Revealed Preference Analysis of Characteristics in Discrete Choice

Roy Allen and John Rehbeck*

rhallen@ucsd.edu jrehbeck@ucsd.edu

Department of Economics

University of California-San Diego

Last Updated: November 1, 2016

Abstract

This paper studies a descriptive model of stochastic choice that includes information about observable characteristics. The model can describe behavioral phenomena observed in datasets while retaining properties similar to the standard consumer demand problem. Necessary and sufficient conditions for the model to describe a dataset are formalized using a system of linear inequalities. We perform an empirical analysis and find that the model can describe individual stated preference data on flight choice from Louviere et al. (2013). After performing a power correction, a model with linear utility over characteristics often provides the most powerful description of individual datasets.

*We thank Zachary Brieg, Richard T. Carson, Christopher P. Chambers, Andrew Ellis, Federico Echenique, Simone Galperti, Jordan J. Louviere, Mark Machina, Anthony Marley, Seung-Keun Martinez, Joel Sobel, and John Quah for helpful comments and advice. Any remaining errors are our own.

1 Introduction

Individuals often confront discrete choice problems. For example, an individual may choose a health care provider offered by an employer, a lottery offered by an experimenter, or a flight to take a vacation. For these examples, the (indexed) alternatives generally differ along observable dimensions. Health care providers differ in number of specialists, number of locations, and copays. Lotteries differ in probabilities and prizes. Flights differ in travel times, prices, and airlines. A *characteristic* is an observable dimension along which an indexed alternative may vary. In this paper, we develop a model of stochastic choice that explicitly incorporates information from characteristics. This approach allows us to perform an empirical analysis on existing individual discrete choice datasets.

We model discrete choice as stochastic since individuals often make different choices when they face the same decision problem.[1] To understand why individual choice appears stochastic, a large theoretical literature has emerged.[2] Much of this literature takes an axiomatic approach to formalize models of stochastic choice, but the axioms are often tenuous or refuted by data. Rather than explain why choice is stochastic, we aim to describe choices individuals make. To accomplish this goal, we present an "as if" model of stochastic choice from costly attention that nests existing models. The model nests approaches accounting for rational inattention (Caplin and Dean, 2015), agents uncertain of their utility (Fudenberg et al., 2015), and unobservable characteristics (Manski, 1977). Moreover, the model can describe behavioral effects seen in data.

Since we take a descriptive approach, we use characteristics recorded in discrete choice datasets. This differs from many theoretical models that are silent about characteristics. This approach is valuable since characteristics are often used in laboratory and thought experiments to generate counterexamples to stochastic choice models. By explicitly incorporating characteristics, the model clarifies often implicit similarity assumptions on indexed alternatives, produces predictions when characteristics vary, and ensures researchers are using the same observables. The model also allows us to examine the descriptive power of different model specifications.

[1] Individuals are robustly shown to choose different lotteries from the same decision problem experimentally. For example, the works of Mosteller and Nogee (1951), Tversky (1969), and Agranov and Ortoleva (Forthcoming) have documented this behavior. In fact, Agranov and Ortoleva (Forthcoming) observes some individuals pay to randomize their choices.

[2] Classical examples include Thurstone (1927), Luce (1959), Block and Marschak (1960), and Machina (1985). Recent work includes Gul et al. (2014), Manzini and Mariotti (2014), Fudenberg et al. (2015), and Brady and Rehbeck (2016) among many others.

This paper proposes and empirically examines a model of stochastic choice generated by costly attention that explicitly incorporates characteristics. Choice probabilities from the model are interpreted "as if" the individual is maximizing a non-expected utility function. In particular, choice probabilities are the unique maximizers of a utility function with an expected utility component that depends on characteristics less a costly attention function that is independent of characteristics. When the costly attention function is strictly convex, we refer to the model as a *strict perturbed utility model of stochastic discrete choice* or strict PUM. The model is formalized by a system of linear inequalities whose feasibility is necessary and sufficient for a strict PUM to describe a set of observed choice probabilities (Theorem 1). The inequalities prevent the existence of "utility pumps". We use these inequalities to examine when individual choice data is described by different specifications of strict PUMs.

For a descriptive model of stochastic choice, it is advantageous that the model nests approaches already deemed useful in applications. Additive random utility models are a common class of models used in applications.[3] We show that perturbed utility models nest additive random utility models (Section 2.1).[4] Thus, if the data are inconsistent with a perturbed utility model, then an additive random utility model with characteristics is also inconsistent.

Strict PUMs can also describe behavior often seen in data but ruled out in many models. For example, behavior consistent with an attraction effect can be generated by strict PUMs. The attraction effect is often documented adding an alternative to a binary choice set. Attraction effects state that if the added alternative is intuitively dominated by exactly one alternative from the binary choice set, then the probability of choosing the dominating alternative increases. Intuitively, adding a dominated alternative draws attention to the dominating alternative, which can cause the dominating alternative to be chosen more often. The attraction effect violates a regularity condition that is imposed by all random utility models.[5] Huber et al. (1982) first documented the attraction effect, while other studies clarify conditions when the effect occurs and document robustness of the effect (Doyle et al., 1999; Ratneshwar et al., 1987; Heath and Chatterjee, 1995). Additional behavioral effects are discussed in Section 2.3.

[3]Additive random utility models include the logit model. See McFadden (1974), Boskin (1974), and Radner and Miller (1970) for early applications of additive random utility models for commute choice, labor decisions, and college choice.

[4]We do not impose strict convexity for this result for technical reasons described in Section 2.1.

[5]Regularity states that the probability an alternative is chosen weakly decreases as additional alternatives are added to a choice set.

In addition to examining strict PUMs theoretically, we perform an empirical analysis to examine when a strict PUM can generate observed individual choice datasets. The empirical analysis uses individual stated preference data on flight choice from Louviere et al. (2013). The analysis provides answers to the following questions: "Can strict PUMs describe individual choice datasets?", "Are there common properties of specifications that best describe the data?", "Which characteristics best describe individual datasets?". For data from Louviere et al. (2013), we find: (1) Strict PUMs often can describe individual choices, (2) Strict PUMs with linear utility over characteristics often provide the most powerful description of datasets, (3) Price of flight is an important characteristic, but descriptive power often improves by explicitly modeling additional characteristics.

Since strict PUMs with linear utility over characteristics often provide powerful descriptions of datasets, we check if there are utility parameters with intuitive monotonicity properties that can describe individual choice datasets. When imposing monotonicity constraints, we find that strict PUMs are still able to describe many individual choice datasets. Lastly, the descriptive power of a strict PUM with linear utility over price relatively improves as the number of alternatives increases. This provides some evidence that individuals may pay attention to fewer characteristics when making decisions from discrete choice problems with more alternatives.

1.1 Relation to Literature

Thus far, we have not discussed how this paper is related to work on revealed preference. We use a revealed preference approach to formalize the refutable aspects of the model using a system of linear inequalities. These inequalities are similar to those used by Afriat (1967) and Varian (1983) for studying the standard consumer demand problem. There has recently been a renewed interest in using a revealed preference approach to formalize models and analyze datasets. For example, the work of Cherchye et al. (2007) gives refutable conditions for when a household's aggregate demand is observed but individual demand is unobserved. Polisson et al. (2015) provide a test of rational behavior on contingent consumption from risky states and provide an empirical analysis using experimental data on portfolio choice.

While there is a large literature on stochastic choice, the interpretation of stochastic choice generated "as if" it were generated from deterministic choice of lotteries is gaining renewed interest. Recall that a strict PUM models an individual with preferences over lotteries represented by an expected utility component less a costly attention function. We show in Section 2.2 this approach is essentially a first order approximation of utility including observables. Machina (1985) is one of the earliest to consider stochastic choice generated from deterministic preferences over lotteries. Recent work that examines stochastic choice using this framework is Swait and Marley (2013), Fudenberg et al. (2015), and Cerreia-Vioglio et al. (2015). This interpretation is desirable since it allows stochastic choice problems to be treated as standard maximization problems. Moreover, choice probabilities behave like demand functions from the standard consumer problem.

A convenient functional form of preferences over lotteries considers a utility function with an expected utility component less a costly attention function. In the setting of preferences over lotteries, the costly attention function can be interpreted as a cost to ensure a desired object is chosen. The idea of using costly attention functions has seen widespread use in information processing environments. For example, Sims (2003) uses the Shannon entropy function to model limited information flows for consumption-labor decisions over time. More recently, Matejka and McKay (2014) study optimal information acquisition when an individual has a prior distribution and entropy costs of information acquisition. Caplin and Dean (2015) suggest the cost function may be of unknown structure and provide revealed preference conditions for choosing information structures without specifying structure on the cost function. Similar to Caplin and Dean (2015) we study general costly attention functions, but abstract from the information acquisition process. This allows us to take these ideas to datasets in which their observables are unavailable.

While we interpret the separable nonlinear utility component as costly attention, it has been interpreted many ways. Similar to our interpretation, Mattsson and Weibull (2002) interpret the nonlinear component as costly effort is necessary to ensure an outcome occurs. Fudenberg et al. (2015) interpret the nonlinear component as preference to avoid regret. This interpretation is similar to individuals choosing risky portfolios with mean-variance preferences studied in Markowitz (1952). Alternatively, Swait and Marley (2013) interpret the nonlinear component as a preference for exploration. While there is some disagreement about what the non-expected utility component represents, there is agreement that it encodes behavioral properties.

Our study of strict perturbed utility models in stochastic choice is related to a more general study of perturbation functions. Hofbauer and Sandholm (2002) helped bring attention to nonlinear perturbation functions in their study of potential games. McFadden and Fosgerau (2012) consider a "perturbed consumer" and provide a study of demand systems with different assumptions. In addition, McFadden and Fosgerau (2012) interpret the expected utility component as a perturbation, while the object of welfare interest is the nonlinear component. In contrast, we follow Hofbauer and Sandholm (2002) and Fudenberg et al. (2015) in interpreting the expected utility component as the object of interest for welfare on the margin. Perturbed utility models also appear in a study of demand systems generated from general entropy functions in Fosgerau and de Palma (2015).

This paper is also related to the literature including additional observable information into microeconomic models. Recently, Rubinstein and Salant (2008) argue to include additional observables in revealed preference studies. However, the idea of including additional observables into models has been considered since at least the work of Lancaster (1966). Lancaster (1966) considers an agent choosing consumption bundles where commodities only enter the utility function through the value of characteristics when commodities have fixed conversion rates to characteristics. Blow et al. (2008) use a revealed preference approach to study the model from Lancaster (1966) and find individuals can be described by the model in an empirical analysis. An alternative approach by Gul et al. (2014) considers subjective descriptors of indexed alternatives in a model of stochastic choice. This approach is elegant and provides insight on when alternatives may be considered similar. However, the approach fails to speak about what should be done with observed characteristics in discrete choice models.

Rather than using a strict PUM, one could consider introducing characteristics in a random utility model. For example, McFadden and Richter (1990) and McFadden (2005) provide general revealed preference analyses of stochastic choice data generated by random utility models. A natural way to include characteristics in random utility models is to treat each unique set of characteristics as a distinct alternative and perform the tests developed in these papers. Kitamura and Stoye (2013) follow this approach to develop statistical test of random utility models and find limited statistical evidence against random utility behavior. However, modifying the approach in these papers to test different model specifications seems challenging and is an open question. In contrast, the formalization of strict PUMs provides a simple way to examine different model specifications (see Section 3).

One could alternatively study strict PUMs using a statistical approach. Allen and Rehbeck (2016b) examine identifying perturbed utility models using symmetry properties and test a structured perturbed utility model against a null hypothesis that parameters are consistent with additive random utility. The test rejects the null hypothesis of additive random utility parameters. This provides some evidence for perturbed utility models using the statistical approach. One may also consider adapting results from additive random utility models to strict PUMs. We note that Shi et al. (2015) study the identifying power of cyclic monotonicity as an implication of additive random utility models. For a strict PUM, strict cyclic monotonicity is a necessary and sufficient condition for the model. Because we use similar implications as Shi et al. (2015), one could consider estimating linear utility parameters over characteristics for a strict PUM following their procedure.

The remainder of the paper proceeds as follows. Section 2 defines the individual choice datasets, defines the model, and provides examples. Section 3 provides theoretical results. Section 4 performs an empirical analysis of strict PUMs using data from Louviere et al. (2013). Section 5 contains our concluding remarks.

2 Definitions and Model

We present a model of stochastic choice that arises from an individual choosing an optimal lottery. To operationalize this idea, it is necessary to choose an indexing of alternatives. In general, individual choice datasets are collections of observables and choices. It is up to the researcher to specify what an individual is choosing when performing an economic analysis. The choice of an index will not be innocuous in the model we present. The index specifies the dimension along which costly attention affects the individual. Later in an application, we choose list position as the index of alternatives to captures the intuition that list position imposes costs that lead to stochastic choice. We denote the indexed alternatives by $a = 1, \ldots, A$. For simplicity, we refer to these as alternatives.

Each indexed alternative has additional structure from observables called characteristics.[6] Characteristics can take many forms. For example, characteristics can be numeric such as price or time. A characteristic could be ordered numeric such as safety rating or quantity. Alternatively, a characteristic could be categorical such as brand or color. We assume each alternative a has d_a (finite) characteristics. The set $X_{a,j}$ contains values that the j-th characteristic of alternative a can take. The characteristic values an alternative can take are encoded to lie in a vector space.[7] Vector notation is suppressed and inner products

[6]This is similar to the standard consumer problem where one chooses an index of commodities that are the same regardless of price. One could instead consider a model of consumer behavior where individuals have preferences over the index and price. However, this approach removes the similarity inherent in a commodity. We feel the same logic holds in discrete choice environments with characteristics and that one should be explicit when indexing alternatives.

[7]We think of $\mathcal{X}_a \subseteq \mathbb{R}^{d_a}$. In this case, quantitative variables have a relevant domain and categorical variables can be defined using indicators.

are represented using dot product notation. Thus, $x_a = (x_{a,1}, \ldots, x_{a,d_a}) \in \mathcal{X}_a = \prod_{j=1}^{d_a} X_{a,j}$ is a vector of characteristic values taken by alternative a.

We call a *menu* a collection of characteristic values for each alternative.[8] We denote a menu $x = (x_1, \ldots, x_A) \in \mathcal{X} = \prod_{a=1}^{A} \mathcal{X}_a$. We denote the probability simplex over indexed alternatives as $\Delta = \{p \in \mathbb{R}^A \mid \sum_{a=1}^{A} p_a = 1 \text{ and } p_a \geqslant 0 \text{ for all } a\}$. We consider datasets of observed menus and choice probabilities from the observed menus denoted $\{(x^n, p(x^n))\}_{n=1}^{N}$. For a dataset, menus are distinct, but it is possible for $p(x^r) = p(x^s)$ for $r \neq s$ and $r, s \in \{1, \ldots, N\}$.

We now define a *strict perturbed utility model of discrete choice* (strict PUM). A strict PUM is represented by

$$p^*(x) = (p_1^*(x), \ldots, p_A^*(x)) = \operatorname*{argmax}_{p \in \Delta} \left(\sum_{a=1}^{A} p_a u_a(x_a) - C(p) \right),$$

where $p^*(x)$ is the optimal distribution of choices when the menu is x, the function $u_a : \mathcal{X}_a \to \mathbb{R}$ gives the utility of characteristic values for alternative a, and $C : \Delta \to \mathbb{R}$ is a strictly convex function which perturbs the expected utility. This model differs from that of Fudenberg et al. (2015) by allowing interactions of choice probabilities in the cost function and explicitly modeling characteristics.[9] The $u_a(x_a)$ term can be interpreted as part of the local utility for an alternative with characteristic values x_a. We discuss this interpretation more in Section 2.2.

We interpret the strictly convex function C as costly attention required to ensure an alternative is chosen. For example, an individual may want to choose an alternative that has the highest utility from observables, but has behavioral biases about the alternatives that generates a distribution of choices. In line with this interpretation, the costly attention function only depends on the indexing of alternatives. The nonseparability of costly attention allows choice probabilities to interact. For example, consider when the indexing of alternatives denotes the position in a list. It may be costly to consider a position in the middle of the list, but once the position is considered the nearby objects are considered more often. Nonseparable cost functions allow this behavior, while separable cost functions rule this behavior out. The costly attention function also encodes substitutability and comple-

[8]This definition will agree with the standard definition of a menu if only availability variation is used. See Appendix I.

[9]The formulation of the cost function in Fudenberg et al. (2015) studies when $C(p) = \sum_{a=1}^{A} c(p_a)$ for a fixed $c(\cdot)$ function that is continuously differentiable. See Appendix C for discussion of a model with symmetric and separable costs. See Appendix D for a discussion about imposing differentiability.

mentarity analogous to standard consumer demand. Other interpretations are provided in Section 1.1.

We briefly describe how this approach relates to theoretical models of stochastic choice that use alternative availability. Alternative availability can be accommodated using characteristics. Assign each alternative that can be chosen a single characteristic which takes the values "available" and "unavailable".[10] Using this mapping, we see that our definition of a menu agrees with the standard terminology using only availability. Any dataset that satisfies positivity with only alternative availability is rationalized by a strict PUM as noted in Machina (1985) (see Appendix I for details and a simple proof).

2.1 Connection to Additive Random Utility Models

We desire strict PUMs to nest approaches already found useful describing data. Additive random utility models are often used in prediction exercises and contain the tractable logit model. Therefore, we consider how strict PUMs are related to additive random utility models. The following discussion considers perturbed utility models with costs not necessarily strictly convex since general additive random utility models can have positive probability of utility ties. In this case, optimal choice distributions need not be unique.

Consider an alternative a with latent utility given by $v_a(x_a) = u_a(x_a) + \varepsilon_a$, where $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_A)$ is unobservable to the researcher and does not depend on the menu. In an additive random utility model, the individual observes ε and chooses the object with the largest latent utility. This can be written as

$$p^*(x,\varepsilon) \in \operatorname*{argmax}_{p \in \Delta} \left(\sum_{a=1}^{A} p_a u_a(x_a) - C(p,\varepsilon) \right), \tag{1}$$

where $C(p,\varepsilon) = -\sum_{a=1}^{A} p_a \varepsilon_a$ and $p^*(x,\varepsilon)$ denotes that the choice depends on the realization of ε.

Allen and Rehbeck (2016a) show under weak regularity conditions that

$$p^*(x) = \mathbb{E}[p^*(x,\varepsilon)] \in \operatorname*{argmax}_{p \in \Delta} \left(\sum_{a=1}^{A} p_a u_a(x_a) - C(p) \right), \tag{2}$$

[10]When C is a bounded function, if $u_a(\text{"unavailable"}) = -K$ for K large, then alternative a is never be chosen when unavailable. The domain of $\mathcal{X}$ excludes the case that all alternatives are unavailable.

where expectations are over ε, $C(p) = \max_{\pi \in \Pi : \mathbb{E}[\pi] = p} \mathbb{E}\left[-\sum_{a=1}^{A} \pi_a(\varepsilon)\varepsilon_a\right]$ where Π is the set of all measurable functions from the support of ε to Δ. This shows that PUMs nest behavior generated by additive random utility models.

The aggregation result in fact applies to *all* random utility models when the only characteristics are alternative availability. To show this result, we require $u_a(\cdot)$ terms to take value $-\infty$. The result now follows letting $u_a(\text{“available”}) = 0$ and $u_a(\text{“unavailable”}) = -\infty$. The distribution of ε induces a distribution over linear orders of the alternatives $\{1, \ldots, A\}$. Since ε does not depend on the menu (which only includes variation in availability), this coincides with random utility models from Block and Marschak (1960). However, there are perturbed utility models that are not additive random utility models.[11] Example 2 gives behavior that is inconsistent with additive random utility models.

2.2 Motivation of Separability and Local Approach

Strict PUMs can also be interpreted as a first order approximation to describing behavior. First, consider a general non-expected utility function as in Machina (1985) given by

$$p^*(x) = \operatorname*{argmax}_{p \in \Delta} V(x, p)$$

where $V(x, p)$ is a continuously differentiable function in p. Similar to the analysis of Machina (1982), we take a first order Taylor expansion around the optimal choice distribution and find that

$$V(x,p) - V(x, p^*(x)) = \sum_{a=1}^{A} V_a(x, p^*(x))[p_a - p_a^*(x)] + o\left(||p - p^*(x)||\right) < 0.^{12}$$

Examining the comparison near the optimum, the $o\left(||p - p^*(x)||\right)$ term is small so

$$\sum_{a=1}^{A} V_a(x, p^*(x)) p_a < \sum_{a=1}^{A} V_a(x, p^*(x)) p_a^*(x).$$

Therefore, $V_a(x, p^*(x))$ represents the local utility of alternative a. However, if values recorded in menu x are informative about the true utility value of an alternative, then

[11] Hofbauer and Sandholm (2002) show under a stronger set of regularity conditions that the additive random utility model of (1) implies a perturbed utility representation. They also provide an example of a perturbed utility model which cannot be represented by an additive random utility model.

[12] The term $||\cdot||$ is the standard Euclidean norm.

it is natural to make additional assumptions. First, it is natural to assume that x_a only provides information about the local utility of alternative a. In this case, local utility is represented by

$$V_a(x, p^*(x)) = V_a(x_a, p^*(x)).$$

A second natural assumption imposes that characteristic values impact local utility independently of choice distributions. This assumption says if observables x_a occur in another problem, then the local utility is the same. Now we can decompose the local utility into a component that depends on the observables x_a and a component that depends on the lottery choice, so

$$V_a(x, p^*(x)) = u_a(x_a) + V_a(p^*(x)).$$

Since other terms are of smaller order locally, a first order approach suggests that characteristic values only effect local utility. However, these assumptions bring us to a utility function

$$V(x, p) = \sum_{a=1}^{A} u_a(x_a) p_a + \tilde{V}(p).$$

A sufficient condition for $p^*(x)$ to be a unique maximizer of $V(x, p)$ is that $\tilde{V}(p)$ is strictly concave. This final assumption yields a strict PUM by setting $C(p) = -\tilde{V}(p)$. From this procedure, it is clear that C captures behavioral properties associated with the indexing of alternatives.

Note that $u_a(x_a)$ is only part of the local utility for an alternative. Therefore, $u_a(x_a)$ does not give information about absolute welfare for alternative a with characteristics x_a. However, $u_a(x_a)$ provides information about marginal welfare and may be of policy interest. For example, an individual may have a predisposition to avoid buses when traveling that is encoded through the costly attention function. However, a policy maker could consider changing the attractiveness of buses through observables to change individual behavior at the margin.

2.3 Behavioral Effects

This section details behavior potentially seen in data that can be described by the model. Two common behavioral effects seen in data are the attraction and compromise effect. Details of the attraction effect are provided in the introduction. The commonly observed compromise effect states that if an alternative is added to a binary choice set and is "between" the two alternatives, then it is chosen more often. For more details about the attraction and compromise effect, we refer the reader to Simonson (1989). While these are robustly documented effects, they need not always occur. Thus, a descriptive model of stochastic choice should allow either type of behavior. We provide an example decision problem below where either type of behavior could occur.

Consider an employee choosing between health care provider A, B, and C offered by their employer. Health care providers only differ in the number of locations and specialists. In this example, alternatives are indexed by health care providers A, B, and C. The characteristics for each alternative are the number of locations and specialists. From this description, the indexing of alternatives specifies a priori what is chosen by the individual and supposes similarity for different values of characteristics. The costly attention function only depends on the health care provider. Costly attention may capture prior information that is unobserved such as reputation, advertising effects, and word of mouth. An individual chooses between providers A, B, and C, but an employer may consider changing the coverage at the same cost to the individual. Let C_ℓ, C_m, and C_h be different potential coverages offered by provider C with low, medium, and high numbers of specialists respectively. The different coverages are pictured in Figure 1. Rather than adding alternatives to a choice set, we consider attraction and compromise effects using continuous changes in characteristic values.

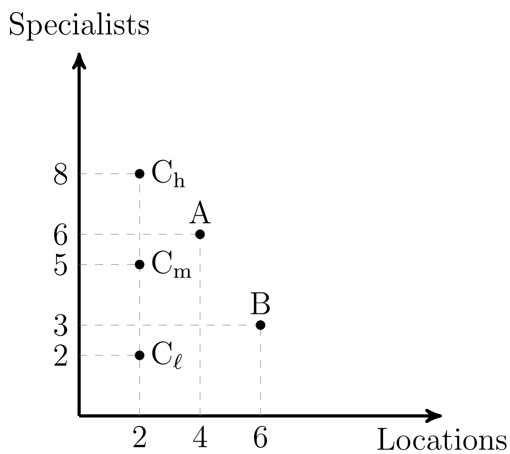

Figure 1: Insurance Provider Decision Problems

If the coverage of provider C changes from C_ℓ to C_m or C_m to C_h, there are plausible reasons to expect the choice probability of provider A to increase or decrease. The direction depends on which behavioral effect dominates. First, consider the choice of provider when the coverages are given by A,B, C_ℓ in Figure 1. Since C_ℓ is strictly dominated by both providers in characteristics, we may expect that provider C is chosen with low or zero probability. If coverage C_ℓ is replaced with the coverage C_m, provider C may become more attractive and be chosen more often. In addition, provider A dominates C_m so A may be chosen more often, in line with an attraction effect. However, provider A may be chosen less since provider C is a better competitor. If the number of specialists for provider C increases further to coverage C_h, then we again expect provider C to be chosen more often. Provider A could be chosen less often if provider C is viewed as a competitor. However, provider A could be viewed as a compromise and chosen more often. Using conditions from Section 3, Example 4 shows that as long as there is a direct effect on the choice probability of the alternative with varying characteristic values, either type of behavior can be described.

2.4 Examples

Although this paper focuses on a revealed preference analysis of the model, we provide some examples of cost functions that highlight the versatility of the model. These examples place a parametric structure on the cost function that allows for analysis of comparative statics and prediction.

Example 1 (Choice from Lists)**.** *Let a be the order of results displayed in a list which has A options. Let $\eta > 0$ and consider the perturbation function*

$$C(p) = \eta \sum_{a=1}^{A} p_a \ln(p_a) - \sum_{a=1}^{A} \gamma_a p_a,$$

where $\gamma_a \in \mathbb{R}$. Assume that each position has the same set of characteristics and $u_a(x_a) = u(x_a)$ for each a. This implies that utility received from characteristic values does not depend on list position. Using the above assumptions,

$$p_a^*(x) = \frac{e^{(u(x_a)+\gamma_a)/\eta}}{\sum_{b=1}^{A} e^{(u(x_b)+\gamma_b)/\eta}}.$$

A high value of γ_a acts as a "boost" to choosing position a. When $\gamma_a = 0$ for each a, this is the logit formula. The modified logit formula is due to Mattsson and Weibull (2002).[13] *Following the interpretation of Matejka and McKay (2014), γ_a may be non-zero to reflect prior beliefs about quality of items in position a.*

Example 2 (Choice from Lists with Linked Position Effects)**.** *Consider choice from an*

[13]Mattsson and Weibull (2002) consider the relative entropy cost $C(p) = \eta \sum_{a=1}^{A} p_a \ln(p_a/q_a)$, where $q = (q_1, \ldots, q_A)$ is a reference measure. This formulation is made equivalent to the above cost function by rewriting $\gamma_a = \eta \ln(q_a)$.

ordered list as in Example 1. Assume $A > 2$. Now suppose that if the last position (A) is chosen with high probability, then position $A-1$ is also chosen with high probability. One perturbation function consistent with this behavior is (with $\eta, \gamma > 0$)

$$C(p) = \eta \sum_{a=1}^{A} p_a \ln(p_a) + \gamma \max\{p_A - p_{A-1}, 0\}.$$

Suppose at menu $x = (x_1, \ldots, x_A)$, $U(x) = (u(x_1), \ldots, u(x_A))$ equals the zero vector. Since $\gamma > 0$ we obtain $p_a^*(x) = 1/A$ for each a. Suppose characteristic values change only for position A so $\tilde{x} = (x_1, \ldots, x_{A-1}, \tilde{x}_A)$ with $u(\tilde{x}_A) > 0$. For fixed η, there exists γ sufficiently large such that $p_A^*(\tilde{x}) > p_A^*(x)$ and $p_{A-1}^*(\tilde{x}) > p_{A-1}^*(x)$.[14] *Thus, there can be position effects that are linked.*

Example 3 (Sparse Stochastic Choice)**.** *In data, one often finds only a few alternatives chosen with positive probability. However, many models of stochastic choice including the logit model assume each alternative is chosen with positive probability. The following perturbed utility function incorporates a penalty so only a few alternatives have positive probabilities. For $\eta > 0$ the function*

$$C(p) = p_1^2 + \sum_{a=2}^{A} \left(p_a^2 + \eta p_a\right)$$

will induce a subset of alternatives to be chosen with zero probability when η is large enough.

3 Revealed Preference Analysis

We take a revealed preference approach to check when a dataset $\{(x^n, p(x^n))\}_{n=1}^N$ can be described by a strict PUM.[15] Conditions developed in this section allow us to give example datasets that cannot be described by a strict PUM. Any proofs not in the main text are in Appendix A.

Definition 1 (Rationalization of Strict PUM)**.** *The dataset $\{(x^n, p(x^n))\}_{n=1}^N$ is rationalized by a strict perturbed utility model if there exist $u_a : \mathcal{X}_a \to \mathbb{R}$ for all $a \in \{1, \ldots, A\}$ and a*

[14] This is an example of stochastic complementarity between objects in positions A and $A-1$ using the definition from Allen and Rehbeck (2016a).

[15] A perturbed utility model with a weakly convex perturbation can rationalize any dataset. Details are in Appendix A.

strictly convex $C : \Delta \to \mathbb{R}$ *such that*

$$p(x^n) = \operatorname*{argmax}_{p \in \Delta} \sum_{a=1}^{A} p_a u_a(x_a^n) - C(p)$$

for all $n \in \{1, \dots, N\}$.

First, we present a lemma on properties of optimizers from perturbed utility models. The result implies that utilities are monotone in probabilities and only relies on the existence of a unique maximizer. Let $U(x) = (u_1(x_1), \dots, u_A(x_A))$ be the vector of utilities generated by menu x. Let $g : \Delta \to \mathbb{R}$ be a function on the simplex.

Lemma 1. *Consider the dataset* $\{(x^n, p(x^n))\}_{n=1}^N$ *and suppose that* $p(x^n) = \operatorname{argmax}_{p \in \Delta} \left(\sum_{a=1}^{A} p_a u_a(x_a^n) - g(p) \right)$ *is a singleton . Then for any* $x, \tilde{x} \in \{x^n\}_{n=1}^N$ *such that* $p(x) \neq p(\tilde{x})$,

$$(p(x) - p(\tilde{x})) \cdot U(\tilde{x}) < g(p(x)) - g(p(\tilde{x})) < (p(x) - p(\tilde{x})) \cdot U(x)$$

Proof. Uniqueness of the argmax set requires that

$$\begin{aligned} p(\tilde{x}) \cdot U(x) - g(p(\tilde{x})) &< p(x) \cdot U(x) - g(p(x)) \\ p(x) \cdot U(\tilde{x}) - g(p(x)) &< p(\tilde{x}) \cdot U(\tilde{x}) - g(p(\tilde{x})). \end{aligned}$$

The result follows by rearrangement. □

Next, consider taking a sequence of probabilities chosen from different menus. Let $x[m] = (x_1[m], \dots, x_A[m]) \in \{x^n\}_{n=1}^N$ for $m \in \{1, \dots, M\}$ be an element in a sequence of observed menus. Let $U(x[m]) = (u_1(x_1[m]), \dots, u_A(x_A[m])$. Assume at least one pair of choice distributions chosen is not equal and let $x[M+1] = x[1]$. Summing inequalities from Lemma 1 yields

$$\sum_{m=1}^{M} (p(x[m+1]) - p(x[m])) \cdot U(x[m]) < \sum_{m=1}^{M} g(p(x[m+1])) - g(p(x[m])) = 0.$$

If all choice distributions from a sequence are equal, then the above sums both equal zero and provide no restriction on utilities. Rearranging the probabilities and still assuming $x[M+1] = x[1]$, we find that

$$\sum_{m=1}^{M} p(x[m+1]) \cdot U(x[m]) < \sum_{m=1}^{M} p(x[m]) \cdot U(x[m]).$$

This condition states that there does not exist a strict utility pump when comparing the probabilities chosen from a menu to other observed probabilities. In light of the discussion in Section 2.2, we interpret this as being able to find no local utility improvements from a cycle.[16] We show that the no strict utility pump condition is necessary and sufficient for a dataset to be rationalized by a strict PUM.

Theorem 1. *Consider the dataset $\{(x^n, p(x^n))\}_{n=1}^N$. The following are equivalent:*

(i) $\{(x^n, p(x^n))\}_{n=1}^N$ *is rationalized by a strict PUM.*

(ii) There exist utility functions $u_a : \mathcal{X}_a \to \mathbb{R}$ *for all* $a \in \{1, \ldots, A\}$ *and a function* $g : \Delta \to \mathbb{R}$ *such that for all* $n \in \{1, \ldots, N\}$, $p(x^n) = \operatorname{argmax}_{p \in \Delta} \sum_{a=1}^A p_a u_a(x_a^n) - g(p)$.

(iii) There exist numbers $\{u_a^n\}_{n=1}^N$ *for all* $a \in \{1, \ldots, A\}$ *and* $\{g^n\}_{n=1}^N$, *such that for all* $(s, r) \in \{1, \ldots, N\} \times \{1, \ldots, N\}$ *with* $p(x^s) \neq p(x^r)$ *then*

$$\sum_{a=1}^A p_a(x^s) u_a^r - g^s < \sum_{a=1}^A p_a(x^r) u_a^r - g^r,$$

and for all $r, s \in \{1, \ldots, N\}$

$$\begin{aligned} u_a^r &= u_a^s \quad \textit{if} \quad x_a^r = x_a^s \\ g^r &= g^s \quad \textit{if} \quad p(x^r) = p(x^s) \end{aligned}$$

(iv) There exist numbers $\{u_a^n\}_{n=1}^N$ *for all* $a \in \{1, \ldots, A\}$ *such that for all finite sequences* $\{x[m]\}_{m=1}^M$ *where all* $x[m] \in \{x^n\}_{n=1}^N$ *and* $p(x[m]) \neq p(x[m+1])$ *for some* m

$$\sum_{m=1}^M p(x[m+1]) \cdot U[m] < \sum_{m=1}^M p(x[m]) \cdot U[m]$$

where $x[M+1] = x[1]$ *and* $U[m] = (u_1[m], \ldots, u_A[m])$ *where* $u_a[m]$ *is the* u_a^r *term associated to* $x_a[m] = x_a^r$;
and for all $r, s \in \{1, \ldots, N\}$ *and for all* $a \in \{1, \ldots, A\}$

$$u_a^r = u_a^s \quad \textit{if} \quad x_a^r = x_a^s$$

(v) Let $\mathcal{S} = \{(\tilde{x}, x) \in \{x^n\}_{n=1}^N \times \{x^n\}_{n=1}^N \mid p(\tilde{x}) \neq p(x)\}$. *There is no* $\{\pi_{(\tilde{x},x)}\}_{(\tilde{x},x)\in\mathcal{S}}$ *with*

[16] This above inequality is a version of strict cyclic monotonicity as developed in Rockafellar (1970).One early application of cyclic monotonicity is the development of necessary and sufficient conditions to implement action profiles in a quasi-linear environment by Rochet (1987).

$\pi_{(\tilde{x},x)} \geqslant 0$ *and* $\sum_{(\tilde{x},x)\in\mathcal{S}} \pi_{(\tilde{x},x)} = 1$ *such that for all* $\hat{x}_a \in \{\{x_a^n\}_{a=1}^A\}_{n=1}^N$

$$\sum_{\{(\tilde{x},x)\in\mathcal{S}|x_a=\hat{x}_a\}} \pi_{(\tilde{x},x)} p_a(\tilde{x}) = \sum_{\{(\tilde{x},x)\in\mathcal{S}|x_a=\hat{x}_a\}} \pi_{(\tilde{x},x)} p_a(x)$$

and for all $\hat{x} \in \{x^n\}_{n=1}^N$

$$\sum_{\{(\tilde{x},x)\in\mathcal{S}|p(\tilde{x})=p(\hat{x})\}} \pi_{(\tilde{x},x)} = \sum_{\{(\tilde{x},x)\in\mathcal{S}|p(x)=p(\hat{x})\}} \pi_{(\tilde{x},x)}$$

That (i) implies (ii) is clear since a strictly convex perturbation function yields argmax sets that are singletons. That (ii) implies (iii) is an implication of Lemma 1. That (iii) implies (iv) is immediate from the discussion preceeding the theorem.[17] We provide a constructive proof that (iv) implies (i). Condition (iii) is equivalent to (v) by a theorem of the alternative.

We briefly mention some highlights of these conditions. Condition (ii) states that one cannot differentiate strict PUMs from a more general class of PUMs with unique maximizers. The inequalities in (iii) are similar to those from Afriat (1967) to test standard consumer demand where the utility numbers u_a^r play a role similar to prices. The inequalities in (iv) are similar to the no attention improving cycles condition in Caplin and Dean (2015) and the revealed preference test for utility maximization in incomplete markets provided by Green and Srivastava (1986). Condition (v) provides a set of universal conditions that can refute strict PUMs.[18] Moreover, condition (v) has a connection to second stage expected utility maximization that we discuss in Appendix B. We now state a simple corollary of the above result.

Corollary 1. *Consider the dataset* $\{(x^n, p(x^n))\}_{n=1}^N$. *If* $p(x^n) = \hat{p}$ *for all* $n \in \{1, \ldots, N\}$, *then the dataset is rationalized by a strict perturbed utility model.*

The corollary states that at least one pair of distinct choice distributions are needed to refute a strict PUM. The result follows from Theorem 1 since there are no comparisons of different probabilities. A simple rationalization of the dataset is generated by setting $U(x) = 0$ for all $x \in \mathcal{X}$ and letting $C(p) = \sum_{a=1}^A (p_a - \hat{p}_a)^2$. This rationalization predicts choice probabilities that never change for any set of characteristic values.

Theorem 1(i)-(iv) can be used to state additional theorems to rationalize strict PUMs

[17](iii) implies (iv) even if $g^n \neq g^{\tilde{n}}$ when $p(x^n) = p(x^{\tilde{n}})$ so that this holds under slightly weaker assumptions. We prefer the above representation for transparency and since it will eliminate variables in the application.

[18]Condition (v) for strict PUMs is also UNCAF as defined in Chambers et al. (2014).

with various utility functions over characteristics. For example, looking for rationalizations with $u_a(x_a) = \sum_{j=1}^{d_a} u_{a,j}(x_{a,j})$ or $u_a(x_a) = \beta_a \cdot x_a$ both yield systems of linear inequalities on $u_{a,j}(x^n_{a,j})$ or β_a, respectively.[19] If characteristics are the same for all alternatives, the a subscript can be dropped to examine utility over characteristics that is independent of the indexing of alternatives. For the linear model, one can test strict monotonicity of characteristics by imposing additional inequalities on the β_a terms. See Appendix A for a statement of these results.

We now provide an example illustrating some restrictions imposed by strict PUMs.

Example 4. *[Monotonicity Violation] Let $A = 3$ and suppose menus $x = (x_1, x_2, x_3)$ and $\tilde{x} = (x_1, x_2, \tilde{x}_3)$ are observed with $p_1(x) > p_1(\tilde{x})$, $p_2(x) < p_2(\tilde{x})$, and $p_3(x) = p_3(\tilde{x})$. This is inconsistent with a strict PUM. To be consistent, there must exist values $u_3(x_3), u_3(\tilde{x}_3)$ such that,*

$$(p_3(\tilde{x}) - p_3(x))(u_3(x_3) - u_3(\tilde{x}_3)) < 0,$$

which is impossible. This restriction says if a change of characteristic values for alternative three changes the choice probability of a different alternative, then the choice probability of alternative three must also change.

If instead $p_3(x) \neq p_3(\tilde{x})$, then the equation imposes ordinal information on $u_3(\cdot)$. For example, if $p_3(x) < p_3(\tilde{x})$, then $u_3(x_3) < u_3(\tilde{x}_3)$. Now, consider characteristic movements as in Figure 1. This says that behavior consistent with the attraction and compromise effect are rationalized by a strict PUM only if the alternative whose characteristic value changes also experiences a change in choice probability.

Example 4 suggests a strict PUM may always rationalize data when there is no overlap of characteristic values. This intuition is formalized below.

Proposition 1. *If $\{x^n\}_{n=1}^N$ are menus such that $x_a^r \neq x_a^s$ for all $a \in \{1, \ldots, A\}$ and for all $r, s \in \{1, \ldots, N\}$ such that $r \neq s$, then $\{(x^n, p(x^n))\}_{n=1}^N$ can always be rationalized by a strict PUM.*

Thus, there must be some alternatives with the same characteristic values in multiple menus to refute a strict PUM. However, the choice of characteristics to include in the model is at the discretion of the researcher. A common characteristic that is dropped from many revealed preference analysis is time for exactly this reason. Thus, when choosing an indexing of alternatives or which characteristics to explicitly model, one must be wary of the "garbage-in, garbage-out" concerns of Myerson (1981). If many characteristics are modeled, then it

[19] Other revealed preference tests of separable models often generate nonlinear restrictions. See for example Varian (1983) and Cherchye et al. (2015).

is more likely to be in the case of Proposition 1 where any dataset of choice probabilities can be generated from a strict PUM. Therefore, one needs to be judicious when choosing observables to explicitly model. However, the ability to refute strict PUMs is regained by imposing additional structure. Below is an example dataset which refutes linear utility over characteristics for a strict PUM, but trivially satisfies a fully nonparametric model.

Example 5 (Refutation of Strict Linear Perturbed Models)**.** *Consider the stochastic choice dataset in Table 1.*

Table 1: Refutation of a Strict Linear Perturbed Model

x_1	x_2	p_1	p_2
0	0	2/3	1/3
1	1	1/3	2/3
2	2	2/3	1/3

Using Theorem 4(iv) in Appendix A for linear utility over characteristics, consider the restriction involving the first two rows so

$$(2/3 - 1/3)\beta_1 + (1/3 - 2/3)\beta_2 < 0,$$

and hence $\beta_1 < \beta_2$. The restriction of the second and third rows imposes $\beta_2 < \beta_1$. Therefore, this dataset is inconsistent with a strict PUM with linear utility over characteristics. Since all characteristics are distinct for an alternative, a strict PUM rationalizes the dataset from Proposition 1.

4 Application: Stated Preference Data

In this section, we test different specifications of strict PUMs using individual choice data from Louviere et al. (2013). A *specification* of a strict PUM includes a functional form of utility over characteristics and a set of relevant characteristics. A *relevant characteristic* is a characteristic that is explicitly modeled. We examine different specifications to address the questions mentioned earlier: "Can strict PUMs describe individual choice datasets?", "Are there common properties of specifications that best describe the data?", "Which characteristics best describe individual datasets?".

The data from Louviere et al. (2013) was collected from an opt-in web survey asking questions about flight choice.[20] We describe the structure of the survey from Louviere et al. (2013). Each question asked the individual to choose their most-preferred flight from a list of flights named "Flight A", "Flight B", etc. ordered left to right. The list is of fixed size for each individual. The characteristics used for individual questions are fixed, while characteristic values vary for each question. After an individual chooses their most-preferred flight, they are asked whether they would purchase any flight. Given the variety of data, we focus on one subset of data in the main text. Specifically, we examine individual choices from survey questions with lists of size four and restrict the analysis to flights the individual would actually purchase. Results for lists of size three and five, as well as results for data that includes choices individuals would not purchase are in Appendix H.

We provide details to map datasets into a strict PUM framework. We index alternatives by "Flight A", "Flight B", etc. used in the survey. This also coincides with list position from left to right. The leftmost object is denoted $a = 1$. This choice of indexing can be interpreted as costly attention acting through the list position.[21] This index captures how costly it is to pay attention to different sections of the list. Since the analysis is at the individual level, each individual is allowed to process the list in a potentially different way.

Next, we select relevant characteristics. We focus on five characteristics that are included in all survey questions.[22] These characteristics are price, time, brand, stops, and beverage. The values the characteristics can take are given in Table 2. We test the following sets of relevant characteristics: All combinations of price, time, and brand, as well as all five characteristics. We say the set of relevant characteristics is "Full" when all five characteristics are treated as relevant.

Table 2: Values of Flight Characteristics

Characteristic	Description	Values			
Price	Round trip airfare*	\$350	\$450	\$550	\$650
Time	Total travel time	4 hr	5 hr	6 hr	7 hr
Brand	Airline	Qantas	Virgin Blue	Jetstar	Oz Jet
Stops	Number of stops	0	1		
Beverage	Juice/water/soft drinks	Not available(0)	All free(1)		

* Round trip airfare excludes taxes

We also need to specify a functional form of the utility function over characteristics. We

[20]Each individual was asked 16 or 32 discrete choice questions. Louviere et al. (2013) reports that the number of discrete choice survey questions normally ranges from 4-8.

[21]There are other ways to choose an index. We discuss some of these in Appendix J.

[22]The number of characteristics for an alternative varies from six to twelve.

focus on five utility functions over characteristics. These utility functions are fully nonparametric $u_a(x_a)$, additively separable $\sum_{j=1}^{d} u_{a,j}(x_{a,j})$, additively separable and independent of list position $\sum_{j=1}^{d} u_j(x_{a,j})$, linear $\beta_a \cdot x_a$, and linear and independent of position $\beta \cdot x_a$.[23]

Lastly, we discuss how to generate the choice probabilities $p(x^n)$ for observed menus. Once we choose a set of relevant characteristics, we average an individual's choices across menus with the same characteristic values. Thus, $p(x^n)$ is a vector of sample averages.[24] Using averages highlights data trade-offs when designing a test. For example, treating an additional characteristic (that takes on at least two values) as relevant reduces the number of choices to average over when generating choice probabilities. We show in Appendix E that when data is deterministic, testing a strict PUM is equivalent to looking for a strict ordering over the chosen alternatives. This result is analogous to a result for the model in Fudenberg et al. (2015).

We provide some basic descriptive statistics for the four alternative dataset of focus. The average age of individuals is 40.6 years old. The population of individuals tested is 54.7 % female. In the population, 86.9% of individuals have an annual income less than 104,000 Australian Dollars. Descriptive statistics for all datasets are in Appendix G.

4.1 Rationalizability Results

All analysis in this section is for choices an individual would purchase from a list of size four. First, we present the percentage of individual datasets that can be rationalized by a strict PUM for the various specifications in Table 3. We call these percentages pass rates. A number of 0.760 means that 76% of the data sets can be rationalized by a strict PUM. The test is performed using Theorem 1(iii) and the implementation is detailed in Appendix F. The sample size of individuals is denoted S. We refer to sets of relevant characteristics in text as represented in the tables.

Many tests have a high pass rate. In particular, the fully nonparametric model has pass rates up to 95%. This provides some evidence that strict PUMs are able to describe individual choices. Moreover, high pass rates suggest that the indexing of alternatives we use is sensible. At first glance, the results seem to align with intuition. For example, specifications which

[23]Price, time, stops, and beverage have natural implementations in the utility functions. However, there is flexibility when including brand as a relevant characteristic. For the separable and linear tests of rationality, we normalize brand by introducing three indicators which give utility relative to Qantas.

[24]The idea of choice probabilities generated from averaging behavior is studied theoretically in Ahn et al. (2016).

Table 3: Pass Rates for Four Alternatives

	$u_a(x_a)$	$\sum_{j=1}^{d} u_{a,j}(x_{a,j})$	$\sum_{j=1}^{d} u_j(x_{a,j})$	$\beta_a \cdot x_a$	$\beta \cdot x_a$	S
Price	0.760	0.760	0.710	0.729	0.575	221
Brand	0.751	0.751	0.543	0.751	0.543	221
Time	0.751	0.751	0.624	0.688	0.357	221
Price & Brand	0.882	0.652	0.348	0.579	0.303	221
Price & Time	0.869	0.620	0.416	0.448	0.262	221
Brand & Time	0.819	0.348	0.118	0.231	0.068	221
Price & Time & Brand	0.955	0.787	0.520	0.670	0.348	221
Full	0.955	0.837	0.643	0.801	0.511	221

include price tend to describe many individuals. In contrast, specifications for a given utility function with Brand & Time often describe the smallest percentage of individual datasets.

When checking if strict PUMs can describe individual data, there are three effects that occur when adding relevant characteristics. First, the number of free parameters increases which makes violations of rationality less likely. Second, there is an increase in the number of menus to compare probabilities which makes violations of rationality more likely. Third, there is a decrease in the number observations to generate an average choice distribution which has an ambiguous effect. Therefore, pass rates do not necessarily increase with the number of relevant characteristics (see for example going from Price to Price & Brand with additively separable utility).

While pass rates are high and suggest strict PUMs can describe individual choice datasets, we are concerned that pass rates are high simply because there is little characteristic overlap as discussed in Section 3. To account for these concerns, we perform a power correction using the measure of predictive success (MPS) from Beatty and Crawford (2011). An MPS for an individual is given by

$$\text{Measure of Predictive Success}(MPS) = \mathbb{1}\{\text{data described by strict PUM}\} - \text{correction term}.$$

A MPS can take values between negative one and one. For an individual, an MPS close to zero means that if someone were randomizing without considering relevant characteristics, they would just as likely pass the test.

We use two corrections based on Bronars (1987). The correction term is interpreted as how likely a dataset from some class would pass the test. The first power correction examines choice distributions generated anywhere on the probability simplex. The second power correction examines choice distributions that are possible from the number of choices an individual made from a menu for given relevant characteristics. Thus, the second procedure considers permissiveness relative to choices that could have been observed and may provide a stronger power correction for the dataset. The details of the corrections are as follows:

(i) For each menu $x \in \{x^n\}_{n=1}^N$ generate a variable $Z = (z_1, \ldots, z_A)' \sim \text{Uniform}[0,1]^A$. Generate probabilities $\hat{p}_a(x) = z_a / \sum_{a=1}^A z_a$. Check if randomly generated probabilities can be generated by a strict PUM and record the result. Repeat this procedure 100 times. Use the percentage of samples that can be generated by a strict PUM as an estimate of the correction term.

(ii) For each menu $x \in \{x^n\}_{n=1}^N$, examine how many decisions N_x were made from the menu. Generate random variables $Z_i = (z_{i,1}, \ldots, z_{i,A})' \sim \text{Uniform}[0,1]^A$ for $i = 1, \ldots, N_x$. Generate sample probabilities $\hat{p}_a(x) = \frac{1}{N_x} \sum_{i=1}^{N_x} \mathbb{1}\{z_{i,a} > z_{i,b} \text{ for all } a \neq b\}$. Check if randomly generated probabilities can be generated by a strict PUM and record the result. Repeat this procedure 100 times. Use the percentage of samples that can be generated by a strict PUM as an estimate of the correction term.

We refer to the first measure as "basic" since it is the same procedure regardless of the dataset. We refer to the second procedure as "adaptive" since it adapts to properties of the dataset.

We present the average basic MPS in Table 4 and the average adaptive MPS in Table 5. There are many differences between pass rates and average MPS. Although specifications with nonparametric utility have high pass rates, they have average MPS close to zero. This indicates that nonparametric utility has low descriptive power for the datasets observed. Other specifications with high pass rates have average MPS close to zero. For example, the average MPS is nearly zero for specifications with only Brand as a relevant characteristic. For simplicity, we use MPS to refer to the average MPS in the remaining discussion.

Next, we examine how many combinations of relevant characteristics have an MPS greater than 0.10 for at least one utility function over characteristics.[25] There are six combinations of relevant characteristics with basic MPS and adaptive MPS greater than 0.10. The only combinations of relevant characteristics that fail to meet this threshold MPS in both cases are Brand and Brand & Time. For relevant characteristics that exceed the threshold for basic MPS, we find that a linear model always has the highest basic MPS. For relevant characteristics that exceed the threshold for adaptive MPS, a linear model has the highest

[25]There is no standard precedent to evaluate what level of average MPS should be used as a threshold. However, if over 10% of the dataset could be plausibly described by a model after correcting for power, we believe it may be of some interest.

adaptive MPS except for the sets of relevant characteristics Price & Time & Brand and Full. In these cases, the model with additively separable utility independent of list position has the highest adaptive MPS. The highest basic MPS is 0.518 occurs for the specification with linear utility and the relevant characteristics Price & Time & Brand. The highest adaptive MPS is 0.516 for a linear utility independent of list position with the relevant characteristic Price. This means over 50% of individuals can be described by some strict PUM after correcting for power.

Table 4: Measure of Predictive Success for Four Alternatives: Basic

	$u_a(x_a)$	$\sum_{j=1}^{d} u_{a,j}(x_{a,j})$	$\sum_{j=1}^{d} u_j(x_{a,j})$	$\beta_a \cdot x_a$	$\beta \cdot x_a$	S
Price	-0.008	-0.008	0.220	0.086	0.493	221
Brand	-0.010	-0.010	0.022	-0.010	0.022	221
Time	-0.013	-0.013	0.133	0.041	0.279	221
Price & Brand	-0.026	0.307	0.202	0.447	0.262	221
Price & Time	-0.012	0.339	0.362	0.381	0.235	221
Brand & Time	-0.076	0.005	-0.027	0.098	0.027	221
Price & Time & Brand	-0.045	0.358	0.363	0.518	0.302	221
Full	-0.045	0.314	0.472	0.490	0.453	221

Table 5: Measure of Predictive Success for Four Alternatives: Adaptive

	$u_a(x_a)$	$\sum_{j=1}^{d} u_{a,j}(x_{a,j})$	$\sum_{j=1}^{d} u_j(x_{a,j})$	$\beta_a \cdot x_a$	$\beta \cdot x_a$	S
Price	0.014	0.014	0.296	0.115	0.516	221
Brand	0.002	0.002	0.098	0.002	0.099	221
Time	0.006	0.006	0.204	0.071	0.298	221
Price & Brand	0.033	0.331	0.283	0.425	0.263	221
Price & Time	0.040	0.348	0.355	0.364	0.237	221
Brand & Time	-0.015	0.042	0.054	0.075	0.031	221
Price & Time & Brand	0.022	0.293	0.420	0.377	0.301	221
Full	0.019	0.257	0.491	0.332	0.443	221

These results suggest that linear models of utility over characteristics in a strict PUM may describe many important aspects of individual choice. For example, the model of Price with linear utility independent of position is among the top two models according to the basic and adaptive MPS. Lastly, while basic and adaptive MPS are typically similar when analyzing strict PUMs, they can qualitatively differ. For example, the basic MPS for $\beta_a \cdot x_a$ with relevant characteristics Price & Time & Brand is 0.518, while it drops to 0.377 using the adaptive MPS. While a difference of 0.13 is large, it is known that different correction procedures can yield different results.[26]

[26]See Andreoni et al. (2013) for a variety of other procedures one could use to calculate correction terms.

4.2 Monotonicity Restriction Results

Given the relative success of describing individual choice datasets using strict PUMs with linear utility over characteristics, a natural question is: "Can we find parameters which have the correct intuitive sign?". For example, it is intuitive that coefficients on price, time, and number of stops would be negative, while the coefficient on beverage would be positive. We impose these coefficient constraints while checking for a rationalization by a strict PUM with linear utility over characteristics. The results are presented in Table 6.

Table 6: Linear Monotonicity Results for Four Alternatives

	Pass Rates		Basic MPS		Adaptive MPS	
	$\beta_a \cdot x_a$	$\beta \cdot x_a$	$\beta_a \cdot x_a$	$\beta \cdot x_a$	$\beta_a \cdot x_a$	$\beta \cdot x_a$
Price	0.674	0.575	0.400	0.529	0.472	0.539
Time	0.525	0.348	0.248	0.301	0.315	0.314
Price & Brand	0.570	0.303	0.484	0.269	0.493	0.269
Price & Time	0.353	0.258	0.314	0.240	0.321	0.240
Brand & Time	0.208	0.068	0.126	0.038	0.139	0.037
Price & Time & Brand	0.633	0.348	0.540	0.312	0.528	0.313
Full	0.760	0.430	0.611	0.384	0.556	0.383

While imposing constraints, there are only modest decreases in pass rates. The largest decrease in pass rate is approximately 16%, while the remaining specifications drop less than 10%. In fact, the pass rate is the same regardless of monotonicity constraints for several specifications. For position independent linear utilities, pass rates are unchanged for the specifications of Price, Price & Brand, Brand & Time, and Price & Time & Brand. This is evidence that if an individual's choices can be rationalized, then they often can be made to have utilities that are monotonic over characteristics in intuitive ways.

Adding monotonicity constraints weakly decreases the pass rates and the correction term, so individual MPS can increase or decrease. The average MPS primarily increases for utilities that depend on position and remains almost unchanged for position independent utilities. This may occur because the correction term for the position independent test is of smaller magnitude before imposing the monotonicity constraints. The specification that has the highest basic and adaptive MPS uses the Full set of relevant characteristics with position dependent linear utilities. Therefore, accounting for *a priori* information can change the MPS ranking of specifications and the descriptive power of strict PUMs. We note that the specification with only Price and position independent utility still has high average MPS.

4.3 Comparison to Three and Five Alternative Results

We the same analysis for individual choice datasets while conditioning on the flights an individual would purchase from lists of size three and five. Results are reported in Appendix H. We find that the MPS ranking of a strict PUM with only Price and position independent linear utility weakly improves as list size increases. This result also holds when examining MPS rankings while imposing monotonicity restrictions. This provides some evidence that as the number of alternatives increases, individual behavior may best be described using fewer relevant characteristics. Moreover, specifications using only Time and Brand do not consistently improve in MPS ranking as list size grows. This suggests that individuals may have a priority ordering over which characteristics are valued as the number of alternatives to choose from increases.

5 Conclusion

We provide a study of characteristics in a model of stochastic choice using strict perturbed utility models. This class of models restricts behavioral effects to occur through a costly attention function that only depends on the indexing of alternatives. We show that the characteristic approach generalizes variation in alternative availability and the strict PUMs nest behavior from additive random utility models. Moreover, we show a strict PUM can be thought of as accounting for characteristics using a first order approximation. In addition, we provide a system of inequalities which completely characterizes behavior from strict PUMs. These inequalities can be interpreted as being able to find utilities which do not produce a utility pump.

Finally, we perform tests of strict PUMs for various specifications of utility functions and relevant characteristics using stated preference data from Louviere et al. (2013). We find that fully nonparametric utility functions over characteristics often rationalize the data. However, after applying a power correction for the permissiveness of the test, we find that linear specifications provide more powerful descriptions of individual choice datasets. A linear specification with only Price is often a powerful descriptor of the data, but can often be improved on by adding other characteristics. There is also some evidence of a priority order on characteristics as the number of alternatives to choose from increases. These results also may suggest examining simple perturbed utility models with linear utility over a few characteristics may provide insights in applications.

Appendix A Proofs of Main Results

Before proving results in the main text, we show that any dataset can be rationalized by a PUM when the set of maximizers is not a singleton.

Definition 2 (Rationalization of Weak PUM). *The dataset $\{(x^n, p(x^n))\}_{n=1}^N$ is rationalized by a weak perturbed utility model if there exist $u_a : \mathcal{X}_a \to \mathbb{R}$ for all $a \in \{1, \ldots, A\}$ and a convex $\tilde{C} : \Delta \to \mathbb{R}$ such that*

$$p(x^n) \in \operatorname*{argmax}_{p \in \Delta} \sum_{a=1}^{A} p_a u_a(x_a^n) - \tilde{C}(p)$$

for all $n \in \{1, \ldots, N\}$.

Theorem 2. *Any dataset $\{(x^n, p(x^n))\}_{n=1}^N$ is rationalized by a weak PUM.*

Proof. For all $a \in \{1, \ldots, A\}$ set $u_a(x_a) = 0$ for all $x_a \in \mathcal{X}_a$. Set $\tilde{C}(p) = 0$ for all $p \in \Delta$. Since all probabilities yield the same utility, any observation $p(x^n)$ is a maximizer. □

We note that condition (iv) of Theorem 1 can be re-written as strict cyclic monotonicity. We present the definition of strict cyclic monotonicity here since it is useful when constructing a cost function in the proof of Theorem 1.

Definition 3 (Strict Cyclic Monotonicity). *A function $\rho : \mathbb{R}^K \to \mathbb{R}^K$ satisfies strict cyclic monotonicity if for every positive integer $M \in \mathbb{Z}_+$ and sequence $y[1], \ldots, y[M], y[M+1] = y[1] \in \mathbb{R}^K$ with at least two $y[m]$'s distinct, then*

$$\sum_{m=1}^{M} (y[m+1] - y[m])' \rho(y[m]) < 0.$$

Proof of Theorem 1. ((*iv*) ⇒ (*i*)) Assume that $\{(x^n, p(x^n))\}_{n=1}^N$ satisfies (iv). Let Σ be the set of all finite sequences $\{x[m]\}_{m=1}^M$ with $x[m] \in \{x^n\}_{n=1}^N$ such that $p(x[m]) \neq p(x[m+1])$ for some m. By (*iv*) there exist numbers $\{u_a^n\}_{n=1}^N$ for all $a \in \{1, \ldots, A\}$ such that for all $s, r \in \{1, \ldots, N\}$ if $x_a^s = x_a^r$ then $u_a^s = u_a^r$ and

$$\sum_{m=1}^{M} (p(x[m+1]) - p(x[m])) \cdot U[m] < 0$$

where $x[M+1] = x[1]$ and $U[m] = (u_1[m], \ldots, u_A[m])$ where $u_a[m]$ is the u_a^r term associated to $x_a[m] = x_a^r$. Since the inequality is strict, there exists $\varepsilon_0 > 0$ small enough so that there exist numbers $\{u_a^n\}_{n=1}^N$ for all $a \in \{1, \ldots, A\}$ such that for all $s, r \in \{1, \ldots, N\}$ if $x_a^s = x_a^r$ then

$u_a^s = u_a^r$ and

$$\sum_{m=1}^{M} (p(x[m+1]) - p(x[m])) \cdot U[m] + \varepsilon_0 < 0 \tag{3}$$

for all sequences in Σ with $x[M+1] = x[1]$.

Consider the function $f : \mathbb{R}^A \to \mathbb{R}$ given by $f(y) = (y_1^2 + \ldots + y_A^2 + T)^{1/2} - T^{1/2}$ for $T > 0$. This function is used in Matzkin and Richter (1991). In particular, $f(\cdot)$ is strictly convex, differentiable, $f(0) = 0$, $f(y) > 0$ if $y \neq 0$, and $\left[\frac{\partial f}{\partial y_a}(y)\right] < 1$ for all y and $a \in \{1, \ldots, A\}$. From Equation 3, there exists $\varepsilon > 0$ small enough so that there exist numbers $\{u_a^n\}_{n=1}^N$ for all $a \in \{1, \ldots, A\}$ such that for all $s, r \in \{1, \ldots, N\}$ if $x_a^s = x_a^r$ then $u_a^s = u_a^r$ and

$$\sum_{m=1}^{M} [(p(x[m+1]) - p(x[m])) \cdot U[m] + \varepsilon g(p(x[m+1]) - p(x[m]))] < 0 \tag{4}$$

where $x[M+1] = x[1]$.

Next, consider the function $\phi_\sigma : \Delta \to \mathbb{R}$ for each sequence $\sigma \in \Sigma$ given by

$$\begin{aligned}\phi_\sigma(p) = &\sum_{m=1}^{M-1} [(p(x[m+1]) - p(x[m]) \cdot U[m] + \varepsilon g(p(x[m+1]) - p(x[m]))] \\ &+ (p - p(x[M])) \cdot U[M] + \varepsilon g(p - p(x[M])).\end{aligned}$$

Each $\phi_\sigma(\cdot)$ is strictly convex on the simplex since it is the sum of an affine and strictly convex function restricted to a convex domain.

Using the ϕ_σ functions, we use a constructive procedure from Rockafellar (1970) Theorem 24.8. First, choose an arbitrary $p(x_0) \in \{p(x^n)\}_{n=1}^N$ and let Σ_0 be the set of sequences which begin with $p(x_0)$. Next, define a function $C : \Delta \to \mathbb{R}$ given by

$$C(p) = \max_{\sigma_0 \in \Sigma_0} \{\phi_{\sigma_0}(p)\}.$$

$C(\cdot)$ is defined as a max of strictly convex functions, so it is strictly convex.

All that remains is to show that the numbers $\{\{u_a^n\}_{a=1}^A\}_{n=1}^N$ used to satisfy Equation 4 and the $C(\cdot)$ function rationalize the data $\{(x^n, p(x^n))\}_{n=1}^N$. To see this let σ_0^n be the sequence where $C(p(x^n))$ achieves the maximum. For $p \neq p(x^n)$ and $p \in \Delta$ we have

$$\begin{aligned}
\sum_{a=1}^{A} p_a u_a^n - C(p) &= \sum_{a=1}^{A} p_a u_a^n - \max_{\sigma_0 \in \Sigma_0}\{\phi_{\sigma_0}(p)\} \\
&\leqslant \sum_{a=1}^{A} p_a u_a^n - \left[\sum_{a=1}^{A}(p_a - p_a(x^n))u_a^n + \varepsilon g(p - p(x^n)) + \phi_{\sigma_0^n}(p(x^n))\right] \\
&= \sum_{a=1}^{A} p_a(x^n)u_a^n - \varepsilon g(p - p(x^n)) - \phi_{\sigma_0^n}(p(x^n)) \\
&< \sum_{a=1}^{A} p_a(x^n)u_a^n - \phi_{\sigma_0^n}(p(x^n)) \\
&= \sum_{a=1}^{A} p_a(x^n)u_a^n - C(p(x^n))
\end{aligned}$$

where the first inequality comes by choosing the sequence which ends with $p(x^n)$ and begins with the largest cost sequence for $p(x^n)$ and the second inequality is from $\varepsilon g(p - P(x^n)) > 0$ for $p \neq p(x^n)$. The rationalization can be extended to all of $\mathcal{X}$ by choosing any real number for utilities associated with unobserved characteristic values.

((*iii*) $\Leftrightarrow$ (*v*)) We prove this result in Appendix B when discussing second stage expected utility. □

The following results can be proved using the same approach in Theorem 1 when imposing structure on $u_a(x_a)$. We provide results imposing structure so that $u_a(x_a)$ is equal to $\sum_{j=1}^{d_a} u_{a,j}(x_{a,j})$ or $\beta_a \cdot x_a$. We do not provide a version of Theorem 1(v) for brevity, however a similar statement will hold by application of a theorem of the alternative as in the proof of Theorem 1.

Theorem 3. *Consider the dataset $\{(x^n, p(x^n))\}_{n=1}^N$. The following are equivalent:*

(i) $\{(x^n, p(x^n))\}_{n=1}^N$ *is strictly rationalized by a perturbed utility model with* $u_a(x_a) = \sum_{j=1}^{d_a} u_{a,j}(x_{a,j})$.

(ii) There exist utility functions $u_a(x_a) = \sum_{j=1}^{d_a} u_{a,j}(x_{a,j})$ *for all* $a \in \{1, \ldots, A\}$ *and a function* $g : \Delta \to \mathbb{R}$ *such that* $p(x^n) = \operatorname{argmax}_{p \in \Delta} \sum_{a=1}^{A} \sum_{j=1}^{d_a} p_a u_{a,j}(x_{a,j}^n) - g(p)$ *for all* $n \in \{1, \ldots, N\}$.

(iii) There exist numbers $\{\{u_{a,j}^n\}_{j=1}^{d_a}\}_{n=1}^N$ *for all* $a \in \{1, \ldots, A\}$ *and* $\{g^n\}_{n=1}^N$, *such that for all*

$(s,r) \in \{1,\ldots,N\} \times \{1,\ldots,N\}$ with $p(x^s) \neq p(x^r)$ then

$$\sum_{a=1}^{A}\sum_{j=1}^{d_a} p_a(x^s)u^r_{a,j} - g^s < \sum_{a=1}^{A}\sum_{j=1}^{d_a} p_a(x^r)u^r_{a,j} - g^r$$

and for all $r,s \in \{1,\ldots,N\}$ and $a \in \{1,\ldots,A\}$

$$\begin{aligned} u^r_{a,j} = u^s_{a,j} \quad &\textit{if} \quad x^r_{a,j} = x^s_{a,j} \qquad \textit{for all } j \in \{1,\ldots,d_a\} \\ g^r = g^s \quad &\textit{if} \quad p(x^r) = p(x^s). \end{aligned}$$

(iv) There exist numbers $\{\{u^n_{a,j}\}_{j=1}^{d_a}\}_{n=1}^{N}$ for all $a \in \{1,\ldots,A\}$ such that for all finite sequences $\{x[m]\}_{m=1}^{M}$ with $x[m] \in \{x^n\}_{n=1}^{N}$ and $p(x[m]) \neq p(x[m+1])$ for some m

$$\sum_{m=1}^{M}\left(\sum_{a=1}^{A}\sum_{j=1}^{d_a}(p_a(x[m+1]) - p_a(x[m]))u_{a,j}[m]\right) < 0$$

where $x[M+1]) = x[1]$ and $u_{a,j}[m]$ is the $u^r_{a,j}$ term associated to $x_{a,j}[m] = x^r_{a,j}$; and for all $r,s \in \{1,\ldots,N\}$ and $a \in \{1,\ldots,A\}$

$$u^r_{a,j} = u^s_{a,j} \quad \textit{if} \quad x^r_{a,j} = x^s_{a,j} \qquad \textit{for all } j \in \{1,\ldots,d_a\}.$$

Theorem 4. *Consider the dataset $\{(x^n, p(x^n))\}_{n=1}^{N}$. The following are equivalent:*

(i) $\{(x^n, p(x^n))\}_{n=1}^{N}$ is strictly rationalized by a perturbed utility model with $u_a(x_a) = \beta_a \cdot x_a$.

(ii) There exist utility functions $u_a(x_a) = \beta_a \cdot x_a$ for all $a \in \{1,\ldots,A\}$ and a function $g : \Delta \to \mathbb{R}$ such that $p(x^n) = \operatorname{argmax}_{p \in \Delta} \sum_{a=1}^{A} p_a(\beta_a \cdot x_a) - g(p)$ for all $n \in \{1,\ldots,N\}$.

(iii) There exist numbers $\beta_a \in \mathbb{R}^d_a$ for all $a \in \{1,\ldots,A\}$ and $\{g^n\}_{n=1}^{N}$, such that for all $(s,r) \in \{1,\ldots,N\} \times \{1,\ldots,N\}$ with $p(x^s) \neq p(x^r)$ then

$$\sum_{a=1}^{A} p_a(x^s)(\beta_a \cdot x^r_a) - g^s < \sum_{a=1}^{A} p_a(x^r)(\beta_a \cdot x^r_a) - g^r$$

and for all $r,s \in \{1,\ldots,N\}$

$$g^r = g^s \quad \textit{if} \quad p(x^r) = p(x^s).$$

(iv) There exist numbers $\beta_a \in \mathbb{R}^d_a$ for all $a \in \{1,\ldots,A\}$ such that for all finite sequences

$\{x[m]\}_{m=1}^{M}$ *with* $x[m] \in \{x^n\}_{n=1}^{N}$ *and* $p(x[m]) \neq p(x[m+1])$ *for some* m

$$\sum_{m=1}^{M}\left(\sum_{a=1}^{A}(p_a(x[m+1]) - p_a(x[m]))(\beta_a \cdot x[m])\right) < 0$$

where $x[M+1]) = x[1]$.

There are other natural results when $\mathcal{X}_m = \mathcal{X}$ for all $m \in \{1,\ldots,M\}$ and the utility is independent of the alternative. In this case, one considers utilities $u(x_a)$, $\sum_{j=1}^{d} u_j(x_{a,j})$, and $\beta \cdot x_a$ appropriately. One can also impose monotonicity constraints on the utility numbers.

Proof of Proposition 1 . Any dataset with $p(x^n) = \hat{p}$ for all $n \in \{1,\ldots,N\}$ can be rationalized by Corollary 1. Next, let $\{x^n\}_{n=1}^{N}$ be characteristics such that $x_a^r \neq x_a^s$ for all $a \in \{1,\ldots,A\}$ and for all $r,s \in \{1,\ldots,N\}$ such that $r \neq s$. To show $\{(x^n, p(x^n))\}_{n=1}^{N}$ can be strictly rationalized it suffices to show that there exist $\{\{u_a^n\}_{a=1}^{A}\}_{n=1}^{N}$ which satisfy Theorem 1(iv). Let $U^n = (u_1^n,\ldots,u_A^n)$ for each $n \in \{1,\ldots,N\}$ and let $v = (U^1,\ldots,U^N)$ be the vector of utilities associated with all observed characteristic values. Utility numbers differ for each menu x^n by assumption. We can rewrite the system of inequalities given by Theorem 1(iv) using a matrix R as $Rv < 0$ where the strict inequality holds component-wise. Let $\gamma_a^{s,r} = p_a(x^s) - p_a(x^r)$. For example, when $N = M = 3$ and probabilities are distinct for each menu, the matrix R is given by

$$R = \begin{array}{c} \\ (1,2) \\ (1,3) \\ (2,3) \\ (1,2,3) \\ (1,3,2) \\ \\ \end{array} \begin{array}{c} \begin{array}{ccccccccc} u_1^1 & u_2^1 & u_3^1 & u_1^2 & u_2^2 & u_3^2 & u_1^3 & u_2^3 & u_3^3 \end{array} \\ \left[\begin{array}{ccccccccc} \gamma_1^{2,1} & \gamma_2^{2,1} & \gamma_3^{2,1} & \gamma_1^{1,2} & \gamma_2^{1,2} & \gamma_3^{1,2} & 0 & 0 & 0 \\ \gamma_1^{3,1} & \gamma_2^{3,1} & \gamma_3^{3,1} & 0 & 0 & 0 & \gamma_1^{1,3} & \gamma_2^{1,3} & \gamma_3^{1,3} \\ 0 & 0 & 0 & \gamma_1^{3,2} & \gamma_2^{3,2} & \gamma_3^{3,2} & \gamma_1^{2,3} & \gamma_2^{2,3} & \gamma_3^{2,3} \\ \gamma_1^{2,1} & \gamma_2^{2,1} & \gamma_3^{2,1} & \gamma_1^{3,2} & \gamma_2^{3,2} & \gamma_3^{3,2} & \gamma_1^{1,3} & \gamma_2^{1,3} & \gamma_3^{1,3} \\ \gamma_1^{3,1} & \gamma_2^{3,1} & \gamma_3^{3,1} & \gamma_1^{1,2} & \gamma_2^{1,2} & \gamma_3^{1,2} & \gamma_1^{2,3} & \gamma_2^{2,3} & \gamma_3^{2,3} \\ \\ \end{array}\right] \end{array}.$$

We show that for any R generated by Theorem 1(iv) with no characteristic value overlap, the set $Rv < 0$ is always feasible.

Assume that the inequalities given by $Rv < 0$ is infeasible, then by a theorem of the alternative[27] there exists $\lambda \gneq 0$ such that

$$\lambda' R = (0, \ldots, 0) \quad \text{and} \quad \sum_{\sigma \in \Sigma} \lambda_\sigma = 1,^{28} \tag{5}$$

where Σ be the set of all finite sequences used to generate the inequalities in Theorem 1(iv). The $(a + (n-1)A)$-th column of R for $a \in \{1, \ldots, A\}$ and $n \in \{1, \ldots, N\}$ contains entries in sequences which are associated with the u_a^n term. Let $\text{Col}_i(R)$ be the i-th column of R. Replace values of 0 in $\text{Col}_{a+(n-1)A}(R)$ by $(p_a(x^n) - p_a(x^n))$. Let $p_a(x^\sigma)$ be the probability for the σ-th sequence in column $a + (n-1)A$ with the zero entries replaced as mentioned before. Thus

$$\lambda' \text{Col}_{a+(n-1)A}(R) = \sum_{\sigma \in \Sigma} \lambda_\sigma (p_a(x^\sigma) - p_a(x^n)) = 0.$$

Recall that $\sum_{\sigma \in \Sigma} \lambda_\sigma = 1$ so that the above implies that

$$\sum_{\sigma \in \Sigma} \lambda_\sigma p_a(x^\sigma) = p_a(x^n). \tag{6}$$

Let $G_a = \{a + (n-1)A \mid p_a(x^n) \geqslant p_a(x^{\tilde{n}}) \text{ for all } \tilde{n} \neq n\}$ be the set of indices such that the probability of choosing alternative a is greatest among the observed probabilities. For $a+(n-1)A \in G_a$, from Equation 5 for column $\text{Col}_{a+(n-1)A}(R)$ is taking an average over points that are weakly less than $p_a(x^n)$, so that $\lambda_\sigma > 0$ for $\sigma \in \Sigma$ if and only if $p_a(x^\sigma) = p_a(x^n)$. Denote the subset of indices with $\lambda_\sigma > 0$ as Σ^a.

Let $\Sigma^* = \bigcap_{a=1}^{A} \Sigma^a$. Suppose that $\Sigma^* \neq \varnothing$. Since σ was included in R there must be $p(x[m+1]) \neq p(x[m])$. However, $\sigma \in \Sigma^*$ so examining the columns associated to utilities generated by menu $x[m]$ we see that $\lambda_\sigma > 0$ implies that $p_a(x[m+1]) = p_a(x[m])$ for all $a \in \{1, \ldots, A\}$ which is a contradiction. Thus, it must be that $\Sigma^* = \varnothing$ and the dual system is infeasible. Therefore, $Rv > 0$ is always feasible and we can construct a strict rationalization using the method in Theorem 1. □

Appendix B Second Stage Expected Utility Relation

Throughout the paper, we have considered a strict PUM as a model of stochastic choice. However, we could use this model to describe preferences over a finite set of risky outcomes. In this section, we consider how to relate the inequalities from Theorem 1(v) to results on

[27] See for example Border (2013) Corollary 15.

[28] We note that R' denotes the transpose of the matrix R.

expected utility from Fishburn (1975). The preferences of strict PUMs are not equivalent to expected utility. However, we can consider preferences in the spirit of second stage expected utility from Segal (1990). Thus, we can consider when preferences over the second stage of randomness are represented by expected utility.

To make the comparison concrete, consider an environment with A outcomes. The value of each outcome depends on observables given by $x = (x_1, \ldots, x_A)$. Let p be a probability distribution over the A outcomes. The pair (x, p) defines a lottery with observables x. Next, consider the second stage lottery over M pairs (x, p) given by $\{(\pi_m, (x[m], p[m]))\}_{m=1}^M$ where $\pi_m \geqslant 0$ for all m and $\sum_{m=1}^M \pi_m = 1$. One can imagine an individual offered a pair of second stage lotteries given by $\{(\pi_m, (x[m], p[m]))\}_{m=1}^M$ and $\{(\tilde{\pi}_\ell, (\tilde{x}[\ell], \tilde{p}[\ell]))\}_{\ell=1}^L$, suppose that the agent has non-expected utility preferences over first stage lotteries given by $V : \mathcal{X} \times \Delta \to \mathbb{R}$ and the second stage satisfies expected utility so

$$\{(\tilde{\pi}_\ell, (\tilde{x}[\ell], \tilde{p}[\ell]))\}_{\ell=1}^L \preceq \{(\pi_m, (x[m], p[m]))\}_{m=1}^M \quad \text{if and only if} \\ \sum_{\ell=1}^L \tilde{\pi}_\ell V(\tilde{x}[\ell], \tilde{p}[\ell]) \leqslant \sum_{m=1}^M \pi_m V(x[m], p[m])$$

and

$$\{(\tilde{\pi}_\ell, (\tilde{x}[\ell], \tilde{p}[\ell]))\}_{\ell=1}^L \prec \{(\pi_m, (x[m], p[m]))\}_{m=1}^M \quad \text{if and only if} \\ \sum_{\ell=1}^L \tilde{\pi}_\ell V(\tilde{x}[\ell], \tilde{p}[\ell]) < \sum_{m=1}^M \pi_m V(x[m], p[m])$$

This formulation can now be related to work on expected utility based tests of demand by Kubler et al. (2014), Echenique and Saito (2015), and Chambers et al. (2016).

Definition 4. *We say that an individual satisfies second stage expected utility preferences if there exists a function $V : \mathcal{X} \times \Delta \to \mathbb{R}$ such that $\{(\pi_m, (x[m], p[m]))\}_{m=1}^M$ and $\{(\tilde{\pi}_\ell, (\tilde{x}[\ell], \tilde{p}[\ell]))\}_{\ell=1}^L$ such that*

$$\{\{(\tilde{\pi}_\ell, (\tilde{x}[\ell], \tilde{p}[\ell]))\}_{\ell=1}^L \preceq (\prec) \{(\pi_m, (x[m], p[m]))\}_{m=1}^M \quad \textit{if and only if} \\ \sum_{\ell=1}^L \tilde{\pi}_\ell V(\tilde{x}[\ell], \tilde{q}[\ell]) \leqslant (<) \sum_{m=1}^M \pi_m V(x[m], q[m])$$

We provide one interpretation of the second stage expected utility formulation. Consider an individual facing a decision between two degenerate lotteries on (x, p) and $(\tilde{x}, \tilde{p})$. Suppose that the A alternatives are risky assets that have observables given by x and $\tilde{x}$. Now, p and $\tilde{p}$ can be interpreted as portfolios over the A risky assets with properties x and $\tilde{x}$. An individual may prefer mixtures of assets depending on the observables x and $\tilde{x}$ when evaluating simple lotteries. However, if an individual is offered a choice between mixtures of portfolios given by $\{(\pi_m, (x[m], p[m])\}_{m=1}^M$ and $\{(\pi_\ell, (x[\ell], p[\ell])\}_{\ell=1}^L$, the decision environment has become more complex. The individual may be able to calculate how much they value each (x, p) lottery, but then evaluate the more complex environment by taking an average. Thus, second stage expected utility preferences allow individuals rich preferences in simple decision environments, but simpler preferences from more complex decision environments.

Now we impose restrictions from a strict PUM, so

$$V(x,p) = \sum_{a=1}^{A} u_a(x_a)p_a - C(p).$$

For the dataset $\{(x^n, p(x^n)\}_{n=1}^N$, recall that $V(x^n, p(x^n)) > V(x^n, p)$ for all $p \in \Delta$ with $p \neq p(x^n)$. Theorem 1(iii) says that it suffices to looking at binary comparisons of probabilities from $\{p(x^n)\}_{n=1}^N$ to find a strict PUM. If we select menus $x[m], \tilde{x}[m] \in \{x^n\}_{n=1}^N$, then for all lotteries $\{(\pi_m, (x[m], p(x[m])))\}_{m=1}^M$ and $\{(\pi_m, (x[m], p(\tilde{x}[m])))\}_{m=1}^M$ an individual with second stage expected utility must also satisfy

$$\sum_{m=1}^{M} \pi_m \Big(\sum_{a=1}^{A} u_a(x_a[m]) p_a(\tilde{x}[m]) - C(p(\tilde{x}[m])) \Big) \leqslant \sum_{m=1}^{M} \pi_m \Big(\sum_{a=1}^{A} u_a(x_a[m]) p_a(x[m]) - C(p(x[m])) \Big).$$

Using this formulation, the terms of $\pi_m, p(x[m])$, and $p(\tilde{x}[m])$ define lotteries over the utility terms $\{\{u_a(x_a[m])\}_{a=1}^A\}_{m=1}^M$ and $\{C(p(x[m])\}_{m=1}^M$. Moreover, if at least one set of probabilities differs the inequality is strict.

We note that second stage expected utility is still an expected utility problem. Therefore, we use the results of Fishburn (1975). This shows that an individual satisfies second stage expected utility with a strict perturbed utility if and only if there does not exist $\pi_m \geqslant 0$ such that $\sum_{m=1}^M \pi_\ell = 1$ such that

$$\sum_{m=1}^{M} \pi_m \Big(\sum_{a=1}^{A} u_a(x_a[m]) p_a(\tilde{x}[m]) - C(p(\tilde{x}[m])) \Big) = \sum_{m=1}^{M} \pi_m \Big(\sum_{a=1}^{A} u_a(x_a[m]) p_a(x[m]) - C(p(x[m])) \Big).$$

with $\pi_m > 0$ for some m with $p(x[m]) \neq p(\tilde{x}[m])$.

We see that any pair with $p(x[m]) = p(\tilde{x}[m])$ can be removed from the comparisons without loss of generality since the utility difference for any lottery is zero. Thus, we restrict attention to the subset of ordered comparisons $(\tilde{x}, x)$ with $p(\tilde{x}) \neq p(x)$. We denote this set of comparisons as

$$\mathcal{S} = \{(\tilde{x}, x) \in \{x^n\}_{n=1}^N \times \{x^n\}_{n=1}^N \mid p(\tilde{x}) \neq p(x)\}.$$

The $\tilde{x}$ term references a choice distribution, while x will reference a choice distribution and the environment for an inequality from Theorem 1(iii). Thus, if we can find a distribution $\{\pi_{(\tilde{x},x)}\}_{(\tilde{x},x)\in\mathcal{S}}$ with $\pi_{(\tilde{x},x)} \geqslant 0$ and $\sum_{(\tilde{x},x)\in\mathcal{S}} \pi_{(\tilde{x},x)} = 1$ such that

$$\sum_{(\tilde{x},x)\in\mathcal{S}} \pi_{(\tilde{x},x)}\Big(\sum_{a=1}^{A} u_a(x_a)p_a(\tilde{x}) - C(p(\tilde{x}))\Big) = \sum_{(\tilde{x},x)\in\mathcal{S}} \pi_{(\tilde{x},x)}\Big(\sum_{a=1}^{A} u_a(x_a)p_a(x) - C(p(x))\Big)$$

then second stage expected utility preferences are refuted. Since we are already restricting to strict utility comparisons, the condition that $\pi_{(\tilde{x},x)} > 0$ for some $p(\tilde{x}) \neq p(x)$ is automatically satisfied.

So far, we have assumed that the $u_a(x_a)$ and $C(p)$ terms are observed. However, utility numbers are unobservable. It is clear that a sufficient condition to find a refutation is that the sums of extended lotteries on $u_a(x_a)$ and $C(\cdot)$ terms are equal. We show this condition is also necessary.

Theorem 5. *Consider the dataset $\{(x^n, p(x^n))\}_{n=1}^N$. The following are equivalent:*

(i) $\{(x^n, p(x^n))\}_{n=1}^N$ *is rationalized by a strict PUM.*

(ii) *Let* $\mathcal{S} = \{(\tilde{x}, x) \in \{x^n\}_{n=1}^N \times \{x^n\}_{n=1}^N \mid p(\tilde{x}) \neq p(x)\}$. *There is no* $\{\pi_{(\tilde{x},x)}\}_{(\tilde{x},x)\in\mathcal{S}}$ *with* $\pi_{(\tilde{x},x)} \geqslant 0$ *and* $\sum_{(\tilde{x},x)\in\mathcal{S}} \pi_{(\tilde{x},x)} = 1$ *such that for all* $\hat{x}_a \in \{\{x_a^n\}_{a=1}^A\}_{n=1}^N$

$$\sum_{\{(\tilde{x},x)\in\mathcal{S} \mid x_a = \hat{x}_a\}} \pi_{(\tilde{x},x)}(p_a(\tilde{x}) - p_a(x)) = 0$$

and for all $\hat{x} \in \{x^n\}_{n=1}^N$

$$\sum_{\{(\tilde{x},x)\in\mathcal{S} \mid P(\tilde{x}) = P(\hat{x})\}} \pi_{(\tilde{x},x)} = \sum_{\{(\tilde{x},x)\in\mathcal{S} \mid P(x) = P(\hat{x})\}} \pi_{(\tilde{x},x)}.$$

(iii) $\{(x^n, p(x^n))\}_{n=1}^N$ *are rationalized by a second stage expected utility model where* $V(x,p)$ *is represented by a strict PUM.*

It is clear that (i) is equivalent to (iii). If the individual is strictly rationalized by perturbed utility model, then the constructed perturbed utility function can be used to generate second stage expected utility preferences. Moreover, if the individual has second stage expected utility preferences with a strictly perturbed utility function, then they have strictly perturbed utility preferences over degenerate first stage lotteries. We also have that (iii) implies (ii). If there exists a distribution that satisfies the equalities in (ii), then it provides a violation for any second stage expected utility with a strict PUM. Therefore, it suffices to show that (ii) implies (i). In fact, (ii) is equivalent to (i) via a theorem of the alternative using the inequalities from Theorem 1(iii). This result also proves that Theorem 1(iii) is equivalent to Theorem 1(v).

Proof of Theorem 5. To prove (ii) is equivalent to (i), we use the inequalities from Theorem 1(iii). We know that $\{(x^n, p(x^n))\}_{n=1}^N$ is strictly rationalized by a perturbed utility model if there exist numbers $\{u_a^n\}_{n=1}^N$ for all $a \in \{1, \ldots, A\}$ and $\{g^n\}_{n=1}^N$ such that for all $(s, r) \in \{1, \ldots, N\} \times \{1, \ldots, N\}$ with $p(x^s) \neq p(x^r)$ then

$$\sum_{a=1}^{A} p_a(x^s) u_a^r - g^s < \sum_{a=1}^{A} p_a(x^r) u_a^r - g^r$$

and for all $r, s \in \{1, \ldots, N\}$

$$\begin{aligned} u_a^r = u_a^s \quad &\text{if} \quad x_a^r = x_a^s \\ g^r = g^s \quad &\text{if} \quad p(x^r) = p(x^s). \end{aligned}$$

Let W_a be the number of unique u_a^n terms for each a. Therefore, for every a there are W_a unique terms of $\{u_a^n\}_{n=1}^N$. These terms will be denoted $u_a^{w_a}$ for $w_a = \{1, \ldots, W_a\}$. Let L be the number of unique $p(x^n)$, so there are L unique terms of $\{g^n\}_{n=1}^N$. These terms will be denoted by g_ℓ for $\ell = \{1, \ldots, L\}$. Let e_x be the vector with a one in the column corresponding to the term of g for $p(x)$. Similarly, let $(p(\tilde{x}) - p(x))_{U(x)}$ be the difference of the probability vectors projected onto the columns corresponding to the utility terms for menu x.We collect the inequalities using the matrix

$$Q = \begin{array}{c} \\ \vdots \\ (\tilde{x},x) \\ \vdots \end{array} \begin{array}{c|c} g_1 \cdots g_L & u_1^1 \cdots u_1^{W_1}\, u_2^1 \cdots u_A^{W_A} \\ \left[\begin{array}{c} \vdots \\ e_x - e_{\tilde{x}} \\ \vdots \end{array}\right. & \left.\begin{array}{c} \vdots \\ (p(\tilde{x}) - p(x))_{U(x)} \\ \vdots \end{array}\right] \end{array}.$$

Recall that $\mathcal{S} = \{(\tilde{x}, x) \in \{x^n\}_{n=1}^N \times \{x^n\}_{n=1}^N \mid p(\tilde{x}) \neq p(x)\}$. Thus, Q is an $|\mathcal{S}| \times \left(L + \sum_{a=1}^A W_a\right)$ dimensional matrix.

Now, we examine conditions when the inequalities from Theorem 1(iii) is infeasible. Using a version of the theorem of the alternative (see for example Border (2013) Corollary 15), the system inequalities from Theorem 1(iii) is infeasible if and only if

$$\{\lambda \in \mathbb{R}^{|\mathcal{S}|} \mid \lambda' Q = 0', 1'\lambda = 1, \lambda \geqslant 0\} \neq \varnothing.$$

Let $\pi \in \{\lambda \in \mathbb{R}^{|\mathcal{S}|} \mid \lambda' Q = 0, 1'\lambda = 1, \lambda \geqslant 0\}$. The vector π satisfies the definition of a second stage lottery that gives equal utility. Take the inner product of π with the columns of Q corresponding to the g_z terms, then for all $\hat{x} \in \{x^n\}_{n=1}^N$

$$\sum_{\{(\tilde{x},x)\in\mathcal{S} \mid p(\tilde{x})=p(\hat{x})\}} \pi_{(\tilde{x},x)} = \sum_{\{(\tilde{x},x)\in\mathcal{S} \mid p(x)=p(\hat{x})\}} \pi_{(\tilde{x},x)}.$$

Take the inner product of π with the columns of Q corresponding to the $u_a^{w_a}$ terms, then for all $\hat{x}_a \in \{\{x_a^n\}_{a=1}^A\}_{n=1}^N$

$$\sum_{\{(\tilde{x},x)\in\mathcal{S} \mid x_a=\hat{x}_a\}} \pi_{(\tilde{x},x)}(p_a(\tilde{x}) - p_a(x)) = 0.$$

Therefore, there is no rationalization by a strict PUM for $\{(x^n, p(x^n))\}_{n=1}^N$ if and only if we can find a distribution over $\{p(x^n)\}_{n=1}^N$ that satisfies the equalities in Theorem 5(ii). Take the contrapositive and we have Theorem 5(i) if and only if Theorem 5(ii). □

We note that this condition is essentially a version of Fishburn (1975) which imposes that we find a distribution where equalities hold in appropriate dimensions.

Appendix C Separable and Symmetric Cost Function

We now consider how to find rationalizations when there is characteristic information when the cost function is separable and symmetric. This characterization is closely related to the functional form studied in Fudenberg et al. (2015). A *kinked additive perturbed utility model* (kinked additive PUM) of discrete choice has the representation

$$p^*(x) = \operatorname*{argmax}_{p\in\Delta} \sum_{a=1}^{A} (p_a u_a(x_a) - c(p_a))$$

for some $u_a : \mathcal{X}_a \to \mathbb{R}$ functions that give the utility of characteristic values and $c : [0,1] \to \mathbb{R}$ is a strictly convex function that perturbs the expected utility. This differs from the *weak additive perturbed utility* model from Fudenberg et al. (2015) which imposes that the cost function is also continuously differentiable.

We characterize a kinked additive PUM using a *strict acyclicity* condition. The strict acyclicity condition checks for cycles of choice probabilities across alternatives and menus. For a dataset $\{(x^n, p(x^n))\}_{n=1}^N$, we say a finite sequence $\{(a_m, x[m]), (b_m, z[m])\}_{m=1}^M$ is *admissible* if (i) for all m that $a_m, b_m \in \{1, \ldots, A\}$ and $x[m], z[m] \in \{x^n\}_{n=1}^N$, (ii) the sequence $\{b_m\}_{m=1}^M$ is a permutation of $\{a_m\}_{m=1}^M$, (iii) the sequence $\{z_{b_m}[m]\}_{m=1}^M$ is a permutation of $\{x_{a_m}[m]\}_{m=1}^M$, and (iv) $\{z[m]\}_{m=1}^M$ is a permutation of $\{x[m]\}_{m=1}^M$.

In words, admissibility states that alternatives, characteristic values for each alternative, and menus show up the same number of times in both sequences from the observed dataset. The difference from the acyclicity condition from Fudenberg et al. (2015) is that we must account for alternatives and characteristic values. We now define strict acyclicity.

Definition 5. *A dataset $\{(x^n, p(x^n))\}_{n=1}^N$ satisfies* strict acyclicity *if there is no admissible sequence such that*

$$p_{a_m}(x[m]) > p_{b_m}(z[m]) \text{ for all } m \in \{1, \ldots, M\}$$

We consider when the dataset $\{(x^n, p(x^n))\}_{n=1}^N$ can be rationalized by a kinked additive PUM using sub-differential first order conditions of the Lagrangian.

Theorem 6. *Consider the dataset $\{(x^n, p(x^n))\}_{n=1}^N$. The following are equivalent:*

(i) $\{(x^n, p(x^n))\}_{n=1}^N$ is rationalized by a kinked additive perturbed utility model.

(ii) $\{(x^n, p(x^n))\}_{n=1}^N$ satisfies strict acyclicity.

(iii) There exist numbers $\{u_a^n\}_{n=1}^N$ for all $a \in \{1, \ldots, A\}$ and numbers $\{\lambda^n\}_{n=1}^N$, such that for all finite sequences $\{(a[m], x[m])\}_{m=1}^M$ with $x[m] \in \{x^n\}_{n=1}^N$, $a[m] \in \{1, \ldots, A\}$, and $p_{a[m]}(x[m]) \neq p_{a[m+1]}(x[m+1])$ for some m then

$$\sum_{m=1}^{M} (p_{a[m+1]}(x[m+1]) - p_{a[m]}(x[m]))(u_{a[m]}[m] + \lambda[m]) < 0$$

with $x[M+1] = x[1]$, $u_{a[m]}[m] = u_{a[m]}^r$ such that $x_{a[m]}[m] = x_{a[m]}^r$, and $\lambda[m] = \lambda^r$ such that $x[m] = x^r$;

and for all $a \in \{1, \ldots, A\}$ and for all $r, s \in \{1, \ldots, N\}$

$$u_a^r = u_a^s \quad \text{if} \quad x_a^r = x_a^s.$$

(iv) There exist $\{\{u_a^n\}_{a=1}^A\}_{n=1}^N$ *and* $\{\lambda^n\}_{n=1}^N$*, such that for all* $r, s \in \{1, \ldots, N\}$ *and* $a, b \in \{1, \ldots, A\}$

$$\text{if} \quad p_a(x^r) > p_b(x^s) \quad \text{then} \quad u_a^r + \lambda^r > u_b^s + \lambda^s$$

and for all $a \in \{1, \ldots, A\}$ *and for all* $r, s \in \{1, \ldots, N\}$

$$u_a^r = u_a^s \quad \text{if} \quad x_a^r = x_a^s.$$

First, we note that convexity is not innocuous for a kinked additive PUM. The implication (i) implies (iii) follows from sub-differential first order conditions. Next, (iii) implies (i) from a constructive procedure similar to Theorem 1. The equivalence of (iii) and (iv) is a property of monotonicity in one dimension. We have (iv) implies (ii) since if there were a strict cycle, then summing the cycle elements of (iv) would equal zero a contradiction. Lastly, (ii) implies (iv) from a rational Farkas lemma which is stated here.

Lemma 2. *Let* $k \in \mathbb{Q}^S$ *and* V *be a linear subspace of* $\mathbb{Q}^S$*. Exactly one of the following holds*

1. *Theres exists* $v \in V$ *such that* $v \leqslant k$*.*
2. *There exists* $w \in \mathbb{Q}_+^S$ *such that* $w \perp V$ *and* $\langle w, k \rangle < 0$*.*

Proof of Theorem 6. To show (i) implies (iii) we give sub-differential first order conditions for the Lagrangian. Consider the Lagrangian given by

$$\max_{p \in \mathbb{R}^A} \sum_{a=1}^{A} \left[u_a(x_a)p_a - c(p_a)\right] + \lambda(x)\left(\sum_{a=1}^{A} p_a - 1\right) + \sum_{a=1}^{A} \left[\mu_a^0(x)p_a + \mu_a^1(x)(1 - p_a)\right]$$

with $\mu_a^0, \mu_a^1 \geqslant 0$ for all $a \in \{1, \ldots A\}$. The terms $\lambda(x), \mu_a^0(x)$ and $\mu_a^1(x)$ are the multipliers on $\sum_{a=1}^A p_a = 1$, $p_a \leqslant 1$ and $p_a \geqslant 0$ respectively. Let $\partial c(\cdot)$ denote the subdifferential of the function $c(\cdot)$.[29] Karush-Kuhn-Tucker conditions require that for each a

$$0 \in u_a(x_a) - \partial c(p_a) + \lambda(x) + \mu_a^0(x) - \mu_a^1(x)$$

and complementary slackness conditions hold for each a so

$$\mu_a^0(x)p_a = \mu_a^1(x)(1 - p_a) = 0.$$

[29] The subdifferential of $c : [0, 1] \to \mathbb{R}$ at a point $\hat{p}_a \in (0, 1)$ is defined as $\partial c(\hat{p}_a) = \{z \in \mathbb{R} \mid c(p_a) - c(\hat{p}_a) \geqslant z(p_a - \hat{p}_a)\}$.

We re-write the first inequality so

$$u_a(x_a) + \lambda(x) + \mu_a^0(x) - \mu_a^1(x) \in \partial c(p).$$

Since c is strictly convex, the subgradient satisfies strict cyclic monotonicity. Since we assume the data are rationalized by a weak kinked PUM, for all finite sequences $\{(a[m], x[m])\}_{m=1}^M$ with $x[m] \in \{x^n\}_{n=1}^N$, $a[m] \in \{1, \ldots, A\}$, and $p_{a[m]}(x[m]) \neq p_{a[m+1]}(x[m+1])$ for some m, then

$$\sum_{m=1}^{M} (p_{a[m+1]}(x[m+1]) - p_{a[m]}(x[m])) \left[u_{a[m]}(x_a[m]) + \lambda(x[m]) + \mu^0_{a[m]}(x[m]) - \mu^1_{a[m]}(x[m])\right] < 0$$

where $x[M+1] = x[1]$.

However, the μ terms can be removed to consider strict cyclic monotonicity in $\lambda(x) + u_a(x_a)$. First, if $p_{a[m]}(x[m]) \in (0,1)$ then $\mu^0_{a[m]}(x[m]) = \mu^1_{a[m]}(x[m]) = 0$ by complementary slackness. If $p_{a[m]}(x[m]) = 1$, then $\mu^0_{a[m]}(x[m]) = 0$ and $\mu^1_{a[m]}(x[m]) \geqslant 0$ from complementary slackness. Extracting the term on $\mu^1_{a[m]}(x[m])$ for each term in the sum, we find that

$$(1 - p_{a[m+1]}(x[m+1]))\mu^1_{a[m]}(x[m]) \geqslant 0$$

since all probabilities are weakly less than one and the multiplier is non-negative. Therefore, we remove this term and still have strict cyclic monotonicity. Similarly, if $p_{a[m]}(x[m]) = 0$, then $\mu^0_{a[m]}(x[m]) \geqslant 0$ and $\mu^1_{a[m]}(x[m]) = 0$ from complementary slackness. Extracting the term on $\mu^0_{a[m]}(x[m])$ we find that

$$p_{a[m+1]}(x[m+1])\mu^0_{a[m]}(x[m]) \geqslant 0$$

since all probabilities are non-negative and the multiplier is non-negative. Therefore, we can remove the multiplier term while retaining strict cyclic monotonicity.

Thus, for all finite sequences $\{(a[m], x[m])\}_{m=1}^M$ with $x[m] \in \{x^n\}_{n=1}^N$, $a[m] \in \{1, \ldots, A\}$, and $p_{a[m]}(x[m]) \neq p_{a[m+1]}(x[m+1])$ for some m, then

$$\sum_{m=1}^{M} (p_{a[m+1]}(x[m+1]) - p_{a[m]}(x[m])) \left(u_{a[m]}(x_a[m]) + \lambda(x[m])\right) < 0$$

where $x[M+1] = x[1]$. Using the numbers from utility functions and Lagrange multipliers from the optimization procedure, we satisfy (iii).

We show (iii) implies (i) from a constructive procedure similar to the proof of Theorem 1. Let Σ be the set of all finite sequences $\{(a[m], x[m])\}_{m=1}^{M}$ with $x[m] \in \{x^n\}_{n=1}^N$, $a[m] \in \{1, \ldots, A\}$, and $p_{a[m]}(x[m]) \neq p_{a[m+1]}(x[m+1])$ for some m. From (iii) there exist numbers $\{u_a^n\}_{n=1}^N$ for all $a \in \{1, \ldots, A\}$ and numbers $\{\lambda^n\}_{n=1}^N$ such that

$$\sum_{m=1}^{M} (p_{a[m+1]}(x[m+1]) - p_{a[m]}(x[m]))(u_{a[m]}[m] + \lambda[m]) < 0$$

where $x[M+1] = x[1]$, $u_{a[m]}[m] = u^r_{a[m]}$ such that $x_{a[m]}[m] = x^r_{a[m]}$, and $\lambda[m] = \lambda^r$ such that $x[m] = x^r$. Since this is a strict inequality, there exists $\varepsilon_0 > 0$ small enough so that there exist numbers $\{u_a^n\}_{n=1}^N$ for all $a \in \{1, \ldots, A\}$ and numbers $\{\lambda^n\}_{n=1}^N$ such that

$$\sum_{m=1}^{M} (p_{a[m+1]}(x[m+1]) - p_{a[m]}(x[m]))(u_{a[m]}[m] + \lambda[m]) + \varepsilon_0 < 0 \tag{7}$$

for all sequences in Σ with $x[M+1] = x[1]$, $u_{a[m]}[m] = u^r_{a[m]}$ such that $x_{a[m]}[m] = x^r_{a[m]}$, and $\lambda[m] = \lambda^r$ such that $x[m] = x^r$.

Again, consider the function $f : \mathbb{R} \to \mathbb{R}$ from Matzkin and Richter (1991) given by $f(y) = (y^2 + T)^{1/2} - T^{1/2}$ for $T > 0$. In particular, $f(\cdot)$ is strictly convex, differentiable, $f(0) = 0$, $f(y) > 0$ if $y \neq 0$, and $\left[\frac{\partial f}{\partial y}(y)\right] < 1$ for all y. From Equation 7, there exists $\varepsilon > 0$ small enough so there exist numbers $\{u_a^n\}_{n=1}^N$ for all $a \in \{1, \ldots, A\}$ and numbers $\{\lambda^n\}_{n=1}^N$ such that

$$\sum_{m=1}^{M} (p_{a[m+1]}(x[m+1]) - p_{a[m]}(x[m]))(u_{a[m]}[m] + \lambda[m]) + \varepsilon f\left(p_{a[m+1]}(x[m+1]) - p_{a[m]}(x[m])\right) < 0 \tag{8}$$

where $x[M+1] = x[1]$, $u_{a[m]}[m] = u^r_{a[m]}$ such that $x_{a[m]}[m] = x^r_{a[m]}$, and $\lambda[m] = \lambda^r$ such that $x[m] = x^r$.

Next, consider the function $\phi_\sigma : [0, 1] \to \mathbb{R}$ for each sequence $\sigma \in \Sigma$ given by

$$\begin{aligned}\phi_\sigma(\tilde{p}) = &\sum_{m=1}^{M-1} (p_{a[m+1]}(x[m+1]) - p_{a[m]}(x[m]))(u_{a[m]}[m] + \lambda[m]) + \varepsilon f\left(p_{a[m+1]}(x[m+1]) - p_{a[m]}(x[m])\right) \\ &+ (\tilde{p} - p_{a[m]}(x[m]))(u_{a[m]}[m] + \lambda[m]) + \varepsilon f\left(\tilde{p} - p_{a[m]}(x[m])\right).\end{aligned}$$

Each $\phi_\sigma(\cdot)$ is strictly convex on the simplex since it is the sum of an affine and strictly convex function restricted to a convex domain.

Using the ϕ_σ functions, we use a constructive procedure from Rockafellar (1970) Theorem 24.8. First, choose an arbitrary $p_{a_0}(x^0) \in \{\{p_a(x^n)\}_{a=1}^A\}_{n=1}^N$ and let Σ_0 be the set of sequences which begin with $p_{a_0}(x^0)$. Next, define a function $c : [0,1] \to \mathbb{R}$ given by

$$c(\tilde{p}) = \max_{\sigma_0 \in \Sigma_0} \{\phi_{\sigma_0}(\tilde{p})\}.$$

$c(\cdot)$ is defined as a max of strictly convex functions, and so it is strictly convex.

All that remains is to show that the numbers $\{\{u_a^n\}_{a=1}^A\}_{n=1}^N$ and $\{\lambda^n\}_{n=1}^N$ used to satisfy Equation 4 and the $c(\cdot)$ function rationalize the data $\{(x^n, p(x^n))\}_{n=1}^N$. To show this result, let $\sigma_{0,a}^n \in \Sigma_0$ be the sequence where each $c(p_a(x^n))$ achieves the maximum. For $p \neq p(x^n)$ and $p \in \Delta$ we have

$$\begin{aligned}
\sum_{a=1}^A (u_a^n p_a - c(p_a)) &= \sum_{a=1}^A \left(u_a^n p_a - \max_{\sigma_0 \in \Sigma_0} \{\phi_{\sigma_0}(p_a)\} \right) \\
&\leqslant \sum_{a=1}^A \left(u_a^n p_a - (p_a - p_a(x^n)) \left[u_a^n + \lambda^n \right] + \varepsilon f(p_a - p_a(x^n)) + \max_{\sigma_{0,a}^n \in \Sigma_0} \{\phi_{\sigma_0}(p_a(x^n))\} \right) \\
&= \sum_{a=1}^A \left(u_a^n p_a(x^n) + \lambda^n (p_a - p_a(x^n)) + \varepsilon f(p_a - p_a(x^n)) + \max_{\sigma_{0,a}^n \in \Sigma_0} \{\phi_{\sigma_0}(p_a(x^n))\} \right) \\
&= \sum_{a=1}^A \left(u_a^n p_a(x^n) + \varepsilon f(p - p_a(x^n)) + \max_{\sigma_{0,a}^n \in \Sigma_0} \{\phi_{\sigma_0}(p_a(x^n))\} \right) \\
&< \sum_{a=1}^A \left(u_a^n p_a(x^n) + \max_{\sigma_{0,a}^n \in \Sigma_0} \{\phi_{\sigma_0}(p_a(x^n))\} \right) \\
&= \sum_{a=1}^A \left(u_a^n p_a(x^n) + c(p_a(x^n)) \right).
\end{aligned}$$

The first inequality comes by choosing the sequence for each a that ends with $p_a(x^n)$ and begins with the largest cost sequence for $p_a(x^n)$. The second equality follows by rearrangement. The third equality follows since the multiplier is the same and probabilities sum to one. The second inequality follows from $\varepsilon f(p_a - p_a(x^n)) > 0$ for at least one a. Now by the second set of equalities on the number $\{u_a^n\}_{n=1}^n$ for all $a \in \{1, \ldots, A\}$, we can create functions for utilities over characteristic values. One can extend the utility functions to $\mathcal{X}$ by choosing arbitrary numbers for unobserved values.

Last, we show that (ii) implies (iv). Let Q^* be the vector space over the field of rational numbers whose coordinates correspond to pairs $(\nu, \xi) = \{(a_\nu, x^\nu), (b_\xi, x^\xi)\}$ with $p_{a_\nu}(x^\nu) > p_{b_\xi}(x^\xi)$. We will use a vector w to count the number of times a relation appears. There is an admissible strict cycle when the comparisons made for each alternative, characteristic values, and menu can appear on the ν and ξ sides the same number of times. Define $k \in \mathbb{Q}^*$ as the vector of all entries equal to negative one. For $w \in \mathbb{Q}^*$, $\langle w, k \rangle < 0$ if and only if at least one comparison in a collection of probability comparisons is strict. This conditions is automatically satisfied unless all choice distributions are uniform over alternatives.

For each $a \in \{1, \ldots, A\}$ let $\hat{x}_a \in \{x_a^n\}_{n=1}^N$, define $v^{\hat{x}_a} \in \mathbb{Q}^*$ as

$$v^{\hat{x}_a}(\nu, \xi) = \begin{cases} -1 & \text{if} \quad x_{a_\nu}^\nu = \hat{x}_a \quad \text{and} \quad x_{b_\xi}^\xi \neq \hat{x}_a \\ 1 & \text{if} \quad x_{a_\nu}^\nu \neq \hat{x}_a \quad \text{and} \quad x_{b_\xi}^\xi = \hat{x}_a \\ 0 & \text{if} \quad \text{otherwise.} \end{cases}$$

Now, $w \perp v^{\hat{x}_{\hat{a}}}$ if and only if $\hat{x}_{\hat{a}}$ is included equally many times on the ν and ξ sides. Moreover, if the collection of $w \perp \{v^{\hat{x}_{\hat{a}}}\}_{\hat{x}_a \in \{x_a^n\}_{n=1}^N}$ then the alternative a shows up equally many times on the ν and ξ sides.

For each $\hat{x} \in \{x^n\}_{n=1}^N$, define $v^{\hat{x}} \in \mathbb{Q}^*$ as

$$v^{\hat{x}}(\nu, \xi) = \begin{cases} -1 & \text{if} \quad x^\nu = \hat{x} \quad \text{and} \quad x^\xi \neq \hat{x} \\ 1 & \text{if} \quad x^\nu \neq \hat{x} \quad \text{and} \quad x^\xi = \hat{x} \\ 0 & \text{if} \quad \text{otherwise} \end{cases}$$

Similarly, $w \perp v^{\hat{x}}$ if and only if $\hat{x}$ is included equally many times on the ν and ξ sides. Let $V = \{\{v^{\hat{x}_a}\}_{\hat{x}_a \in \{x_a^n\}_{n=1}^N}\}_{a=1}^A \cup \{v^{\hat{x}}\}_{\hat{x} \in \{x^n\}_{n=1}^N}$. Thus, $w \in \mathbb{Q}^*$ represents a strict cycle if and only if $w \perp V$ and $\langle w, k \rangle < 0$.

Since strict acyclicity holds, we have $w \perp V$ and $\langle w, k \rangle < 0$, so by Lemma 2 there exists $v \in V$ such that $v \leqslant k$. This means there exists numbers $\{\{u_a(\hat{x}_a)\}_{\hat{x}_a \in \{x_a^n\}_{n=1}^N}\}_{a=1}^A$ and $\{\lambda(x^n)\}_{n=1}$ in $\mathbb{Q}$ such that

$$v = \sum_{a=1}^{A} \sum_{\hat{x}_a \in \{x_a^n\}_{n=1}^N} u_a(\hat{x}_a) v^{\hat{x}_a} + \sum_n \lambda(x^n) v^{x^n}$$

which enforces the utility for each alternative to be unique at observed characteristic values. Moreover, if $p_a(x^r) > p_b(x^s)$ then $v((a, x^r), (b, x^s)) = -u_a(x^r) - \lambda(x^r) + u_b(x^s) + \lambda(x^s) < k((a, x^r), (b, x^s)) = -1$. Thus, $u_a(x^r) + \lambda(x^r) > u_b(x^s) + \lambda(x^s)$ and (iv) is satisfied.

□

For a kinked additive PUM with fully nonparametric utility over characteristics, the conditions make direct comparisons across alternatives with different choice probabilities and utilities. Thus, unlike a general strict PUM, choosing an index is innocuous for kinked additive PUMs with fully nonparametric utility. The equivalence of (i), (iii), and (iv) holds even if we assume that $u_a(x_a) = \sum_{j=1}^{d_a} u_{a,j}(x_{a,j})$ and $u_a(x_a) = \beta_a \cdot x_a$. In these cases, the choice of indexing alternatives is not innocuous due to potentially different characteristics. A strict acyclicity condition can be shown equivalent when $u_a(x_a)$ is separable in characteristics since the weightings on the utility numbers are integers. Alternatively, when $u_a(x_a)$ is linear there is a weighted cycles condition similar to Theorem 1(v). Currently, we are researching if the weighted cycles condition can be converted to an acyclicity condition. We are also currently working on examining data using Theorem 6(iv).

Appendix D Differentiable Cost Functions

We consider placing conditions on the differentiability of the cost functions. In this case, imposing differentiability has a behavioral interpretation. Without differentiability, the cost function is kinked so there can be utility changes from characteristic values without a change in the choice distribution. This behavior relates to the notion of *just noticeable differences*. For example, if utility is linear over characteristic values, there may be coarse levels of perception where an individual has the same behavior for characteristic values in some range of characteristic values. For example, consider "grains of sugar" as a characteristic when choosing between coffee and tea. Many individuals treat a cup of coffee with no sugar the same as cup of coffee with one grain of sugar. Therefore, an individual may have the same choice probabilities for a range of characteristic values. However, individuals often distinguish a cup of coffee with no sugar and a cup of coffee with a packet of sugar. If differentiability is imposed on the cost function with linear utility over characteristic values, then a small change in characteristic values causes a small change choice probabilities which rules out behavior associated with just noticeable differences.

Now, we desire the cost function to be continuously differentiable, so the subgradient of the cost function has unique utility values at each choice distribution. This implication imposes conditions similar to the strong version of the strong axiom of revealed preference from Chiappori and Rochet (1987). Therefore, one can follow the proofs of Theorem 1 and Theorem 6 imposing this additional constraint and apply the convolution methods of Chiappori and Rochet (1987) to get an infinitely differentiable cost function. We present the results for nonparametric utility over characteristics without proof. We drop the results on cycles conditions to avoid additional definitions.

Theorem 7. *Consider the dataset* $\{(x^n, p(x^n))\}_{n=1}^N$. *The following are equivalent:*

(i) $\{(x^n, p(x^n))\}_{n=1}^N$ *is rationalized by an infinitely differentiable strict PUM.*

(ii) There exist utility functions $u_a : \mathcal{X}_a \to \mathbb{R}$ *for all* $a \in \{1, \ldots, A\}$ *and a continuously differentiable function* $g : \Delta \to \mathbb{R}$ *such that* $p(x^n)$ *is the unique argmax from* $\max_{p\in\Delta} \sum_{a=1}^A p_a u_a(x_a^n) - g(p)$ *for all* $n \in \{1, \ldots, N\}$.

(iii) There exist numbers $\{u_a^n\}_{n=1}^N$ *for all* $a \in \{1, \ldots, A\}$ *and* $\{g^n\}_{n=1}^N$, *such that for all* $(s, r) \in \{1, \ldots, N\} \times \{1, \ldots, N\}$ *with* $p(x^s) \neq p(x^r)$ *then*

$$\sum_{a=1}^A p_a(x^s)u_a^r - g^s < \sum_{a=1}^A p_a(x^r)u_a^r - g^r$$

and for all $r, s \in \{1, \ldots, N\}$

$$\begin{aligned} u_a^r &= u_a^s \quad \text{if} \quad x_a^r = x_a^s \\ g^r &= g^s \quad \text{if} \quad p(x^r) = p(x^s) \\ \text{for all } a \in \{1, \ldots, A\} \quad u_a^r &= u_a^s \quad \text{if} \quad p(x^r) = p(x^s). \end{aligned}$$

(iv) There exist numbers $\{u_a^n\}_{n=1}^N$ *for all* $a \in \{1, \ldots, A\}$ *such that for all finite sequences* $\{x[m]\}_{m=1}^M$ *where all* $x[m] \in \{x^n\}_{n=1}^N$ *and* $p(x[m]) \neq p(x[m+1])$ *for some* m

$$\sum_{m=1}^M p(x[m+1]) \cdot U[m] < \sum_{m=1}^M p(x[m]) \cdot U[m]$$

where $x[M+1] = x[1]$ *and* $U[m] = (u_1[m], \ldots, u_A[m])$ *where* $u_a[m]$ *is the* u_a^r *term associated to* $x_a[m] = x_a^r$;
and for all $r, s \in \{1, \ldots, N\}$ *and for all* $a \in \{1, \ldots, A\}$

$$\begin{aligned} u_a^r &= u_a^s \quad \text{if} \quad x_a^r = x_a^s \\ \text{for all } a \in \{1, \ldots, A\} \quad u_a^r &= u_a^s \quad \text{if} \quad p(x^r) = p(x^s). \end{aligned}$$

Theorem 8. *Consider the dataset* $\{(x^n, p(x^n))\}_{n=1}^N$. *The following are equivalent:*

(i) $\{(x^n, p(x^n))\}_{n=1}^N$ *is rationalized by an infinitely differentiable additive perturbed utility model.*

(ii) There exist numbers $\{u_a^n\}_{n=1}^N$ *for all* $a \in \{1, \ldots, A\}$ *and numbers* $\{\lambda^n\}_{n=1}^N$*, such that for all finite sequences* $\{(a[m], x[m])\}_{m=1}^M$ *with* $x[m] \in \{x^n\}_{n=1}^N$*,* $a[m] \in \{1, \ldots, A\}$*, and* $p_{a[m]}(x[m]) \neq p_{a[m+1]}(x[m+1])$ *for some* m *then*

$$\sum_{m=1}^{M} (p_{a[m+1]}(x[m+1]) - p_{a[m]}(x[m]))(u_{a[m]}[m] + \lambda[m]) < 0$$

with $x[M+1] = x[1]$*,* $u_{a[m]}[m] = u^r_{a[m]}$ *such that* $x_{a[m]}[m] = x^r_{a[m]}$*, and* $\lambda[m] = \lambda^r$ *such that* $x[m] = x^r$*;*
and for all $a \in \{1, \ldots, A\}$ *and for all* $r, s \in \{1, \ldots, N\}$

$$\begin{aligned} & u_a^r = u_a^s \quad \text{if} \quad x_a^r = x_a^s \\ \text{for all } a \in \{1, \ldots, A\} \quad & u_a^r = u_a^s \quad \text{if} \quad p(x^r) = p(x^s). \end{aligned}$$

(iii) There exist $\{\{u_a^n\}_{a=1}^A\}_{n=1}^N$ *and* $\{\lambda^n\}_{n=1}^N$*, such that for all* $r, s \in \{1, \ldots, N\}$ *and* $a, b \in \{1, \ldots, A\}$

$$\text{if} \quad p_a(x^r) > p_b(x^s) \quad \text{then} \quad u_a^r + \lambda^r > u_b^s + \lambda^s$$

and for all $a \in \{1, \ldots, A\}$ *and for all* $r, s \in \{1, \ldots, N\}$

$$\begin{aligned} & u_a^r = u_a^s \quad \text{if} \quad x_a^r = x_a^s \\ \text{for all } a \in \{1, \ldots, A\} \quad & u_a^r = u_a^s \quad \text{if} \quad p(x^r) = p(x^s). \end{aligned}$$

Appendix E Deterministic Choice

We examine when a dataset consists only of deterministic choice. Choices are deterministic when the choice distributions consist of zeros and ones. Thus, the dataset $\{(x^n, p(x^n))\}_{n=1}^N$ is deterministic when every $p(x^n)$ has some $a \in \{1, \ldots, A\}$ such that $p_a(x^n) = 1$. We note that Fudenberg et al. (2015) provide conditions for deterministic choice, so we begin by studying conditions to rationalize deterministic data with a kinked additive PUM with nonparametric utility over characteristics. We focus our study looking on acyclicity conditions that formalize kinked additive PUMs.

For deterministic choice data, one only needs to look for cycles over alternatives or cycles over menus to refute the model. We begin by defining strict item acyclicity. This condition says we cannot have preference cycles over alternatives with different characteristic values.

Definition 6. *A dataset* $\{(x^n, p(x^n))\}_{n=1}^N$ *satisfies* strict item acyclicity *if all finite sequences* $\{(a[m], x[m])\}_{m=1}^M$ *with* $x[m] \in \{x^n\}_{n=1}^N$*,* $a[m] \in \{1, \ldots, A\}$*, and such that* $x_{a[m+1]}[m] = x_{a[m+1]}[m+1]$ *for all* $m = 1, \ldots, M-1$ *and* $x_{a[1]}[1] = x_{a[1]}[M]$ *satisfy*

$$p_{a[1]}(x[1]) > p_{a[2]}(x[1]), \ldots, p_{a[M-1]}(x[M-1]) > p_{a[M]}(x[M-1]) \quad \textit{implies} \quad p_{a[M]}(x[M]) \ngtr p_{a[1]}(x[M]).$$

Alternatively, one could look for cycles holding the alternatives fixed in comparisons while varying the menu. In this case, we arrive at the following definition of strict menu acyclicity.

Definition 7. *A dataset* $\{(x^n, p(x^n))\}_{n=1}^N$ *satisfies* strict menu acyclicity *if all finite sequences* $\{(a[m], x[m])\}_{m=1}^M$ *with* $x[m] \in \{x^n\}_{n=1}^N$*,* $a[m] \in \{1, \ldots, A\}$*, and such that* $x_{a[m]}[m] = x_{a[m]}[m+1]$ *for all* $m = 1, \ldots, M-1$ *and* $x_{a[M]}[1] = x_{a[M]}[M]$*.*

$$p_{a[1]}(x[1]) > p_{a[1]}(x[2]), \ldots, p_{a[M-1]}(x[M-1]) > p_{a[M-1]}(x[M]) \quad \textit{implies} \quad p_{a[M]}(x[M]) \ngtr p_{a[M]}(x[1]).$$

Looking for either a strict item cycle or a strict menu cycle refutes deterministic choice. In this case, the indexing of alternatives places no additional structure on behavior. This result again highlights that if there is no structure on how characteristics enter utility, we return to standard models of decision theory. The following proposition can be deduced from Fudenberg et al. (2015), but we provide details of the result below.

Proposition 2. *Assume that the dataset* $\{(x^n, p(x^n))\}_{n=1}^N$ *is deterministic. The following conditions are equivalent:*

1. $\{(x^n, p(x^n))\}_{n=1}^N$ *satisfies strict item acyclicity.*
2. $\{(x^n, p(x^n))\}_{n=1}^N$ *satisfies strict menu acyclicity.*
3. $\{(x^n, p(x^n))\}_{n=1}^N$ *satisfies strict acyclicity.*
4. *There exists an injective function* $v : \bigcup_{a=1}^A \{x_a^n\}_{n=1}^N \to \mathbb{R}$ *such that* $p_a(x^n) = 1$ *if and only if* $v(x_a^n) = \max_{x_a \in \bigcup_{a=1}^A \{x_a^n\}} v(x_a^n)$*.*

Proof. First, we show (i) if and only if (iv). When the $\{p(x^n)\}_{n=1}^N$ are deterministic, we define a single valued choice function $K : \{x^n\}_{n=1}^N \to \bigcup_{a=1}^A \{x_a^n\}_{n=1}^N$ which takes menus to a chosen alternative from the menu. Thus, $K(x) = x_a$ when $p_a(x^n) = 1$. Thus, a dataset satisfies strict item acyclicity if and only if there is no sequence

$$x_{a[1]}[1] = K(x[1]) \neq x_{a[2]}[1], x_{a[2]}[2] = K(x[2]) \neq x_{a[3]}, \ldots$$
$$x_{a[M]}[M] = K(x[M]) \neq x_{a[1]}[M]$$

where $x_{a[m+1]}[m] = x_{a[m+1]}[m+1]$ for all $m = 1, \ldots, M-1$ and $x_{a[1]}[1] = x_{a[1]}[M]$. Thus strict item acyclicity is equivalent to, the congruence axiom from Richter (1966). As shown by Richter (1966), congruence is equivalent to the existence of a preference relation over $\bigcup_{a=1}^{A}\{x_a^n\}_{n=1}^N$ such that for each $x \in \mathcal{X}$, $K(x)$ is the set of most preferred elements. Since $\bigcup_{a=1}^{A}\{x_a^n\}_{n=1}^N$ is finite and K is single valued, this is equivalent to a strict utility function over $\bigcup_{a=1}^{A}\{x_a^n\}_{n=1}^N$ that rationalizes the choice function K.

Next, we show (iv) implies (iii). Let v be an injective function such that $p_a(x) = 1$ if $v(x_a^n) = \max_{x_a \in \bigcup_{a=1}^{A}\{x_a^n\}} v(x_a^n)$. If strict acyclicity is violated, then there is an admissible sequence such that

$$p_{a[m]}(x[m]) > p_{b[m]}(z[m]) \text{ for all } m \in \{1, \ldots, M\}.$$

For this sequence, pick an arbitrary $x_{a[m]}[m]$. By admissibility (iii), there is an element $z_{b[\tilde{m}]}[\tilde{m}] = x_{a[m]}[m]$. Since $\{p(x^n)\}_{n=1}^N$ is deterministic, we can take all comparisons in the permutation from m to $\tilde{m}$ and conclude that $v_{a[m]}(x_{a[m]}[m]) > v_{b[\tilde{m}]}(z_{b[\tilde{m}]}[\tilde{m}]) = v_{a[m]}(x_{a[m]}[m])$. However, this contradicts the strict ordering of utilities.

Note that (iii) implies (ii) by fixing the appropriate elements in a cycle. Lastly (ii) implies (i). Suppose that strict item acyclicity is violated by the sequence

$$p_{a[1]}(x[1]) > p_{a[2]}(x[1]), \ldots, p_{a[M-1]}(x[M-1]) > p_{a[M]}(x[M-1]) \quad \text{implies} \quad p_{a[M]}(x[M]) > p_{a[1]}(x[M])$$

such that $x_{a[m+1]}[m] = x_{a[m+1]}[m+1]$ for all $m = 1, \ldots, M-1$ and $x_{a[1]}[1] = x_{a[1]}[M]$. However, then

$$0 = p_{a[m+1]}(x[m]) < p_{a[m+1]}(x[m+1]) = 1 \quad \text{for all} \quad m = 1, \ldots, M-1$$

and $x_{a[m+1]}(x[m]) = x_{a[m+1]}(x[m+1])$. In addition, $0 = p_{a[1]}(x[M]) < p_{a[1]}(x[1]) = 1$ with $x_{a[1]}[M] = x_{a[1]}[1]$ so this is a menu cycle.

□

Appendix F Implementation of Tests

To analyze a dataset $\{(x^n, p(x^n))\}_{n=1}^{N}$, we examine whether the data can be described by a strict PUM. We operationalize checking for rationalization by a strict PUM using the inequalities from Theorem 1(iii).[30] We focus on the following five specifications of utilities over characteristics: Nonparametric $u_a(x_a)$, additively separable $\sum_{j=1}^{d} u_{a,j}(x_{a,j})$, additively separable and independent of list position $\sum_{j=1}^{d} u_j(x_{a,j})$, linear $\beta_a \cdot x_a$, and linear and independent of position $\beta \cdot x_a$. First, consider testing strict PUM with nonparametric utility over characteristics. Let Q be matrix generated by Theorem 1(iii) for the dataset $\{(x^n, p(x^n))\}_{n=1}^{N}$ which places restrictions on the vector of unknowns, U, associated to $\left\{\{u_a^n\}_{a=1}^{A}, g^n\right\}_{n=1}^{N}$ after imposing the equality conditions on u. By Theorem 1(iii), the dataset is rationalized by a strict PUM if $\{U \mid QU < 0\} \neq \varnothing$. Using a theorem of the alternative (see for example Border (2013) Corollary 15),

$$\{U \mid QU < 0\} = \varnothing \quad \Leftrightarrow \quad \{\lambda \mid Q'\lambda = 0, \mathbf{1}'\lambda = 1, \lambda \geqslant 0\} \neq \varnothing.$$

We check whether the data are described by a strict PUM by examining if solutions exist to the quadratic program

$$\begin{aligned} \min_{\lambda} \quad & \sum_{i=1}^{r_Q} \lambda_i^2 \\ \text{s.t.} \quad & Q'\lambda = 0 \\ & \mathbf{1} \cdot \lambda = 1 \\ & \lambda \geqslant 0, \end{aligned}$$

where r_Q is the number of rows of Q and $\mathbf{1}$ is a vector of ones. If solutions exist to the above problem, then the dual system is non-empty and there are no utility numbers which rationalize the data. Let the solutions be denoted λ_i^*. From the formulation of the problem, λ_i^* is strictly greater than zero only in the presence of violations. As the number of violations increases, there are more $\lambda_i^* \in (0, 1]$ and the minimum decreases. Therefore, one could use the optimal value to measure violations of rationality, where heuristically a smaller value means "less rational".

Similarly, we construct matrices of restrictions for separable and linear utilities using

[30] We could have applied Theorem 1(iv) when testing rationality by fixing the length of the sequence. Fixing the length to be at most two or three, we find numbers with pass rates which always exceed those of the full test. We find that these differences are at most approximately a 6% difference.

Theorem 3(iii) and Theorem 4(iii). We perform a similar procedure when utility over characteristics is independent of an alternative. We refer to these matrices as Q without loss of generality. We also test intuitive monotonicity restrictions on linear utility parameters (e.g. $\beta_{price} < 0$). In this case, let C be the matrix that generates the monotonicity constraints. We denote the matrix that generates rationality and monotonicity restrictions by

$$\tilde{Q} = \begin{bmatrix} Q \\ C \end{bmatrix}.$$

We jointly test rationality and monotonicity restrictions by replacing Q with $\tilde{Q}$ in the quadratic program. However, now there are λ_i terms associated with monotonicity violations. We can operationalize the test of kinked additive PUM by checking the strict monotonicity conditions imposed by Theorem 6(iv). To run this test with additively separable utility or linear utility over characteristics, we can run the test imposing the additional restrictions on utility numbers.

Appendix G Descriptive Statistics

We provide tables of individual descriptive statistics for the raw data and the purchase data. The differences between the raw and purchase datasets are small, so we focus on the differences as the number of alternatives changes. Regardless of the number of alternatives, the average age is approximately 40 years old. The three and four alternative datasets have slightly more women, while the five alternative dataset has slightly more men. The three and five alternative datasets have similar distributions of wealth and are slightly wealthier than the four alternative dataset.

Table 7: Purchase Data Descriptive Statistics

	Number of Alternatives		
	3	4	5
Age (Years)	41.3	40.6	39.8
Male	46.4%	45.3%	57.0%
Income (AUD)			
$0-51,999	30.6%	44.8%	32.6%
$52,000-103,999	46.0%	42.1%	43.4%
$104,000 or above	23.4%	13.1%	24.0%
Respondents	222	221	221

Table 8: Raw Data Descriptive Statistics

	Number of Alternatives		
	3	4	5
Age (Years)	41.4	40.3	39.9
Male	46.1%	46.3%	57.1%
Income (AUD)			
$0-51,999	31.3%	45.4%	33.2%
$52,000-103,999	45.7%	41.9%	42.5%
$104,000 or above	23%	12.8%	24.3%
Respondents	230	227	226

Appendix H Empirical Analysis: Strict PUM

H.1 Purchase Data

Here we present the rationalization analysis of strict PUMs for datasets from lists with three and five flight after restricting the dataset to flights the individual would actually purchase. We often refer to the analysis with three or five alternatives, so the indexing by list position is implicit. For three and five alternatives, the raw pass rates are similar lists with four alternatives. We note that pass rates slightly decrease for many specifications as the size of the list increases increase. Next, we examine the fraction of sets of relevant characteristics that have an MPS above the threshold 0.10 for at least one utility specification. For three alternatives, only the sets Brand and Brand & Time fail to pass the threshold for basic MPS and only Brand & Time fails to pass the threshold adaptive MPS. For five alternatives, the set of relevant characteristics Brand and Brand & Time fail to pass the threshold for basic and adaptive MPS. These results are similar to those for the four alternative survey.

Next, we examine if a linear utility model has more descriptive power from the sets of relevant characteristics that pass the 0.10 MPS threshold and excluding the specification with only Brand. We exclude Brand alone since many of these tests have identical numbers. For three alternatives, a linear utility over characteristics has the highest basic MPS for 5/6 cases and highest adaptive MPS for 5/6 cases. For five alternatives, a linear utility over characteristics has the highest basic MPS for 5/6 cases and the highest adaptive MPS for 4/6 cases.

We next consider which specification has the most descriptive power after performing the correction. For three alternatives, the highest basic MPS specification is a position dependent linear utility model with Price & Time & Brand and the highest adaptive MPS is a position independent linear utility for the Full set of characteristics. For five alternatives, a position independent linear utility for only Price yields the highest basic and adaptive MPS among all menus.

For the monotonicity tests with linear utility, there are only modest drops in pass rates for the three and five alternative case. The largest difference occurs for the specification with only Time. The decrease for the three and five alternative datasets are approximately 14% and 15%, respectively. We see that position independent utility experiences little change to MPS, while the position dependent utility MPS experiences changes that can be up to approximately 0.36. For three alternatives, the highest basic and adaptive MPS is for the

Full specification of characteristics with position dependent utility. In contrast, for five alternatives the highest basic and adaptive MPS is for the specification with only Price and position independent utility. This along with the information in the last paragraph is some evidence that as list size increases individuals consider a smaller number of characteristics.

Table 9: Pass Rates for Three Alternatives

	$u_a(x_a)$	$\sum_{j=1}^{d} u_{a,j}(x_{a,j})$	$\sum_{j=1}^{d} u_j(x_{a,j})$	$\beta_a \cdot x_a$	$\beta \cdot x_a$	S
Price	0.806	0.806	0.797	0.788	0.378	222
Brand	0.757	0.757	0.622	0.757	0.626	222
Time	0.748	0.748	0.662	0.676	0.243	222
Price & Brand	0.901	0.613	0.414	0.532	0.297	222
Price & Time	0.914	0.703	0.518	0.518	0.293	222
Brand & Time	0.865	0.252	0.099	0.176	0.045	222
Price & Time & Brand	0.982	0.815	0.635	0.725	0.428	222
Full	0.982	0.874	0.757	0.829	0.622	222

Table 10: Measure of Predictive Success for Three Alternatives: Basic

	$u_a(x_a)$	$\sum_{j=1}^{d} u_{a,j}(x_{a,j})$	$\sum_{j=1}^{d} u_j(x_{a,j})$	$\beta_a \cdot x_a$	$\beta \cdot x_a$	S
Price	0.046	0.046	0.271	0.247	0.348	222
Brand	0.004	0.004	0.076	0.004	0.081	222
Time	-0.004	-0.004	0.148	0.151	0.220	222
Price & Brand	0.037	0.419	0.272	0.443	0.265	222
Price & Time	0.045	0.532	0.460	0.472	0.271	222
Brand & Time	0.005	0.058	-0.044	0.088	0.012	222
Price & Time & Brand	-0.011	0.501	0.469	0.613	0.389	222
Full	-0.011	0.438	0.577	0.588	0.562	222

Table 11: Measure of Predictive Success for Three Alternatives: Adaptive

	$u_a(x_a)$	$\sum_{j=1}^{d} u_{a,j}(x_{a,j})$	$\sum_{j=1}^{d} u_j(x_{a,j})$	$\beta_a \cdot x_a$	$\beta \cdot x_a$	S
Price	0.050	0.050	0.307	0.249	0.332	222
Brand	0.007	0.007	0.116	0.007	0.122	222
Time	-0.005	-0.005	0.182	0.150	0.204	222
Price & Brand	0.057	0.345	0.338	0.389	0.260	222
Price & Time	0.045	0.422	0.435	0.439	0.268	222
Brand & Time	0.015	-0.015	0.023	0.033	0.008	222
Price & Time & Brand	0.048	0.343	0.473	0.441	0.381	222
Full	0.050	0.315	0.498	0.389	0.539	222

Table 12: Linear Monotonicity Results for Three Alternatives

	Pass Rates		Basic MPS		Adaptive MPS	
	$\beta_a \cdot x_a$	$\beta \cdot x_a$	$\beta_a \cdot x_a$	$\beta \cdot x_a$	$\beta_a \cdot x_a$	$\beta \cdot x_a$
Price	0.748	0.374	0.553	0.352	0.566	0.343
Time	0.541	0.225	0.349	0.205	0.369	0.196
Price & Brand	0.514	0.297	0.458	0.268	0.439	0.265
Price & Time	0.428	0.284	0.398	0.269	0.394	0.266
Brand & Time	0.113	0.045	0.054	0.015	0.035	0.013
Price & Time & Brand	0.644	0.405	0.573	0.373	0.520	0.368
Full	0.788	0.541	0.682	0.502	0.570	0.493

Table 13: Pass Rates for Five Alternatives

	$u_a(x_a)$	$\sum_{j=1}^{d} u_{a,j}(x_{a,j})$	$\sum_{j=1}^{d} u_j(x_{a,j})$	$\beta_a \cdot x_a$	$\beta \cdot x_a$	S
Price	0.765	0.765	0.679	0.733	0.584	221
Brand	0.756	0.756	0.520	0.756	0.516	221
Time	0.765	0.765	0.561	0.679	0.308	221
Price & Brand	0.842	0.647	0.308	0.593	0.285	221
Price & Time	0.873	0.701	0.371	0.443	0.199	221
Brand & Time	0.842	0.348	0.081	0.253	0.081	221
Price & Time & Brand	0.896	0.719	0.403	0.652	0.312	221
Full	0.914	0.796	0.579	0.760	0.466	221

Table 14: Measure of Predictive Success for Five Alternatives: Basic

	$u_a(x_a)$	$\sum_{j=1}^{d} u_{a,j}(x_{a,j})$	$\sum_{j=1}^{d} u_j(x_{a,j})$	$\beta_a \cdot x_a$	$\beta \cdot x_a$	S
Price	-0.018	-0.018	0.170	0.018	0.479	221
Brand	-0.025	-0.025	-0.027	-0.025	-0.026	221
Time	-0.021	-0.021	0.057	-0.040	0.211	221
Price & Brand	-0.082	0.046	0.144	0.415	0.223	221
Price & Time	-0.048	0.129	0.298	0.345	0.149	221
Brand & Time	-0.079	-0.254	-0.084	0.076	0.020	221
Price & Time & Brand	-0.041	0.212	0.227	0.439	0.246	221
Full	-0.070	0.257	0.393	0.340	0.387	221

Table 15: Measure of Predictive Success for Five Alternatives: Adaptive

	$u_a(x_a)$	$\sum_{j=1}^{d} u_{a,j}(x_{a,j})$	$\sum_{j=1}^{d} u_j(x_{a,j})$	$\beta_a \cdot x_a$	$\beta \cdot x_a$	S
Price	0.007	0.007	0.277	0.044	0.506	221
Brand	0.002	0.002	0.079	0.002	0.089	221
Time	0.006	0.006	0.171	-0.010	0.239	221
Price & Brand	0.005	0.275	0.237	0.423	0.228	221
Price & Time	0.052	0.382	0.300	0.353	0.156	221
Brand & Time	0.016	-0.013	0.013	0.091	0.028	221
Price & Time & Brand	0.010	0.260	0.312	0.351	0.251	221
Full	0.020	0.262	0.461	0.306	0.387	221

Table 16: Linear Monotonicity Results for Five Alternatives

	Pass Rates		Basic MPS		Adaptive MPS	
	$\beta_a \cdot x_a$	$\beta \cdot x_a$	$\beta_a \cdot x_a$	$\beta \cdot x_a$	$\beta_a \cdot x_a$	$\beta \cdot x_a$
Price	0.701	0.584	0.290	0.517	0.373	0.530
Time	0.529	0.290	0.122	0.225	0.216	0.242
Price & Brand	0.561	0.281	0.445	0.225	0.477	0.230
Price & Time	0.353	0.167	0.286	0.130	0.299	0.133
Brand & Time	0.208	0.081	0.094	0.028	0.126	0.032
Price & Time & Brand	0.579	0.303	0.455	0.246	0.468	0.251
Full	0.715	0.412	0.505	0.347	0.498	0.352

H.2 Raw Data

Here we present the rationalization analysis of strict PUMs for datasets from lists with three, four, and five positions without conditioning on choices an individual would purchase. Again, we often refer to the analysis with three, four, or five alternatives, so the indexing by list position is implicit. Pass rates using the unconditioned data are lower in most specifications. This could be because using purchase data reduces noise. However, this decrease could be mechanical since the purchase datasets contain fewer comparisons. Inspecting basic and adaptive MPS, there are no systematic increases or decreases between results for the raw datasets or those which include only purchase data. Many of the MPS differences between raw and purchase data are small (< 0.05 difference in MPS). The largest differences of MPS between the raw and purchase datasets is approximately 0.10 for a specification with position independent linear utility with a basic MPS.

Next, we examine fraction of the sets of relevant characteristics that pass a 0.10 MPS threshold for at least one utility specification. For three alternatives, Brand and Brand & Time fail to pass the threshold for basic MPS and only Brand & Time fails to pass the threshold for adaptive MPS. For four alternatives, only Brand fails to pass the threshold for basic MPS and only Brand & Time fails to pass the threshold for adaptive MPS. For five alternatives, only Brand fails to pass the threshold for basic MPS and adaptive MPS. Like the purchase data, only the relevant characteristic sets of Brand and Brand & Time ever fail to pass MPS thresholds.

Next, we examine if a linear utility model has more descriptive power from the sets of relevant characteristics that pass the 0.10 MPS threshold and excluding the specification with only Brand. Again, only Brand is excluded since many tests are identical. For three alternatives, a linear utility has the highest basic MPS for 4/6 cases and the highest adaptive MPS 2/6 cases. For four alternatives, a linear utility has the highest basic MPS 6/7 cases and the highest adaptive MPS for 4/6 cases. For five alternatives, a linear utility has the highest the highest basic MPS 7/7 cases and the highest adaptive MPS for 5/6 cases. Thus, linear utility specifications tend to also have more descriptive power for the raw data.

Again, there is some evidence that size of the list is important when performing this analysis. For example, separable utility over characteristics performs well for lists of size three. Therefore, individual utility from characteristics may be more subtle when lists are smaller. For four and five alternatives, a simple specification of a position independent linear utility over Price yields the highest adaptive MPS. This result provides some evidence that individuals may use simple descriptions of the alternatives as list size increases.

For the monotonicity tests with linear utility, there are only modest drops in pass rates for three, four, and five alternatives. The largest difference occurs for the specification with only Time for the three, four, and five alternative tests. The difference in pass rate is approximately 0.13-0.17. The position independent utility experiences little change to MPS, while the position dependent utility experiences MPS changes that can be up to 0.36. The Full specification of characteristics with position dependent utility has the highest basic MPS for the three, four, and five alternative datasets. The ranking by adaptive MPS differs from the basic MPS ranking of specifications. For three alternatives, the highest adaptive MPS occurs for the specification with position independent utility over Price. The four alternative adaptive MPS is largest with the Full set of relevant characteristics with position dependent utilities. For five alternatives, the highest adaptive MPS occurs for the specification with position dependent utility over Price.

Table 17: Pass Rates for Three Alternatives

	$u_a(x_a)$	$\sum_{j=1}^{d} u_{a,j}(x_{a,j})$	$\sum_{j=1}^{d} u_j(x_{a,j})$	$\beta_a \cdot x_a$	$\beta \cdot x_a$	S
Price	0.774	0.774	0.752	0.757	0.300	230
Brand	0.735	0.735	0.583	0.735	0.583	230
Time	0.730	0.730	0.643	0.665	0.270	230
Price & Brand	0.870	0.448	0.270	0.374	0.178	230
Price & Time	0.891	0.570	0.396	0.422	0.191	230
Brand & Time	0.843	0.113	0.017	0.061	0.009	230
Price & Time & Brand	0.970	0.704	0.487	0.587	0.291	230
Full	0.970	0.752	0.604	0.704	0.461	230

Table 18: Measure of Predictive Success for Three Alternatives: Basic

	$u_a(x_a)$	$\sum_{j=1}^{d} u_{a,j}(x_{a,j})$	$\sum_{j=1}^{d} u_j(x_{a,j})$	$\beta_a \cdot x_a$	$\beta \cdot x_a$	S
Price	0.039	0.039	0.257	0.238	0.300	230
Brand	0.004	0.004	0.059	0.004	0.059	230
Time	-0.000	-0.000	0.148	0.150	0.270	230
Price & Brand	0.057	0.391	0.178	0.369	0.178	230
Price & Time	0.073	0.553	0.396	0.422	0.191	230
Brand & Time	0.037	0.057	-0.074	0.057	0.009	230
Price & Time & Brand	-0.019	0.572	0.396	0.585	0.291	230
Full	-0.019	0.493	0.513	0.644	0.461	230

Table 19: Measure of Predictive Success for Three Alternatives: Adaptive

	$u_a(x_a)$	$\sum_{j=1}^{d} u_{a,j}(x_{a,j})$	$\sum_{j=1}^{d} u_j(x_{a,j})$	$\beta_a \cdot x_a$	$\beta \cdot x_a$	S
Price	0.040	0.040	0.313	0.241	0.286	230
Brand	0.004	0.004	0.125	0.004	0.123	230
Time	-0.001	-0.001	0.206	0.151	0.255	230
Price & Brand	0.065	0.329	0.265	0.331	0.178	230
Price & Time	0.057	0.465	0.393	0.419	0.191	230
Brand & Time	0.041	0.008	0.013	0.023	0.009	230
Price & Time & Brand	0.063	0.379	0.450	0.442	0.291	230
Full	0.064	0.329	0.504	0.417	0.455	230

Table 20: Linear Monotonicity Results for Three Alternatives

	Pass Rates		Basic MPS		Adaptive MPS	
	$\beta_a \cdot x_a$	$\beta \cdot x_a$	$\beta_a \cdot x_a$	$\beta \cdot x_a$	$\beta_a \cdot x_a$	$\beta \cdot x_a$
Price	0.722	0.300	0.551	0.300	0.571	0.293
Time	0.543	0.261	0.377	0.261	0.394	0.254
Price & Brand	0.357	0.178	0.356	0.178	0.347	0.178
Price & Time	0.343	0.191	0.343	0.191	0.343	0.191
Brand & Time	0.035	0.009	0.034	0.009	0.027	0.009
Price & Time & Brand	0.496	0.278	0.496	0.278	0.466	0.278
Full	0.648	0.396	0.643	0.396	0.558	0.394

Table 21: Pass Rates for Four Alternatives

	$u_a(x_a)$	$\sum_{j=1}^{d} u_{a,j}(x_{a,j})$	$\sum_{j=1}^{d} u_j(x_{a,j})$	$\beta_a \cdot x_a$	$\beta \cdot x_a$	S
Price	0.753	0.753	0.678	0.722	0.555	227
Brand	0.727	0.727	0.511	0.727	0.515	227
Time	0.727	0.727	0.577	0.678	0.379	227
Price & Brand	0.846	0.511	0.242	0.485	0.229	227
Price & Time	0.846	0.515	0.317	0.374	0.216	227
Brand & Time	0.762	0.185	0.053	0.128	0.044	227
Price & Time & Brand	0.947	0.692	0.401	0.573	0.286	227
Full	0.947	0.744	0.533	0.709	0.419	227

Table 22: Measure of Predictive Success for Four Alternatives: Basic

	$u_a(x_a)$	$\sum_{j=1}^{d} u_{a,j}(x_{a,j})$	$\sum_{j=1}^{d} u_j(x_{a,j})$	$\beta_a \cdot x_a$	$\beta \cdot x_a$	S
Price	0.023	0.023	0.207	0.086	0.495	227
Brand	-0.000	-0.000	0.009	-0.000	0.013	227
Time	-0.003	-0.003	0.108	0.040	0.318	227
Price & Brand	-0.008	0.361	0.150	0.465	0.229	227
Price & Time	0.023	0.436	0.317	0.374	0.216	227
Brand & Time	-0.072	0.033	-0.040	0.109	0.044	227
Price & Time & Brand	-0.053	0.419	0.308	0.560	0.286	227
Full	-0.053	0.390	0.440	0.558	0.419	227

Table 23: Measure of Predictive Success for Four Alternatives: Adaptive

	$u_a(x_a)$	$\sum_{j=1}^{d} u_{a,j}(x_{a,j})$	$\sum_{j=1}^{d} u_j(x_{a,j})$	$\beta_a \cdot x_a$	$\beta \cdot x_a$	S
Price	0.026	0.026	0.299	0.114	0.523	227
Brand	0.000	0.000	0.104	0.000	0.109	227
Time	-0.001	-0.001	0.199	0.068	0.347	227
Price & Brand	0.037	0.346	0.241	0.437	0.229	227
Price & Time	0.055	0.405	0.317	0.373	0.216	227
Brand & Time	-0.026	0.036	0.052	0.082	0.044	227
Price & Time & Brand	0.029	0.329	0.395	0.406	0.286	227
Full	0.029	0.292	0.508	0.385	0.418	227

Table 24: Linear Monotonicity Results for Four Alternatives

	Pass Rates		Basic MPS		Adaptive MPS	
	$\beta_a \cdot x_a$	$\beta \cdot x_a$	$\beta_a \cdot x_a$	$\beta \cdot x_a$	$\beta_a \cdot x_a$	$\beta \cdot x_a$
Price	0.661	0.555	0.409	0.524	0.480	0.539
Time	0.507	0.370	0.257	0.339	0.331	0.352
Price & Brand	0.463	0.229	0.459	0.229	0.458	0.229
Price & Time	0.308	0.216	0.308	0.216	0.308	0.216
Brand & Time	0.115	0.044	0.112	0.044	0.109	0.044
Price & Time & Brand	0.529	0.278	0.527	0.278	0.507	0.278
Full	0.656	0.344	0.642	0.344	0.570	0.343

Table 25: Pass Rates for Five Alternatives

	$u_a(x_a)$	$\sum_{j=1}^d u_{a,j}(x_{a,j})$	$\sum_{j=1}^d u_j(x_{a,j})$	$\beta_a \cdot x_a$	$\beta \cdot x_a$	S
Price	0.739	0.739	0.642	0.721	0.544	226
Brand	0.730	0.730	0.434	0.730	0.434	226
Time	0.730	0.730	0.491	0.668	0.279	226
Price & Brand	0.805	0.553	0.199	0.509	0.181	226
Price & Time	0.819	0.575	0.265	0.350	0.124	226
Brand & Time	0.796	0.221	0.022	0.155	0.018	226
Price & Time & Brand	0.867	0.633	0.288	0.571	0.195	226
Full	0.872	0.686	0.469	0.659	0.350	226

Table 26: Measure of Predictive Success for Five Alternatives: Basic

	$u_a(x_a)$	$\sum_{j=1}^d u_{a,j}(x_{a,j})$	$\sum_{j=1}^d u_j(x_{a,j})$	$\beta_a \cdot x_a$	$\beta \cdot x_a$	S
Price	0.009	0.009	0.174	0.010	0.487	226
Brand	-0.001	-0.001	-0.069	-0.001	-0.069	226
Time	-0.003	-0.003	0.022	-0.044	0.216	226
Price & Brand	-0.069	0.085	0.106	0.456	0.181	226
Price & Time	-0.052	0.163	0.265	0.350	0.124	226
Brand & Time	-0.070	-0.247	-0.071	0.105	0.018	226
Price & Time & Brand	-0.040	0.287	0.195	0.509	0.195	226
Full	-0.097	0.326	0.376	0.405	0.350	226

Table 27: Measure of Predictive Success for Five Alternatives: Adaptive

	$u_a(x_a)$	$\sum_{j=1}^d u_{a,j}(x_{a,j})$	$\sum_{j=1}^d u_j(x_{a,j})$	$\beta_a \cdot x_a$	$\beta \cdot x_a$	S
Price	0.009	0.009	0.303	0.037	0.518	226
Brand	0.000	0.000	0.055	0.000	0.064	226
Time	0.000	0.000	0.159	-0.014	0.251	226
Price & Brand	0.012	0.326	0.199	0.449	0.181	226
Price & Time	0.047	0.418	0.265	0.348	0.124	226
Brand & Time	0.012	-0.007	0.022	0.101	0.018	226
Price & Time & Brand	0.012	0.303	0.286	0.394	0.195	226
Full	0.016	0.285	0.461	0.348	0.349	226

Table 28: Linear Monotonicity Results for Five Alternatives

	Pass Rates		Basic MPS		Adaptive MPS	
	$\beta_a \cdot x_a$	$\beta \cdot x_a$	$\beta_a \cdot x_a$	$\beta \cdot x_a$	$\beta_a \cdot x_a$	$\beta \cdot x_a$
Price	0.664	0.544	0.295	0.513	0.388	0.532
Time	0.487	0.265	0.118	0.234	0.207	0.253
Price & Brand	0.478	0.181	0.467	0.181	0.471	0.181
Price & Time	0.274	0.111	0.274	0.111	0.274	0.111
Brand & Time	0.119	0.018	0.110	0.018	0.114	0.018
Price & Time & Brand	0.487	0.181	0.478	0.181	0.467	0.181
Full	0.606	0.292	0.545	0.292	0.508	0.292

Appendix I Availability Variation

We consider the special case when a menu only contains information about which alternatives are available. We show if all available alternatives in a menu are chosen with positive probability, then we can always rationalize the data with a strict perturbed utility model. These datasets are mentioned in Machina (1985) and it is stated that one can always always find a non-expected utility function that rationalizes the data. We show here that a strict PUM is equivalent to the non-expected utility function mentioned there.

For each a, let $\mathcal{X}_a = \{0, 1\}$ where 0 is interpreted as "unavailable" and 1 as "available". Let $\mathcal{X} = \left(\prod_{a=1}^{A} \mathcal{X}_a\right) \backslash (0, \ldots, 0)$. Consider subsets of alternatives given by $\mathcal{M} \subseteq \{1, \ldots, A\}$ and $\mathcal{M} \neq \varnothing$. Let $e_a \in \mathbb{R}^A$ be the a-th standard basis element with one in the a-th position and zeros elsewhere. We denote a menu with the alternatives in $\mathcal{M}$ available as $x^{\mathcal{M}} = \sum_{a \in \mathcal{M}} e_a$. Let $\mathcal{D} \subseteq \{\mathcal{M} \subseteq \{1, \ldots, A\} \mid \mathcal{M} \neq \varnothing\}$ be a collection of subsets of alternatives. When the characteristics are as defined above, we say a dataset $\{(x^{\mathcal{M}}, p(x^{\mathcal{M}}))\}_{\mathcal{M} \in \mathcal{D}}$ has *availability-variation* when for all $\mathcal{M} \in \mathcal{D}$, if $a \notin \mathcal{M}$ then $p_a(x^{\mathcal{M}}) = 0$.

An availability-variation dataset $\{(x^{\mathcal{M}}, p(x^{\mathcal{M}}))\}_{\mathcal{M} \in \mathcal{D}}$ satisfies *positivity* if for all $\mathcal{M} \in \mathcal{D}$, if $a \in \mathcal{M}$ then $p_a(x^{\mathcal{M}}) > 0$. This amounts to all probabilities being positive on the face of the simplex associated with $\mathcal{M}$. From Theorem 5(ii), it suffices to look at whether or not one can find a convex combination of probabilities that satisfies a system of equalities.

Proposition 3. *Let $\{(x^{\mathcal{M}}, p(x^{\mathcal{M}}))\}_{\mathcal{M} \in \mathcal{D}}$ be an availability-variation dataset that satisfies positivity, then there exists a strict perturbed utility model that rationalizes the dataset.*

Proof. For a dataset with a single observation, the data are rationalized by the argument for Corollary 1. Therefore, we consider when $|\mathcal{D}| \geqslant 2$. We prove the result by showing no distribution satisfying the equalities in Theorem 5(ii) exists.

First, we examine the relevant objects of Theorem 5(ii) for an availability-variation dataset. Consider the set $\mathcal{S} = \{(x^{\mathcal{N}}, x^{\mathcal{M}}) \mid \mathcal{M}, \mathcal{N} \in \mathcal{D} \text{ with } p(x^{\mathcal{N}}) \neq p(x^{\mathcal{M}})\}$. For $\mathcal{M} \neq \mathcal{N}$ there exists $a \in \mathcal{M} \backslash \mathcal{N}$ or $a \in \mathcal{N} \backslash \mathcal{M}$, so that $p_a(x^{\mathcal{N}}) \neq p_a(x^{\mathcal{M}})$ from positivity. Therefore, $\mathcal{S} = \{(\mathcal{N}, \mathcal{M}) \mid \mathcal{M}, \mathcal{N} \in \mathcal{D} \text{ and } \mathcal{M} \neq \mathcal{N}\}$. We label the distribution $\{\pi_{(\mathcal{N},\mathcal{M})}\}_{(\mathcal{N},\mathcal{M}) \in \mathcal{S}}$ with $\pi_{(\mathcal{N},\mathcal{M})} \geqslant 0$ and $\sum_{(\mathcal{N},\mathcal{M}) \in \mathcal{S}} \pi_{(\mathcal{N},\mathcal{M})} = 1$.

Suppose by contradiction that π satisfies the equalities in Theorem 5(ii). For each $a \in \{1, \ldots, A\}$, first equality of Theorem 5(ii) implies that

$$\sum_{\{(\mathcal{N},\mathcal{M}) \in \mathcal{S} \mid x_a^{\mathcal{M}} = 0\}} \pi_{(\mathcal{N},\mathcal{M})} p_a(x^{\mathcal{N}}) = 0$$

since $p_a(x^{\mathcal{M}}) = 0$ if $x_a^{\mathcal{M}} = 0$. Therefore, if $a \notin \mathcal{M}$ and $a \in \mathcal{N}$ then $\pi_{(\mathcal{N},\mathcal{M})} = 0$, otherwise the left hand side would be strictly positive from positivity. This means positive mass cannot be put on comparisons when $\mathcal{M}$ does not have alternatives that $\mathcal{N}$ contains. Therefore, $\pi_{(\mathcal{N},\mathcal{M})} \geqslant 0$ if $\mathcal{N} \subsetneq \mathcal{M}$, otherwise $\pi_{(\mathcal{N},\mathcal{M})} = 0$. If $\mathcal{D}$ contains no pairs $\mathcal{N}, \mathcal{M} \in \mathcal{D}$ with $\mathcal{N} \subsetneq \mathcal{M}$, then $\pi_{(\mathcal{N},\mathcal{M})} = 0$ for all $(\mathcal{N}, \mathcal{M}) \in \mathcal{S}$. This contradicts the assumption of π being a probability distribution.

We consider the final case when there exists $\mathcal{N}, \mathcal{M} \in \mathcal{D}$ such that $\mathcal{N} \subsetneq \mathcal{M}$. Recall for any $\mathcal{M}, \mathcal{N} \in \mathcal{D}$ that $p(x^{\mathcal{M}}) \neq p(x^{\mathcal{N}})$. The second equality of Theorem 5(ii) states for a fixed $\mathcal{M}$ that

$$\sum_{\mathcal{N}} \pi_{(\mathcal{M},\mathcal{N})} = \sum_{\mathcal{N}} \pi_{(\mathcal{N},\mathcal{M})}$$

since $x^{\mathcal{M}}$ has a distinct choice distribution. Define the set of alternatives in $\mathcal{D}$ with the largest cardinality to be

$$\mathcal{D}_{\max} = \{\mathcal{M} \in \mathcal{D} \mid \max_{\mathcal{M} \in \mathcal{D}} \{|\mathcal{M}|\}\}.$$

For $\mathcal{M}_{\max} \in \mathcal{D}_{\max}$, then

$$\sum_{\mathcal{N}} \pi_{(\mathcal{M}_{\max},\mathcal{N})} = 0$$

since no $\mathcal{N}$ contain $\mathcal{M}_{\max}$. However, the second equality of Theorem 5(ii) allows us to conclude for $\mathcal{M}_{\max} \in \mathcal{D}_{\max}$ that

$$\sum_{\mathcal{N}} \pi_{(\mathcal{N},\mathcal{M}_{\max})} = 0$$

so if $\mathcal{N} \subsetneq \mathcal{M}_{\max}$ then $\pi_{(\mathcal{N},\mathcal{M}_{\max})} = 0$. We can repeat this procedure inductively to show that $\pi_{(\mathcal{N},\mathcal{M})} = 0$ for all $\mathcal{N} \subsetneq \mathcal{M}$. However, then all $\pi_{(\mathcal{N},\mathcal{M})} = 0$ that contradicts π being a probability distribution. □

We now describe how to relate strict PUMs to a nonexpected utility approach. In the setup of a strict PUM, there are no explicit constraints placed on choice alternatives. Alternatively, one could consider placing explicit constraints on choice probabilities from menu information as in Machina (1985). For example, consider generating a nonexpected utility function from a strict PUM rationalization of an availability-variation dataset that satisfies positivity. We create the function

$$V(p) = \sum_{a \in A} v_a p_a + \tilde{V}(p)$$

where $v_a = u_a(1)$ and $\tilde{V}(p) = -C(p)$. We can explicitly enforce constraints for a menu $x^{\mathcal{M}}$

that all $b \notin \mathcal{M}$ satisfy $p_b = 0$ with a Lagrangian given by

$$\sum_{a \in A} v_a p_a + \tilde{V}(p) + \lambda_{\mathcal{M}} \left(\sum_{a \in \mathcal{M}} p_a - 1 \right) + \sum_{b \notin \mathcal{M}} \lambda_{(\mathcal{M},b)} p_b$$

with $\lambda_{\mathcal{M}}, \lambda_{(\mathcal{M},b)} \in \mathbb{R}$ and boundary constraints are excluded because of positivity. When $p_b = 0$ for all $b \notin \mathcal{M}$, the solutions are given by $p(x^{\mathcal{M}})$. Moreover, the construction of $C(\cdot)$ imposes that the v_a terms are in the sub-gradient so the $\lambda_{\mathcal{M}}$ terms can be assumed zero. The setup of the Lagrangian allows us to interpret the unavailability utilities. The unavailable utility of an alternative must satisfy $u_a(0) < u_a(1) + \min_{\mathcal{M} \in \mathcal{D}} \left(\lambda_{(\mathcal{M},a)} \right)$ so that the unavailable utility must be less than the available utility plus the worst case shadow cost of being unable to choose a.[31] Therefore, there is an equivalence between strict PUMs and strictly concave nonexpected utility for datasets of availability-variation with positivity.

Appendix J   Indexing and Two Stage Models

In the main text, we noted that the indexing of alternatives imposes structure on behavioral effects and complementarity/substitution patterns. As mentioned, the indexing implicitly states that these commodities are the same up to the additional structure imposed by characteristics. We note that there are some PUMs where the indexing can be thought of as an additional characteristic. For example, kinked additive PUMs with nonparametric utility over characteristics treats an alternative index the same as a characteristic. However, the indexing may be treated asymmetrically by placing more structure on utility for an indexed alternative over characteristics (such as linearity).

One may wonder what happens if there are two items which truly have the same index. For example, consider an individual choosing a flight from a list. One may want to use the airline to index alternatives, but the airline may show up several times on any list. One way to get around this issue would be to index the alternatives by both the position in the list and the airline. However, there are other approaches one could take.

For example, an individual may choose an airline based on some cognitive process, but then flights offered by the same airline are substitutable. For the choice of a flight from a list, we could sum up all choice probabilities from a list with the same airline to generate airline choice probabilities. We could then use some aggregation method to generate an

[31] This means the utility of an alternative never chosen cannot be distinguished from unavailable alternatives.

average characteristic value for the airline. This would allow us to test if a strict PUM to rationalizes the airline choice probabilities. We could then test the second hypothesis of substitution within an airline by imposing a kinked additive PUM for the probabilities of choosing different flights with the observed characteristic values after conditioning on the airline choice probability. Therefore, one can finely test different behavioral processes about individual choices by imposing different restrictions at different levels of indexing and aggregation. The choice of which models are suitable to test is at the discretion of the researcher.

References

Afriat, S. N. (1967): "The construction of utility functions from expenditure data," *International economic review*, 8, 67–77.

Agranov, M. and P. Ortoleva (Forthcoming): "Stochastic Choice and Preferences for Randomization," *Journal of Political Economy.*

Ahn, D., F. Echenique, and K. Saito (2016): "Average Choice," Caltech SS Working Paper 1409.

Allen, R. and J. Rehbeck (2016a): "Complementarity in Perturbed Models," *Working Paper.*

——— (2016b): "Identification of Average Demand Models," *Working Paper.*

Andreoni, J., B. Gillen, and W. T. Harbaugh (2013): "The power of revealed preference tests: Ex-post evaluation of experimental design," *Unpublished manuscript.*

Beatty, T. K. and I. A. Crawford (2011): "How demanding is the revealed preference approach to demand?" *The American Economic Review*, 101, 2782–2795.

Block, H. D. and J. Marschak (1960): "Random orderings and stochastic theories of responses," *Contributions to probability and statistics*, 97–132.

Blow, L., M. Browning, and I. Crawford (2008): "Revealed preference analysis of characteristics models," *The Review of Economic Studies*, 75, 371–389.

Border, K. C. (2013): "Alternative linear inequalities," *Manuscript, Division of Humanities and Social Sciences, California Institute of Technology.*

Boskin, M. J. (1974): "A conditional logit model of occupational choice," *Journal of Political Economy*, 82, 389–398.

Brady, R. L. and J. Rehbeck (2016): "Menu-Dependent Stochastic Feasibility," *Econometrica*, 84, 1203–1223.

Bronars, S. G. (1987): "The power of nonparametric tests of preference maximization," *Econometrica: Journal of the Econometric Society*, 693–698.

Caplin, A. and M. Dean (2015): "Revealed Preference, Rational Inattention, and Costly Information Acquisition," *American Economic Review*, 105, 2183–2203.

Cerreia-Vioglio, S., D. Dillenberger, P. Ortoleva, and G. Riella (2015): "Deliberately stochastic," *Manuscript Columbia University.*

Chambers, C. P., F. Echenique, and E. Shmaya (2014): "The axiomatic structure of empirical content," *The American Economic Review*, 104, 2303–2319.

Chambers, C. P., C. Liu, and S.-K. Martinez (2016): "A test for risk-averse expected utility," *Journal of Economic Theory*, 163, 775–785.

Cherchye, L., B. De Rock, and F. Vermeulen (2007): "The collective model of household consumption: a nonparametric characterization," *Econometrica*, 75, 553–574.

Cherchye, L., T. Demuynck, B. De Rock, and P. Hjertstrand (2015): "Revealed preference tests for weak separability: an integer programming approach," *Journal of econometrics*, 186, 129–141.

Chiappori, P.-A. and J.-C. Rochet (1987): "Revealed preferences and differentiable demand," *Econometrica*, 687–691.

Doyle, J. R., D. J. O'Connor, G. M. Reynolds, and P. A. Bottomley (1999): "The robustness of the asymmetrically dominated effect: buying frames, phantom alternatives, and in-store purchases," *Psychology & Marketing*, 16, 225–243.

Echenique, F. and K. Saito (2015): "Savage in the Market," *Econometrica*, 83, 1467–1495.

Fishburn, P. C. (1975): "Separation theorems and expected utilities," *Journal of Economic Theory*, 11, 16–34.

Fosgerau, M. and A. de Palma (2015): "Demand systems for market shares," .

Fudenberg, D., R. Iijima, and T. Strzalecki (2015): "Stochastic choice and revealed perturbed utility," *Econometrica*, 83, 2371–2409.

Green, R. C. and S. Srivastava (1986): "Expected utility maximization and demand behavior," *Journal of Economic Theory*, 38, 313–323.

Gul, F., P. Natenzon, and W. Pesendorfer (2014): "Random choice as behavioral optimization," *Econometrica*, 82, 1873–1912.

Heath, T. B. and S. Chatterjee (1995): "Asymmetric decoy effects on lower-quality versus higher-quality brands: Meta-analytic and experimental evidence," *Journal of Consumer Research*, 22, 268–284.

Hofbauer, J. and W. H. Sandholm (2002): "On the global convergence of stochastic fictitious play," *Econometrica*, 70, 2265–2294.

Huber, J., J. W. Payne, and C. Puto (1982): "Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis," *Journal of consumer research*, 9, 90–98.

Kitamura, Y. and J. Stoye (2013): "Nonparametric analysis of random utility models: testing," Working Paper.

Kubler, F., L. Selden, and X. Wei (2014): "Asset demand based tests of expected utility maximization," *The American Economic Review*, 104, 3459–3480.

Lancaster, K. J. (1966): "A new approach to consumer theory," *The Journal of Political Economy*, 132–157.

Louviere, J. J., R. T. Carson, L. Burgess, D. Street, and A. Marley (2013): "Sequential preference questions factors influencing completion rates and response times using an online panel," *Journal of Choice Modelling*, 8, 19–31.

Luce, R. D. (1959): *Individual Choice Behavior: A Theoretical Analysis*, New York: Wiley.

Machina, M. J. (1982): "" Expected Utility" Analysis without the Independence Axiom," *Econometrica: Journal of the Econometric Society*, 277–323.

——— (1985): "Stochastic choice functions generated from deterministic preferences over lotteries," *The economic journal*, 575–594.

Manski, C. F. (1977): "The structure of random utility models," *Theory and decision*, 8, 229–254.

Manzini, P. and M. Mariotti (2014): "Stochastic choice and consideration sets," *Econometrica*, 82, 1153–1176.

Markowitz, H. (1952): "Portfolio selection," *The journal of finance*, 7, 77–91.

Matejka, F. and A. McKay (2014): "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *The American Economic Review*, 105, 272–298.

Mattsson, L.-G. and J. W. Weibull (2002): "Probabilistic choice and procedurally bounded rationality," *Games and Economic Behavior*, 41, 61–78.

Matzkin, R. L. and M. K. Richter (1991): "Testing strictly concave rationality," *Journal of Economic Theory*, 53, 287–303.

McFadden, D. (1974): "The measurement of urban travel demand," *Journal of public economics*, 3, 303–328.

McFadden, D. and M. K. Richter (1990): "Stochastic rationality and revealed stochastic preference," *Preferences, Uncertainty, and Optimality, Essays in Honor of Leo Hurwicz, Westview Press: Boulder, CO*, 161–186.

McFadden, D. L. (2005): "Revealed stochastic preference: a synthesis," *Economic Theory*, 26, 245–264.

McFadden, D. L. and M. Fosgerau (2012): "A theory of the perturbed consumer with general budgets," Tech. rep., National Bureau of Economic Research.

Mosteller, F. and P. Nogee (1951): "An Experimental Measurement of Utility," *The Journal of Political Economy*, 371–404.

Myerson, R. B. (1981): "Optimal auction design," *Mathematics of operations research*, 6, 58–73.

Polisson, M., J. K.-H. Quah, L. Renou, et al. (2015): "Revealed preferences over risk and uncertainty," *Institute for Fiscal Studies Working Paper W*, 15.

Radner, R. and L. S. Miller (1970): "Demand and supply in US higher education: A progress report," *The American Economic Review*, 60, 326–334.

Ratneshwar, S., A. D. Shocker, and D. W. Stewart (1987): "Toward understanding the attraction effect: The implications of product stimulus meaningfulness and familiarity," *Journal of Consumer Research*, 13, 520–533.

Richter, M. K. (1966): "Revealed preference theory," *Econometrica*, 635–645.

Rochet, J.-C. (1987): "A necessary and sufficient condition for rationalizability in a quasi-linear context," *Journal of Mathematical Economics*, 16, 191–200.

Rockafellar, R. T. (1970): *Convex analysis*, 28, Princeton university press.

Rubinstein, A. and Y. Salant (2008): "Some thoughts on the principle of revealed preference," *The foundations of positive and normative economics: A handbook*, 115–124.

Segal, U. (1990): "Two-stage lotteries without the reduction axiom," *Econometrica*, 349–377.

Shi, X., M. Shum, and W. Song (2015): "Estimating Multinomial Choice Models using Cyclic Monotonicity," .

Simonson, I. (1989): "Choice based on reasons: The case of attraction and compromise effects," *Journal of consumer research*, 16, 158–174.

Sims, C. A. (2003): "Implications of rational inattention," *Journal of monetary Economics*, 50, 665–690.

Swait, J. and A. Marley (2013): "Probabilistic choice (models) as a result of balancing multiple goals," *Journal of Mathematical Psychology*, 57, 1–14.

Thurstone, L. L. (1927): "A law of comparative judgment." *Psychological review*, 34, 273.

Tversky, A. (1969): "Intransitivity of preferences." *Psychological review*, 76, 31.

Varian, H. R. (1983): "Non-parametric tests of consumer behaviour," *The review of economic studies*, 50, 99–110.