COUNTERFACTUAL ESTIMATION IN SEMIPARAMETRIC DISCRETE-CHOICE MODELS

KHAI X. CHIONG[1], YU-WEI HSIEH[2], AND MATTHEW SHUM[3]

Abstract. We show how to construct bounds on counterfactual choice probabilities in semiparametric discrete-choice models. Our procedure is based on cyclic monotonicity, a convex-analytic property of the random utility discrete-choice model. These bounds are useful for typical counterfactual exercises in aggregate discrete-choice demand models. In our semiparametric approach, we do not specify the parametric distribution for the utility shocks, thus accommodating a wide variety of substitution patterns among alternatives. Computation of the counterfactual bounds is a tractable linear programming problem. We illustrate our approach in a series of Monte Carlo simulations and an empirical application using scanner data.

Keywords: Semiparametric Discrete-choice Models; Counterfactuals; Convex Analysis, Cyclic Monotonicity; Linear Programming

JEL Classification: C14, C25, C53.

Date: *Initial draft:* November 2016; *This draft:* June 8, 2017.

[1]USC Dornsife INET, Department of Economics, University of Southern California. kchiong@usc.edu

[2]University of Southern California. E-mail: yuwei.hsieh@usc.edu

[3]California Institute of Technology. E-mail: mshum@caltech.edu

Acknowledgement: We thank Whitney Newey and seminar participants in Ohio State, Texas-Austin, UBC, and UCLA for helpful comments.

1. Introduction

Since the seminal work of McFadden (1978, 1981, 1984), discrete choice modeling based on the random utility framework has been at the core of modern empirical research in many areas of economics, including labor, public finance, industrial organization, and health economics. Many applications of discrete choice models utilize strong parametric assumptions (such as probit or logit), and hence are not robust to mis-specification. Moreover, the parametric assumptions can lead to unrealistic substitution patterns in the subsequent counterfactual analysis. There is now a growing literature on semiparametric estimation of discrete choice models[1] aiming to resolve the robustness issue in estimation.

Our contribution is to address the problem of counterfactual evaluations in semiparametric discrete-choice models. We show how to obtain counterfactual choice probabilities or market shares in these models even if the error distribution is not specified. To the best of our knowledge, this is the first paper to address the estimation of counterfactuals in a semiparametric *multinomial* choice model. In the case of semiparametric *binary* choice models, Hausman et al. (1998) propose an approach based on isotonic regression, which can be used to construct counterfactual choice probabilities without knowledge of the distribution of the random error terms. This idea, however, only works the binary case. We provide an approach which works for both multinomial and binary choice settings.[2]

Counterfactual evaluations is often the primary goal in industrial organization and marketing, such as predicting the market shares under the new pricing scheme, or conducting merger simulations. Despite the potential pitfall of imposing strong assumptions on substitution patterns, parametric models such as mixed logit or probit remain the dominating approach for practitioners[3] because counterfactuals can be easily obtained via simulation. The applicability of semiparametric estimators, by contrast, is largely limited by the absence of methods for counterfactual evaluations.

In this paper, we propose using *cyclic monotonicity*, a *cross-market* restriction on market-level data derived from discrete-choice theory, to bound the counterfactual market shares in a semiparametric multinomial choice framework. We do not impose a functional form for the probability distribution of the utility shocks, and hence the counterfactual market

[1]Examples include Manski (1975), Han (1987), among many others.

[2]Recently, Allen and Rehbeck (2016) study the nonparametric identification of counterfactuals in a class of perturbed utility models, which includes random utility multinomial choice models.

[3]see, for instance, Train (2003), Ackerberg et al. (2007), Allenby and Rossi (2005)

shares cannot be obtained even via simulations. Since this approach is semiparametric, it accommodates rich substitution patterns among alternatives. Such generality comes at the cost of losing point identification in general. We also show that our construction of counterfactual bounds is a tractable Linear Programming (LP) formulation.

In the next section we lay out the assumptions and key results on cyclic monotonicity. In section 3 we discuss how to construct bounds for counterfactual market shares, the computational issue, as well as the economic meaning of cyclic monotonicity. In section 4 we show that bound is very informative using a Monte Carlo experiment. In section 5 we show our method can produce very different substitution patterns using actual data.

2. Setup

There are $\mathcal{M} = \{1, \ldots, M\}$ markets. In each market, there are $\mathcal{J} = \{1, \ldots, J\}$ alternatives or products. The latent indirect utility that a consumer i derives from product j at market m is given by $U_{ij}^m = \boldsymbol{X}_j^m \boldsymbol{\beta} + \epsilon_{ij}^m$, where $\boldsymbol{X}_j^m$ is a $1 \times b$ vector of observed product-specific attributes, $\boldsymbol{\beta}$ is a $b \times 1$ vector of unknown parameters, and $\boldsymbol{\epsilon}_i^m = (\epsilon_{i1}^m, \ldots, \epsilon_{iJ}^m)'$ is a vector of latent utility shocks. Further, let the market share of product j in market m be $s_j^m = \Pr(\boldsymbol{X}_j^m \boldsymbol{\beta} + \epsilon_{ij}^m \geq \max_{k \neq j}\{\boldsymbol{X}_k^m \boldsymbol{\beta} + \epsilon_{ik}^m\})$. The model is semiparametric as we do not specify the distribution of the utility shocks $\boldsymbol{\epsilon}_i^m$. Similar to Berry (1994), we denote the mean utility as $\delta_j^m \equiv \boldsymbol{X}_j^m \boldsymbol{\beta}$.

The basic idea is, under the assumption that shocks across individuals and markets are identical and independently distributed (Assumption 1), the mean utilities and the observed market shares must satisfy a set of inequalities (Proposition 1). To conduct a counterfactual exercise, suppose then the counterfactual market is characterized by $\tilde{\boldsymbol{X}}_j, j = 1, \ldots, J$. Its mean utility vector can be obtained by $\tilde{\boldsymbol{X}}_j \hat{\boldsymbol{\beta}}$. We then use the aforementioned cross-market restrictions to infer the set of counterfactual market shares that are consistent with (i) all the mean utilities, and (ii) all the observed market shares.

We now formally state Assumption 1 and Proposition 1 which serve as the backbone of our counterfactual procedure.

Assumption 1. The vector of utility shocks $\boldsymbol{\epsilon}_i^m$ is distributed identically and independently across individual i and market $m = 1, \ldots, M$, with the joint distribution F. Further, F does not depend on $(\boldsymbol{X}_j^m)_{j=1,\ldots,J}$, that is, $\boldsymbol{X}_j^m$ and $\boldsymbol{\epsilon}_i^m$ are independent.

Our assumption allows the utility shocks to follow an unknown joint distribution that can be arbitrarily correlated among different products j. This accommodates many discrete-choice demand model specifications in the literature, including multinomial logit, nested logit (Goldberg (1995)), cross-nested logit (Bresnahan et al. (1997)), and multinomial probit (Goolsbee and Petrin (2004)).[4]

The key restriction we use for constructing counterfactual market shares is the property of *cyclic monotonicity*, which we define next. Let $\boldsymbol{s}^m$ be the vector of market shares evaluated at the mean utilities $\boldsymbol{\delta}^m$. That is, $s_j^m = \Pr(\delta_j^m + \epsilon_{ij}^m \geq \max_{k \neq j}\{\delta_k^m + \epsilon_{ik}^m\})$, and $\boldsymbol{s}^m = (s_1^m, \ldots, s_J^m)$.

Definition 1 (Cyclic Monotonicity)**:** Define a cycle of length K as a permutation of $K-1$ distinct elements from $\{1, 2, \ldots, M\}$. Denote a generic cycle of length K by $(l_1, l_2, \ldots, l_K, l_{K+1})$ with $l_{K+1} = l_1$. The market shares $\boldsymbol{s}^m$ and mean utilities $\boldsymbol{\delta}^m$ satisfy cyclic monotonicity if the following inequality (1) holds for all possible cycles of length K and for all $K \geq 2$.

$$\sum_{k=1}^{K} (\boldsymbol{\delta}^{l_{k+1}} - \boldsymbol{\delta}^{l_k}) \cdot \boldsymbol{s}^{l_k} \leq 0 \tag{1}$$

Cyclic monotonicity is a defining property of the gradients of vector-valued convex functions, analogous to how monotonicity is a property of the derivatives of scalar-valued convex functions. We exploit cyclic monotonicity to construct counterfactual market shares using the next proposition.

Proposition 1. *Under Assumption 1, the market shares $\boldsymbol{s}^m$ and mean utilities $\boldsymbol{\delta}^m$ for all markets $m = 1, \ldots, M$ satisfy* cyclic monotonicity.

Proof. Proposition 1 arises from the convexity properties of the *social surplus function* (or the expected indirect utility) of the discrete choice problem (McFadden (1978, 1981)):[5]

$$\mathcal{G}(\boldsymbol{\delta}) = \mathbb{E}_F \left[\max_{j \in \{1,\ldots,J\}} (\boldsymbol{X}_j \boldsymbol{\beta} + \epsilon_{ij}) \middle| \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\delta} \right],$$

[4]However, the framework does not apply to random-coefficient logit models of demand (Berry et al. (1995)).

[5]See Fosgerau and De Palma (2015), Shi et al. (2016) and Chiong and Shum (2016) for full details

where $\boldsymbol{X\beta} = (\boldsymbol{X}_1\boldsymbol{\beta}, \ldots, \boldsymbol{X}_J\boldsymbol{\beta})$, and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_J)$. Under Assumption 1 we have:

$$\mathcal{G}(\boldsymbol{\delta}) = \mathbb{E}_F \left[\max_{j \in \{1,\ldots,J\}} (\delta_j + \epsilon_{ij}) \right],$$

which is a convex function of the vector of mean utilities $\boldsymbol{\delta}$. The social surplus function has an important property. Define the function $\boldsymbol{s} : \mathcal{U} \subset \mathbb{R}^J \to \Delta^J$ as the mapping from mean utilities to market shares, where the j-th component of $\boldsymbol{s}(\boldsymbol{\delta})$ is $s_j(\boldsymbol{\delta}) = \Pr(\delta_j + \epsilon_{ij}^m \geq \max_{k \neq j}\{\delta_k + \epsilon_{ik}^m\})$, and $\mathcal{U}$ is a convex super-set of the unknown true mean utilities $\{\boldsymbol{\delta}^1, \ldots, \boldsymbol{\delta}^m\}$. By the Williams-Daly-Zachary Theorem[6], $\boldsymbol{s}(\boldsymbol{\delta})$ lies in the subgradient of $\mathcal{G}$ evaluated at $\boldsymbol{\delta}$, for all $\boldsymbol{\delta} \in \mathcal{U}$. That is,

$$\boldsymbol{s}(\boldsymbol{\delta}) \in \partial\mathcal{G}(\boldsymbol{\delta}). \tag{2}$$

By a well-known result in convex analysis (Rockafellar (1970)), the subgradient of a convex function satisfies *cyclic monotonicity*. More precisely, because the mapping $\boldsymbol{s}$ is a gradient function of a convex function, it follows that for every cycle $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n, \boldsymbol{x}_{n+1} = \boldsymbol{x}_0$ in $\mathcal{U}$, we have $\boldsymbol{s}(\boldsymbol{x}_0) \cdot (\boldsymbol{x}_1 - \boldsymbol{x}_0) + \boldsymbol{s}(\boldsymbol{x}_1) \cdot (\boldsymbol{x}_2 - \boldsymbol{x}_1) + \cdots + \boldsymbol{s}(\boldsymbol{x}_n) \cdot (\boldsymbol{x}_{n+1} - \boldsymbol{x}_n) \leq 0$.[7] In particular, we can take the cycle in $\mathcal{U}$ consisting of the unknown mean utilities. This proves Proposition 1. □

For binary-choice models, semiparametric estimation using the maximum score or maximum rank correlation approaches (Han (1987); Manski (1975)) exploits the feature that the (scalar-valued) choice probability function $s(\delta)$ is increasing in the mean utility δ. Cyclic monotonicity is a natural extension of this insight into multinomial-choice models, where cyclic monotonicity formalizes the idea that $\boldsymbol{s}(\boldsymbol{\delta})$ is monotonic in $\boldsymbol{\delta}$ in a vector-sense.

---

[6] See Rust (1994). Chiong et al. (2016) generalize it to the case when the social surplus function may be non-differentiable, corresponding to cases where the utility shocks $\boldsymbol{\epsilon}$ have bounded support or follow a discrete distribution.

[7] See Rockafellar (1970, Theorem 23.5); Villani (2003). Conversely, any function that satisfies cyclic monotonicity must be a subgradient of some convex function.

3. Counterfactuals Implied by Cyclic Monotonicity

In this section, we show how to compute the counterfactual market shares using the cyclic monotonicity restriction (Proposition 1). We assume that the first-step estimation of the parameters $\boldsymbol{\beta}$ has been carried out by some semiparametric estimator such as Fox (2007) or Shi et al. (2016).[8] The researchers observe $\boldsymbol{s} = \left(\boldsymbol{s}^1, \ldots, \boldsymbol{s}^M\right)$ vectors of market shares in markets $m = 1, \ldots, M$, and the corresponding mean utilities $\boldsymbol{\delta} = \left(\boldsymbol{\delta}^1, \ldots, \boldsymbol{\delta}^M\right)$ with $\delta_j^m = \boldsymbol{X}_j^m \boldsymbol{\beta}$. The counterfactual market, index by market $M+1$, is characterized by its mean utility vector $\boldsymbol{\delta}^{M+1}$. The corresponding market share vector is denoted by $\boldsymbol{s}^{M+1}$. Since $\boldsymbol{\delta}^{M+1}$ is known by construction, the remaining problem is to determine $\boldsymbol{s}^{M+1}$. First, market shares must be non-negative and must add up to one, we have the following constraints:

$$\sum_{j=1}^{J} s_j^{M+1} = 1$$
$$s_j^{M+1} \geq 0 \quad \forall j$$

Second, the counterfactual market $(\boldsymbol{\delta}^{M+1}, \boldsymbol{s}^{M+1})$ must satisfy the cyclic monotonicity for any cycle containing $M+1$. That is, given any cycle of length K, $(l_1, l_2, \ldots, l_K, l_1)$, such that $l_K = M+1$ corresponds to the counterfactual market, cyclic monotonicity implies that

$$\sum_{k=1}^{K} (\boldsymbol{\delta}^{l_{k+1}} - \boldsymbol{\delta}^{l_k}) \cdot \boldsymbol{s}^{l_k} \leq 0. \tag{3}$$

Notice that where the cycle starts and ends is irrelevant to the definition of cyclic monotonicity as only the "sum" matters. For example, the cycle $(1,2,3,1)$ is equivalent to $(2,3,1,2)$ and $(3,1,2,3)$ when computing (3). Therefore, without loss of generality one can always put the counterfactual market at the end of the cycle.

[8]While Fox's (2007) estimator is based on the *rank-order property*, which involves utility comparisons amongst all pairs of options in the choice set, it turns out that the bounds on market shares implied by the rank-order property have undesirable properties for the evaluation of counterfactual choice probabilities, which we describe in Appendix A. Because of that we focus on the implications of cyclic monotonicity in this paper.

3.1. Computing Bounds for Market Shares

Given the constraints in the previous section, the upper bound for the counterfactual market share of the i-th good is given by the linear programming (LP) problem:

$$\begin{aligned} & \max s_i^{M+1} \\ \text{s.t.} \quad & \sum_{j=1}^{J} s_j^{M+1} = 1 \\ & \sum_{k=1}^{K} (\boldsymbol{\delta}^{l_{k+1}} - \boldsymbol{\delta}^{l_k}) \cdot \boldsymbol{s}^{l_k} \leq 0;\ l_K = M+1;\ 2 \leq K \leq M \\ & s_j^{M+1} \geq 0 \quad \forall j \end{aligned} \tag{4}$$

Similarly, the lower bound can be found by changing maximization to minimization. Several policy relevant counterfactuals can be computed by choosing a suitable objective function and mean utilities $\boldsymbol{\delta}^{M+1}$. As examples, the elasticity of substitution matrix is often a key input for merger simulations (Nevo (2000)); counterfactual market shares resulting from large price changes are also used to evaluate the welfare benefits of new product introductions (Hausman (1996), Petrin (2002)). For evaluating the effects of price changes, we proceed by introducing a counterfactual market, which we label $M+1$, which is identical to the benchmark market m, except that the price of product i in market $M+1$ is higher than p_i^m. By solving (4) for all $j \in \mathcal{J}$ one is able to bound the effect of the price increase $p_i^{M+1} - p_i^m$ on market shares of product j.

More complicated counterfactuals are possible. For example, a multi-product firm may want to predict the maximum total share $(s_1^{M+1} + s_2^{M+1})$ or revenue $(p_1^{M+1} \cdot s_1^{M+1} + p_2^{M+1} \cdot s_2^{M+1})$ under the new pricing scheme (p_1^{M+1}, p_2^{M+1}), holding other things equal. This can be done by choosing a benchmark market first, and change the objective function.

3.2. Linear Programming Formulation

At the first glance the LP problem of (4) appears to be computationally intensive, because even considering cycles up to length 3 would result in $M + \binom{M}{2} \times 2!$ constraints, which is proportional to M^2. To better understand the complexity of the LP problem, we first need to express the constraints associated with cyclic monotonicity in terms of

$$\boldsymbol{A} \cdot (\boldsymbol{s}^{M+1})' \leq \boldsymbol{b} \tag{5}$$

$\boldsymbol{A}$ and $\boldsymbol{b}$ will be referred as the constraint matrix and the right-hand side vector, and $(\boldsymbol{s}^{M+1})'$ are the unknown parameters (the vector of counterfactual market shares) to be determined.[9] The dimension of $\boldsymbol{b}$, and the number of rows in $\boldsymbol{A}$, is equal to the number of cyclic monotonicity inequalities, i.e. the number of possible cycles. For a given cycle $(l_1, l_2, \ldots, M+1, l_1)$, we will next derive the corresponding row and entry in $\boldsymbol{A}$ and $\boldsymbol{b}$ corresponding to this cycle. Now consider the following $K \times J$ matrices $\boldsymbol{D}$ and $\boldsymbol{S}$ stacked by the relevant utility and market share row vectors:

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{\delta}^{l_2} \\ \boldsymbol{\delta}^{l_3} \\ \vdots \\ \boldsymbol{\delta}^{M+1} \\ \hdashline \boldsymbol{\delta}^{l_1} \end{bmatrix} - \begin{bmatrix} \boldsymbol{\delta}^{l_1} \\ \boldsymbol{\delta}^{l_2} \\ \vdots \\ \boldsymbol{\delta}^{l_{K-1}} \\ \hdashline \boldsymbol{\delta}^{M+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{D_1} \\ \hdashline \boldsymbol{\delta}^{l_1} - \boldsymbol{\delta}^{M+1} \end{bmatrix}$$

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{s}^{l_1} \\ \boldsymbol{s}^{l_2} \\ \vdots \\ \boldsymbol{s}^{l_{K-1}} \\ \hdashline \boldsymbol{s}^{M+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{S_1} \\ \hdashline \boldsymbol{s}^{M+1} \end{bmatrix}.$$

The cyclic monotonicity inequality corresponding to this cycle (see Equation (3)) can be written as

$$(\boldsymbol{\delta}^{l_1} - \boldsymbol{\delta}^{M+1}) \cdot (\boldsymbol{s}^{M+1})' \leq -\mathbf{1}'(\boldsymbol{D}_1 \circ \boldsymbol{S}_1)\mathbf{1}, \tag{6}$$

where $\mathbf{1}$ is a $(K-1) \times 1$ vector of 1s and $\circ$ is the Hadamard product. For this cycle, then, the corresponding row of $\boldsymbol{A}$ and entry in $\boldsymbol{b}$ is given by (6).

It turns out that, in this way, by rewriting the cyclic monotonicity inequalities for each cycle, we can show that the number of effective constraints in the LP problem is much smaller than the number of cycles. The following Lemma shows that our counterfactual

[9] For notational convenience, in this section, we will take the vectors of mean utilities and market shares ($\boldsymbol{\delta}^i$ and $\boldsymbol{s}^i$) as $1 \times J$ row vectors.

problem can always be formulated as a LP problem with at most M distinct constraint inequalities, where M is the number of markets (the LP problem also includes J non-negativity constraints).

Lemma 1. *Expressing the LP constraints imposed by cyclic monotonicity in the form of (5), there are at most M distinct rows in $\boldsymbol{A}$.*

Proof. From (6), the constraint coefficient always takes the form $\boldsymbol{\delta}^i - \boldsymbol{\delta}^{M+1}$, where $i \in \{1, \ldots, M\}$. Since there are M markets, there are at most M distinct constraint coefficients.

□

Lemma 1 implies that our counterfactual procedure is a computationally easy LP problem.[10] Although the number of constraints appears to increase exponentially as the number of markets increases, Lemma 1 implies that most of them are parallel to each other, which are then easily eliminated in a pre-solved step.

3.3. Gross Substitution

In this section, we analytically demonstrate a simple counterfactual using the LP formulation above. Suppose we increase the price of product j in market m, our counterfactual framework shows that its market share weakly decreases. In another words, there is at least one product $i \in \mathcal{J}\backslash\{j\}$ such that products i, j are weakly substitute.

Proposition 2. *Suppose the regressors contain price for each product p_j, and the price coefficient $\beta_p < 0$ (normal goods). Consider a counterfactual market $M+1$, which is identical to the benchmark market m, except that δ_j^{M+1} differs from δ_j^m due to the price change of product j. Cyclic monotonicity implies that $(p_j^{M+1} - p_j^m)(s_j^{M+1} - s_j^m) \leq 0$. Since the market share of product j weakly decreases when its price increases, there must be some product in $J\backslash\{j\}$ whose market share weakly increases.*

Proof. Consider the cycle of length 2 containing both the benchmark and counterfactual market $(m, M+1)$. We have $(\boldsymbol{\delta}^{M+1} - \boldsymbol{\delta}^m) \cdot \boldsymbol{s}^m + (\boldsymbol{\delta}^m - \boldsymbol{\delta}^{M+1}) \cdot \boldsymbol{s}^{M+1} \leq 0$. $\because (\boldsymbol{\delta}^{M+1} - \boldsymbol{\delta}^m) =$

[10] The main computational bottleneck is to find all possible cycles itself (generate constraints) due to its combinatorial nature, not linear programming.

$[0, \ldots, \beta_p(p_j^{M+1} - p_j^m), \ldots, 0]$, $\therefore \beta_p(p_j^{M+1} - p_j^m)\boldsymbol{s}_j^m - \beta_p(p_j^{M+1} - p_j^m)\boldsymbol{s}_j^{M+1} \leq 0$. If $\beta_p \leq 0$, we have $(p_j^{M+1} - p_j^m) \cdot (s_j^{M+1} - \boldsymbol{s}_j^m) \leq 0$. □

While our framework asserts that the market share of a product is weakly decreasing in its price, it does not assert gross substitution more broadly. That is, the requirement that whenever the price of j increases, the market shares for *all* other products weakly increase. Our framework thus allows for some degree of complementarity.

When desired, we can easily impose gross substitution in our counterfactual framework. Suppose we are interested in the counterfactual of increasing the price of product v in market m. Then the upper bound counterfactual market share for products $i = 1, \ldots, J$ is given by:

$$
\begin{aligned}
&\max s_i^{M+1} \\
\text{s.t.}\quad &\sum_{j=1}^{J} s_j^{M+1} = 1 \\
&\sum_{k=1}^{K} (\boldsymbol{\delta}^{l_{k+1}} - \boldsymbol{\delta}^{l_k}) \cdot \boldsymbol{s}^{l_k} \leq 0;\ l_K = M+1;\ 2 \leq K \leq M \\
&s_j^{M+1} \geq s_j^m \quad \forall j \neq v \quad \text{(gross substitution constraints)} \\
&s_v^{M+1} \geq 0
\end{aligned} \tag{7}
$$

When we conduct the counterfactual of decreasing the price of v, then the gross substitution constraints are $s_j^{M+1} \leq s_j^m$ for all $j \neq v$. For some problems, it is reasonable to assume that all products are gross substitutes.[11] Imposing gross substitution also narrows down the identified set.

4. Monte Carlo simulations

In this section we conduct a Monte Carlo simulation to study the identified set of counterfactuals. The main finding is that our counterfactual bounds always cover the true Logit counterfactual market shares in all 100 runs, therefore it has a coverage probability of

[11] Berry et al. (2013) show that certain forms of gross substitution, in particular connected substitution, are sufficient for the invertibility of demand, which is fundamental for many demand estimation methods such as Berry et al. (1995).

100%. We also find that the length of the bounds is typically tight, with the worst-case average of around 5% of market shares (see Table 1).

Table 1. Monte Carlo simulation: (upper bound − lower bound)

	$M = 200$			$M = 500$			$M = 1000$		
	s_1	s_2	s_3	s_1	s_2	s_3	s_1	s_2	s_3
p_1 +1%	0.0417	0.0263	0.0525	0.0297	0.0161	0.0340	0.0233	0.0114	0.0256
	(.0012)	(.0007)	(.0013)	(.0008)	(.0004)	(.0013)	(.0007)	(.0004)	(.0008)
p_2 +1%	0.0566	0.0151	0.0581	0.0355	0.0113	0.0364	0.0263	0.0090	0.0283
	(.0018)	(.0005)	(.0018)	(.0009)	(.0003)	(.0009)	(.0008)	(.0002)	(.0008)
p_3 +1%	0.0474	0.0260	0.0416	0.0334	0.0165	0.0306	0.0240	0.0122	0.0231
	(.0013)	(.0007)	(.0013)	(.0009)	(.0005)	(.0008)	(.0008)	(.0003)	(.0006)

Reported figures are averages across 100 replications. We report the width of the bounds rather than coverage, because our bounds always cover the true (Logit) counterfactual, and so the coverage probability is 100%.

We first generate (market-invariant) product-specific regressors $X_j = (x_{j1}, x_{j2}, x_{j3})$ from multivariate normal with

$$\mu = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & -0.7 & 0.3 \\ \cdot & 1 & 0.3 \\ \cdot & \cdot & 1 \end{bmatrix}.$$

The price of product j, p_j^m, in market m is generated according to

$$p_j^m = |1.1(x_{j1} + x_{j2} + x_{j3}) + \epsilon_j^m|$$

$$\epsilon_j^m \sim \mathcal{N}(0, 0.3^2).$$

The mean utility δ_j^m is then computed as $[X_j, p_j^m]\boldsymbol{\beta}$, where $\boldsymbol{\beta} = [1.5, 1.5, 0.8, -2.2]'$. The market shares are computed under the Logit model, assuming a Type-I Extreme Value distribution. We set the number of products to be $J = 3$. We first generate one benchmark market, which is then held fixed across Monte Carlo repetitions. For each Monte Carlo repetition, we generate another $M - 1$ markets from the above DGP. We then compute the upper and lower bounds counterfactuals s_i^{M+1} for $i = 1, 2, 3$ under $+1\%$ increase in $p_j^{\bar{m}}$, $j = 1, 2, 3$ of the benchmark market $\bar{m}$. We repeat this procedure 100 times, and report the average length of the bounds in Table 1. The standard deviations across the 100 repetitions are reported in parentheses.

We also vary the number of markets, that is, $M \in \{200, 500, 1000\}$. The results are summarized in Table 1. We use only cycles of length 2, but using cycles up to length 3 does not make much of a difference.

5. Empirical illustration: Supermarket sales of coffee

For our empirical application, we use the IRI Marketing Dataset (Bronnenberg et al. (2008)) – which is a retail scanner dataset containing weekly pricing, sales volume, and promotion data of participating retail stores. We will focus on the product category of *coffee*, within the time-frame of the year 2005.

Each observation in the IRI dataset is at the level of store-product-week. We take all stores and aggregate them at the level of 6 major metropolitan areas: Boston, Chicago, Dallas, Houston, Milwaukee, Washington DC. As such, the unit of observation is (j, t), where j is the product, and t is the market – each market is defined as a week-metro combination. For instance, Houston during Week 3 is one such market. Therefore, there are 318 markets in total.

The consumer's choice set consists of 7 brands: *Chock Full O Nuts, Eight O Clock, Folgers, Maxwell House, Private Label, Seattle's Best, Starbucks.* Now *Private Label* is the brand of coffee carried by the participating individual stores. We also include an *Outside Option* as a composite good consisting of all other smaller coffee brands. The average market share of a coffee brand is tabulated in Table 2.

The dependent variable is $Share$, which is the number of units sold in a market divided by the total number of all units sold in a market. The main covariate is $Price$, which is the average retail price paid per unit of the product in a market. We have three other covariates measuring the degree of product promotions: *PR, Display,* and *Feature.* The variable PR is a binary variable indicating a temporary price reduction of 5% or more. The variable $Display$ takes values in $\{0, 1, 2\}$, and is defined by IRI as the degree to which the product is exhibited prominently in the store. The variable $Feature$ takes values in $\{0, 1, \dots, 4\}$ which are coded as the degree of advertisement featured by the store retailer.

These product-specific variables are aggregated to the level of week-metro by means of weighted average across stores, the weights used are the number of units sold.[12]

Brands	Average Market Share (%)
Chock Full O Nuts	1.86
Eight O Clock	6.10
Folgers	27.38
Maxwell House	15.59
Outside	29.33
Private Label	10.53
Seattles Best	1.22
Starbucks	7.99

Table 2. Average market share of a brand across 318 markets.

5.1. Result

First, we obtain point estimates of the model coefficients using the semiparametric method of Shi, Shum and Song (2017). Then, we run a series of counterfactual exercises as follows: for each brand, we increase its price in the median market by $1\%, 2\%, \ldots, 20\%$, and then compute the changes in the market shares of all brands in the median market.[13]

The counterfactual market share is partially identified. In fact, our identified set is multi-dimensional (one dimension for each product). We slice the identified set in the following way to further narrow down the identified set. First, we look at the set of counterfactual market shares that are consistent with Gross Substitution (as discussed in Section 3.3). That is, we restrict attention to only those counterfactual market shares such that s_j is weakly increasing for all $j \neq i$ when we increase the price of i. Second, we look at the component-wise midpoints of the identified set.

The main result is given in Figure 1. We plot the absolute increases in market shares for different brands as a function of percentage increases in the price of a reference brand.

[12] For example, PR_{jt}, it is constructed as the number of units of j sold that had a temporary price-reduction in market t divided by the total number of units of j sold in market t.

[13] The median market in the data is determined using the algorithm PAM (Partitioning Around Medoid).

The changes in market shares are relative to the pre-counterfactual benchmark market, which is the median market.

In Figure 1(c), we increase the price of the reference brand, Folgers, and examine the substitution effects on other brands. (We label the lines in Figures 1 and 2 such that the brand with the highest substitution strength appears first in the graph legend, and so on.) We see that Seattle's Best is the coffee brand that benefits the least, its market share increases from 1% to 4.5% as the price of Folgers increases from 1% to 20%. In comparison, Starbucks is a stronger substitute – its market share increases from 1.5% to 7% at the same range.

Crucially, our framework allows for a much richer pattern of substitutions than the Logit model. The Logit model implies that the elasticity of substitution from product j to i does not depend on the identity of j. This leads to Figure 2, where the ordering of brands in terms of substitution strengths is fixed. In contrast, the ordering of how strongly one brand substitutes another brand changes widely depending on the reference brand.

For example, our analysis shows that (i) Starbucks is a very strong substitute to Seattle's Best while being a weak substitute to Eight O Clock. (ii) Folgers is a strong substitute to all other brands except Seattle's Best and Starbucks. Using the Logit model, such remark is not possible, Starbucks is the weakest substitute no matter which reference brand we are considering. We also observe that the substitution lines can cross one another, that is, one brand could be a stronger substitute when price increase is modest, but could be a weaker substitute when price increase is large. In all the figures, there is a noticeable upward jump at the zero percent price change (corresponding to the observed market shares in the data). This is because we are reporting the *midpoint* of the identified interval for each market share.

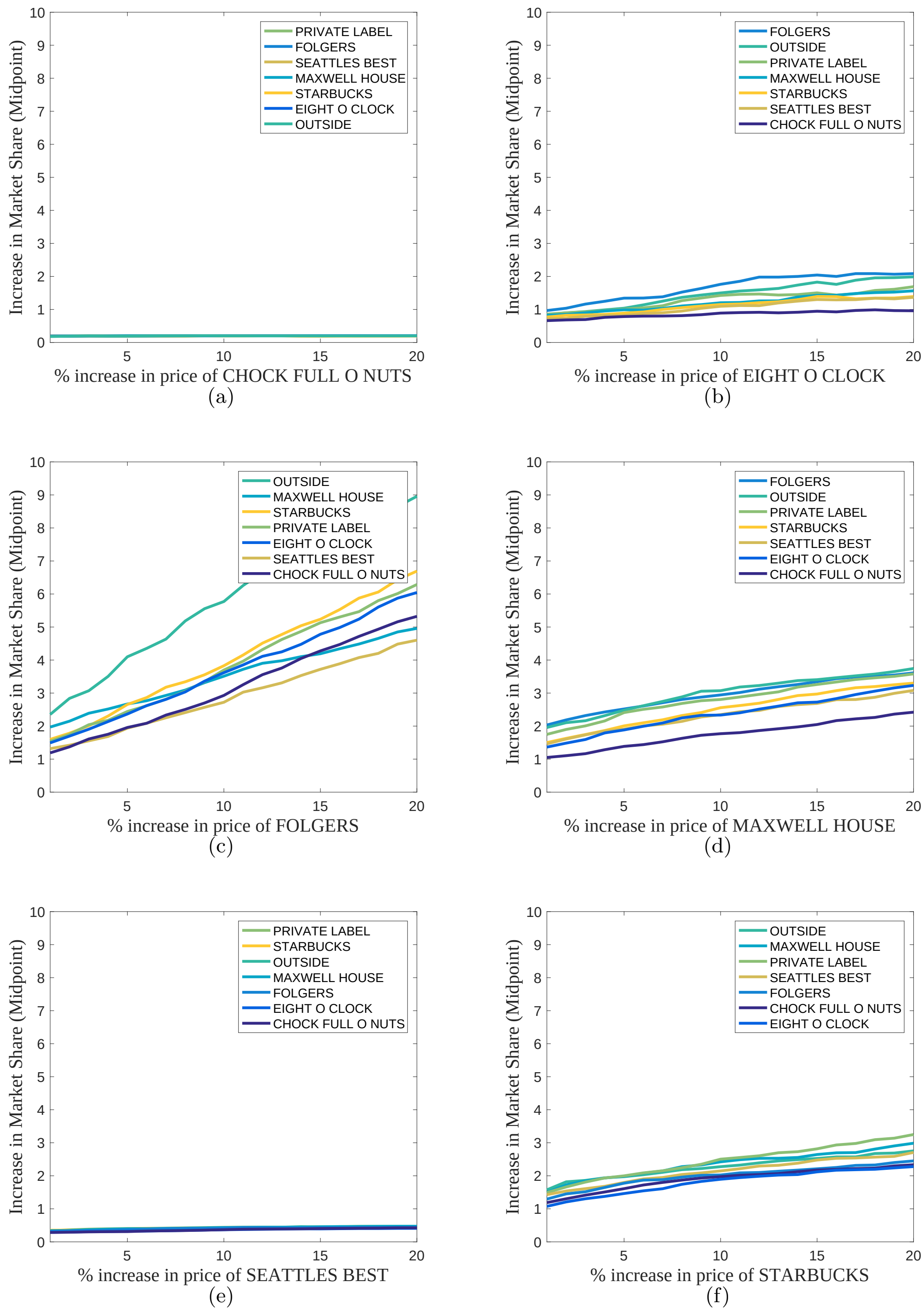

FIGURE 1. Absolute increases in market shares for different brands as a result of percentage increases in the price of a reference brand. Each line corresponds to a brand. The labeling of lines is such that the brand corresponding to the lowest line (weakest substitution) appears last in the graph legend, etc.

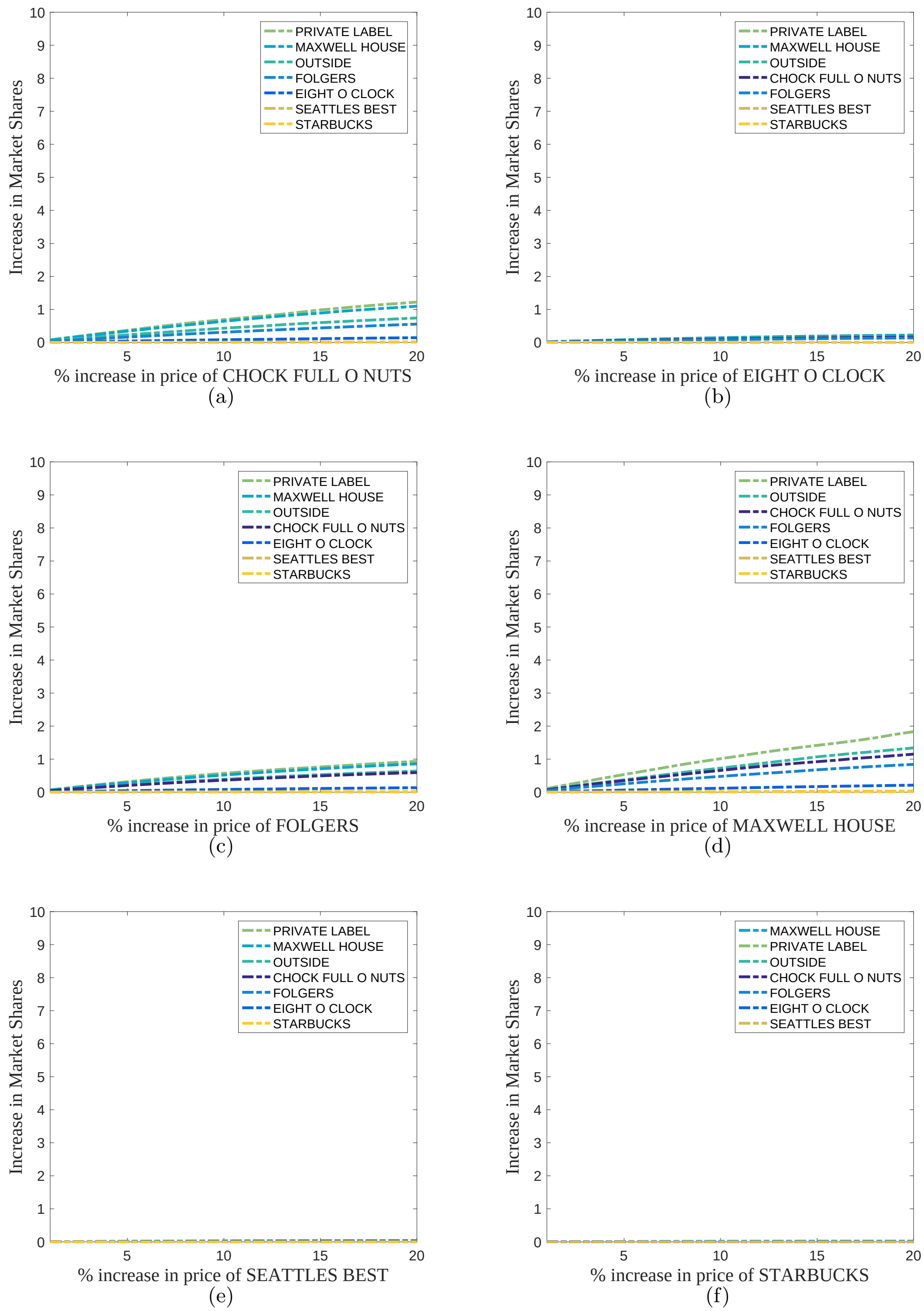

FIGURE 2. Substitution between products assuming the logit model. The labeling of lines is such that the brand corresponding to the lowest line (weakest substitution) appears last in the graph legend, etc.

6. Concluding remarks

While the literature on semiparametric discrete-choice models is quite large and mature, empirical applications utilizing these methods have been relatively sparse. This may be due in part to the absence of methods for counterfactual evaluation utilizing results from these semiparametric models. Our paper fills in this gap by providing an approach for evaluating bounds for counterfactual choice probabilities for semiparametric multinomial choice models, in which the systematic components of utility are specified as a single-index but the distribution of the error terms are left unspecified. Exploiting the property of *cyclic monotonicity* (a convex-analytic property of the random utility choice model), we derive upper and lower bounds on choice probabilities for hypothetical counterfactual scenarios which may lie outside the range of the observed data. Monte Carlo simulations and an empirical illustration using scanner data show that our method works well in practice.

References

Ackerberg, D., Benkard, L., Berry, S. and Pakes, A. Econometric tools for analyzing market outcomes. *Handbook of Econometrics*, volume 6, pp. 4171-4276.

Allen, R. and Rehbeck, J. Identification of Average Demand Models. Working paper, 2016.

Allenby, G. and Rossi, P. (2005). *Bayesian Statistics and Marketing*. Wiley.

Berry, S., Gandhi, A., and Haile, P. (2013). Connected substitutes and invertibility of demand. *Econometrica*, 81(5):2087–2111.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, pages 841–890.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *RAND Journal of Economics*, pages 242–262.

Bresnahan, T., Stern, S., and Trajtenberg, M. (1997). Market segmentation and the sources of rents from innovation: personal computers in the late 1980s. *RAND Journal of Economics*, pages S17-S44.

Bronnenberg, B. J., Kruger, M. W., and Mela, C. F. (2008). Database paper: The iri marketing data set. *Marketing Science*, 27(4):745–748.

Chiong, K. X., Galichon, A., and Shum, M. (2016). Duality in dynamic discrete-choice models. *Quantitative Economics*, 7(1):83–115.

Chiong, K. X. and Shum, M. (2016). Random projection estimation of discrete-choice models with large choice sets. *Available at SSRN 2748700*.

Fosgerau, M. and De Palma, A. (2015). Generalized entropy models.

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

Goldberg, P. (1995). Product differentiation and oligopoly in international markets: The case of the US automobile industry. *Econometrica*, pages 891–951.

Goolsbee, A. and Petrin, A. (2004). The consumer gains from direct broadcast satellites and the competition with cable TV. *Econometrica*, pages 351–381.

Han, A. (1987). Nonparametric analysis of a generalized regression model. *Journal of Econometrics*, 35: 303–316.

Hausman, J. (1996). Valuation of new goods under perfect and imperfect competition. In *The Economics of New Goods*, University of Chicago Press.

Hausman, J., Abrevaya, J., and Scott-Morton, F. (1998). Misclassificaion of the dependent variable in a discrete-response setting. *Journal of Econometrics*, 87: 239–269.

McFadden, D. L. (1978). Modeling the choice of residential location. In A. Karlqvist et. al., editors, *Spatial Interaction Theory and Residential Location*, North-Holland.

McFadden, D. L. (1981). Economic models of probabilistic choice. In C. Manski and D. McFaddenm editors, *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press.

McFadden, D. L. (1984). Econometric analysis of qualitative response models. *Handbook of Econometrics*, 2:1395–1457.

Manski, C. (1975). The maximum score estimation of the stochastic utility model. *Journal of Econometrics*, 3: 205–228.

Nevo, A. (2000). Mergers with differentiated products: The case of the ready-to-eat cereal industry. *RAND Journal of Economics*, pages 395–421.

Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy*, 110: 705–729.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press.

Rust, J. (1994). Structural estimation of Markov decision processes. *Handbook of Econometrics*, Volume 4.

Shi, X., Shum, M., and Song, W. (2016). Estimating semi-parametric panel multinomial choice models using cyclic monotonicity. *Available at SSRN 2748647*.

Train, K. (2003). *Discrete choice methods with simulation*. Cambridge University Press.

Villani, C. (2003). *Topics in optimal transportation*. American Mathematical Society.

A. Additional remarks: Rank-order property

Here we study the *rank order* condition studied by Fox (2007): $\delta_i > \delta_j \Leftrightarrow s_i > s_j$ and consider its implications for evaluating counterfactual choice probabilities. The Rank order condition can be viewed as a generalization of the IIA property, and is implied by the exchangeable, continuously-distributed error terms with full support. When applying to the counterfactual evaluation, rank order condition implies that s^{M+1} should satisfy

Assumption 2. (Rank Order): Suppose σ is the permutation on $\{1, 2, \dots, J\}$ such that $\delta_{\sigma(1)} \geq \delta_{\sigma(2)} \geq \dots \geq \delta_{\sigma(J)}$. The rank order condition implies that the same permutation also sorts the market shares in decreasing order: $s_{\sigma(1)} \geq s_{\sigma(2)} \geq \dots \geq s_{\sigma(J)}$.

By definition, rank order is a *within-market* restriction: it only places restriction directly on $(\boldsymbol{\delta}^{M+1}, \boldsymbol{s}^{M+1})$, regardless of other sample information as well as the magnitude of the price change. To investigate further, we rewrite these inequalities in matrix form. Suppose s^{M+1} is a $J \times 1$ column vector, and P is the $J \times J$ permutation matrix representing σ: $P[j, \sigma(j)] = 1, j = 1, \dots, J$ and zero otherwise.

$$D \cdot P \cdot s^{M+1} \leq \mathbf{0},$$

where D is the $(J-1) \times J$ bi-diagonal matrix with the elements $(-1, 1)$ in the $(j, j+1)$ positions in row j. From this representation, we see that the constraint matrix $D \cdot P$ only contains $(-1, 1, 0)$ and hence does not depend on other market-level data $(\boldsymbol{\delta}^m, \boldsymbol{s}^m)$. As a result, its identified set will be *invariant* to the the sample size and the magnitude of the price change. Moreover, the rank order condition implies a simple component-wise upper bound for s_j:

Proposition 3. *$s_{\sigma(j)}$ is bounded above by $\frac{1}{j}$, regardless of the number of alternative J.*

proof: It is trivial to show that the upper bound for the market share $s_{\sigma(1)}$ corresponding to the largest mean utility $\delta_{\sigma(1)}$ is 1, as $\boldsymbol{s}_\sigma = (1, 0, \dots, 0)$ does not violate the rank order condition. Similarly, $\max s_{\sigma(j)}$ is at most $\frac{1}{j}$, otherwise $s_{\sigma(j)} > s_{\sigma(1)}$, violating the rank order condition. ■