

Partially observable latent class analysis (POLCA): An application to serial participation in mosquito control in Madison, Wisconsin

Zachary Brown, Katherine L. Dickinson, Susan Paskewitz

Center for Environmental and Resource Economic Policy
Working Paper Series: No. 16-015
November 2016

Suggested citation: Brown, Z., Dickinson, K.L. and S. Paskewitz (2016). Partially observable latent class analysis (POLCA): An application to serial participation in mosquito control in Madison, Wisconsin. (CEEnREP Working Paper No. 16-015). Raleigh, NC: Center for Environmental and Resource Economic Policy.



Partially observable latent class analysis (POLCA): an application to serial participation in mosquito control in Madison, Wisconsin.

November 2016

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

Abstract

Serial nonparticipation in nonmarket valuation using choice data is a frequently observed pattern of behavior in which an individual always appears to choose the status quo or ‘no program’ alternative. In choice models serial nonparticipation may be viewed as belonging to a class of deterministic choice patterns, other examples of which include serial participation and lexicographic preferences. While common in the context of environmental goods unfamiliar to respondents, logit-based choice models are ill-equipped for identifying such preferences, because predicted choice probabilities cannot take a value of zero or one. We extend latent class analysis (LCA) of preference heterogeneity to address this issue, for each class specifying a subset of alternatives that are avoided with certainty. We are then able to partially observe class membership, knowing with certainty that an individual does not belong to a class if she selects any alternatives excluded by that class. We apply our model to a discrete choice experiment on mosquito control programs to reduce West Nile virus risk and nuisance disamenities in Madison, Wisconsin. We find that partially observable latent class analysis (POLCA) obtains the same goodness of fit as LCA with fewer parameters. Adjusting for the need to re-specify the reference alternative when the status quo is excluded, our relative valuation measures are significantly different than those obtained from LCA. We argue that our model is useful for detecting and addressing alternative-specific nonidentification in a given dataset, thus reducing the risk of invalid inference from discrete choice data.

Keywords: random utility models, preference heterogeneity, latent class analysis, partial observability, serial nonparticipation, serial participation, deterministic choice patterns, 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 *serial nonparticipation* in both choice experiment and revealed preference data. These effects arise when a substantial subset of decision-makers choose the *status quo* (SQ) 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, for predicting individuals' responses to new policies, and for understanding which valuation measures can be identified using a given dataset. In this paper we provide a novel econometric method we call Partially Observable Latent Class Analysis (POLCA) for addressing serial nonparticipation and the more general phenomenon of deterministic choice patterns among subgroups of decision makers in discrete choice data.

Deterministic choice patterns are cases in which assumptions underlying the standard random utility model (RUM) for discrete choice data are violated. Serial nonparticipation is an example of such a pattern, because serial nonparticipants by definition select the SQ with certainty. Conversely, serial participation – avoiding the SQ with certainty – is also deterministic choice behavior, even if such behavior is otherwise consistent with a RUM (exhibiting random choices among the subset of non-SQ alternatives). Other deterministic choice patterns include cases in which a subset of alternatives are selected with probability equal to zero. For example, in discrete choice recreation demand models, subsets of recreation sites may be inaccessible or unknown to subgroups of the population (Haab & Hicks 1997). Other examples of nonrandom

choice can arise from discontinuous preferences. For example, lexicographic choice patterns – while consistent with Rational Choice Theory – exhibit choice patterns clearly inconsistent with a RUM, first ordering alternatives strictly by the levels of their most- to least-important attributes. The alternative with the highest level of the most-important attribute is selected with probability equal to one.

These examples illustrate a range of explanations for why deterministic choice patterns may arise. One explanation is that the decision maker is rational (in the choice theory sense), but her preferences are truly deterministic. This will always imply degenerate choice probabilities, regardless of the exact preferences. Lexicographic preferences provide an empirically relevant example of this possibility (e.g. capturing dietary preferences such as vegetarianism or choices based religious beliefs).

Another explanation for deterministic choice, most applicable to revealed preference data, is unobservable variation in the set of individuals' feasible choice sets. This may be due to inaccessibility of excluded alternatives or because some individuals do not know about them. This explanation corresponds to the recreation demand example given above. Other methods have been developed for addressing this specific issue have been developed, such as the Independent Availability Logit (IAL) model (Haab & Hicks 1997; Li et al. 2015). We review this model in more detail below, as it relates to POLCA.

Other reasons for deterministic choice may occupy a gray area between deterministic preferences and unobservable choice set restrictions. For example, serial (non)participation behavior may reflect persistent bias for or against the *status quo*. Status quo biases have been extensively researched in economics and psychology and may arise for a number of reasons (Kahneman 2003). One psychological explanation for a bias towards the SQ is that that actively

considering non-SQ alternatives in a decision may require too much cognitive effort. Such cognitive demands are likely higher in complex choice tasks with hypothetical, unfamiliar alternatives (Meyerhoff & Liebe 2009): a frequent feature of stated preference studies of environmental goods and services. 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).

Serial participation is somewhat more difficult to explain, as such behavior implies bias *away* from the SQ. However, a natural conjecture explaining this phenomena is that preferences for non-SQ alternatives strongly dominate the SQ. An extreme example would be a willingness-to-accept (WTA) study of equivalent variation, in which one of the non-SQ alternatives consisted of money payment to decision makers and the other non-SQ alternatives consisted of non-monetary improvements in environmental quality. Few, if any, individuals would be expected to select the SQ in such a choice task. A less extreme example could include a subgroup of individuals with a willingness to pay (WTP) for environmental quality which is orders of magnitude above the tradeoffs asked of them in an observed choice task. Even if a RUM in theory applies to such individuals, practically speaking their choice patterns will appear deterministic with respect to the SQ: they will avoid it with certainty.

Our aim in this paper is not to test these competing explanations for deterministic choice patterns, but instead to extend the conventional RUM econometric framework to allow for subgroups of decision makers who exhibit deterministic patterns in their choices. All of the behavioral patterns described above are ill-suited for discrete choice econometric models based only on logit (or probit) choice probabilities. The essence of the problem is that data in which one alternative is selected or avoided with certainty imply best-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 where maximum likelihood estimates of logit model coefficients blow up to infinity. As we show below, such cases provide a ‘smoking gun’ that some individuals follow deterministic decision rules.

The POLCA we describe and perform 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). POLCA consists of hypothesizing the existence of precise subgroups of individuals exhibiting deterministic choice patterns. 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 groups, 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 uses the concepts of *exclusion sets* within observed choice sets to address a larger range of deterministic preference patterns than addressed by prior approaches. This includes serial participation, as we show below.

The aim of this research is also supported by recent laboratory-based experimental tests of RUMs (Regenwetter et al. 2011; McCausland & Marley 2014; McCausland et al. 2016). Such studies, through precisely controlled choice sets and in-depth observation of individual subjects over many choice tasks, generally identify individual behavior as consistent with or violating the assumptions of a RUM. Our POLCA model may be viewed as a complement to this research effort, being most suitable for large-scale but less in-depth and focused choice data (either experimental or observational). Instead of exactly estimating individuals' choice structures through a long series of choice tasks, POLCA classifies individuals into a relatively small set of predefined choice models (RUM and non-RUM), and is suitable for larger cross-sections of shorter individual choice series.

After presenting the formulation, we perform a POLCA 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 and the nuisance costs associated with mosquitoes. We find that our POLCA 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 POLCA estimates imply that there are three partially observable 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 levels of West Nile risk but who behave according to a logit model in a high-risk setting. We also find relative WTP values that are somewhat 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, two observable factors (and

their interaction effect) acquire statistical precision that is lacking in the conventional LCA model: 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 in favor of enhanced mosquito control programs).

Previous approaches for addressing deterministic behavior 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 indirect utilities to be estimated, which are typically decomposed into the marginal utilities of each alternative's attributes, including environmental and health benefits and risks, and (in economics) monetary attributes such as cost.¹ For identification, one of these alternative-specific constants must be normalized (usually 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:

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

¹ 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 C_{htn} is an indicator variable for whether individual n selected alternative h in task t .

Serial (non)participation specifically – and deterministic choice patterns 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 fact typically does not pose problems in a CL model estimated on a sample of individuals who ‘on average’ 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. An individual level logit model (now subscripting by n) implies a choice probability equal to:

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

where α_{htn} is the alternative-specific, logit-based indirect utility 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 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 $v_{\tilde{h}tn}$, 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 to restrict 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 using latent class logit (described below). But all of these models posit the same basic logit structure to choice probabilities. Clearly, individuals ascribed a logit preference structure but whose choices do not follow such a pattern (as described above) could significantly bias the results of any logit model estimated on the full sample of data. This could have important consequences both for predicted choice behavior and for 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. Because our proposed model merges these two approaches, we provide some detail below.

The hurdle approach uses a binary probability model for participation (e.g. probit or logit), and then models choices using a conditional logit framework.² The formulation is as follows: for each individual n , let d_n be an indicator for whether that respondent is a

² As VHMA point out, this model could easily be extended to allow for randomly varying conditional logit taste parameters within the participant subsample.

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

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

Note that the first summation is the log-likelihood function for the CL model, restricted to the participant subsample, 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, due to sheer chance. Thus, the predicted probability for observed 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). This is a key point in the design of our generalized latent class model presented below. In their application (using both stated and revealed preference datasets), 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 latent class analysis (LCA) to probabilistically classify respondents across a range of logit-based choice models. Beyond issues of serial nonparticipation and the like, LCA has 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 (Greene & Hensher 2003). In BR’s application, one class is hypothesized – though not restricted – to represent serial nonparticipants.

LCA posits the existence of a finite number M of latent classes, indexed throughout this paper by m (for *model*), according to which every individual can be classified. 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 ($v_{SQ,t}^m = 0 \forall m, t$). 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 . LCA then use a separate multinomial probability model (usually also logit-based) for these predicted probabilities. An expectation-maximization (E-M) algorithm is often used to treat class membership as unobserved data, which can be imputed 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 LCA 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 LCA: it captures a wide variety of behavioral patterns beyond only serial nonparticipation.

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 $v_{SQ}^{SNP} = \infty$, 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 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 conventional LCA is that it ignores information that can make class membership 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.

In the revealed preference literature, a related approach has been the use of endogenous choice sets, sometimes referred to as endogenous 'consideration' sets to distinguish from the observed choice set posed in an experiment or posited by the researcher in an observational study (Manski 1977; Swait & Ben-Akiva 1987; Haab & Hicks 1997; Li et al. 2015). This approach is most commonly applied in the independent availability logit (IAL) model. The IAL model specifies an independent probability of inclusion for every alternative (not just ones for which serial behavior is observed) that is independent of the choice set.

Partially observable latent class analysis

Our approach jointly addresses the issue of unbounded valuation estimates and partial observability of SP or SNP for subgroups of respondents, as well as a wide variety of other possibly deterministic behaviors. As with LCA, POLCA supposes that there are M classes,

indexed by m , capturing individual preference heterogeneity. POLCA departs from the standard approach because of a single additional assumption: the specification of 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 deterministic preferences, especially lexicographic 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 preferences, even if in theory they satisfy the assumptions of a RUM. The excluded alternative could also be systematically overlooked by some individuals due to a lack of survey engagement in stated preference context (Hess & Stathopoulos 2013) or because they are not even aware of the alternative in a revealed preference context (Haab & Hicks 1997).

The rest of this section simply carries out the mathematical formulation implied by this logic. With respect to nonexcluded alternatives, POLCA assumes individuals exhibit standard logit choice probabilities, as in LCA. In order to retain a focus on valuation, we also explicitly separate the marginal utility of money η_m from the alternative-specific constant α_{ht}^m (embodying

the alternatives' other attributes): $v_{ht}^m = \alpha_{ht}^m - \eta_m q_{ht}$, where q_{ht} is the cost of alternative h in task t . The class-specific predicted choice probabilities for POLCA 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 LCA 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.

The partial observability of individual class membership in the model consists of the following. 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 nine choice tasks, for example, can only be probabilistically classified as a serial nonparticipant, because on the 10th

choice task she may have chosen a non-SQ alternative. Consequently, for these individuals, we use the same Bayesian imputation formula as in conventional LCA.

This logic is captured in the following specification for class membership. For each m , define the observed class membership variable d_n^m as follows:

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

Then we only use Bayesian imputation for the missing class membership information. The imputed value 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:

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

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

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

As with LCA, 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). In practice, X_n should contain characteristics that predict a respondent's behavioral profile: for example, serial participants in a program to reduce an environmental health risk may exhibit other behaviors that either increase their baseline exposure to that risk or indicate a much higher disutility from the risk. It is *not* advisable to include attitudinal variables directly in X_n , even though it is best practice in choice experiments to include a set of attitudinal debriefing questions: while these questions are useful for descriptive analysis of the data, they can pose endogeneity problems when used to explain, for example, stated choices (Adamowicz & Deshazo 2006).³

As with LCA, estimation can proceed using the iterative E-M algorithm or direct MLE, employing the gradient and Hessian of the full likelihood function (eq. 6) 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 LCA conditional logit 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 POLCA bridges the advantages of hurdle models and latent class models in addressing not only serial nonparticipation, but also a much broader range of 'weird'

³ One approach to dealing with the endogeneity between stated choices and attitudes is to jointly model these variables as explained by a common set of latent preference classes (Morey et al. 2006). While this is a promising direction to extend the model described in this paper, it is beyond the present scope of our analysis.

preferences. POLCA reduces to a standard LCA by setting $\phi_{m,t} = \emptyset$ 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), POLCA 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 $\phi_{SNP,t} = \{h \neq SQ\}$ 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 $l_{n|SNP} = \prod_t \left(p_{SQ,t,SNP}^{C_{n,SQ,t}} \times \prod_{h \neq SQ} p_{h,t,SNP}^{C_{nht}} \right) = \prod_t (1^1 \times \prod_{h \neq SQ} 0^0) = 1$. Consequently, the POLCA log-likelihood function restricted to the hurdle model reduces to:

$$LL^{POLCA \rightarrow Hurdle} = \tag{7}$$

$$(1 - \tilde{d}_n^{SNP}) \sum_n \log l_{n|CL} + \sum_n [\tilde{d}_n^{SNP} \log p_n^{SNP} + (1 - \tilde{d}_n^{SNP}) \log(1 - p_n^{SNP})]$$

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 a POC model 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.

It is more difficult to econometrically compare POLCA with IAL models, because the latter has not (to our knowledge) been applied to repeated choice data, e.g. in a stated preference setting, and has not yet been combined with a latent class approach to preference heterogeneity. Consequently, while the IAL model is able to allow for choice probabilities equal to zero *and* estimate all alternative-specific constants in a conditional logit model, it is unsuitable in its conventional form for directly addressing SP (for example), where inference about the *status quo* alternative's utility among serial participants necessarily relies on the behavior of other groups of individuals. The IAL model is also clearly unsuitable for lexicographic preferences, where the consideration probability of a given alternative obviously depends on the other potential items in the choice set. For example, a 'health-first' lexicographic preference ordering would exclude all alternatives from consideration that did not provide the highest level health. But as soon as the healthiest alternatives in the original choice set were removed, such an ordering would actively consider (e.g. via the standard conditional logit model) the alternatives in the choice set sharing the next-highest level for the health attribute. This behavior is inconsistent with the IAL model, but can easily be addressed in our POLCA framework.

Valuation estimates

As with LCA, WTP estimates in POLCA-based conditional logit model 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 POLCA approach, but only used here because of the above setup, which

was chosen for ease of exposition and because of the nature of the data in our application (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). 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. POLCA acknowledges that the data simply may not permit estimating valuation estimates which are of most interest to the researchers, due to violation of the RUM assumptions among discrete subgroups. 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, we can only obtain the value of alternative h 's availability relative to other, non-SQ alternatives. This is not a bug but a feature of POLCA (we argue): it acknowledges and helps uncover the limits of the data.

Therefore the appropriate conditional choice probability to be used in computing CV must take 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 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 POLCA and LCA models. There is a growing literature on the unresolved question of how to choose the ‘right’ number of classes in LCA, as well as how to choose between an LCA specification or another parametric specification for preference heterogeneity (Strazzera et al. 2013). The approach continuing to prevail is to select the number of classes via 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 POLCA, it is important to note that POLCA conditional logit model restricts a set of alternative-specific constants to $-\infty$ within an LCA structure. Such an observation would motivate the use of a likelihood ratio (LR) test of the restricted POLCA specification against the more general (though possibly unidentified) LCA. Such a test applied to our case would take POLCA as the null hypothesis and test whether this could be rejected in favor of its unrestricted LCA counterpart. Unfortunately, the POLCA model 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 POLCA. 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 POLCA against LCA.

Empirical application: mosquito control in Madison, Wisconsin

We demonstrate application of POLCA 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 in Madison 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 a web-based survey of homeowners in six different neighborhoods in Madison in the summer of 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 275 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.

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

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. At the beginning of the choice experiment exercise, respondents were told that one way to reduce mosquitoes in Madison would be through an expanded citywide control program, and that the purpose of this section of the survey was to gather information about Madison homeowners' preferences for such a program. Background information was provided on West Nile virus and the difference between nuisance and West Nile virus-transmitting mosquitoes. Respondents were also told that the expanded control program would use a combination of environmentally safe control methods targeting breeding sites and mosquito larvae. Regarding the program's impacts, respondents were told that the program would reduce the number of mosquito bites a Madison resident receives by about 90% and were told that the cost of the program would be funded through an increase in annual 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 (pointed out to respondents) of 1 in 250,000 for a first set of choice tasks, then increased to a hypothetical background risk of 10 in 250,000 and then 100 in 250,000, asking respondents to choose between pairs of hypothetical control programs in each task. 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).

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

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

Table 2 presents the regression diagnostics for the standard LCA (along with those for POLCA models, which are discussed below). These diagnostics suggest that the three- and four-class specifications, according to the AIC, CAIC and BIC, provide 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 LCA 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 LCA is converging to a boundary solution, suggesting deterministic

choice patterns.⁵ Class 1 in the LCA, 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 LCA results therefore suggest three behavioral patterns, two of which are deterministic and not representable by a conditional logit model. The first class contains serial participants (SPs). The second contains serial nonparticipants (SNPs), except evidently in situations of high WNV risk (where the coefficient magnitudes are within normal ranges).

Table 4 shows how these patterns lend themselves to a POLCA. First, we specify the exclusion sets $\emptyset_{m,t}$ for each class m and then 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

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 'lcmlogit' 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).

$\emptyset_{SP,t} = \{SQ\} \forall t$, 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 POLCA model estimates. Based on the regression diagnostics in Table 2, we present estimates for the best-performing, three-class POLCA 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 perform a POLCA, assuming 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, comparing POLCA with LCA. In qualitative terms, Class 1 in the LCA corresponds most closely to the unrestricted class in the POLCA 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 LCA is comparable to the SP class in the POLCA. Note that the SP class specification in the POLCA model 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 LCA with respect to Class 2 warrants caution in any interpretation of these coefficients (which is precisely the motivation for the POLCA model). Also that in both POLCA and LCA, 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 POLCA model 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 ‘WNV 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 LCA and POLCA 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 LCA classes and their counterparts in the POLCA model are within a few percentage points of each other. Notably, the POLCA model 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 LCA). This arises precisely because of the partial observability of class membership in POLCA. 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 POLCA of belonging to the ‘well-behaved’ CL class (class 1 in the LCA). Examining the full distribution of the posterior class membership variables (**Figure 2**), we again see an extremely tight parallel between LCA (panel a) and POLCA (panel b). However, we can also see that POLCA produces posterior membership variables which convey greater certainty. Only 16% of respondents in the POLCA have posterior membership probabilities which are between 0.01 and 0.99, compared to 23% of respondents in the LCA.

Valuation

We compute mean CV from the results of POLCA, LCA and a standard conditional logit model. 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, LCA and POLCA), the mean CV for the ‘Both’ alternative is always greater than the ‘WNV only’ alternative. Relative to the basic conditional logit model, the POLCA’s CV estimates are significantly lower in all programs and contexts. Relative to the LCA specification, the POLCA estimates are qualitatively similar for most risk-program combinations, with around a 10-20% difference in magnitude (the POLCA estimate being sometimes higher or lower than the LCA estimate for different alternatives). An exception to

this pattern is in the low-risk, ‘WNV only’ program. In this scenario the LCA CV estimate is nine times higher than the POLCA 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 LCA 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 POCM are more precise than the conditional logit model.

[Figure 3 here.]

Predicting class membership with respondent characteristics

We also perform a POLCA 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 for predicting class membership, 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, as well as the interaction of hours outside with mosquito abundance. In general, respondents who spend more time outside and who have more mosquitoes around their homes are more likely to belong to be serial participants who generally have a higher relative WTP for mosquito control across all programs. 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, relative to the ‘Nuisance only’ program 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 There is also some evidence for a negative interaction effect between outdoor time and mosquito abundance, with the effect of mosquito abundance on SP class membership and relative WTP decreasing with greater outdoor time. This interaction effect may seem paradoxical, but could reflect heterogeneity in individuals’ tolerance for different levels of mosquito abundance.

Discussion

This paper develops and demonstrates the POLCA for dealing with and identifying behavioral patterns reflecting a variety of deterministic decision rules 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 LCA and conjecturing from the results as to which alternatives are systematically avoided by some individuals, our proposed model can then be estimated to overcome the identification problem. In the application to the mosquito control choice experiment in Madison, our results first imply that overall, population mean valuations for specific programs can only be accurately obtained from these data relative to a non-SQ program, since the valuation estimates for the serial participants are effectively unidentified. Moreover, while following the same qualitative pattern across programs, even the relative valuation estimates differ significantly in magnitude between POLCA and LCA. Given the identification challenges for LCA, we argue that it is preferable to use the POLCA estimates and

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

to take seriously the apparent lack of identification of mean absolute program values (i.e. relative to the ‘no program’ status quo option).

References

- Adamowicz, W. & Deshazo, J.R., 2006. Frontiers in Stated Preferences Methods: An Introduction. *Environmental and Resource Economics*, 34(1), pp.1–6.
- 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., Giergiczy, 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*, 37(8), pp.681–698.
- Haab, T.C. & Hicks, R.L., 1997. Accounting for Choice Set Endogeneity in Random Utility Models of Recreation Demand. *Journal of Environmental Economics and Management*, 34(2), pp.127–147.
- 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.
- Kahneman, D., 2003. Maps of bounded rationality: psychology for behavioral economists. *American Economic Review*, 93(5), pp.1449–1475.
- 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.
- Li, L., Adamowicz, W. & Swait, J., 2015. The Effect of Choice Set Misspecification on Welfare Measures in Random Utility Models. *Resource and Energy Economics*.
- 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.
- Manski, C.F., 1977. The structure of random utility models. *Theory and Decision*, 8(3), pp.229–254.

- McCausland, W.J. et al., 2016. *Testing the Random Utility Hypothesis Directly*, Montreal, Canada.
- McCausland, W.J. & Marley, A.A.J., 2014. Bayesian inference and model comparison for random choice structures. *Journal of Mathematical Psychology*, 62–63, pp.33–46.
- McConnell, K.E., 1995. Consumer surplus from discrete choice models. *Journal of Environmental 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>.
- Regenwetter, M., Busemeyer, J.R. & Mellers, B.A., 2011. Transitivity of Preferences. *Psychological Review*, 118, pp.42–56.
- 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. *Journal of Agricultural and Resource Economics*, 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*.
- Swait, J. & Ben-Akiva, M., 1987. Incorporating random constraints in discrete models of choice set generation. *Transportation Research Part B: Methodological*, 21(2), pp.91–102.
- 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	LCA			POLCA		
					<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: POLCA, LCA and conditional logit choice model estimates, 3 classes. Robust standard errors, clustered at the individual level, in parentheses. SP = serial participant, SNP = serial nonparticipant. For the POLCA, 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	LCA			POLCA		
			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)	--
Medium risk 10/250K	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)	--
High risk 100/250K	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)
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: POLCA parameterization 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: POLCA class membership model with covariates. *POLCA 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 POCM 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 POCM specification in Table 3 (which contains a constant-only model for class membership).*

	POLCA classes	
	<i>All alternatives class</i> Coeff. (Std. Err.)	<i>SP class</i> Coeff. (Std. Err.)
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

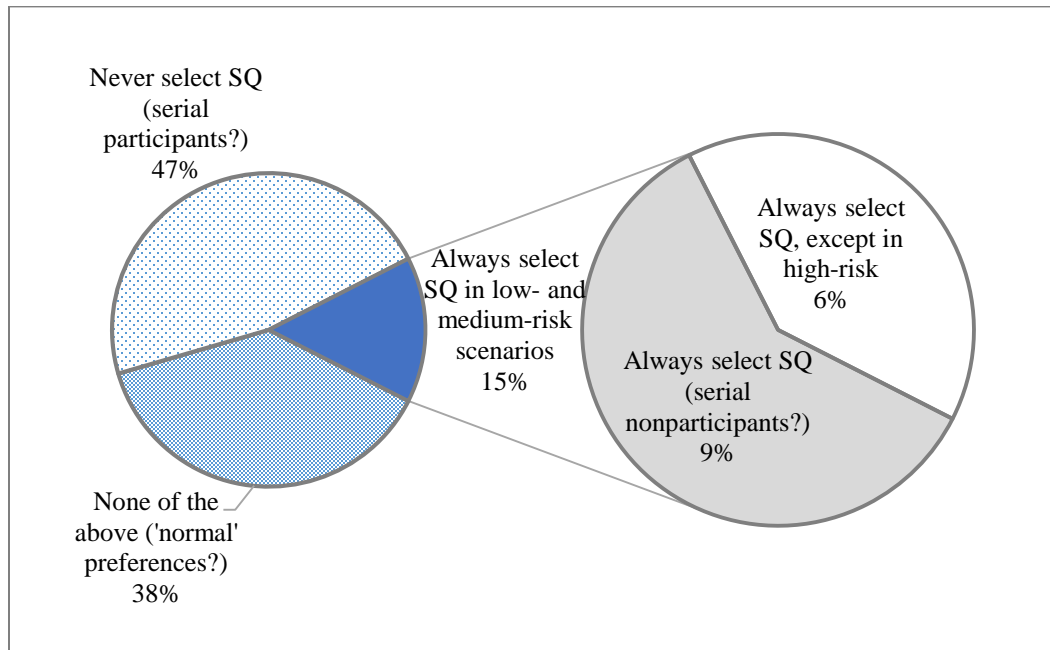
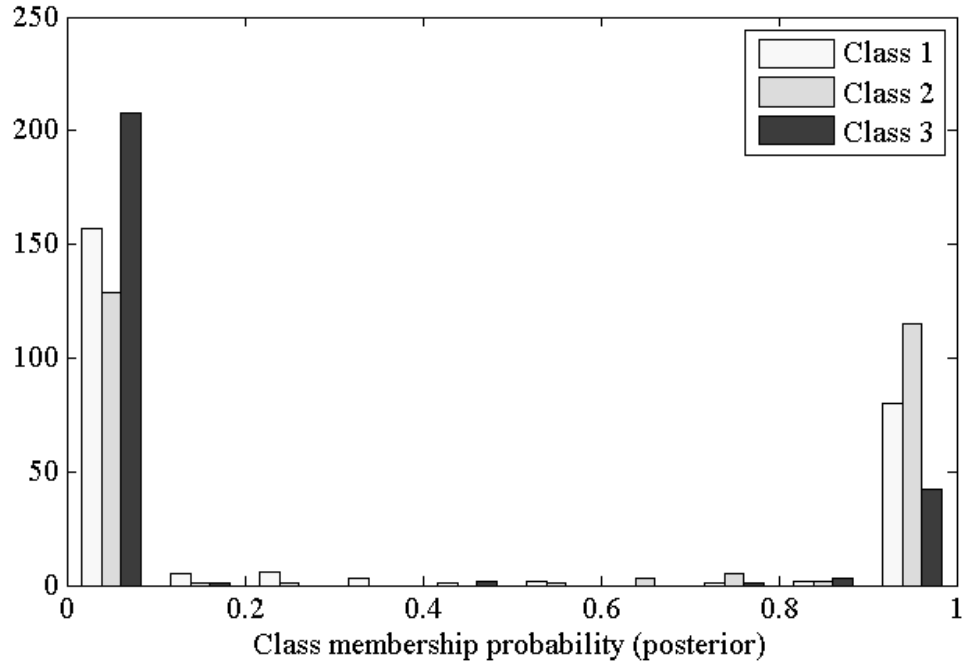


Figure 2: Posterior class membership probabilities

a) LCA (3 classes)



b) POLCA (3 classes)

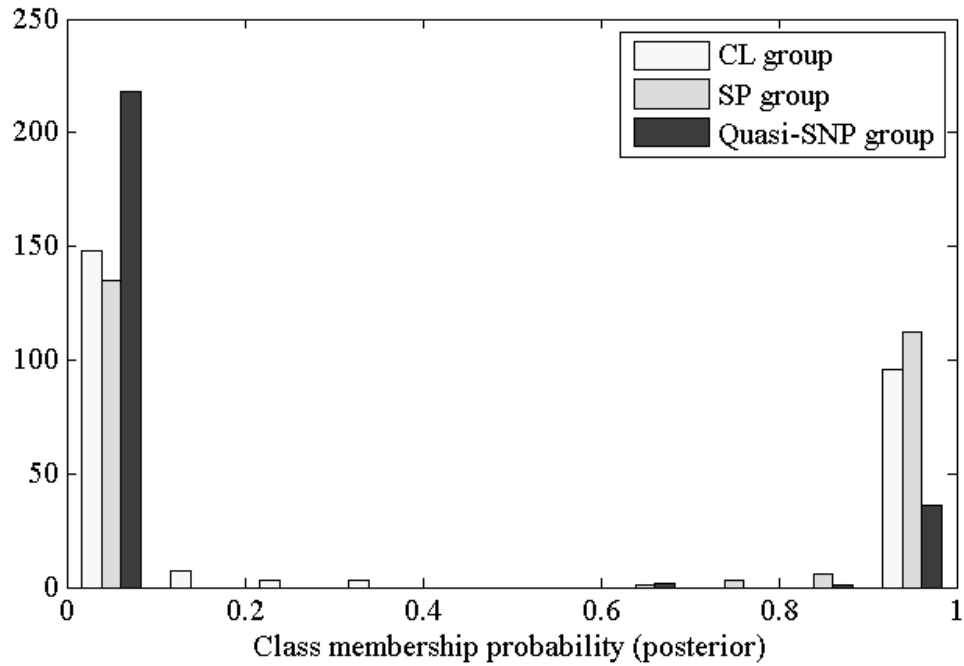


Figure 3: Compensating variation (CV) estimates. *Alternative-specific estimates presented relative to a condition in which the nuisance only program is already assumed available. 90% confidence intervals presented in brackets.*

