# RANDOM QUASI-LINEAR UTILITY

ERYA YANG[1] AND IGOR KOPYLOV[2]

Abstract. We propose a *random quasi-linear utility model* (RQUM) where stochastic choices maximize quasi-linear utility functions that are randomly drawn via some probability distribution $\pi$. Utility ties are allowed and broken by a convenient lexicographic rule. Our main result characterizes RQUM and identifies the probability measure $\pi$ uniquely and explicitly in terms of stochastic choice data. McFadden's (1973) additive random utility model is obtained as a special case where ties have probability zero. Another distinct case captures finite populations and derives $\pi$ with a finite support. Our main axioms constrain aggregate effects of cost variations on choice probabilities. In particular, context and reference dependence are prohibited. We also characterize RQUM through a suitable 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 [23]), recreational fishing (Train [30]), selection of appliance efficiency levels (Revelt and Train [26]), and many other settings. A single agent's choices can be also random due to intertemporal planning (Rust [27]) or spontaneous variations in her 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 $\pi$. Such $\pi$ is interpreted in terms of heterogeneous preferences. More formally, $\pi$ is defined over a suitable set $\Theta$ 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*: January 3, 2023.

*JEL Classification Code.* D01, D90.

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

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.
The authors thank Professors Mira Frick, Ryota Iijima, Asen Kochov, Stergios Skaperdas, Michael McBride, Jay Lu, Andrew Ellis, Miguel Zerecero, and Jiawei Chen for their helpful suggestions on this paper. We would also like to thank the participants at seminars in UCI, Rochester, Yale, Sun Yat-sen, and SWUFE Universities.

In the *classic* RUM of Block and Marschak [7] (henceforth BM), the domain $X$ is finite, and $\Theta$ is the set of all *total orders* (i.e. complete, transitive, anti-symmetric preferences over $X$). Falmagne [10] characterizes the classic RUM via non-negativity of BM polynomials. McFadden and Richter [24] provide another characterization based on their Axiom of Revealed Stochastic Preference (ARSP).

In many applications, it is convenient to associate the set $\Theta$ with some particular class of utility functions on $X$. Most importantly, McFadden's [21] 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 $\alpha$, and the set $\Theta$ 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. Quasi-linearity is also very common in mechanism design, auction theory, bargaining theory, public economics etc.

To make the additive RUM well-defined, it is necessary that "the probability of ties is zero" (McFadden [22, p. S15]). Therefore, $\pi$ 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.

Under suitable continuity assumptions, the additive model is well-defined, but its axiomatic meaning is still unclear from a decision theoretic perspective. Indeed, the well-known characterization result by Daly and Zachery [9] (henceforth, DZ) relies heavily on differentiation of choice probability functions. In particular, DZ require that

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

where $\rho_k$ and $\rho_j$ 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 conditions cannot be refuted by empirical data because partial derivatives like $\frac{\partial \rho_k(c)}{\partial c_j}$ are unobservable. Moreover, the DZ 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 and achieves several objectives. First, it allows any Borel probability measure $\pi$ over $\Theta$ or more formally, over the Euclidean space $\mathbb{R}^n$ that parametrizes $\Theta$.

Second, RQUM is characterized via novel axioms that do not use differentiation, but constrain aggregate effects of cost variations on choice probabilities.

Third, the probability measure $\pi$ is identified uniquely and explicitly, which allows to interpret its parametric structure in terms of the observable $\rho$.

[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 $\rho(y, \{x, y\})$.

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.

RQUM represents stochastic choices via

$$\rho(x, A) = \pi(\{v \in \mathbb{R}^n : x \text{ has the lowest grade among maxima of } R_v \text{ in } A\}) \tag{2}$$

where $\pi$ is a Borel probability measure on the Euclidean space $\mathbb{R}^n$, and the *grade* of any pair $(i, \alpha)$ is defined as $i$. In other words, we combine the additive RUM with the tie-breaking rule that favors alternatives with lower grades.

Our main result (Theorem 1) characterizes RQUM via axioms that do not assume or imply differentiability for the functions $\rho_k(c)$. Roughly speaking, our main axioms prohibit context dependence, reference dependence, and a more complicated non-monotonic pattern that would contradict the identification of the probability distribution $\pi$ below.

A major benefit of RQUM is that $\pi$ can be uniquely and explicitly derived from the observed stochastic choice rule $\rho$. Indeed, the *cumulative distribution function* of $\pi$ for all $v \in \mathbb{R}^n$ must satisfy

$$F_\pi(v) = \rho((0,0), A) \tag{3}$$

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 (2) 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 $A$. Obviously, formula (3) implies the uniqueness of $\pi$, 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 $\pi$. Apesteigua, Ballester, and Lu [3] use a strong version of single-crossing to derive $\pi$ uniquely in terms of choices in binary menus.

Special cases of Theorem 1 include (but not limited to)

- McFadden's additive RUM where "the probability of ties is zero",
- finite populations where $\pi$ has a finite support.

To illustrate the analytical power of (3), we provide another quick example where the distribution $\pi$ is derived in the multivariate *exponential* form when the function $\rho$ in (3) satisfies a multiplicative Cauchy equation. Identification (3) is substantially simpler than the counterpart in the classic RUM where the construction of $\pi$ employs a multi-step procedure based on BM polynomials. The identification (3) is also the cornerstone of our proofs, but the full argument is complicated and invokes some results from probability theory (e.g. Billingsley [6, Theorem 12.5]) rather than differentiability techniques (see Koning and Ridder [15] and Forgerau et al. [11] for recent proofs and discussions of the differentiation approach).

Our next result (Theorem 3) characterizes RQUM via McFadden and Richter's [24] linear programming approach. This approach extends to finite datasets where the identification of $\pi$ is based on the Farkas Lemma rather than the formula (3). We argue that there are no observable distinctions between grading procedures if all ties are broken by any permutation of the set $\{0, 1, \ldots, n\}$. So the bias in favor of goods with low grades can be motivated by Occam's razor: it simplifies the grading procedure without changing its observable meaning.

Our work contributes to the growing list of refinements of RUM. In the *random expected utility model* (REUM) of Gul and Pesendofer [13], the domain $X$ consists of lotteries over deterministic prizes, and $\Theta$ is the class of preferences that have expected utility representations. In this case, the distribution $\pi$ is determined uniquely by $\rho$, but the identification of $\pi$ relies on compactness arguments from real analysis. Gul and Pesendorfer consider only the regular case where utility ties have probability zero. Piermont [25] combines REUM with various tie-breaking rules, but his extensions do not identify $\pi$ in terms of $\rho$ and impose consistency conditions on a pair $(\pi, \rho)$ instead. Besides the REUM of Gul and Pesendofer [13] and the single-crossing RUM in Apesteigua, Ballester, and Lu [3], there are applications to random attention in Manzini and Marriotti [20], choices over state-contingent acts in Lu [16, 17], dynamic choices in Frick, Iijima, and Strzalecki [12], and various other settings.

Our model includes the additive RUM as a special case and hence, can be combined with the *pure characteristic* models (see, e.g. Berry and Pakes [5]). When taken to data, pure characteristic models are usually written as an average utility plus a random error. The special case where the error term follows the extreme type I distribution while the average utility has the logistic form, is equivalent to the Luce model [18].[2] The Luce model is the foundation of the discrete choice literature, and its variations such as the *random coefficient logit model* (aka the *mixed logit model*), are used in many demand estimation papers in empirical industrial organization. The Berry-Levison-Pakes estimator [4] is a classic method for estimating demand functions. Recent econometric literature (e.g. Shi, Shum, and Song [29], Khan, Ouyang, and Tamer [14]) provides many other estimation methods for the mixed logit model.

In contrast with the econometric literature, we focus on testable conditions for observable choice data that must hold whenever the noise distribution is invariant of the consumption menu. Thus we do not discuss the separation of the average utility and the error term for the RQUM.

## 2. Primitives and Functional Forms

Let $N = \{0, 1, \ldots, n\}$ 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 $[1, n] = \{1, \ldots, n\}$.

[2]This equivalence result is attributed to Holman and Marley in Luce [19], and is also shown in Yellot [32] and McFadden [21].

Let $X = \{x, y, \dots\}$ be the set $N \times \mathbb{R}$ of all pairs $(i, \alpha)$ that combine some consumption $i \in N$ with a monetary cost $\alpha \in \mathbb{R}$. If 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 $\Omega$ 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*.

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

$$\sum_{x \in A} \rho(x, A) = 1 \qquad \text{for all } A \in \mathcal{A}. \tag{4}$$

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.

2.1. **Quasi-Linear Orders.** 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 $x = y$.

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[3]

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

and such $v \in \mathbb{R}^n$ is determined uniquely by $R \in \mathcal{Q}$. This parametrization is analytically simpler than its counterparts for expected utility functions used by Gul and Pesendrofer and other authors. Their parameter spaces are topologically equivalent to a sphere, which is more complicated (e.g. for tie-breaking) than the Euclidean space $\mathbb{R}^n$.

2.2. **Main Representation.** Next, we adapt the random utility model for quasi-linear orders or equivalently, quasi-linear utility functions.

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 $y \in A \setminus x$.

[3] 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 $\mathcal{Q}$.

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

Let $\Pi = \{\pi, \dots\}$ be the set of all Borel probability measures on $\mathbb{R}^n$. A Borel probability measure $\pi \in \Pi$ is called a *regular* representation for a stochastic choice rule $\rho$ if for all trials $(x, A) \in \Omega$,

$$\begin{aligned} &\rho(x, A) = \pi(M(x, A)) \qquad \text{where} \\ &M(x, A) = \{v \in \mathbb{R}^n : x \text{ maximizes } R_v \text{ in } A\ \} \end{aligned} \tag{5}$$

In other words, the observed likelihood of any trial $(x, A)$ should equal the probability that the measure $\pi$ 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 (5) refines the general form (1) for $\Theta = \mathcal{Q}$.

For any $\pi \in \Pi$, representation (5) is consistent with the definition of a stochastic choice rule if and only if[4] for all $(x, A) \in \Omega$,

$$\begin{aligned} &\pi(M(x, A)) = \pi(S(x, A)) \qquad \text{where} \\ &S(x, A) = \{v \in \mathbb{R}^n : x \text{ strictly maximizes } R_v \text{ in } A\ \}. \end{aligned} \tag{6}$$

This condition requires that $\pi$ should assign a zero probability to quasi-linear utility ties. In particular, if $\pi$ has a finite support over $\mathbb{R}^n$, then (6) 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 random utility model 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 $A$.

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

$$\begin{aligned} &\rho(x, A) = \pi(L(x, A)) \qquad \text{where} \\ &L(x, A) = \{v \in \mathbb{R}^n : x \text{ is a low maximum for } R_v \text{ in } A\}. \end{aligned} \tag{7}$$

Representation (7) is well-defined for any $\pi \in \Pi$ and $A \in \mathcal{M}$ because for any vector $v \in \mathbb{R}^n$, the quasi-linear order $R_v$ has a unique low maximum in $A$ and hence,

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

[4]For any $(x, A) \in \Omega$, the set $M(x, A)$ contains $S(x, A)$. Thus (5) and (4) 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, (6). On the other hand, (5) and (6) 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 $\mathbb{R}^n$.

is a partition of the Euclidean space $\mathbb{R}^n$. Obviously, the low representation (7) implies the regular one (5) when $\pi$ satisfies (6). In this case, for all $(x, A) \in \Omega$,

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

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

We call representation (7) the *random quasi-linear utility model* (RQUM), and refer to the tie-breaking rule in (7) as the *grading procedure.*

2.3. **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 $\rho$, 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 $A(c)$. Thus the reduction $\rho^*$ restricts the stochastic choice rule $\rho$ to assortments.

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

$$\begin{aligned} \rho^*_k(c) &= \pi(L_k(c)) \qquad \text{where} \\ L_k(c) &= L\left((k, c_k), A(c)\right). \end{aligned} \tag{8}$$

This representation can be applied if $\rho^*$ is given as a primitive without $\rho$.

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 (8) for $\rho^*$ the *reduced* RQUM.

2.4. **Identification.** RQUM identifies the measure $\pi$ via a transparent formula.

For any vectors $w, v \in \mathbb{R}^n$, write $v \geqq w$ if $v_i \geq w_i$ for all $i \in [1, n]$. For any measure $\pi \in \Pi$, its *cumulative distribution function* (cdf) is defined as

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

It is easy to check that for any $v \in \mathbb{R}^n$,

$$\{w \in \mathbb{R}^n : v \geqq w\} = L((0, 0), A(0, v)) = L_0(0, v)$$

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

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

Thus the definition of cdfs and the reduced representation (8) imply that

$$F_\pi(v) = \rho_0^*(0, v). \tag{9}$$

For any vectors $v, w \in \mathbb{R}^n$ such that $v \geqq w$, formula (9) can be extended to the *rectangle*

$$(w, v] = \{r \in \mathbb{R}^n : w_i < r_i \leq v_i \quad \text{for all } i \in [1, n]\}$$

that consists of all vectors bounded by $v$ from above and strictly bounded by $w$ from below. Indeed, it is well-known (e.g. Billingsley [6, Section 12]) that

$$\pi((w, v]) = \sum_{K \subset [1,n]} (-1)^{|K|} F_\pi(wKv), \tag{10}$$

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

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

Obviously, the probability of any single vector $v \in \mathbb{R}^n$ can be found as

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

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

More generally, formula (10) establishes that $\pi$ is uniquely derived by its cdf for all Borel sets because the semiring of rectangles $(w, v]$ generates the entire Borel $\sigma$-algebra. Thus the probability measure $\pi$ is uniquely determined by $\rho$, or even by the component $\rho_0^*$ of the reduction $\rho^*$. Here the special role of the zeroth component $\rho_0^*$ is an artifact of the grading procedure where any tie between $(0, 0)$ and $(k, v_k)$ is broken in favor of the former alternative.

Finally, formula (9) can be rewritten in terms of choices in menus where the cost of good zero is arbitrary.

For any $\alpha \in \mathbb{R}$ and vector $v \in \mathbb{R}^n$, let $v + \alpha$ be the vector in $\mathbb{R}^n$ such that $(v + \alpha)_i = v_i + \alpha$ for all $i \in [1, n]$. Similarly, define $c + \alpha$ when $c \in \mathbb{R}^N$. Then for all $v \in \mathbb{R}^n$, RQUM implies that

$$F_\pi(v) = \rho_0^*(\alpha, v + \alpha) \tag{11}$$

because $L_0(0, v) = L_0(\alpha, v + \alpha)$.

## 3. Main Representation Results

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

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 [28]). The second part of NC requires that $x$ should be *never* 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 $\rho^*$ and rely on quasi-linearity in a more substantial way. The reduction $\rho^*$ 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 $\gamma$.

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

$$\rho^*_k(c) - \rho^*_k\left(c - \gamma\vec{j}\right) \geq \rho^*_j\left(c + \gamma\vec{k}\right) - \rho^*_j(c).$$

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

A function $F : \mathbb{R}^n \to [0,1]$ is called *jointly monotone* if for all vectors $v, w \in \mathbb{R}^n$,

$$v \geqq w \quad \Rightarrow \quad \sum_{K \subset [1,n]} (-1)^{|K|} F(wKv) \geq 0.$$

For any $\alpha \in \mathbb{R}$ and $v \in \mathbb{R}^n$, let

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

By (11), RQUM implies that each $G_\alpha : \mathbb{R}^n \to [0,1]$ should equal the cdf $F_\pi$ of some Borel probability measure $\pi$. As all cdfs must satisfy (10), then RQUM implies

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

For example, if $n = 2$ and $(v_1, v_2) \geqq (w_1, w_2)$, then JM requires that

$$G_\alpha(v_1, v_2) - G_\alpha(v_1, w_2) - G_\alpha(w_1, v_2) + G_\alpha(w_1, w_2) \geq 0.$$

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 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 good $k$ over other alternatives regardless of monetary costs.

Say that $\rho^*$ 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 $\gamma$ is constrained to be non-negative.

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

Here the special direction $(0, 1, \ldots, n)$ reflects the grading procedure.[5] The meaning of Axioms 1–5 and their logical independence are clarified further by several examples after our main result below.

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

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

*(i) uniquely identified by the reduction $\rho^*$ via* (11)*,*
*(ii) regular if and only if $\rho^*$ is continuous,*
*(iii) finite-ranged if and only if $\rho^*$ has a finite range.*

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

To prove that Axioms 1–5 are sufficient for the low representation (7), we proceed in three broad steps. First, we use JM, AC, and GC to construct a Borel probability measure $\pi \in \Pi$ with a cumulative distribution function such that

$$F_\pi(v) = G_0(v) = \rho^*_0(0, v)$$

for all $v \in \mathbb{R}^n$. The existence of such $\pi$ follows from Billingsley [6, Theorem 12.5].

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

$$\rho^*_k(c) = \pi(L_k(c)).$$

CPN is crucial here. This step takes the most effort in the proof of Theorem 1.

Third, we use NC to establish that the low representation (7) holds for all menus $A \in \mathcal{M}$ rather than just for assortments $A(c)$ with cost vectors $c \in \mathbb{R}^N$. All details are in the appendix.

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

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

**Corollary 2.** *A reduced scr $\rho^*$ satisfies Axioms 2–5 if and only if $\rho^*$ is represented by* (8) *for some $\pi \in \Pi$. Moreover, there is a unique stochastic choice rule $\rho$ that satisfies NC and has $\rho^*$ as its reduction.*

Here the identification (11) still applies, and the probability measure $\pi$ can be used as a low representation for the unique extension $\rho$. Similarly, the regular and finite-ranged cases can be characterized in terms of $\rho^*$ as well.

Theorem 1 can be refined further by imposing parametric structures on the distribution $\pi$ via the endogenous cdf $G_0$. To illustrate, suppose that for all $v, w \in \mathbb{R}^n_+$,

$$1 - G_0(v + w) = (1 - G_0(v))(1 - G_0(w)) < 1. \tag{13}$$

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

$$G_0(v) = \begin{cases} \prod_{i=1}^n (1 - \exp(-\lambda_i v_i)) & \text{for all } v \in \mathbb{R}^n_+ \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

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 (13). This specification is an example of the *exponomial* discrete choice model under the convention that the value $v_0$ is unperturbed.

### 3.1. RQUM vs the classic RUM.

Similarly to the classic RUM, RQUM can be characterized in terms of a Farkas-style condition.

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

$$\sum_{k=1}^m \rho(x_k, A_k) \le \max_{v \in \mathbb{R}^n} |\{k \in \{1, \ldots, m\} : x_k \textit{ strictly maximizes } R_v \textit{ in } A_k\}|. \tag{15}$$

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 $\rho$,*

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

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 the discussion section below).

Theorem 3 clarifies the connection of RQUM to the classic RUM. Obviously, ARSQ strengthens McFadden and Richter's[24]

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

$$\sum_{k=1}^{m} \rho(x_k, A_k) \leq \max_{total\ R \in \mathcal{R}} |\{k \in \{1, \ldots, m\} : x_k \ maximizes\ R\ in\ A_k\}|. \tag{16}$$

To see that (15) implies (16), 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 $(i, \alpha), (j, \beta) \in X$,

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

Recall that ARSP is equivalent to the non-negativity of Block-Marschak polynomials written for any finite menu $A \in \mathcal{M}$. This equivalence follows from the characterizations of the classic RUM by Falmagne [10] and McFadden and Richter [24] (see also Chambers and Echenique [8, Theorem 7.2].) Thus ARSQ implies that BM polynomials should be non-negative in any finite menu $A \subset X$.

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

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

Suppose that the stochastic choice rule $\rho$ 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 $\mathcal{F} \subset \mathcal{M}$ 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* $(x, A) \in \Omega(\mathcal{F})$,

$$\rho(x, A) = \pi(L(x, A)). \tag{17}$$

*Moreover, it is without loss in generality to take* $\pi$ *regular or finite-rangged.*

This result identifies $\pi$ via the Farkas Lemma. In contrast with Theorem 1, the identification of $\pi$ is not unique, and 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 relies on any permutation of $N$ to break ties. Indeed, ARSQ is invariant of 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.

## 4. Discussion

To clarify Theorem 1 further, it is useful to illustrate the logical independence of Axioms 1–5 and connect them to other conditions.

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 [28]). Note that Axioms 2–5 need not be violated by context effects because the reduction $\rho^*$ 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 $\rho$ that satisfies NC and has $\rho^*$ as its reduction.[6] So in these examples, it is enough to specify $\rho^*$.

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 $\rho_0^*$ 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 $\rho_0^*$ from 0 to 1, but decreasing the cost of good 0 by one unit does not change $\rho_1^*$ at all.

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 $\rho^*$ is determined by the difference $c_1 - c_0$. AC and GC are obvious. However, JM does not hold because $\rho_0^*$ is not increasing with respect to $c_1$. One can interpret $\rho^*$ 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

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

[6] 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.

for all $c \in \mathbb{R}^N$. Then Axioms 1–5 hold, except for AC that is obviously false. The reason is that $\rho^*$ 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 be also 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 obvious, 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 $\rho^+$ is generated by RQUM with one type $R_0$, and $\rho^-$ 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 $\rho^+$ and $\rho^-$ satisfy RQUM, then $\rho$ satisfies RQUM as well.

4.1. **Other Axioms.** Besides Axioms 1–5 in Theorem 1, RQUM implies other important conditions for stochastic choice rules. It is easy to conclude from (8) that the reduction $\rho^*$ 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 5.1 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.

Finally, suppose that the reduction $\rho^*$ is *continuously differentiable* and has all continuous partial derivatives up to order $n$. Then RQUM implies DZ's axioms, which assert that for all $c \in \mathbb{R}^N$ and distinct $k, j \in N$,

$$\frac{\partial \rho^*_k(c)}{\partial c_j} = \frac{\partial \rho^*_j(c)}{\partial c_k} \tag{18}$$

$$\frac{\partial^n \rho^*_k}{\partial c_0 \partial c_1 \ldots \partial c_{k-1} \partial c_{k+1} \ldots \partial c_n} \geq 0. \tag{19}$$

In this case, one can establish[7] that the combination of conditions (18)–(19) is equivalent to our CPN and JM.

Unlike the DZ conditions, both CPN and JM are written in terms of observed choice probabilities rather than their derivatives and put no resrictions on the

[7] This argument is available upon request.

Borel probability measure $\pi$ over the vectors $v \in \mathbb{R}^n$. Moreover, the DZ Theorem is stated in terms of the reduced scr and hence, does not apply in menus where some consumption goods are unavailable at any cost.

Finally, consider an example where ARSP holds, but ARSQ does not. Let $N = \{0, 1\}$, and

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

Let $\rho$ be the unique extension of $\rho^*$ 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, $\rho$ satisfies ARSP because $\rho$ selects the maximizer of a total order $R$ that is represented by a utility function

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

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

## 5. APPENDIX

Show Theorem 1.

Suppose that a Borel probability measure $\pi \in \Pi$ is a low representation (7) for a stochastic choice rule $\rho$. 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 (7) for $\rho$ implies (8) for its reduction $\rho^*$.

Show CPN. By (8),

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

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

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 $L_k(c)$. Moreover, (8) implies another set inclusion

$$\left[L_j\left(c + \gamma\vec{k}\right) \setminus L_j(c)\right] \subset \left[L_k(c) \setminus L_k\left(c - \gamma\vec{j}\right)\right]. \tag{20}$$

Indeed, 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 $A\left(c + \gamma\vec{k}\right)$. By (8), CPN follows from the set inclusion (20).

Show JM. Take any $\alpha \in \mathbb{R}$. By (11), the function

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

is the cdf of $\pi$. Thus every $G_\alpha$ is jointly monotone.

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

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

These sets are monotonically decreasing, $V_1 \supset V_2 \supset \ldots$, and satisfy $\bigcap_{k=1}^\infty V_k = \emptyset$. As $\pi$ 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 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 $v_0$ 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 $m = 1, 2, \ldots$, let

- $W_m$ be the set of all $v \in \mathbb{R}^n$ such that
$$\begin{aligned} &v_k - c_k > v_i - c_i + \tfrac{1}{m} && \text{for all } i \in N \text{ such that } i < k \\ &v_k - c_k \geq v_j - c_j && \text{for all } j \in N \text{ such that } j \geq k. \end{aligned}$$
- $W'_m$ be the set of all $v \in \mathbb{R}^n$ such that
$$\begin{aligned} &v_k - c_k > v_i - c_i && \text{for all } i \in N \text{ such that } i < k \\ &v_k - c_k \geq v_j - c_j - \tfrac{1}{m} && \text{for all } j \in N \text{ such that } j \geq k. \end{aligned}$$

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

As $L_k(c) = \bigcup_{m \to \infty} W_m = \bigcap_{m \to \infty} W'_m$ and $\pi$ 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 $d = (0, 1, \ldots, n)$. Then for all $0 < \gamma < \frac{1}{mn}$,

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

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

Suppose that $\pi$ satisfies the regularity condition (6). Show that $\rho^*$ 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 $\pi$ has finite range. Then it is obvious that $\rho^*$ has a finite range as well because each of its components $\rho^*_i$ has a finite range.

5.1. **Sufficiency of Axioms.** Suppose that $\rho$ 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, $G_\alpha$ is jointly monotone. By AC, for all $j \in [1, n]$,

$$\lim_{v_j\to-\infty} G_\alpha(v) = 0 \tag{21}$$

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

$$\lim_{\gamma\to+\infty} G_\alpha(\gamma, \ldots, \gamma) = 1 \tag{22}$$

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

$$\lim_{\gamma\to+\infty} \rho^*_k(\alpha, \gamma + \alpha, \gamma + \alpha, \ldots, \gamma + \alpha) = 0$$

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

Argue that $G_\alpha$ 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 $G_\alpha$ is jointly monotone. This contradicts the definition of $\varepsilon$. Thus $G_\alpha$ is weakly increasing with respect to all of its variables.

Show that $G_\alpha$ 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 $G_\alpha$ 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 $\varepsilon$ is arbitrary, then $G_\alpha$ is continuous from above.

Besides the joint monotonicity and continuity from above, the function $G_\alpha : \mathbb{R}^n \to [0, 1]$ satisfies the asymptotic normalizations (21)–(22). Billingsley's Theorem 12.5 implies that $G_\alpha$ is the cdf of some Borel measure $\pi_\alpha \in \Pi$.

The next lemma establishes that the functions $G_\alpha$—and hence, the corresponding measures $\pi_\alpha$—are invariant of $\alpha$.

**Lemma 5.1.** *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

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

Similarly,

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

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 $v \in \mathbb{R}^n$,

$$\left| \int_{w \in \mathbb{R}^n : v+\gamma \geqq w \geqq v} [G_\alpha(w) - G_\beta(w)] dw \right| \leq (\alpha - \beta)^2. \tag{23}$$

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

Before proving (23), observe that (23) 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 $G_\alpha$ 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 $\gamma$. However, (23) 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 (23). The proof invokes the Fubini theorem as in Billingsley [6, Theorem 18.3].

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

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

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

$$g(r,w)+(n-1)g(r,w+\varepsilon)-\sum_{k=1}^{n} g(r,w+\varepsilon-\varepsilon\vec{k}) \leq \sum_{i,j\in[1,n]:i>j} h(r,w) \tag{24}$$

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 (24) 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 (24) 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 (24) is $p-1$, and the right hand side is $\frac{p(p-1)}{2}$. Thus (23) must hold.

Conclude the proof of (23) by the Fubini theorem.

$$\begin{aligned}
&\int_{w\in\mathbb{R}^n: v+\gamma\geqq w\geqq v} [G_\alpha(w)-G_\beta(w)]dw \leq \\
&\qquad \int_{w\in\mathbb{R}^n: v+\gamma\geqq w\geqq v}\left[\int_{r\in\mathbb{R}^n}\sum_{i,j\in[1,n]:i>j} h(r,w)d\pi_\alpha(r)\right]dw = \quad \text{Fubini} \\
&\qquad\quad \int_{r\in\mathbb{R}^n}\left[\int_{w\in\mathbb{R}^n: v+\gamma\geqq w\geqq v}\sum_{i,j\in[1,n]:i>j} h(r,w)dw\right]d\pi_\alpha(r) \leq \\
&\qquad\qquad\qquad \int_{r\in\mathbb{R}^n}\frac{n(n-1)}{2}\gamma^{n-2}\varepsilon^2 d\pi(r) \leq \varepsilon^2.
\end{aligned}$$

because 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 $\beta < \alpha$. $\square$

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 5.1,

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

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

$$\rho_k^*(c) \geq \pi(S_k(c)) \tag{25}$$

*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 $t = 1, 2, \dots,$ 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:

$$V_1 \subset V_2 \subset V_3 \subset \dots$$

Take any $v \in V_t$. Then there is $m \in \{0, 1, \dots, 4^t - 1\}$ 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 $v \in L_0 \left( c + \tfrac{1}{2^{t+1}}\vec{k} - \tfrac{2m}{2^{t+1}}\vec{0} \right)$, 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 $v \notin L_0 \left( c + \tfrac{1}{2^{t+1}}\vec{k} - \tfrac{2m}{2^{t+1}}\vec{0} \right)$, then $v \in V_{t+1}$ because

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

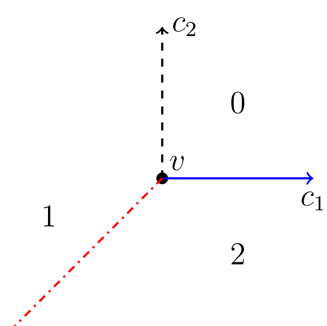
Next, show that the union $\bigcup_{t=1}^{\infty} V_t$ contains $S_k(c)$. Suppose that $v \in \mathbb{R}^n$ is such that $(k, c_k)$ is a strict maximum for $R_v$ in $A(c)$. Then $v_k - c_k > -c_0$. Take $t$ such that

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

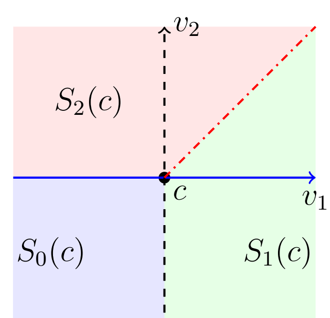
Pick $m \in \{0, \dots, 4^t - 1\}$ such that

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

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, $v \in V_m$.




(A) Type $v$'s choice of $0, 1, 2$ partitions the space $\mathbb{R}^2$




(B) Partition of the type space $\mathbb{R}^2$ according to choices at cost $c$. Green types choose good 1, red types choose good 2, and blue types choose the status quo 0.

FIGURE 1. Partition by $v$ and by $c$, where $v_0 = 0, c_0 = 0$ are fixed

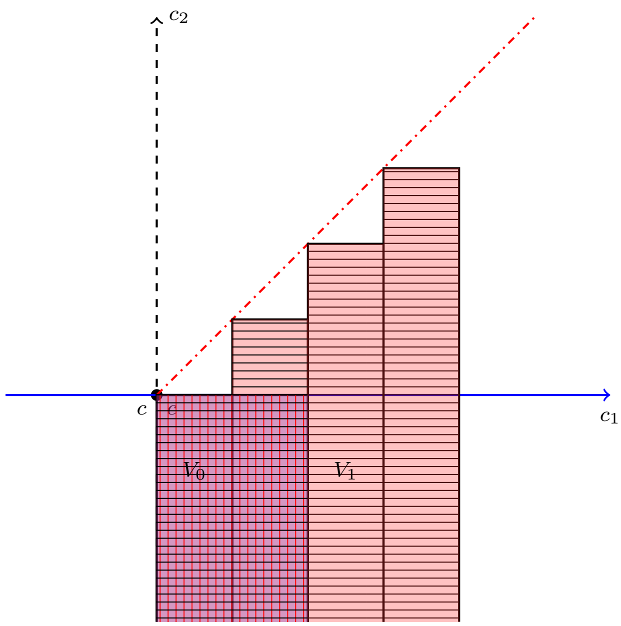



FIGURE 2. Fix $c = 0$ and $k = 1$. Then $V_0$ is the area shaded by vertical lines; $V_1$ is the area shaded by vertical lines. Here the width of $V_0$ is one monetary unit, and the width of $V_1$ is two monetary units.

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

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

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

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 $\pi$ is countably additive.

This proof can be illustrated in Figure 1. WLOG let $N = \{0, 1, 2\}$. Fix $v_0 = 0, c_0 = 0$, and let each vector in $(c_1, c_2) \in \mathbb{R}^2$ represent the the costs on goods 1 and 2, respectively. For simplicity, consider regular distribution of types. Figure 1(A) shows that if the cost vector falls in regions $0, 1$, or 2, then type $v$ chooses $0, 1$, or 2, respectively. Figure 1(B) shows that a given cost $c = (c_1, c_2)$ partitions the $\mathbb{R}^2$ space into $S_0(c), S_1(c)$ and $S_2(c)$. Figure 2 illustrates $V_t$ for $t = 0, 1$.

□

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

**Lemma 5.3.** *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 $\pi$ 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 5.3. For all $k \in N$ and $t = 1, 2, \ldots,$ Lemma 5.2 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 $\rho^{**}$ be the reduced scr that has $\pi$ as a low representation. Both $\rho^*$ and $\rho^{**}$ satisfy GC. Thus

$$\rho^*(c) = \lim_{t\to\infty} \rho^*(c + \gamma_t d)) = \lim_{t\to\infty} \rho^{**}(c + \gamma_t d)) = \rho^{**}(c). \tag{26}$$

Thus $\pi$ is a low representation for $\rho^*$.

Extend this representation for the entire $\rho$. 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 $t = 1, 2, \ldots,$ define a cost vector $c(t) \in \mathbb{R}^N$ as

$$c_i(t) = \begin{cases} \min\{\alpha \in \mathbb{R} : (i, \alpha) \in B\} & \text{if } (i, \alpha) \in B \text{ for some } \alpha \in \mathbb{R} \\ t & \text{if } (i, \alpha) \notin B \text{ for all } \alpha \in \mathbb{R}. \end{cases} \tag{27}$$

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 $\pi$ is a low representation for $\rho$.

If $\rho^*$ is continuous, then $\pi$ must satisfy the regularity condition (6). 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 (27). 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 $\rho_i^*$ 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 (6) must hold.

Finally, suppose that $\rho^*$ has a finite range. Show that the support of $\pi$ 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 $F_\pi$ and hence, belongs to the closure of the finite range of $\rho_0^*$. The closure of a finite set is the same finite set. Thus $F_i$ 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 (26), which was derived above from Axioms 2–5 without NC.

## 6. Proof of Theorem 3

Suppose that $\rho$ satisfies Axioms 1–5. By Theorem 1, $\rho$ 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 (7) 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,

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

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 $\rho$ satisfies ARSQ because

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

Show next that ARSQ implies Axioms 1–3.

**Lemma 6.1.** *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 $v \in \mathbb{R}^n$,*

$$\sum_{k=1}^{m} l_v(x_k, A_k) \leq \sum_{i=1}^{t} l_v(y_i, B_i), \tag{28}$$

*then ARSQ implies that*

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

*Proof.* Inequality (28) 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 $i = 1, \dots, t$. 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 (29), $\rho(x, A\cup y) \le \rho(x, A)$. 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 $v \in \mathbb{R}^n$,

$$l_v\left((k, c_k), A\left((c - \gamma\vec{j}\right)\right) + l_v\left((j, c_j), A\left(c + \gamma\vec{k}\right)\right) \le l_v((k, c_k), A(c)) + l_v((j, c_j), A(c)). \tag{30}$$

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 (30) 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 (30) must hold.

Suppose that $l_v\left((j, c_j), A\left(c + \gamma\vec{k}\right)\right) = 1$. Similarly, 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, (30) must hold.

CPN follows from (30) and Lemma 6.1.

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

$$S = \sum_{K\subset[1,n]} (-1)^{|K|} l_v((0, \alpha), A(\alpha, wKr)) \ge 0. \tag{31}$$

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 $S = l_v((0, \alpha), A(\alpha, r)) \ge 0$.

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 (31),

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

By (29),

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

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 $G_\alpha$,

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

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

## 7. Proof of Theorem 4

Take a stochastic dataset $\rho : \Omega(\mathcal{F}) \to [0, 1]$. Low representation (17) implies ARSQ by the same argument as for stochastic choice rules.

Suppose instead that $\rho$ satisfies ARSQ. Then Lemma 6.1 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 [8, 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

$$\sum_{(x,A) \in \Omega(\mathcal{F})} z(x, A) l_w(x, A) = 0 \tag{32}$$

for all $w \in W$, but

$$\sum_{(x,A) \in \Omega(\mathcal{F})} z(x, A) \rho(x, A) < 0. \tag{33}$$

It follows from (32) 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 6.1,

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

which contradicts (33).

Thus Case 1 must hold.

Show that the identification of $\pi$ 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 $\pi$ by a continuous distribution $\sigma = \sum_{w \in W} \pi(w) \delta_w$ where $\delta_w \in \Pi$ has density

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

where $\lambda(B)$ is the Lebesgue volume of the neighborhood $B_w$. Then the dataset $\rho$ has $\sigma$ 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] S. Berry and A. Pakes. The pure characteristics demand model. *International Economic Review*, 48(4):1193–1225, 2007.
[6] P. Billingsley. *Probability and Measure.* Wiley, New York, third edition, 1995.
[7] 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.
[8] C. P. Chambers and F. Echenique. *Revealed preference theory*, volume 56. Cambridge University Press, 2016.
[9] 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.
[10] J.-C. Falmagne. On a class of probabilistic conjoint measurement models: Some diagnostic properties. *Journal of Mathematical Psychology*, 19(2):73–88, 1979.
[11] M. Fosgerau, D. McFadden, and M. Bierlaire. Choice probability generating functions. *Journal of Choice Modelling*, 8:1–18, 2013.
[12] M. Frick, R. Iijima, and T. Strzalecki. Dynamic random utility. *Econometrica*, 87(6):1941–2002, 2019.
[13] F. Gul and W. Pesendorfer. Random expected utility. *Econometrica*, 74(1):121–146, 2006.
[14] S. Khan, F. Ouyang, and E. Tamer. Inference on semiparametric multinomial response models. *Quantitative Economics*, 12:743—-777, 2021.
[15] R. H. Koning and G. Ridder. Discrete choice and stochastic utility maximization. *The Econometrics Journal*, 6(1):1–27, 2003.
[16] J. Lu. Random choice and private information. *Econometrica*, 84(6):1983–2027, 2016.
[17] J. Lu. Random ambiguity. *Theoretical Economics*, 16(2):539–570, 2021.
[18] R. D. Luce. Individual choice behavior, 1959.
[19] 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.
[20] P. Manzini and M. Mariotti. Stochastic choice and consideration sets. *Econometrica*, 82(3):1153–1176, 2014.
[21] D. McFadden. Conditional logit analysis of qualitative choice behavior. pages 105–142. Academic Press, New York, 1973.
[22] D. McFadden. Econometric models for probabilistic choice among products. *The Journal of Business*, 53:13–29, 1980.
[23] D. McFadden. Economic choices. *American Economic Review*, 91(3):351–378, 2001.
[24] 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.
[25] E. Piermont. Disentangling strict and weak choice in random expected utility models. *Journal of Economic Theory*, 1202:1–34, 2022.
[26] 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.

[27] J. Rust. Optimal replacement of GMC bus engines: An empirical model of harold zurcher. *Econometrica*, 55:999–1033, 1987.
[28] E. Shafir, I. Simonson, and A. Tversky. Reason-based choice. *Cognition*, 49:11—-36, 1993.
[29] X. Shi, M. Shum, and W. Song. Estimating semi-parametric panel multinomial choice models using cyclic monotonicity. *Econometrica*, 86:737–761, 2018.
[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] 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.