



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
<http://ageconsearch.umn.edu>
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

**A generalized latent class logit model of discontinuous preferences in repeated discrete choice data:
an application to mosquito control in Madison, Wisconsin.**

Zachary S. Brown^{1*}, Katherine L. Dickinson^{2,3}, Susan Paskewitz⁴

¹ Assistant Professor of Agricultural and Resource Economics, North Carolina State University

² Project Scientist, University Corporation for Atmospheric Research

³ Research Scientist, University of Colorado at Boulder

⁴ Professor of Entomology, University of Wisconsin at Madison

* Corresponding author: zack_brown@ncsu.edu

Selected Paper prepared for presentation at the 2015 Agricultural & Applied Economics
Association and Western Agricultural Economics Association Annual Meeting, San Francisco,
CA, July 26-28

Copyright 2015 by the authors. All rights reserved. Readers may make verbatim copies of this
document for non-commercial purposes by any means, provided that this copyright notice
appears on all such copies.

Abstract

Serial nonparticipation in nonmarket valuation using choice data is a pattern of behavior in which an individual always appears to choose the status quo or ‘no program’ alternative. From a choice modelling perspective serial nonparticipation may be viewed as belonging to a class of ‘discontinuous preferences,’ which also includes other behavioral patterns, such as serial participation (*never* choosing the status quo), as well as lexicographic preferences (e.g. always choosing the alternative with the greatest health benefit). Discontinuous preferences are likely to be especially relevant in the context of environmental goods, due to the lack of familiarity that individuals have with valuing these goods in markets. In the case of discrete choice data, logit-based choice models are ill-equipped for identifying such preferences, because conditional logit choice probabilities cannot take a value of zero or one for any finite parameter estimates. Here we extend latent class choice models to account for discontinuous preferences. Our methodological innovation is to specify for each latent class a subset of alternatives that are avoided with certainty. This results in class membership being partially observable, since we then know with certainty that an individual does not belong to a class if she selects any alternatives avoided by that class. We apply our model to data from a discrete choice experiment on mosquito control programs to reduce West Nile virus risk and nuisance disamenities in Madison, Wisconsin. We find that our ‘generalized latent class model’ (GLCM) outperforms standard latent class models in terms of information criteria metrics, and provides significantly different estimates for willingness-to-pay. We also argue that GLCMs are useful for identifying some alternatives for which valuation estimates may not be identified in a given dataset, thus reducing the risk of invalid inference from discrete choice data.

Keywords: discrete choice econometrics, latent class models, partial observability, serial nonparticipation, serial participation, discontinuous preferences, E-M algorithm

JEL codes: Q51, C35

Introduction

A current challenge in using discrete choice methods to value nonmarket goods is the consistent treatment of alternatives which are excluded by different subgroups in a population in ways that are not fully observable to the econometrician. Well-studied examples of this issue include ‘status quo’ (SQ) effects and ‘serial nonparticipation’ in both choice experiment and revealed preference data. These effects arise when a substantial subset of decision-makers choose the ‘status quo’ or ‘opt out’ alternative repeatedly, often to the exclusion of alternatives whose attributes are the objects of study. The converse phenomenon – ‘serial participation’ or the complete avoidance of the SQ alternative – is also possible. The standard class of logit-based choice models are ill-suited for these situations. Yet methods for dealing with such effects are important for reducing bias in valuation estimates and for predicting individuals’ responses to new policies. In this paper we provide a novel econometric method for estimating SQ effects, and the broader phenomena of ‘discontinuous preferences,’ in discrete choice data.

Preference discontinuities are important to consider in nonmarket valuation of environmental goods. Individuals are often not familiar with the goods under study, particularly in the case of nonuse values where stated preference methods are most frequently used. Consequently, individuals may be more likely to rely on prior attitudes and beliefs in valuing these goods, and in deciding whether or not to participate in hypothetical markets for such goods. In stated preference choice experiments, SQ effects have been conjectured to represent protest responses on the part of those who do not see any value in the goods or services being evaluated, or who reject the nature of the hypothetical choice task in the survey (von Haefen et al. 2005). It has also been proposed that choice task complexity gives rise to SQ effects, and that repeated selection of the SQ represents a lack of understanding or engagement in a choice experiment (Meyerhoff & Liebe 2009).

It is important to emphasize that the types of choice discontinuities, such as biases for or against the SQ, do not necessarily imply irrational choice behavior. In both a stated and revealed preference setting, lexicographic preferences among subgroups of individuals can give rise to a high frequency of SQ selection, as well as a complete avoidance of specific alternatives.¹ Burton and Rigby (2008) illustrate this point in a hypothetical choice experiment studying preferences about genetically modified (GM) food in the UK. In a revealed preference setting, von Haefen et al. (2005) observe that serial nonparticipation is also a common phenomenon, observed as subgroups of the population who are nonusers of the public goods under study, e.g. respondents who never recreate at any of the sites studied in a travel cost survey.

SQ effects are only one instance within a broad set of issues arising in discrete choice datasets which contain observations of individuals' repeated choices. In general, as choices are observed more frequently, the more likely we are to notice some individuals who appear to completely avoid – or gravitate towards – specific alternatives (Lancsar & Louviere 2006; Louviere 2013). In the case of serial nonparticipation, individuals avoid the SQ alternative, whereas serial participants never select the SQ. Yet in general individuals may exhibit a variety of 'discontinuous' choice behaviors: in the presence of contextual factors, such as background risk or ambiguity, individual behavior may suggest a threshold above which individuals choose a risk-mitigating alternative, or below which they never choose such an alternative.

All of these are situations which econometric models based only on logit choice probabilities are ill-equipped for addressing, as we illustrate later in this paper. The essence of the problem is that data in which one alternative is selected or avoided with certainty imply best-

¹ Lancsar and Louviere (2006) note how lexicographic preferences can still satisfy the Weak and Strong Axioms of Revealed Preference. As such, they argue against excluding from the estimation sample respondents who have been labelled as irrational only on the basis of evident serial nonparticipation behavior. The method we introduce here can provide a post-hoc approach for dealing with these situations.

fit logit coefficients which are not finite. In a standard conditional logit framework, in which one representative set of preferences is estimated for the entire dataset, this problem rarely arises as long as there is a broad cross-section of individuals in the sample. However, as logit-based discrete choice models have been embellished to provide more detailed characterization of individual preference heterogeneity, researchers increasingly find situations wherein maximum likelihood estimates of logit model coefficients blow up to infinity. As we show below, such cases provide a ‘smoking gun’ that some individuals are exhibiting discontinuous preferences.

The model we describe and estimate in this paper extends and unifies previous approaches – latent class models and hurdle models – for dealing with serial nonparticipation and status quo effects (von Haefen et al. 2005; Burton & Rigby 2008; Thiene et al. 2012). Our approach consists of hypothesizing the existence of subgroups of individuals who appear to exhibit discontinuous choice patterns. Then, once these hypotheses are clearly defined, a variation of a latent class choice model is used to probabilistically classify each individual as belonging to the hypothesized group, as well as to other groups (including a standard conditional logit choice model). Following Thiene et al. (2012) who note that the latent class analysis is a form of data imputation and thus adaptable to cases in which class membership is observable, we modify the standard latent class approach to treat class membership as partially observable: for example, in the case of serial nonparticipation, we know definitely that an individual is not a serial nonparticipant if she selects a non-SQ alternative. In this case, our model is almost identical to a double hurdle model (von Haefen et al. 2005). However, our approach is designed for a much larger range of preference discontinuities – including serial participation, as shown below – than addressed by prior approaches. Consequently, we dub our model Generalized Latent Class logit model (GLCM).

After presenting the formulation of our model, we demonstrate its estimation on a dataset gathered from a discrete choice experiment (DCE) examining preferences for mosquito control programs in Madison, Wisconsin, aimed at reducing the risk of West Nile virus as well as the nuisance costs associated with mosquitoes. We find that our GLCM provides the best balance between fit and parsimony in modeling the data (based on information criteria statistics), as compared to standard latent class models (which we also estimate). Our GLCM estimates imply that there are three partially latent classes of respondents in the data: serial participants (estimated at 46% of the sample), those for which a standard logit model applies (39%), and finally a group (the remaining 15%) who are serial nonparticipants for low and medium West Nile risk but who behave according to a logit model in a high risk setting. We also find relative willingness-to-pay values that are more reasonable and precise, due to an identifying restriction related to the reference alternative specified in the choice model. When we include individual covariates in the model as predictors of class membership, only two factors (and their interaction effect) are found to be statistically significant: individuals with greater mosquito densities around their homes (as measured via entomological surveys) and those who report spending more time indoors are more likely to be serial participants (i.e. those who avoid the status quo alternative entirely).

Previous approaches for addressing discontinuous preferences in discrete choice data

The vast majority of discrete choice econometric analyses are based in some way on the conditional logit (CL) model (McFadden 1973). When individual-level covariates are excluded in the regression, then utility-maximizing behavior in the CL model implies that the probability of choosing alternative h in choice task t for any individual is:

$$p_{ht}^{CL} \equiv \frac{\exp \alpha_{ht}}{\sum_{h \in t} \exp \alpha_{ht}}$$

The α_{ht} 's are alternative-specific constants to be estimated, which are typically decomposed into alternative attributes, including environmental and health benefits and risks, and (in economics) monetary attributes such as cost.² For identification, one of these alternative-specific effects must be normalized to zero; typically the SQ-specific constant serves as this reference alternative. If a respondent participates, the conditional logit specification implies that the likelihood of observing their sequence of choices is:

$$l_n^{CL} \equiv \prod_t \prod_{h \in t} (P_{ht}^{CL})^{C_{htn}}$$

where C_{htn} is an indicator variable for whether individual n selected alternative h in task t .

SQ effects specifically – and discontinuous preferences more generally – pose problems for logit-based models for a very simple reason: logit-based choice probabilities cannot take a value of zero or unity over any finite values for the taste parameters α_{ht} . This issue typically does not present itself in a CL model estimated on a sample of individuals who as a whole exhibit ‘well-behaved’ choices across alternatives. But the issue is common when considering choice probabilities at an individual level. Let P_{htn} be the predicted probability that individual n chooses alternative h in choice task t out of T tasks per respondent in total. An individual level logit model (now subscripting by n) implies a choice probability equal to:

$$P_{htn} \equiv \frac{\exp \alpha_{htn}}{\sum_{h \in t} \exp \alpha_{htn}}$$

² We index the α_{ht} 's by task t , to allow task-specific factors, such as background risk. Later, we consider monetary attributes explicitly in the model we formulate below. All logit-based models like this emerge from a random utility framework, in which case the α_{ht} 's can be interpreted as the expected indirect utility from alternative h in task t and Extreme Value Type I random utility component for each alternative.

where α_{htn} is the alternative-specific logit taste parameter for individual n . If we had a large number of choice tasks for each respondent, then we could estimate this choice model directly for each individual.

If we did perform such an estimation, suppose some individual n never chooses alternative k in any of the many choice tasks in which k was included in the choice set. In this case, it is well known that maximum likelihood estimation (MLE) will often not converge to a finite solution. This is because in order for $P_{ktn} = 0$ it is necessary for $\alpha_{kn} = -\infty$: the log-likelihood function can be continually increased by decreasing α_{htn} (or by continually increasing all other $\alpha_{h'tn}$, if h is the reference alternative). Such a model is not identified.

However, it is not standard practice to fully estimate conditional logit models at the individual level (Louviere 2013 describes some exceptions; also see Czajkowski et al. 2010). Instead, assumptions are made on how α_{htn} varies across individuals. This variation can be specified in terms of individuals' observable covariates or in terms of unobservable heterogeneity. The latter can be modeled parametrically using mixed or generalized logit, or nonparametrically, for example using latent class logit (described below). But all of these models posit the same basic logit structure to choice probabilities. Clearly, individuals whose choices may not follow a logit pattern of choice behavior (as described above) could significantly bias the results of any logit model estimated on the full sample of data, which can have important consequences both for predicted choice behavior and for willingness to pay (WTP) measures which are the main outputs of choice modeling for economic analysis.

A number of approaches have been considered to address this type of bias, especially in the case of serial participation. The most direct approach is to simply estimate the preferred model on the sample of those with 'well-behaved' choice patterns. However, Lancsar and

Louviere (2006) strongly advise against this approach, as doing so can introduce bias through the researcher having to assess ex ante whose choice patterns are ‘valid.’ Von Haefen et al. (2005) and Burton and Rigby (2008), hereafter VHMA and BR, each propose alternative approaches. VHMA propose hurdle models as a way to address the specific SQ effect of serial nonparticipation (the ‘hurdle’ in this case is whether a respondent chooses a non-SQ alternative in any of the choice tasks to which she is exposed). In contrast, BR argue that the application of standard latent class (LC) choice models can address not only serial nonparticipation but also other irregularities in choices. Our model merges these two approaches.

The hurdle approach uses a binary probability model for participation (e.g. probit or logit), and then models choices using a conditional logit framework.³ The single hurdle (SH) model allows for correlation between the participation decision and the subsequent choice decision. The formulation is as follows: for each individual n , let d_n be an indicator for whether that respondent is a nonparticipant (in which case $d_n = 1$), and let p_n^{SH} be the predicted probability that a respondent is a nonparticipant (which may be modeled using a binary probability model, such as probit or logit. The motivation behind modeling a separate participation process is that certain individuals may select the SQ repeatedly as a way to reject the entire premise of the choice task (e.g. protest responses in stated choice data) or because they belong to a group of people for whom the choice task is irrelevant (e.g. non-users in recreation demand studies). This predicted probability of serial nonparticipation is indexed by n , to allow for individual covariates (or, possibly, unobserved heterogeneity). The likelihood of observing individual n ’s sequence of choices under the single hurdle model is then $(p_n^{SH})^{d_n}[(1 -$

³ As VHMA point out, this model could easily be extended to allow for randomly varying conditional logit taste parameters within the participant subsample.

$p_n^{SH})l_n^{CL}]^{1-d_n}$. The log-likelihood function to be maximized with respect to parameters contained in p_n and l_n^{CL} can then be expressed as:

$$LL^{SH} \equiv (1 - d_n) \sum_n \log l_n^{CL} + \sum_n [d_n \log p_n^{SH} + (1 - d_n) \log(1 - p_n^{SH})]$$

Note that the first summation is the log-likelihood function for the CL model and the second summation fully contains the parameters for the serial nonparticipation probability model. Thus the maximization of the SH log-likelihood is the same as estimating a CL model on a subsample of participants, and separately estimating a probability model for serial nonparticipation. The SH model therefore ignores correlation between the participation probability and the subsequent CL choice probabilities.

VHMA thus propose the double hurdle (DH) model as one approach for allowing for more integration between the two processes – participation and subsequent choice. This model, following Shonkwiler and Shaw (1996), acknowledges that there are two ways for a respondent to appear as a serial nonparticipant in the dataset: one is through a separate probability process (as in the SH case above) and the other is for a ‘standard’ decisionmaker, subject to the CL choice probabilities, to repeatedly choose the SQ, perhaps due to extreme (but finite) values for the CL taste parameters α_n . Thus, the probability for serial nonparticipation in the DH model is:

$$p_n^{DH} \equiv p_n^{SH} + (1 - p_n)^{SH} \left(\prod_t P_{SQ,t}^{CL} \right)$$

where $P_{SQ,t}^{CL}$ is the predicted probability, within the CL model, of choosing the SQ in choice task t . The log-likelihood function for the DH model replaces the p_n^{SH} terms in the LL^{SH} formula above with p_n^{DH} . In this way the DH model is aimed at consistently capturing the problem of ‘excess’ SQ choices in the way that zero-inflated Poisson models capture excess zeros in count

data (Shonkwiler & Shaw 1996). The DH model therefore acknowledges that, given a finite number of choice tasks observed per individual, there is some positive probability that serial nonparticipation can emerge from a standard CL model. This is a key point in the design of our generalized latent class model presented below. In their application (using both a stated and revealed preference dataset), VHMA find that the SH and DH models more parsimoniously represent the data than standard approaches, including the CL model and mixed logit.

BR take a markedly different approach for studying serial nonparticipation. They use a latent class model (LCM) to probabilistically classify respondents across a range of logit-based choice models. Beyond issues of serial nonparticipation and the like, LCMs have been extensively used as a way to model individual heterogeneity in choice data (Morey et al. 2006; Boxall & Adamowicz 2002). LCMs are sometimes contrasted with mixed and generalized multinomial logit models as a way to nonparametrically capture preference heterogeneity (e.g. Greene & Hensher 2003). In BR's application, one of the classes is hypothesized – though not restricted – to represent serial nonparticipants.

LCMs posit the existence of finite number M of latent classes, indexed throughout this paper by m (for *model*), to which every individual belongs. Conditional on an individual belonging to class m , the choice probability takes the usual conditional logit form, but with the parameters now indexed by class:

$$p_{ht}^m \equiv \frac{\exp \alpha_{ht}^m}{\sum_{h \in t} \exp \alpha_{ht}^m}$$

The above requires the usual identifying restriction of having one of the alternative-specific effects restricted to zero; usually this is the SQ effect ($\alpha_{SQ,t}^m = 0 \forall m, t$). Conditional on the individual belonging to class m , the likelihood of an individual's observed sequence of choices is therefore:

$$l_{n|m} = \prod_t \prod_{h \in t} (P_{ht}^m)^{c_{htn}}$$

Let p_n^m be the predicted probability of individual n belonging to class m . LCMs then use a separate multinomial probability model (usually also logit-based) for these predicted probabilities. Because class membership is assumed to be latent, LCMs impute class membership using Bayes' rule. The imputed value \hat{d}_n^m for an individuals' likelihood of belonging to class m , conditional on the predicted probability p_n^m and the conditional likelihood of their observed choices $l_{n|m}$ is:

$$\hat{d}_n^m \equiv \frac{l_{n|m} p_n^m}{\sum_{\tilde{m}} l_{n|\tilde{m}} p_n^{\tilde{m}}}$$

Using these imputed values in place of the 'missing' class membership data, the log-likelihood function for an LCM is:

$$LL^{LC} \equiv \sum_m \sum_n \hat{d}_n^m \log l_{n|m} + \sum_m \sum_n \hat{d}_n^m \log p_n^m$$

Some rough comparison can be made between the LL function for this model and the SH model above: if class membership were directly observable, rather than imputed in \hat{d}_n^m , the first double-summation over m would correspond to summing over separable CL models for each class, and the second double-summation would correspond to the LL function for estimating a multinomial class membership probability model. That is, if we had observed d_n^m as data, rather than imputing \hat{d}_n^m , the above would equate to estimating each CL model separately, as well as separately estimating the class membership model. However, imputing class membership as above eliminates this separability, and – as with the DH model – links the class membership and choice probability models.

Using this LC approach in an application to a DCE related to genetically modified (GM) food, BR find evidence for multiple classes who exhibit different types of SQ effects: one class appears to contain genuine serial nonparticipants, with a very high estimated SQ effect, whereas another class is strongly biased against GM food, and thus appears as otherwise similar to serial nonparticipants. This highlights a key advantage of the LC approach: it can capture a wide variety of behavioral patterns beyond simply serial nonparticipation. While not arising in BR's analysis, serial participation can also be captured in a LC framework (as we demonstrate below).

Econometrically, serial nonparticipation here is captured by large estimated values for α_{SQ}^{SNP} , the SQ effect in the serial nonparticipant (SNP) class. As BR point out, the “high and limiting value for that parameter” captures the fact “all those within the [SNP] class have a very high probability of selecting that option throughout the choice sequence, regardless of other attribute values.” This high, limiting value – while useful for exploratory examination of the data – can imply a lack of identification, for the reasons discussed at the beginning of this section. While BR estimate finite coefficients for the SQ effect in the SNP class, it seems reasonable that this effect should in fact be such that $\alpha_{SQ}^{SNP} = \infty$, so that true members of this class select the SQ with probability equal to one. In terms of economic valuation, members of such a class in principle have an infinite willingness-to-accept (WTA) a non-SQ alternative: no amount of money offered to SNPs (or at least no finite amount identified within the scope of the choice experiment) would induce them away from the SQ.

Another drawback of the standard LC approach is the fact that class membership is partially observable. In the case of serial nonparticipation, an individual who selects a non-SQ alternative can be classified with certainty as *not* belonging to the SNP class. Thiene et al (2012) note this observability of SNP behavior and incorporate this into a LC framework, by specifying

an SNP class with fully observable membership probabilities (and leaving the other preferences as latent and unobservable). Yet this approach does not acknowledge the fact that there is a positive probability (albeit decreasing with the number of observed choice tasks per respondent) that an SNP behavioral pattern can emerge with some positive probability from a standard conditional logit model of preferences. In this way, Thiene et al.’s approach can be viewed as a latent class extension of the single hurdle model of VHMA.

A generalized latent class logit model

Our approach jointly addresses this issue of unbounded valuation estimates and partial observability of SP or SNP for subgroups of respondents (as well as variety of other possibly discontinuous preferences). As with the standard latent class model, our GLCM supposes that there are M classes, indexed by m , which capture individual preference heterogeneity. The GLCM departs from the standard approach because of one additional assumption. The GLCM requires the econometrician to specify a subset $\emptyset_{m,t}$ of alternatives which are excluded from class- m individuals’ choice sets in task t . An implication of this assumption is that class membership is partially observable, since we then know with certainty that an individual does *not* belong to class m if she chooses an alternative h from the exclusion set $\emptyset_{m,t}$.

There are a number of causal hypotheses as to why nonempty exclusion sets may be justified for certain classes. Individuals may exclude alternatives from their choice sets for a number of reasons: members of such classes could exhibit truly discontinuous preferences (e.g. always choosing the alternative that offers a positive health benefit, regardless of the amount that it costs). Or members of these classes may have finite, logit-based taste parameters, but ones that are so large in relation to the excluded alternatives’ attributes that these alternatives are never selected in the observed tasks. (In the case of choice experiments, this would imply a failure of

the experimental design to properly account for individuals with more extreme – if still continuous – preferences.) The excluded alternative could be systematically overlooked by some individuals (e.g. due to a lack of survey engagement, see Hess & Stathopoulos 2013). Finally, the alternative could simply be inapplicable to a subset of respondents for an unobserved reason, for example, nonusers of recreational resources (von Haefen et al. 2005).

The rest of this section simply carries out the mathematical formulation implied by this logic. With respect to nonexcluded alternatives, individuals are assumed to exhibit standard logit choice probabilities, as in the LC logit model. In order to retain a focus on valuation, we also explicitly separate monetary attributes from the alternative-specific constant (embodying the alternatives' other attributes). We assume that the cost of alternative h in task t is q_{ht} and the marginal utility of money for class m is η_m . The class-specific predicted choice probabilities for the GLCM are therefore:

$$P_{htm} \equiv \begin{cases} 0 & , \quad h \in \emptyset_{m,t} \\ \frac{\exp(\alpha_{ht}^m - \eta_m q_{ht})}{\sum_{h \notin \emptyset_{m,t}} \exp(\alpha_{ht}^m - \eta_m q_{ht})} & , \quad h \notin \emptyset_{m,t} \end{cases} \quad (1)$$

As with LCMs, the likelihood of observing an individual n 's sequence of choices, conditional on them belonging to class m is then $l_{n|m} \equiv \prod_t \prod_{h \in t} P_{htm}^{C_{nht}}$, which provides the basis for MLE. This is a general formulation that captures serial nonparticipation (with the excluded set consisting of everything but the SQ), serial participation (with the excluded set consisting *only* of the SQ), as well as variety of other behaviors (e.g. excluding the alternative conveying the lowest health improvement in a given choice task). Note that this formulation requires the econometrician to take extra care with identification in specifying the reference alternative (for each class). If the SQ is excluded from consideration, for example with serial participants, then a new reference

alternative must be specified. This means that WTP or WTA values within each class can only be estimated relative to the (potentially class-specific) reference alternative. The implications of this point are discussed below, in the empirical application.

We now turn to the implied partial observability of individual class membership in the GLCM: if an individual is ever observed selecting an alternative $h \in \emptyset_{m,t}$, then we can be sure that she does not belong to class m . Nevertheless, individuals who *never* select an alternative from the excluded set $\emptyset_{m,t}$ cannot be classified with certainty, because we only ever observe a finite number of decisions: An individual who always selects the SQ out of 9 choice tasks, for example, can only be probabilistically classified as a serial nonparticipant, because on the tenth choice task she may have chosen a non-SQ alternative. Consequently, for these individuals, we use the same Bayesian imputation formula as in the standard LCM.

This logic is captured in the following specification for class membership. Let d_n^m be the observed class membership variables:

$$d_n^m \equiv \begin{cases} 0 & \text{if } C_{nht} = 1 \text{ for any } h \in \emptyset_{m,t} \\ \text{missing} & \text{otherwise} \end{cases} \quad (2)$$

Then we only use Bayesian imputation for the missing class membership information. The imputed class membership indicator when d_n^m is missing, using Bayes' rule is:

$$\hat{d}_n^m \equiv \frac{l_{n|m} p_n^m}{\sum_{\{\tilde{m} | d_n^{\tilde{m}} \text{ missing}\}} l_{n|\tilde{m}} p_n^{\tilde{m}}} \quad (3)$$

Then the class membership variable \tilde{d}_n^m used in estimation, based on observed values where possible and imputation otherwise is:

$$\tilde{d}_n^m \equiv \begin{cases} 0 & \text{if } C_{nht} = 1 \text{ for any } h \in \emptyset_{m,t} \\ \hat{d}_n^m & \text{otherwise} \end{cases} \quad (4)$$

Given this formulation, the log-likelihood for the GLCM is the same as for the standard LCM, except using equations (1) – (4) in place of the standard formulation:

$$LL^{GLCM} \equiv \sum_m \sum_n \tilde{d}_n^m \log l_{n|m} + \sum_m \sum_n \tilde{d}_n^m \log p_n^m \quad (5)$$

As in the standard LCM, p_n^m is the predicted probability that individual n belongs to class m .

Usually, p_n^m is parameterized using a multinomial logit probability function:

$$p_n^m = \frac{\exp \gamma_m X_n}{\sum_{\tilde{m}} \exp \gamma_{\tilde{m}} X_n} \quad (6)$$

where X_n is a vector of respondent covariates that may predict class membership (including a vector of ones to allow for a regression constant), and the γ_m 's are conformable vectors of coefficients to be estimated (with the identifying restriction of one class's coefficients set to zero).

As with LCMs, estimation can use both the iterative EM algorithm, as well as direct maximum-likelihood estimation, employing the gradient and Hessian of the full likelihood function (eq. 5) to compute regular or robust standard errors (Train 2009). In supplementary material, we provide a formula for the analytical gradient of the above likelihood function, which is a minor generalization of the standard LC gradient. This greatly reduces computation time in estimation and in the computation of robust standard errors, as can be verified in the Matlab computer code implementing the model (also included as supplementary material).

Relationship to standard latent class and hurdle models

The GLCM bridges the advantages of hurdle models and latent class models in addressing not only serial nonparticipation, but also a much broader range of ‘weird’ preferences. GLCM reduces to a standard LCM by setting $\emptyset_{m,t} = \emptyset$ for all t and m , i.e. all alternatives retain a positive probability of selection in all tasks and classes. In this special case, we do not observe any information about class membership (since no alternatives are excluded from any of the choice sets), and so class membership is completely imputed via Bayes’ rule.

In the case of serial nonparticipation (SNP), GLCM can also be reduced to a hurdle model. This can be seen by supposing there are only two classes $m \in \{CL, SNP\}$: a normal conditional logit (CL) class and a SNP class, with $\emptyset_{SNP,t} = \{h \neq SQ\}$ for all tasks t (i.e. all alternatives but the SQ are excluded from the SNP choice set). The conditional likelihood for the SNP class in this case is degenerate, with $\log l_{n|SNP} = 0$, because $l_{n|SNP} = \prod_t (P_{SQ,t,SNP}^{C_{n,SQ,t}} \times \prod_{h \neq SQ} P_{h,t,SNP}^{C_{n,h,t}}) = \prod_t (1^1 \times \prod_{h \neq SQ} 0^0) = 1$. Consequently, the GLCM function reduces to:

$$LL^{GLCM \rightarrow Hurdle} = (1 - \tilde{d}_n^{SNP}) \sum_n \log l_{n|CL} + \sum_n [\tilde{d}_n^{SNP} \log p_n^{SNP} + (1 - \tilde{d}_n^{SNP}) \log(1 - p_n^{SNP})] \quad (7)$$

This is almost identical to the log-likelihood function for the hurdle model, presented in the literature review above, except that the fully observable indicator d_n for SNP in the hurdle model is replaced with the partially observable indicator \tilde{d}_n^{SNP} (defined in equations 3 and 4). This formulation is therefore similar in spirit to the DH model, because the hurdle/class probability model and the CL model must be estimated jointly. In GLC logit this joint estimation arises because $l_{n|CL}$ is used in Bayes’ rule to impute missing values for \tilde{d}_n^{SNP} , whereas the DH model

considers SNP classification itself to be fully observed but with two latent processes giving rise to SNP, one of which is determined by the CL choice model.

Valuation estimates from the GLCM

As with the standard LCM, WTP estimates are class-specific and the overall sample mean WTP is the weighted mean across classes using the posterior class membership probabilities as weights (e.g. Scarpa et al. 2005). Because the choice structure here assumes only alternative-specific constants α_{ht}^m in conjunction with a monetary attribute with taste parameter η_m , we estimate nonmarginal values for each alternative. This is not a necessary feature for the general GLCM structure, but only used because of the above setup, which was chosen for ease of exposition and because of the nature of the data (described below).

To obtain aggregate, nonmarginal welfare measure across classes, we can calculate the compensating variation (CV) associated with making alternative h available in task t . For class m , CV is $\log(1 - \tilde{P}_{htm}) / \eta_m$ (McConnell 1995), where \tilde{P}_{htm} is the probability of a class m individual selecting alternative h in task t in an *appropriately specified* baseline situation (defined below). Averaging over the sample, accounting for heterogeneous preferences, the mean CV across the sample is then:

$$\overline{CV}_{ht} \equiv -\frac{1}{N} \sum_{n=1}^N \sum_m \tilde{d}_n^m \frac{\log(1 - \tilde{P}_{htm})}{\eta_m} \quad (8)$$

Care is required in defining the baseline situation. A key point of the GLCM is that the data simply may not permit estimating valuation estimates which are of most interest to the researchers, due to discontinuities in preferences. Consider the case of serial participants ($m = \text{"SP"}$): in this case $P_{0,t,SP} = 0$, i.e. the probability of SPs selecting the status quo is zero. If we are considering a valuation scenario in which *any* alternative h is made available relative to a

baseline case of no alternatives being available, then the probability of selecting any alternative is one and the CV for this class is infinite, which carries through to the whole sample so that $\overline{CV}_{ht} = \infty$. Thus, when the data support the existence of SPs, then we can only obtain the value of alternative h 's availability relative to other, non-SQ alternatives. This is not a limitation of the GLCM (we argue). Rather it is a limitation implied by the data, which the GLCM can uncover.

Therefore the appropriate conditional choice probability to be used in computing CV is:

$$\tilde{P}_{htm} = \frac{\exp \alpha_{th}^m}{\sum_{h \in \tilde{H}_t} \exp \alpha_{th}^m} \quad (9)$$

where \tilde{H}_t is some appropriately specified baseline set of all alternatives *common to all classes*.

The only rule on \tilde{H}_t required to maintain a finite CV estimate is that there are at least two alternatives in \tilde{H}_t that are not in *any* of the exclusion sets $\emptyset_{m,t}$ (i.e. which have a nonzero predicted probability of selection by members of all latent classes).

Model selection

An additional consideration is how the econometrician should choose between competing specifications for the GLCM and LCMs. There is a growing literature on the unresolved question of how to choose the 'right' number of classes in standard LCMs, as well as how to choose between an LC specification or another parametric specification for preference heterogeneity (such as mixed logit, Strazzera et al. 2013). What continues to be the typical approach for selecting the number of classes in LCMs is to use the various information criteria statistics such as the AIC, CAIC and BIC (Morey et al. 2006; Rungie et al. 2012).

For the purposes of testing different specifications of our GLCM, it is important to note that a GLCM can be viewed as a restriction of a standard LCM. This is because a standard LCM can be written as a GLCM with empty exclusion sets ($\emptyset_{m,t} = \emptyset$ for all m, t). In other words a

GLCM can be viewed as a LCM with some alternative-specific constants restricted to $-\infty$. Such an observation would motivate the use of a likelihood ratio (LR) test of the restricted GLCM against the more general (though possibly unidentified) LCM. Such a test applied to our case would take the GLCM as the null hypothesis and test whether this could be rejected in favor of its unrestricted LCM counterpart. Unfortunately, the GLCM restrictions lie at the boundary of the parameter space, making a standard LR test using a χ^2 distribution invalid (Greene 2011).⁴ Furthermore, using a standard χ^2 distribution in such a test has been found to be too conservative in structural equation models (Stoel et al. 2006): in our case this would mean that such a test would too frequently fail to reject the GLCM. We therefore set aside for future work the derivation of the distribution of the LR statistic in this nonstandard case, and settle for comparing the log-likelihood and information criteria of GLCMs against comparable LCMs (and alternative GLCM specifications).

Empirical application: Mosquito control in Madison, Wisconsin

We demonstrate application of the GLCM by analyzing data from a stated discrete choice experiment (DCE) analyzing the value of hypothetical mosquito control programs designed to reduce mosquito abundance in Madison, Wisconsin. In particular, the DCE was designed to measure the value of two types of mosquito-related disamenities: (1) nuisance and (2) disease risk associated with exposure to West Nile virus (WNV). The average risk of WNV at the time of data collection (2009) was one illness in 250,000 per year. Because nuisance and WNV-transmitting mosquitoes are distinct species in Madison, mosquito control programs could be designed to differentially target the different mosquito types. The DCE was conducted as part of

⁴ A Vuong test (of the differences in BIC) could in principle be used to test between the GLCM and LCM without specifying either as the null or alternative, though a nonstandard distribution for this statistic would still be required.

a web-based survey of homeowners in Madison in 2009. Mosquito abundance in the vicinity of the surveyed households was also measured. This mosquito exposure data, along with respondents' characteristics in the survey, are used below as the predictor X_n variables in the class membership model (according to equation 6 above). Complete data for 257 respondents was obtained, and this is the sample we analyze here. Details on the overall survey are included in the supplementary material of the paper.

The format of the choice experiments was as follows. Respondents read a short background section informing them of the fact that there are multiple types of mosquitoes in Madison, some of which are simply a nuisance while others are capable of transmitting WNV. We explained that a hypothetical citywide mosquito control program, which would use environmentally-friendly methods to control mosquito larvae, could target nuisance mosquitoes, WNV mosquitoes, or all mosquitoes. Respondents were told that the cost of the program would be funded through an increase in property taxes of between \$10 and \$200 per household. We also told respondents the level of West Nile disease risk (set at the current level of 1 in 250,000 for the first three choice tasks, then increased to 10 in 250,000 and then 100 in 250,000), and then asked respondents to choose between pairs of hypothetical control programs. The attribute descriptions and levels are presented in Table 1. A representative choice task is shown in the supplementary material. Importantly, respondents were allowed to “opt out” and choose neither program, which defines the status quo (SQ) alternative.

[Table 1 here.]

Results from the survey pre-test and focus group discussion suggested that it was possible to conduct nine choice tasks with respondents in the final survey, where each task consisted of a comparison of two hypothetical mosquito control programs under a specified disease risk level.

To generate these choice sets, we used a fractional factorial design (Johnson et al. 2007). First, we constructed all of the unique programs consisting of 1) type of mosquito controlled, and 2) cost. Since there were three different mosquito types (nuisance, vector, both) and four cost levels, this produced 12 possible programs. The set of choice tasks was further reduced by eliminating dominated alternatives. This included eliminating tasks with any alternative which controlled all mosquitoes at lowest cost, as well as any tasks in which two alternatives were identical except for one being cheaper. This narrowed the number of unique, undominated program pairs down to 28. Of these, we selected 18 pairs in which 6 compared nuisance and WNV programs, 6 compared nuisance programs with programs controlling all mosquitoes, and 6 compared WNV and all mosquito programs, and verified that the resulting attribute matrix had full rank and was orthogonal (Johnson et al. 2007). We then interacted these 18 program pairs with the three WNV risk levels (current risk, slight increase, large increase) to create 54 total choice pairs, and divided these into six unique sets of nine choice tasks. That is, we created six different versions of the survey in which each version contained three choice tasks at each of the three West Nile risk levels. Finally, in distributing the survey, we ensured that recruits from each of the six targeted neighborhoods were distributed across the six survey versions (described in the supplementary material).

[Table 1 here.]

Summary inspection of respondents' behavioral patterns (Figure 1) in the choice experiment reveal that 47% of the sample are potential serial participants: these respondents avoided the SQ in all nine of their choice tasks. Potential serial nonparticipants (always selecting the SQ) comprise 9% of the sample. However, an additional 6% always selected the SQ in the low- and medium risk scenarios (comprising six of the nine choice tasks), but did opt for some

control program in the three tasks pertaining to the high-risk scenario. This ‘quasi-SNP’ behavior is found to be important for reasons discussed below.

[Figure 1 here.]

Econometric analysis and estimation results

The recommended approach to implementing the GLCM is not only to visually inspect the data, but to also first implement a standard LC estimation routine. Such estimation is exploratory. In considering application of a GLC approach the researcher is searching the standard LCM for very large coefficient estimates (as well as nonconvergence of the estimation routine). A rigorous LC analysis is likely to be more systematic than *ad hoc* visual inspection at identifying segments of the population who appear to behave according to discontinuous preference structures. In point of fact the non-logit choice patterns illustrated in Figure 1 were originally discovered in the dataset by estimating an LCM. Once these preference patterns are revealed, the GLCM is implemented by imposing the implied restrictions on the excluded sets and reference alternatives.

Table 2 presents the regression diagnostics for the standard LCM (along with those for GLCMs, which are discussed below). These diagnostics suggest that the three- and four-class specifications, according to the AIC, CAIC and BIC, provides the best balance between model fit and parsimony. In choosing between the three and four class model, we opt for parsimony and focus on the three-class specification, buoyed by support from the CAIC and BIC criteria (which both give more weight to model parsimony).

[Table 2 here.]

Coefficient estimates from the three-class standard LCM are shown in the first three columns of Table 3. Roughly speaking, coefficient magnitudes greater than three or four in our

experiment raise concerns that the LCM is converging to a boundary solution, suggesting discontinuous preferences.⁵ Class 1 in the LCM, estimated to comprise 35% of the sample, appears to exhibit continuous preferences that are well-captured by a conditional logit representation. Class 2 (estimated at 48% of the sample) appears to exhibit preferences suggesting serial participation: all alternative-specific constants (ASCs) are greater than four. Moreover, the ASCs for the low WNV risk scenarios are essentially unbounded, suggesting a flat likelihood function, difficulty in the convergence of the optimization routine and a lack of identification.⁶ Similarly, Class 3 (18% of the sample) suggests serial nonparticipation, except in the high-risk scenarios, where there appears to be a slight preference for programs which reduce abundance of both nuisance and WNV-transmitting mosquitoes. Note how closely the estimated class shares for this model correspond to the raw frequencies of these behaviors observed in the data (Figure 1).

[Table 3 here.]

The LC logit results therefore suggest three behavioral patterns, two of which are discontinuous and not representable by a conventional conditional logit model. The first discontinuous class is serial participants (SPs). The second is comprised of individuals who are serial nonparticipants (SNPs), except evidently in situations of high WNV risk (where the coefficient magnitudes are within normal ranges).

⁵ For example, an alternative-specific constant (ASC) equal to four implies behavior in which that alternative is selected relative to the SQ option 98% of the time. Statistically, such an individual would need to be observed 50 times before one selection of the SQ is observed, whereas respondents in the analyzed DCE only completed nine choice tasks. An ASC equal to three would imply the alternative is selected over the SQ 95% of the time so that a task would need to be repeated 20 times before observing one SQ selection.

⁶ All estimates shown in this paper are generated from the authors' own Matlab code, provided as supplementary material to this paper. The LCM code was compared to estimation output from standard software packages, such as the 'llogit' package in Stata (Pacífico & Hong il Yoo 2013). The Matlab output was identical with the Stata output in all cases, except when the models had trouble converging (though when either the Matlab or the Stata implementation did not converge, neither did the other).

Table 4 shows how these discontinuities may be represented using a GLCM. What is required is first to specify the exclusion sets $\emptyset_{m,t}$ for each class m and then to specify the reference alternatives as needed (i.e. when the SQ alternative is excluded). For example the excluded set pertaining to the SP class is $\emptyset_{SP,t} = \{SQ\} \forall t$, i.e. the SQ option is by assumption selected with probability zero by members of this class. Furthermore, since the subsequent conditional logit model for this model is estimated without the SQ alternative, a new reference alternative must be specified by the researcher. We specify the ‘Nuisance-only’ program as the reference alternative whenever the SQ is excluded, in which case its ASC is restricted to zero (e.g. $\alpha_{Nuis,t}^{SP} = 0 \forall t$ for the SP class).

[Table 4 here.]

Results from this estimation are presented in Table 3, alongside the GLC estimates. Based on the regression diagnostics in Table 2, we present estimates for the best-performing, three-class GLC specification, with one class corresponding to ‘normal’ CL preferences over all alternatives, the second class corresponding to SPs and the third corresponding SNPs except in the high-risk scenario. For robustness we also estimated an alternative GLCM with SNPs over *all* scenarios. The summary statistics for this regression are presented in the last column of Table 2, which shows (based on all three reported information criteria) that this specification is overly restrictive.

A number of insights emerge from Table 3, when comparing the LCM and GLCM. In qualitative terms, Class 1 in the LCM corresponds most closely to the unrestricted class in the GLC specification (the ‘CL group’). For example, in the tasks with low WNV risk, both of these classes value the ‘nuisance only’ program more than the ‘WNV only’ program, which is in turn

valued more than the ‘Both’ program. This preference ordering essentially reverses in the high-risk choice tasks.

Class 2 in the LCM is comparable to the SP class in the GLCM. Note that the SP class specification in the GLCM uses the ‘nuisance only’ program as the reference alternative, so that the coefficient estimates in this column are only interpretable relative to this program. Seen from this perspective, the results are in fact similar to the Class 2 estimates: for example, if we look at the high-risk scenarios and look at the difference between the ‘WNV only’ and ‘Nuisance only’ programs ($5.46 - 5.05 = 0.41$), we see that the coefficient estimate lines up nicely with the GLC-SP ‘WNV only’ estimate of 0.426. Similar patterns emerge in comparing the other coefficient estimates, though clearly the lack of identification in the LCM with respect to Class 2 warrants caution in any interpretation of these coefficients (which is precisely the motivation for the GLCM). Also that in both the LCM and GLCM the SP class is the only one for which the program cost attribute is statistically insignificant (though still negatively signed).

The last column in the GLCM corresponds to the class whose preferences are most akin to a serial nonparticipant, the only difference being that this class does appear to participate in high-risk scenarios (a specification supported by the model fit comparisons in Table 2). As such we can only estimate logit preference parameters for the high-risk choice tasks (hence the blank entries for all the parameters in the low- and medium-risk scenarios). This ‘quasi-SNP’ class is most comparable to Class 3: both classes appear to prefer the ‘Both’ program over either the ‘WV only’ or ‘Nuisance only’ programs, though these differences are not statistically significant. Even in this scenario with high WNV risk, support for any of the programs among members of this class appears weak.

Comparing the class membership probabilities between the GLCM and LCM also confirms the clear mapping of classes between the two models. The class membership shares (the means of the posterior class membership variables in eq. 4) in the LCM classes and their counterparts in the GLCM are within a few percentage points of each other. Notably, the GLCM produces class membership shares which are respectively 2% smaller for the SP class (class 2 in the LCM) and 4% smaller in the quasi-SNP class (class 3 in the LCM). This arises precisely because of the partial observability of class membership in the GLCM: for example, respondents who select the SQ in any of their nine choice tasks are excluded from the SP class with certainty, whereas respondents who *do* appear to behave as SPs still have some small positive probability in the GLCM of belonging to the ‘well-behaved’ CL class (class 1 in the LCM). Examining the full distribution of the posterior class membership variables (**Figure 2**), we again see an extremely tight parallel between the LCM (panel a) and GLCM (panel b). However, we can also see that the GLCM produces posterior membership variables which convey greater certainty: only 16% of respondents in the GLCM have posterior membership probabilities which are between 0.01 and 0.99, compared to 23% of respondents in the LCM.

Valuation

We compute mean CV from the results of the GLCM, and compare to the LCM model and conditional logit. Results are presented in Figure 3. As noted above, computing a valuation estimate for the full sample requires care, because values can only be estimated relative to a reference alternative not excluded by any classes. We therefore estimate CV for the addition of a ‘WNV only’ or ‘Both’ program as an option in the choice set, relative to a baseline situation in which a ‘Nuisance only’ program is in effect. Across all three choice contexts (low, medium and high risk) and models (CL, LCM and GLCM), the mean CV for the ‘Both’ alternative is always

greater than the ‘WNV only’ alternative. Relative to the basic conditional logit model, the GLCM model CV estimates are significantly lower in all programs and contexts. Relative to the LCM specification, the GLCM estimates are qualitatively similar for most risk-program combinations, with around a 10 to 20 percent difference in magnitude (the GLCM estimate being sometimes higher or lower than the LCM estimate for different alternatives). An exception to this pattern is in the low-risk, ‘WNV only’ program: In this scenario the LCM CV estimate is nine times higher than the GLCM estimate. Looking back at the regression results in Table 3, we can see that this discrepancy is most likely related to the very large (and clearly poorly identified) regression estimates for LCM model in the low-risk scenario, the result of the serial participation behavior of this class. Examining the 90% confidence intervals around the mean CV estimates in Figure 3, we also see that mean CV estimates for the GLCM are more precise than the CL model.

[Figure 3 here.]

Predicting class membership with respondent characteristics

We also estimate a GLCM otherwise identical to that in Table 3, but which uses respondent characteristics as regressors in the class membership model. This model uses a multinomial logit specification, as in equation (6). Table 5 presents multinomial logit coefficients from this class membership model (the choice coefficients are omitted because they are nearly the same as the last three columns of Table 3). Only two characteristics are found to predict class membership (and only at a statistical significance of 10%). These are the number of hours the respondent spent outside on a typical day in the summer and mosquito abundance, as measured via entomological sampling in the respondent’s neighborhood. In general, respondents who spend more time outside and who have more mosquitoes around their homes are more likely to

belong to the SP class in the GLCM model, relative to the standard CL class and the quasi-SNP class. The SP class generally has a higher willingness-to-pay for mosquito control across all programs: hence the estimated average marginal effect of a 1% change in hours spent outside is estimated at roughly a \$0.70 increase in the CV associated with the ‘Both’ alternative in the low-risk scenario (for example). Similarly, an increase in mosquito abundance of 1% is associated with an increase of \$0.60 increase in CV for the ‘Both’ alternative in the low-risk scenario (again, relative to a ‘Nuisance only’ program).⁷

Discussion

This paper develops and demonstrates the application of a generalized latent class model (GLCM) for dealing with and identifying behavioral patterns reflecting a variety of discontinuous preferences in discrete choice data.

The core element of the proposed methods lies in permitting the analyst to restrict the modelled choice probabilities for some respondents to zero for a subset of alternatives. By first conducting an exploratory estimation of a standard LCM and conjecturing from the results as to which alternatives are systematically avoided by some individuals, a confirmatory GLCM can then be estimated which overcomes the identification problem. In this paper we compare the GLCM to the LCM using information criteria.⁸ In the application to the West Nile virus choice experiment in Madison, results imply that overall valuations for specific programs can only be accurately obtained from these data relative to a non-SQ program. Moreover, while following the

⁷ The details on this computation and full table of average marginal effects for alternative-specific CV are available on request to the authors.

⁸ Ideally, because the GLCM can be viewed as a restriction of an LCM, a likelihood ratio test would be used to test the (less restricted, but possibly unidentified) LCM against the GLCM. However, the restrictions imposed by the GLCM lie at the boundary of the parameter space (i.e. effectively setting some alternative-specific constants to negative infinity in some classes), which precludes the use of a standard chi-squared distribution in such a test. Deriving the distribution of the likelihood ratio in such circumstances is therefore a topic for future research.

same qualitative pattern across programs, even the relative valuation estimates differ significantly in magnitude between the GCLM and LCM. Given the identification challenges for the standard LCM, we argue that it is preferable to use the GLCM estimates and to take seriously the apparent lack of identification of mean absolute program values (i.e. relative to the ‘no program’ status quo option).

References

- Boxall, P.C. & Adamowicz, W.L., 2002. Understanding Heterogeneous Preferences in Random Utility Models : A Latent Class Approach. *Environmental and Resource Economics*, 23(2), pp.421–446.
- Burton, M. & Rigby, D., 2008. Hurdle and Latent Class Approaches to Serial Non-Participation in Choice Models. *Environmental and Resource Economics*, 42(2), pp.211–226.
- Czajkowski, M., Giergiczny, M. & Greene, W.H., 2010. Learning and Fatigue Effects Revisited : Investigating the Effects of Accounting for Unobservable Preference and Scale Heterogeneity. *Land Economics*, 90(2), pp.324–351.
- Greene, W.H., 2011. *Econometric Analysis* 7th ed., Prentice Hall.
- Greene, W.H. & Hensher, D. a., 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological*, 37(8), pp.681–698.
- Von Haefen, R.H., Massey, D.M. & Adamowicz, W.L., 2005. Serial Nonparticipation in Repeated Discrete Choice Models. *American Journal of Agricultural Economics*, 87(4), pp.1061–1076.
- Hess, S. & Stathopoulos, A., 2013. Linking response quality to survey engagement: A combined random scale and latent variable approach. *Journal of Choice Modelling*, 7, pp.1–12.
- Johnson, F.R. et al., 2007. Experimental design for stated-choice studies. In *Valuing environmental amenities using stated choice studies*. Springer, pp. 159–202.
- Lancsar, E. & Louviere, J., 2006. Deleting “irrational” responses from discrete choice experiments: a case of investigating or imposing preferences? *Health Economics*, 15(8), pp.797–811.
- Louviere, J., 2013. Modeling Single Individuals: The Journey from Psych Lab to the App Store. In *Choice Modelling: The State of the Art and the State of Practice*. Northhampton, MA (USA): Edward Elgar Publishing.

- McConnell, K.E., 1995. Consumer surplus from discrete choice models. *Journal of Environmentl Economics and Management*, 29, pp.263–270.
- McFadden, D., 1973. Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, pp.105–142.
- Meyerhoff, J. & Liebe, U., 2009. Status Quo Effect in Choice Experiments: Empirical Evidence on Attitudes and Choice Task Complexity. *Land Economics*, 85(3), pp.515–528.
- Morey, E., Thacher, J. & Breffle, W., 2006. Using Angler Characteristics and Attitudinal Data to Identify Environmental Preference Classes: A Latent-Class Model. *Environmental and Resource Economics*, 34(1), pp.91–115.
- Pacifico, D. & Hong il Yoo, 2013. LCLOGIT: Stata module to fit latent class conditional logit models via EM algorithm. Available at: <https://ideas.repec.org/c/boc/bocode/s457313.html>.
- Rungie, C.M., Coote, L. V. & Louviere, J.J., 2012. Latent variables in discrete choice experiments. *Journal of Choice Modelling*, 5(3), pp.145–156.
- Scarpa, R., Willis, K.G. & Acutt, M., 2005. Individual-specific welfare measures for public goods: a latent class approach to residential customers of Yorkshire Water. *Econometrics Informing Natural Resource Management*, Edward Elgar Publisher, Cheltenham, UK and Nottingham, MA, USA.
- Shonkwiler, J.S. & Shaw, W.D., 1996. Hurdle Count-Data Models in Recreation Demand Analysis. , 21(2), pp.210–219.
- Stoel, R.D. et al., 2006. On the likelihood ratio test in structural equation modeling when parameters are subject to boundary constraints. *Psychological methods*, 11(4), pp.439–455.
- Strazzera, E., Contu, D. & Ferrini, S., 2013. Check it out! A Monte Carlo analysis of the performance of selection criteria and tests for models fitted to Choice Experiments data. In *Proceedings of the International Choice Modelling Conference*.
- Thiene, M., Meyerhoff, J. & De Salvo, M., 2012. Scale and taste heterogeneity for forest biodiversity: Models of serial nonparticipation and their effects. *Journal of Forest Economics*, 18(4), pp.355–369.
- Train, K.E., 2009. *Discrete Choice Methods with Simulation*, Cambridge University Press.

Tables

Table 1: Mosquito program attribute descriptions and levels

Attributes	Description	Levels
<i>Mosquitoes Targeted</i>	Type(s) of mosquitoes that would be targeted by the mosquito control program	All Nuisance only WNV vectors only
<i>Cost</i>	Increase in annual property taxes used to fund the mosquito control program	\$10 \$50 \$100 \$200
<i>West Nile Virus risk</i> (task-spanning attribute)	Risk of contracting West Nile virus in Madison, WI	Low risk (Madison status quo): 1 in 250,000, or 1 case per year Medium risk: 10 in 250,000, or 10 cases per year High risk: 100 in 250,000, or 100 cases per year

Table 2: Regression summary statistics

	Conditional logit	Standard LCMs			GLCMs		
					<i>with serial participants (SPs)</i>	<i>with SPs & Quasi-SNPs</i>	<i>with SPs & SNPs</i>
Number of Classes	1	2	3	4	3	3	3
Log-likelihood	-2,405	-1,960	-1,847	-1,829	-1,849	-1,851	-1,877
Deg. of freedom	10	21	32	43	29	23	20
Respondents	257	257	257	257	257	257	257
Respondents X Tasks	2,313	2,313	2,313	2,313	2,313	2,313	2,313
AIC	4,790	3,878	3,630	3,572	3,640	3,656	3,714
CAIC	4,875	4,051	3,890	3,920	3,875	3,840	3,872
BIC	4,866	4,037	3,872	3,897	3,859	3,830	3,936

Table 3: Generalized and standard latent class logit choice model estimates, 3 classes. *Robust standard errors, clustered at the individual level, in parentheses. SP = serial participant, SNP = serial nonparticipant. For the GLCM, see companion Table 4 on the restriction of reference alternatives. Class membership probabilities here estimated with a constant-only model. *, ** and *** indicate statistical significance respectively at the 10%, 5% and 1% levels.*

WNV risk level	Program alternative	Conditional logit	Standard latent class model (LCM)			Generalized latent class model (GLCM)		
			Class 1	Class 2	Class 3	CL group (all alternatives)	SPs	Quasi-SNPs
Low risk 1/250K	WNV only	0.014 (0.144)	0.52** (0.229)	236 (1.78 x 10 ³)	-2.66** (1.31)	0.342 (0.216)	-1.05*** (0.177)	--
	Nuisance only	0.956*** (0.144)	1.5*** (0.261)	237 (1.78 x 10 ³)	-14.4*** (0.842)	1.26*** (0.236)	- ref. alt. -	--
	Both	0.732*** (0.187)	0.0557 (0.508)	237 (1.78 x 10 ³)	-14.6*** (2.84)	0.209 (0.411)	-0.159 (0.232)	--
	WNV only	0.491*** (0.139)	1.05*** (0.218)	4.3*** (1.16)	-2.31*** (0.794)	0.842*** (0.195)	0.187 (0.156)	--
	Nuisance only	0.35** (0.146)	0.813*** (0.235)	4.14*** (1.15)	-3.09*** (1.12)	0.662*** (0.207)	- ref. alt. -	--
	Both	1.25*** (0.181)	1.84*** (0.378)	5.24*** (1.08)	-13.0*** (1.08)	1.71*** (0.317)	1.01*** (0.188)	--
Medium risk 10/250K	WNV only	0.928*** (0.152)	1.3*** (0.287)	5.46*** (1)	-0.896 (0.618)	1.21*** (0.263)	0.426*** (0.152)	-1.97 (1.87)
	Nuisance only	0.676*** (0.160)	1.31*** (0.27)	5.05*** (1.01)	-0.982* (0.508)	1.15*** (0.234)	- ref. alt. -	-1.34** (0.522)
	Both	1.59*** (0.181)	2.68*** (0.31)	5.89*** (1.03)	0.998 (0.753)	2.49*** (0.284)	0.829*** (0.24)	0.116 (0.604)
	WNV only	0.928*** (0.152)	1.3*** (0.287)	5.46*** (1)	-0.896 (0.618)	1.21*** (0.263)	0.426*** (0.152)	-1.97 (1.87)
	Nuisance only	0.676*** (0.160)	1.31*** (0.27)	5.05*** (1.01)	-0.982* (0.508)	1.15*** (0.234)	- ref. alt. -	-1.34** (0.522)
	Both	1.59*** (0.181)	2.68*** (0.31)	5.89*** (1.03)	0.998 (0.753)	2.49*** (0.284)	0.829*** (0.24)	0.116 (0.604)
High risk 100/250K	Program cost (\$)	-0.00593*** (0.001)	-0.0177*** (0.0021)	-0.00243 (0.00151)	-0.0218*** (0.00667)	-0.0172*** (0.00202)	-0.00206 (0.00143)	-0.0134** (0.0052)
	Predicted class prob.	--	0.32	0.49	0.19	0.36	0.48	0.15
	Imputed class shares	--	34%	48%	19%	39%	46%	15%

Table 4: Parameterizing the GLCM for behaviors observed in the choice experiment

Class / Behavior	Exclusion sets ($\emptyset_{t,m}$)	Identification restrictions
Standard conditional logit (CL) behavior	$\emptyset_{CL,t} = \emptyset \forall t$	$\alpha_{SQ,t}^{CL} = 0 \forall t$
Serial participants (SPs)	$\emptyset_{SP,t} = \{SQ\} \forall t$	$\alpha_{Nuis,t}^{SP} = 0 \forall t$
Serial nonparticipants (SNPs)	$\emptyset_{SNP,t} = \{h \neq SQ\} \forall t$	(degenerate case)
SNPs, except in high-risk scenarios	$\emptyset_{SNP,low} = \{h \neq SQ\}$	(degenerate case)
	$\emptyset_{SNP,medium} = \{h \neq SQ\}$	(degenerate case)
	$\emptyset_{SNP,high} = \emptyset$	$\alpha_{SQ,high}^{SNP} = 0$

Table 5: Class membership model with covariates. GLCM estimated as in Tables 3 and 4, but additionally including respondent covariates in a multinomial logit class prediction model (choice model coefficients from this specification not presented but very similar to GLCM presented in Table 3, available from authors on request). *, ** and *** indicate statistical significance respectively at the 10%, 5% and 1% levels. Statistical significance of the log-likelihood value obtained from a likelihood ratio test of this model against the GLCM specification in Table 3 (which contains a constant-only model for class membership).

	Latent class membership model	
	<i>All alternatives class</i>	<i>SP class</i>
	<i>Coeff. (Std. Err.)</i>	<i>Coeff. (Std. Err.)</i>
Constant	-3.68 (5.49)	-9.56* (5.50)
Kids	0.206 (0.55)	0.313 (0.50)
Female	0.202 (0.42)	0.487 (0.40)
Married	-0.489 (0.58)	-0.238 (0.60)
Age	0.0231 (0.02)	0.00239 (0.02)
log(income)	0.309 (0.38)	0.322 (0.37)
Post-secondary degree	0.0536 (0.56)	0.422 (0.53)
log(1 + hours outside)	0.253 (1.86)	3.45* (2.02)
log(1 + mosquitoes)	0.261 (0.74)	1.54* (0.79)
log(1 + hours outside) X log(1 + mosquitoes)	-0.219 (0.46)	-0.821* (0.48)
Log likelihood (full model)	-1,835**	
Deg. of freedom (full model)	41	

Figures

Figure 1: Frequency of different choice patterns in the sample data

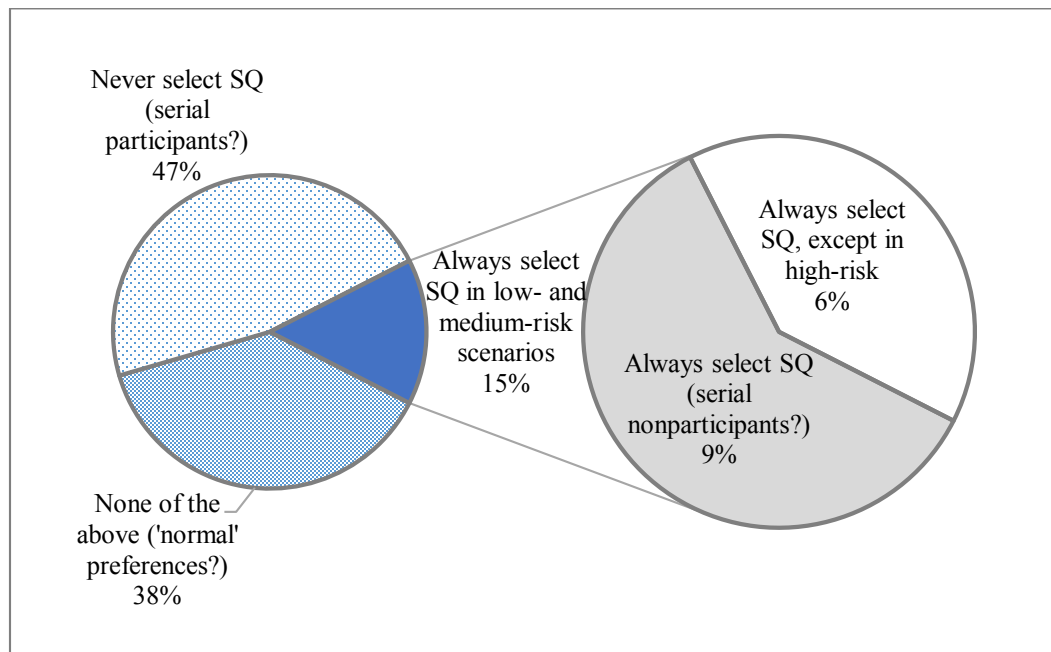
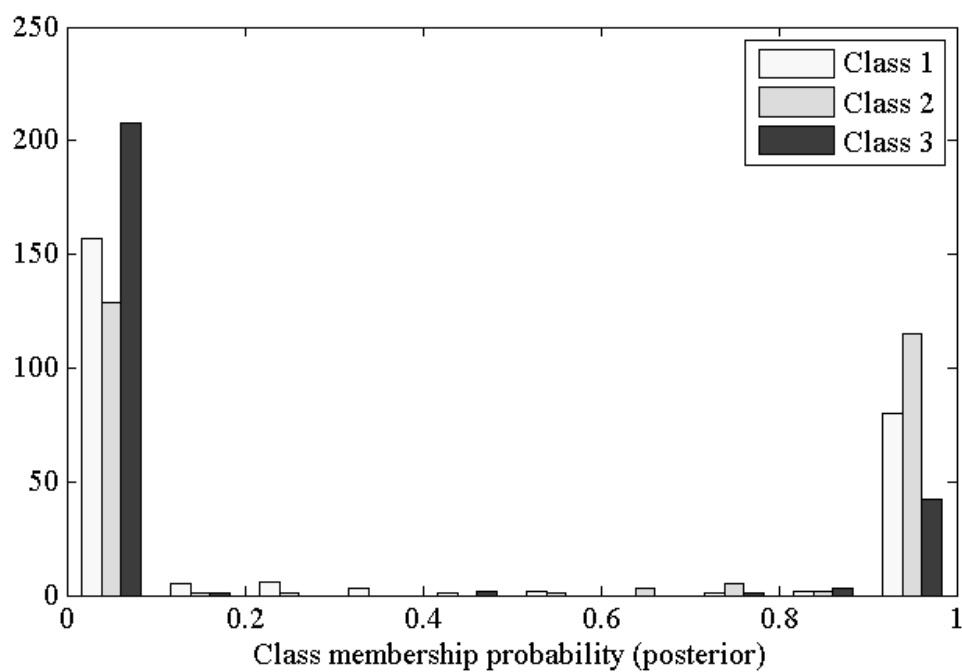


Figure 2: Posterior class membership probabilities

a) LCM (3 class) model



b) GLCM (3 class) model

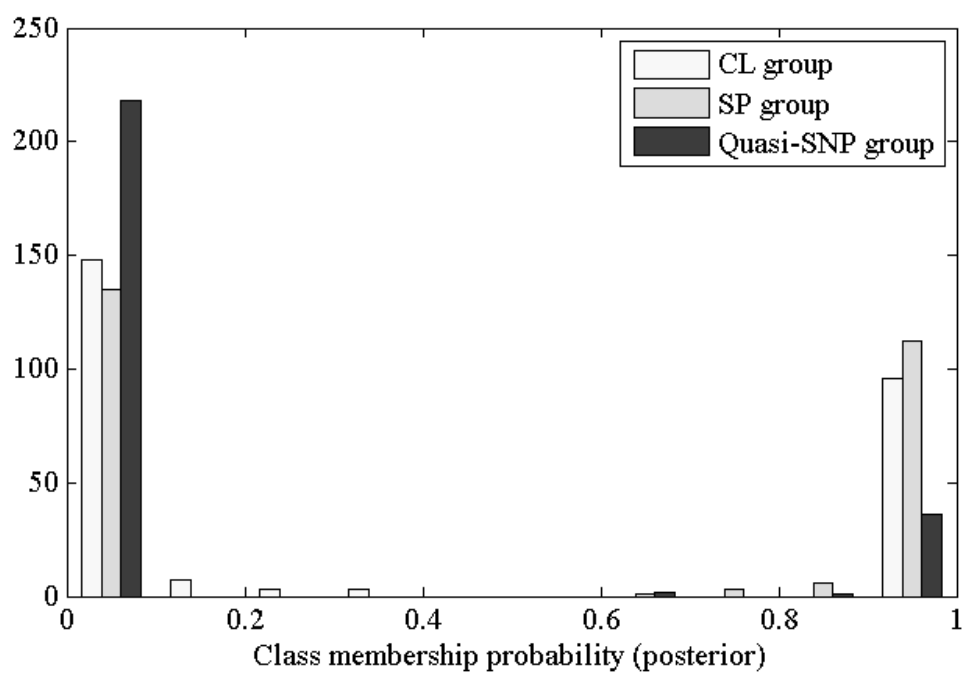


Figure 3: Compensating variation (CV) estimates across models. *Alternative-specific estimates presented relative to a condition in which the nuisance only program is already assumed available. 90% confidence intervals presented in brackets.*

