Valuing Policy Characteristics and New Products using a Simple Linear Program

H. Spencer Banzhaf*

March 2024

Abstract

The Random Utility Model (RUM) is a workhorse model for valuing new products or changes in public goods. But RUMs have been faulted along two lines. First, for including idiosyncratic errors that imply unreasonably high values for new alternatives and unrealistic substitution patterns. Second, for involving strong restrictions on functional forms for utility. This paper shows how, starting with a RUM framework, one can nonparametrically set-identify the answers to policy questions using only the Generalized Axiom of Revealed Preference (GARP). When GARP is satisfied, the approach set identifies a pure characteristics model. When GARP is violated, the approach recasts the RUM errors as departures from GARP, to be minimized using a minimum-distance criterion. This perspective provides another avenue for nonparametric identification of discrete choice models. The paper illustrates the approach by estimating bounds on the values of ecological improvements in the Southern Appalachian Mountains using survey data.

* Professor, Department of Agricultural and Resource Economics, North Carolina State University. Email: spencer_banzhaf@ncsu.edu.

I thank seminar participants at Georgetown U., Georgia State U., LSE, NC State U., UC Santa Barbara, and the U. of Chicago for comments on an earlier draft.

Valuing Policy Characteristics and New Products using a Simple Linear Program

H. Spencer Banzhaf

1. Introduction

Valuing policy-induced changes in the attributes of public goods is a common goal in public economics and related fields like health and environmental economics. Examples include Medical plans, local public goods, outdoor recreation, and ecosystem services (Keane et al. 2021, Bayer et al. 2016, English et al. 2018, Banzhaf et al. 2016). Similarly, estimating consumers' values for new products is a common goal in industrial organization and marketing. Examples include new automobiles, consumer appliances, and computers (Berry, Levinsohn, and Pakes 2004, Revelt and Train 1998, Bajari and Benkard 2005).

For some 40 years, the standard approach to this problem has been to use a discrete choice random utility model (RUM), such as a multinomial logit. However, these models raise two difficulties. First, they give a large role to additive idiosyncratic errors, which in turn can imply unrealistic substitution patterns. It also implies that expected utility always increases in the number of alternatives, so adding an alternative—even one dominated in the characteristic space—always has value. Consequently, products always have market power, so markups above marginal cost can never be competed away. Accordingly, Bajari and Benkard (2005) and Berry and Pakes (2007) have introduced a pure characteristics model with no idiosyncratic errors at all, only unobserved tastes and product characteristics. However, these models are computationally challenging and can be quite sensitive to outlying observations (Athey and Imbens 2007, Gandhi and Nevo 2021).

A second criticism of discrete choice models is that, in practice, most estimators are highly parametric. Typically, they specify distributions for the additive errors (e.g. logit) and the distribution of tastes for characteristics, as well as a functional form for the utility function. Much of the identification from discrete choice models comes from such assumptions.

This paper shows how to overcome both problems by relating the RUM framework to the Generalized Axiom of Revealed Preference (GARP). Whereas the RUM framework looks for a utility function with patterns of heterogeneity and errors that best explain the observed choices, the GARP framework tests whether there is any utility function that can perfectly explain the choices without errors. If so, it can weakly identify indifference curves and the answers to policy questions—in the sense that it can point identify these objects as the sample goes to infinity, but for finite samples the estimator can only recover sets of parameters and bound answers to policy questions (Lewbel 2019). For discrete choices, this approach assumes only that utility is strictly monotonic and quasi-concave.[1]

To see its connection to the RUM approach, two long-standing features of the GARP approach are especially salient. First, economists can test GARP using a linear program (LP), which also provides a way to construct a candidate utility index non-parametrically (Afriat 1967). Second, if the tests reject GARP, economists can quantify the degree of the rejection using indices known as "critical cost efficiency" (Afriat 1972, Varian 1985, 1990). Such measures typically ask how much expenditure would have to be wasted to make the data consistent with GARP, and typically appear in an additively separable form. These two features of the GARP literature are well established for competitive budget sets. However, their implications for RUMs with discrete alternatives has not previously been noted.

With those features in mind, my argument can be summarized in the following steps. First, consider micro-level data with a panel of discrete choices that allows for non-trivial revealed preference tests. The "panel" structure of the data may comprise repeated choice observations for each individual or a cross section of data but with the individuals considered as part of a panel at a group level (e.g. Berry and Haile 2023). Second, note that if a set of discrete choices is consistent with GARP in characteristics space, then we can weakly identify the pure characteristics demand model and bound answers to policy questions. However, it is likely that GARP will be violated in most data sets big enough to be of practical use, in which case the pure characteristics model is incoherent. Accordingly, the third step of the logic is to generalize the model by introducing the adjustments associated with critical cost efficiency. Crucially, I show that these adjustments are observationally equivalent to the additively separable, idiosyncratic errors of a RUM. That is, optimization mistakes in a model with no idiosyncratic shocks are equivalent to a model without mistakes, but with idiosyncratic tastes, a "moment's fancy," or quality shocks.

[1] The revealed preference approach, begun by Samuelson and extended by Houthakker, Afriat, Varian, and others, is now quite large and has been extended to firm behavior, Nash equilibria, and other contexts. See e.g. Chambers and Echenique (2016) and Hands (2014) for overviews and discussion.

Unfortunately, introducing such errors nonparametrically has the potential to make the model radically incomplete (Kuminoff 2009). For *any* candidate utility functions, there are always some errors that could rationalize the observed data. In other words, the identified sets are unbounded. However, this problem can be overcome by assuming that the additive idiosyncratic errors of the RUM are, like other econometric errors, objects to be minimized. They are only there to explain why we cannot otherwise perfectly satisfy GARP in characteristics space. As with critical cost efficiency, they are the minimum adjustment to utility (potentially measured in dollars) that would rationalize the data. Completing the link, the model can then be estimated with an LP, with the objective to minimize some norm of the errors. Rather than fitting market shares conditional on an error structure, this way of thinking about the problem inverts the logic and finds the pure characteristics model that comes closest to the data, with the residuals being the unexplained gaps.

Even when minimizing the errors, with this minimal structure we may still only be able to bound indifference curves. However, one can also bound functions of indifference curves, and thus set-identify the answer to policy questions, in the spirit of Manski (2007, 2010). Moreover, the bounds on the policy questions can conveniently be obtained by "searching" over a domain defined by the values which are consistent with GARP, directly within the LP. Specifically, I show how to bound willingness to pay (WTP) for policy changes in characteristic space or new products.

I illustrate this approach with an application to the valuation of ecosystem services in the Southern Appalachian Mountain area of the southeastern United States, previously analyzed using traditional logit models (Banzhaf et al. 2016). The ecosystems in that region have been damaged by years of acid precipitation. A survey of the area's residents elicited preferences for various investments to improve different dimensions of the ecosystem (including the health of streams, forests, and bird populations). The survey was set up as a series of choice experiments, over which respondents chose their preferred option from a menu. Analyzing these data using the nonparametric approach posed here, and allowing for heterogeneity across households, I find the average household's WTP for a given set of improvements can be bounded between \$44 and \$210, compared to \$49 for a mixed logit, which imposes stronger assumptions. The average absolute error in the non-parametric model, when monetized, is about \$0.46, compared to \$75 in the mixed logit. Moreover, 98% of observations require no error et all, and 88% of households require no error for any of the alternatives they face. Thus, in this application, the pure characteristics model is much closer to the data that standard RUMs give it credit for.

This paper is related to the existing literature in at least four ways. First, it relates to work that extends GARP to discrete goods. Blow, Browning, and Crawford (2008) showed how GARP can be extended from goods space to Gorman-Lancaster characteristics space. They considered the value of new products, but still assumed choice sets were continuous. Polisson and Quah (2013) and Cosaert and Demuynck (2015) extend GARP from the usual competitive budget sets, which are convex, to discrete choice sets. I rely on their results, but also extend them by introducing additive errors and by showing how their results can help bound economic questions in characteristics space.

Second, this paper is related to recent work on identification of discrete choice models. Berry and Haile (2016, 2023) emphasize that identification of demand in such models is built on nonparametric foundations, and several results have pushed in that direction. Fox et al. (2012) consider semi-parametric identification of the distribution of tastes for characteristics, within the mixed logit model. Shi, Shum, and Song (2018) show how to semi-parametrically identify average tastes for characteristics with minimal assumptions on the distribution of idiosyncratic errors, except for additive separability and a continuous distribution function. However, they impose a linear form for the utility function and restrict the admissible distribution of observed characteristics. Allen and Rehbeck (2019) nonparametrically identify average demands via a representative consumer. Their approach bears some similarity to mine insofar as it imposes the integrability conditions on this consumer, which are comparable to GARP. However, for welfare analysis, it requires restrictions to guarantee that the representative consumer is normative. In contrast, my approach makes no assumption on the distribution or continuity of the distribution of idiosyncratic errors (which, in fact, will have a large probability mass at zero), makes no assumption about the functional form for utility, and recovers information on the distribution of tastes. However, in general it only weakly identifies demands.

One limitation of my approach is that, although it can account for an index of unobserved product characteristics, it assumes they are independent of observed characteristics. This assumption is consistent with McFadden's early work on logit models and with more recent work by Lewbel, Linton, and McFadden (2011) and Allen and Rehbeck (2019). It is applicable to cases where one is working with plausibly exogenous distances as the price (e.g., Beckert, Christensen, and Collier 2012, English et al. 2018) or experimentally controlled data as in a survey, which I use in my empirical application. However, it will be problematic in cases where price or other characteristics are endogenous. Future work might consider imposing additional exclusion restriction, as emphasized by Berry and Haile (2016, 2023) and in the spirit of Blundell, Browning, and Crawford (2003).

Third, this paper relates to other work on non-market valuation in environmental and other settings. Adamowicz and Graham-Tomasi (1991) use revealed preference methods to test rationality in outdoor recreation behavior in the context of the travel cost model. Crooker and Kling (2000) use them to gauge the plausibility of functional forms for parametric methods. Blow and Blundell (2018) use them to gauge the plausibility of benefits transfer exercises. Of especial relevance, Kuminoff (2009) set identifies parameters within a structural model, following Manski (2007). I extend this body of work by introducing RUM errors as critical cost efficiency indices of GARP violations. I also extend it by considering non-parametric bounds on WTP, without functional form assumptions.

Finally, this paper relates to the non-parametric testing of RUMs (McFadden 2005, Kitamura and Stoye 2018) using an axiom of revealed stochastic preference. However, in some ways it takes the opposite approach. Rather than test axiomatically whether an entire sample of data could be generated by a random utility function, it tests whether individual responses could be generated from non-random utility functions, and *generates* a random utility function through least deviations from such functions.

The paper proceeds along the following outline. Section 2 introduces the intuition with a simple linear-in-parameters structural model, set identified through GARP rather than point identified with distributional assumptions on the errors. It begins by assuming the observed choice behavior does satisfy GARP, and hence that there is no need for additive errors. It then derives bounds on the policy questions of interest. Finally, it introduces unobserved product characteristics into the model, as in Bajari and Benkard (2005) and Berry and Pakes (2007).

Section 3 introduces idiosyncratic errors into the linear model, to rationalize the data when GARP is violated, using the logic of critical cost efficiency. Section 4 relaxes the linearity assumption, making the model fully nonparametric, with only the assumption of additive errors.

Section 5 offers an application, estimating WTP for improvements to ecosystem service in the southern Adirondack mountains in the United States. It compares the GARP-based approach to more traditional logit models. Section 6 concludes.

2. Set Identification of the Linear Pure Characteristics RUM

This section introduces and develops the basic intuition using the canonical linear RUM, before eventually relaxing any functional form restrictions in Sections 4.

2.1. Set Identifying a Random Coefficients RUM

Consider the following utility function:

$$(1) \qquad v_{ijt} = \alpha(y_{it} - p_{jt}) + \bar{\boldsymbol{\beta}}'\mathbf{x}_{jt} + u_{ijt},$$

where v_{ijt} is the utility for household i conditional on choosing alternative j in period t, y_{it} is income in period t and p_{jt} is the price of product j in period t (so $y_{it} - p_{jt}$ is numeraire consumption) and $\mathbf{x}_{jt}$ is a vector of K other observed characteristics ($x_1, \ldots, x_k, \ldots, x_K$) for the product/period. The j alternatives could be products, residential locations, or simply policy scenarios. The utility parameters include $\alpha > 0$ and $\bar{\boldsymbol{\beta}} \geq 0$. These sign restrictions imply utility is strictly increasing in the numeraire and non-decreasing in all $\mathbf{x}$. (This is a harmless normalization as we can always model bads as absence of goods.) Note additionally that, though this is a very simple linear model, it easily can be extended to any linear-in-parameters model in the usual way. Given the sign restrictions and the linearity assumption, utility is strictly monotonic and concave.[2] Finally, an outside option ($p_{0t} = 0$, $\mathbf{x}_{0t} = \mathbf{0}$) is normalized to have $v_{i0t} = 0\ \forall\ i,t$.

As written in the general form of Equation (1), the errors u_{ijt} may have at least three interpretations. First, they could be random idiosyncratic errors: $u_{ijt} = \varepsilon_{ijt}$, as for example in the multinomial logit model (McFadden 1981). Second, they could reflect unobserved product attributes: $u_{ijt} = \mu_j$ (Berry, Levinsohn, and Pakes 1995, 2004, Berry and Pakes 2007, Bajari and Benkard 2005). Third, they could reflect heterogeneity in tastes for $\mathbf{x}$, as in the mixed logit model (Berry,

---

[2] With competitive budget sets, monotonicity and concavity are observationally equivalent to local non-satiation, given a finite set of observations (Afriat 1967). However, as Cosaert and Demuynck (2015) discuss, this equivalence no longer holds with discrete budget sets, because an open ball around some choice alternative might not be in the choice set. Accordingly, I impose strict monotonicity (utility is non-decreasing in all goods and increasing in income) and quasi-concavity throughout this paper.

Levinsohn, and Pakes 1995, Revelt and Train 1998, McFadden and Train 2000). In that case, $u_{ijt} = \tilde{\boldsymbol{\beta}}_i'\mathbf{x}_{jt}$, where individual i's taste vector is $\boldsymbol{\beta}_i = (\bar{\boldsymbol{\beta}} + \tilde{\boldsymbol{\beta}}_i)$, with $\bar{\boldsymbol{\beta}}$ interpreted as the population average taste and $\tilde{\boldsymbol{\beta}}_i$ the departure from the average. Or any combination of these. From the perspective of this paper, the first of the three interpretations (idiosyncratic errors) differs from the other two in that there are *always* values for such errors that guarantee we can rationalize the data.

To develop the argument, in this section I initially assume the data are rationalizable without needing either idiosyncratic errors or unobserved product characteristics to explain observed choices, and thus temporarily omit them from the model. I thus assume the errors reflect only heterogeneity in $\boldsymbol{\beta}_i$. This is a reasonable starting point, as much of the previous work on revealed preference in characteristics space also has taken this approach, including Manski (2007), Blow, Browning, and Crawford (2008), Polisson and Quah (2013), and Cosaert and Demuynck (2015).

With these initial assumptions, we can write the model as:

(2) $$v_{ijt} = \alpha_i(y_{it} - p_{jt}) + \boldsymbol{\beta}_i'\mathbf{x}_{jt},$$

with $\alpha_i > 0$ and $\boldsymbol{\beta}_i \geq 0$. Furthermore, we can take an affine transformation of this utility function by dividing by α_i and subtracting y_{it}:

(3) $$v_{ijt} = -p_{jt} + \boldsymbol{\beta}_i'\mathbf{x}_{jt}.$$

With this renormalization the coefficients $\boldsymbol{\beta}_i$ can be interpreted as individual-specific marginal WTP for the attributes $\mathbf{x}$, as they are now relative to the marginal utility of money. Additionally, we've conveniently eliminated the strict inequality constraint on α_i.[3]

Note this model nests several special cases, including the simplest case of total homogeneity ($\boldsymbol{\beta}_i = \boldsymbol{\beta}\ \forall\ i$) or observed heterogeneity ($\boldsymbol{\beta}_i = \boldsymbol{\beta}_z'\mathbf{z}_i$ for some vector of observables $\mathbf{z}_i$). Thus, although revealed preference approaches always require repeated observations, the model allows for several kinds of data sets, each with its respective interpretation of the panel. If only a cross section of data (or repeated cross sections) are observed, one could take one of these two special cases. Or, one could define a group by some set of observables, and think of i as the group, with individuals within the group being treated like repeated observations from the same preference

[3] Technically, the $\boldsymbol{\beta}$ in (3) might better be written by $\breve{\boldsymbol{\beta}}$, but for simplicity I abuse the notation.

ordering. The most general case is individual-level panel data, which allows for unobserved heterogeneity (i.e., $\boldsymbol{\beta}_i$ as random coefficients).[4] In my application to ecosystem services, these data are available. Accordingly, for concreteness, I will use the language of this kind of setup, though the others are also possible.

Each household i (i = 1…I) faces $T(i)$ choice occasions indexed by $t(i)$=1…$T(i)$, chooses from a choice set $\mathcal{J}(i,t)$ at each choice occasion t, and actually chooses alternative $j'(i,t) \in \mathcal{J}(i,t)$ at choice occasion $t(i)$. At each choice occasion t the household chooses alternative $j'(t)$ if and only if $v_{i,j'(t)} \geq v_{i,j}\ \forall\, j \in \mathcal{J}(i,t)$. Substituting Equation (3) into this condition, re-arranging, and considering all choice occasions in the data, we have:

$$\text{(4)} \qquad \boldsymbol{\beta}_i'(\mathbf{x}_{j'(i,t)t} - \mathbf{x}_{jt}) \;\geq\; p_{j'(i,t)t} - p_{jt},\;\; \forall\, i, \forall\, t, \forall\, j \in \mathcal{J}(i,t).$$

That is, household i's utility for the product chosen at any choice occasion must be at least as great as the utility of all other products in that occasion's choice set.

We will say the data are “rationalizable by a linear utility function,” or “linearly rationalizable” for short, if there is a linear utility function that explains the choice patterns observed in the data, as well as patterns of monotonicity (A is preferred to B, regardless of observed choices, if it dominates in every dimension). Following the revealed preference approach, utility maximization is equivalent to satisfying a set of choice axioms, to the effect that if households choose a bundle A on one occasion when B is available (so that A is revealed to be at least as good as B), then there is no sequence of choices such that B is ever indirectly revealed preferred to be strictly better than A. More formally, using the notation A $\succsim_R$ B to mean “A revealed at least as good as B” and A $\succ_R$ B to mean “A revealed strictly preferred to B,” then A $\succsim_R$ B implies there is no sequence B $\succsim_R$ C $\succsim_R$ . . . $\succsim_R$ D $\succsim_R$ A with one of the relations being strict in the sequence. Thus, GARP assures that there is a no-cycling condition on choice patterns. Simplifying the problem, Afriat (1967) showed that the no-cycling condition is equivalent to the existence of a solution to

[4] Panel data are necessary to identify random coefficients using the non-parameteric GARP approach of this paper, whereas, by identifying off functional form and distributional assumptions, the conventional MLE approach to RUMs can identify even random coefficients models in a cross section. Similar issues arise in other contexts for testing GARP and bounding policy questions, motivating Blundell, Browning and Crawford (2003, 2008) to nonparametrically adjust budget sets for income differences, to make different choice occasions comparable. While motivated by a similar data problem, their solution is not directly relevant to my application, as I focus on a single differentiated product, for which budget-balancedness need not hold.

an LP. His approach relies on the equivalence of GARP and, for competitive budget sets, the existence of utilities for bundles and marginal utilities of income that explain choice patterns. Polisson and Quah (2013) and Cosaert and Demuynck (2015) extend Afriat's approach to discrete choices.

The linearity assumption in this section greatly simplifies the problem: The data are linearly rationalizable if and only if $\exists\ \boldsymbol{\beta}_i \geq \mathbf{0}$ satisfying Expression (4). Equivalently, there must be some solution (possibly a continuum of solutions) to an LP to minimize the degenerate objective $\boldsymbol{\beta}_i'\mathbf{0}$, subject to the inequalities given by Expression (4) and the non-negativity constraints. In other words, Expression (4) represents the Afriat inequalities for this model. Without additional distributional assumptions about the random coefficients, this is *all* we can say about the $\boldsymbol{\beta}_i$. That is, the set of $\{\boldsymbol{\beta}_i \mid \boldsymbol{\beta}_i \geq \mathbf{0}\}$ consistent with utility maximization (and the functional form specified in (3)) are all those satisfying Expression (4). However, intuitively, the sets shrink as we observe more data. This intuition introduces an alternative argument for identification of the discrete choice model.

This logic can be summarized more formally by the following theorem:

Theorem 1 (weak identification). If the data are generated from utility-maximizing individuals with linear utility, as in expression (3), then the data are linearly rationalizable for each individual i, and the distribution of $\boldsymbol{\beta}_i$ is set identified by $\{\boldsymbol{\beta}_i \mid \boldsymbol{\beta}_i \geq \mathbf{0}$ and $\boldsymbol{\beta}_i'(\mathbf{x}_{j'(i,t)t} - \mathbf{x}_{jt}) \geq p_{j'(i,t)t} - p_{jt}\}$. Moreover, if $\mathbf{x}$ is distributed over continuous support, then as the set of observed alternatives in the data increases and becomes dense, each set $\{\boldsymbol{\beta}_i\}$ converges to its true value. Thus, the individual random coefficients are point identified in the limit.

Proof: The first part of the theorem is just an application of Afriat's Theorem (Afriat 1967) and follows immediately from Expression (4). To prove the second part, on convergence, let $\{\mathbf{x}_{j'(i,t)t}, \mathcal{J}(i,t)\}_t \rightarrow \{\boldsymbol{\beta}_i\}$ be the correspondence mapping the data to the set of parameters that satisfy GARP (which by the premise of the theorem is non-empty). Next, consider some true $\tilde{\boldsymbol{\beta}}_i$, some $\mathbf{x}_j \in \mathcal{J}(i,t)$, the set $\sim\mathbf{x}_j = \{\mathbf{x} \mid \tilde{\boldsymbol{\beta}}_i'\mathbf{x} = \tilde{\boldsymbol{\beta}}_i'\mathbf{x}_j\}$ and the set $\{\hat{\mathbf{x}}_j\} = \{\mathbf{x} \mid \boldsymbol{\beta}'\mathbf{x} = \boldsymbol{\beta}'\mathbf{x}_j$, for all $\boldsymbol{\beta} \in \{\boldsymbol{\beta}_i\}\}$. The first of these sets is the true indifference set for $\mathbf{x}_j$; the second comprises the candidates for the indifference set given the data. Now, consider a sequence l which adds additional data points to the choice set $\mathcal{J}(i,t)$. Given the continuous support for $\mathbf{x}$, as $l \rightarrow \infty$, there will be $\mathbf{x} \in \{\hat{\mathbf{x}}_j\}_{l-1}, \mathcal{J}(i,t)_l$, which, by completeness, is either in the no-worse-than set or no-better-than set for $\mathbf{x}_j$, thus removing regions from $\{\boldsymbol{\beta}_i\}$. As the minimum distance in $\mathbb{R}^K$ between the hyperplane defined by $\sim\mathbf{x}$ and an element in $\mathcal{J}(i,t)_l$ shrinks to zero, the set $\{\boldsymbol{\beta}_i\} \rightarrow \tilde{\boldsymbol{\beta}}_i$. In the limit, as $\mathcal{J}(i,t)_l$ becomes dense, there are K-1 linearly independent elements from $\sim\mathbf{x}_j$ and $\{\boldsymbol{\beta}_i\} = \tilde{\boldsymbol{\beta}}_i$.

The basic intuition of the proof follows the logic first introduced by Mas-Collel (1978) for competitive budget sets, adapted to the case of linear utility and discrete choices in characteristics space. The proof uses large-J asymptotics, but a similar argument could be constructed using large-T, i.e., repeated choice occasions, so long as each choice occasion involves an independent draw from the space of possible choice sets.

As noted above, if the data are rationalizable, all we can accomplish with finite data and without further structural assumptions is to set identify each $\boldsymbol{\beta}_i$. Figure 1 illustrates this point with an example in K=1 and three alternatives, a status quo option (SQ), $\{p_1, x_1\}$, and $\{p_2, x_2\}$. In this example, we observe $\{p_1, x_1\} \succ_R$ SQ $\succ_R \{p_2, x_2\}$. The shaded area depicts the set $\{\boldsymbol{\beta}_i\}$ consistent with these data. The linear indifference curve going through the SQ option could lie in any part of this shaded area. The absolute values of the slopes of such indifference curves reflect the possible WTP for attribute x.

Nevertheless, it is reassuring that, as the data set becomes large for each i, the identified set converges. This second part of the proposition is illustrated in Figure 2. In this example, x_3 and x_4 now enter $\mathcal{J}$, and we observe $\{p_3, x_3\} \succ_R$ SQ $\succ_R \{p_4, x_4\}$. Now the light shaded areas can be eliminated, and the set $\{\boldsymbol{\beta}_i\}$ shrinks to the darker shaded area, closing in on the true indifference curve shown. However, the rate of convergence is not known.

Last minute addition: insights on convergence from simulations – see appendix

Upshot -- The simulations illustrate weak convergence in principle but also highlight the practical importance of optimal sampling, so that the new observations are informative.

2.2. Bounding WTP

Even when indifference curves are only set identified, we can construct bounds on policy questions defined by scalar-valued functions of $\boldsymbol{\beta}$ (Manski 2010). To do so, we need only change the objective function of the above LP, while satisfying the same Afriat inequalities. In this sub-section, I consider two such policy questions: (i) the WTP to switch to a new policy scenario, such as a change in the vector of public goods; and (ii) the highest price that can be charged for a product competing against a set of substitutes.

Value of a new policy scenario

Suppose we are interested in the WTP for an alternative with characteristics $\mathbf{x}^*$ when the only other alternative is the outside option. For example, suppose $\mathbf{x}$ is a vector of public goods and we wish to know individuals' WTP for a policy moving $\mathbf{x}$ from the status quo to some $\mathbf{x}^*$. Note that in this linear model the status quo can always be renormalized to the outside option by subtracting the status quo values of $\mathbf{x}$ from each alternative, so characteristics now represent changes.

Since each candidate utility function implies a unique WTP, it is easy to see that since utility is weakly identified in this model so is WTP. Moreover, the following theorem shows how we can construct bounds on WTP for such a policy question.

> **Theorem 2**. If the data are linearly rationalizable and the set of $\{\boldsymbol{\beta}_i\}$ rationalizing the data is bounded, then the WTP for some counterfactual scenario characterized by $\mathbf{x}^*$, relative to the status quo, can be bounded by solving the two linear programs:
>
> $\text{Max}_{\{\boldsymbol{\beta}_i\}} \sum_i \boldsymbol{\beta}_i'\mathbf{x}^*$ and respectively $\text{Min}_{\{\boldsymbol{\beta}_i\}} \sum_i \boldsymbol{\beta}_i'\mathbf{x}^*$,
>
> subject to $\boldsymbol{\beta}_i \geq \mathbf{0}$ and inequalties (4).
>
> *Proof.* Note first that the set $\{\boldsymbol{\beta}_i \mid \boldsymbol{\beta}_i'(\mathbf{x}_{j'(i,t)t} - \mathbf{x}_{jt})\} \geq p_{j'(i,t)t} - p_{jt}\} \ \forall\ t, \forall\ j \in \mathcal{J}(i,t)$ is the intersection of half spaces, so it is a convex polytope. Because it is bounded as well, the set of $\{\boldsymbol{\beta}_i\}$ satisfying GARP is compact. Then the objective described in the proposition is to maximize (resp. minimize) a continuous function defined on a compact set. By the extreme value theorem, there is a solution to this program, and it is the arg-max (resp. arg-min) of the function.

Figure 3 illustrates Theorem 2, assuming as in Figure 1 that we observe $\{p_1, x_1\} \succ_R \text{SQ} \succ_R \{p_2, x_2\}$. Steeper candidate indifference curves in the shaded region represent higher marginal WTP for x; flatter curves represent lower marginal WTP. Therefore, we simply "search" over all indifference curves satisfying GARP to find the highest and lowest. The figure shows the highest and lowest WTP, consistent with the data, that would make the individual indifferent to x^* or the status quo—that is, the range of feasible WTP.

To implement this LP, we need to assure that the set of $\{\boldsymbol{\beta}_i\}$ satisfying GARP is bounded. Without additional restrictions, it may not be. For example, the (finite) data for a particular household may show it always chooses the alternative with the highest value of some attribute *k,* regardless of price. In that case the set of $\{\beta_{ik}\}$ satisfying GARP would be unbounded from above. A practical solution is to set some arbitrary (but defensible) upper bound on the support for $\boldsymbol{\beta}$ through additional inequalities. Equivalently, one could set an upper bound through one additional constraint on the maximal permissible WTP, $\boldsymbol{\beta}'\mathbf{x}^* \leq \text{WTP}^{\max}$. When the $\{\boldsymbol{\beta}_i\}$ are actually unbounded but such a constraint is imposed, the above procedure will then return the upper bound of the support. At the researchers' discretion, values at the boundary could be flagged.

Pricing a product

Another policy question of interest is the highest price a consumer would pay for a discrete product with characteristics $\mathbf{x}^*$ when facing some choice set $\mathcal{J}^*$ in some period t^*. This is still a WTP, but in the presence of more substitutes. Such a question might still be useful in the context of public policy, when an agency is offering a program into which households can select, or in the context of a local government changing a vector of local public goods when households can sort into that jurisdiction or others. But it also pertains to industrial organization, where we are interested in the price the market will bear for a product. The following two-part theorem shows how we can construct bounds for such questions.

> **Theorem 3A**. If the data are linearly rationalizable and the set of $\{\boldsymbol{\beta}_i\}$ rationalizing the data is bounded, then the upper bound on the WTP for a product $\mathbf{x}^*$ is determined by the linear program:
>
> $\text{Max}_{\{\boldsymbol{\beta}_i, p_i^*\}} \sum_i p_i^*$, subject to:
>
> $\boldsymbol{\beta}_i \geq \mathbf{0}$,
>
> Inequalities (4), and

(5) $\quad -p_i^* + \boldsymbol{\beta}_i'\mathbf{x}^* \geq -p_{jt} + \boldsymbol{\beta}_i'\mathbf{x}_{jt}, \ \forall\, i, \forall\, j \in \mathcal{J}^*$, and for $t = t^*$.

That is, we are looking for the highest price p_i^* that would satisfy the participation constraint (5) that household i would choose the product $\mathbf{x}^*$ from the choice set $\mathcal{J}^*$. At the same time, we are free to choose $\boldsymbol{\beta}_i$ to make the product as valuable as possible, so long as the $\boldsymbol{\beta}_i$ are consistent with the observed choice patterns in the sense of satisfying GARP, i.e. consistent with expressions (4). Mechanically, Expression (5) is just like (4), only it is not conditioned by the observed data, but rather a hypothetical condition requiring the new product to be (weakly) preferred when priced at WTP. Accordingly, we can just add an additional "observation" to the data that states the household chooses $\mathbf{x}^*$ at a price p_i^* out of choice set $\mathcal{J}^*$, and then find the highest value of p_i^* that still satisfies GARP.

To find the lower bound, we cannot just minimize the same objective, because we are not looking for the lowest possible price we could charge for a product such that consumers buy it (which might be low, indeed!). We are looking for the lower bound, consistent with the data, on the *highest* price we can charge. However, we can obtain this by reversing the sign in the participation constraint.

> **Theorem 3B**. If the data are linearly rationalizable and the set of $\{\boldsymbol{\beta}_i\}$ rationalizing the data is bounded, then the lower bound on the WTP for a product $\mathbf{x}^*$ is determined by:
>
> $p_i^* = \min_{j\in\mathcal{J}^*}\left\{\min_{\boldsymbol{\beta}_i\geq 0}\{\boldsymbol{\beta}_i'(\mathbf{x}^* - \mathbf{x}_{jt^*}) + p_{jt^*}\}\right\}$, subject to:
>
> $\boldsymbol{\beta}_i \geq \mathbf{0}$ and
>
> Inequalities (4).
>
> *Proof.* For a given $\boldsymbol{\beta}_i$, the highest price we can charge i for $\mathbf{x}^*$ in period t^* is $\boldsymbol{\beta}_i'\mathbf{x}^* - \max_{j\in\mathcal{J}^*}\{-p_{jt^*} + \boldsymbol{\beta}_i'\mathbf{x}_{jt^*}\}$, which is the consumer surplus for $\mathbf{x}^*$ relative to the best alternative. This is equivalent to $\min_{j\in\mathcal{J}}\{\boldsymbol{\beta}_i'(\mathbf{x}^* - \mathbf{x}_{jt^*}) + p_{jt^*}\}$. By continuity and monotonicity of v, the greatest lower bound on the household's maximum willingness to pay is the lowest possible value of this expression, subject to GARP. Thus, we seek $\min_{\boldsymbol{\beta}_i\geq 0}\left\{\min_{j\in\mathcal{J}^*}\{\boldsymbol{\beta}_i'(\mathbf{x}^* - \mathbf{x}_{jt^*}) + p_{jt^*}\}\right\} = \min_{j\in\mathcal{J}^*}\left\{\min_{\boldsymbol{\beta}_i\geq 0}\{\boldsymbol{\beta}_i'(\mathbf{x}^* - \mathbf{x}_{jt^*}) + p_{jt^*}\}\right\}$ such that inequalities (4) are satisfied.

While the double min{} function may at first look daunting, note that, in practice, we can simply solve the LP $\min_{\boldsymbol{\beta}_i\geq 0}\{\boldsymbol{\beta}_i'(\mathbf{x}^* - \mathbf{x}_{jt^*}) + p_{jt^*}\}$ separately for each alternative $j \in \mathcal{J}^*$ and then find the minimal value of the solution over $\mathcal{J}^*$. Potentially, the value may be negative, but this is a completely valid finding: we might have to pay a household to induce it to use a product instead of its favorite alternative.

Figure 4 illustrates the procedure, for the same data as Fig. 3 (but relabeling the status quo as the "outside option).[5] A new product is shown with higher x than any other existing product. Because $\{p_1, x_1\}$ is the customer's current choice, that is the one the new product must outcompete. Because the new product has higher x than x_1, we know for sure we can charge a higher price than p_1. In fact, we can charge at least $p_{\min}$ shown in the figure, because that is on the linear indifference curve associated with the lowest value x_1 consistent with the data and GARP. We can charge no more than $p_{\max}$, as that is associated with the highest WTP for x consistent with the data. How do we know this? The slope of the line going through x_2 represents the highest value of x consistent

[5] The label "status quo" may be somewhat inapt here, but I keep it for comparability to the previous examples. In can be thought of as an arbitrary product, or perhaps the firm's current product design.

with the data, but we need to outcompete x_1, not x_2. So we translate that marginal willingness to pay through x_1 as shown by the dashed blue line in the figure.

The solutions bound the revenue that a firm could obtain under perfect price discrimination. Bounding revenue for a single-price monopolist (or oligopolist in Bertrand competition) would involve taking the solutions as bounds on discrete demand functions and searching for the revenue-maximizing prices.

2.3. Introducing Unobserved Quality into the Model

So far, we have assumed the econometrician observes all relevant information about the choice alternatives faced by households. However, if they are reacting to unobserved characteristics, households may appear to violate GARP even when they are behaving rationally, and the sets of $\boldsymbol{\beta}$ will be misidentified. In their application to computers, for example, Bajari and Benkard (2005) found that, even with 19 characteristics included in $\mathbf{x}$, over half of the data could not be rationalized by a linear utility function with no errors (see also Kuminoff 2009 for discussion). One response to this problem is to introduce idiosyncratic errors, as in most RUMs, which I will consider in Section 3. First, however, consider the possibility of adding only a vertically differentiated unobserved quality attribute.

In particular, consider the following pure characteristics model from Bajari and Benkard (2005) and Berry and Pakes (2007):

$$(6) \qquad v_{ijt} = -p_{jt} + \boldsymbol{\beta}_i'\mathbf{x}_{jt} + \lambda_i\xi_j,$$

where ξ_j is an unobserved time-invariant characteristic and $\lambda_i \geq 0$ is a household-specific weight.[6] The unobserved characteristic ξ can be thought of as an index of multiple characteristics, so long as the underlying index is the same for everybody. Additionally, utility is monotonic in ξ, so the unobservables differentiate products vertically. Nevertheless, it is easy to see how the model can be interpreted as a RUM with $u_{ijt} = \lambda_i\xi_j$.

We can still bound the value of existing products or policy scenarios using this model. For example, following the outline of the previous sub-section, we can bound the value of some product $\{\mathbf{x}^*, \xi^*\}$ relative to an outside option by solving the following program:

[6] Note we can normalize $\lambda_i = 1$ for one household and $\xi_j = 0$ for one product (such as the outside option).

$\text{Max}_{\{\boldsymbol{\beta}_i,\lambda_i,\xi_j\}}$ (respectively $\text{Min}_{\{\boldsymbol{\beta}_i,\lambda_i,\xi_j\}}$): $\sum_i(\boldsymbol{\beta}_i'\mathbf{x}^*)$

subject to

$\boldsymbol{\beta}_i \geq \mathbf{0}$, $\lambda_i \geq 0$, and

(7) $\quad -p_{j'(i,t)t} + \boldsymbol{\beta}_i'\mathbf{x}_{j'(i,t)} + \lambda_i\xi_{j'(i,t)} \ \geq\ -p_{jt} + \boldsymbol{\beta}_i'\mathbf{x}_{jt} + \lambda_i\xi_j,\ \forall\ i,\ \forall\ t,\ \forall\ j \in \mathscr{J}(i,t).$

Here, the Inequalities (7) replace (4) as the Afriat inequalities, simply adding $\lambda_i\xi_j$. It would be straightforward to answer other economic questions using a similar strategy.

Three remarks are worth making. First, in taking $\sum_i\boldsymbol{\beta}_i'\mathbf{x}^*$ as the objective, I assume here researchers will want to set $\xi^*=0$ for the new product, but this is not necessary. One could incorporate an arbitrary value of ξ^* or place it at some percentile of the distribution. Second, the bilinear terms $\lambda_i\xi_j$ in (7) require quadratic (rather than linear) programming. However, the objective remains convex. Third, we now must impose an assumption of mean independence between the ξ_j and $\mathbf{x}$. Normally, in econometric practices, such an assumption is simply assumed. Here, because the ξ_j are solved for, we need to impose it in the program. In practice, we can do this by binning the joint distribution of $\mathbf{x}$ into cells c=1…C, and imposing the additional constraint that $\frac{1}{\sum_{j\in c} 1}\sum_{j\in c}\xi_j$ is equal for all c (i.e., the mean ξ_j is equal across all cells).

It may be possible to generalize the model further. Using the notation $\boldsymbol{\beta}_i = (\overline{\boldsymbol{\beta}} + \widetilde{\boldsymbol{\beta}}_i)$ and $\mu_{ij} = f_i(\xi_j) + \overline{\boldsymbol{\beta}}'\mathbf{x}_j$, where $f_i(\ )$ is any increasing function, we can re-write Equation (6) as:

(8) $\quad v_{ijt} \ =\ -p_{jt} + \widetilde{\boldsymbol{\beta}}_i'\mathbf{x}_j + \mu_{ij},$

with the restriction:

(9) $\quad (\mu_{ij} - \mu_{ij'})(\mu_{i'j} - \mu_{i'j'}) \ \geq\ 0\ \forall\ i,i',\ \forall\ j,j'.$

That is, households rank the ξ_j in the same order: For any two households i,i' and any two products j,j', if $f_i(\xi_j)+\overline{\boldsymbol{\beta}}'\mathbf{x}_j \geq f_i(\xi_{j'})+\overline{\boldsymbol{\beta}}'\mathbf{x}_{j'}$ then $f_{i'}(\xi_j)+\overline{\boldsymbol{\beta}}'\mathbf{x}_j \geq f_{i'}(\xi_{j'})+\overline{\boldsymbol{\beta}}'\mathbf{x}_{j'}$. However, in general this quadratic program is non-convex and may be difficult to solve.[7]

[7] Too, without imposing additional identifying restrictions on the joint distribution of $\{\mathbf{x}, \xi\}$, we would no longer identify the mean tastes for the product $\overline{\boldsymbol{\beta}}$ separately from $f_i(\xi_j)$. However, recovering the parameter μ_{ij} is sufficient information for many policy questions.

3. When GARP is Not Satisfied: Rationalizing the Data with Minimal Idiosyncratic Errors

If the data do not satisfy GARP, the model to this point will be incoherent: no values of $\boldsymbol{\beta}$ will satisfy the constraints. Consequently, the model needs to be generalized to allow for otherwise unexplained departures from GARP. We can do this by introducing additive idiosyncratic errors, as is common practice in discrete choice modeling. We now have the utility function:

$$(10) \quad v_{ijt} = -p_{jt} + \boldsymbol{\beta}_i'\mathbf{x}_{jt} + \lambda_i\xi_j + \varepsilon_{ijt}.$$

In addition to the $\boldsymbol{\beta}_i$, this model has $J(i,t)$ parameters per choice occasion, per person (namely, the ε_{ijt} terms). The Afriat inequalities become:

$$(11) \quad -p_{j'(i,t)} + \boldsymbol{\beta}_i'\mathbf{x}_{j'(i,t)} + \lambda_i\xi_{j'(i,t)} + \varepsilon_{ij'(i,t)t} \geq -p_j + \boldsymbol{\beta}_i'\mathbf{x}_j + \lambda_i\xi_j + \varepsilon_{ijt}, \ \forall\, i, \forall\, t, \forall\, j \in \mathcal{J}(i,t).$$

If we view these idiosyncratic errors through the lens of the revealed preference literature (and estimate the model with an LP), the ε_{ijt} are now taken as parameters (in econometrics parlance, "variables" in programming parlance) which allow us to satisfy these inequalities for every choice occasion.[8] That is, we are not just fitting "market shares" or the probabilities that consumer chooses a product given some distribution of $\boldsymbol{\varepsilon}_i$, we are jointly estimating permissible sets of $\boldsymbol{\beta}_i$ and $\boldsymbol{\varepsilon}_i$ that perfectly explain each observed choice.[9] As Kuminoff (2009) discusses, this makes such models radically incomplete. The set of $\{\boldsymbol{\beta}_i\}$ consistent with GARP are now unbounded: For any $\boldsymbol{\beta}_i$ there are always idiosyncratic errors that could explain a choice. Accordingly, we need some rule for resolving, or at least reducing, this incompleteness, ideally a rule with economic and/or statistical foundations (Lewbel 2019).

A natural choice is to select the subset of permissible $\{\boldsymbol{\beta}_i, \lambda_i, \xi_j, \boldsymbol{\varepsilon}_i\}$ that minimizes some norm of $\boldsymbol{\varepsilon}$, not unlike a least squares problem. In fact, the revealed preference literature has a long tradition of doing just that, to find what is known as "critical cost efficiency" (Afriat 1972, Varian 1985, 1990). The basic idea is to find the smallest adjustment to the data needed to make them consistent with GARP. To understand these adjustments, consider as a simple example a violation

[8] In this paper, I use the language of econometrics, "variables" are given and the objects to be estimated are "parameters." Confusingly, in the language of programming, "parameters" are given and the objects to be estimated are "variables."

[9] In this sense, the "solved" errors are comparable to the conditional distribution of parametric RUM errors, where one conditions on observed choices (von Haefen 2003).

of Samuelson's Weak Axiom of Revealed Preference (WARP). If WARP is violated, the household chooses bundle A when B is available but on another occasion chooses B when A is available. Equivalently, the household behaves as if either B was not available the first time or A was not available the second time. With competitive (convex) budget sets the adjustments are equivalent to adjustments to the household's income: Either it wasted expenditure in the first case (by buying A when it could have done at least as well with the cheaper B) or vice versa. With discrete alternatives in characteristic space, we can adapt the idea of the efficiency index by considering how much money a household must seemingly throw away when it makes a particular choice. That is, how much does it understate the *cost* of products it buys, or overstate those it does not buy?

That is, suppose we simply re-write (11) by regrouping some terms:

$$-(p_{j'(i,t)t} - \varepsilon_{ij'(i,t)t}) + \boldsymbol{\beta}_i'\mathbf{x}_{j'(i,t)t} + \lambda_i\xi_{j'(i,t)} \geq -(p_j - \varepsilon_{ijt}) + \boldsymbol{\beta}_i'\mathbf{x}_j + \lambda_i\xi_j, \ \forall\, i, \forall\, t, \forall\, j \in \mathcal{J}(i,t).$$

Written this way, the relationship between the errors and money becomes transparent. The errors either lower the effective price of the chosen product or raise it for the products not chosen (or both).

Figure 5 illustrates the procedure. Again, $\{p_1, x_1\} \succ_R \text{SQ} \succ_R \{p_2, x_2\}$, but now these choices are not linearly rationalizable. The indifference curve going through SQ must pass below $\{p_1, x_1\}$, which puts $\{p_2, x_2\}$ in the strictly-better than set for SQ. We can explain this apparent contradiction with a positive error on the first bundle, as if it had a cheaper price, $\{p_1 - \varepsilon_1, x_1\}$, which shows up as a vertical displacement in the figure, until it appears in the upper contour of the lowest candidate indifference curve going through x_2. Minimizing the errors in this way can be thought of as bringing the model as close as possible to the pure characteristics model, introducing idiosyncratic errors only when that model violates GARP. Thus, this generalization nests the pure characteristics model, but also can accommodate the concerns raised by Athey and Imbens (2007) about the difficulty of forcing the pure characteristics model on data that violate it.

In the application below I minimize the sum of absolute errors.[10] I also impose an inde-

[10] Varian (1990) suggests this is computationally difficult when using his proposed algorithms. However, as I show in the following section, it can easily be built into the objective functions for LP tests of discrete choices proposed by Cosaert and Demuynck (2015). Although the usual absolute value function is non-

pendence assumption that the ε's are mean zero in each of 33 data cells, as well as globally orthogonal to p and $\mathbf{x}$. Note also that the distribution of ε can—and typically will—have a large probability mass at zero (indeed, a degenerate distribution with all mass at zero when GARP is satisfied), so the recovered distribution function of ε will not be continuous.

As McFadden (2014) has observed, in a cross section, the interpretation of idiosyncratic errors as fixed inter-agent heterogeneity or as intra-agent heterogeneity across choice alternatives (based on a "moment's fancy") are observationally equivalent. By the same token, so too in a panel are the interpretation of the errors as either transient tastes and/or shifting quality, or alternatively as irrational behavior in the face of unchanging tastes and quality. That is, ε could be a measure of the error, either by the agents or the analyst in observing prices, or it could be a measure of how wrong the pure characteristics model is in the first place, how mistaken we are to ignore idiosyncratic tastes.[11] We can *never* tell the difference between a model where people do not violate GARP, but where there are transient shocks to tastes and/or transient unobserved shocks to product quality, and one where there are no such shocks and people do violate GARP. This observational equivalence is why we can interpret the errors through either the lens of critical cost efficiency or though the lens of random utility. To my knowledge, this connection between critical efficiency in the GARP framework and errors in the RUM framework has not previously been pointed out.

Even starting from the RUM framework rather than the GARP framework, arguably the errors still are best interpreted as just a retrofit. They are a way to improve the fit of an empirical model when tastes for characteristics alone cannot fully explain observed choice patterns. Thus, minimizing a metric of the errors makes sense regardless of the interpretation. If they are interpreted as actual violations of GARP, they should be minimized following the logic of Afriat (1972) and Varian (1990). Alternatively, if they are interpreted as idiosyncratic tastes, they still should be minimized so that the pure characteristics portion of the model, $\boldsymbol{\beta}_i'\mathbf{x}$, gives the best fit to the data

---

linear, it is well known that it can be reformulated as an LP by replacing ε_i with $(\varepsilon_i^+ - \varepsilon_i^-)$ in the constraints, where ε_i^+ and ε_i^- are both defined as positive variables, and minimizing $(\varepsilon_i^+ + \varepsilon_i^-)$.

[11] Of course, as the argument has been developed so far, they also could be a measure of how wrong the linearity assumption is. However, because I have only introduced linearity for simplicity, I ignore this possibility here. Below, I relax linearity, taking a non-parametric approach.

possible. However, when adopting the GARP framework, "the best fit to the data" becomes equivalent to "satisfying GARP," or finding the pure characteristics model that comes as close to the data as possible (i.e., when ignoring the errors). This is different than the standard RUM approach of integrated over the errors to fit market shares.

In practice, this procedure sometimes will point identify $\boldsymbol{\beta}_i$ for any individual (or group) that violates GARP, as the focus is on limiting the norm of $\boldsymbol{\varepsilon}$. Returning to the policy questions, this means that, say, the sum of WTP may involve the sum over point-identified WTP values (for individuals that violate GARP) and over set-identified WTP values (for individuals that don't), which of course still results in a set-identified aggregate WTP. In theory, we cannot rule out the possibility that $\boldsymbol{\beta}_i$ is only set identified even for individuals who do violate GARP (if two solutions are tied on the criteria of minimizing the norm of ε but result in different WTP values). In this case, one can minimize a weighted sum of the norm of ε and WTP, with sufficient weight on the ε's to guarantee they are minimized, and the range of WTP found only within the set of parameters where the norm of ε is tied.

4. A Non-Parametric Approach

So far, to develop the argument on familiar ground, I have relied on a linear utility function. But eschewing such functional form restrictions has been a central feature of the revealed preference approach since Samuelson first introduced it. Figure 6 illustrates the potential importance of relaxing linearity. It repeats the example from Figure 5, with $\{p_1, x_1\} \succ_R \text{SQ} \succ_R \{p_2, x_2\}$. As previously noted, this choice behavior is not linearly rationalizable. However, it is easily rationalizable by a quasi-concave utility function, such as the indifference curve shown in the figure. Thus, these data are actually consistent with the most general expression of GARP; only the linearity assumption is violated. With competitive budget sets, Afriat's theorem allows us to easily test whether there is *any* locally non-satiated utility function that rationalizes the data, using LP. With discrete choices, we can still test for the existence of any strictly monotonic, quasi-concave utility function that rationalizes the data (Cosaert and Demuynck 2015). In Figure 6, the shaded area represents the valid region for the indifference curve passing through SQ, given the choice behavior, monotonicity, and quasi-concavity.

Extending Afriat's basic logic, Cosaert and Demuynck (2015, Thm. 3) formally demonstrate that, for a household i, discrete choice data satisfy GARP, and can be represented by a weakly monotonic concave utility function, if and only if there exist numbers $v_{ij'(i,t)}$, v_{ij}, $\hat{\alpha}_{ij}$, $\hat{\boldsymbol{\pi}}_{ij}$ such that

(12) $v_{ij'(i,t)} \geq v_{ij}$, $\forall\, t, \forall\, j \in \mathcal{J}(i,t)$.

(13) $v_{ij} - v_{ik} \geq -\hat{\alpha}_{ij}(p_j - p_k) + \hat{\boldsymbol{\pi}}_{ij}(\mathbf{x}_j - \mathbf{x}_k)$, $\forall\, t, \forall\, j,k \in \mathcal{J}(i,t)$.

The first condition simply states that we must find utilities for chosen bundles j' that exceed those for other bundles available when it was chosen. The second condition states that for any alternative j, there is a supporting hyperplane, given by $\hat{\alpha}_{ij}$, $\hat{\boldsymbol{\pi}}_{ij}$, to the indifference curve going through j. This hyperplane represents a vector of shadow prices on the characteristics (times the marginal utility of income). Although the set of alternatives is discrete, the shadow prices represent the budget constraint that *would* have induced a purchase of j over any other alternative k, if, hypothetically, we were in a competitive budget setting. Importantly, they include the marginal utilities of income, explaining why a person might choose a bundle A on one choice occasion even though it is dominated by a bundle B available on *another* occasion. These conditions differ from the usual Afriat inequalities (with competitive budget sets) in that in this case the $\hat{\alpha}$, $\hat{\boldsymbol{\pi}}$ are shadow values that must be solved for rather than observed prices. However, whereas in the usual case we can solve for marginal utilities of income given the observed prices, here we solve for shadow prices that are not identified to scale separately from the marginal utilities of income.

I extend Cosaert and Demuynck's model in two ways. First, I introduce additive errors representing critical efficiency, so Expression (12) now becomes

(14) $v_{ij'(i,t)} + \varepsilon_{ij'(i,t)} \geq v_{ij} - \varepsilon_{ij}$, $\forall\, t, \forall\, j \in \mathcal{J}(i,t)$,

which are the Afriat inequalities accounting for the errors.

Second, I introduce the policy as an artificial alternative j^*. Maximum WTP can be found by solving for the maximum value of WTP such that

(15) $v_{ij^*} \geq v_{i0}$,

(16) $v_{ij^*} - v_{i0} \geq -\hat{\alpha}_{ij^*}[(p_{j^*} + \text{WTP}) - p_0] + \hat{\boldsymbol{\pi}}_{ij^*}(\mathbf{x}_{j^*} - \mathbf{x}_0)$

(17) $v_{i0} - v_{ij^*} \geq -\hat{\alpha}_{i0}[p_0 - (p_{j^*} - \text{WTP})] + \hat{\boldsymbol{\pi}}_{i0}(\mathbf{x}_0 - \mathbf{x}_{j^*})$,

where 0 represents the baseline or status-quo state for comparison. The intuition is that we are looking for the highest possible price that would be associated with the policy and allow us to choose it over the status quo, while still satisfying GARP. Minimum WTP can be found by maximizing the value of WTP when the inequality in Expression (15) reversed. In that case, we are looking for the lowest possible price that would allow us to choose the status quo over the policy while satisfying GARP. The two values bracket WTP. As written, this problem is non-linear because it involves the product of two variables, $\widehat{\alpha}$ and WTP. In my application, the NLP converged quickly to a solution using modern non-linear solvers like Baron (Khajavirad and Sahinidis 2018). However, if analysts find it difficult in their applications, the problem can also be solved by nesting an LP within a line search over WTP.[12]

5. Empirical Application to Ecosystem Services

I apply the methods proposed in this paper to the valuation of improvements in ecosystem services in the Southern Appalachian Mountains of the United States. The data are based on a "conjoint" survey, or choice experiment, of households in the southeastern US. Choice experiments ask respondents to choose an alternative from a hypothetical menu of options, with each option described by a set of attributes and a price.[13] Such data make for an appropriate illustration of the proposed methods, for three reasons. First, they involve a panel of choices at the individual level, making it possible to set identify each β_i. Second, the data involve experimentally controlled characteristics $\mathbf{x}$ as well as prices, sidestepping the need for instruments in this case. Finally, data from choice experiments are widely used in marketing studies for new products, as well as in applications to environmental policy (e.g. Banzhaf et al. 2016, Lewbel, Linton, and McFadden 2013) and health (Mühlbacher and Johnson 2016).

The area studied includes portions of West Virginia, Virginia, North Carolina, Tennessee, and Georgia. The survey was distributed to residents of those states both online and by mail from

---

[12] Specifically, for the case of maximizing WTP, first bound WTP and begin with an initial guess. Then, simply test for GARP using the Cosaert-Demuynck LP, based on the observed data plus the augmented "observation" that the household votes for the policy (or buys the product) at the trial WTP value. If these augmented data satisfy GARP, this WTP value is consistent with the observed behavior. The initial guess can be set at the lower bound and a new, higher guess made. If the data are not consistent with GARP, the guess is too high and can be set at the upper bound. This process can be repeated until the WTP range converges to a tolerable limit. A similar process can be used for minimizing WTP. When the initial data do not satisfy GARP, the same procedure can be used, constraining the sum of absolute errors to its value without the augmented policy "observation."

[13] See, e.g., Ben-Akiva, McFadden, and Train (2019), Johnston et al. (2017), or Mühlbacher and Johnson (2016) for an overview of choice experiments.

October 2009 to December 2010. It introduces the state of ecosystem services in the region, with quantitative information in three dimensions: streams and fish, forests, and birds. Based on scientific models, it describes 20 percent of small streams in high-elevation areas as being affected by acid precipitation (about 60,000 streams). Additionally, the survey describes six species of fish affected in the streams: brook trout, rainbow trout, mottled sculpin, longnose dace, rosyside dace, and fantail darter. It also describes 3% of the forested area as affected by acidification. Finally, it describes three bird species (waterthrush, crossbill, and ovenbird) as harmed by acidification, with their populations now at 65% of what they once were.

The survey next tells respondents that a program has been proposed to improve the streams, forests, and affected bird populations in the region. After describing the program, the survey explains that the respondent's state government will fund its share of the liming program with a revenue bond that will be paid off by additional state income tax payments for the next 10 years. Next, the survey describes a set of improvements to streams, forests, and bird populations that the program could achieve over a ten-year period. It provides a sequence of six choice sets, each including the status quo or "no program" option plus one to two variants of the program. The choice alternatives range from no improvement in each dimension, to a maximal improvement of 45,000 streams (15% of all streams), 625,000 acres of forest improvement (2.5% of all forests) or a 30 pp improvement in the three bird populations (from 65% of what they once were to 95%). Respondents indicated whether they would be willing to support the program and the indicated annual cost to them over ten years. Collectively, 1,512 respondents evaluated 293 unique bundles of the three attributes on at least one choice occasion, or 1,651 bundles when also considering different costs.[14] The alternatives were selected randomly using a Bayesian experimental design with updating, to maximize the statistical efficiency of the variable space. See Banzhaf et al. (2016) for additional details about the survey and its implementation.

Based on these data, I evaluate the mean WTP of one bundle of possible improvements to ecosystem services from a reduction in acidification from air pollution, in particular improving

[14] An additional 1,678 respondents faced one or two choice occasions in a traditional dichotomous choice contingent valuation format. These respondents were included in the original parametric analysis of the data in Banzhaf et al. (2016) but were excluded here.

15,000 streams in the region, 250,000 acres of forest, and 15 pp of bird populations. These improvements are within the range of the data considered by the survey. I assume they take 50 years to reach full maturity, and that they then continue indefinitely. I report annualized WTP for this full stream of services (i.e., as of year 50) in present value terms, using a discount rate of 3%.

For comparison purposes, I first compute values using a simple parametric RUM, a linear logit model imposing homogeneity in taste parameters, with heterogeneity entering only through the additive errors. Parameters include the marginal utility of money α and a vector $\boldsymbol{\beta}$ for streams, forest, and birds, plus a dummy variable for the status quo option. This dummy can be interpreted as representing the effect of a reference point, the taste for "doing something" if positive or "not interfering" if negative. Alternatively, it can simply be interpreted as adding a kink to an otherwise linear indifference curve, to improve the fit to the data. Other than this dummy, I do not include product-specific unobservables ξ_j in the model, as in this application there are no obvious "products" to model. After the logit, I estimate the linear GARP alternative, with no distributional assumptions on the errors, based on the logic of Theorem 2. This model finds the vector $\boldsymbol{\beta}$ that minimizes the additive errors needed to satisfy GARP for a representative consumer. In this case, the panel structure of the data can be thought of as individuals representing separate observations of a representative consumer.

Table 1 displays the results. Because the policy relevance of tastes for the status quo is controversial (Banzhaf et al. 2016), the table shows WTP estimates respectively ignoring or including this parameter. The first column shows the results for the standard multinomial model. Mean WTP is \$52 if we ignore the effect of the status quo dummy when calculating WTP, with a 95% confidence interval (based on standard errors clustered by individual respondent) spanning \$40 to \$64. If we allow for the disutility of interfering with the status quo when evaluating the policy, mean WTP falls to \$21, with a confidence interval of \$7 to \$35. The average absolute difference in logit errors between any two options, normalized by the marginal utility of income, is \$146, indicating substantial randomness in this RUM relative to the fit of the parameters.

The second column shows the results for the GARP approach to estimating a linear RUM proposed in this paper, again imposing homogeneity in $\boldsymbol{\beta}$. This model is essentially the same as the logit version, but imposes no distributional assumptions on the additive errors. Because this model imposes homogeneity in the face of what is likely substantial heterogeneity, all the burden of the model is on satisfying GARP as close to possible, with no weight given to altering the parameters to minimize or maximize WTP. Thus, this model is point identified. Mean WTP is estimated to be \$25 when omitting the status quo and \$40 when including it. The confidence intervals, based on cluster-bootstrapping the data by respondent and recalculating the LP for each bootstrapped sample, are fairly tight, at about +/- \$3. These estimates are in the same range as the linear logit model, but they flip the relative order of the models with and without the status quo. As we would expect, the GARP-based LP approach drives the errors much lower. In other words, it comes much closer to the pure characteristics model. The average $|\varepsilon|$ is only \$12 versus \$146 for the logit, and 80% of observations and 7% of households require no ε at all. However, it has a similar record as the logit at predicting choices, with each predicting about 50% correctly based only on the characteristics. When integrated over the error distribution, the GARP model does somewhat better at predictions in this case, at 49% correct vs. 42%.

The assumption of homogeneity imposed in these models is a strong one, but it can be relaxed both with parametric methods (e.g. random coefficients, or mixed, logit) as well as with the GARP-based methods suggested in this paper. Table 2 presents results allowing for unobserved heterogeneity. Again for comparison purposes, the first column shows results using a logit model, with random coefficients on the status quo dummy distributed normal and random coefficients on streams and forests distributed log-normal (to impose positive values). (The coefficient on birds is kept homogeneous, as attempts to estimate heterogeneity were fragile.) Heterogeneity has little effect on the estimate when ignoring the status quo dummy, which is now \$49 (CI \$35 to \$64), compared to \$52 in the model with homogeneity. However, it now increases it substantially when including the status quo dummy, from \$21 to \$77 (CI \$55 to \$100). Including heterogeneity in the random parameters allows the scale of the additive logit errors, relative to the model, to fall, with the average absolute value of the difference in errors now at \$75 instead of \$146.

The next pair of columns shows the estimates using the linear GARP-based approach with random coefficients, as in Theorem 2. As most individuals satisfy GARP, the model is now only set identified, with a continuum of $\boldsymbol{\beta}_i$ satisfying GARP but having different implications for WTP. The two columns in the pair show the results for the argmin and argmax in $\{\boldsymbol{\beta}_i\}$ for the WTP objective. Plausible mean WTP ranges from \$36 to \$259 when ignoring the status quo dummy, and -\$130 to \$190 when including it. Both values bracket the respective range from the logit model, although this result is not guaranteed.[15] The WTP range reflects model uncertainty, which is large, but sampling variance remains small, with tight confidence intervals. (To compute standard errors, we can now simply take the sampling distribution across households rather than bootstrap.) Again, we see that the model shrinks the additive errors much more than the logit model assumes, with average errors now \$2 rather than \$75 for logit and \$10 for the GARP approach without heterogeneity, and with about 97% of observations having no need of errors to fit the data. Seventy-four percent of households have choice patterns that are linearly rationalizable, so require no error at all to fit the data. Finally, the GARP model can find 93% of choices correctly focusing only on the characteristics. This prediction rate falls to 81-82% when integrating over the error distribution, but this is substantially better than the mixed logit model, which correctly predicts only 39% of choices.[16]

The remaining columns use the non-parametric GARP-based approach of Section 4. In this case, separate models were estimated with and without the status quo dummy. Not surprisingly, these models are able to reduce GARP violations, with mean errors falling to about \$0.36 to \$0.47 depending on the model, and with 98% of observations and 88 – 92% of households having no error. The plausible range of mean WTP actually is now \$44 to \$210 for the model without status quo, which is not much different than the linear model. However, the range grows rapidly for the model with the status quo, at -\$282 to \$224. Because, in this application, so much of the mass of data is at the status quo, it is apparently difficult to nonparametrically identify WTP for policy improvements (without any linear interpolations between data points) while controlling for the status quo nonparametrically as well. Nevertheless, it should be noted that these values do continue to bracket the point estimates from the mixed logit model. Finally, in the bottom rows we see that the models' predictions are better than the linear version, and remain much better than the mixed logit.

For policy analysis, the non-parametric model without the status quo dummy is probably the most appropriate, as it is the most general but does not allow an arbitrary reference point to

[15] There is nothing that guarantees the GARP-based approach will always bound the logit estimates. Consider for example a case where all the data satisfy GARP. The GARP-approach will set identify a range of WTP using a pure characteristics model (with no idiosyncratic errors). The logit estimator conditions on its distributional assumption, which can take it outside the range of the GARP model.

[16] Note it is not possible to compute the predictions in pure characteristics space for the mixed logit, except at the "average tastes," since the value of the coefficients for any one given respondent is unknown.

affect policy. With 13.3 million households in study area, and using a discount rate of 3%, even the most conservate estimate of \$13.25 implies an aggregate present value of \$10.4 billion for ecosystem service improvements that might be realized from a reduction in air pollution. When added to the substantial health benefits of most air pollution control policies, these benefits are likely to outweigh the costs of most policies.

5. Conclusions

This paper shows how, starting with a RUM framework, one can nonparametrically set-identify the answers to policy questions using only the Generalized Axiom of Revealed Preference (GARP). This approach recasts the RUM errors as departures from GARP, to be minimized using a minimum-distance criterion, and provides another avenue for nonparametric identification of the RUM model. When GARP is satisfied, this approach easily estimates the pure characteristics model. When GARP is violated, it gets as close to that model as possible.

Given my use of survey data, it was plausible to treat all variables as exogenous. As Berry and Haile (2016, 2023) have emphasized, valid instruments are the core foundation of RUMs. Future work might consider how to introduce additional exclusion restrictions. For example, if prices p are endogenous one might estimate $E[p|\mathbf{x},\mathbf{z}]$ for additional instruments $\mathbf{z}$ and then minimize errors to satisfy GARP over the conditional expectation of p rather than p itself, not unlike the approach taken by Blundell, Browning, and Crawford (2003). The potential to apply generalized instrumental variables (Chesher and Rosen 2017) might be another fruitful avenue to pursue.

Future work might also consider the implications of this approach for designing choice experiments. When using surveys or conducting lab experiments, researchers have considered algorithms to update the bundles of characteristics offered, to provide the best information for estimating a parametric utility function (e.g. Sándor and Wedel 2005). Much earlier, using a revealed preference approach, Mas-Colell (1978) showed that, by increasing the number of observations, one can identify the underlying preferences for the consumer, in the case of competitive budget sets. Further work might consider how Mas-Colell's insight might be applied to discrete choice sets. As highlighted by the simulations, with experimental data prices can thus be adjusted efficiently, while holding incomes and tastes constant. In contrast to controlled experiments, in the real world incomes and tastes may change between observations. Blundell, Browning and Crawford's (2003, 2008) suggestions to adjust budget sets for changes in income might also be combined with the approaches suggested in this paper.

The paper illustrates the approach by estimating bounds on the values of ecological improvements in the Southern Appalachian Mountains. In the most general model, WTP is bounded between \$44 and \$210 using the GARP-based approach, and point identified at \$77 using a mixed logit model. While such bounds may be less satisfying than a point identified model, they represent an honest appraisal of modeling uncertainty and can help take the "con" out of econometrics (Leamer 1983).

References

Adamowicz, Wiktor L., and Theodore Graham-Tomasi. 1991. "Revealed preference tests of nonmarket goods valuation methods." *Journal of Environmental Economics and Management* 20(1): 29-45.

Afriat, S.N. 1967. "The construction of a utility function from demand data." *International Economic Review* 8(1): 67-77.

Afriat, S.N. 1972. "Efficiency estimation of production functions." *International Economic Review* 13(3): 568-98.

Allen, Roy, and John Rehbeck. 2019. "Identification with additively separable heterogeneity." *Econometrica* 87(3): 1021-54.

Athey, Susan, and Guido W. Imbens. 2007. "Discrete Choice Models with Multiple Unobserved Choice Characteristics." *International Economic Review* 48(4): 1159-92.

Banzhaf, H. Spencer, Dallas Burtraw, Susie Chung Criscimagna, Bernard J. Cosby, David A. Evans, Alan J. Krupnick, Juha V. Siikamäki. 2016. "Valuation of ecosystem services in the southern Appalachian Mountains." *Environmental Science and Technology* 50(6): 2830-36.

Bajari, Patrick, and C. Lanier Benkard. 2005. "Demand estimation with heterogeneous consumers and unobserved product characteristics: A hedonic approach." *Journal of Political Economy* 113(6): 1239-76.

Bayer, Patrick, Robert McMillan, Alvin Murphy, and Christopher Timmins. 2016. "A dynamic model of demand for houses and neighborhoods." *Econometrica* 84 (3): 893–942.

Ben-Akiva, Moshe, Daniel McFadden and Kenneth Train. 2019. "Foundations of Stated Preference Elicitation: Consumer Behavior and Choice-based Conjoint Analysis." *Foundations and Trends in Econometrics* 10(1-2): 1–144.

Berry, Steven T., and Philip A. Haile. 2023. "Nonparametric identification of differentiated products demand using micro data." Working Paper.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile prices in market equilibrium." *Econometrica* 23: 242-62.

Berry, Steven, James Levinsohn, and Ariel Pakes. 2004. "Differentiated products demand systems from a combination of micro and macro data: The new car market." *Journal of Political Economy* 112(1): 68-105.

Berry, Steven, and Ariel Pakes. 2007. "The pure characteristics demand model." *International Economic Review* 48(4): 1193-1225.

Blow, Laura, and Richard Blundell. 2018. “A nonparametric revealed preference approach to measuring the value of environmental quality.” *Environmental and Resource Economics* 69(3): 503-27.

Blow, Laura, Martin Browning, and Ian Crawford. 2008. “Revealed preference analysis of characteristics models.” *Review of Economic Studies* 75(2): 371-89.

Blundell, Richard W., Martin Browning, and Ian A. Crawford. 2003. “Nonparametric Engel curves and revealed preference.” *Econometrica* 71(1): 205-40.

Blundell, Richard, Martin Browning, and Ian Crawford. 2008. “Best nonparametric bounds on demand responses.” *Econometrica* 76(6): 1227-62.

Chambers, Christopher P., and Federico Echenique. 2016. *Revealed Preference Theory*. Cambridge: Cambridge University Press.

Chesher, Andrew, and Adam M. Rosen. 2017. “Generalized instrumental variable models.” *Econometrica* 85(3): 959-89.

Cosaert, Sam, and Thomas Demuynck. 2015. “Revealed preference theory for finite choice sets.” *Economic Theory* 59(1): 169-200.

Crooker, John, and Catherine L. Kling. 2000. “Nonparametric bounds on welfare measures: a new tool for nonmarket valuation.” *Journal of Environmental Economics and Management* 39(2): 145-61.

English, Eric, Roger H. von Haefen, Joseph Herriges, Christopher Leggett, et al. 2018. "Estimating the value of lost recreation days from the Deepwater Horizon oil spill." *Journal of Environmental Economics and Management* 91: 26-45.

Fox, Jeremy T., Kyoo il Kim, Stephen P. Ryan, and Patrick Bajari. 2012. “The random coefficients logit model is identified.” *Journal of Econometrics* 166(2): 204-12.

Gandhi, Amit, and Aviv Nevo. 2021. “Empirical models of demand and supply in differentiated products industries.” In *Handbook of Industrial Organization Vol. 4*, ed. by Kate Ho, Ali Hortaçsu, and Alessandro Lizzeri, 63-139. Elsevier.

Hands, D. Wade. 2014. “Paul Samuelson and revealed preference theory.” *History of Political Economy* 46(1): 85-116.

Johnston, Robert J., Kevin J. Boyle, Wiktor Adamowicz, Jeff Bennett, et al. 2017. “Contemporary guidance for stated preference studies.” *Journal of the Association of Environmental and Resource Economists* 4(2): 319-405.

Khajavirad, Aida, and Nikolaos V. Sahinidis. 2018. “A hybrid LP/NLP paradigm for global optimization relaxations.” *Mathematical Programming Computation* 10: 383-421.

Keane, Michael, Jonathan Ketcham, Nicolai Kuminoff and Timothy Neale. 2021. "Evaluating consumers' choices of Medicare Part D plans: A study in behavioral welfare economics." *Journal of Econometrics* 222: 107-40.

Kitamura, Yuichi, and Jörge Stoye. 2018. "Nonparametric analysis of random utility models." *Econometrica* 86(6): 1883-1909.

Kuminoff, Nicolai V. 2009. "Decomposing the structural identification of non-market values." *Journal of Environmental Economics and Management* 57(2): 123-39.

Leamer, Edward E. 1983. "Let's take the con out of econometrics." *American Economic Review* 73(1): 31-43.

Lewbel, Arthur. 2019. "The identification zoo: Meanings of identification in econometrics." *Journal of Economic Literature* 57(4): 835-903.

Lewbel, Arthur, Daniel McFadden, and Oliver Linton. 2011. "Estimating features of a distribution from binomial data." *Journal of Econometrics* 162(2): 170-88.

Manski, Charles F. 2007. "Partial identification of counterfactual choice probabilities." *International Economic Review* 48(4): 1393-1410.

Manski, Charles F. 2010. "Partial identification in economics." In *Microeconometrics*, ed. by Steven N. Durlauf and Lawrence E. Blume. Macmillan.

Mas-Colell, Andreu. 1978. "On revealed preference analysis." *Review of Economic Studies* 45(1): 121-31.

McFadden, Daniel. 1981. "Structural discrete probability models derived from theories of choice." In *Structural Analysis of Discrete Data and Econometric Applications*, ed. by Charles F. Manski and Daniel L. McFadden. MIT Press, pp. 198-272.

McFadden, Daniel L. 2005. "Revealed stochastic preference: A synthesis." *Economic Theory* 26: 245-64.

McFadden, Daniel L. 2014. "The new science of pleasure: Consumer choice behavior and the measure of well-being." In *Handbook of Choice Modelling*, ed. by Stephane Hess and Andrew Daly, pp. 7-48. Edward Elgar.

McFadden, Daniel, and Kenneth Train. 2000. "Mixed MNL models for discrete response." *Journal of Applied Econometrics* 15(5): 447-70.

Mühlbacher, Axel, and F. Reed Johnson. 2016. "Choice experiments to quantify preferences for health and healthcare: state of the practice." *Applied Health Economics and Health Policy* 14: 253-66.

Polisson, Matthew, and John K-H. Quah. 2013. "Revealed preference in a discrete consumption space." *American Economic Journal: Microeconomics* 5(1): 28-34.

Revelt, David, and Kenneth Train. 1998. "Mixed logit with repeated choices: households' choices of appliance efficiency level." *Review of Economics and Statistics* 80(4): 647-657.

Sándor, Zsolt, and Michel Wedel. 2005. "Heterogeneous conjoint choice designs." *Journal of Marketing Research* 42(2): 210-18.

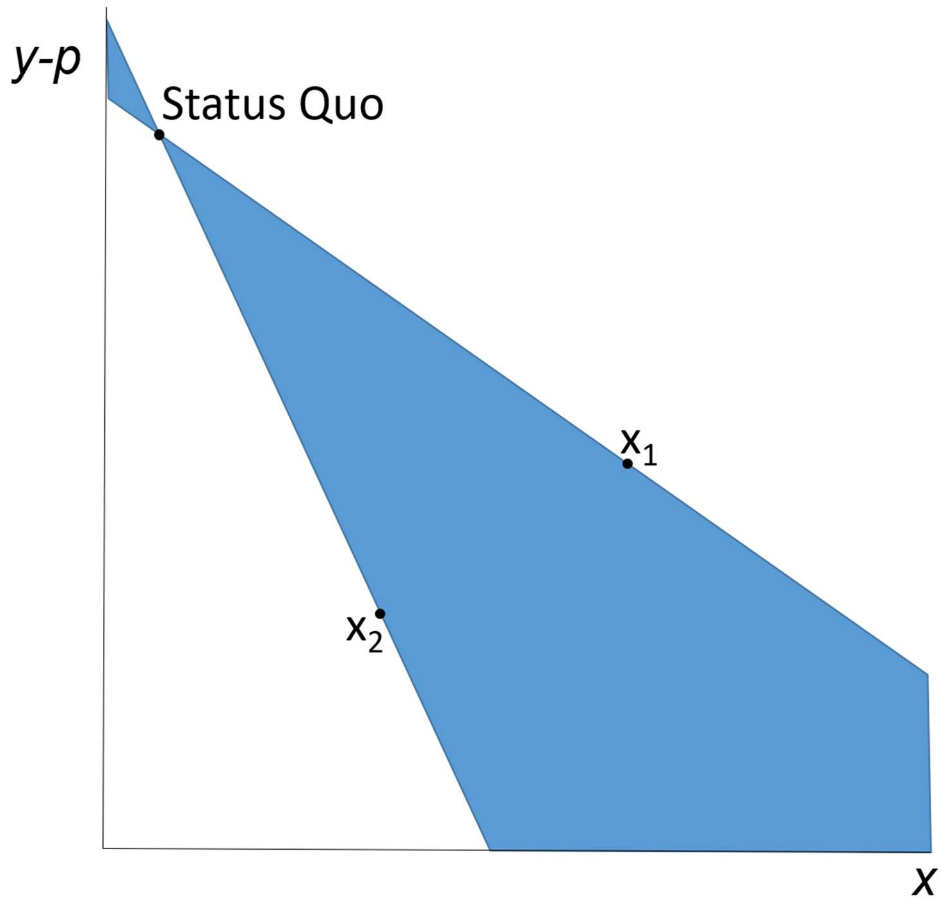
Shi, Xiaoxia, Matthew Shum, and Wei Song. 2018. "Estimating semi-parametric panel multinomial choice models using cyclic monotonicity." *Econometrica* 86(2): 737-61.

Varian, Hal R. 1985. "Non-parametric analysis of optimizing behavior with measurement error." *Journal of Econometrics* 30: 445-58.

Varian, Hal R. 1990. "Goodness-of-fit in optimizing models." *Journal of Econometrics* 46(1-2): 125-140.

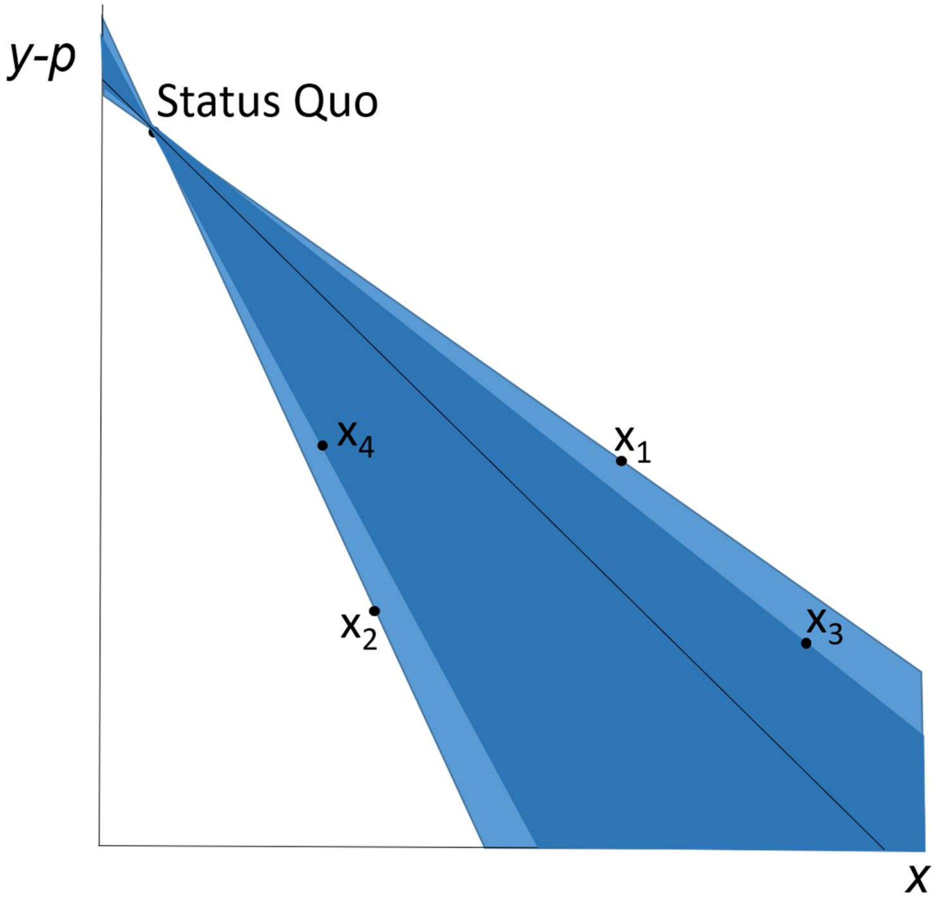
von Haefen, Roger H. 2003. "Incorporating observed choice into the construction of welfare measures from random utility models." *Journal of Environmental Economics and Management* 45: 145-65.

Figure 1. Set Identifying the Linear RUM with Random Coefficients



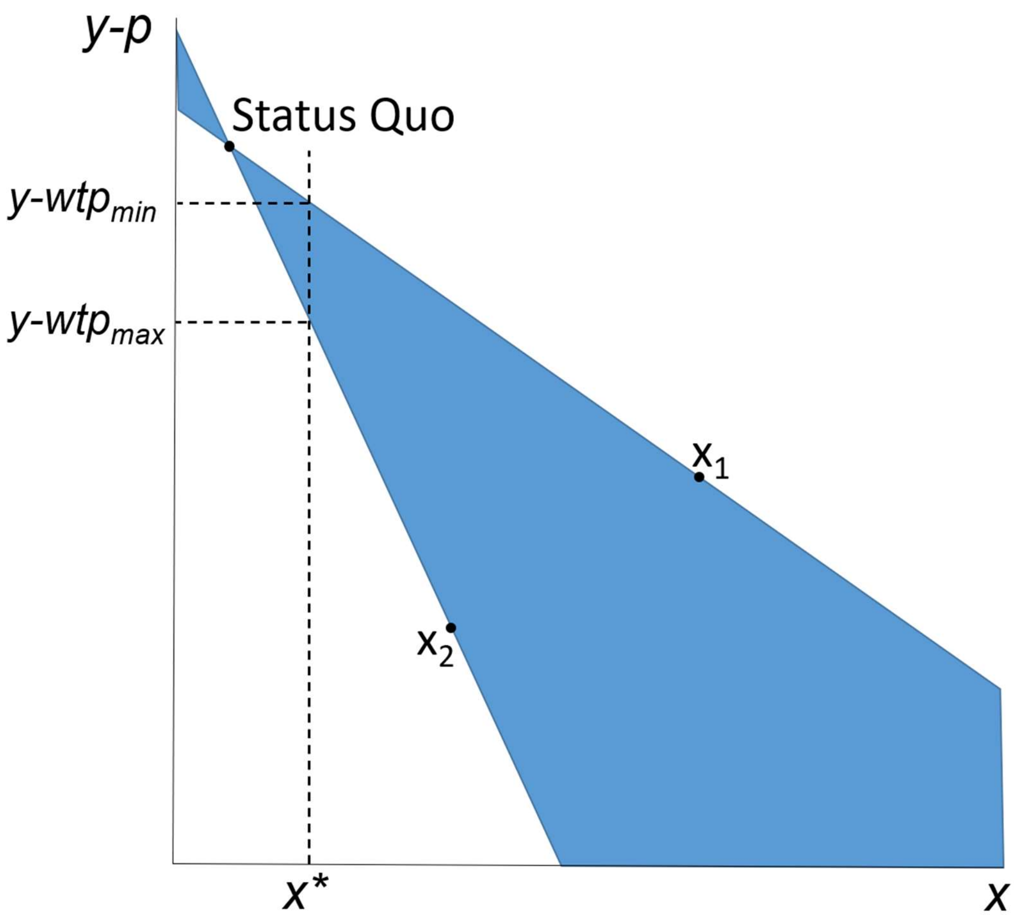
In this example, we observe $\{p_1, x_1\} \succ_R \text{SQ} \succ_R \{p_2, x_2\}$. The shaded area depicts the set $\{\boldsymbol{\beta}_i\}$ of linear taste parameters consistent with these data.

Figure 2. **Asymptotic Point identification of the Linear RUM with Random Coefficients**



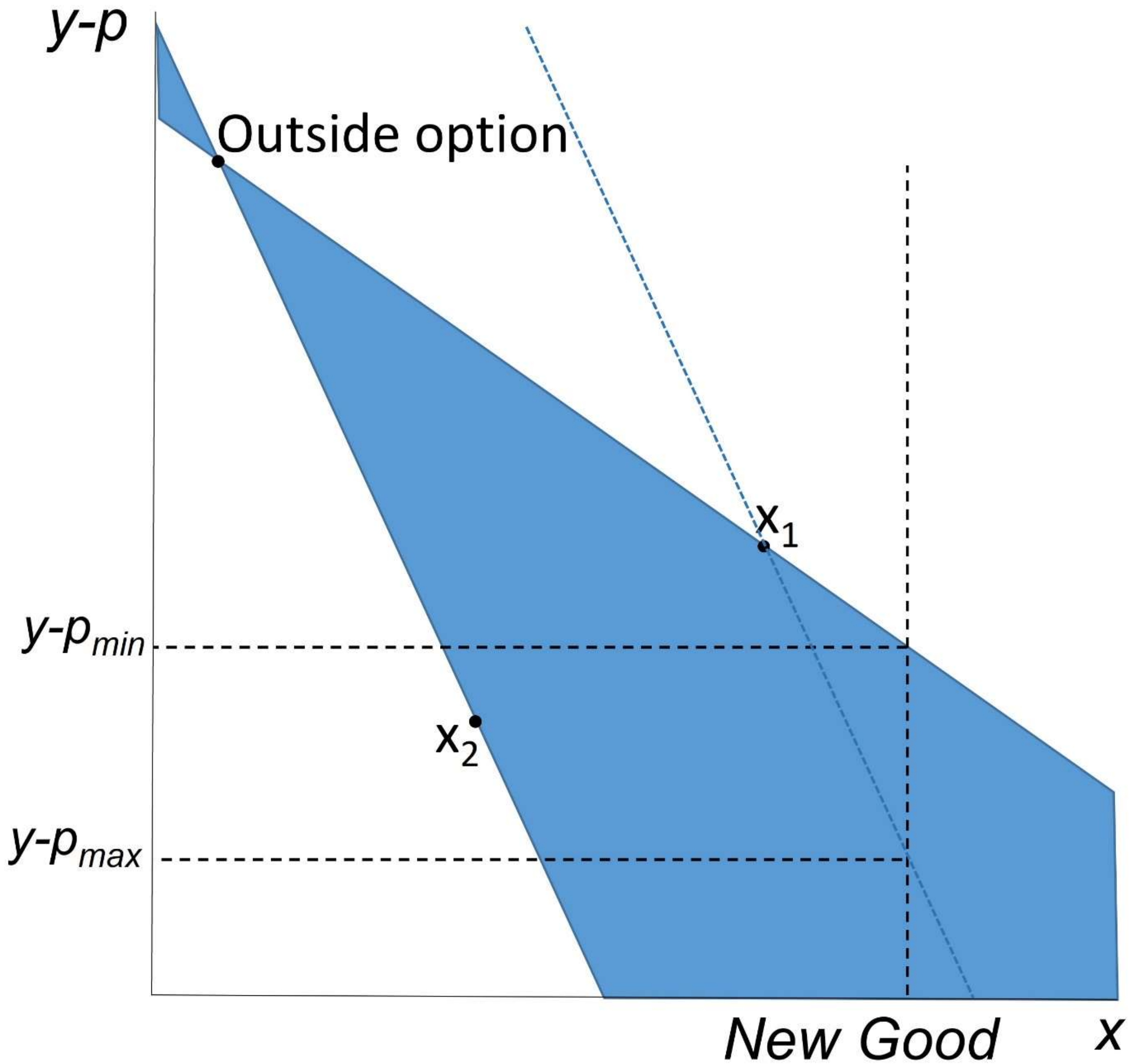
In this example, we observe $\{p_1, x_1\} \succ_R \{p_3, x_3\} \succ_R$ SQ $\succ_R \{p_4, x_4\} \succ_R \{p_2, x_2\}$. The dark shaded area depicts the set $\{\boldsymbol{\beta}_i\}$ consistent with these data, which has shrunk from the light shaded area of Fig. 1.

Figure 3. Bounding Willingness to Pay Relative to the Status Quo



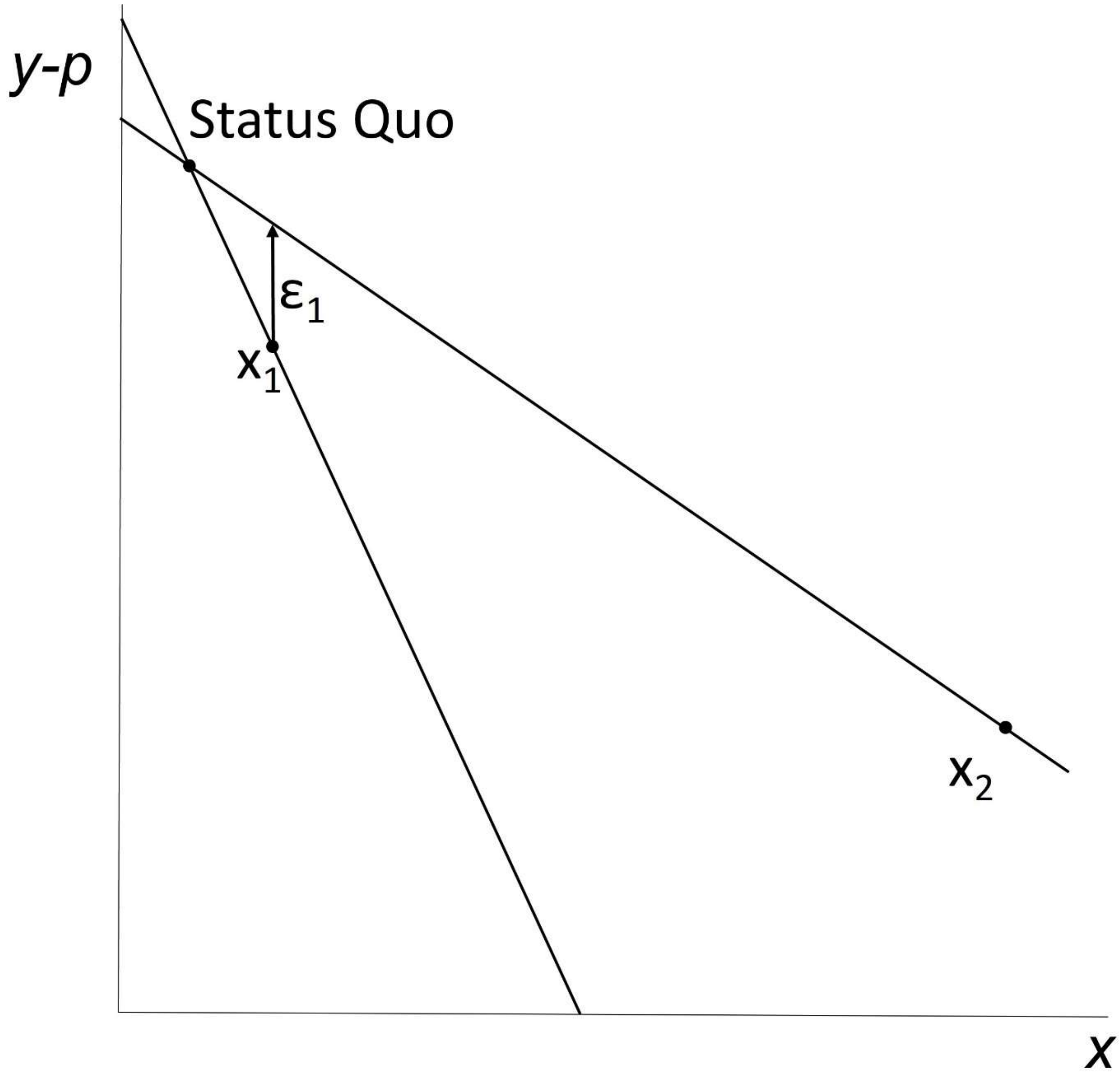
The figure illustrates the set of plausible WTP values for moving from the status quo to a bundle $\mathbf{x}^*$.

Figure 4. Bounding the Highest Price for a New Product



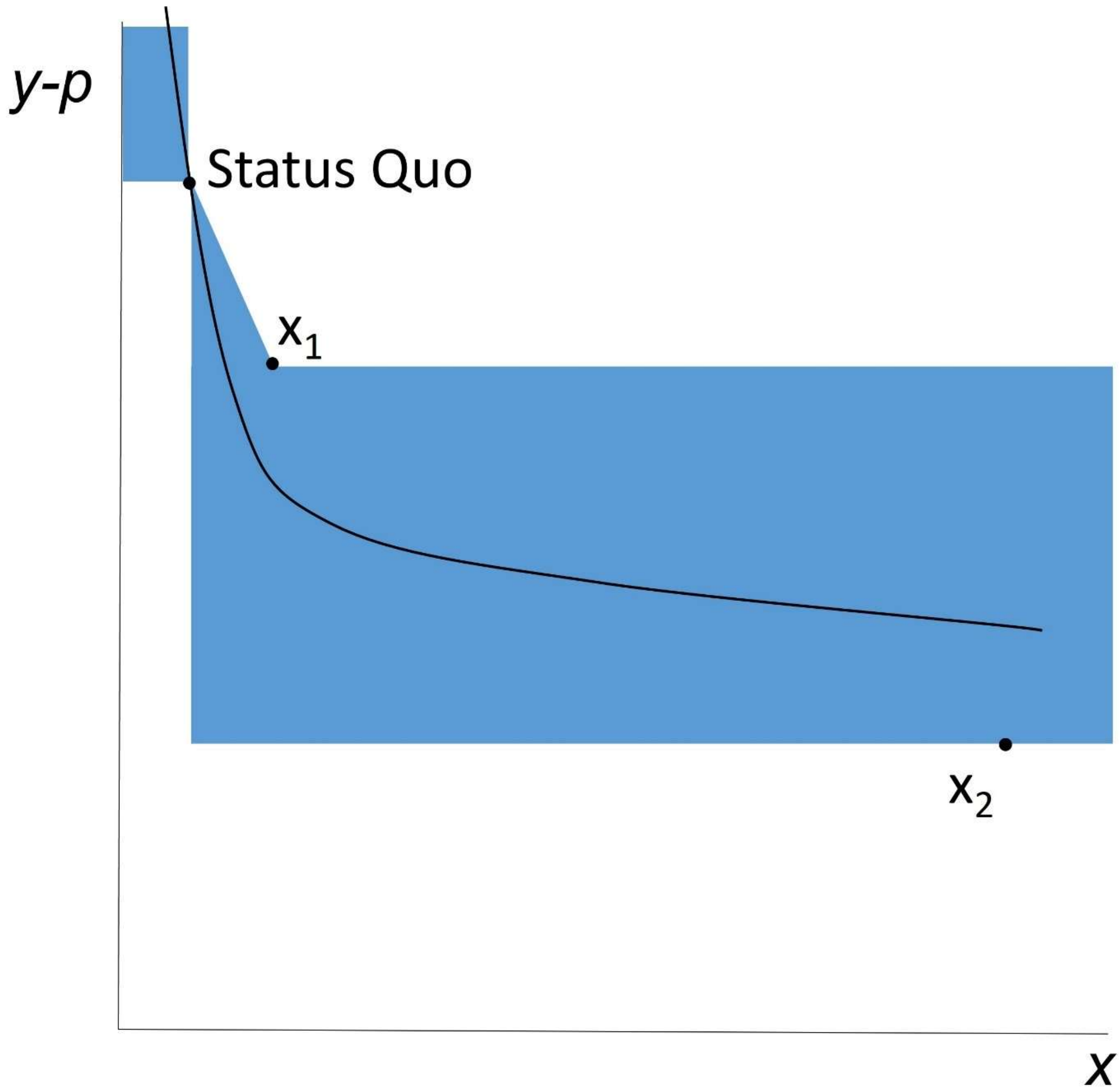
The figure illustrates the plausible range of the highest price for an unobserved new good when, in this example, we observe that $\mathbf{x}_1$ is currently the preferred good and the outside option is preferred to $\mathbf{x}_2$.

Figure 5. Rationalizing the Data with Idiosyncratic Errors



In this example, we observe $\{p_1, x_1\} \succ_R SQ \succ_R \{p_2, x_2\}$, which is not linearly rationalizable. The vertical distance ε_1 represents the minimum adjustment to p_1 that would rationalize the data.

Figure 6. Rationalizable but not Linearly Rationalizable



This figure repeats the example of Fig. 5. The blue shaded area shows the permissible area for indifference curves going through the status quo.

Table 1. Results imposing Homogeneous Preferences

	Multinomial Logit	Linear GARP
Mean WTP without Status Quo	\$52.04 (\$40.32 - \$63.76)	\$25.31 (\$21.31 - \$29.30)
Mean WTP with Status Quo	\$20.55 (\$6.53 - \$34.57)	\$40.12 (\$36.86 - \$43.38)
Average $\lvert\varepsilon\rvert$	\$146.04	\$11.83
Pct ε=0	0%	80.3%
Pct of hholds $\Sigma\lvert\varepsilon\rvert$=0	0%	6.7%
Pct Correctly Predicted, Pure Characteristics	50.1%	50.7%
Pct Correctly Predicted, Conditioning on F(ε)	41.7%	49.4%

The numbers in parentheses represent 95% confidence intervals. All standard errors clustered by household. For Linear GARP, the standard errors are based on a clustered bootstrap with 500 draws.

Table 2. Results with Heterogeneous Preferences

	Mixed Logit	Linear GARP		Non-Parametric GARP (without status quo)		Non-Parametric GARP (with status quo)	
		Min	Max	Min	Max	Min	Max
Mean WTP without Status Quo	\$49.07 (34.56 - 63.58)	\$36.48 (33.33 - 39.62)	\$259.40 (249.36 - 269.43)	\$44.35 (40.98 - 47.73)	\$210.37 (199.50 - 221.23)	N/A	N/A
Mean WTP with Status Quo	\$77.40 (55.08 - 99.71)	-\$129.90 (-143.34 - -116.46)	\$189.95 (178.75 - 201.15)	N/A	N/A	-\$281.74 (-295.83 – -267.66)	\$223.61 (212.68 – 234.54)
Avg $\lvert\varepsilon\rvert$	\$74.96	\$2.00	\$2.00	\$0.46	\$0.47	\$0.36	\$0.37
Pct ε=0	0%	97.2%	97.2%	97.6%	97.7%	98.4%	98.4%
Pct of hholds $\Sigma\lvert\varepsilon\rvert = 0$	0%	74.0%	73.6%	88.0%	88.0%	92.4%	92.4%
Pct Correctly Predicted, Pure Characteristics	N/A	92.8%	92.8%	95.4%	95.4%	97.0%	97.0%
Pct Correctly Predicted, Conditioning on $F(\varepsilon)$	38.5%	82.2%	81.3%	92.7%	92.4%	95.2%	95.0%

The numbers in parentheses represent 95% confidence intervals. All standard errors clustered by household.

Appendix: Results from Simulations

To gauge the weak identification of the GARP-based approach, I employ simulations and estimate the rate at which the identified bounds on WTP converge to the known true value.

Following the premise of Theorem 1, I assume that utility takes a linear form and that households choices are rationalizable. Specifically, I assume they have tastes for ecosystem improvements comparable to those estimated in Section 5, with tastes for improvements to streams, forests, and bird populations close to those estimated in the simplest model of that section, but slightly adjusted so that the true WTP for the policy in the Southern Adirondacks is \$30. (As discussed in Section 5, the policy involves an improvement in 15,000 streams, 250,000 acres of forest, and 15 pp of bird populations.)

I randomly draw choice alternatives with these three characteristics, on a uniform distribution between 0 and 2x the policy-level improvements. In one version of the simulations, I also randomly draw prices on a uniform distribution between zero and \$180. I then create choice occasions with three choice alternatives each, two of these draws plus a status quo option of no change (at no cost). Next, using the true utility parameters, I compute which alternative the households would choose at each choice occasion. This constitutes all the data for the simulations. Next, I estimate the model with the following total numbers of choice occasions: {5, 10, 15, …, 50, 60, 70, …, 150, 200, 250, …, 500, 600, 700, …, 1000, 2000, 3000, …, 10000}. Last, I repeat this process 100 times.

One potential criticism of this simulation is that the prices are drawn at random and are even independent of the other attributes. In an experimental setting, such as that of Section 5, we would normally compute a better guess at the price, e.g. with higher-quality alternatives being assigned higher prices. Too, in a market setting we would expect the equilibrium to have converged on prices that are closer to WTP. Accordingly, in a second version of the simulations, at each addition of choice occasions to the data, I use the previous iteration's estimates to take a guess at the WTP for the new alternatives coming in, confronting the households with those prices in their new choice sets.

Figure A1 shows the results. The figure in the upper-left panel summarizes the upper and lower bounds across simulations. The inner (solid) lines show the median upper and lower bounds respectively, across simulations. The outer (dashed) lines show the 95th percentile of all upper bounds across simulations and the 5th percentile of all lower bounds, at each observed number of choice occasions. (Note that each data point in this figure could come from any simulation.) As shown in the figure, the spreads are very wide at first and do converge rapidly, but improvements begin to stall. The probability of drawing the needed information becomes very low. The upper right panel shows a similar figure but with "smart" prices (i.e. updated based on previous rounds within the iteration). We now get much more rapid convergence, with, for all practical purposes, point identification at about 50 choice occasions.

The lower panels show the respective results for one representative simulation.[17] For the totally random sampled prices, we can see flat portions in the curves, where the new data does not help at all. Moreover, we still have not reached perfect point identification at 1,000 choice occasions. In contrast, in the "smart" sample, there is steady convergence, with point identification at about 50 choice occasions.

Table A1 presents the simulation results in tabular form. Each row of the table is a given number of choice occasions. The columns represent the 5th, 50th, and 95th percentiles (across simulations) in the gap between the estimated max and min of WTP. With only five choice occasions, the median range in the identified set of WTP is \$52, but it can be over \$100 in some simulations. By 50 choice occasions, the range falls to a typical level of \$9 with random sampling, which may be as good as many applications in the literature today when accounting for model uncertainty. However it falls to \$0.04 with "smart" sampling. From there, it quickly falls to zero with "smart" sampling, but with random sampling it is still \$0.04 with as many as 10,000 choice occasions.

The simulations illustrate weak convergence in principle but also highlight the practical importance of optimal sampling.

[17] To pick a representative example, I first computed the rate of convergence for each simulation. Keeping the middle quintile, I then computed, for each remaining simulation, the average gap between the upper and lower bounds on estimated WTP at each number of choice occasions, took the average over the number of choice occasions, and found the simulation closest to that average.

Figure A1. Convergence to point identification in simulations.

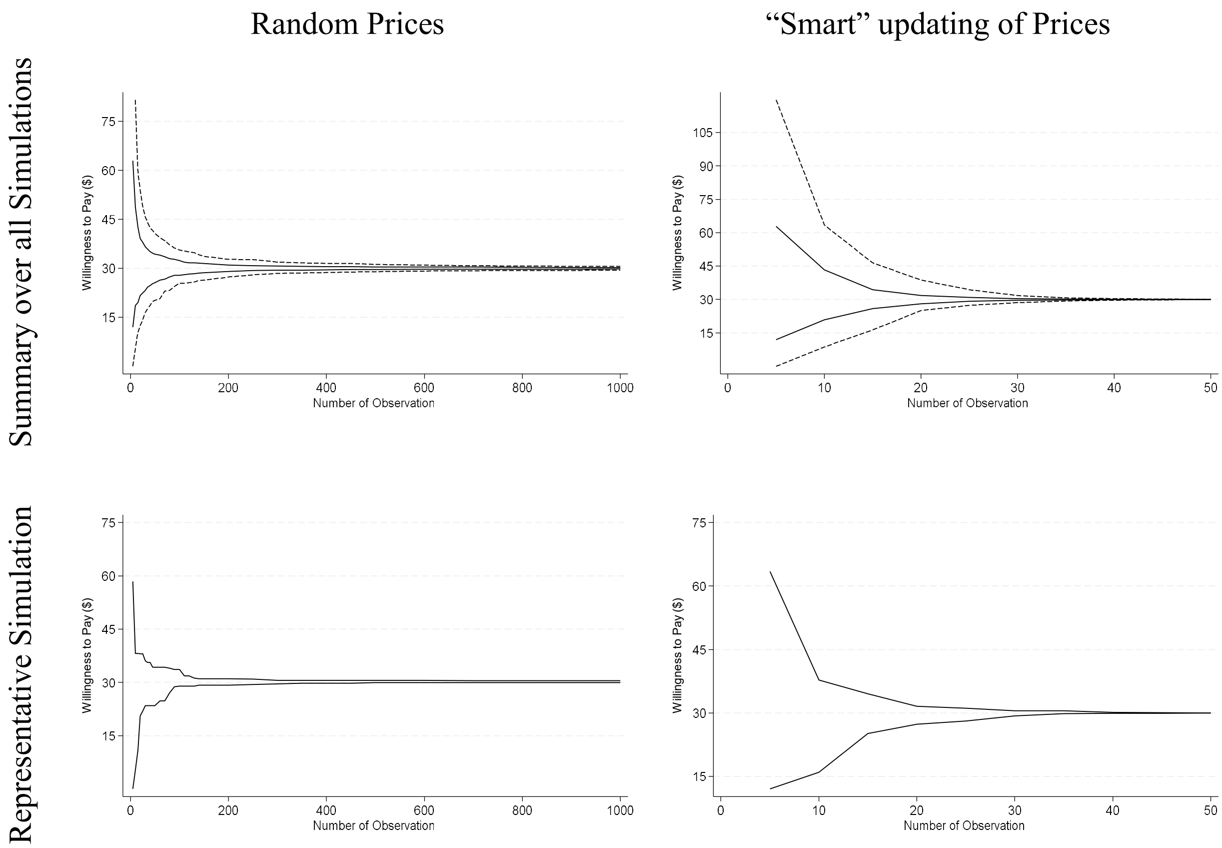


Table A1. Convergence to point identification in simulations.

# choice occasions	Random sampling of prices: 5th percentile	Median	95th percentile	"Smart" sampling of prices: 5th percentile	Median	95th percentile
5	$19.21	$52.28	$101.88	$19.21	$52.28	$101.88
10	$14.78	$32.70	$65.18	$8.58	$21.35	$50.94
25	$7.40	$16.19	$30.43	$0.69	$1.88	$5.88
50	$3.95	$9.05	$16.88	$0.01	$0.04	$0.13
100	$1.84	$4.74	$9.63	$0	$0	$0
200	$1.00	$2.26	$4.58	$0	$0	$0
300	$0.78	$1.46	$3.02	$0	$0	$0
400	$0.50	$1.08	$2.20	$0	$0	$0
500	$0.36	$0.87	$1.67	$0	$0	$0
1000	$0.17	$0.45	$0.90	$0	$0	$0
2000	$0.07	$0.22	$0.53	$0	$0	$0
3000	$0.06	$0.16	$0.33	$0	$0	$0
4000	$0.05	$0.11	$0.26	$0	$0	$0
5000	$0.03	$0.08	$0.20	$0	$0	$0
10,000	$0.02	$0.04	$0.12	$0	$0	$0

The table shows the gap between the upper and lower bounds on estimated WTP, by number of choice occasions included, at three percentiles across the distribution of simulations.