RANDOM QUASI-LINEAR UTILITY

ERYA YANG[1] AND IGOR KOPYLOV[2]

Abstract. We propose a *random quasi-linear utility model* (RQUM) where quasi-linear utility functions are drawn randomly via some probability distribution π, and utility ties are broken by a convenient lexicographic rule. We characterize RQUM and identify π uniquely in terms of stochastic choice data. McFadden's (1973) additive random utility model is obtained as a special case where utility ties have a zero probability in all menus. Another distinct case of RQUM captures finite populations and derives π with a finite support. Our main axioms are testable. They prohibit context and reference dependence, and also modify the non-negativity of Block-Marschack polynomials for monetary cost variations. We also characterize RQUM through a stronger version of McFadden and Richter's (1990) axiom of revealed stochastic preferences (ARSP). This approach extends to incomplete datasets.

1. Introduction

Empirical observations of consumers' aggregate choices are stochastic in transportation (McFadden [21]), recreational fishing (Train [30]), selection of appliance efficiency levels (Revelt and Train [25]), and many other settings. A single agent's choices can also be random due to intertemporal planning (Rust [26]) or spontaneous variations in their tastes (e.g., Agranov and Ortoleva [2]).

Random utility models (RUM) represent stochastic choices by maximization of utility functions that are randomly drawn via some probability distribution π. Such π is interpreted in terms of heterogeneous preferences. More formally, π is defined over some set Θ of complete and transitive preferences on some consumption space X. Then any alternative x in any finite menu $A \subset X$ should be chosen with probability

$$\rho(x, A) = \pi(R \in \Theta : \ x \text{ maximizes } R \text{ in } A). \tag{1}$$

Date: March 28, 2023.

JEL Classification Code. D01, D90.

Key words and phrases. Random utility, random choice, quasi-linear utility, tie-breaking, finite datasets, ARSP, Block-Marschak polynomials.

1: Business School, Sun Yat-sen University, 518107 Shenzhen, China. yangery@mail.sysu.edu.cn. Corresponding author. 2: Department of Economics, University of California, Irvine. ikopylov@uci.edu.
We thank Mira Frick, Ryota Iijima, Asen Kochov, Stergios Skaperdas, Michael McBride, Jay Lu, Andrew Ellis, Christopher Turansick, David Brownstone, Miguel Zerecero, and Jiawei Chen for their helpful suggestions on this paper. We are grateful to the editor Faruk Gul and two anonymous referees; their insightful comments helped significantly in improving this paper. We would also like to thank participants at seminars at UCI, Rochester, Yale, Sun Yat-sen, and SWUFE Universities.

In the *classic* RUM of Block and Marschak [6] (henceforth BM), the domain X is finite, and Θ is the set of all *total orders* (i.e., complete, transitive, antisymmetric preferences over X). Falmagne [9] characterizes the classic RUM via non-negativity of BM polynomials. McFadden and Richter [22] provide another characterization via the axiom of revealed stochastic preference (ARSP).

In many applications, it is convenient to associate the set Θ with a particular class of utility functions on X. Most importantly, McFadden's [19] additive RUM adopts representation (1) where the domain

$$X = \{(i, \alpha) :\ i \in \{0, 1, ..., n\} \text{ and } \alpha \in \mathbb{R}\}$$

consists of pairs of consumption goods i and monetary costs α, and the set Θ consists of all *quasi-linear* preferences. By definition, such preferences can be represented by quasi-linear utility functions that are standard in discrete choice theory and estimation methods (e.g. Nocke and Schutz [23]). Quasi-linearity is also very common in mechanism design, auction theory, bargaining theory, public economics, etc. The quasi-linear structure implies that Θ can be parametrized by the Euclidean space $\mathbb{R}^n$ and hence, π can be modelled as a *Borel probability measure* over $\mathbb{R}^n$.

To make the additive RUM well-defined, it is necessary that "the probability of ties is zero" (McFadden [20, p. S15]). Therefore, π cannot have atoms[1] and hence, cannot have a finite support either. Thus, finite populations are inconsistent with the additive RUM, which can be problematic for welfare analysis and other applications.

Williams [32] and Daly and Zachery [8] (henceforth, WDZ) characterize the additive RUM in terms of derivatives of choice probability functions. In particular, the WDZ Theorem assumes a symmetry condition

$$\frac{\partial \rho_k(c)}{\partial c_j} = \frac{\partial \rho_j(c)}{\partial c_k} \tag{2}$$

where ρ_k and ρ_j denote the probabilities of choosing goods k and j, respectively when $c = (c_0, c_1, \ldots, c_n)$ is the cost vector. Such differential equations cannot be refuted by empirical data because partial derivatives like $\frac{\partial \rho_k(c)}{\partial c_j}$ are unobservable. Moreover, the WDZ theorem does not accommodate some familiar continuous distributions (e.g., uniform or exponential) for which partial derivatives $\frac{\partial \rho_k(c)}{\partial c_j}$ do not exist at some cost vectors c.

Our random quasi-linear utility model (RQUM) extends the additive RUM characterization and achieves several objectives.

(i) RQUM accommodates any Borel probability measure π over $\mathbb{R}^n$. Utility ties are broken by a convenient lexicographic tie-breaking rule.

[1]Suppose that $\pi(R) > 0$ for some preference R with a quasi-linear utility representation. Then R should be indifferent between some distinct alternatives $x, y \in X$. By (1), $\rho(x, \{x, y\}) + \rho(y, \{x, y\}) \geq 1 + \pi(R) > 1$ because $\pi(R)$ is counted both in $\rho(x, \{x, y\})$ and in $\rho(y, \{x, y\})$.

(ii) RQUM is characterized via novel axioms that do not use differentiation, but are written instead in terms of discrete cost variations. All of our axioms are testable, except for continuity conditions.
(iii) The endogenous probability distribution π has a unique and explicit identification in terms of the observable stochastic choice rule ρ.

To formulate RQUM, associate each vector $v \in \mathbb{R}^n$ with the quasi-linear preference R_v that is represented over X by the function $q_v(i, \alpha) = v_i - \alpha$. Here $v_1, \ldots, v_n$ reflect *reservation values* for consumption goods $i = 1, \ldots, n$ respectively, and $v_0 = 0$ by convention.

Our main representation for stochastic choices is

$$\rho(x, A) = \pi(\{v \in \mathbb{R}^n : x \text{ has the lowest grade among maxima of } R_v \text{ in } A\}) \tag{3}$$

where π is a Borel probability measure on $\mathbb{R}^n$, and the *grade* of any pair (i, α) is defined as i. Thus all utility ties are assumed to be broken in favor of alternatives with lower grades.[2]

Our main result (Theorem 1) characterizes RQUM via axioms that do not assume or imply differentiability for the functions $\rho_k(c)$. The first two axioms (No Complementarity and Cross-Price Neutrality) prohibit behaviors like context dependence and reference dependence. The third axiom (Joint Monotonicity) is more complicated, but still testable. It asserts roughly that the revealed probability of any bounded rectangle in the type space $\mathbb{R}^n$ should be non-negative. This condition converges to the non-negativity of some BM polynomial when the rectangle expands to an unbounded orthant. We close the model with continuity assumptions and use them to deliver refinements where (i) "the probability of ties is zero" as in McFadden's additive RUM, or (ii) the support of π is finite, which accommodates finite populations.

Next, we illustrate how π can be uniquely and explicitly derived from the observed stochastic choice rule ρ. The key observation is that the *cumulative distribution function* of π for all $v \in \mathbb{R}^n$ must satisfy

$$F_\pi(v) = \rho((0,0), A) \tag{4}$$

where the menu

$$A = \{(0,0), (1, v_1), (2, v_2), \ldots, (n, v_n)\}$$

provides all goods $i = 0, 1, \ldots, n$ at costs $0, v_1, \ldots, v_n$, respectively. Here it follows from (3) that for any vector $w \in \mathbb{R}^n$, the comparisons $v_i \geq w_i$ should hold for all $i = 1, \ldots, n$ if and only if the preference R_w is maximized by the alternative $(0, 0)$ in the menu A. Formula (4) implies the uniqueness of π, which is not guaranteed by the classic RUM. Turansick [31] shows that such uniqueness can be only obtained under stringent single-crossing conditions on the support of π. Apesteigua, Ballester, and Lu [3] use a strong version of single-crossing to derive π uniquely in terms of choices in binary menus. Identification (4) is substantially

[2]We establish later that the observable implications of our model are unchanged if ties are broken by any other permutation of consumption goods. By Occam's razor, we adopt the tie-breaking rule with the simplest notation.

simpler than the counterpart in the classic RUM, where the construction of π employs a multi-step procedure based on BM polynomials. The identification (4) is also the cornerstone of our proofs, but the full argument invokes results from probability theory (e.g., Billingsley [5, Theorem 12.5]) rather than differentiability techniques. Corollary 2 applies our Theorem 1 to the reduced domain that appears in the WDZ theorem.

Our next result (Theorem 3) characterizes RQUM via McFadden and Richter's [22] linear programming approach. This approach extends to finite datasets in Theorem 4 where the identification of π is based on the Farkas Lemma rather than the formula (4). We argue that there are no observable distinctions between grading procedures if all ties are broken by any permutation of the set $\{0, 1, \ldots, n\}$.

Finally, we derive the WDZ Theorem as a corollary and discuss several other examples that satisfy or reject RQUM.

1.1. **Related Literature.** Our work contributes to the growing list of refinements of RUM in the settings of preference over menus and choices in menus. In the *random expected utility model* (REUM) of Gul and Pesendorfer [12], the domain X consists of lotteries over deterministic prizes, and Θ is the class of preferences that have expected utility representations. In this case, the distribution π is determined uniquely by ρ, but the identification of π relies on compactness arguments from real analysis. Gul and Pesendorfer [12] consider only the regular case where utility ties have a probability of zero. In the supplement of their paper [11], Gul and Pesendorfer discuss non-regular random choice rules and deal with ties by defining a tie-breaking rule. Piermont [24] combines REUM with various tie-breaking rules. However, his extensions do not identify π in terms of ρ, but instead impose consistency conditions on a pair (π, ρ). Besides REUM, refinements of RUM in the menu domain include the single-crossing RUM in Apesteigua, Ballester, and Lu [3], applications of RUM to random attention in Manzini and Mariotti [18], to choices over state-contingent acts in Lu [14, 15], to dynamic choices in Frick, Iijima, and Strzalecki [10], and various other settings.

Our work also contributes to the literature on the integrability problem in market demand analysis. The integrability problem (see discussions in Hurzicz and Uzawa [13]) assumes that a given demand system is generated by utility maximization and aims to recover such a utility.[3] In empirical demand analysis, additive RUM can be written as an average utility plus a random error. The error is assumed to be known, and the observed choice is generated by a social surplus function.[4]

A result closely related to the WDZ theorem and widely used in empirical estimation in the integrability problem is the differential WDZ lemma (e.g. McFadden

[3]Nocke and Schutz [23] provide a recent discussion for the integrability problem with a quasilinear utility for a demand system.

[4]The special case where the error term follows the extreme type I distribution is equivalent to the Luce model [16], which is the foundation of the discrete choice literature. This equivalence result is attributed to Holman and Marley in Luce [17], and it is also shown in Yellot [33] and McFadden [19].

[22]). This result characterizes stochastic choice rules to be the gradient of the social surplus function. The fact of being the gradient of some function imposes some constraints on the stochastic choice rule, and the WDZ theorem uses these constraints to characterize the additive RUM. Moreover, the differential WDZ lemma is the basis for the first-stage estimation in the Berry-Levison-Pakes (BLP) estimator [4]. Shi, Shum, and Song [28, Lemma 2.1] provide another estimation methods for the mixed logit model where the differential WDZ lemma is combined with cyclic monotonicity conditions to construct a novel estimator for panel data.

Sorensen and Fosgerau [29] extend the WDZ lemma to accommodate any random utility distribution and discuss the identification of the heterogeneous utilities without assuming differentiability. However, their conditions are written in terms of subgradients that have the same observability issues as derivatives. By contrast, our conditions and proofs do not use any kind of marginal analysis, which is replaced by integral methods of the measure theory.

2. Preliminaries

Let $N = \{0, 1, \dots, n\}$ be a finite set of consumption goods. Assume that $n \geq 1$ so that N has at least two elements. The subset of goods with positive indices is written as $[1, n] = \{1, \dots, n\}$.

Let $X = \{x, y, \dots\}$ be the set $N \times \mathbb{R}$ of all pairs (i, α) that combine some consumption $i \in N$ with a monetary cost $\alpha \in \mathbb{R}$. If a good i is paired with a positive reward $\beta > 0$, then its cost $\alpha = -\beta$ is negative.

Let $\mathcal{A} = \{A, B, \dots\}$ be the set of all *menus*—finite non-empty subsets $A \subset X$. Singleton menus $\{x\}$ are written without curly brackets hereafter.

Let Ω be the set of all pairs (x, A) such that $A \in \mathcal{A}$ and $x \in A$, that is, x is a feasible element in a menu A. Such pairs are called *trials*.

Let $\mathcal{R} = \{R, \dots\}$ be the set of all *orders*—complete and transitive relations on X. An order $R \in \mathcal{R}$ is called *total* if for all $x, y \in X$, $xRyRx$ implies $x = y$.

For any order $R \in \mathcal{R}$ and trial $(x, A) \in \Omega$, say that x:

- *maximizes* R in A if xRy for all $y \in A$,
- *strictly maximizes* R in A if yRx does not hold for any $y \in A \setminus x$.

Say that x is also a *maximum* or a *strict maximum* for R in A, respectively.

2.1. Quasi-Linearity.

A function $q : X \to \mathbb{R}$ is called *quasi-linear* if

$$q(i, \alpha) = q(i, 0) - \alpha \qquad \text{for all } (i, \alpha) \in X.$$

Let $\mathcal{Q} \subset \mathcal{R}$ be the set of all orders that have quasi-linear utility representations. Such orders are called *quasi-linear* as well.

The set $\mathcal{Q}$ has a convenient parametrization by the Euclidean space $\mathbb{R}^n$. For any vector $v = (v_1, \dots, v_n) \in \mathbb{R}^n$, let R_v be represented by the quasi-linear function

$$q_v(i, \alpha) = v_i - \alpha \quad \text{for all } (i, \alpha) \in X,$$

where $v_0 = 0$ by convention. Here, the vector v specifies *reservations values* for goods $i \in [1, n]$ when v_0 is normalized to zero. It is easy to check that

$$R \in \mathcal{Q} \quad \Leftrightarrow \quad R = R_v \quad \text{for some } v \in \mathbb{R}^n,$$

and such $v \in \mathbb{R}^n$ is determined uniquely by $R \in \mathcal{Q}$. Indeed, if $R_v = R_w$, then $v = w$. For example, if $v_i > w_i$, then $(i, v_i) R_v (0, 0)$, but $(i, v_i) R_w (0, 0)$ does not hold. Moreover, if R is represented by some quasi-linear utility function q, then $R = R_v$ where $v_i = q(i, 0) - q(0, 0)$ for all $i \in [1, n]$. Thus $v \leftrightarrow R_v$ is a bijective mapping between the Euclidean space $\mathbb{R}^n$ and the space $\mathcal{Q}$.

2.2. **Jointly Monotone Functions.** For any function $F : \mathbb{R}^n \to \mathbb{R}$ and vectors $w, v \in \mathbb{R}^n$, define the *joint variation*

$$\Delta(F, w, v) = \sum_{K \subset [1,n]} (-1)^{|K|} F(wKv)$$

where for any $K \subset [1, n]$, $wKv \in \mathbb{R}^n$ denotes the *composite vector* such that

$$(wKv)_i = \begin{cases} w_i & \text{if } i \in K \\ v_i & \text{if } i \in [1, n] \setminus K. \end{cases}$$

For any vectors $w, v \in \mathbb{R}^n$, write $v \geqq w$ if $v_i \geq w_i$ for all $i \in [1, n]$.

Say that $F : \mathbb{R}^n \to \mathbb{R}$ is *jointly monotone* if for all vectors $v, w \in \mathbb{R}^n$,

$$v \geqq w \quad \Rightarrow \quad \Delta(F, w, v) \geq 0.$$

For example, if $n = 2$, then the joint monotonicity requires that for all $w, v \in \mathbb{R}^2$,

$$\begin{aligned} \Delta(F, w, v) &= F(w\emptyset v) - F(w\{1\}v) - F(w\{2\}v) + F(w\{1, 2\}v) = \\ &= F(v_1, v_2) - F(w_1, v_2) - F(v_1, w_2) + F(w_1, w_2) \geq 0. \end{aligned}$$

2.3. **Cumulative Distribution Functions.** Let $\Pi = \{\pi, \dots\}$ be the set of all *Borel probability measures* on $\mathbb{R}^n$. By definition, Π consists of all countably additive probability measures on the minimal σ algebra that contains all opens sets in $\mathbb{R}^n$.

For any probability measure $\pi \in \Pi$, its *cumulative distribution function* (cdf) $F_\pi : \mathbb{R}^n \to [0, 1]$ is defined for all $v \in \mathbb{R}^n$ as

$$F_\pi(v) = \pi(\{r \in \mathbb{R}^n : v \geqq r\}).$$

For any vectors $w, v \in \mathbb{R}^n$, write $v \gg w$ if $v_i > w_i$ for all $i \in [1, n]$, and define the *rectangle*

$$(w, v] = \{r \in \mathbb{R}^n : v \geqq r \gg w\}.$$

It is well-known (e.g., Billingsley [5, Section 12]) that the probability measure π is uniquely determined by its cdf F_π. In particular, for any $w, v \in \mathbb{R}^n$ such that $v \geqq w$, the rectangle $(w, v]$ has probability

$$\pi((w, v]) = \Delta(F_\pi, w, v), \tag{5}$$

and the singleton $\{v\}$ has probability

$$\pi(\{v\}) = \lim_{m \to \infty} \Delta(F_\pi, w^m, v),$$

where $w^m = \big(v_1 - \frac{1}{m}, v_2 - \frac{1}{m}, \ldots, v_n - \frac{1}{m}\big)$.

Identity (5) implies that F_π must be jointly monotone as well.

3. Functional Form and Identification

A function $\rho : \Omega \to [0,1]$ is called a *stochastic choice rule* (scr) if

$$\sum_{x \in A} \rho(x, A) = 1 \qquad \text{for all } A \in \mathcal{A}. \tag{6}$$

Here, the probability $\rho(x, A)$ of any trial $(x, A) \in \Omega$ is interpreted as the likelihood of x being chosen when the menu A is feasible. Next, we adapt the RUM to represent ρ by heterogenous quasi-linear orders, or equivalently, quasi-linear utility functions.

A Borel probability measure $\pi \in \Pi$ is called a *regular* representation for a stochastic choice rule ρ if for all trials $(x, A) \in \Omega$,

$$\begin{aligned} &\rho(x, A) = \pi(M(x, A)) \qquad \text{where} \\ &M(x, A) = \{v \in \mathbb{R}^n : x \text{ maximizes } R_v \text{ in } A\ \}. \end{aligned} \tag{7}$$

In other words, the observed likelihood of any trial (x, A) should equal the probability that the measure π assigns to all types $v \in \mathbb{R}^n$ for which the order R_v—or equivalently, the function q_v—is maximized by x in the menu A. Therefore, representation (7) refines the general form (1) for $\Theta = \mathcal{Q}$.

For any $\pi \in \Pi$, representation (7) is consistent with the definition of a stochastic choice rule if and only if[5] for all $(x, A) \in \Omega$,

$$\begin{aligned} &\pi(M(x, A)) = \pi(S(x, A)) \qquad \text{where} \\ &S(x, A) = \{v \in \mathbb{R}^n : x \text{ strictly maximizes } R_v \text{ in } A\ \}. \end{aligned} \tag{8}$$

This condition requires that π should assign a zero probability to quasi-linear utility ties. In particular, if π has a finite support over $\mathbb{R}^n$, then (8) must be violated because for any $v \in \mathbb{R}^n$ such that $\pi(v) > 0$, there are trials $(x, A) \in \Omega$ such that x maximizes R_v in A, but not strictly so.

To combine the RUM with any Borel probability measure $\pi \in \Pi$, consider a convenient tie-breaking rule.

Define the *grade* of any alternative $(i, \alpha) \in X$ as i. Say that x is a *low maximum* for an order $R \in \mathcal{R}$ in a menu A if x maximizes R in A and has the lowest grade among all maxima of R in A.

---

[5]For any $(x, A) \in \Omega$, the set $M(x, A)$ contains $S(x, A)$. Thus (7) and (6) imply

$$\rho(x, A) = 1 - \sum_{y \in A \setminus x} \rho(y, A) = 1 - \sum_{y \in A \setminus x} \pi(M(y, A)) \leq \pi(S(x, A)) \leq \pi(M(x, A)) = \rho(x, A)$$

and hence, (8). On the other hand, (7) and (8) imply that for any $A \in \mathcal{A}$,

$$1 = \pi(\mathbb{R}^n) \leq \sum_{x \in A} \pi(M(x, A)) = \sum_{x \in A} \rho(x, A) = \sum_{x \in A} \pi(S(x, A)) \leq 1$$

because the sets $S(x, A)$ are disjoint, and the sets $M(x, A)$ cover $\mathbb{R}^n$.

Say that π is a *low* representation for an scr ρ if for all trials $(x, A) \in \Omega$,

$$\rho(x,A) = \pi(L(x,A)) \qquad \text{where} \tag{9}$$
$$L(x,A) = \{v \in \mathbb{R}^n : x \text{ is a low maximum for } R_v \text{ in } A\}.$$

Representation (9) is well-defined for any $\pi \in \Pi$ and $A \in \mathcal{A}$ because for any vector $v \in \mathbb{R}^n$, the quasi-linear order R_v has a unique low maximum in A and hence,

$$\mathbb{R}^n = \bigcup_{x \in A} L(x,A)$$

is a partition of the Euclidean space $\mathbb{R}^n$. Obviously, the low representation (9) implies the regular one (7) when π satisfies (8). In this case, for all $(x, A) \in \Omega$,

$$\pi(L(x,A)) \le \pi(M(x,A)) = \pi(S(x,A)) \le \pi(L(x,A))$$

because $S(x, A) \subset L(x, A) \subset M(x, A)$.

Representation (9) is called the *random quasi-linear utility model* (RQUM). The tie-breaking rule in (9) is called the *grading procedure*.

3.1. **Reduced Form.** Let $\mathbb{R}^N$ be the set of all functions $c : N \to \mathbb{R}$. Such functions $c = (c_0, c_1, \ldots, c_n) \in \mathbb{R}^N$ are called *cost vectors*. Obviously, $\mathbb{R}^N$ is isomorphic to the Euclidean space $\mathbb{R}^{n+1}$.

For any $c \in \mathbb{R}^N$, define its *assortment*

$$A(c) = \bigcup_{k \in N} (k, c_k)$$

as a menu that provides all goods in N at the costs $c_0, c_1, \ldots, c_n$ respectively.

For any scr ρ, define its *reduction* as the function $\rho^* : \mathbb{R}^N \to \mathbb{R}^N$ such that for any good $k \in N$ and cost vector $c \in \mathbb{R}^N$,

$$\rho^*_k(c) = \rho\left((k, c_k), A(c)\right)$$

is the probability of choosing alternative (k, c_k) in the assortment $A(c)$. Therefore, the reduction ρ^* restricts the stochastic choice rule ρ to assortments.

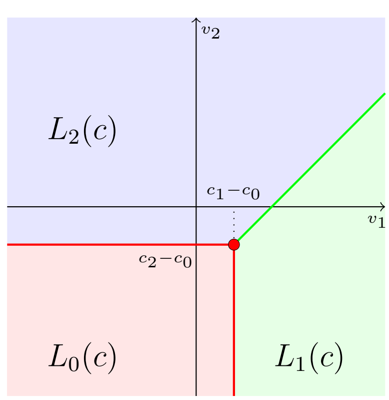

FIGURE 1. Partition $\mathbb{R}^2 = L_0(c) \cup L_1(c) \cup L_2(c)$ at $c = (c_0, c_1, c_2)$.

RQUM implies that for any $c \in \mathbb{R}^N$,

$$\rho^*_k(c) = \pi(L_k(c)) \qquad \text{where} \\ L_k(c) = L\left((k, c_k), A(c)\right). \tag{10}$$

This representation can be applied if ρ^* is given as a primitive without ρ. The partition of the type space $\mathbb{R}^n$ into the sets $L_i(c)$ is illustrated by Figure 1 for the case $n = 2$.

Say that $\rho^* : \mathbb{R}^n \to [0, 1]$ is a *reduced stochastic choice function* (reduced scr) if for all $c \in \mathbb{R}^N$,

$$\sum_{k \in N} \rho^*_k(c) = 1.$$

We call representation (10) for ρ^* the *reduced* RQUM.

3.2. **Identification and Uniqueness.** The Borel probability measure π in RQUM can be identified in terms of the observed scr ρ via a transparent formula.

Fix any $\alpha \in \mathbb{R}$. For any vector $v \in \mathbb{R}^n$, let

$$G_\alpha(v) = \rho^*_0(\alpha, v + \alpha)$$

where $v + \alpha$ denotes the vector in $\mathbb{R}^n$ such that $(v + \alpha)_i = v_i + \alpha$ for all $i \in [1, n]$. If π is a low representation for ρ, then

$$F_\pi = G_\alpha. \tag{11}$$

Indeed, for all $v \in \mathbb{R}^n$,

$$\{w \in \mathbb{R}^n : v \geqq w\} = L_0(\alpha, v + \alpha)$$

because for any $w \in \mathbb{R}^n$, the dominance $v \geqq w$ holds if and only if the alternative $(0, \alpha)$ is a low maximum for R_w in the assortment

$$A(\alpha, v + \alpha) = \{(0, \alpha), (1, v_1 + \alpha), (2, v_2 + \alpha), \ldots, (n, v_n + \alpha)\}.$$

Thus, the definition of cdfs and the reduced representation (10) imply that

$$F_\pi(v) = \rho^*_0(\alpha, v + \alpha) = G_\alpha(v).$$

The identity $F_\pi = G_\alpha$ implies that the probability measure π is *uniquely* determined by ρ, or even by the component ρ^*_0 of the reduction ρ^*. Moreover, each function G_α must be jointly monotone, which becomes one of our axioms below. The special role of the zeroth component ρ^*_0 is an artifact of the grading procedure where any tie between $(0, \alpha)$ and $(k, v_k + \alpha)$ is broken in favor of the former alternative in the assortment $A(\alpha, v + \alpha)$.

4. Main Representation Results

RQUM has several implications for stochastic choice rules ρ. To wit, let $\pi \in \Pi$ be a low representation for ρ.

Say that $x \in X$ is *discounted* by $y \in X$ if $x = (i, \alpha)$ and $y = (i, \beta)$ for some $i \in N$ and $\beta < \alpha$. Such y provides the same good i as x at a discounted cost.

Axiom 1 (No Complementarity (NC))**.** *For all* $(x, A) \in \Omega$ *and* $y \in X$,

$$\rho(x, A \cup y) \leq \rho(x, A), \tag{12}$$

and if x *is discounted by* y, *then* $\rho(x, A \cup y) = 0$.

Inequality (12) is inherited from the classic RUM. It asserts that adding any extra option y to a menu A should not increase the probability of choosing any feasible $x \in A$. Thus, NC excludes complementarities across distinct consumption goods. Moreover, it excludes context effects where the presence of y can make x more likely to be chosen due to increased attention or reason-based heuristics (e.g., Shafir, Simonson, and Tversky [27]). The second part of NC requires that x should never be chosen in the presence of a discounted alternative y. It is assumed here that all choices should reveal a perfect perception with respect to monetary costs. Note that NC remains plausible in any random utility model where all types should strictly prefer more money to less money.

Other axioms for RQUM are formulated in terms of the reduction ρ^* and rely on quasi-linearity in a more substantial way. The reduction ρ^* makes it convenient to analyze the effects of changing monetary costs on stochastic choices.

For any $k \in N$, let $\vec{k} \in \mathbb{R}^N$ be a cost vector such that $\vec{k}_k = 1$ and $\vec{k}_i = 0$ for all $i \in N \setminus k$. The difference $\rho^* \left(c + \gamma\vec{k}\right) - \rho^*(c)$ describes how stochastic choices are affected when the cost of good k varies by γ.

Axiom 2 (Cross-Price Neutrality (CPN))**.** *For any* $\gamma > 0$, *cost vector* $c \in \mathbb{R}^N$, *and distinct goods* $k, j \in N$,

$$\rho_k^*(c) - \rho_k^* \left(c - \gamma\vec{j}\right) \geq \rho_j^* \left(c + \gamma\vec{k}\right) - \rho_j^*(c).$$

By CPN, the effect of *decreasing* the cost of good j by some $\gamma > 0$ on the demand for good k should be greater than or equal to the effect of *increasing* the cost of k by the same γ on the demand for j. Informally, CPN assumes that the perception of money is linear and has no reference points. For example, increasing c_k from 0 to γ should not be viewed as more prominent than decreasing c_j from 0 to $-\gamma$.

In particular, if $\gamma \to +\infty$, then CPN converges the inequality[6]

$$\rho(x, A) \geq \rho(y, A \setminus x) - \rho(y, A)$$

where $x = (k, c_k)$, $y = (j, c_j)$, and $A = A(c)$. This inequality is a special case of NC and typically holds strictly because

$$\rho(x, A) > \rho(y, A \setminus x) - \rho(y, A)$$

whenever $\rho(z, A \setminus x) - \rho(z, A) > 0$ for some $z \in A \setminus \{x, y\}$.

Axiom 3 (Joint Monotonicity (JM))**.** *For any* $\alpha \in \mathbb{R}$, G_α *is jointly monotone.*

[6]Here the limits

$$\lim_{\gamma \to \infty} \rho_k^* \left(c - \gamma\vec{j}\right) = 0 \quad \text{and} \quad \lim_{\gamma \to \infty} \rho_j^* \left(c + \gamma\vec{k}\right) = \rho(y, A(c) \setminus x)$$

are implied by RQUM.

This condition follows from the identity (11). The asymptotic meaning of JM for large cost variations can be related to the non-negativity of BM polynomials. Indeed, as $\gamma \to +\infty$, then the condition $\Delta(G_0, w, w+\gamma) \geq 0$ converges to

$$\sum_{K \subset [1,n]} (-1)^{|K|} \rho(x, A_K) \geq 0 \tag{13}$$

where $x = (0,0)$ and $A_K = x \cup \bigcup_{i \in K} (i, w_i)$. Condition (13) requires the non-negativity of a particular BM polynomial. In the classic RUM, this polynomial computes the probability that x is the worst choice in the menu $A(0,w)$. This asymptotic observation also explains the mathematical similarity between the weights in JM and BM non-negativity condition. Both systems of weights arrive via the Möbius inversion on the Boolean lattice of all subsets of the set $[1,n]$. Of course, the weights are applied to distinct observations in the two models and hence, the overlap is achieved only when cost variations are large that essentially makes some goods not feasible. For $\gamma < \infty$, JM requires the Mobius inversion at vertex v to be nonnegative for any rectangle $(w,v] \subseteq R^n$. JM is written only for the special good 0, and CPN passes the restrictions imposed by JM to the other goods.

Next, consider two continuity conditions.

Axiom 4 (Archimedean Continuity (AC))**.** *For any $\varepsilon > 0$, there is $\delta > 0$ such that for all $c \in \mathbb{R}^N$ and $k, j \in N$,*

$$c_k - c_j > \delta \quad \Rightarrow \quad \rho^*_k(c) < \varepsilon.$$

This axiom asserts that any possible type should reject a good k if it is feasible to get some other good j with a sufficiently high discount. In particular, AC excludes lexicographic types who would choose a good k over other alternatives regardless of monetary costs.

Say that ρ^* is *continuous in a direction* $d \in \mathbb{R}^N$ if for all $c \in \mathbb{R}^N$,

$$\lim_{\gamma \to 0,\ \gamma \geq 0} \rho^*(c + \gamma d) = \rho^*(c),$$

where the parameter γ is constrained to be nonnegative.

Axiom 5 (Grading Continuity (GC))**.** *ρ^* is continuous in the direction* $(0, 1, \ldots, n)$.

Here the special direction $(0, 1, \ldots, n)$ reflects the grading procedure.[7] The meaning of Axioms 1–5 and their logical independence are clarified further by several examples in the discussion section.

A low representation $\pi \in \Pi$ is called *finite-ranged* if π has a finite range.

Theorem 1. *A stochastic choice rule ρ satisfies Axioms 1–5 if and only if ρ has a low representation $\pi \in \Pi$. This representation is*

(i) uniquely identified by the reduction ρ^* via (11),

(ii) regular if and only if ρ^* is continuous,

[7]GC can be naturally adapted if the grading procedure minimizes some permutation $p : N \to N$ to break ties. Then ρ^* should be continuous in the direction $(p(0), p(1), \ldots, p(n))$.

(iii) finite-ranged if and only if ρ^* has a finite range.

This result characterizes RQUM. The two special cases where the representation is either regular or finite-ranged require that ρ^* is continuous or finite-ranged, respectively. These conditions are mutually exclusive. Note also that the continuity of ρ^* implies GC, and hence, it can replace GC in the list of axioms for the regular representation.

To prove that Axioms 1–5 are sufficient for the low representation (9), we proceed in three broad steps. First, we use JM, AC, and GC to construct a Borel probability measure $\pi \in \Pi$ such that $F_\pi = G_0$. The existence of such π follows from Billingsley [5, Theorem 12.5].

Second, we use CPN to show that for all cost vectors $c \in \mathbb{R}^N$ and goods $k \in N$,

$$\rho_k^*(c) = \pi(L_k(c)).$$

Third, we use NC to establish that the low representation (9) holds for all menus $A \in \mathcal{A}$ rather than just for assortments $A(c)$ with cost vectors $c \in \mathbb{R}^N$. All details are in the appendix.

In the above outline, NC is invoked only at the last step to extend a low representation from the reduction ρ^* to the entire scr ρ. Thus, Theorem 1 can be rewritten in a reduced form as follows.

Corollary 2. *A reduced scr ρ^* satisfies Axioms 2–5 if and only if ρ^* is represented by* (10) *for some $\pi \in \Pi$. Moreover, there is a unique scr ρ that satisfies NC and has ρ^* as its reduction.*

Here, the identification (11) still applies, and the probability measure π can be used as a low representation for the unique extension ρ. Similarly, the regular and finite-ranged cases can be characterized in terms of ρ^* as well. In fact, the domain of assortments is exactly the domain considered in the WDZ theorem. In fact, their result can be obtained from ours as outlined in Section 5.

4.1. **RQUM via Linear Programming.** Similar to the classic RUM, RQUM can be derived via the linear programming techniques.

Axiom 6 (Axiom of Revealed Stochastic Quasi-linearity (ARSQ))**.** *For any finite sequence of trials $\{(x_k, A_k) \in \Omega\}_{k=1}^m$,*

$$\sum_{k=1}^{m} \rho(x_k, A_k) \le \max_{v \,\in\, \mathbb{R}^n} |\{k \in \{1, \ldots, m\} : x_k \textit{ strictly maximizes } R_v \textit{ in } A_k\}|. \tag{14}$$

Note that the sequence $\{(x_k, A_k) \in \Omega\}_{k=1}^m$ may include multiple copies of the same trial $(x, A) \in \Omega$. ARSQ follows from RQUM and implies all testable properties in Theorem 1.

Theorem 3. *For any stochastic choice rule ρ,*

$$\textit{Axioms 1–5} \quad \Rightarrow \quad \textit{ARSQ} \quad \Rightarrow \quad \textit{Axioms 1–3.}$$

An immediate corollary of Theorems 1 and 3 is that

$$\text{RQUM} \quad \Leftrightarrow \quad \text{ARSQ, AC, GC.}$$

Note that ARSQ alone is not sufficient for RQUM because it does not imply the continuity conditions AC and GC (see examples in Section 4).

Theorem 3 clarifies the connection of RQUM to the classic RUM. ARSQ strengthens McFadden and Richter's ARSP [22]:

Axiom 7 (Axiom of Revealed Stochastic Preference (ARSP)). *For any finite sequence of trials* $\{(x_k, A_k) \in \Omega\}_{k=1}^m$,

$$\sum_{k=1}^{m} \rho(x_k, A_k) \leq \max_{\textit{total } R \in \mathcal{R}} |\{k \in \{1, \dots, m\} : x_k \textit{ maximizes } R \textit{ in } A_k\}|. \tag{15}$$

To see that (14) implies (15), note that x is a strict maximum for some R_v in A if and only if x is a maximum for the total order T_v such that for all $(i, \alpha), (j, \beta) \in X$,

$$(i, \alpha) T_v (j, \beta) \quad \Leftrightarrow \quad \text{either } q_v(i, \alpha) > q_v(j, \beta), \text{ or } q_v(i, \alpha) = q_v(j, \beta) \text{ and } i \leq j.$$

On the other hand, ARSP does not imply ARSQ (see an example in Section 5). To summarize, RQUM strengthens the classic RUM by refining ARSP to ARSQ and adding the continuity conditions AC and GC.

4.2. **RQUM in Finite Datasets.** Another benefit of ARSQ is that it can still characterize RQUM in finite datasets where Axioms 1–5 may all be vacuous and hence, insufficient for any representation.

Suppose that the stochastic choice rule ρ is observed only in a subclass of menus $\mathcal{F} \subset \mathcal{A}$. Define the corresponding set of trials as

$$\Omega(\mathcal{F}) = \{(x, A) : A \in \mathcal{F} \text{ and } x \in A\}.$$

A pair $(\rho, \mathcal{F})$ is called a *stochastic dataset* if $\mathcal{F} \subset \mathcal{A}$ and $\rho : \Omega(\mathcal{F}) \to [0, 1]$ is a function such that

$$\sum_{x \in A} \rho(x, A) = 1 \qquad \text{for all } A \in \mathcal{F}.$$

RQUM and ARSQ can be adopted as is.

Theorem 4. *A stochastic dataset* $(\rho, \mathcal{F})$ *satisfies ARSQ if and only if there is* $\pi \in \Pi$ *such that for all* $(x, A) \in \Omega(\mathcal{F})$,

$$\rho(x, A) = \pi(L(x, A)). \tag{16}$$

Moreover, it is without loss in generality to take π *as either regular or finite-ranged.*

This result identifies π via the Farkas Lemma. In contrast with Theorem 1, the identification of π is not unique, and it arrives as a solution to a linear program rather than the explicit formula (11).

Theorem 4 also implies that there is no empirical difference between RQUM and its modifications where the grading procedures rely on any permutation of N to break ties. Indeed, ARSQ is invariant with such permutations as long as the grading procedure is used in all menus. It is only the GC axiom that can be affected by the permutation of N in the grading procedure.

5. Discussion

Next, we explain how the WDZ Theorem can be derived from our results.

Call a function $F : \mathbb{R}^n \to \mathbb{R}$ *smooth* if it has a continuous partial derivative

$$D(F,r) = \frac{\partial^n F(r)}{\partial r_1 \partial r_2 \dots \partial r_n}$$

for all $r \in \mathbb{R}^n$. If F is smooth, then for all $w, v \in \mathbb{R}^n$,

$$\Delta(F,w,v) = \int_{w_n}^{v_n} \cdots \int_{w_2}^{v_2} \int_{w_1}^{v_1} D(F,r) dr_1 dr_2 \dots dr_n \tag{17}$$

by the iterative use of the Newton-Leibniz formula.[8] Accordingly, a smooth function F is jointly monotone if and only[9]

$$D(F,r) \geq 0 \quad \text{for all } r \in \mathbb{R}^n.$$

For example, if F_π is a smooth cdf, then it has a continuous density $D(F_\pi, r) \geq 0$ for all $r \in \mathbb{R}^n$.

Suppose next that a reduced scr ρ^* is differentiable, and the functions G_α are smooth for all $\alpha \in \mathbb{R}$. If $\gamma \to 0$, then by CPN, $\frac{\partial \rho_k^*(c)}{\partial c_j} \geq \frac{\partial \rho_j^*(c)}{\partial c_k}$ and

$$\frac{\partial \rho_k^*(c)}{\partial c_j} = \frac{\partial \rho_j^*(c)}{\partial c_k} \tag{18}$$

because j and k can be switched. The symmetry (18) is one of the WDZ assumptions. Moreover, JM is equivalent to the non-negativity

$$D(G_\alpha, v) \geq 0 \qquad \text{for all } v \in \mathbb{R}^n, \tag{19}$$

which is also assumed in the WDZ Theorem.

Conversely, one can show that conditions (18)–(19) together with AC imply CPN.[10] Therefore, Theorem 1 and Corolary 2 remain valid if CPN and JM are replaced by the differential conditions (18) and (19), and the Borel measure π is required to have a continuous density on $\mathbb{R}^n$.

The main benefits of our approach is the added generality and the formulation of axioms CPN and JM in terms of observed choice probabilities rather than their derivatives.

[8] For example, if $n = 2$, then $\Delta(F,w,v) = \int_{w_2}^{v_2} \int_{w_1}^{v_1} \frac{\partial^n F(r_1,r_2)}{\partial r_1 \partial r_2} dr_1 dr_2$.

[9] Indeed, if $D(F,w) < 0$ for some $w \in \mathbb{R}^n$, then by continuity, $D(F,r) < 0$ in some ε-neighborhood of w, which makes the integral in (17) negative when $v \gg w$ belongs to this ε-neighorhood.

[10] We omit the proof of this claim.

5.1. **Exponomial Example.** To illustrate the flexibility of our results, we derive an example where the distribution π does not have a continuous density, but still has a convenient parametric structure. Suppose that for all $v, w \in \mathbb{R}^n_+$,

$$1 - G_0(v + w) = (1 - G_0(v))(1 - G_0(w)) < 1. \tag{20}$$

This multiplicative version of the Cauchy functional equation implies (e.g., Aczel[1, Theorem 1, p. 215] that the cdf G_0 must have the form

$$G_0(v) = \begin{cases} \prod_{i=1}^n (1 - \exp(-\lambda_i v_i)) & \text{for all } v \in \mathbb{R}^n_+ \\ 0 & \text{otherwise} \end{cases} \tag{21}$$

for some positive parameters $\lambda_1, \lambda_2, \ldots, \lambda_n > 0$. Thus, G_0 is a multivariate exponential cdf. Theorem 1 (or Corollary 2) characterizes RQUM with the exponential distribution G_0 via Axioms 1–5 and the additional condition (20). This specification is an example of the *exponomial* discrete choice model under the convention that the value v_0 is unperturbed.

5.2. **Axioms 1–5 are logically independent.** NC can be violated by *context effects*. For example, if x discounts y, then the presence of y can make x more likely to be chosen due to increased attention or reason-based heuristics (e.g., Shafir, Simonson, and Tversky [27]). Note that Axioms 2–5 need not be violated by context effects because the reduction ρ^* is restricted to assortments that have the same size $n + 1$.

Several examples below use $N = \{0, 1\}$ with two elements. In this binary framework, any reduced scr $\rho^* : \mathbb{R}^2 \to [0, 1]$ determines a unique scr ρ that satisfies NC and has ρ^* as its reduction.[11] So in these examples, it is enough to specify ρ^*.

Without CPN, the evaluation of monetary costs need not be linear, and some types can exhibit reference dependence where positive costs can appear more significant than negative ones (i.e., rewards). To illustrate, let $N = \{0, 1\}$ and

$$\rho_0^*(c_0, c_1) = \begin{cases} 1 & \text{if } u(c_1) - u(c_0) \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

where $u(\alpha) = \alpha$ for all $\alpha \geq 0$ and $u(\alpha) = \frac{\alpha}{2}$ for all $\alpha < 0$. JM is trivial here because ρ_0^* is increasing in c_1. AC and GC are also obvious. However, CPN does not hold. For example, if $c = (0, 0)$, then increasing the cost of good 1 by one unit increases ρ_0^* from 0 to 1, whereas decreasing the cost of good 0 by one unit does not change ρ_1^* at all.

[11] In any menu A that provides only one good, 0 or 1, the cheapest option is selected with probability one. In any menu A where c_0 and c_1 are the smallest available costs for goods 0 and 1, the pairs $(0, c_0)$ and $(1, c_1)$ must be chosen with probabilities $\rho_0^*(c_0, c_1)$ and $\rho_1^*(c_0, c_1)$ respectively. All other alternatives must be chosen with probability zero.

Without JM, the weights of some possible types can become negative. To illustrate, let $N = \{0, 1\}$, and for all $c \in \mathbb{R}^N$,

$$\rho_0^*(c_0, c_1) = \begin{cases} 1 & \text{if } c_1 - c_0 \in [0, 1) \cup [2, +\infty) \\ 0 & \text{otherwise.} \end{cases}$$

Here CPN holds because ρ^* is determined by the difference $c_1 - c_0$. AC and GC are obvious. However, JM does not hold because ρ_0^* is not increasing with respect to c_1. One can interpret ρ^* as an aggregation of three quasi-linear orders with unit weights: positive ones R_0 and R_2 and a negative one R_1.

Without AC, there can be possible types who do not care about money at all. To illustrate, let $N = \{0, 1\}$ and

$$\rho_0^*(c_0, c_1) = 1$$

for all $c \in \mathbb{R}^N$. Then Axioms 1–3 and 5 hold, whereas AC is false. The reason is that ρ^* is produced by an agent who is not willing to reject good 0 regardless of its cost. ARSQ holds here because any finite dataset that is generated when $\rho_0^*(c_0, c_1) = 1$ can also be generated by RQUM with a single type $R_{-\alpha}$ for sufficiently large $\alpha > 0$.

Without GC, there can be other tie-breaking rules that are consistent with Axioms 1–4. To illustrate, let $N = \{0, 1\}$ and

$$\rho_0^*(c_0, c_1) = \begin{cases} 1 & \text{if } c_1 > c_0 \\ \frac{1}{2} & \text{if } c_1 = c_0 \\ 0 & \text{if } c_1 < c_0. \end{cases}$$

Then Axioms 1–4 are holds, but GC is violated at $c = (0, 0)$ because $\rho^*(0, 0) = \frac{1}{2}$, but $\rho_0^*((0, 0) + \gamma(0, 1)) = 1$ for all $\gamma > 0$. This example corresponds to the *uniform* tie-breaking rule that is distinct from our grading procedure. Note that ARSQ holds in this example as well. To see this, note that

$$\rho = \tfrac{1}{2}\rho^+ + \tfrac{1}{2}\rho^-,$$

where ρ^+ is generated by RQUM with one type R_0, and ρ^- is generated by the mirror version of RQUM with one type R_0 where all ties are broken in favor of good 1. As both ρ^+ and ρ^- satisfy RQUM and ARSQ, ρ satisfies ARSQ as well.

5.3. **Wealth Invariance is redundant.** It follows from (10) that the reduction ρ^* should be invariant to wealth variations where the cost differentials across all goods in N are unchanged.

Axiom 8 (Wealth Invariance)**.** *For all $c \in \mathbb{R}^N$ and $\gamma \in \mathbb{R}$, $\rho^*(c) = \rho^*(c + \gamma)$.*

Axioms 1–5 imply Wealth Invariance, but this claim is not trivial and requires a technical Lemma 6.1 in the proofs. This lemma could be omitted if Wealth Invariance is just added to the assumptions in Theorem 1, but then the list of our axioms would become redundant.

5.4. **ARSQ vs ARSP.** Consider an example where ARSP holds but ARSQ does not. Let $N = \{0, 1\}$, and

$$\rho_0^*(c_0, c_1) = \begin{cases} 1 & \text{if } c_0 \le 0 \\ 0 & \text{if } c_0 > 0. \end{cases}$$

Let ρ be the unique extension of ρ^* that satisfies NC. ARSQ does not hold here because CPN is violated:

$$1 = \rho_1^*(1, 0) - \rho_1^*(0, 0) > \rho_0^*(0, 0) - \rho_0^*(0, -1) = 0.$$

However, ρ satisfies ARSP because ρ selects the maximizer of a total order R that is represented by a utility function

$$u(i, \alpha) = \begin{cases} -3 - \alpha & \text{if } i = 0 \text{ and } \alpha > 0 \\ 3 - \alpha & \text{if } i = 0 \text{ and } \alpha \le 0 \\ \arctan(-\alpha) & \text{if } i = 1. \end{cases}$$

Of course, this total order does not have a quasi-linear utility representation.

6. APPENDIX

Show Theorem 1.

Suppose that a Borel probability measure $\pi \in \Pi$ is a low representation (9) for a stochastic choice rule ρ. Then NC holds because for all $(x, A) \in \Omega$ and $y \in X$,

$$L(x, A \cup y) \subset L(x, A)$$

and $L(c, A \cup y) = \emptyset$ if y discounts x.

Representation (9) for ρ implies (10) for its reduction ρ^*.

Show CPN. By (10),

$$\rho_k^*(c) - \rho_k^*\left(c - \gamma\vec{j}\right) = \pi\left[L_k(c) \setminus L_k\left(c - \gamma\vec{j}\right)\right]$$
$$\rho_j^*\left(c + \gamma\vec{k}\right) - \rho_j^*(c) = \pi\left[L_j(c + \gamma\vec{k}) \setminus L_j(c)\right]$$

because $L_j(c)$ is a subset of $L_j\left(c + \gamma\vec{k}\right)$, and $L_k\left(c - \gamma\vec{j}\right)$ is a subset of $L_k(c)$. To derive CPN, show the set inclusion

$$\left[L_j\left(c + \gamma\vec{k}\right) \setminus L_j(c)\right] \subset \left[L_k(c) \setminus L_k\left(c - \gamma\vec{j}\right)\right].$$

Take any type $v \in \mathbb{R}^n$ such that its quasi-linear order R_v is maximized by (j, c_j) in the assortment $A\left(c + \gamma\vec{k}\right)$, but not in $A(c)$. Then (k, c_k) should maximize R_v in $A(c)$. By quasi-linearity, (k, c_k) cannot maximize R_v in $A\left(c - \gamma\vec{j}\right)$ because then (j, c_j) would not maximize R_v in $A\left(c + \gamma\vec{k}\right)$.

JM is implied by the identity $G_\alpha = F_\pi$.

Show AC. For all $m = 1, 2, \dots,$ let

$$V_m = \{v \in \mathbb{R}^n : \max_{k \in [1,n]} |v_k| \geq m\}.$$

These sets are monotonically decreasing, $V_1 \supset V_2 \supset \dots$, and satisfy $\bigcap_{k=1}^{\infty} V_k = \emptyset$. As π is countably additive, then

$$\lim_{m \to \infty} \pi(V_m) = 0.$$

Take any $\varepsilon > 0$. Pick m such that $\pi(V_m) < \varepsilon$. Let $\delta = 2m$. Take any $c \in \mathbb{R}^N$ and $k, i \in [1, n]$ such that $c_k - c_i > \delta$. Suppose that $v \in L_k(c)$. Then (k, c_k) maximizes q_v in $A(c)$. It follows that $v_k - c_k \geq v_i - c_i$ where $v_0 = 0$ by convention. Thus $v_k - v_i \geq c_k - c_i > 2m$ and hence, either $v_k \geq m$ or $v_i \leq -m$. In either case, $v \in V_m$, and hence,

$$\rho_k^*(c) = \pi(L_k(c)) \leq \pi(V_m) < \varepsilon.$$

Show GC. Take any $c \in \mathbb{R}^N$ and $k \in N$. For all $m = 1, 2, \dots,$ let:

- W_m be the set of all $v \in \mathbb{R}^n$ such that

$$v_k - c_k > v_i - c_i + \tfrac{1}{m} \qquad \text{for all } i \in N \text{ such that } i < k$$
$$v_k - c_k \geq v_j - c_j \qquad \text{for all } j \in N \text{ such that } j \geq k.$$

- W'_m be the set of all $v \in \mathbb{R}^n$ such that

$$v_k - c_k > v_i - c_i \qquad \text{for all } i \in N \text{ such that } i < k$$
$$v_k - c_k \geq v_j - c_j - \tfrac{1}{m} \qquad \text{for all } j \in N \text{ such that } j \geq k.$$

The set inclusions $W_1 \subset W_2 \subset \dots$ and $W'_1 \supset W'_2 \supset \dots$ are obvious.

As $L_k(c) = \bigcup_{m \to \infty} W_m = \bigcap_{m \to \infty} W'_m$ and π is countably additive, then

$$\rho_k^*(c) = \pi(L_k(c)) = \lim_{m \to \infty} \pi(W_m) = \lim_{m \to \infty} \pi(W'_m).$$

Let $d = (0, 1, \dots, n)$. Then for all $0 < \gamma < \frac{1}{mn}$,

$$W_m \subset L_k(c + \gamma d) \subset W'_m.$$

Thus $\lim_{\gamma \to 0, \gamma \geq 0} \rho_k^*(c + \gamma d) = \lim_{\gamma \to 0, \gamma \geq 0} \pi(L_k(c + \gamma d)) = \rho_k^*(c)$. As $k \in N$ is arbitrary, then ρ^* is continuous in the direction d.

Suppose that π satisfies the regularity condition (8). Show that ρ^* is continuous. For any $c \in \mathbb{R}^n$ and $k \in N$,

$$\lim_{\gamma \to 0, \gamma \geq 0} \pi(L_k(c - \gamma \vec{k})) = \pi(M((k, c_k), A(c))) = \pi(L_k(c)) =$$
$$\pi(S((k, c_k), A(c))) = \lim_{\gamma \to 0, \gamma \geq 0} \pi(L_k(c + \gamma \vec{k})).$$

Take any sequence $c(m) \in \mathbb{R}^n$ such that $\lim_{m \to \infty} c(m) = c$. Take any $\varepsilon > 0$. Then there is $\gamma > 0$ such that

$$\pi(L_k(c - \gamma \vec{k})) - \pi(L_k(c + \gamma \vec{k})) < \varepsilon.$$

Then for all sufficiently large m, $||c(m) - c|| < \gamma$ and hence,

$$L_k(c + \gamma \vec{k}) \subset L_k(c(m)) \subset L_k(c - \gamma \vec{k}).$$

As $L_k(c+\gamma\vec{k}) \subset L_k(c) \subset L_k(c-\gamma\vec{k})$ as well, then

$$|\pi(L_k(c(m))) - \pi(L_k(c))| < |\pi(L_k(c-\gamma\vec{k})) - \pi(L_k(c+\gamma\vec{k}))| < \varepsilon.$$

Thus $\lim_{m\to\infty} \rho_k^*(c(m)) = \rho_k^*(c)$.

Suppose that π has a finite range. It is then obvious that ρ^* has a finite range as well because each of its components ρ_i^* has a finite range.

6.1. **Sufficiency of Axioms.** Suppose that ρ satisfies Axioms 1–5. Take any $\alpha \in \mathbb{R}$. For all $v \in \mathbb{R}^n$, let

$$G_\alpha(v) = \rho_0^*(\alpha, v+\alpha).$$

By JM, G_α is jointly monotone. By AC, for all $j \in [1,n]$,

$$\lim_{v_j \to -\infty} G_\alpha(v) = 0 \tag{22}$$

because the cost differential $\alpha-(v_j+\alpha)$ between goods 0 and j becomes arbitrarily large in this limit. Similarly, by AC,

$$\lim_{\gamma\to+\infty} G_\alpha(\gamma,\dots,\gamma) = 1 \tag{23}$$

because $G_\alpha(\gamma,\dots,\gamma) = 1 - \sum_{k=1}^n \rho_k^*(\alpha, \gamma+\alpha, \gamma+\alpha, \dots, \gamma+\alpha)$. Here

$$\lim_{\gamma\to+\infty} \rho_k^*(\alpha, \gamma+\alpha, \gamma+\alpha, \dots, \gamma+\alpha) = 0$$

for all $k \in [1,n]$ because the cost of good k exceeds the cost of good 0 by γ which becomes arbitrarily large.

Argue that G_α is monotonically increasing with respect to each of its variables. Suppose that

$$G_\alpha\left(v-\gamma\vec{k}\right) - G_\alpha(v) > 0$$

for some $v \in \mathbb{R}^n$, $k \in [1,n]$, and $\gamma > 0$. Let $\varepsilon = G_\alpha\left(v-\gamma\vec{k}\right) - G_\alpha(v)$. By AC, there is $\delta > 0$ such that for all $c \in \mathbb{R}^N$ and $j \in N$,

$$c_0 - c_j \geq \delta \quad \Rightarrow \quad \rho_0^*(c) < \tfrac{\varepsilon}{2^n}.$$

Take $w \in \mathbb{R}^n$ such that $w_k = v_k - \gamma$ and $w_i = -\delta$ for all other $i \in [1,n]\setminus k$. Then for any $K \subset [1,n]$ such that $K \neq \emptyset$ and $K \neq \{k\}$,

$$G_\alpha(wKv) = \rho_0^*(\alpha, (w+\alpha)K(v+\alpha)) \leq \tfrac{\varepsilon}{2^n}$$

because $\alpha - (w_j+\alpha) \geq \delta$ for $j \in K\setminus k$. Thus

$$G_\alpha(v) - G_\alpha\left(v-\gamma\vec{k}\right) = G_a(w\emptyset v) - G_\alpha(w\{k\}v) =$$

$$\sum_{K\subset[1,n]} (-1)^{|K|} G_\alpha(wKv) - \sum_{K\subset[1,n]:K\neq\emptyset \text{ and } K\neq\{k\}} (-1)^{|K|} G_\alpha(wKv) \geq$$

$$\sum_{K\subset[1,n]} (-1)^{|K|} G_\alpha(wKv) - (2^n-2)\tfrac{\varepsilon}{2^n} > -\varepsilon$$

because $v \geqq w$ and G_α is jointly monotone. This contradicts the definition of ε. Thus, G_α is weakly increasing with respect to all its variables.

Show that G_α satisfies *continuity from above.* To do so, take any vectors $w, v(1), v(2), \dots \in \mathbb{R}^n$ such that $v(1) \geqq v(2) \geqq \dots$ and

$$\lim_{k\to\infty} v(k) = w.$$

Let $d = (1, 2, \dots, n)$. Take any $\varepsilon > 0$. By GC, there is $\gamma > 0$ such that

$$G_\alpha(w + \gamma d) \leq G_\alpha(w) + \varepsilon.$$

The convergence $\lim_{k\to\infty} v(k) = w$ implies that there is m such that $w + \gamma d \geqq v(k) \geqq w$ for all $k \geq m$. As G_α is weakly increasing in all variables, then

$$G_\alpha(w) \leq G_\alpha(v(k)) \leq G_\alpha(w + \gamma) \leq G_\alpha(w) + \varepsilon.$$

As ε is arbitrary, then G_α is continuous from above.

Besides JM and continuity from above, the function $G_\alpha : \mathbb{R}^n \to [0, 1]$ satisfies the asymptotic normalizations (22)–(23). Billingsley's Theorem 12.5 implies that G_α is the cdf of some Borel measure $\pi_\alpha \in \Pi$.

The next lemma establishes that the functions G_α—and hence, the corresponding measures π_α—are invariant of α.

Lemma 6.1. *For all $\alpha \in \mathbb{R}$, $G_\alpha = G_0$.*

Proof. Suppose first that $n = 1$. Take any $\alpha \geq 0$ and $v \in \mathbb{R}$. Then

$$\begin{aligned} G_0(v) - G_\alpha(v) &= \rho_0^*(0, v) - \rho_0^*(\alpha, v + \alpha) = \\ &\rho_0^*(0, v) - \rho_0^*(0, v + \alpha) + \rho_0^*(0, v + \alpha) - \rho_0^*(\alpha, v + \alpha) =_{\text{as } n=1, \text{ then } \rho_0^* = 1 - \rho_1^*} \\ &\rho_0^*(0, v) - \rho_0^*(0, v + \alpha) + [\rho_1^*(\alpha, v + \alpha) - \rho_1^*(0, v + \alpha)] \leq_{\text{by CPN}} \\ &\rho_0^*(0, v) - \rho_0^*(0, v + \alpha) + [\rho_0^*(0, v + \alpha) - \rho^*(0, v)] = 0. \end{aligned}$$

Similarly,

$$\begin{aligned} G_\alpha(v) - G_0(v) &= [\rho_0^*(\alpha, v + \alpha) - \rho_0^*(\alpha, v)] + [\rho_1^*(0, v) - \rho_1^*(\alpha, v)] \leq_{\text{by CPN}} \\ &[\rho_1^*(\alpha, v) - \rho_1^*(0, v)] + [\rho_1^*(0, v) - \rho_1^*(\alpha, v)] = 0. \end{aligned}$$

Thus $G_\alpha(v) = G_0(v)$.

Let $n \geq 2$. Take any $\gamma \in \left(0, \frac{2}{n(n-1)}\right]$. Show that for all $\alpha, \beta \in \mathbb{R}$ and $v \in \mathbb{R}^n$,

$$\left| \int_{w \in \mathbb{R}^n : v + \gamma \geqq w \geqq v} [G_\alpha(w) - G_\beta(w)] dw \right| \leq (\alpha - \beta)^2, \tag{24}$$

where the integration is taken over all vectors $w \in \mathbb{R}^n$ such that $v + \gamma \geqq w \geqq v$.

Before proving (24), we see that (24) implies $G_\alpha = G_0$. Indeed, suppose that $G_\alpha(v) \neq G_0(v)$ for some $\alpha \in \mathbb{R}$ and $v \in \mathbb{R}^n$. Both G_α and G_0 are continuous from above, and hence,

$$\int_{w \in \mathbb{R}^n : v + \gamma \geqq w \geqq v} G_\alpha(w) dw \neq \int_{w \in \mathbb{R}^n : v + \gamma \geqq w \geqq v} G_0(w) dw$$

for all sufficiently small γ. However, (24) implies that for any $m = 1, 2, \ldots$,

$$\left| \int_{w \in \mathbb{R}^n : v+\gamma \geqq w \geqq v} [G_\alpha(w) - G_0(w)] dw \right| \leq m \tfrac{\alpha^2}{m^2} = \tfrac{\alpha^2}{m}.$$

As m is arbitrary, then the inequality $G_\alpha(v) \neq G_0(v)$ is impossible.

So, it remains to show (24). The proof invokes the Fubini theorem as in Billingsley [5, Theorem 18.3].

Without loss in generality, let $\beta = \alpha + \varepsilon$ for some $\varepsilon \geq 0$. Then

$$\begin{aligned}
G_\alpha(w) - G_\beta(w) &= \rho_0^*(\alpha, w+\alpha) - \rho_0^*(\beta, w+\beta) = \\
&\rho_0^*(\alpha, w+\alpha) - \rho_0^*(\alpha, w+\beta) + \rho_0^*(\alpha, w+\beta) - \rho_0^*(\beta, w+\beta) = \\
&\rho_0^*(\alpha, w+\alpha) - \rho_0^*(\alpha, w+\beta) + \sum_{k=1}^n [\rho_k^*(\beta, w+\beta) - \rho_k^*(\alpha, w+\beta)] \leq \text{ (By CPN)} \\
&\rho_0^*(\alpha, w+\alpha) - \rho_0^*(\alpha, w+\beta) + \sum_{k=1}^n [\rho_0^*(\alpha, w+\beta) - \rho_0^*((\alpha, w+\beta) - \varepsilon\vec{k})] = \\
&G_\alpha(w) + (n-1)G_\alpha(w+\varepsilon) - \sum_{k=1}^n G_\alpha(w+\varepsilon-\varepsilon\vec{k}) = \\
&\int_{r\in\mathbb{R}^n} \left[g(r,w) + (n-1)g(r,w+\varepsilon) - \sum_{k=1}^n g(r, w+\varepsilon-\varepsilon\vec{k}) \right] d\pi_\alpha(r) \leq \\
&\int_{r\in\mathbb{R}^n} \sum_{i,j\in[1,n]:i>j} h(r,w) d\pi_\alpha(r)
\end{aligned}$$

where $g(r,w)$ and $h(r,w)$ are indicator functions such that

$$g(r,w) = \begin{cases} 1 & \text{if } w \geqq r \\ 0 & \text{otherwise} \end{cases} \qquad h(r,w) = \begin{cases} 1 & \text{if } r_i - w_i \in (0,\varepsilon] \text{ and } r_j - w_j \in (0,\varepsilon] \\ 0 & \text{otherwise.} \end{cases}$$

Here, the inequality

$$g(r,w) + (n-1)g(r,w+\varepsilon) - \sum_{k=1}^n g(r, w+\varepsilon-\varepsilon\vec{k}) \leq \sum_{i,j\in[1,n]:i>j} h(r,w) \tag{25}$$

must hold for all $r, w \in \mathbb{R}^n$. To see this, consider several cases.

If $w \geqq r$, then $g(r,w) = g(r,w+\varepsilon) = g(r,w+\varepsilon-\varepsilon\vec{k}) = 1$ for all $k \in [1,n]$, and hence, the left-hand side of (25) is zero.

If $w+\varepsilon \geqq r$ is not true, then $g(r,w) = g(r,w+\varepsilon) = g(r,w+\varepsilon-\varepsilon\vec{k}) = 0$ for all $k \in [1,n]$, and hence, the left-hand side of (25) is zero.

So suppose that $w+\varepsilon \geqq r$ and there is a positive count p of variables $i \in [1,n]$ such that $r_i - w_i \in (0,\varepsilon]$. Then the left-hand side of (25) is $p-1$, and the right-hand side is $\frac{p(p-1)}{2}$. Thus (24) must hold.

Conclude the proof of (24) by the Fubini theorem.

$$\int_{w\in\mathbb{R}^n:v+\gamma\geqq w\geqq v}[G_\alpha(w)-G_\beta(w)]dw \leq$$

$$\int_{w\in\mathbb{R}^n:v+\gamma\geqq w\geqq v}\left[\int_{r\in\mathbb{R}^n}\sum_{i,j\in[1,n]:i>j}h(r,w)d\pi_\alpha(r)\right]dw = \quad \text{Fubini}$$

$$\int_{r\in\mathbb{R}^n}\left[\int_{w\in\mathbb{R}^n:v+\gamma\geqq w\geqq v}\sum_{i,j\in[1,n]:i>j}h(r,w)dw\right]d\pi_\alpha(r) \leq$$

$$\int_{r\in\mathbb{R}^n}\tfrac{n(n-1)}{2}\gamma^{n-2}\varepsilon^2 d\pi(r) \leq \varepsilon^2,$$

since for any fixed $r\in\mathbb{R}^n$, the Lebesgue measure of the intersection of the set $\{w\in\mathbb{R}^n : v+\gamma\geqq w\geqq v\}$ with the constraints $r_i-w_i\in(0,\varepsilon]$ and $r_j-w_j\in(0,\varepsilon]$ for any distinct i,j can be bounded by $\gamma^{n-2}\varepsilon^2$. As $\gamma\leq\frac{2}{n(n-1)}$, then $\gamma\leq 1$ and hence $\frac{n(n-1)}{2}\gamma^{n-2}\varepsilon^2\leq\varepsilon^2$.

Similarly, $\int_{w\in\mathbb{R}^n:v+\gamma\geqq w\geqq v}[G_\beta(w)-G_\alpha(w)]dw\leq\varepsilon^2$ when $\beta<\alpha$. □

Let $\pi=\pi_0$. Take any cost vector $c\in\mathbb{R}^N$. Let $v=(c_1-c_0,c_2-c_0,\ldots,c_n-c_0)$. By Lemma 6.1,

$$\rho_0^*(c)=G_{c_0}(v)=G_0(v)=\pi(\{w\in\mathbb{R}^n : w\leqq v\})=\pi(L_0(c))$$

because $\{w\in\mathbb{R}^n : w\leqq v\}=L_0(c_0,v+c_0)=L_0(c)$.

Extend the low representation to all other goods $k>0$.

Lemma 6.2. *For all $c\in\mathbb{R}^n$ and $k\in[1,n]$,*

$$\rho_k^*(c)\geq\pi(S_k(c)) \tag{26}$$

where $S_k(c)=\{v\in\mathbb{R}^n : (k,c_k)$ strictly maximizes R_v in $A(c)\}$.

Proof. Take any $c\in\mathbb{R}^N$ and $k\in[1,n]$. For any $t=1,2,\ldots$, let

$$V_t=\bigcup_{m=0}^{4^t-1}\left[L_0\left(c+\tfrac{1}{2^t}\vec{k}-\tfrac{m}{2^t}\vec{0}\right)\setminus L_0\left(c-\tfrac{m}{2^t}\vec{0}\right)\right].$$

Show that these sets are nested:

$$V_1\subset V_2\subset V_3\subset\ldots$$

Take any $v\in V_t$. Then there is $m\in\{0,1,\ldots,4^t-1\}$ such that

$$v\in L_0\left(c+\tfrac{1}{2^t}\vec{k}-\tfrac{m}{2^t}\vec{0}\right)\setminus L_0\left(c-\tfrac{m}{2^t}\vec{0}\right)$$

If $v\in L_0\left(c+\frac{1}{2^{t+1}}\vec{k}-\frac{2m}{2^{t+1}}\vec{0}\right)$, then $v\in V_{t+1}$ because

$$v\in L_0\left(c+\tfrac{1}{2^{t+1}}\vec{k}-\tfrac{2m}{2^{t+1}}\vec{0}\right)\setminus L_0\left(c-\tfrac{2m}{2^{t+1}}\vec{0}\right).$$

If $v \notin L_0\left(c+\frac{1}{2^{t+1}}\vec{k}-\frac{2m}{2^{t+1}}\vec{0}\right)$, then $v \in V_{t+1}$ because

$$v \in L_0\left(c+\frac{1}{2^{t+1}}\vec{k}+\frac{1}{2^{t+1}}\vec{k}-\frac{2m}{2^{t+1}}\vec{0}\right) \setminus L_0\left(c+\frac{1}{2^{t+1}}\vec{k}-\frac{2m}{2^{t+1}}\vec{0}\right).$$

Next, show that the union $\bigcup_{t=1}^{\infty} V_t$ contains $S_k(c)$. Suppose that $v \in \mathbb{R}^n$ is such that (k, c_k) is a strict maximum for R_v in $A(c)$. Then $v_k - c_k > -c_0$. Take t such that

- $v_k - c_k \leq 2^t - c_0$, and
- $v_k - c_k - \frac{1}{2^t} > v_i - c_i$ for all $i \in [1, n] \setminus k$.

Pick $m \in \{0, \ldots, 4^t - 1\}$ such that

$$\frac{m}{2^t} - c_0 < v_k - c_k \leq \frac{m+1}{2^t} - c_0.$$

Then $v \in L_0\left(c+\frac{1}{2^t}\vec{k}-\frac{m}{2^t}\vec{0}\right) \setminus L_0\left(c-\frac{m}{2^t}\vec{0}\right)$ and hence, $v \in V_m$.

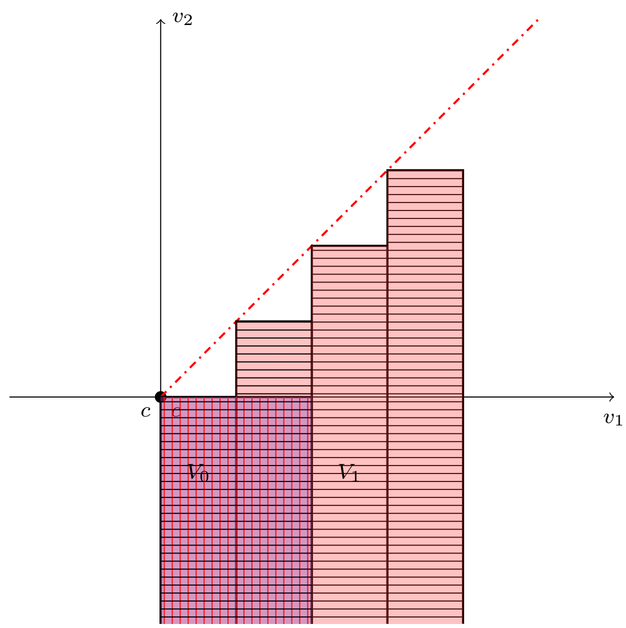

FIGURE 2. Fix $c = (0, 0, 0)$ and $k = 1$, then V_0 is the area shaded by horizontal lines; V_1 is the area shaded by vertical lines. The width of V_1 is double the width of V_0.

For each $t = 1, 2, \ldots,$

$$\rho_k^*(c) - \rho_k^*(c - 2^t\vec{0}) = \sum_{m=0}^{4^t-1}\left[\rho_k^*\left(c-\frac{m}{2^t}\vec{0}\right) - \rho_k^*\left(c-\frac{m+1}{2^t}\vec{0}\right)\right] \geq \text{ (by CPN)}$$

$$\sum_{m=0}^{4^t-1} \rho_0^*\left(c+\frac{1}{2^t}\vec{k}-\frac{m}{2^t}\vec{0}\right) - \rho_0^*\left(c-\frac{m}{2^t}\vec{0}\right) = \pi(V_t)$$

As $t \to \infty$, the limit of $\rho_k^*(c - 2^t\vec{0})$ should be zero by AC, and hence,

$$\rho_k^*(c) = \lim_{t\to\infty} \pi(V_t) \geq \pi(S_k(c))$$

because π is countably additive.

□

Let $d = (0, 1, 2, \ldots, n)$.

Lemma 6.3. *For any $c \in \mathbb{R}^n$, there is a sequence $\{\gamma_t \geq 0\}_{t=1}^{\infty}$ such that*

(1) for all $k \in N$ and $t = 1, 2, \ldots,$

$$\pi(L_k(c + \gamma_t d)) = \pi(S_k(c + \gamma_t d))$$

(2) $\lim_{t\to\infty} \gamma_t = 0$.

Proof. Suppose that the lemma is not true. Then there are $c \in \mathbb{R}^N$, $k \in N$, and $\varepsilon > 0$ such that

$$\pi(L_k(c + \gamma d)) > \pi(S_k(c + \gamma d))$$

for all $\gamma \in [0, \varepsilon]$. Note that

$$L_k(c + \gamma d) \setminus S_k(c + \gamma d) \subset \bigcup_{i\in N\setminus k} T(i, \gamma)$$

where $T(i, \gamma) = \{v \in \mathbb{R}^n : q_v(i, c_i + \gamma i) = q_v(k, c_k + \gamma k)\}$. As $i \neq k$, then $T(i, \gamma) \cap T(i, \alpha) = \emptyset$ for all $\alpha \neq \gamma$. As π is countably additive, then there can be only countably many points $\gamma \in \mathbb{R}$ such that $\pi(T(i, \gamma)) > 0$. Thus, there can be countably many points $\gamma \in \mathbb{R}$ such that

$$\pi(L_k(c + \gamma d)) > \pi(S_k(c + \gamma d))$$

can hold. Thus, this inequality does not hold for some $\gamma \in [0, \varepsilon]$. □

Take any $c \in \mathbb{R}^n$. Take a sequence $\{\gamma_t \geq 0\}_{t=1}^{\infty}$ that satisfies Lemma 6.3. For all $k \in N$ and $t = 1, 2, \ldots,$ Lemma 6.2 implies

$$1 = \sum_{k=0} \rho_k^*(c + \gamma_t d)) \geq \sum_{k=0} \pi(S_k(c + \gamma_t d)) = \sum_{k=0} \pi(L_k(c + \gamma_t d)) = 1$$

and hence,

$$\rho_k^*(c + \gamma_t d)) = \pi(L_k(c + \gamma_t d)).$$

Let ρ^{**} be the reduced scr that has π as a low representation. Both ρ^* and ρ^{**} satisfy GC. Thus

$$\rho^*(c) = \lim_{t\to\infty} \rho^*(c + \gamma_t d)) = \lim_{t\to\infty} \rho^{**}(c + \gamma_t d)) = \rho^{**}(c). \tag{27}$$

Thus, π is a low representation for ρ^*.

Extend this representation for the entire ρ. Take any trial $(x, A) \in \Omega$. Let $B \subset A$ consist of all alternatives $(i, \alpha) \in A$ such that for all $\beta < \alpha$, $(i, \beta) \notin A$. If $x \in A \setminus B$, then x is discounted by some $y \in A$ and hence, by NC

$$\rho(x, A) = 0 = \pi(\emptyset) = \pi(L(x, A)).$$

Suppose that $x \in B$. For each $i \in N$ and $t = 1, 2, \dots,$ define a cost vector $c(t) \in \mathbb{R}^N$ as

$$c_i(t) = \begin{cases} \min\{\alpha \in \mathbb{R} : (i,\alpha) \in B\} & \text{if } (i,\alpha) \in B \text{ for some } \alpha \in \mathbb{R} \\ t & \text{if } (i,\alpha) \notin B \text{ for all } \alpha \in \mathbb{R}. \end{cases} \tag{28}$$

Then B is a subset of the assortment $A(c(t))$. Take any $x = (i,\alpha) \in B$. By NC,

$$\rho(x, B) \geq \rho_i^*(c(t)).$$

The sets $L_i(c(t))$ are nested

$$L_i(c(1)) \subset L_i(c(2)) \subset \dots$$

and $\bigcup_{t=1}^{\infty} L_i(c(t)) = L(x, B)$. Thus

$$\rho(x,B) \geq \lim_{t\to\infty} \rho_i^*(c(t)) = \lim_{t\to\infty} \pi(L_i(c(t))) = \pi(L(x,B)) = \pi(L(x,A)).$$

By NC, $\rho(y, B) \geq \rho(y, A)$ for all $y \in B$. Thus

$$1 = \sum_{y\in B} \rho(y,B) \geq \sum_{y\in B} \rho(y,A) = \sum_{y\in B} \rho(y,A) + \sum_{y\in A\setminus B} \rho(y,A) = 1.$$

Thus, $\rho(y, B) = \rho(y, A)$ for all $x \in B$. Similarly, for all $y \in A$, $\rho(y, A) \geq \pi(L(y, A))$ and hence,

$$1 = \sum_{y\in A} \rho(y,A) \geq \sum_{y\in A} \pi(L(y,A)) = 1.$$

Thus, $\rho(y, A) = \pi(L(y, A))$, which implies that π is a low representation for ρ.

If ρ^* is continuous, then π must satisfy the regularity condition (8). Indeed, suppose that $\pi(M(x, A)) > \pi(S(x, A))$ for some $(x, A) \in \Omega$. Let $x = (i, \alpha)$ for some $i \in N$ and $\alpha \in \mathbb{R}$. Construct the menu B and vectors $c(t)$ as in (28). As $\pi(M(x, A)) > 0$, then $x \in B$ and

$$\pi(M(x,B)) = \pi(M(x,A)) > \pi(S(x,A)) = \pi(S(x,B)).$$

For all $t = 1, 2, \dots,$ let $M_t(x, B) \subset M(x, B)$ be the set of all $v \in M(x, B)$ such that $q_v(x) > q_v(k, t)$ for all $k \in N$. By countable additivity,

$$\lim_{t\to\infty} \pi(M_t(x,B)) = \pi(M(x,B)).$$

Take t such that $\pi(M_t(x, B)) > \pi(S(x, B))$. Note that $M_t(x, B) \subset L_i(c(t) - \gamma\vec{i})$ for all $\gamma > 0$ because any $x \in M_t(x, B)$ is a maximum for R_v in $A(c(t))$, and hence a strict maximum for R_v in $A(c(t) - \gamma\vec{i})$.

Moreover, $L_i(c(t) + \gamma\vec{i}) \subset S(x, B)$ for all $\gamma > 0$. As ρ_i^* is continuous, then

$$\pi(M_t(x,B)) \leq \lim_{\gamma\to 0} L_i(c(t) - \gamma\vec{i}) = \rho_i^*(c(t)) = \lim_{\gamma\to 0} L_i(c(t) + \gamma\vec{i}) \leq \pi(S(x,B))$$

which contradicts $\pi(M_t(x, B)) > \pi(S(x, B))$. Thus, the regularity (8) must hold.

Finally, suppose that ρ^* has a finite range. Show that the support of π must be finite. To show this, define the marginal cdfs for all $i \in N$ and $\alpha \in \mathbb{R}$ as

$$F_i(\alpha) = \pi\{v \in \mathbb{R}^n : v_i \leq \alpha\}.$$

Each of these values is the limit of the values of the joint cdf F_π and hence, belongs to the closure of the finite range of ρ_0^*. The closure of a finite set is the same finite set. Therefore, F_i has a finite range as well, and hence, finitely many discontinuity points $D_i \subset \mathcal{R}$. By construction,

$$\pi(\{v \in \mathbb{R}^n : v_i \in D_i\}) = 1$$

and hence, $\pi(D_0 \times D_1 \times \cdots \times D_n) = 1$ as well.

Corollary 2 asserts equality (27), which was derived above from Axioms 2–5 without NC.

6.2. **Proof of Theorem 3.** Suppose that ρ satisfies Axioms 1–5. By Theorem 1, ρ has a low representation $\pi \in \Pi$. For any type $v \in \mathbb{R}^n$, define its *indicator* $l_v : \Omega \to \{0, 1\}$ for all $(x, A) \in \Omega$ as

$$l_v(x, A) = \begin{cases} 1 & \text{if } x \text{ is a low maximum for } R_v \text{ in } A \\ 0 & \text{otherwise.} \end{cases}$$

Then the low representation (9) implies that for all trials (x_k, A_k),

$$\rho(x_k, A_k) = \int_{v \in \mathbb{R}^n} l_v(x_k, A_k) d\pi(v)$$

and hence,

$$\sum_{k=1}^{m} \rho(x_k, A_k) = \int_{v \in \mathbb{R}^n} \left[\sum_{k=1}^{m} l_v(x_k, A_k)\right] d\pi(v) \le$$
$$\max_{v \in \mathbb{R}^n} \sum_{k=1}^{m} l_v(x_k, A_k) = \max_{v\, \in\, \mathbb{R}^n} |\{k \in \{1, \ldots, m\} : v \in L(x_k, A_k)\}|.$$

Take any $w \in \mathbb{R}^n$ such that

$$|\{k \in \{1, \ldots, m\} : w \in L(x_k, A_k)\}| = \max_{v\, \in\, \mathbb{R}^n} |\{k \in \{1, \ldots, m\} : v \in L(x_k, A_k)\}|.$$

Then for sufficiently small $\gamma > 0$,

$$w \in L(x_k, A_k) \quad \Leftrightarrow \quad [w - \gamma(0, 1, \ldots, n)] \in S(x_k, A_k).$$

Therefore, ρ satisfies ARSQ because

$$\sum_{k=1}^{m} \rho(x_k, A_k) \le |\{k \in \{1, \ldots, m\} : [w - \gamma(0, 1, \ldots, n)] \in S(x_k, A_k)\}|.$$

Show next that ARSQ implies Axioms 1–3.

Lemma 6.4. *If $\{(x_k, A_k) \in \Omega\}_{k=1}^m$ and $\{(y_i, B_i) \in \Omega\}_{i=1}^t$ are finite sequences of trials such that for all $v \in \mathbb{R}^n$,*

$$\sum_{k=1}^{m} l_v(x_k, A_k) \le \sum_{i=1}^{t} l_v(y_i, B_i), \tag{29}$$

then ARSQ implies that

$$\sum_{k=1}^{m} \rho(x_k, A_k) \le \sum_{i=1}^{t} \rho(y_i, B_i). \tag{30}$$

Proof. Inequality (29) implies

$$\sum_{k=1}^{m} l_v(x_k, A_k) + \sum_{i=1}^{t} \sum_{z \in B_i \setminus y_i} l_v(z, B_i) \le t$$

because $\sum_{z \in B_i \setminus y_i} l_v(z, B_i) = 1 - l_v(y_i, B_i)$ for all $i = 1, \ldots, t$. Thus, t is the maximal number of low maxima—and a fortiori, strict maxima—for the type $v \in \mathbb{R}^n$ in the sequence of all trials in the left-hand side of the above inequality. By ARSQ,

$$\sum_{k=1}^{m} \rho(x_k, A_k) + \sum_{i=1}^{t} \sum_{z \in B_i \setminus y_i} \rho(z, B_i) \le t$$

and hence, $\sum_{k=1}^{m} \rho(x_k, A_k) \le t - \sum_{i=1}^{t} \sum_{z \in B_i \setminus y_i} \rho(z, B_i) = \sum_{i=1}^{t} \rho(y_i, B_i)$. □

Assume ARSQ. Show NC. For all $(x, A) \in \Omega$ and $y \in X$, $l_v(x, A \cup y) \le l_v(x, A)$ for all $v \in \mathbb{R}^n$. By (30), $\rho(x, A \cup y) \le \rho(x, A)$. If y discounts x, then $l_v(x, A \cup y) = 0$ for all $v \in \mathbb{R}^n$. Thus $\rho(x, A \cup y) = 0$.

Show CPN. Take any $\gamma > 0$, $c \in \mathbb{R}^N$, and distinct goods $k, j \in N$. We claim that for all $v \in \mathbb{R}^n$,

$$\begin{gathered} l_v\left((k, c_k), A\left((c - \gamma\vec{j}\right)\right) + l_v\left((j, c_j), A\left(c + \gamma\vec{k}\right)\right) \le \\ l_v((k, c_k), A(c)) + l_v((j, c_j), A(c)). \end{gathered} \tag{31}$$

If $l_v\left((k, c_k), A\left((c - \gamma\vec{j}\right)\right) = l_v\left((j, c_j), A\left(c + \gamma\vec{k}\right)\right) = 0$, then (31) is trivial.

Suppose that $l_v\left((k, c_k), A\left((c - \gamma\vec{j}\right)\right) = 1$. As (k, c_k) is a low maximum in $A\left((c - \gamma\vec{j}\right)$, then (j, c_j) cannot be a low maximum in $A\left(c + \gamma\vec{k}\right)$ because the cost difference between the two goods is unchanged. Therefore, $l_v\left((j, c_j), A\left(c + \gamma\vec{k}\right)\right) = 0$. As (k, c_k) must be a low maximum in $A(c)$, then $l_v((k, c_k), A(c)) = 1$. Thus (31) must hold.

Suppose that $l_v\left((j, c_j), A\left(c + \gamma\vec{k}\right)\right) = 1$. Similar to the previous case, $l_v\left((k, c_k), A\left((c - \gamma\vec{j}\right)\right) = 0$. Moreover, the low maximum in $A(c)$ must be either (k, c_k) or (j, c_j). Therefore, $l_v((k, c_k), A(c)) + l_v((j, c_j), A(c)) \ge 1$ and hence, (31) must hold.

CPN follows from (31) and Lemma 6.4.

Show JM. Take any $\alpha \in \mathbb{R}$ and show that G_α is jointly monotone. Take any vectors $r, w \in \mathbb{R}^n$ such that $r \geqq w$. Then for all $r \in \mathbb{R}_n$

$$S = \sum_{K \subset [1,n]} (-1)^{|K|} l_v((0, \alpha), A(\alpha, wKr)) \ge 0. \tag{32}$$

To show this claim, consider two cases. Suppose first that $v_i + \alpha > w_i$ for all $i \in [1,n]$. Then for all non-empty $K \subset [1,n]$, $(0,\alpha)$ is not a maximum for R_v in $A(\alpha, wKr)$ because $v_i - w_i > -\alpha$ for $i \in K$. Thus $S = l_v((0,\alpha), A(\alpha, r)) \geq 0$. Suppose next that $v_i + \alpha \leq w_i$ for some $i \in [1,n]$. Take any $K \subset [1,n] \setminus i$. Then $(0,\alpha)$ is a low maximum for R_v in $A(\alpha, wKr)$ if and only if it is a low maximum for R_v in $A(\alpha, w(K \cup i)r)$ because $v_i - w_i \leq -\alpha$ and hence, (i, w_i) cannot be a low maximum in the presence of $(0,\alpha)$. Thus

$$l_v((0,\alpha), A(\alpha, wKr)) = l_v((0,\alpha), A(\alpha, w(K \cup i)r))$$

and hence,

$$S = \sum_{K \subset [1,n] \setminus i} (-1)^{|K|} \left[l_v((0,\alpha), A(\alpha, wKr)) - l_v((0,\alpha), A(\alpha, w(K \cup i)r)) \right] = 0.$$

Thus $S = 0$. By (32),

$$\sum_{\text{even } K \subset [1,n]} l_v((0,\alpha), A(\alpha, wKr)) \geq \sum_{\text{odd } K \subset [1,n]} l_v((0,\alpha), A(\alpha, wKr)).$$

By (30),

$$\sum_{\text{even } K \subset [1,n]} \rho((0,\alpha), A(\alpha, wKr)) \geq \sum_{\text{odd } K \subset [1,n]} \rho((0,\alpha), A(\alpha, wKr))$$

and hence,

$$\sum_{\text{even } K \subset [1,n]} \rho_0^*(\alpha, wKr) \geq \sum_{\text{odd } K \subset [1,n]} \rho_0^*(\alpha, wKr).$$

By definition of the function G_α,

$$\sum_{K \subset [1,n]} (-1)^{|K|} G_\alpha((w-\alpha)K(r-\alpha)) \geq 0.$$

Substitute $w + \alpha$ for w and $r + \alpha$ for r to argue that G_α is jointly monotone. JM follows.

6.3. **Proof of Theorem 4.** Take a stochastic dataset $\rho : \Omega(\mathcal{F}) \to [0,1]$. Low representation (16) implies ARSQ by the same argument as for stochastic choice rules.

Suppose instead that ρ satisfies ARSQ. Then Lemma 6.4 holds as is.

As $\Omega(\mathcal{F})$ is finite, then there are only finitely many functions $l : \Omega(\mathcal{F}) \to \{0,1\}$ such that $l = l_v$ for some $v \in \mathbb{R}^n$. Pick a finite set $W \subset \mathbb{R}^n$ such that for every $v \in \mathbb{R}^n$, there is $w \in W$ such that $l_v = l_w$. Use the Integer-Real Farkas Lemma (Chambers and Echenique [7, Lemma 1.13]) to conclude the proof. By that result, exactly one of the following cases must hold.

Case 1. The stochastic dataset $\rho : \Omega(\mathcal{F}) \to [0,1]$ can be written as

$$\rho = \sum_{w \in W} \pi(w) l_w$$

where $\pi(w) \geq 0$ for all $w \in W$. Then the probabilistic normalization $\sum_{w \in W} \pi(w) = 1$ holds because for any $A \in \mathcal{F}$,

$$1 = \sum_{x,A} \rho(x, A) = \sum_{x,A} \sum_{w \in W} \pi(w) l_w(x, A) = \sum_{w \in W} \pi(w) \left[\sum_{x,A} l_w(x, A) \right] = \sum_{w \in W} \pi(w).$$

Case 2. There exists an integer-valued function $z : \Omega(\mathcal{F}) \to \mathbb{Z}$ such that

$$\sum_{(x,A) \in \Omega(\mathcal{F})} z(x, A) l_w(x, A) = 0 \tag{33}$$

for all $w \in W$, but

$$\sum_{(x,A) \in \Omega(\mathcal{F})} z(x, A) \rho(x, A) < 0. \tag{34}$$

It follows from (33) that

$$\sum_{(x,A) \in \Omega(\mathcal{F}), z(x,A) \geq 0} z(x, A) l_w(x, A) \geq \sum_{(x,A) \in \Omega(\mathcal{F}), z(x,A) < 0} (-z(x, A)) l_w(x, A)$$

for all $w \in W$. By Lemma 6.4,

$$\sum_{(x,A) \in \Omega(\mathcal{F}), z(x,A) \geq 0} z(x, A) \rho(x, A) \geq \sum_{(x,A) \in \Omega(\mathcal{F}), z(x,A) < 0} (-z(x, A)) \rho(x, A),$$

which contradicts (34).

Thus, Case 1 must hold.

Show that the identification of π can be modified to be regular. For each $w \in W$, there exists a sufficiently small $\gamma > 0$ such that

$$w \in L(x_k, A_k) \quad \Leftrightarrow \quad [w - \gamma(0, 1, \ldots, n)] \in S(x_k, A_k).$$

Let B_w be a small open neighborhood of $w - \gamma(0, 1, \ldots, n)$ such that for all $v \in B_w$ and $k \in \{1, \ldots, m\}$,

$$w \in L(x_k, A_k) \quad \Leftrightarrow \quad v \in S(x_k, A_k).$$

Replace π with a continuous distribution $\sigma = \sum_{w \in W} \pi(w) \delta_w$, where $\delta_w \in \Pi$ has density

$$f(v) = \begin{cases} \frac{1}{\lambda(B_w)} & \text{if } v \in B_w \\ 0 & \text{if } v \notin B_w \end{cases}$$

and where $\lambda(B_w)$ is the Lebesgue volume of the neighborhood B_w. Then the dataset ρ has σ as a low representation, as well.

References

[1] J. Aczel. *Lectures on Functional Equations and Their Applications*. Dover Publications, Mineola, New York, third edition, 1966.

[2] M. Agranov and P. Ortoleva. Stochastic choice and preferences for randomization. *Journal of Political Economy*, 125:40–68, 2017.

[3] J. Apesteguia, M. A. Ballester, and J. Lu. Single-crossing random utility models. *Econometrica*, 85(2):661–674, 2017.

[4] S. Berry, J. Levinsohn, and A. Pakes. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 63(4):841–890, 1995.

[5] P. Billingsley. *Probability and Measure*. Wiley, New York, third edition, 1995.

[6] H. D. Block and J. Marschak. Random orderings and stochastic theories of response. Discussion papers 289, Cowles Foundation for Research in Economics, Yale University, 1959.

[7] C. P. Chambers and F. Echenique. *Revealed preference theory*, volume 56. Cambridge University Press, 2016.

[8] A. Daly and S. Zachery. Improved multiple choice models. In D. Hensher and Q. Dalvi, editors, *Identifying and Measuring the Determinants of Mode Choice*. London: Teakfield, 1979.

[9] J.-C. Falmagne. On a class of probabilistic conjoint measurement models: Some diagnostic properties. *Journal of Mathematical Psychology*, 19(2):73–88, 1979.

[10] M. Frick, R. Iijima, and T. Strzalecki. Dynamic random utility. *Econometrica*, 87(6):1941–2002, 2019.

[11] F. Gul and W. Pesendorfer. Supplement to 'random expected utility', 2005.

[12] F. Gul and W. Pesendorfer. Random expected utility. *Econometrica*, 74(1):121–146, 2006.

[13] L. Hurwicz and Uzawa. On the integrability of demand functions. *Preferences, utility and demand*, 1971.

[14] J. Lu. Random choice and private information. *Econometrica*, 84(6):1983–2027, 2016.

[15] J. Lu. Random ambiguity. *Theoretical Economics*, 16(2):539–570, 2021.

[16] R. D. Luce. Individual choice behavior, 1959.

[17] R. D. Luce and P. Suppes. Preference, utility, and subjective probability. In R. Bush and E. Galanter, editors, *Handbook of Mathematical Psychology*, volume 3, pages 249–410. Wiley New York, 1965.

[18] P. Manzini and M. Mariotti. Stochastic choice and consideration sets. *Econometrica*, 82(3):1153–1176, 2014.

[19] D. McFadden. Conditional logit analysis of qualitative choice behavior. In P. Zarembka, editor, *Frontiers in econometrics*, pages 105–142. Academic Press, New York, 1973.

[20] D. McFadden. Econometric models for probabilistic choice among products. *The Journal of Business*, 53:13–29, 1980.

[21] D. McFadden. Economic choices. *American Economic Review*, 91(3):351–378, 2001.

[22] D. McFadden and M. K. Richter. Stochastic rationality and revealed stochastic preference. In J. S. Chipman, D. McFadden, and K. Richter, Marcel, editors, *Preferences, Uncertainty, and Optimality, Essays in Honor of Leo Hurwicz*, pages 161–186. Westview Press, 1990.

[23] V. Nocke and N. Schutz. Quasi-linear integrability. *Journal of Economic Theory*, 169:603–628, 2017.

[24] E. Piermont. Disentangling strict and weak choice in random expected utility models. *Journal of Economic Theory*, 1202:1–34, 2022.

[25] D. Revelt and K. Train. Mixed logit with repeated choices: Households' choices of appliance efficiency level. *The Review of Economics and Statistics*, 80:647–657, 1998.

[26] J. Rust. Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica*, 55:999–1033, 1987.

[27] E. Shafir, I. Simonson, and A. Tversky. Reason-based choice. *Cognition*, 49:11—-36, 1993.

[28] X. Shi, M. Shum, and W. Song. Estimating semi-parametric panel multinomial choice models using cyclic monotonicity. *Econometrica*, 86:737–761, 2018.

[29] J. R.-V. Sωrensen and M. Fosgerau. How McFadden met Rockafellar and learned to do more with less. *Journal of Mathematical Economics*, 100:102629, 2022.

[30] K. Train. Recreation demand models with taste differences over people. *Land Economics*, 74:230–239, 1998.

[31] C. Turansick. Identification in the random utility model. *arXiv Preprint arXiv:2102.05570*, 2021.

[32] H. Williams. On the formulation of travel demand models and economic measures of user benefit. *Environment and Planning A*, 9:285–344, 1977.

[33] J. I. Yellott Jr. The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology*, 15(2):109–144, 1977.