird member is currently employed, whether they were willing to report their income in our eBird survey, the level of that income, their gender, their membership in three broad age brackets and three broad educational attainment categories, as well as whether they specifically express *no* interest in each of the five categories of bird species. Comparing Model 2 to Model 1, the use of our weights—based on relative fitted engagement-level probabilities in the general population, as opposed to this eBird member survey sample—results in only minor differences in the parameter estimates. The only notable difference with weighting is that the coefficient on the female indicator, which was negative and statistically significant at the 5% level in the unweighted model, becomes statistically insignificant in Model 2. Given this difference, we retain these weights in subsequent specifications and focus our main discussion on Models 3, 4, and 5.

IMR coefficients. Models 3 and 4 in Table 6 are the two IMR-based selection-corrected models that rely on the strong assumption of bivariate normal errors for the latent engagement intensity variable and the outcome variable. The main coefficient of interest in IMR-based selection-corrected models is the coefficient on the fitted inverse Mills ratio. In two-stage methods, this coefficient is the estimate of $\rho\sigma_\epsilon = \beta_\lambda$, as in equation (4). Given that the error standard deviation, σ_ϵ, must be positive, the sign of this compound parameter implies the sign of ρ, the correlation between the errors in the selection and outcome equations. Our negative IMR coefficients imply that unobserved factors that make a respondent more likely to be intensely engaged with eBird also make them willing to travel less far on a typical one-day birding trip. For both Models 3 and 4, which treat these second-step fitted IMR variables as non-stochastic (thereby understating the amount of noise in the model), these negative IMR coefficient are strongly strongly statistically significant.[22]

[22]We acknowledge that the use of an estimated IMR variable in two-step estimators, employed without corrections to the second-stage parameter variance-covariance matrix, may qualify any inferences we might try to make. FIML estimation of the joint model for engagement propensity and maximum travel distance could

Employment coefficient. Models 1 and 2 in Table 6 both suggest that employment status has no statistically significant effect on the market extent for birders (i.e. on how far the respondent is willing to travel for a typical one-day birding trip). However, Models 3 and 4 suggest that being employed decreases the maximum one-way distance for a birding day-trip by a statistically significant 13% to 14%. Model 5, using our ad hoc correction of interacting each of the regressors with the demeaned engagement propensity variable, suggests that maximum distance decreases by about 20% if the respondent is currently employed. This is not surprising as employed individuals have less free time for all leisure activities, including birding day-trips.

Income availability indicator and level of income, if known. Compared to respondents who decline to provide income data in the eBird member survey, those who do provide income data report a maximum distance for a birding day-trip is lower by about 37% to 58%. However, this negative effect is offset by the positive effect of income (when reported) on maximum travel distance—a 1% higher income corresponds to an extent-of-market that is greater by about 0.2% (i.e. there is an income elasticity of about 0.2 for maximum distance willingly traveled to see birds). Higher-income respondents likely have less-binding budget constraints for travel expenses. Compared to respondents who withhold their income data, the positive effect of greater income overcomes the negative effect associated with the provision of any income data when income reaches roughly $21,400 to $26,200. Given that the mean reported household income in the sample is about $87,200, and the minimum reported income is $18,000, the effect of additional income on maximum travel distance is positive for most of the sample.

Gender coefficient. Whereas the effect of gender (i.e. being female) on the outcome variable is estimated to be negative in Models 1 and 2 (and is statistically significantly

remedy this problem and should be pursued in future applications where hypothesis testing is particularly important.

negative in Model 1), the effect of gender changes sign for IMR-based selection-corrected estimates in Models 3 and 4. In these two models, being female appears to increase maximum travel distance by 10% to 12%, significant at the 10% level. However, the coefficient is not statistically different from zero in Model 5.

Age coefficients. Models 1 and 2 suggest that being less than 45 years old (compared to being between 45 and 64 years old) is associated with a maximum travel distance that is greater by about 18% to 21%, and that being 65 years or older has a negative, but not statistically significant relationship to maximum travel distance. These age effects are reversed in Models 3 and 4, where younger eBird members are willing to travel less far, but older eBird members are willing to travel distances that are about 13% to 32% greater. In Model 3, the age effect appears to be strongly significant. It loses significance in Model 4, and neither of the two age effects are statistically different from zero in Model 5, the ac hoc correction specification.

Education coefficients. Relative to eBird members with college degrees, Models 1 and 2 imply that lesser or greater educational attainment has no statistically discernible effect on maximum travel distances for birding trips. None of the selection-corrected models suggests that educational attainment less than a college degree has any statistically significant effect on the outcome variable, but all three selection-corrected models find a negative effect of post-graduate education, where maximum travel distance is lower by about 12 to 22 percent for highly educated respondents (including graduate and professional degrees). These results may reflect smaller amounts of leisure time for this group, or different leisure-time pursuits.

Disinterest in particular categories of species. Across all five specifications, respondents to our eBird member survey who reveal that they are *not* interested in a given type of bird, among five distinct categories of bird species, are willing to travel statistically significantly less far on a birding day-trip (although it is unlikely that anyone is not interested in *all* categories because they are all members of eBird). In Models 3, 4 and 5, a lack

of interest in perching birds is associated with about a 43% to 47% lesser value for maximum travel distance and a lack of interest in game birds other than waterfowl, other things equal, is associated with maximum travel distances lower by about 43% to 50%. A lack of interest in wading birds, waterfowl, or birds of prey does not have a statistically significant effect on the outcome variable, although the point estimates are considerably smaller in the first two cases for the IMR-based Models 3 and 4.

Model 5's interaction terms. Models 3 and 4 rely upon a strong assumption of bivariate normal errors. This assumption permits a single IMR term, with an unrestricted coefficient, to yield slope coefficients that are uncontaminated by sample selection. This strategy can be described as a structural approach to sample selection correction. In contrast, Model 5 is ad hoc, unstructured and highly flexible. This approach makes a different, but perhaps equally strong assumption—that each of the parameters of the outcome model varies linearly with the respondent's predicted propensity to be included in the estimating sample. In this case, we treat this propensity as equivalent to the respondent's propensity to engage with eBird, the latent continuous variable that drives either the binary-probit model for eBird membership, or the ordered-probit model for various eBird engagement-intensity levels. The linear relationship between the estimated coefficients and the predicted engagement propensity may be positive or negative or statistically zero. The counterfactual we wish to simulate is the set of outcome-model parameters that would obtain if everyone in the estimating sample shared the mean engagement propensity in the general population.

We have transformed each respondent's predicted engagement propensity by taking its deviation from population mean (i.e. from its mean in the qBus sample). In the population, the demeaned engagement propensity variable would be zero, but our estimating sample is not representative of that population. It is a selected sample. Table 5 shows that the average demeaned engagement propensity in the estimating sample is positive, at about 0.45, rather than being zero. The people who appear in our estimating sample have a higher-

than-average propensity to engage with eBird, which is not surprising since they are all eBird members. With our demeaned interaction terms, the baseline coefficients on the main variables in the outcome model can be interpreted as the simulated values of those coefficients *at the mean engagement propensity in the population.* In the bottom section of Table 6, for Model 5, we show the estimated systematic effects of predicted engagement propensity on selected basic coefficients in the outcome model. The other interactions terms, involving 1(Employed), 1(Female), 1(Education - Some College or Less) and all five of the indicators for "Not Interested in" different species group, bear coefficients that are not statistically significantly different from zero, and are thus not reported, although they are included in the model.

The estimated coefficients on the interaction terms in Model 5 that are included in Table 6 show that the estimated coefficient on the indicator for "Have income data" is statistically significantly larger for respondents who are more engaged with eBird. Also, the estimated coefficient for the effect of being younger than 45 is statistically significantly smaller for a respondent who has a greater propensity to engage with eBird. In contrast, the coefficient for the effect of being older than 64 is larger with a greater propensity to engage with eBird. Finally, the estimated effect of having "Education - Some Grad School of More" is larger for respondents with higher engagement intensities.

In terms of the qualitative estimates, we can consider the extent of the market implied by these results. The intercept parameter in each of the models in Table 6 gives the logarithm of the expected maximum travel distance for a respondent having zero values for all of the other variables in the model. This would correspond to someone who is not employed, with missing income data, male, aged 45-60 years, with a college degree, who is nevertheless at least somewhat interested in all five categories of birds. Zeroing-out all of the correction terms in each model, the intercept estimates at the bottom of Table 6, suggest that median value of maximum travel distance for such a person ranges from about 60 miles (Models 1,

2, and 5) to 96 miles (Model 4) and a much greater 228-mile distance according to Model 3 (with its binary-probit selection model).[23]

How do the implications of Model 5, the ad hoc correction approach, compare to those from the other four models? All five models in Table 6 produce coefficient estimates with the same sign and level of significance for some variables: the indicator for income data availabilty and the income variable itself, and for indicators of disinterest in each of the five categories of bird species. Model 5 does better than Models 1 and 2 in identifying the statistical significance of the indicator for employed status (assuming that Models 3 and 4 are more "correct" than Models 1 and 2). At least in terms of coefficient signs, Model 5 is consistent with Models 3 and 4 (and contrary to Models 1 and 2) in the case of 1(Female), 1(Age $>$ 64) and 1(Education - Some College or Less). Finally, Model 5 is consistent with Models 3 an 4 in finding a statistically significant negative coefficient on 1(Education - Some Grad School or more) whereas Models 1 and 2 suggest a positive (although statistically insignificant effect).

Provided that the assumption of bivariate normal errors for the selection and outcome equations is valid, and conditional on the assumption that it is appropriate to use the selection equations fitted on our auxiliary qBus sample with the data from our eBird sample, our preferred specification is probably Model 4, with its more-general selection equation. Table 6 certainly suggests that the choice between a binary-probit and an ordered-probit selection equation might prove to be very important, given the marked differences in the fitted median of the maximum travel distance for the baseline respondent.

[23]Recall that exponentiation of a fitted log value yields the median of the fitted level. One must multiply by the fitted value of $(\sigma^2/2)$ to recover the mean of the fitted conditional distribution, due to the skewness of the implied log-normal distribution.

6 Discussion and Directions for Future Research

6.1 Some comments and qualifications

6.1.1 Sign of error correlation

A priori, we expected (if anything) that the propensity to participate in eBird would be positively associated with the outcome variable that we consider in this analysis, since latent birding avidity could be a potentially important omitted variable. It is thus somewhat counter-intuitive to find negative coefficients on our sample-selection correction terms, suggesting that people who are more-intensively engaged with the eBird project are actually willing to travel *less* far to observe birds on a typical one-day birding trip. Thus the most important omitted variable affecting both equations seems not to be latent birding avidity. Instead, these unobserved factors may include the opportunity cost of time, age-related technical sophistication in responding to online surveys, or some other factor.

6.1.2 Other possible layers of selection

Our estimating sample for the illustrative market-extent model in this paper consists of respondents to our eBird member survey who provided complete data for all except the (typically sensitive) detailed income variable. The selection-correction strategies we feature in this paper presume that this group of eBird member survey respondents is representative of eBird members, an assumption we make to permit us to focus on the problem of systematic selection into eBird engagement at different levels of intensity.

Of course, there may be a variety of reasons why invited eBird members decide not to participate in our survey. In other research, we have sought to control for heterogeneity across all eBird members by linking the center of gravity of their birding trips to a specific census tract, which we impute to be their home census tract. We have employed census tract

attributes as proxies for possible systematic variation across birder characteristics because there is no sociodemographic information available for eBird members who did not respond to our survey. We then employ these various census tract attributes to construct a "propensity of an invited eBird member to respond to our survey" and use this propensity in an attempt to control crudely for selection into our estimating sample for destination choice models. We do not attempt to overlay that correction procedure in addition to the strategies employed here. It is possible that the selection of eBird members into our survey, based on the attributes of their home zipcode relative to those of the general population, is dominated by the selection of the general population into eBird participation, but this is an empirical question that is beyond the scope of this study.

Even with the illustrative example concerning the extent of the market for birding excursions, it is possible that sample selection may occur along more than one dimension. Respondents to our eBird survey may have unobserved characteristics that make them simultaneously more likely (than otherwise expected) to be members of eBird and also more-likely than average to participate in travel of more than one mile from home to observe birds. It is this second behavior upon which we base our models of the "active" recreational use of opportunities to watch wild birds. In addition to the eBird engagement-level variable captured by our ordinal variable $CS6_i$, we can also distinguish between birders who do, or do not, travel more than one mile to see birds, in both the qBus sample and the eBird sample. Three categories of actual bird-watching might be distinguished: no birding trips, trips only less than 1 mile, and trips of one mile or more. Thus the ordinal eBird engagement levels might be supplemented by a second (presumably correlated) ordinal variable that captures heterogeneity in the actual bird-watching behavior of respondents in both surveys. Models with selection on two (correlated) latent variables would require working with trivariate normal joint distributions of the error terms. These models are beyond the scope of the current paper.

Nevertheless, the key insight from our analysis remains. Sample selection has the potential to create distortions in the inferences that researchers may try to draw, if they rely upon convenience samples of participants in CS projects rather that surveys of the general population. However, no government agency, to the best of our knowledge, currently surveys people about their levels of engagement with CS projects (although government agencies run several general population surveys related to recreational activities and also provide funding for numerous CS projects concerning environmental goods that may affect people's utility derived from those recreational activities).

6.2 Issues still to be addressed

6.2.1 When the outcome equation is a conditional logit choice model, rather than a regression model

Our broader research project with the eBird sample involves destination choice models and inferred preferences for site attributes, employed to estimate non-market benefits associated with wild birds. However, Heckman selection-correction models are not appropriate when the outcome variable of interest is a discrete choice, because the latent choice propensity in a multiple discrete-choice model is *not* conditionally normally distributed.[24] When there is no bivariate normal error term to justify the use of a fitted IMR from a selection model as an additional regressor in an outcome equation, it is nevertheless still possible to explore the more ad hoc correction that accommodates systematic differences in selection propensities across respondents by allowing second-stage parameters to differ systematically with deviations of fitted respondent selection propensities from the average propensity in the general population. This ad hoc correction is used in Cameron and DeShazo (2013), Johnston and

[24]Some researchers (e.g. Yuan et al. (2015)) have inserted a fitted inverse Mills ratio (IMR) into a second-stage discrete-choice model, although there seems to be no statistical justification for this particular transformation of the fitted selection propensity.

Abdulrahman (2017), Kolstoe and Cameron (2017) and Kolstoe et al. (2018).

6.2.2 Estimated regressors and inference in a second-stage model

Of course, any two-step estimation process that does not account for the estimated property of the $\hat{\gamma}^q$ parameters embodied in the calculated IMR terms—as in Models 3 and 4 (or the fitted de-meaned response propensities, as in Model 5)—can risk some bias in the inferences to be drawn in the second step. The IMR term (or fitted de-meaned predicted response propensity) is an "estimated regressor" that overstates the amount of information in the data. It may be straightforward (if tedious), to implement an appropriate FIML estimator in the case where one does not need to contend with any missing data in the dataset to which the selection specification is to be transferred. Recall that in this case, it was necessary to estimate 24 different selection equation specifications using most combinations of none, some, or all of the categories of indicators in the full selection model.

If all variables for the selection equation were available for every observation in the CS sample having the outcome variable of interest, one could define the log-likelihood function over the full set of parameters: $\gamma, \beta, \sigma_\eta, \sigma_\epsilon, \rho$. The structure of the two-step model could be preserved, but the two equations could be estimated simultaneously, constraining the γ parameters to be the same in both the selection equation using the qBus sample and the outcome equation using the IMR term, where the index for the IMR variable is constructed using the γ parameters combined with the Z_j variables for the eBird dataset. The matter of how to construct the weights would need to be resolved, of course. We do not attempt this approach here, because of the significant amount of missing data for the selection equation applied to the eBird sample, and corresponding proliferation of different specifications necessary to provide predicted propensities, IMRs, and weights that maximize our use of the available data for each eBird respondent.

7 Conclusions and Recommendations

We adapt methods from the sample selection, transportation and marketing literatures to (1) explain sample selection techniques we have developed to handle the use of data from auxiliary surveys to correct for sample selection present in citizen science data, and (2) model the market extent of regular bird-watching day-trips in Oregon and Washington states. We correct for the propensity to be an eBird member as well as the intensity level of engagement with eBird by its members. Our results for the market-extent model demonstrate that corrections for sample selection and weighting for engagement intensity may be important to the generalizablity of models estimated with citizen science data.

The key takeaway from our market-extent "outcome" model is the potential importance of non-random selection into citizen science projects. Preferences in the general population are important if government agencies, for example, are to make good decisions about the efficient allocation of resources to protect wild birds, a public good. How to provide the appropriate amount of wild bird habitat is an increasingly relevant policy question because land-cover change and climate change present significant threats to wild bird populations. Changes in bird populations will affect birdwatcher welfare (see Kolstoe et al. (2018) for an illustrative example). To limit the loss of bird populations and bird biodiversity, multiple agencies at all levels of government will likely need to work together. It is important to recognize—especially in the case of migratory species such as birds—that actions in one location have the potential to affect outcomes at other locations. Existing programs, such as the National Wildlife Refuge System and the Urban Bird Treaty Program, make a good start but appear not to have been sufficient, given that avian biodiversity remains a concern (in light of changes in land cover and the climate). Conservation solutions must account for the fact that political jurisdictions may not align with the spatial "market extent" for non-market demands for conservation.

The need for a qBus-type sample to permit sample selection corrections in this instance highlights the potential value of broad-based surveys of bird-watching trip behavior and citizen-science engagement. Information about trip-taking behavior has been gathered by the U.S. Fish and Wildlife Service through their quinquennial survey on Fishing, Hunting and Wildlife Watching. However, as of 2016, the information began to be reported only at the census division level, rather than the state-level as had been done in prior surveys. The loss of geographic resolution from this decision limits the usefulness of FHWAR information to city-, county- and state-level government agencies.

In addition, the FHWAR survey is the most appropriate survey to which a detailed question could be added about participation in outdoor-based CS projects (assuming U.S. Fishing, Hunting & Wildlife Watching Survey continues in the future). Also, given that the federal registry now documents more than 400 CS projects (see `www.citizenscience.gov`), it would benefit other agencies, such as the National Oceanic Atmospheric Administration (NOAA) or US Geological Survey (USGS), that could also exploit data from CS projects on recreational behavior, for example Watch for Whales (NOAA), Geocache for a Good Cause (NOAA), and Nature's Notebook (USGS).

To be most useful, existing general-population surveys could (and should) include questions about citizen science engagement in projects related to ecosystems services that are valued for active recreational activities. Also, this engagement information would be useful in any special-purpose surveys fielded to members of CS projects to help researchers understand both active and passive use values for a wide range of environmental public goods. Without general-population information, it will continue to be very difficult to scale to the general population any empirical findings based solely on surveys fielded to "convenience samples" of CS participants.

References

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

Cameron, T. A. and J. Englin (1997). Respondent experience and contingent valuation of environmental goods. *Journal of Environmental Economics and Management 33*(3), 296–313.

De Valck, J. and J. Rolfe (2018). Spatial heterogeneity in stated preference valuation: Status, challenges and road ahead. *International Review of Environmental and Resource Economics 11*(4), 355–422.

Greene, W. H. (2012). *Econometric Analysis (7 th)*. Upper Saddle River, NJ: Prentice Hall.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, 153–161.

Johnston, R. J. and A. S. Abdulrahman (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., S. A. Wood, S. Polasky, C. Kling, C. T. Filstrup, and J. A. Downing (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 T. A. Cameron (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., T. A. Cameron, and C. Wilsey (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.

Loomis, J. B. (1996). How large is the extent of the market for public goods: evidence from a nationwide contingent valuation survey. *Applied economics 28*(7), 779–782.

Rodrigue, J.-P., C. Comtois, and B. Slack (2016). *The geography of transport systems*. Routledge.

Rosenberg, K. V., A. M. Dokter, P. J. Blancher, J. R. Sauer, A. C. Smith, P. A. Smith, J. C. Stanton, A. Panjabi, L. Helft, M. Parr, and P. P. Marra (2019). Decline of the north american avifauna. *Science*.

Sonter, L. J., K. B. Watson, S. A. Wood, and T. H. Ricketts (2016). Spatial and temporal dynamics and value of nature-based recreation, estimated via social media. *PLoS one 11*(9), e0162372.

Terza, J. V. (1998). Estimating count data models with endogenous switching: Sample selection and endogenous treatment effects. *Journal of Econometrics 84* (1), 129–154.

Vella, F. (1998). Estimating models with sample selection bias: a survey. *Journal of Human Resources*, 127–169.

Wood, S. A., A. D. Guerry, J. M. Silver, and M. Lacayo (2013). Using social media to quantify nature-based tourism and recreation. *Scientific reports 3*, srep02976.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.-2002.

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

Complete contents of paper and online appendices:

How to Use Auxiliary General-Population Samples to Reduce Sample Selection Bias in Modeling Relevant Geographic Extents for Citizen Scientists

Contents

List of Figures

List of Tables

Appendices
(Online supplementary material)

A Survey questions posed to general-population qBus sample concerning activities related to wild birds

Table A1: Share of 1050 qBus respondents reporting at least one day over the last year of engagement with the following wild-bird-related activities (response format: slider with labels at 0, 61, 122, 183, 243, 304 and 365 days); mean days per year and lower and upper quartiles of days. We find these counts to be rather high. This may be an artifact of using the Qualtrics' sliders to elicit numbers of days.

Description	N	At least 1 day	Mean days/ year	Lower quartile (days)	upper quartile (days)
Positive non-consumptive engagement with wild birds:					
Pause what you are doing to observe wild birds	999	0.881	92.8	10	161
Put out food for wild birds	960	0.783	98.4	3	181
Seek opportunities to learn more about wild birds	940	0.728	69.7	0	112
Photograph wild birds	945	0.747	68.5	0	107
Visit public parks/areas less than one mile from home to see, photograph or feed wild birds	926	0.703	67.2	0	109
Travel more than one mile from home to see, photograph or feed wild birds	914	0.667	62.4	0	90
— Any days, any of the above?	1050	0.878	-	-	-
Other interactions with wild birds:					
Employ measures to keep wild birds from harming your garden or property	910	0.624	60.3	0	92
Hunt wild birds for sport or for food	895	0.517	49.7	0	48

B Review of the usual context for Heckman's two-stage binary selection correction

For the qBus general population panel, consider a binary indicator for eBird participation CS_i and a set of available regressors Z_i, for a representative sample. If we also had data, for these eBird participants, on an outcome variable of interest, y_i, and a set of regressors, X_i for a subset of this same sample, we would proceed as follows. Suppose the latent propensity to participate in eBird in this qBus sample is a linear-in-parameters function of the Z_i variables, $CS_i^* = Z_i\hat{\gamma} + \eta_i$, then the standard Heckman two-step sample-selection correction procedure involves two terms constructed from $Z_i\hat{\gamma}$.[25] Define:

$$\begin{aligned} \lambda\left(\alpha_{CS_i}\right) &= \lambda\left(-Z_i\hat{\gamma}\right) = \frac{\phi\left(-Z_i\hat{\gamma}\right)}{1-\Phi\left(-Z_i\hat{\gamma}\right)} = \frac{\phi\left(Z_i\hat{\gamma}\right)}{\Phi\left(Z_i\hat{\gamma}\right)} \\ \delta\left(\alpha_{CS_i}\right) &= \delta\left(-Z_i\hat{\gamma}\right) = \lambda\left(-Z_i\hat{\gamma}\right)\left[\lambda\left(-Z_i\hat{\gamma}\right) - \left(-Z_i\hat{\gamma}\right)\right] \end{aligned} \tag{4}$$

With sample selection, the conditional expected value and the error variance of the outcome variable y_i are no longer given simply by $E\left[y_i\right] = X_i\beta$ and $Var\left[y_i\right] = \sigma_y^2$. Instead, we need the expected value and variance of the *marginal* distribution of y_i *conditional* on y_i being observed (i.e. when $CS_i = 1$). If we can assume that the latent propensity variable CS_i^* is distributed bivariate normal with the outcome variable y_i, but the joint distribution is truncated below at $-Z_i\hat{\gamma}$ in the CS_i^* dimension, the formulas for the expected value and variance of the relevant marginal distribution of y_i for this singly truncated bivariate normal distribution are as follows, as in Greene (2012, p. 836):

$$\begin{aligned} E\left[y_i|y_i\ observed\right] &= E\left[y_i|CS_i^* > -Z_i\hat{\gamma}\right] = X_i\beta + \rho\sigma_\epsilon\lambda\left(-Z_i\hat{\gamma}\right) = X_i\beta + \beta_\lambda\lambda\left(-Z_i\hat{\gamma}\right) \\ Var\left[y_i|y_i\ observed\right] &= Var\left[y_i|CS_i^* > -Z_i\hat{\gamma}\right] = \sigma_y^2\left[\left(1-\rho^2\delta\left(-Z_i\hat{\gamma}\right)\right]\right. \end{aligned} \tag{5}$$

These formulas provide the rationale for the Heckman two-step approach and why, once this augmented second-stage model has been estimated, we would have unbiased estimates of the expected value of y_i when y_i is observed under the counterfactual conditions where the correlation between the errors in these two equations is zero. For *uncorrelated* bivariate normal variables, the conditional distributions are everywhere equal to the marginal distribution, so we want to *simulate* the absence of any such error correlation. Based on the augmented regression model, therefore, we can set $\rho = 0$ to get:

$$\begin{aligned} E\left[y_i|y_i\ observed\right] &= X_i\beta + (0\sigma_\epsilon)\lambda\left(-Z_i\hat{\gamma}\right) = X_i\beta \\ Var\left[y_i|y_i\ observed\right] &= \sigma_y^2\left[\left(1-(0)^2\delta\left(-Z_i\hat{\gamma}\right)\right]\right. = \sigma_y^2 \end{aligned} \tag{6}$$

[25] Typically, however, attention is focused primarily on the λ term.

C Legitimate use of Inverse Mills Ratios to correct for sample selection

Over the last several decades, empirical researchers have become accustomed to the idea that estimating a sample-selection model via maximum likelihood methods, calculating the IMR, and including that estimated IMF into the desired "outcome" equation of interest will (somehow) purge the parameters of that outcome equation of any bias due to sample selection. However, it is crucial to remember that the IMR offers an appropriate correction for sample-selection bias only under some very specific conditions. Confidence that "including an IMR term" will "fix" selection bias hinges on the assumption that the selection equation and the outcome equation have error terms that are jointly normally distributed.

The joint normality assumption is critical because the IMR correction derives entirely from the formula for the expected value of a singly truncated bivariate normal distribution. If the conditional distribution latent variable in the selection equation is not normal and the conditional distribution of the dependent variable in the outcome equation is not normal (either observed or censored in some way, perhaps after some transformation), then the needed expected value of the singly truncated joint distribution of the errors in the selection equation and the outcome equation cannot automatically be assumed to be given by the usual IMR formulas.

Ideally, selection and outcome equations should be estimated jointly, in which case a wide variety of joint distributions for the two error terms can be assumed/employed, provided that the joint density can be derived and written down. In some cases, it is convenient to write the conditional joint distributions of the selection propensity and the outcome variable as the product of a conditional distribution and a marginal distribution.[26]

This insight is especially relevant for researchers who wish to estimate conditional logit "outcome" models based on people's choices across alternatives with different attributes. Nothing stops the analyst from estimating a binary probit sample selection model and calculating the usual IMR term from the fitted parameters. However, there is no rigorous statistical rationale for including this fitted IMR term like other respondent characteristics as a variable that might shift one or more slope characteristics or the coefficient on the status quo indicator variable, as is done in Yuan et al. (2015). Some types of joint models where IMR correction terms can make sense, statistically, include the following:

- The usual OLS outcome regression with a continuous dependent variable that is con-

[26]Stata now includes the "heckpoisson" estimator, following Terza (1998). Appropriately, this estimator is available only as a FIML estimator, not as a two-step estimator that relies on an IMR term. Jointly distributed variables that are not both normal have also bee used in a FIML model that combines a participation/experience variable (that is distributed either Poisson or zero-inflated-Poisson) with a censored-normal outcome variable is estimated jointly in Cameron and Englin (1997).

ditionally normal, perhaps after some transformation

- A Tobit outcome model (censored anywhere, at the bottom, the top, or both) with a partially censored normal propensity variable
- An interval-data outcome variable censored between known thresholds
- An ordered probit outcome model with a normally distributed latent propensity variable
- A censored normal outcome model with different censored points across observations

Simply appending an estimated IMR variable to a second-step outcome equation of interest cannot be assumed to be correct in any of the following cases:

- Count data models: Poisson, negative binomial
- Conditional logit models: fixed or random parameters
- Any other statistical model for the "outcome" equation, where the (perhaps latent) dependent is not normal (or at least normal after transformation)

D Additional complications to estimating IMR: Dealing with missing values for Z_j variables in the eBird sample

D.1 qBus sociodemographic variables have few missing values

Any empirical application of this methodology may have to confront the problem of what to do when there are missing values of some variables in one sample or the other. If the correction is based upon the standard sociodemographic variables available for qBus panel members, the data for those variables can be expected to be relatively complete. Any missing values in the qBus sample might be expected to be missing at random.[27]

D.2 eBird sociodemographics match Census, but have more missing values

Missing values in the citizen-science eBird sample, for the sociodemographic variables that conform to the set available the qBus sample, are likely to be more of a problem. For example, due to time constraints for our survey of eBird members, we elected not to ask about individuals' political ideologies. Had we anticipated being able to employ qBus questions to build sampling weights and estimated response propensities, it would have been prudent to be sure that the citizen science members were asked *every* standard sociodemographic question, verbatim, that is available with the qBus responses. For this first example of our procedure, we can assemble conformable measures for gender, race, ethnicity, income brackets, four regions of the U.S., employment status and educational attainment. Some aggregation of categories has been required in each sample to produce matching categories. In future applications of this method, it would be prudent to minimize this type of aggregation. In the eBird data, we used categories that matched the U.S. Census, which would facilitate more-conventional comparisons of marginal distributions in the eBird sample to marginal distributions in the general population. However, the U.S. Census does not provide any information about engagement in citizen science, so our special-purpose qBus sample is much superior in that way.

D.3 Using maximal available Z_j regressors for each eBird observation

Suppose there were no data in the eBird sample on any of the same sociodemographic regressors, Z_i provided by the qBus sample. There would still be valuable information in the

[27] If other key variables intended to serve as regressors, Z_i, in our weighting strategy have been posted to qBus participants, however, it is entirely possible that there may be item non-response for some of those variables. We do not consider that issue in this analysis, since we rely exclusively on the standard sociodemographics available for all qBus panelists.

qBus sample that could help construct either probability weights or propensity corrections. If one runs an ordered probit model to explain the engagement outcome in the eBird data, but use *no* explanatory variables, the result is a set of estimates for only the three cut-points between the four outcome levels in in that eBird data. If one then calculates the predicted probabilities for each of the four participation intensities, the means of these probabilities, across the sample, match the proportions of the sample observed at each level.[28]

D.4 If there are no Z_j regressors available for some eBird respondents

With no Z_j regressors available for some (small) subset of observations in the eBird sample, the best available option for weighting the observations at each level of participation intensity must be derived solely from (a) the predicted probabilities for each of the four relevant participation-intensity levels in the qBus sample (also estimated without regressors) relative to (b) the analogous predicted probabilities for the same four participation intensity levels in the qBus sample. The model being used to predict participation intensities, in this case, has no Z regressors, so there is no basis for observable systematic heterogeneity in these probabilities. The weights will differ across the four observed participation intensity levels, but will be the same for every person who has no available Z variables in the eBird sample.

D.5 If only some subset of Z_j regressors is available for some eBird respondents

The most-general approach to weighting by participation intensity level or correcting parameters for different-from-average participation intensity would exploit the maximum information available in both samples, on an observation-by-observation basis for the eBird sample. To simplify, assume that only three basic factors are available as explanatory variables. In practice, each factor may be captured by a set of indicators for the categories of that factor, but we will assume for now that there is one continuous variable per factor such that the universe of potential Z variables consists of Z_1, Z_2, and Z_3. All three variables (groups of variables) are available for each qBus observation, but different observations in the eBird sample have missing values for either one, two, or all three of these variables.

To fully exploit the available information, it is necessary to estimate an array of models for the qBus sample so that one of these models will be appropriate to transfer to every observation in the eBird sample. Suppose that we have indicators for the presence or absence of values for each of these three Z variables in the eBird sample. The number of necessary

[28]For binary probit and logit models, the means of the fitted probabilities will be either extremely close to the observed proportions, or exactly equal to those proportions, as can be proven by the algebra of the first-order conditions for the maximum likelihood estimation algorithm.

models using the qBus data could then be calculated using the sum of all the relevant combinations:

$$C_0^3 + C_1^3 + C_2^3 + C_3^3 = 1 + 3 + 3 + 1 = 8 \tag{7}$$

Of course, as the number of potential factors increases, the number of potentially relevant models to explain participation intensities in the qBus data can increase dramatically. With six different factors potentially missing for at least some observations in the eBird sample, the number of potentially relevant models would be 64.

E Six-level ordered-probit eBird engagement-level models for all relevant subsets of regressors (using data from the qBus Sample); needed to compute weights, as well as to predict IMRs for eBird member survey sample

Table A2: qBus sample: Model 1-3 (of 24) to accommodate eBird missing values

	Model 1	Model 2	Model 3
Engagement-level indicator			
Region: Northeast	0.215***	0.181***	0.215***
	(0.0647)	(0.0662)	(0.0681)
Region: Midwest	-0.0754	-0.0661	-0.0259
	(0.0656)	(0.0672)	(0.0692)
Region: South	-0.0453	-0.0156	0.0378
	(0.0576)	(0.0591)	(0.0608)
Employment status: Part time		-0.0281	
		(0.0643)	
Employment status: Looking for work		-0.101	
		(0.0935)	
Employment status: Unemployed		-0.345***	
		(0.0697)	
Employment status: Retired		-0.877***	
		(0.0751)	
Education: High school		0.129**	-0.0460
		(0.0639)	(0.0648)
Education: Some college		-0.116**	-0.195***
		(0.0586)	(0.0603)
Education: Masters degree		0.347***	0.371***
		(0.0726)	(0.0748)
Education: Doctoral degree		0.371***	0.307***
		(0.110)	(0.113)
Gender = male			0.325***
			(0.0460)
Age 24 years or less			0.721***
			(0.0877)

Continued on next page

Table A2 – continued from previous page

Age 25 to 34 years			0.731***
			(0.0785)
Age 35 to 44 years			0.539***
			(0.0810)
Age 55 to 64 years			-0.435***
			(0.100)
Age 65 years and up			-0.563***
			(0.108)
/			
cut1	0.861***	0.736***	1.362***
	(0.0456)	(0.0615)	(0.0932)
cut2	1.217***	1.111***	1.766***
	(0.0473)	(0.0629)	(0.0947)
cut3	1.399***	1.305***	1.975***
	(0.0488)	(0.0641)	(0.0960)
cut4	1.620***	1.543***	2.231***
	(0.0513)	(0.0660)	(0.0980)
cut5	1.995***	1.943***	2.655***
	(0.0581)	(0.0719)	(0.103)
Observations	4161	4161	4161
Max. log-likelihood	-3274.80	-3159.02	-3014.04

t in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A3: qBus sample: Model 4-6 (of 24) to accommodate eBird missing values

	Model 4	Model 5	Model 6
Engagement-level indicator			
Gender = male	0.288***	0.287***	
	(0.0469)	(0.0471)	
Age 24 years or less	0.749***	0.751***	
	(0.0908)	(0.0914)	
Age 25 to 34 years	0.730***	0.728***	
	(0.0795)	(0.0800)	
Age 35 to 44 years	0.533***	0.533***	
	(0.0817)	(0.0818)	
Age 55 to 64 years	-0.416***	-0.412***	
	(0.103)	(0.103)	
Age 65 years and up	-0.487***	-0.487***	
	(0.125)	(0.125)	
Region: Northeast	0.217***	0.219***	0.253***
	(0.0683)	(0.0684)	(0.0666)
Region: Midwest	-0.0293	-0.0297	-0.0525
	(0.0694)	(0.0696)	(0.0678)
Region: South	0.0366	0.0375	-0.0254
	(0.0610)	(0.0612)	(0.0595)
Employment status: Part time	-0.0421	-0.0394	
	(0.0694)	(0.0705)	
Employment status: Looking for work	-0.185*	-0.182*	
	(0.0971)	(0.0984)	
Employment status: Unemployed	-0.333***	-0.327***	
	(0.0736)	(0.0757)	
Employment status: Retired	-0.157	-0.157	
	(0.0994)	(0.100)	
Education: High school	0.0266	0.0307	
	(0.0669)	(0.0698)	
Education: Some college	-0.165***	-0.164***	
	(0.0610)	(0.0623)	
Education: Masters degree	0.363***	0.369***	
	(0.0749)	(0.0760)	
Education: Doctoral degree	0.301***	0.319***	

Continued on next page

Table A3 – continued from previous page

	(0.113)	(0.115)	
Income less than 25K		-0.00878	
		(0.0798)	
Income 25 K to 50 K		0.00622	
		(0.0723)	
Income 75 K to 100 K		0.0736	
		(0.0792)	
Income 100 K or more		-0.0253	
		(0.0723)	
Travel 1 mile data available			0.538***
			(0.0461)
Some 1+ mile trip for birds			1.418***
			(0.101)
/			
cut1	1.295***	1.303***	2.403***
	(0.0949)	(0.105)	(0.114)
cut2	1.700***	1.709***	2.788***
	(0.0964)	(0.107)	(0.115)
cut3	1.911***	1.920***	2.984***
	(0.0976)	(0.108)	(0.117)
cut4	2.169***	2.178***	3.219***
	(0.0996)	(0.110)	(0.118)
cut5	2.597***	2.606***	3.611***
	(0.105)	(0.115)	(0.122)
Observations	4161	4161	4161
Max. log-likelihood	-3002.21	-3001.27	-3118.07

t in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A4: qBus sample: Model 7-9 (of 24) to accommodate eBird missing values

	Model 7	Model 8	Model 9
Engagement-level indicator			
Travel 1 mile data available	0.546***	0.540***	0.542***
	(0.0465)	(0.0469)	(0.0472)
Some 1+ mile trip for birds	1.422***	1.429***	1.426***
	(0.101)	(0.103)	(0.103)
Region: Northeast	0.230***	0.243***	0.217***
	(0.0670)	(0.0677)	(0.0681)
Region: Midwest	-0.0539	-0.0426	-0.0401
	(0.0681)	(0.0691)	(0.0694)
Region: South	-0.0185	-0.00480	0.00462
	(0.0598)	(0.0607)	(0.0610)
Education: High school	-0.0229		0.0529
	(0.0631)		(0.0660)
Education: Some college	-0.175***		-0.141**
	(0.0589)		(0.0604)
Education: Masters degree	0.320***		0.348***
	(0.0732)		(0.0745)
Education: Doctoral degree	0.295***		0.323***
	(0.111)		(0.113)
Employment status: Part time		-0.114*	-0.0734
		(0.0650)	(0.0660)
Employment status: Looking for work		-0.164*	-0.120
		(0.0955)	(0.0969)
Employment status: Unemployed		-0.400***	-0.348***
		(0.0692)	(0.0719)
Employment status: Retired		-0.885***	-0.892***
		(0.0773)	(0.0781)
/			
cut1	2.398***	2.217***	2.253***
	(0.118)	(0.117)	(0.121)
cut2	2.788***	2.618***	2.659***
	(0.120)	(0.119)	(0.123)
cut3	2.988***	2.823***	2.868***
	(0.122)	(0.120)	(0.124)
cut4	3.230***	3.070***	3.122***

Continued on next page

Table A4 – continued from previous page

	(0.123)	(0.121)	(0.126)
cut5	3.633***	3.479***	3.541***
	(0.127)	(0.125)	(0.130)
Observations	4161	4161	4161
Max. log-likelihood	-3089.97	-3034.60	-3008.29

t in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A5: qBus sample: Model 10-12 (of 24) to accommodate eBird missing values

	Model 10	Model 11	Model 12
Engagement-level indicator			
Travel 1 mile data available	0.525***	0.533***	0.535***
	(0.0480)	(0.0484)	(0.0483)
Some 1+ mile trip for birds	1.444***	1.452***	1.451***
	(0.107)	(0.107)	(0.106)
Age 24 years or less	0.710***	0.757***	0.733***
	(0.0923)	(0.0932)	(0.0910)
Age 25 to 34 years	0.716***	0.715***	0.724***
	(0.0810)	(0.0815)	(0.0811)
Age 35 to 44 years	0.542***	0.530***	0.532***
	(0.0833)	(0.0839)	(0.0832)
Age 55 to 64 years	-0.384***	-0.393***	-0.425***
	(0.105)	(0.106)	(0.103)
Age 65 years and up	-0.328***	-0.416***	-0.518***
	(0.125)	(0.128)	(0.111)
Region: Northeast	0.271***	0.242***	0.247***
	(0.0693)	(0.0698)	(0.0697)
Region: Midwest	-0.0276	-0.0258	-0.0235
	(0.0710)	(0.0714)	(0.0713)
Region: South	0.0281	0.0369	0.0454
	(0.0623)	(0.0626)	(0.0625)
Employment status: Part time	-0.194***	-0.139**	
	(0.0692)	(0.0702)	
Employment status: Looking for work	-0.313***	-0.252**	
	(0.0985)	(0.0998)	
Employment status: Unemployed	-0.484***	-0.407***	
	(0.0718)	(0.0744)	
Employment status: Retired	-0.260***	-0.216**	
	(0.101)	(0.102)	
Education: High school		-0.0471	-0.0788
		(0.0687)	(0.0705)
Education: Some college		-0.202***	-0.212***
		(0.0626)	(0.0634)
Education: Masters degree		0.377***	0.381***

Continued on next page

Table A5 – continued from previous page

		(0.0766)	(0.0776)
Education: Doctoral degree		0.312***	0.322***
		(0.116)	(0.117)
Income less than 25K			-0.143*
			(0.0789)
Income 25 K to 50 K			-0.0418
			(0.0737)
Income 75 K to 100 K			0.104
			(0.0814)
Income 100 K or more			0.0230
			(0.0738)
/			
cut1	2.639***	2.654***	2.740***
	(0.138)	(0.142)	(0.149)
cut2	3.067***	3.089***	3.174***
	(0.140)	(0.144)	(0.151)
cut3	3.286***	3.314***	3.397***
	(0.141)	(0.146)	(0.153)
cut4	3.549***	3.585***	3.665***
	(0.143)	(0.148)	(0.155)
cut5	3.978***	4.027***	4.102***
	(0.147)	(0.152)	(0.159)
Observations	4161	4161	4161
Max. log-likelihood	-2910.11	-2877.53	-2890.44

t in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A6: qBus sample: Model 13-15 (of 24) to accommodate eBird missing values

	Model 13	Model 14	Model 15
Engagement-level indicator			
Travel 1 mile data available	0.533***	0.541***	0.548***
	(0.0484)	(0.0462)	(0.0466)
Some 1+ mile trip for birds	1.454***	1.423***	1.424***
	(0.107)	(0.101)	(0.101)
Age 24 years or less	0.769***		
	(0.0938)		
Age 25 to 34 years	0.724***		
	(0.0822)		
Age 35 to 44 years	0.531***		
	(0.0840)		
Age 55 to 64 years	-0.392***		
	(0.106)		
Age 65 years and up	-0.419***		
	(0.128)		
Income less than 25K	-0.0453		
	(0.0816)		
Income 25 K to 50 K	-0.0216		
	(0.0742)		
Income 75 K to 100 K	0.0974		
	(0.0816)		
Income 100 K or more	0.0173		
	(0.0741)		
Region: Northeast	0.246***	0.264***	0.240***
	(0.0699)	(0.0669)	(0.0672)
Region: Midwest	-0.0266	-0.0367	-0.0407
	(0.0715)	(0.0680)	(0.0682)
Region: South	0.0411	-0.00797	-0.00468
	(0.0628)	(0.0597)	(0.0600)
Employment status: Part time	-0.124*		
	(0.0714)		
Employment status: Looking for work	-0.234**		
	(0.101)		
Employment status: Unemployed	-0.388***		

Continued on next page

Table A6 – continued from previous page

	(0.0767)		
Employment status: Retired	-0.202**		
	(0.103)		
Education: High school	-0.0218		-0.00742
	(0.0715)		(0.0633)
Education: Some college	-0.186***		-0.154***
	(0.0638)		(0.0591)
Education: Masters degree	0.372***		0.308***
	(0.0778)		(0.0734)
Education: Doctoral degree	0.313***		0.251**
	(0.117)		(0.111)
Gender = male		0.250***	0.212***
		(0.0445)	(0.0451)
/			
cut1	2.686***	2.549***	2.525***
	(0.151)	(0.117)	(0.122)
cut2	3.122***	2.936***	2.918***
	(0.153)	(0.119)	(0.124)
cut3	3.347***	3.134***	3.119***
	(0.154)	(0.120)	(0.125)
cut4	3.619***	3.373***	3.363***
	(0.156)	(0.122)	(0.127)
cut5	4.061***	3.770***	3.770***
	(0.160)	(0.126)	(0.131)
Observations	4161	4161	4161
Max. log-likelihood	-2876.02	-3102.17	-3078.94

t in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A7: qBus sample: Model 16-18 (of 24) to accommodate eBird missing values

	Model 16	Model 17	Model 18
Engagement-level indicator			
Travel 1 mile data available	0.542***	0.544***	0.519***
	(0.0469)	(0.0473)	(0.0479)
Some 1+ mile trip for birds	1.426***	1.422***	1.423***
	(0.103)	(0.103)	(0.106)
Gender = male	0.239***	0.211***	0.364***
	(0.0464)	(0.0468)	(0.0467)
Region: Northeast	0.250***	0.225***	0.279***
	(0.0679)	(0.0683)	(0.0694)
Region: Midwest	-0.0274	-0.0271	0.00273
	(0.0692)	(0.0695)	(0.0709)
Region: South	0.0117	0.0184	0.0577
	(0.0608)	(0.0611)	(0.0623)
Employment status: Part time	-0.0688	-0.0359	
	(0.0658)	(0.0667)	
Employment status: Looking for work	-0.121	-0.0842	
	(0.0961)	(0.0975)	
Employment status: Unemployed	-0.328***	-0.288***	
	(0.0707)	(0.0732)	
Employment status: Retired	-0.882***	-0.889***	
	(0.0773)	(0.0780)	
Education: High school		0.0526	
		(0.0661)	
Education: Some college		-0.128**	
		(0.0606)	
Education: Masters degree		0.340***	
		(0.0747)	
Education: Doctoral degree		0.285**	
		(0.113)	
Age 24 years or less			0.649***
			(0.0890)
Age 25 to 34 years			0.728***
			(0.0803)
Age 35 to 44 years			0.548***

Continued on next page

Table A7 – continued from previous page

			(0.0829)
Age 55 to 64 years			-0.436***
			(0.103)
Age 65 years and up			-0.482***
			(0.109)
/			
cut1	2.366***	2.383***	2.943***
	(0.121)	(0.125)	(0.140)
cut2	2.769***	2.791***	3.372***
	(0.123)	(0.127)	(0.142)
cut3	2.976***	3.002***	3.592***
	(0.124)	(0.128)	(0.144)
cut4	3.226***	3.257***	3.855***
	(0.126)	(0.130)	(0.146)
cut5	3.639***	3.680***	4.284***
	(0.130)	(0.134)	(0.150)
Observations	4161	4161	4161
Max. log-likelihood	-3021.34	-2998.14	-2906.93

t in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A8: qBus sample: Model 19-21 (of 24) to accommodate eBird missing values

	Model 19	Model 20	Model 21
Engagement-level indicator			
Travel 1 mile data available	0.537***	0.526***	0.536***
	(0.0484)	(0.0481)	(0.0485)
Some 1+ mile trip for birds	1.438***	1.435***	1.442***
	(0.106)	(0.107)	(0.107)
Gender = male	0.325***	0.309***	0.284***
	(0.0473)	(0.0478)	(0.0482)
Age 24 years or less	0.728***	0.725***	0.770***
	(0.0904)	(0.0928)	(0.0936)
Age 25 to 34 years	0.729***	0.735***	0.734***
	(0.0809)	(0.0814)	(0.0819)
Age 35 to 44 years	0.535***	0.545***	0.534***
	(0.0836)	(0.0837)	(0.0843)
Age 55 to 64 years	-0.442***	-0.403***	-0.412***
	(0.104)	(0.106)	(0.106)
Age 65 years and up	-0.558***	-0.384***	-0.465***
	(0.111)	(0.126)	(0.128)
Region: Northeast	0.250***	0.281***	0.253***
	(0.0699)	(0.0696)	(0.0701)
Region: Midwest	-0.000740	-0.00400	-0.00458
	(0.0713)	(0.0712)	(0.0715)
Region: South	0.0610	0.0532	0.0591
	(0.0626)	(0.0625)	(0.0628)
Education: High school	-0.133**		-0.0566
	(0.0669)		(0.0690)
Education: Some college	-0.222***		-0.190***
	(0.0621)		(0.0628)
Education: Masters degree	0.377***		0.369***
	(0.0767)		(0.0768)
Education: Doctoral degree	0.264**		0.259**
	(0.116)		(0.116)
Employment status: Part time		-0.140**	-0.0900
		(0.0701)	(0.0710)
Employment status: Looking for work		-0.267***	-0.209**

Continued on next page

Table A8 – continued from previous page

		(0.0993)	(0.101)
Employment status: Unemployed		-0.399***	-0.331***
		(0.0732)	(0.0757)
Employment status: Retired		-0.220**	-0.179*
		(0.101)	(0.103)
/			
cut1	2.895***	2.831***	2.828***
	(0.145)	(0.142)	(0.146)
cut2	3.333***	3.264***	3.268***
	(0.147)	(0.144)	(0.148)
cut3	3.559***	3.486***	3.496***
	(0.148)	(0.145)	(0.150)
cut4	3.831***	3.754***	3.770***
	(0.150)	(0.147)	(0.152)
cut5	4.275***	4.189***	4.218***
	(0.155)	(0.152)	(0.156)
Observations	4161	4161	4161
Max. log-likelihood	-2871.17	-2889.14	-2860.10

t in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A9: qBus sample: Model 22-24 (of 24) to accommodate eBird missing values

	Model 22	Model 23	Model 24
Engagement-level indicator			
Travel 1 mile data available	0.531***	0.538***	0.532***
	(0.0481)	(0.0485)	(0.0482)
Some 1+ mile trip for birds	1.444***	1.441***	1.448***
	(0.106)	(0.106)	(0.107)
Gender = male	0.333***	0.314***	0.296***
	(0.0473)	(0.0477)	(0.0480)
Age 24 years or less	0.718***	0.752***	0.755***
	(0.0907)	(0.0915)	(0.0936)
Age 25 to 34 years	0.764***	0.741***	0.759***
	(0.0809)	(0.0815)	(0.0820)
Age 35 to 44 years	0.547***	0.536***	0.545***
	(0.0831)	(0.0837)	(0.0838)
Age 55 to 64 years	-0.439***	-0.438***	-0.410***
	(0.103)	(0.104)	(0.106)
Age 65 years and up	-0.489***	-0.556***	-0.404***
	(0.110)	(0.111)	(0.126)
Income less than 25K	-0.127	-0.100	-0.0264
	(0.0773)	(0.0794)	(0.0805)
Income 25 K to 50 K	-0.0539	-0.0317	-0.0296
	(0.0736)	(0.0740)	(0.0740)
Income 75 K to 100 K	0.157*	0.0889	0.148*
	(0.0807)	(0.0817)	(0.0809)
Income 100 K or more	0.149**	0.00168	0.136*
	(0.0708)	(0.0742)	(0.0711)
Region: Northeast	0.284***	0.257***	0.281***
	(0.0696)	(0.0701)	(0.0698)
Region: Midwest	-0.00499	-0.000851	-0.0112
	(0.0712)	(0.0715)	(0.0714)
Region: South	0.0657	0.0671	0.0573
	(0.0625)	(0.0628)	(0.0627)
Education: High school		-0.0894	
		(0.0709)	
Education: Some college		-0.199***	

Continued on next page

Table A9 – continued from previous page

		(0.0636)	
Education: Masters degree		0.374***	
		(0.0779)	
Education: Doctoral degree		0.268**	
		(0.118)	
Employment status: Part time			-0.107
			(0.0717)
Employment status: Looking for work			-0.223**
			(0.101)
Employment status: Unemployed			-0.356***
			(0.0766)
Employment status: Retired			-0.182*
			(0.102)
/			
cut1	2.998***	2.908***	2.911***
	(0.148)	(0.152)	(0.150)
cut2	3.430***	3.347***	3.345***
	(0.150)	(0.155)	(0.152)
cut3	3.652***	3.574***	3.569***
	(0.152)	(0.156)	(0.153)
cut4	3.919***	3.847***	3.838***
	(0.154)	(0.158)	(0.155)
cut5	4.353***	4.291***	4.278***
	(0.158)	(0.162)	(0.160)
Observations	4161	4161	4161
Max. log-likelihood	-2896.29	-2868.71	-2883.99

t in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

F Four-level ordered-probit eBird engagement-level models for all relevant subsets of regressors (using data from the eBird member survey sample); needed to compute weights

Table A10: eBird sample: Model 1-3 (of 24) to accommodate missing values

	Model 1	Model 2	Model 3
Engagement-level indicator			
Employment status: Part time		-0.197	
		(0.142)	
Employment status: Looking for work		-0.595	
		(0.426)	
Employment status: Unemployed		-0.144	
		(0.160)	
Employment status: Retired		-0.448***	
		(0.0809)	
Education: High school		0.0641	-0.151
		(0.217)	(0.204)
Education: Some college		-0.226*	-0.179
		(0.120)	(0.109)
Education: Masters degree		0.0742	0.161*
		(0.0900)	(0.0848)
Education: Doctoral degree		0.321***	0.369***
		(0.124)	(0.117)
Gender = male			0.417***
			(0.0699)
Age 24 years or less			0.789***
			(0.267)
Age 25 to 34 years			0.356**
			(0.155)
Age 35 to 44 years			0.184
			(0.141)
Age 55 to 64 years			-0.221**
			(0.107)
Age 65 years and up			-0.366***
			(0.105)

Continued on next page

Table A10 – continued from previous page

/			
cut1	-0.276***	-0.507***	-0.204*
	(0.0387)	(0.0872)	(0.108)
cut2	0.444***	0.255***	0.582***
	(0.0395)	(0.0862)	(0.108)
cut3	1.029***	0.836***	1.195***
	(0.0464)	(0.0895)	(0.112)
Observations	1081	899	1051
Max. log-likelihood	-1422.62	-1159.32	-1332.35

t in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A11: eBird sample: Model 4-6 (of 24) to accommodate missing values

	Model 4	Model 5	Model 6
Engagement-level indicator			
Gender = male	0.394***	0.409***	
	(0.0771)	(0.0859)	
Age 24 years or less	0.899**	0.817**	
	(0.351)	(0.401)	
Age 25 to 34 years	0.384**	0.239	
	(0.171)	(0.186)	
Age 35 to 44 years	0.339**	0.238	
	(0.157)	(0.167)	
Age 55 to 64 years	-0.0677	-0.140	
	(0.125)	(0.135)	
Age 65 years and up	-0.149	-0.0642	
	(0.152)	(0.168)	
Employment status: Part time	-0.112	-0.147	
	(0.148)	(0.161)	
Employment status: Looking for work	-0.625	-0.459	
	(0.432)	(0.465)	
Employment status: Unemployed	-0.0263	-0.150	
	(0.164)	(0.189)	
Employment status: Retired	-0.191	-0.320**	
	(0.122)	(0.136)	
Education: High school	-0.118	-0.0476	
	(0.231)	(0.255)	
Education: Some college	-0.204*	-0.227*	
	(0.121)	(0.136)	
Education: Masters degree	0.125	0.0679	
	(0.0916)	(0.102)	
Education: Doctoral degree	0.326**	0.268*	
	(0.127)	(0.142)	
Income less than 25K		-0.168	
		(0.199)	
Income 25 K to 50 K		0.185	
		(0.126)	
Income 75 K to 100 K		-0.188	

Continued on next page

Table A11 – continued from previous page

		(0.132)	
Income 100 K or more		0.00340	
		(0.117)	
Travel 1 mile data available			0
			(.)
Some 1+ mile trip for birds			1.765***
			(0.232)
/			
cut1	-0.212*	-0.397**	1.397***
	(0.124)	(0.156)	(0.229)
cut2	0.583***	0.466***	2.153***
	(0.125)	(0.156)	(0.231)
cut3	1.181***	1.062***	2.754***
	(0.129)	(0.160)	(0.234)
Observations	896	727	1076
Max. log-likelihood	-1134.30	-927.40	-1375.97

t in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A12: eBird sample: Model 7-9 (of 24) to accommodate missing values

	Model 7	Model 8	Model 9
Engagement-level indicator			
Travel 1 mile data available	0	0	0
	(.)	(.)	(.)
Some 1+ mile trip for birds	1.786***	1.687***	1.711***
	(0.234)	(0.237)	(0.239)
Education: High school	0.284		0.302
	(0.200)		(0.228)
Education: Some college	-0.165		-0.211*
	(0.109)		(0.121)
Education: Masters degree	0.0835		0.0786
	(0.0837)		(0.0911)
Education: Doctoral degree	0.349***		0.322**
	(0.115)		(0.125)
Employment status: Part time		-0.247*	-0.213
		(0.140)	(0.142)
Employment status: Looking for work		-0.658	-0.634
		(0.423)	(0.426)
Employment status: Unemployed		-0.158	-0.0760
		(0.159)	(0.164)
Employment status: Retired		-0.412***	-0.386***
		(0.0807)	(0.0821)
/			
cut1	1.465***	1.067***	1.150***
	(0.239)	(0.240)	(0.251)
cut2	2.239***	1.853***	1.950***
	(0.241)	(0.242)	(0.253)
cut3	2.841***	2.447***	2.547***
	(0.244)	(0.245)	(0.256)
Observations	1050	914	895
Max. log-likelihood	-1334.04	-1152.18	-1121.09

t in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A13: eBird sample: Model 10-12 (of 24) to accommodate missing values

	Model 10	Model 11	Model 12
Engagement-level indicator			
Travel 1 mile data available	0	0	0
	(.)	(.)	(.)
Some 1+ mile trip for birds	1.659***	1.665***	1.828***
	(0.237)	(0.239)	(0.296)
Age 24 years or less	0.787**	0.764**	0.656**
	(0.335)	(0.355)	(0.316)
Age 25 to 34 years	0.329*	0.380**	0.251
	(0.170)	(0.172)	(0.167)
Age 35 to 44 years	0.288*	0.292*	0.0794
	(0.156)	(0.157)	(0.151)
Age 55 to 64 years	-0.0348	-0.0568	-0.312***
	(0.123)	(0.125)	(0.116)
Age 65 years and up	0.00333	-0.0222	-0.278**
	(0.151)	(0.154)	(0.117)
Employment status: Part time	-0.280*	-0.236	
	(0.145)	(0.147)	
Employment status: Looking for work	-0.756*	-0.736*	
	(0.428)	(0.432)	
Employment status: Unemployed	-0.162	-0.0853	
	(0.163)	(0.166)	
Employment status: Retired	-0.316***	-0.267**	
	(0.120)	(0.122)	
Education: High school		0.205	0.321
		(0.243)	(0.242)
Education: Some college		-0.180	-0.152
		(0.122)	(0.122)
Education: Masters degree		0.123	0.0982
		(0.0924)	(0.0942)
Education: Doctoral degree		0.382***	0.367***
		(0.127)	(0.132)
Income less than 25K			-0.161
			(0.171)
Income 25 K to 50 K			0.148

Continued on next page

Table A13 – continued from previous page

			(0.118)
Income 75 K to 100 K			-0.124
			(0.122)
Income 100 K or more			0.130
			(0.105)
/			
cut1	1.122***	1.215***	1.314***
	(0.251)	(0.262)	(0.317)
cut2	1.920***	2.029***	2.164***
	(0.254)	(0.264)	(0.319)
cut3	2.519***	2.630***	2.781***
	(0.256)	(0.267)	(0.321)
Observations	912	894	852
Max. log-likelihood	-1143.39	-1112.77	-1080.05

t in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A14: eBird sample: Model 13-15 (of 24) to accommodate missing values

	Model 13	Model 14	Model 15
Engagement-level indicator			
Travel 1 mile data available	0	0	0
	(.)	(.)	(.)
Some 1+ mile trip for birds	1.810***	1.717***	1.737***
	(0.302)	(0.235)	(0.237)
Age 24 years or less	0.754*		
	(0.403)		
Age 25 to 34 years	0.232		
	(0.187)		
Age 35 to 44 years	0.185		
	(0.168)		
Age 55 to 64 years	-0.141		
	(0.136)		
Age 65 years and up	0.0405		
	(0.169)		
Income less than 25K	-0.147		
	(0.203)		
Income 25 K to 50 K	0.190		
	(0.128)		
Income 75 K to 100 K	-0.170		
	(0.133)		
Income 100 K or more	0.0150		
	(0.117)		
Employment status: Part time	-0.254		
	(0.160)		
Employment status: Looking for work	-0.622		
	(0.462)		
Employment status: Unemployed	-0.230		
	(0.191)		
Employment status: Retired	-0.383***		
	(0.136)		
Education: High school	0.247		0.205
	(0.263)		(0.200)
Education: Some college	-0.189		-0.182*

Continued on next page

Table A14 – continued from previous page

	(0.137)		(0.109)
Education: Masters degree	0.0609		0.0710
	(0.102)		(0.0840)
Education: Doctoral degree	0.328**		0.267**
	(0.143)		(0.117)
Gender = male		0.387***	0.356***
		(0.0686)	(0.0702)
/			
cut1	1.171***	1.505***	1.543***
	(0.329)	(0.233)	(0.242)
cut2	2.052***	2.279***	2.330***
	(0.332)	(0.236)	(0.245)
cut3	2.647***	2.892***	2.944***
	(0.334)	(0.238)	(0.247)
Observations	726	1070	1048
Max. log-likelihood	-911.56	-1354.16	-1320.09

t in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A15: eBird sample: Model 16-18 (of 24) to accommodate missing values

	Model 16	Model 17	Model 18
Engagement-level indicator			
Travel 1 mile data available	0	0	0
	(.)	(.)	(.)
Some 1+ mile trip for birds	1.639***	1.665***	1.642***
	(0.239)	(0.242)	(0.235)
Gender = male	0.368***	0.338***	0.407***
	(0.0758)	(0.0773)	(0.0691)
Employment status: Part time	-0.167	-0.141	
	(0.141)	(0.143)	
Employment status: Looking for work	-0.576	-0.559	
	(0.423)	(0.427)	
Employment status: Unemployed	-0.0442	0.0309	
	(0.161)	(0.166)	
Employment status: Retired	-0.379***	-0.356***	
	(0.0812)	(0.0825)	
Education: High school		0.230	
		(0.228)	
Education: Some college		-0.233*	
		(0.122)	
Education: Masters degree		0.0766	
		(0.0913)	
Education: Doctoral degree		0.261**	
		(0.126)	
Age 24 years or less			0.539**
			(0.255)
Age 25 to 34 years			0.255*
			(0.155)
Age 35 to 44 years			0.134
			(0.141)
Age 55 to 64 years			-0.221**
			(0.107)
Age 65 years and up			-0.304***
			(0.105)
/cut1	1.201***	1.258***	1.286***

Continued on next page

Table A15 – continued from previous page

	(0.245)	(0.255)	(0.246)
cut2	2.000***	2.069***	2.085***
	(0.247)	(0.257)	(0.248)
cut3	2.607***	2.678***	2.708***
	(0.250)	(0.260)	(0.251)
Observations	912	893	1066
Max. log-likelihood	-1139.47	-1110.58	-1331.88

t in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A16: eBird sample: Model 19-21 (of 24) to accommodate missing values

	Model 19	Model 20	Model 21
Engagement-level indicator			
Travel 1 mile data available	0	0	0
	(.)	(.)	(.)
Some 1+ mile trip for birds	1.622***	1.605***	1.612***
	(0.237)	(0.239)	(0.242)
Gender = male	0.374***	0.374***	0.347***
	(0.0707)	(0.0765)	(0.0781)
Age 24 years or less	0.661**	0.722**	0.725**
	(0.269)	(0.335)	(0.355)
Age 25 to 34 years	0.325**	0.313*	0.356**
	(0.156)	(0.171)	(0.173)
Age 35 to 44 years	0.146	0.293*	0.290*
	(0.142)	(0.156)	(0.157)
Age 55 to 64 years	-0.235**	-0.0558	-0.0820
	(0.108)	(0.124)	(0.126)
Age 65 years and up	-0.315***	-0.0691	-0.0953
	(0.106)	(0.152)	(0.155)
Education: High school	0.0809		0.140
	(0.213)		(0.243)
Education: Some college	-0.165		-0.199
	(0.111)		(0.123)
Education: Masters degree	0.164*		0.125
	(0.0857)		(0.0926)
Education: Doctoral degree	0.361***		0.324**
	(0.118)		(0.128)
Employment status: Part time		-0.179	-0.143
		(0.147)	(0.149)
Employment status: Looking for work		-0.678	-0.661
		(0.429)	(0.432)
Employment status: Unemployed		-0.0421	0.0270
		(0.165)	(0.168)
Employment status: Retired		-0.234*	-0.190
		(0.121)	(0.124)
/cut1	1.335***	1.240***	1.308***

Continued on next page

Table A16 – continued from previous page

	(0.254)	(0.255)	(0.265)
cut2	2.150***	2.053***	2.134***
	(0.257)	(0.258)	(0.268)
cut3	2.775***	2.664***	2.746***
	(0.259)	(0.261)	(0.271)
Observations	1046	910	892
Max. log-likelihood	-1297.01	-1130.49	-1101.90

t in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A17: eBird sample: Model 22-24 (of 24) to accommodate missing values

	Model 22	Model 23	Model 24
Engagement-level indicator			
Travel 1 mile data available	0	0	0
	(.)	(.)	(.)
Some 1+ mile trip for birds	1.725***	1.763***	1.693***
	(0.294)	(0.299)	(0.300)
Gender = male	0.378***	0.357***	0.377***
	(0.0768)	(0.0785)	(0.0848)
Age 24 years or less	0.578*	0.628**	0.725*
	(0.308)	(0.316)	(0.395)
Age 25 to 34 years	0.204	0.220	0.212
	(0.165)	(0.167)	(0.185)
Age 35 to 44 years	0.0847	0.0779	0.204
	(0.150)	(0.151)	(0.167)
Age 55 to 64 years	-0.281**	-0.314***	-0.120
	(0.115)	(0.117)	(0.134)
Age 65 years and up	-0.287**	-0.316***	-0.00309
	(0.116)	(0.118)	(0.168)
Income less than 25K	-0.162	-0.116	-0.238
	(0.166)	(0.172)	(0.196)
Income 25 K to 50 K	0.0742	0.138	0.111
	(0.115)	(0.118)	(0.125)
Income 75 K to 100 K	-0.184	-0.160	-0.245*
	(0.121)	(0.123)	(0.132)
Income 100 K or more	0.119	0.0841	0.00779
	(0.104)	(0.106)	(0.116)
Education: High school		0.221	
		(0.243)	
Education: Some college		-0.186	
		(0.123)	
Education: Masters degree		0.0933	
		(0.0945)	
Education: Doctoral degree		0.303**	
		(0.133)	
Employment status: Part time			-0.180

Continued on next page

Table A17 – continued from previous page

			(0.160)
Employment status: Looking for work			-0.474
			(0.460)
Employment status: Unemployed			-0.156
			(0.190)
Employment status: Retired			-0.345**
			(0.136)
/			
cut1	1.281***	1.348***	1.162***
	(0.309)	(0.320)	(0.322)
cut2	2.128***	2.211***	2.043***
	(0.312)	(0.323)	(0.325)
cut3	2.757***	2.841***	2.650***
	(0.314)	(0.325)	(0.327)
Observations	866	851	738
Max. log-likelihood	-1096.53	-1069.65	-924.36

t in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

G Complete estimation results for the Extent-of-Market Model (including interaction terms in Model (5) with statistically insignificant coefficients)

Table A18: Market Extent Models with Intensity Weights and Sample Selection Corrections

	(1) Naive	(2) Weights only	(3) Probit IMR	(4) Ordered probit IMR	(5) Demeaned propensity interact.
Explanatory variables:					
1(Employed)	-0.0400	-0.104	-0.140**	-0.134**	-0.211**
	(0.0693)	(0.0783)	(0.0674)	(0.0675)	(0.0894)
1(Have Self-Reported Income Bracket)	-0.420***	-0.456***	-0.370***	-0.420***	-0.539***
	(0.127)	(0.149)	(0.132)	(0.133)	(0.174)
ln(Income in 10Ks, If Reported)	0.210***	0.226***	0.173***	0.185***	0.230***
	(0.0532)	(0.0608)	(0.0538)	(0.0537)	(0.0811)
1(Female)	-0.126**	-0.0842	0.124*	0.103*	0.145
	(0.0587)	(0.0612)	(0.0648)	(0.0627)	(0.0973)
1(Age $<$ 45)	0.175**	0.205**	-0.292***	-0.158	0.371
	(0.0831)	(0.0867)	(0.110)	(0.0963)	(0.298)
1(Age $>$ 64)	-0.0407	-0.0587	0.315***	0.129*	0.124
	(0.0709)	(0.0768)	(0.0883)	(0.0728)	(0.0877)
1(Education - Some College or Less)	-0.0973	-0.0868	0.0527	0.0899	0.106
	(0.0842)	(0.0945)	(0.0927)	(0.0942)	(0.0977)
1(Education - Some Grad School or More)	0.0700	0.0481	-0.198**	-0.119*	-0.237**
	(0.0672)	(0.0677)	(0.0776)	(0.0712)	(0.105)
1(Not Interested in Perching Birds)	-0.514***	-0.521**	-0.428**	-0.471**	-0.402**

Continued on next page

	(0.169)	(0.204)	(0.200)	(0.200)	(0.196)
1(Not Interested in Wading Birds)	-0.253	-0.254	-0.0743	-0.0352	-0.111
	(0.180)	(0.188)	(0.197)	(0.200)	(0.208)
1(Not Interested in Waterfowl)	-0.216	-0.255	-0.0855	-0.0756	-0.178
	(0.166)	(0.178)	(0.186)	(0.188)	(0.198)
1(Not Interested in Other Game Birds)	-0.447***	-0.427***	-0.492***	-0.503***	-0.480***
	(0.117)	(0.122)	(0.124)	(0.126)	(0.138)
1(Not Interested in Birds of Prey)	-0.243	-0.170	-0.262	-0.244	-0.217
	(0.174)	(0.179)	(0.173)	(0.173)	(0.168)
Different inverse Mills ratio estimates, for Models (3) and (4):					
Binary Probit IMR			-0.620***		
			(0.0932)		
Ordered probit IMR				-1.439***	
				(0.213)	
Selected interactions with demeaned predicted engagement propensity, for Model (5):					
1(Employed) × Demeaned Engagement Propensity					0.0779
					(0.113)
1(Have Self-Reported Income Bracket) × Demeaned Engagement Propensity					0.334
					(0.220)
ln(Income in 10Ks, If Reported) × Demeaned Engagement Propensity					-0.105
					(0.0949)
1(Female) × Demeaned Engagement Propensity					-0.0174
					(0.111)
1(Age $<$ 45) × Demeaned Engagement Propensity					-0.629***
					(0.230)
1(Age $>$ 64)					0.168

Continued on next page

× Demeaned Engagement Propensity					(0.167)
1(Education - Some College or Less)					-0.162
× Demeaned Engagement Propensity					(0.145)
1(Education - Some Grad School or More)					0.127
× Demeaned Engagement Propensity					(0.119)
1(Not Interested in Perching Birds)					-0.0599
× Demeaned Engagement Propensity					(0.243)
1(Not Interested in Wading Birds)					0.00850
× Demeaned Engagement Propensity					(0.229)
1(Not Interested in Waterfowl)					0.0699
× Demeaned Engagement Propensity					(0.202)
1(Not Interested in Other Game Birds)					0.0335
× Demeaned Engagement Propensity					(0.176)
1(Not Interested in Birds of Prey)					0.255
× Demeaned Engagement Propensity					(0.220)
Constant	4.123***	4.133***	5.429***	4.562***	3.898***
	(0.104)	(0.115)	(0.220)	(0.123)	(0.145)
Demeaned Engagement Propensity					0.514**
					(0.212)
$ln(\sigma)$					
Constant	-0.0879***	-0.0946***	-0.121***	-0.123***	-0.127***
	(0.0237)	(0.0270)	(0.0268)	(0.0269)	(0.0270)
Predicted maximum one-way distances for baseline individual:					
median (miles)	61.7	62.4	228.	95.8	49.3
mean (miles)	93.9	94.3	337.	142.	72.6
Observations	1081	1081	1081	1081	1081
Log Likelihood	-2404.01	-2403.52	-2377.36	-2375.24	-2371.15

Continued on next page

AIC	4838.02	4837.05	4786.72	4782.48	4800.29
BIC	4912.81	4911.83	4866.49	4862.25	4944.88
Weighted?	No	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

H Additional Figures

These figures supplement the figures in the main paper

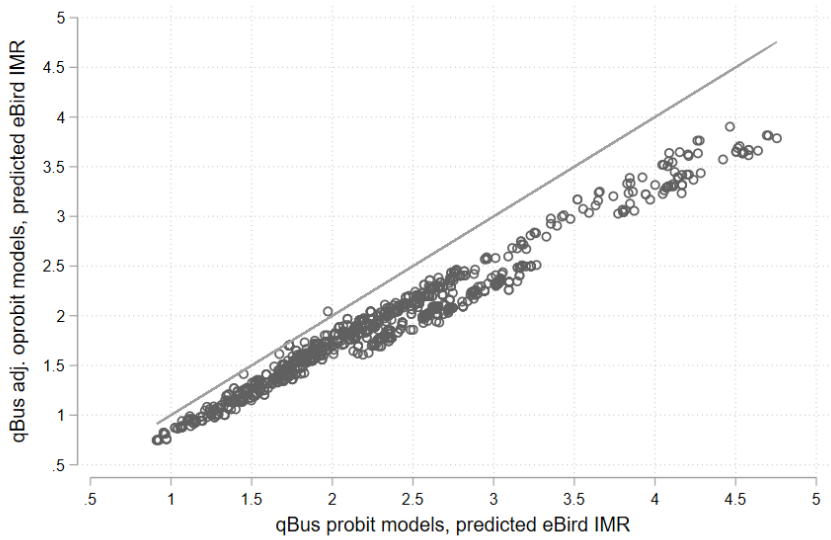

Figure A1: eBird sample: joint distribution of binary-probit-based IMRs and ordered-probit-based IMRs, both calculated from qBus parameters for the most-detailed specification consistent with any missing Z_j data in the eBird sample.

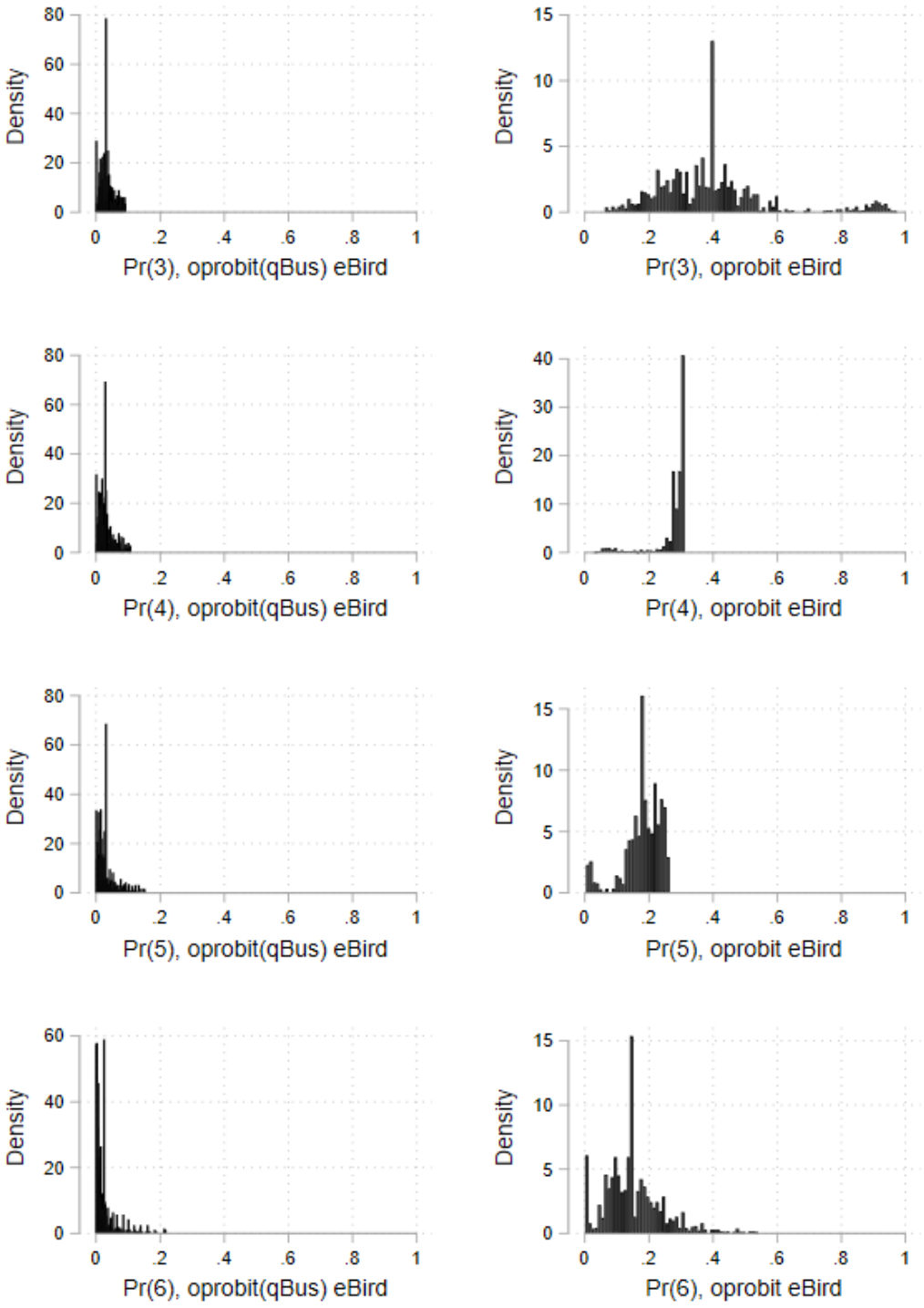

Figure A2: Compare eBird sample expected probabilities based on qBus models to fitted probabilities based on eBird models.

References

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

Cameron, T. A. and J. Englin (1997). Respondent experience and contingent valuation of environmental goods. *Journal of Environmental Economics and Management 33*(3), 296–313.

De Valck, J. and J. Rolfe (2018). Spatial heterogeneity in stated preference valuation: Status, challenges and road ahead. *International Review of Environmental and Resource Economics 11*(4), 355–422.

Greene, W. H. (2012). *Econometric Analysis (7 th)*. Upper Saddle River, NJ: Prentice Hall.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, 153–161.

Johnston, R. J. and A. S. Abdulrahman (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., S. A. Wood, S. Polasky, C. Kling, C. T. Filstrup, and J. A. Downing (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 T. A. Cameron (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., T. A. Cameron, and C. Wilsey (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.

Loomis, J. B. (1996). How large is the extent of the market for public goods: evidence from a nationwide contingent valuation survey. *Applied economics 28*(7), 779–782.

Rodrigue, J.-P., C. Comtois, and B. Slack (2016). *The geography of transport systems*. Routledge.

Rosenberg, K. V., A. M. Dokter, P. J. Blancher, J. R. Sauer, A. C. Smith, P. A. Smith, J. C. Stanton, A. Panjabi, L. Helft, M. Parr, and P. P. Marra (2019). Decline of the north american avifauna. *Science*.

Sonter, L. J., K. B. Watson, S. A. Wood, and T. H. Ricketts (2016). Spatial and temporal dynamics and value of nature-based recreation, estimated via social media. *PLoS one 11*(9), e0162372.

Terza, J. V. (1998). Estimating count data models with endogenous switching: Sample selection and endogenous treatment effects. *Journal of Econometrics 84* (1), 129–154.

Vella, F. (1998). Estimating models with sample selection bias: a survey. *Journal of Human Resources*, 127–169.

Wood, S. A., A. D. Guerry, J. M. Silver, and M. Lacayo (2013). Using social media to quantify nature-based tourism and recreation. *Scientific reports 3*, srep02976.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.-2002.

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