Latent Utility and Permutation Invariance: A Revealed Preference Approach *

Roy Allen
Department of Economics
University of Western Ontario
rallen46@uwo.ca

John Rehbeck
Department of Economics
The Ohio State University
rehbeck.7@osu.edu

October 25, 2023

Abstract

This paper provides partial identification results for latent utility models that satisfy an invariance property on unobservables such as exchangeability. We employ a simple revealed preference argument to "difference out" unobservables and show that this gives identifying inequalities for utility indices. We show the differencing argument is also useful for counterfactual analysis. The framework generalizes existing work in discrete choice by allowing latent feasibility sets and by allowing individuals to purchase multiple (possibly continuous) goods. We present a new framework leveraging nesting structures that generalizes nested logit. In a panel setting, we innovate by allowing preferences for variety.

Keywords: Identification, Latent Utility, Exchangeable, Discrete Choice, Panel.

JEL Code: C5.

*We thank Bo Honoré, Michal Kolesár, Nail Kashaev, Rory McGee, Mikkel Plagborg-Møller, and audiences at the Rochester Mini Conference, 2023 SAET Conference, and Princeton University for helpful comments. This research was supported by the SSHRC Insight Development Grant Program.

1 Introduction

Latent utility models often invoke an assumption that different unobservable components have the "same" distribution in some sense. One leading example is maximum score (Manski, 1975). For multinomial discrete choice, maximum score assumes the distribution of additive unobservables is exchangeable conditional on covariates. That is, the distribution of unobservables is the same when their order is changed (Goeree et al., 2005; Fox, 2007). Another example is nested logit where unobservables are exchangeable within a nest conditional on a nest-specific shock. Lastly, the static panel analysis of Manski (1987) and Shi et al. (2018) assumes the distribution of unobservables is the same across time conditional on fixed effects and covariates.

This paper combines invariance and utility maximization to generate identifying inequalities. A central challenge is that while it is straightforward to write revealed preference inequalities characterizing optimization, it is hard to "remove" the role of unobservables because they are not independent of choices (Pakes et al., 2015; Pakes, 2010). We address this by showing how to leverage invariance conditions to "difference out" the distribution of unobservables. Once these are differenced out, we are left with conditional moment inequalities that provide identifying information for utility indices. Unlike classic revealed preference arguments (e.g. Afriat (1967)), this revealed preference argument allows heterogeneous preferences and holds even when budgets are latent and random. The contribution of this paper is to unify existing work discussed above and present a general toolkit to use revealed preference arguments in new settings.

First, we use this framework to study new cross-sectional models. We present a generalization of nested logit that assumes a within-nest exchangeability condition on unobservables, but not a parametric distribution for unobservables. We show how within-nest exchangability generates within-nest moment inequalities. We then generalize the setup to consider the purchase of multiple goods which can be complements or substitutes. For example, in Ershov et al. (2018) it is assumed that all types of soda are in a nest, all types of chips are in a nest, and the paper allows soda and chips to be complements. More generally, the moment inequalities we obtain involve the conditional mean quantities of goods given covariates and hold for general discrete/continuous models that have a nesting structure.

Next, we turn to panel models. We study invariance conditions concerning the joint distribution of utility shocks across time. We show how to extend the cyclic monotonicity inequalities of Shi et al. (2018) to all perturbed utility models. As a leading example, we cover a fixed effects bundles model allowing complementarity/substitutability, as in Gentzkow (2007).[1] We accommodate a generalization in which quantities may be discrete, continuous, or any mix. We also show that the inequalities of Shi et al. (2018) can be motivated two ways: panel stationarity with time-separable preferences, or a new way that imposes panel exchangeability but allows preferences that are not time-separable, covering preference for variety.

In order to accommodate a variety of choice settings, we use the perturbed utility model (McFadden and Fosgerau, 2012; Allen and Rehbeck, 2019a). This model covers the classic discrete choice additive random utility model (McFadden, 1981), as well as models of matching (Fox et al., 2018), decisions under uncertainty (Agarwal and Somaini, 2018), and purchases of bundles (Gentzkow, 2007).[2] In addition to covering these existing models, perturbed utility models have also been used as direct models of choice. Examples include Fosgerau et al. (2022a), Fosgerau et al. (2022b), and Yan et al. (2022). One key feature of general perturbed utility models that we exploit is additively separable heterogeneity, which allows us to difference out unobservable heterogeneity.

Our identification analysis starts by examining the necessary and sufficient inequalities for optimization. In general, choices and unobservables are related, making a revealed preference approach challenging. In particular, it is not obvious how to remove the role of unobservables such as by an independence or mean-zero condition. We innovate by providing a general way to use invariance assumptions on how the unobservables and choices interact in the utility function. We construct a random variable from the assumed invariance conditions that reproduces the utility-match between the original choice and unobservables. In doing so, we show that the mean utility (of unobservables) of the original utility match and the utility match of the constructed random variable are equivalent. Thus, rather than assume a mean-zero condition on utility unobservables, we assume invariance conditions that ensure an

[1] Our analysis applies with an arbitrary number of goods. Wang (2021) derives sharp bounds for a slightly different version with two goods and binary quantities of each good.

[2] See Allen and Rehbeck (2019a) for how to represent choice models in these papers as perturbed utility models.

equal mean utility match. This is how we difference out the role of unobservables, generating identifying inequalities only concerning observable regressors and choices.

The technique we introduce is robust to latent feasibility sets. By a latent feasibility set, we mean that an individual may only be able to choose from a subset of the candidate choices the econometrician observes. Examples of latent feasibility include random stock outs or limited consideration. We handle latent feasibility sets by working with extended real-valued functions. Thus, we encode "infeasible" choices by setting their utility to $-\infty$.[3] The key assumption we require is that infeasibility satisfies an invariance property. For example, using the panel stationarity assumption of Shi et al. (2018), the invariance assumption requires that the distribution of utility shocks and feasible sets is the same across time, conditional on covariates and fixed effects. This handles the case of individual-specific (but not time varying) latent feasibility as a special case.[4]

This paper is part of broader agenda providing identification results for general perturbed utility models in Allen and Rehbeck (2019a) and Allen and Rehbeck (2020). The key difference is that both previous papers require an average structural function where covariates change and the distribution of unobservables is fixed. This is an invariance condition on the distribution of unobservables; it is implied if covariates are independent of unobservables. Both papers take a demand approach and use the envelope theorem heavily, in the style of classic discrete choice work such as McFadden (1981). The invariance condition employed in those papers rules out many examples that motivate the present paper, such as exchangeability conditional on covariates as in multinomial choice maximum score following Manski (1975). Our analysis here differs in four ways. First, we consider general invariance conditions such as exchangeability *between* unobservables in the same choice problem as in maximum score following Manski (1975), not just independence of covariates and unobservables *across* choices. Second, we use discrete revealed preference techniques whereas Allen and Rehbeck (2019a, 2020) differentiate the envelope theorem and focus on calculus-based techniques. Third, we provide inequalities that typically only partially identify structural objects of interest, whereas Allen and Rehbeck (2019a, 2020) focus on point

[3] This is a standard technique in convex analysis, e.g. Rockafellar (1970).

[4] See also Lu (2022), Crawford et al. (2021), and Aguiar and Kashaev (2019), who use different methods.

identification. Fourth, we focus on the linear index case with constant coefficients, whereas Allen and Rehbeck (2019a) study general nonparametric indices and Allen and Rehbeck (2020) study identification of the distribution of random coefficients. Finally, we study panel settings with fixed effects, which were not considered in Allen and Rehbeck (2019a) or Allen and Rehbeck (2020).

The rest of the paper is organized as follows. Section 2 describes the basic setup. Section 3 presents the main result delivering identifying inequalities by revealed preference. Section 4 show how invariance conditions imply inequalities for counterfactuals. Section presents examples in which invariance holds within a nest. Section 6 presents examples in which invariance holds across time in a panel context. Section 7 discusses the results and potential avenues for future work.

2 Setup

We study *perturbed utility models* where the observed choice $Y \in \mathbb{R}^K$ satisfies

$$Y \in \underset{y \in B}{\operatorname{argmax}} \sum_{k=1}^{K} y_k \beta_k' X_k + D(y, \alpha, \eta). \tag{1}$$

Each good k has utility shifters X_k. These are collected in $X = (X_1', \ldots, X_K')'$. The shifters enter through a utility index $\beta_k' X_k$ where $\beta = (\beta_1', \ldots, \beta_K')'$ is nonrandom. The unobservables are shifters α of unrestricted dimension, and good-specific shifters η of dimension K. We place no restrictions on α. In a panel setting, α can represent fixed effects. The unobservables combine with quantities y through the disturbance function D.

The primary object of interest in the paper is β. The other parts of the utility are "differenced out" by a revealed preference argument and we do not directly study identification of the disturbance function or distribution of unobservables. These objects indirectly show up when we examine counterfactual analysis in Section 4.

The set B and function D jointly encode complementarity/substitutability patterns between goods. For notational simplicity we use B to denote a "budget" constraint. For example, in discrete choice the budget is the probability simplex which we define formally in Example 1. Latent feasibility is allowed and can be encoded in D by letting

$D(y, \alpha, \eta) = -\infty$ for infeasible quantities. Importantly, these latent feasibility sets can be random and unobserved to the analyst as in Manski (1977). Thus, Equation (1) can equivalently be written

$$Y \in \operatorname*{argmax}_{y \in \tilde{B}(\alpha,\eta)} \sum_{k=1}^{K} y_k \beta_k' X_k + D(y, \alpha, \eta),$$

where $\tilde{B}(\alpha, \eta)$ is the set of $y \in B$ such that $D(y, \alpha, \eta) > -\infty$.[5] We focus on Equation (1) and follow the convex analysis approach of encoding budgets in extended real-valued functions. We emphasize this point because there has been extensive recent interest in latent feasibility that could arise because of stock out (Conlon and Mortimer (2013)) or limited consideration of alternatives (Goeree (2008); Fox (2007); Manzini and Mariotti (2014)).[6]

We require the following maintained technical assumptions.

Assumption 1. *Assume the following:*

i. The random variables (Y, X, α, η) *satisfy (1) almost surely, where* $D : \mathbb{R}^K \times A \times \mathbb{R}^K \to \mathbb{R} \cup \{-\infty\}$.

ii. $\Pr(\sup_{y \in B} D(y, \alpha, \eta) > -\infty) = 1$.

iii. Y *is* (X, α, η)*-measurable.*

iv. Y, X, $D(Y, \alpha, \eta)$ *and* $Y_k \beta_k' X_k$ *have finite expectation for each* k.

Part (i) formalizes the choices come from the optimizing model. The disturbance D can be $-\infty$ to reflect stochastic infeasibility of certain quantities, but can never be ∞. Part (ii) formalizes that some y is feasible with probability 1. This condition guarantees that the maximizer has finite utility for every draw of α and η. Part (iii) is for technical convenience and is automatic when the maximizer is unique. Part (iv) is needed because we work with conditional means.

Our key requirement is the following set of invariance conditions.

Assumption 2 ($\mathcal{G}$-Invariance)**.** *For a specified set of transformations* $\mathcal{G}$ *in* $\mathbb{R}^K$*, for*

[5] See also Allen and Rehbeck (2019a), pp. 1025-1026.

[6] Classic work includes Manski (1977), which emphasizes endogeneity in choice sets.

each $\Pi \in \mathcal{G}$, the following conditions hold:

i. The inverse Π^{-1} exists and $\Pi^{-1} \in \mathcal{G}$.

ii. Conditional on (X, α), η and $\Pi(\eta)$ have the same distribution.

iii. For each $y \in B$, $D(y, \alpha, \eta) = D(\Pi(y), \alpha, \Pi(\eta))$ almost surely.

iv. $y \in B$ implies $\Pi(y) \in B$.

v. $\Pi(Y)$ and $\Pi_k(Y)\beta_k' X_k$ have finite expectation for each k.

For the examples in this paper, $\mathcal{G}$ is a set of permutations that change the order of a vector. Condition (ii) is an exchangeability condition. For the main analysis, the α random variable is unobservable and integrated out to partially identify linear utility indices. When Π is a permutation, (iii) and (iv) formalize that "order doesn't matter" and are satisfied in many examples. The basic idea of this paper is that when $\mathcal{G}$-Invariance holds, we can "difference out" the distribution of unobservables using transformations of the data that are in $\mathcal{G}$. In part (v), $\Pi_k(Y)$ is the k-th component of the K dimensional vector $\Pi(Y)$.

By differencing out the unobservables, we can cover a range of existing models. As discussed in the Introduction, Allen and Rehbeck (2019a) show how different choices of disturbance D and budget B cover different existing models. In particular, representations of choice in the following papers are covered by this setup: classic discrete choice additive random utility (McFadden, 1981), matching (Fox et al., 2018), decisions under uncertainty (Agarwal and Somaini, 2018), and bundles (Gentzkow, 2007).[7]

To fix ideas in the familiar setting of discrete choice, we begin with an example where $\mathcal{G}$-Invariance holds for all permutations of indices of the K goods. This example is related to maximum score and related work studied in Manski (1975), Matzkin (1993), Goeree, Holt, and Palfrey (2005), Fox (2007), and Allen and Rehbeck (2019a). We study additional examples in Section 5.

Example 1 (Full Exchangeability)**.** *Consider the discrete choice additive random*

[7] We emphasize that while each of these papers has a utility representation of choice covered by our setup, one also needs to make additional invariance conditions to use the tools of this paper.

utility model, which can be written as a perturbed utility model by writing

$$Y \in \underset{y \in \Delta^{K-1}}{\operatorname{argmax}} \sum_{k=1}^{K} y_k \beta_k' X_k + \sum_{k=1}^{K} y_k \eta_k,$$

where $B = \Delta^{K-1} = \{y \in \mathbb{R}^K \mid y_k \geqslant 0, \sum_{k=1} y_k = 1\}$ *is the probability simplex and let* $D(y, \varepsilon) = \sum_{k=1}^{K} y_k \eta_k$*. With this specification, if alternative* k *is chosen, then the quantity vector* Y *has a* 1 *in the* k*-th dimension and a* 0 *elsewhere, and the individual gets utility* $v_k = \beta_k' X_k + \eta_k$*. The budget set* Δ^{K-1} *allows randomization in the presence of utility ties.*

We make assumptions such that $\mathcal{G}_F$*-Invariance holds with* $\mathcal{G}_F$ *the set of all permutations in* $\mathbb{R}^K$*. We verify each part of Assumption 2. Part (i) holds because permutations are invertible, and the inverse of a permutation is a permutation. Part (ii) is satisfied by assuming* $\eta = (\eta_1, \ldots, \eta_K)$ *is exchangeable, conditional on* X*. (There is no* α *in this example.) In other words, for any permutation* Π*,* $\Pi(\eta) = (\Pi_1(\eta), \ldots, \Pi_K(\eta)))$ *has the same distribution as* η*, conditional on* X*.*[8] *Note here that if components of* η *are independent and identically distributed conditional on* X*, then* η *is automatically exchangeable conditional on* X*. Thus, logit shocks satisfy this assumption. Part (iii) is verified by noting that* $D(Y, \eta) = D(\Pi(Y), \Pi(\eta))$ *by construction of* D*. Part (iv) is satisfied because the probability simplex* Δ^{K-1} *is invariant to permutations. Choice* Y *here is bounded and hence* $\Pi(Y)$ *is as well, and so* $\Pi(Y)$ *has finite expectation. A maintained assumption is that* X *has finite expectation, and so since* Y *is bounded,* $\Pi_k(Y)\beta_k' X_k$ *has finite expectation. Thus part (v) is satisfied and we have verified* $\mathcal{G}_F$*-Invariance.*

3 Identification by Revealed Preference

We show how $\mathcal{G}$-Invariance (see Assumption 2) provides identifying power for slope coefficients by leveraging revealed preference inequalities. To that end, let $Y = (Y_1, \ldots, Y_K)$ be a maximizer given (X, α, η) as in Equation (1). A necessary and

[8]We can encode permutations as permutations of indices. Thus, each Π can be written as $\Pi(\eta) = (\eta_{\pi(1)}, \ldots, \eta_{\pi(K)})$ for a permutation π of the set $\{1, \ldots, K\}$.

sufficient condition for optimality is that for any random variable $\tilde{Y}$ supported on B,

$$\sum_{k=1}^{K} Y_k \beta_k' X_k + D(Y, \alpha, \eta) \geqslant \sum_{k=1}^{K} \tilde{Y}_k \beta_k' X_k + D(\tilde{Y}, \alpha, \eta). \tag{2}$$

Taking expectations conditional on X, we obtain

$$\sum_{k=1}^{K} \mathbb{E}[Y_k - \tilde{Y}_k \mid X] \beta_k' X_k \geqslant \mathbb{E}[D(\tilde{Y}, \alpha, \eta) - D(Y, \alpha, \eta) \mid X] \tag{3}$$

whenever these expectations exist. For brevity we will leave the almost sure qualifications implicit, except in formal results.

The inequality in (3) states that the average change in utility due to the interaction of quantities and observables must outweigh the change due to the interaction with unobservables. A general challenge of operationalizing revealed preference arguments in econometrics is that Y and η are related (Pakes, 2010). This paper addresses this fundamental selection problem by constructing a random variable that "matches" the type of selection between Y and η. Specifically, this paper refines the inequality in Equation (3) by constructing a random variable $\tilde{Y}$ such that the right hand side is zero. We do this by leveraging the following implication of $\mathcal{G}$-Invariance.

Lemma 1. *Let Assumption 1 hold and let the set of transformations $\mathcal{G}$ satisfy Assumption 2. Then for every $\Pi \in \mathcal{G}$ there is a random variable $\tilde{Y}^{\Pi}$ supported on B such that:*

i. $\mathbb{E}[\tilde{Y}^{\Pi} - \Pi(Y) \mid X, \alpha] = 0$ *almost surely.*

ii. $\mathbb{E}\left[D(\tilde{Y}^{\Pi}, \alpha, \eta) - D(Y, \alpha, \eta) \mid X, \alpha\right] = 0$ *almost surely.*

The constructed variable $\tilde{Y}^{\Pi}$ is not observed in the general case. Thus, part (i) is key to identifying the conditional mean of $\tilde{Y}^{\Pi}$. Part (ii) is key to remove the role of unobservables in the inequality (3). Combining Lemma 1 and (3), we have the following result.

Theorem 1. *Let Assumption 1 hold and let the set of transformations $\mathcal{G}$ satisfy*

Assumption 2. For every $\Pi \in \mathcal{G}$,

$$\sum_{k=1}^{K} \mathbb{E}[Y_k - \Pi_k(Y) \mid X, \alpha]\beta_k' X_k \geqslant 0 \qquad a.s.$$

and hence

$$\sum_{k=1}^{K} \mathbb{E}[Y_k - \Pi_k(Y) \mid X]\beta_k' X_k \geqslant 0 \qquad a.s. \tag{4}$$

This result provides identifying information relating β and the joint distribution of observables. Note that this result holds without needing to specify the budget B, disturbance D, or the distribution of (α, η). Instead, revealed preference conditions show that we can "difference out" these components of the distribution when taking expectations. We describe implications of Theorem 1 in an example.

Example 1 (continued)**.** *When $\mathcal{G} = \mathcal{G}_F$ is the set of all permutations, (4) becomes*

$$\sum_{k=1}^{K} \mathbb{E}[Y_k - \Pi_k(Y) \mid X]\beta_k' X_k \geqslant 0$$

for every $\Pi \in \mathcal{G}_F$. In particular, by considering permutations that swap only two indices at a time, we obtain that for each pair k, ℓ,

$$(\mathbb{E}[Y_k \mid X] - \mathbb{E}[Y_\ell \mid X])(\beta_k' X_k - \beta_\ell' X_\ell) \geqslant 0,$$

which is equivalent to

$$\mathbb{E}[Y_k \mid X] > \mathbb{E}[Y_\ell \mid X] \implies \beta_k' X_k \geqslant \beta_\ell' X_\ell.$$

This shows the order of conditional probabilities lines up with the order of utility indices. These are a version of the maximum score inequalities originally studied in discrete choice (Manski, 1975). We differ because we allow utility ties, and these inequalities hold for all perturbed utility models when $\mathcal{G}_F$-Invariance is satisfied. That is, while this example stated an existing discrete choice setup, the resulting inequalities do not use specific features of discrete choice.

We make some additional remarks.

Remark 1 (Scale). The inequalities in Equation (4) do not contain scale information. That is, if β satisfies all such inequalities, then so does $\lambda\beta$ for any $\lambda > 0$. To handle this, we recommend setting $|\beta_{1,1}| = 1$, where the goods are arranged so that the first characteristic of the first good is thought to be relevant on *a priori* grounds. Alternatively, analysis can be done on relative terms such as the ratio of two components of β.

Remark 2 (Exclusion Restrictions). Theorem 1 does not make use of exclusion restrictions. Thus, X_k and X_j can be arbitrarily related; for example they can be identical. In many contexts, exclusion restrictions are important for meaningful analysis. To illustrate this, suppose we can split $X_k = (Z_k', W')'$ into variables Z_k that are good-specific shifters and variables W that are common across goods. For example, Z_k could include price of good k and W could include demographic variables. Suppose further $\beta_k = (\beta_k^{z\prime}, \beta^{w\prime})'$ so that $\beta_k' X_k = \beta_k^{z\prime} Z_k + \beta^{w\prime} W$. Equation (4) then reads

$$\sum_{k=1}^{K} \mathbb{E}[Y_k - \Pi_k(Y) \mid X] \beta_k^{z\prime} Z_k + \beta^{w\prime} W \sum_{k=1}^{K} \mathbb{E}[Y_k - \Pi_k(Y) \mid X] \geqslant 0 \qquad a.s.$$

In discrete choice, quantities sum to 1 and so when Π is a permutation, the second sum (next to W) is zero. Thus, the inequalities are not informative for β^w, though they can be for β_k^z. From this, we conclude that for discrete choice when coefficients for common regressors are the same across goods (β^w has no k subscript here), the inequalities from $\mathcal{G}$-invariance contain no information for the common regressors when the transformations consist of permutations.

Remark 3 (Utility Ties). When Y is the unique maximizer with probability 1, the inequality from Equation (2) is strict whenever $Y \neq \tilde{Y}$. Thus, (4) is strict provided $P(Y \neq \Pi(Y) \mid X) > 0$. Assuming a unique maximizer is common in the literature though we do not require this assumption.

Remark 4 (Smaller Conditioning Set). Suppose Π only alters a subset of alternatives, e.g. the first $L < K$ alternatives. Then we obtain $\mathbb{E}[Y_k - \Pi_k(Y) \mid X] = 0$ for any $k > L$. From this, (4) becomes

$$\sum_{k=1}^{L} \mathbb{E}[Y_k - \Pi_k(Y) \mid X] \beta_k' X_k \geqslant 0 \qquad a.s. \tag{5}$$

so that from the law of iterated expectations,

$$\sum_{k=1}^{L}(\mathbb{E}[Y_k - \Pi_k(Y) \mid X_1, \ldots, X_L])\beta_k' X_k \geqslant 0 \qquad a.s. \tag{6}$$

So we see Equation (5) is stronger than Equation (6). Importantly though, Equation (6) only requires observation of $(X_1, \ldots, X_L)$ and $(Y_1, \ldots, Y_L)$, and not the full vectors X and Y. This is related to a finding in Fox (2007) that certain multinomial choice maximum score inequalities arise when only a subset of alternatives are observed.

Remark 5 (Larger Conditioning Set). This paper focuses on α being an unobservable (such as a fixed effect). The analysis goes through if α is partially or fully observable. Specifically, split $\alpha = (W, \xi)$ where W is observable and ξ is unobservable; ξ can be degenerate in which case α is fully observable. Note in particular that some observables W can then enter the disturbance $D(Y, W, \xi, \eta)$. In addition, invariance conditions then hold conditional on the mix of observables and unobservables (W, ξ). By integrating out ξ, we obtain the inequalities

$$\sum_{k=1}^{K}\mathbb{E}[Y_k - \Pi_k(Y) \mid W, X]\beta_k' X_k \geqslant 0 \qquad a.s. \tag{7}$$

Remark 6 (Beyond Means). Here we only leverage identifying power of conditional means. In discrete choice these are conditional probabilities and this does not lose information, but in some contexts it might. The proof of Lemma 1 actually establishes equality of certain conditional distributions. Specifically, it constructs a random variable $\tilde{Y}^{\Pi}$ such that conditional on (X, α), $\Pi(Y)$ and $\tilde{Y}^{\Pi}$ have the same distribution, and similarly for $D(\tilde{Y}^{\Pi}, \alpha, \eta)$ and $D(Y, \alpha, \eta)$.

Here we only use the implication that means conditional on X are equal (by integrating out α). An interesting direction for future work is to study additional implications of this matching argument.

4 Invariance and Counterfactuals

This section studies how invariance conditions can be used to place restrictions on counterfactuals. To formalize this, the analyst is interested in placing restrictions on quantities at covariates $\tilde{x}$. We write the optimizing quantities as

$$\tilde{Y}(\tilde{x}, \alpha, \eta) \in \operatorname*{argmax}_{y \in B} \sum_{k=1}^{K} y_k \beta_k' \tilde{x}_k + D(y, \alpha, \eta).$$

This assumes the individual optimizes the same utility function as before. It also assumes for technical convenience that the counterfactual quantity is $(\tilde{x}, \alpha, \eta)$-measurable.

The object of interest is the mean quantity at the value of covariates $\tilde{x}$, under the distribution of (α, η) in the counterfactual setting, denoted F^C. We write this as

$$\overline{Y}(\tilde{x}, F^C) = \mathbb{E}_{F^C}[\tilde{Y}(\tilde{x}, \alpha, \eta)].$$

Some restrictions on F^C are needed for meaningful counterfactual analysis. We consider some examples.

First, we can assume the distribution F^C over (α, η) satisfies an invariance condition similar to Assumption 2 before. Namely, let the set of transformations applied to the counterfactual setting be denoted $\mathcal{G}^C$, and for any $\Pi \in \mathcal{G}^C$, the distribution of η is the same as $\Pi(\eta)$ conditional on α. Moreover, $D(\Pi(\tilde{Y}), \alpha, \Pi(\eta)) = D(\tilde{Y}, \alpha, \eta)$ and $y \in B$ implies $\Pi(y) \in B$. Similar to Theorem 1, this leads to inequalities such as

$$\sum_{k=1}^{K} \left(\overline{Y}(\tilde{x}, F^C) - \Pi(\overline{Y}(\tilde{x}, F^C))\right) \beta_k' \tilde{x}_k \geqslant 0. \tag{8}$$

As an example, with $\mathcal{G}^C$ the set of all permutations in $\mathbb{R}^K$, the analysis of Example 1 implies the inequalities

$$\beta_k' \tilde{x}_k > \beta_\ell' \tilde{x}_\ell \implies \overline{Y}_k(\tilde{x}, F^C) \geqslant \overline{Y}_\ell(\tilde{x}, F^C).$$

Suppose β is identified or more generally we can identify whether the left hand side holds. Then this implication provides inequalities relating different components of the counterfactual average quantities.[9]

We now consider an alternative assumption on F^C. Let x be a point in $\text{supp}(X)$, i.e. the support of X.[10] We assume that F^C equals the distribution of (α, η) given $X = x$, which we can write as $F^C_x = F_{(\alpha,\eta)|X=x}$. Since x is a point in the support of X, this formalizes that the counterfactual distribution of latent variables is the same at a certain value x of observable covariates. This invariance condition allows us to use revealed preference inequalities to describe restrictions on $\overline{Y}(\tilde{x}, F^C_x)$. Specifically, for (Y, X, α, η) consistent with utility maximization, we use the following implications of optimization

$$\sum_{k=1}^{K} Y_k \beta_k' X_k + D(Y, \alpha, \eta) \geqslant \sum_{k=1}^{K} \tilde{Y}_k(\tilde{x}, \alpha, \eta) \beta_k' X_k + D(\tilde{Y}(\tilde{x}, \alpha, \eta), \alpha, \eta)$$
$$\sum_{k=1}^{K} \tilde{Y}_k(\tilde{x}, \alpha, \eta) \beta_k' \tilde{x}_k + D(\tilde{Y}(\tilde{x}, \alpha, \eta), \alpha, \eta) \geqslant \sum_{k=1}^{K} Y_k \beta_k' \tilde{x}_k + D(Y, \alpha, \eta).$$

These rearrange to

$$\sum_{k=1}^{K} \left(Y_k - \tilde{Y}_k(\tilde{x}, \alpha, \eta) \right) (\beta_k' X_k - \beta_k' \tilde{x}_k) \geqslant 0.$$

Taking the expectation conditional on $X = x$ yields

$$\sum_{k=1}^{K} \left(\mathbb{E}[Y_k \mid X = x] - \overline{Y}_k(\tilde{x}, F^C_x) \right) (\beta_k' x_k - \beta_k' \tilde{x}_k) \geqslant 0,$$

where we have used the assumption that $F^C_x = F_{(\alpha,\eta)|X=x}$ to conclude that $\overline{Y}(\tilde{x}, F^C_x) = \mathbb{E}_{F_{(\alpha,\eta)|X=x}}[\tilde{Y}(\tilde{x}, \alpha, \eta)]$. We emphasize that while this counterfactual exercise holds the distribution of unobservables fixed and varies covariates, we do *not* require independence between X and unobservables for the general results in the paper.[11]

[9] Appendix A of Chiong et al. (2021) presents similar inequalities for discrete choice.

[10] The support of X is the smallest closed set K such that $P(X \in K) = 1$.

[11] Allen and Rehbeck (2019a) and Chiong et al. (2021) show independence places additional restrictions on counterfactuals. Recall here we only exploit invariance (equality) between two distributions: the distribution of unobservables at $X = x$ and in the (single) counterfactual setting.

5 Nesting Invariance

In this section we describe examples in which $\mathcal{G}$-Invariance holds for a class of transformations due to a nesting structure. We first consider a discrete choice example that generalizes nested logit because it does not require a parametric specification of unobservables. Then we consider nesting structures in discrete/continuous models that allow complementarity between goods.

5.1 Nesting in Discrete Choice

Partition the K goods into m non-overlapping nests $K_1, \ldots, K_M$. For each $m \in \{1, \ldots, M\}$, let $\mathcal{G}_{N,m}$ be the set of permutations of indices of goods in K_m. Let $\mathcal{G}_N = \bigcup_{m=1}^{M} \mathcal{G}_{N,m}$ be the set of all such permutations. Note that transformations in $\mathcal{G}_{N,m}$ only compare elements of the nest K_m.

First, consider a discrete choice model. Let the nest alternative k belongs to be denoted $m(k)$. Specify latent utility for good k as

$$v_k = \beta_k' X_k + \alpha_{m(k)} + \eta_k.$$

The unobservable $\alpha_{m(k)}$ is common to the nest $m(k)$. The unobservable η_k is specific to alternative k. As in Example 1, write $D(y, \alpha, \eta) = \sum_{k=1}^{K} y_k(\alpha_{m(k)} + \eta_k)$. Then any element of $\mathcal{G}_N$ only permutes choices in the same nest. For example, if $\Pi \in \mathcal{G}_{N,m}$, then alternatives in nest m are permuted and we have

$$\begin{aligned} D(\Pi(y), \alpha, \Pi(\eta)) &= \alpha_m \left(\sum_{k \in K_m} y_{\pi(k)} \right) + \sum_{k \in K_m} y_{\pi(k)} \eta_{\pi(k)} + \sum_{k \in K \setminus K_m} y_k (\alpha_{m(k)} + \eta_k) \\ &= \alpha_m \left(\sum_{k \in K_m} y_k \right) + \sum_{k \in K_m} y_k \eta_k + \sum_{k \in K \setminus K_m} y_k (\alpha_{m(k)} + \eta_k) \\ &= D(y, \alpha, \eta). \end{aligned}$$

We see that the disturbance D is invariant to transformations that permute the quantities of the goods within the same nest. Suppose that conditional on (α, X), $\Pi^m(\eta)$ and η have the same distribution for any $\Pi^m \in \mathcal{G}_{N,m}$. That is, Π^m permutes only goods in nest m and we have within-nest exchangeability. Then under minor regularity conditions we satisfy $\mathcal{G}_{N,m}$-Invariance for nest m. Thus, the inequalities of Theorem 1 hold for $\Pi \in \mathcal{G}_{N,m}$. We note nested logit is a special case of this framework (Cardell (1997), Galichon (2022)). Here, we generalize nested logit since we do not specify a parametric distribution for any of the random variables.

5.2 Nesting Outside of Discrete Choice

Now consider a model in which a consumer can buy quantities of different goods. We go beyond discrete choice by allowing complementarity between goods. Goods are partitioned into M nests. For example, Ershov et al. (2018) study demand for soda and potato chips.[12] Here, the partition splits the goods into a nest containing all types of soda, and a nest containing all types of potato chips. More generally, we consider M nests.

To begin, assume quantities are each either 0 or 1, and quantity in each nest is at most 1. Quantities across nests need not sum to 1 and so this is not classic discrete choice. Motivated by the presentation of nesting in discrete choice, suppose utility takes the form

$$\begin{aligned}
&\sum_{k=1}^{K} y_k \beta_k' X_k + \sum_{k=1}^{K} y_k \eta_k + \sum_m \alpha_m \left(\sum_{k \in K_m} y_k \right) \\
&\quad + \sum_{m_1 \neq m_2} \alpha_{m_1,m_2} \left(\sum_{k \in K_{m_1}} y_k \right) \left(\sum_{k \in K_{m_2}} y_k \right) \\
&\quad + \cdots + \sum_{m_1 \neq \cdots \neq m_M} \alpha_{m_1,\ldots,m_M} \left(\sum_{k \in K_{m_1}} y_k \right) \cdots \left(\sum_{k \in K_{m_M}} y_k \right).
\end{aligned} \tag{9}$$

The term $\sum_{k=1}^{K} y_k \beta_k' X_k$ reflects that the utility shifters only shift the marginal utility of each good. The term $\sum_{k \in K} y_k \eta_k$ describes how preferences depend on the good-specific unobservables. The term $\sum_m \alpha_m \left(\sum_{k \in K_m} y_k \right)$ describes how preferences depend on the nest-specific unobservable α_m. This part parallels the discrete choice analysis above. What differs here is that individuals may purchase goods from multiple nests. For this reason, we include terms such as $\sum_{m_1 \neq m_2} \alpha_{m_1,m_2} \left(\sum_{k \in K_{m_1}} y_k \right) \left(\sum_{k \in K_{m_2}} y_k \right)$.

[12] See also Iaria and Wang (2019) and Iaria and Wang (2021) for recent work on bundles in other contexts.

When $\alpha_{m_1,m_2} > 0$, there is a utility "boost" to consuming positive quantities of goods in nest m_1 and nest m_2 at the same time. This is a direct generalization of Gentzkow (2007), which considers bundles of at most two types of goods. Higher order terms allow complementarity/substitutability patterns to depend on total consumption from each nest.

The key structure in (9) that can motivate invariance conditions is that quantities in each nest enter utility either through a pure sum involving a component of α, or the weighted sum involving η. We will study invariance conditions that use this structure. We cast it in a more general setup to handle examples in which quantities are not zero or one. That is, we now assume instead that quantities in each nest are unrestricted, so they can be discrete, continuous, or any mix. Let utility be given by

$$\sum_{k=1}^{K} y_k(\beta_k' X_k) + g\left(\sum_{k=1}^{K} y_k \eta_k, \sum_{k \in K_{m_1}} y_k, \cdots, \sum_{k \in K_{m_M}} y_k, \alpha\right). \tag{10}$$

Here, g is now a general function that depends on the η-weighted sum of all quantities, the summation of quantities in each nest, and the individual-specific term α. The function g encodes general complementarity/substitutability patterns between the goods. Heterogeneity in this relationship is encoded by α. The restriction on complementarity/substitutability is that it enters through sums or a weighted sum (the first argument of g). This setup generalizes the specification in Equation (9).

We now present an invariance condition in this model of multiple purchase. Let m index an arbitrary nest and let Π^m be a permutation of quantities within nest K_m. Suppose that conditional on (α, X), $\Pi^m(\eta)$ and η have the same distribution. Each summation $\left(\sum_{k \in K_m} y_k\right)$ is equal when components of K_m are permuted. Moreover, $\sum_{k \in K_m} y_k \eta_k$ is the same as $\sum_{k \in K_m} y_{\pi(k)} \eta_{\pi(k)}$, where π encodes Π^m in terms of the corresponding permutation of indices. We conclude that with the specification in Equation (10), $\mathcal{G}_{N,m}$-Invariance is satisfied under straightforward conditions. We can directly apply Theorem 1. The resulting inequalities are

$$\sum_{k \in K_m} \mathbb{E}[Y_k - \Pi_k^m(Y) \mid X] \beta_k' X_k \geqslant 0.$$

Note that the sum only involves terms in the nest $k \in K_m$, because these are the only parts of k that are affected by the transformation Π^m. Further, note that by the law of iterated expectations, these inequalities can be applied if we only observe covariates for the goods in nest K_m, and do not require observing the full vector of covariates X.

Additional invariance conditions yield additional inequalities. One example is if substitution/complementarity patterns are similar between nests. This can be modelled by assuming g is invariant to permutations of different arguments, such as arguments 2 and 3 corresponding to the quantities in nests 2 and 3. Common continuous demand systems such as CES exhibit such symmetries.

6 Panel Data and Invariance across Time

We now describe how panel data and across-time invariance conditions fit into this framework. Specifically, we focus on invariance conditions that permute choices across time periods. We first consider when there is separability in preferences over time in Section 6.1. This covers and extends work on discrete choice panel models as studied in Manski (1987) and Shi et al. (2018) to general perturbed utility models. In Section 6.2, we show that we can relax the time-separability of preferences when unobservables are exchangeable across time. This accommodates when individuals have nontrivial preferences for variety.

We emphasize here that while we reference existing work on discrete choice in this section, and use specific examples for familiarity, we cover general perturbed utility models. In particular, this analysis covers "panel versions" of classic discrete choice (McFadden, 1981), matching (Fox et al., 2018), decisions under uncertainty (Agarwal and Somaini, 2018), and bundles (Gentzkow, 2007).

6.1 Time-Separability

Consider a panel setting with T periods. In period t, quantities in $\mathbb{R}^L$ are chosen to satisfy

$$Y^t \in \underset{y^t \in B^t}{\operatorname{argmax}} \sum_{\ell=1}^{L} y_\ell^t \beta_\ell^{t\prime} X_\ell^t + D^t(y^t, \alpha, \eta^t). \tag{11}$$

We use superscripts to index time periods. We interpret Y^t as the choice in period t where there are L goods available each time period. We switch to this notation since we will later show that decisions across time can be written as a single period decision. Here, α is an individual-specific fixed effect and η^t is a individual-time-specific, L-dimensional unobservable.

To write this in our previous framework, define the covariates in period t by $X^t = (X_1^1, \ldots, X_L^t)'$ and the period t linear coefficients by $\beta^t = (\beta_1^t, \ldots, \beta_L^t)'$. In practice it common to make these constant across t but we do not require this, which can accommodate discounting. Collect terms across periods in $Y = (Y^{1\prime}, \ldots, Y^{T\prime})'$, $X = (X^{1\prime}, \ldots, X^{T\prime})'$, $\beta = (\beta^{1\prime}, \ldots, \beta^{T\prime})'$, and $\eta = (\eta^{1\prime}, \ldots, \eta^{T\prime})'$. We can compactly write the collection of all T choice problems as

$$Y \in \underset{(y^{1\prime},\ldots,y^{T\prime}) | y^t \in B^t \forall t}{\operatorname{argmax}} \sum_{t=1}^{T} \sum_{\ell=1}^{L} y_\ell^t (\beta_\ell^{t\prime} X_\ell^t) + \sum_{t=1}^{T} D^t(y^t, \alpha, \eta^t),$$

which we recognize as a time-separable case of our original formulation (1) by setting

$$D^{TS}(y, \alpha, \eta) := \sum_{t=1}^{T} D^t(y^t, \alpha, \eta^t) \tag{12}$$

and $B = \prod_{t=1}^{T} B^t$. To map to our previous notation, we let $K = T \times L$ for this section so that the choice variables are given by all goods over all time periods.

We formalize assumptions on time-separability below.

Assumption 3 (Time-Separability)**.** *Assume the following:*

i. (11) holds for each $t \in \{1, \ldots, T\}$.

ii. Y^t is (X^t, α, η^t)-measurable.

iii. D^t and B^t are the same for each t.

Parts (i) and (ii) ensure full separability across time, except through the individual-specific fixed effect α. Part (iii) ensures the disturbance function and budgets are the same across time. We recall, however, that D^t itself can encode random feasibility sets and so heterogeneous choice sets are possible provided they satisfy the conditions below.

In contrast with the previous results, here we only study transformations that swap time periods. In more detail, we consider Π that satisfy

$$\Pi(Y) = (Y^{\pi(1)\prime}, \ldots, Y^{\pi(T)\prime}),$$

where π is a permutation of time indices. We call such transformations *time swaps.* An example is the transformation that swaps periods 1 and 2,

$$\Pi_{1,2}(Y) = (Y^{2\prime}, Y^{1\prime}, Y^{3\prime}, \ldots, Y^{T})'.$$

Here, $\pi_{1,2}(1) = 2, \pi_{1,2}(2) = 1$, and $\pi_{1,2}(t) = t$ for $t \geqslant 3$. For notational convenience, when Π is a time swap we then write $\Pi^t(Y) = Y^{\pi(t)}$, so that Π^t extracts the time-t components after the transformation Π has been applied.

Our new invariance conditions are as follows.

Assumption 4. *For a specified set of transformations $\mathcal{G}$ in $\mathbb{R}^K$, for each $\Pi \in \mathcal{G}$, the following conditions hold:*

i. The inverse Π^{-1} exists and satisfies $\Pi^{-1} \in \mathcal{G}$.

ii. Each $\Pi \in \mathcal{G}$ is a time swap. Formally, there is a permutation π of time indices $\{1, \ldots, T\}$ such that for every $y = (y^{1\prime}, \ldots, y^{T\prime})' \in \mathbb{R}^K$ with each $y^t \in \mathbb{R}^L$,

$$\Pi(y) = (y^{\pi(1)\prime}, \ldots, y^{\pi(T)\prime})'.$$

iii. For every t, η^t and $\Pi^t(\eta)$ have the same distribution conditional on (α, X).

iv. For each $y \in B$, $\sum_{t=1}^{T} D^t(y^t, \alpha, \eta^t) = \sum_{t=1}^{T} D^t(\Pi^t(y), \alpha, \Pi^t(\eta))$ almost surely.

v. $y \in B$ implies $\Pi(y) \in B$ almost surely.

vi. $\Pi(Y)$ and $\Pi_k(Y)\beta_k' X_k$ have finite expectation for each k.

Part (ii) is an additional restriction relative to Assumption 2. Part (iii) is often referred to as a stationarity assumption, and is used in Manski (1987), Shi et al. (2018), and Pakes and Porter (2021). Part (iii) weakens the previous Assumption 2(ii), which required that η and $\Pi(\eta)$ have the same conditional distribution. Note that in this panel setup, η is a vector of shocks for all goods and for all time periods.

Part (iii) of the new assumption only requires time-t unobservables η^t and the time-swapped part $\Pi^t(\eta)$ to have the same distribution. Given the other assumptions, (iv) is equivalent to

$$D^{TS}(Y, \alpha, \eta) = D^{TS}(\Pi(Y), \alpha, \Pi(\eta)),$$

where D^{TS} represents the sum of all period-t disturbances as in (12). Part (vi) is stated in the previous notation for consistency because it is unchanged; in panel notation, the good k in $\Pi_k(Y)\beta_k' X_k$ corresponds to a time-good tuple (t, ℓ).

We obtain the following counterpart to Theorem 1. We use separability (Assumption 3) to replace Assumption 2 with Assumption 4.

Theorem 2. *Let Assumptions 1 and 3 hold and let $\mathcal{G}$ satisfy Assumption 4. For every $\Pi \in \mathcal{G}$,*

$$\sum_{t=1}^{T} \sum_{\ell=1}^{L} \mathbb{E}[Y_\ell^t - \Pi_\ell^t(Y) \mid X] \beta_\ell^{t\prime} X_\ell^t \geqslant 0 \qquad a.s. \tag{13}$$

Recall the panel notation of (11) uses the notation $Y = (Y^{1\prime}, \ldots, Y^{T\prime})'$, so Y is a $K = T \times L$-dimensional vector.

6.1.1 Cyclic Monotonicity

When $\mathcal{G}$ is the set of all time swaps, the inequalities of (13) are cyclic monotonicity as in Shi et al. (2018).[13] Cyclic monotonicity can be stated in terms of cycles or permutations. A function $f : \mathbb{R}^A \to \mathbb{R}^A$ is cyclically monotone if for any sequence $z^1, \ldots, z^S$ with $z^{S+1} = z^1$, it follows that

$$\sum_{s=1}^{S} f(z^s)(z^s - z^{s+1}) \geqslant 0.$$

Equivalently, f is cyclically monotone if for any such sequence and permutation π of $\{1, \ldots, S\}$,

$$\sum_{s=1}^{S} (f(z^s) - f(z^{\pi(s)}))z^s \geqslant 0.$$

[13] See also McFadden and Fosgerau (2012), which generalizes cyclic monotonicity to accommodate observable budget variation.

Inequalities in (13) can be written in this form by writing them as

$$\sum_{t=1}^{T}\sum_{\ell=1}^{L}\mathbb{E}[Y_\ell^t - Y_\ell^{\pi(t)} \mid X]\beta_\ell^{t\prime}X_\ell^t \geqslant 0 \qquad a.s.,$$

where π encodes the time swap Π as a permutation of time indices.

While Shi et al. (2018) study discrete choice, we present a general framework in which the same identifying inequalities hold for all perturbed utility models. When we specialize our analysis to discrete choice, our assumptions are weaker than those in Shi et al. (2018) because we allow extended real-valued unobservables, and because we do not require existence of a density.[14] Extended real-valued unobservables are not just a technical curiosity. They are needed to accommodate latent feasibility sets.

6.1.2 Latent Feasibility Sets

While the main goal of this paper is to push beyond discrete choice, we illustrate how latent feasibility sets can be handled for discrete choice. This analysis can be adapted to other settings by the general method of encoding latent feasibility sets through restrictions on combinations of quantities and unobservables such that $D(y, \alpha, \eta) = -\infty$.

To illustrate latent feasibility sets in discrete choice, suppose the latent utility of alternative k in period t is given by

$$v_k^t = \beta_k^{t\prime}X_k^t + \alpha_k + \eta_k^t.$$

We define $a - \infty = -\infty$ for any $a < \infty$. If $\alpha_k = -\infty$, then alternative k is infeasible or not considered. More generally, our framework also allows the feasibility set to change across time. Alternative k is stochastically infeasible at time t if $\eta_k^t = -\infty$ can occur with positive probability. We allow random infeasibility as long as it satisfies our invariance assumptions. With time swaps, this means the distribution of infeasible alternatives must be the same across time, conditional on covariates and the fixed effect. When the latent feasibility set is individual-specific but does not change across time, we note that the only restriction is that one alternative must be feasible.

[14] Shi et al. (2018) focus on the inequalities resulting from $T = 2$ or from a time swap that only permutes two time periods.

6.2 Nonseparability across time

We now consider a more general specification of the disturbance that allows nontrivial interactions of choices across time periods. For example, preferences can depend on the total consumption of a good across time. Results here return to the original Theorem 1. The key trade-offs are that nonseparability means we have to match the full conditional distributions of η and $\Pi(\eta)$, whereas additive separability means we only need the conditional distributions of η^t and $\Pi^t(\eta)$ to be equal (Theorem 2).

We begin by analyzing time-swaps as in Section 6.1 and then turn to partial swaps that only swap time indices for a fixed good.

6.2.1 Time Swaps

We show that the cyclic monotonicity inequalities of Shi et al. (2018) hold with nontrivial substitution/complementarity patterns over time when we strengthen the invariance condition. In this section $\Pi \in \mathcal{G}^T$ is a time swap. We require η to have the same distribution as $\Pi(\eta)$, conditional on (X, α). This strengthens Assumption 2, which only required matching the period-t marginal distributions. This is not for free. Invariance of η to time swaps rules out an AR(1) model for η^t because the distribution of η would then depend on the order of time.

To formalize the setup, assume choices Y across all time periods satisfy

$$Y \in \operatorname*{argmax}_{(y^{1\prime},\ldots,y^{T\prime}) \in B} \sum_{t=1}^{T} \sum_{\ell=1}^{L} y_\ell^t (\beta_\ell^{t\prime} X_\ell^t) + D(y, \alpha, \eta).$$

Here, we do not assume $D(y, \alpha, \eta)$ is additively separable across time (Section 6.1), but instead assume it is invariant to time swaps. Formally, we assume

$$D(Y, \alpha, \eta) = D(\Pi(Y), \alpha, \Pi(\eta)), \tag{14}$$

for any $\Pi \in \mathcal{G}^T$. Here $\mathcal{G}^T$ is defined in Assumption 4(iii).

The only restriction on the budget B is that it must be invariant to time swaps. Specifically, Assumption 2(iii) states only that $y \in B$ implies $\Pi(y) \in B$. Here, B is the budget of quantities across all time periods. Importantly, the budget B need not be the Cartesian product of identical period-t budgets; this stronger condition was required in Section 6.1. Here B need only satisfy the time-invariance condition. An example in which B is not a Cartesian product is repeated binary choice in which a good is chosen at most once. Then B can encode vectors $(y^1, \ldots, y^T)$ that sum to either 0 or 1 across all time periods, where each y^t is either 0 or 1. If we force these components to also sum to exactly 1, this budget admits the interpretation as an optimal stopping problem.

By treating the disturbance as a nontrivial function across all time periods, this formulation models an individual who chooses all of Y at once. Specifically, the individual knows the persistent value α as well as covariates X^t and shocks η^t for all t, and then chooses all quantities $Y = (Y^1, \ldots, Y^T)$ at once. This differs from the separable case (11), which requires only that at each t the individual knows (X^t, α, η^t) and chooses Y^t.

Inequalities for this setup are directly covered by Theorem 1. Any transformation satisfying Assumption 2 generates inequalities. Here we focus on time-swaps $\mathcal{G}^T$. For any $\Pi \in \mathcal{G}^T$,

$$\sum_{k=1}^{K} \mathbb{E}[Y_k - \Pi_k(Y) \mid X]\beta_k' X_k \geqslant 0 \qquad a.s.,$$

or in panel notation,

$$\sum_{t=1}^{T} \sum_{\ell=1}^{L} \mathbb{E}[Y_\ell^t - \Pi_\ell^t(Y) \mid X]\beta_\ell^{t\prime} X_\ell^t \geqslant 0 \qquad a.s.$$

Thus, the inequalities of Shi et al. (2018) hold even when there are nontrivial preferences for variety that imply violations of time-separability.

For an example that satisfies Equation (14), consider the weak separability restriction that

$$D^{WS1}(Y, \alpha, \eta) = V(D^1(Y^1, \alpha, \eta^1), \ldots, D^T(Y^T, \alpha, \eta^T), \alpha),$$

where each function D^t is the same across t, and the aggregator function V is symmetric in its first T arguments so that $V(d, \alpha) = V(\Pi(d), \alpha)$ for any permutation Π in T dimensions. Assumption 3(i) is a special case of this in which $V(d, \alpha) = \sum_{t=1}^{T} d^t$.

For a parametric example of weak separability, consider the specification

$$D^{WS2}(Y,\alpha,\eta) = \left(\sum_{t=1}^{T}(\eta^{t\prime}Y^t)^{\alpha}\right)^{1/\alpha}.$$

Here, the period-t utility part is inside the sum, and is aggregated by the α terms. That is, $D^t(Y^t,\alpha,\eta^t) = \eta^{t\prime}Y^t$ and then the aggregator is given by $\left(\sum_{t=1}^{T}(d^t)^{\alpha}\right)^{1/\alpha}$. The case $\alpha = 1$ corresponds to simple summation (and satisfies time-separability), while $\alpha \neq 1$ allows nontrivial relationships across time.

A key restriction of weak separability is that it assumes that substitution/complementarity patterns across time enter only through the index D^t. This rules out substitution/complementarity patterns that directly depend on quantities. To allow such patterns, consider a parametric specification

$$D^{V1}(Y,\alpha,\eta) = \left(\sum_{\ell=1}^{L}\left(\sum_{t=1}^{T}\eta_{\ell}^{t}Y_{\ell}^{t}\right)^{\alpha}\right)^{1/\alpha}.$$

This specification states that preferences depend on an aggregator that depends on the weighted sum of quantities for each good. Another parametric specification is

$$D^{V2}(Y,\alpha,\eta) = \sum_{t=1}^{T}\sum_{\ell=1}^{L}\eta_{\ell}^{t\prime}Y_{\ell}^{t} + \left(\sum_{\ell=1}^{L}\left(\sum_{t=1}^{T}Y_{\ell}^{t}\right)^{\alpha}\right)^{1/\alpha}.$$

When $\alpha < 1$, there is diminishing marginal benefit to repeatedly choosing a good. This is a preference for variety that violates time-separability.

6.2.2 Time Swaps for a Specific Good

Both D^{V1} and D^{V2} satisfy Equation (14) for $\Pi \in \mathcal{G}^T$, and allow nontrivial preferences across time whenever $\alpha \neq 1$. These transformations also are invariant to permutations of time only for a specific good. That is, if we permute the time superscript on all variables with the same subscript ℓ, the function D^{V2} is unchanged. This means these transformations imply inequalities by comparing swaps across time for the same good.

We can write these inequalities as

$$\sum_{t=1}^{T} \mathbb{E}[Y_\ell^t - Y_\ell^{\pi(t)} \mid X]\beta_\ell^{t\prime} X_\ell^t \geqslant 0 \qquad a.s.,$$

where π encodes the transformation of time. Since no other quantities are permuted, the sum of differences only involves the good ℓ that is swapped. Note here that these inequalities are similar to the maximum score inequalities (Example 1). When we only permute time for one good ℓ, the invariance condition on the distribution of η is that conditional on (X, α), $\Pi(\eta)$ has the same distribution as η are unrestricted. Here, Π is a permutation swapping time for one good ℓ, and thus permutes as most T elements of the $K \times T$ dimensional vector η.

7 Discussion

This paper presents a new technique to difference out unobservables and generate identifying inequalities for slope coefficients. We apply the technique to generate inequalities in settings with a nesting structure. The analysis covers nested logit and also models of mixed continuous/discrete choice of multiple goods. We generalize the panel cyclic monotonicity inequalities of Shi et al. (2018) to all perturbed utility models, show they hold with preferences for variety, and show they are robust to latent feasibility sets.[15] We discuss some avenues for future work.

Remark 7 (Future Work on Inference). We outline some relevant dimensions for inference here. These dimensions interact in nontrivial ways.

The first key distinction is whether covariates are treated as discrete or continuous. For discrete covariates with finite support, Theorem 1 describes finitely many moment inequalities that "in principle," can be taken to data using the many tools designed for moment inequalities. For example, one can form a confidence set for β by inverting the test of Chernozhukov et al. (2019) or Bai et al. (2022).[16] Point identification is typically impossible with discrete covariates.[17] With continuous covariates, point

[15]We make this claim for the identifying inequalities, not the estimator used in Shi et al. (2018), which imposes additional regularity conditions.

[16]Note the moment inequalities have a linear structure that can potentially be exploited. There are often many inequalities when the set of transformations is large, e.g. permutations.

[17]Provided they also have finite and not countably infinite support. See Shi et al. (2018).

identification is possible depending on the nature of the invariance condition.[18]

The second key distinction is the specific invariance condition. Here we consider general invariance conditions, in which the number of identifying inequalities equals the cardinality of the set of transformations $\mathcal{G}$. For example, cyclic monotonicity studied in Shi et al. (2018) yields $T!$ inequalities in a panel setup, where T is the number of time periods.[19] Substantive simplifications are possible in this setup.[20] A general analysis of computational simplifications that exploit the specific structure of the set of transformations is beyond the scope of the paper.

The third key distinction is that analysis should depend on the ultimate object of interest. This paper focuses on identifying inequalities for β, and then in Section 4 shows these inequalities can be used for the dual purpose of counterfactual bounds once we have a candidate β. This analysis is amenable to a two-step approach in which one first conducts inference on β, and then does counterfactual inference for each candidate β.[21] An alternative approach is an integrated one-step approach that directly targets counterfactuals as in Christensen and Connault (2023) or Tebaldi et al. (2023). A one-step approach to counterfactuals may be possible by further studying the empirical content of the model, i.e. studying restrictions directly on the population distribution without the intermediate step of constructing inequalities for β. Another interesting topic is welfare analysis, which we do not study here. Welfare analysis is subtle when covariates and unobservables can violate (conditional) independence, because common welfare analysis proceeds by having a stable distribution of unobservable heterogeneity not one that changes with covariates.

Remark 8 (Future Work on Identification). More analysis is needed in panel contexts studying preferences that violate time-separability and allow forward-looking behavior with uncertainty. For example, we do not study binary choice panel models with state dependence like Honoré and Kyriazidou (2000) or Khan et al. (2019). These

[18] When regressors and unobservables are independent, Allen and Rehbeck (2019a) show with a slightly different setup and using different techniques that we can replace utility indices $\beta_k' X_k$ with unknown functions $u_k(X_k)$ and identify these functions with a single scale assumption.

[19] The main analysis of Shi et al. (2018) focuses on inequalities for cycles of length 2, which correspond to time swaps that only swap two time periods.

[20] Theorem S.7.1(iii) in the Supplemental Appendix of Allen and Rehbeck (2019c) shows how to write cyclic monotonicity with order T^2 inequalities. See also Chiong et al. (2021) and Franguridi (2021).

[21] See e.g. Chiong et al. (2021) which estimates a β and then conducts counterfactual analysis.

papers allow choices to depend on a lagged variable but do not model forward-looking behavior. While we do study a forward-looking panel model, we do not study uncertainty as in Aguirregabiria et al. (2021), which studies a multinomial choice fixed effects logit model. We also do not study sharpness. As mentioned above, we provide two separate assumptions that generate the cyclic monotonicity inequalities of Shi et al. (2018) – a panel stationarity assumption with time-separability and panel exchangeability with a preference for variety. It is unknown whether the identified sets coincide under these assumptions for general perturbed utility models.[22]

Appendix A Proof of Main Results

This section provides proofs for the results in the paper.

A.1 Proof of Lemma 1 and Theorem 1

We first prove Lemma 1 and then use it to prove Theorem 1.

Proof of Lemma 1. Define the random variable $\tilde{Y}^{\Pi}$ via

$$\tilde{Y}^{\Pi}(X, \alpha, \eta) = \Pi(Y(X, \alpha, \Pi^{-1}(\eta))),$$

which is possible since Assumption 2(i) ensures Π^{-1} exists. $\tilde{Y}^{\Pi}$ is supported on B by Assumption 2(iv). We obtain the almost sure equalities

$$\begin{aligned}\mathbb{E}[\Pi(Y(X, \alpha, \eta)) \mid X, \alpha] &= \mathbb{E}[\Pi(Y(X, \alpha, \Pi^{-1}(\eta))) \mid X, \alpha] \\ &= \mathbb{E}\left[\tilde{Y}^{\Pi}(X, \alpha, \eta) \mid X, \alpha\right].\end{aligned}$$

Assumption 2(v) ensures $\mathbb{E}[\Pi(Y(X, \alpha, \eta)) \mid X, \alpha]$ is well-defined. The first equality

[22] Pakes and Porter (2021) characterize the sharp identified set for a panel multinomial choice model when $T = 2$ with a panel stationarity condition. Pakes and Porter (2021) shows the inequalities of Shi et al. (2018) are not sharp in that setting. Mbakop (2022) shows when we strengthen the invariance condition to across-time exchangeability and have $T = 2$, there are additional identifying inequalities in that setting. We reiterate that we show the inequalities of Shi et al. (2018) hold for a *strictly* more general model of perturbed utility. In a stochastic choice context in which a single individual chooses a probability distribution over alternatives, Allen and Rehbeck (2019b) characterize the model for which cyclic monotonicity delivers sharp identifying inequalities. The key qualitative difference from classic discrete choice is that this model allows complementarity.

uses Assumption 2(ii), i.e. η and $\Pi^{-1}(\eta)$ have the same conditional distribution. The second equality is a definition. Thus, part (i) holds by subtraction. We also obtain the almost sure equalities

$$\begin{aligned}\mathbb{E}[D(\tilde{Y}^{\Pi}, \alpha, \eta) \mid X, \alpha] &= \mathbb{E}[D(\Pi(Y(X, \alpha, \Pi^{-1}(\eta))), \alpha, \eta) \mid X, \alpha] \\ &= \mathbb{E}[D(\Pi(Y(X, \alpha, \eta)), \alpha, \Pi(\eta)) \mid X, \alpha] \\ &= \mathbb{E}[D(Y(X, \alpha, \eta), \alpha, \eta) \mid X, \alpha].\end{aligned}$$

The first equality is a definition, the second equality uses Assumption 2(ii), and the third equality uses Assumption 2(iii), i.e. $D(\Pi(Y), \alpha, \Pi(\eta)) = D(Y, \alpha, \eta)$. These conditional expectations are well-defined because Assumption 1(iv) states that $D(Y, \alpha, \eta)$ has finite expectation. Thus, $\tilde{Y}^{\Pi}$ satisfies part (ii). □

Proof of Theorem 1. First note that when $\Pi_k(Y)\beta_k' X_k$ has finite expectation for each k, we can reproduce the proof of Lemma 1(i) to obtain

$$\mathbb{E}\left[\Pi_k(Y)\beta_k' X_k \mid X, \alpha\right] = \mathbb{E}\left[\tilde{Y}_k^{\Pi}\beta_k' X_k \mid X, \alpha\right].$$

In particular, since $\Pi_k(Y)\beta_k' X_k$ has finite expectation, so does $\tilde{Y}_k^{\Pi}\beta_k' X_k$.

With this setup, recall the main text establishes the revealed preference inequalities

$$\sum_{k=1}^{K} (Y_k - \tilde{Y}_k^{\Pi})\beta_k' X_k \geqslant D(\tilde{Y}^{\Pi}, \alpha, \eta) - D(Y, \alpha, \eta).$$

The expectations of each term in the sum exists, and the expectation of the right hand side exists by the proof of Lemma 1(ii). This implies existence of conditional expectations that satisfy the almost sure inequality

$$\sum_{k=1}^{K} \mathbb{E}[Y_k - \tilde{Y}_k^{\Pi} \mid X, \alpha]\beta_k' X_k \geqslant \mathbb{E}[D(\tilde{Y}^{\Pi}, \alpha, \eta) - D(Y, \alpha, \eta) \mid X, \alpha].$$

The right hand side is zero from Lemma 1(ii). Using Lemma 1 we modify the left hand side to obtain the almost sure inequality

$$\sum_{k=1}^{K} \mathbb{E}[Y_k - \Pi_k(Y) \mid X, \alpha]\beta_k' X_k \geqslant 0.$$

Integrating out α completes the proof. $\square$

A.2 Proof of Theorem 2

The proof technique is the same as Theorem 1. The change is we need to replace Lemma 1 with the following lemma.

Lemma A.1. *Let Assumption 1 and Assumption 3 hold and the set of transformations $\mathcal{G}$ satisfy Assumption 4. For every $\Pi \in \mathcal{G}$ there is a random variable $\tilde{Y}^{\Pi}$ supported on B such that:*

i. $\mathbb{E}[\tilde{Y}^{\Pi} - \Pi(Y) \mid X, \alpha] = 0$ *almost surely.*

ii. $\mathbb{E}\left[D(\tilde{Y}^{\Pi}, \alpha, \eta) - D(Y, \alpha, \eta) \mid X, \alpha\right] = 0$ *almost surely.*

Proof. The proof of part (i) is similar to that of Lemma 1. As before, define the random variable $\tilde{Y}^{\Pi}$ via

$$\tilde{Y}^{\Pi}(X, \alpha, \eta) = \Pi(Y(X, \alpha, \Pi^{-1}(\eta))).$$

Write

$$Y(X, \alpha, \eta) = (Y^1(X^1, \alpha, \eta^1)', \ldots, Y^T(X^T, \alpha, \eta^T)')',$$

which is possible because Assumption 3(ii) states each Y^t is (X^t, α, η^t)-measurable.

We obtain the almost sure equalities

$$\begin{aligned}
\mathbb{E}[\Pi(Y(X, \alpha, \eta)) \mid X, \alpha] &= \mathbb{E}[(\Pi^1(Y(X, \alpha, \eta))' \ldots \Pi^T(Y(X, \alpha, \eta))')' \mid X, \alpha] \\
&= \mathbb{E}\left[(Y^{\pi(1)}(X^{\pi(1)}, \alpha, \eta^{\pi(1)})', \ldots, Y^{\pi(T)}(X^{\pi(T)}, \alpha, \eta^{\pi(T)})')' \mid X, \alpha\right] \\
&= \mathbb{E}\left[(Y^{\pi(1)}(X^{\pi(1)}, \alpha, \eta^1)', \ldots, Y^{\pi(T)}(X^{\pi(T)}, \alpha, \eta^T)')' \mid X, \alpha\right] \\
&= \mathbb{E}[\Pi(Y(X, \alpha, \Pi^{-1}(\eta))) \mid X, \alpha] \\
&= \mathbb{E}\left[\tilde{Y}^{\Pi}(X, \alpha, \eta) \mid X, \alpha\right].
\end{aligned}$$

Assumption 4(vi) ensures $\mathbb{E}[\Pi(Y(X,\alpha,\eta)) \mid X,\alpha]$ is well-defined. The first equality represents Π in terms of blocks as described in the main text. The second equality uses the fact that Π is a time swap to write the transformation in terms of a permutation π of time indices. The third equality uses the fact that permutations are invertible and so $\eta^t = \eta^{\pi^{-1}(\pi(t))}$. In addition, $\eta^{\pi(t)}$ and η^t have the same conditional distributions by Assumptions 4(iii). The final two equalities are definitions.

Part (ii) follows from the almost sure equalities

$$\begin{aligned}
\mathbb{E}[D(\tilde{Y}^{\Pi},\alpha,\eta) \mid X,\alpha] &= \mathbb{E}[D(\Pi(Y(X,\alpha,\Pi^{-1}(\eta))),\alpha,\eta) \mid X,\alpha] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} D^t(\Pi^t(Y(X,\alpha,\Pi^{-1}(\eta))),\alpha,\eta^t) \mid X,\alpha\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} D^t(\Pi^t(Y(X,\alpha,\eta)),\alpha,\Pi^t(\eta)) \mid X,\alpha\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} D^t(Y(X,\alpha,\eta),\alpha,\eta^t) \mid X,\alpha\right] \\
&= \mathbb{E}[D(Y(X,\alpha,\eta),\alpha,\eta) \mid X,\alpha].
\end{aligned}$$

The first equality is a definition. The second equality rewrites D as formalized in Assumption 3(i), and uses the fact that each time-specific disturbance D^t only depends on the t-specific components of Y and η. To establish the third equality, recall Π is a time swap. Thus, $\Pi^t(Y(X,\alpha,\Pi^{-1}(\eta)) = Y^{\pi(t)}(X^{\pi(t)},\alpha,\eta^t)$ and hence

$$D^t(\Pi^t(Y(X,\alpha,\Pi^{-1}(\eta))),\alpha,\eta^t) := D^t(Y^{\pi(t)}(X^{\pi(t)},\alpha,\eta^t),\alpha,\eta^t).$$

This has the same conditional distribution when we replace η^t with $\eta^{\pi(t)} := \Pi^t(\eta)$, which completes the proof of the third equality. The fourth equality applies Assumption 4(iv). The final equality just rewrites the D^t summation in terms of the overall disturbance D. All expectations are finite by Assumption 1(iv). □

References

Afriat, S. N. (1967): "The Construction of Utility Functions from Expenditure Data," *International Economic Review*, 8, 67–77.

Agarwal, N. and P. Somaini (2018): "Demand Analysis using Strategic Reports: An application to a school choice mechanism," *Econometrica*, 86, 391–444.

Aguiar, V. H. and N. Kashaev (2019): "Identification and Estimation of Discrete Choice Models with Unobserved Choice Sets," *arXiv preprint arXiv:1907.04853*.

Aguirregabiria, V., J. Gu, and Y. Luo (2021): "Sufficient statistics for unobserved heterogeneity in structural dynamic logit models," *Journal of Econometrics*, 223, 280–311.

Allen, R. and J. Rehbeck (2019a): "Identification with additively separable heterogeneity," *Econometrica*, 87, 1021–1054.

——— (2019b): "Revealed Stochastic Choice with Attributes," *Available at SSRN 2818041*.

——— (2019c): "Supplement to "Identification With Additively Separable Heterogeneity"," *Econometrica Supplemental Material*, 87.

——— (2020): "Identification of random coefficient latent utility models," *arXiv preprint arXiv:2003.00276*.

Bai, Y., A. Santos, and A. M. Shaikh (2022): "A two-step method for testing many moment inequalities," *Journal of Business & Economic Statistics*, 40, 1070–1080.

Cardell, N. S. (1997): "Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity," *Econometric Theory*, 13, 185–213.

Chernozhukov, V., D. Chetverikov, and K. Kato (2019): "Inference on causal and structural parameters using many moment inequalities," *The Review of Economic Studies*, 86, 1867–1900.

Chiong, K. X., Y.-W. Hsieh, and M. Shum (2021): "Bounds on counterfactuals in semiparametric discrete-choice models," in *Handbook of Research Methods and Applications in Empirical Microeconomics*, Edward Elgar Publishing.

Christensen, T. and B. Connault (2023): "Counterfactual sensitivity and robustness," *Econometrica*, 91, 263–298.

Conlon, C. T. and J. H. Mortimer (2013): "Demand estimation under incomplete product availability," *American Economic Journal: Microeconomics*, 5, 1–30.

Crawford, G. S., R. Griffith, and A. Iaria (2021): "A survey of preference estimation with unobserved choice set heterogeneity," *Journal of Econometrics*, 222, 4–43.

Ershov, D., J.-W. Laliberté, and S. Orr (2018): "Mergers in a Model with Complementarity," Tech. rep., working Paper.

Fosgerau, M., J. Monardo, and A. De Palma (2022a): "The inverse product differentiation logit model," *Available at SSRN 3141041*.

Fosgerau, M., M. Paulsen, and T. K. Rasmussen (2022b): "A perturbed utility route choice model," *Transportation Research Part C: Emerging Technologies*, 136, 103514.

Fox, J. T. (2007): "Semiparametric estimation of multinomial discrete-choice models using a subset of choices," *The RAND Journal of Economics*, 38, 1002–1019.

Fox, J. T., C. Yang, and D. H. Hsu (2018): "Unobserved heterogeneity in matching games," *Journal of Political Economy*, 126, 1339–1373.

Franguridi, G. (2021): "Efficient counterfactual estimation in semiparametric discrete choice models: a note on Chiong, Hsieh, and Shum (2017)," *arXiv preprint arXiv:2112.04637*.

Galichon, A. (2022): "On the representation of the nested logit model," *Econometric Theory*, 38, 370–380.

Gentzkow, M. (2007): "Valuing new goods in a model with complementarity: Online newspapers," *American Economic Review*, 97, 713–744.

Goeree, J. K., C. A. Holt, and T. R. Palfrey (2005): "Regular quantal response equilibrium," *Experimental Economics*, 8, 347–367.

GOEREE, M. S. (2008): "Limited information and advertising in the US personal computer industry," *Econometrica*, 76, 1017–1074.

HONORÉ, B. E. AND E. KYRIAZIDOU (2000): "Panel data discrete choice models with lagged dependent variables," *Econometrica*, 68, 839–874.

IARIA, A. AND A. WANG (2019): "Identification and Estimation of Demand for Bundles," Tech. rep., working Paper.

——— (2021): "An empirical model of quantity discounts with large choice sets," .

KHAN, S., M. PONOMAREVA, E. T. TAMER, ET AL. (2019): *Identification of dynamic panel binary response models*, Boston College.

LU, Z. (2022): "Estimating multinomial choice models with unobserved choice sets," *Journal of Econometrics*, 226, 368–398.

MANSKI, C. F. (1975): "Maximum score estimation of the stochastic utility model of choice," *Journal of Econometrics*, 3, 205–228.

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

——— (1987): "Semiparametric analysis of random effects linear models from binary panel data," *Econometrica: Journal of the Econometric Society*, 357–362.

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

MATZKIN, R. L. (1993): "Nonparametric identification and estimation of polychotomous choice models," *Journal of Econometrics*, 58, 137–168.

MBAKOP, E. (2022): "Identification in Some Discrete Choice Models: A Computational Approach," *Available at SSRN 4228761*.

MCFADDEN, D. (1981): "Econometric Models of Probabilistic Choice," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. F. Manski and D. McFadden, Cambridge, MA: MIT Press, 198–272.

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

Pakes, A. (2010): "Alternative models for moment inequalities," *Econometrica*, 78, 1783–1822.

Pakes, A. and J. Porter (2021): "Moment Inequalities for Multinomial Choice with Fixed Effects," .

Pakes, A., J. Porter, K. Ho, and J. Ishii (2015): "Moment inequalities and their application," *Econometrica*, 83, 315–334.

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

Shi, X., M. Shum, and W. Song (2018): "Estimating Semi-Parametric Panel Multinomial Choice Models Using Cyclic Monotonicity," *Econometrica*, 86, 737–761.

Tebaldi, P., A. Torgovitsky, and H. Yang (2023): "Nonparametric estimates of demand in the california health insurance exchange," *Econometrica*, 91, 107–146.

Wang, R. (2021): "Semiparametric Identification and Estimation of Substitution Patterns," .

Yan, Z., K. Natarajan, C. P. Teo, and C. Cheng (2022): "A representative consumer model in data-driven multiproduct pricing optimization," *Management Science*, 68, 5798–5827.