DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

Axiomatizations of the Mixed Logit Model

Kota Saito
California Institute of Technology

SOCIAL SCIENCE WORKING PAPER 1433

September 2017

Axiomatizations of the Mixed Logit Model

Kota Saito*
California Institute of Technology

September 15, 2017

Abstract

A mixed logit function, also known as a random-coefficients logit function, is an integral of logit functions. The mixed logit model is one of the most widely used models in the analysis of discrete choice. Observed behavior is described by a random choice function, which associates with each choice set a probability measure over the choice set. I obtain several necessary and sufficient conditions under which a random choice function becomes a mixed logit function. One condition is easy to interpret and another condition is easy to test.

Keywords: Random utility, random choice, mixed logit, random coefficients.

1 Introduction

The purpose of this paper is to provide axiomatizations of the mixed logit model, also known as the random-coefficients logit model. The mixed logit model is one of the most widely used models in the analysis of discrete choice, especially in the empirical literature on marketing, industrial organization, and public economics. I provide several axiomatizations of the mixed logit model. One axiomatization is

*This paper was first presented at the University of Tokyo on July 29, 2017. I appreciate the valuable discussions I had with Kim Border, Federico Echenique, Hidehiko Ichimura, Yimeng Li, Jay Lu, and Matt Shum. Jay Lu also read the manuscript and offered helpful comments. This research is supported by Grant SES1558757 from the National Science Foundation.

useful to understand the behavioral implications of the mixed logit model. Another axiomatization is useful to test the mixed logit model.

In this paper, the observed behavior is described by a random choice function ρ that assigns to each choice set D a probability distribution over D. The number $\rho(D, x)$ is the probability that an alternative x is chosen from a choice set D. The function ρ is called a *mixed logit function* if there exists a probability measure m such that

$$\rho(D, x) = \int \frac{\exp(u(x))}{\sum_{y \in D} \exp(u(y))} dm(u). \tag{1}$$

The mixed logit model has been popular for several reasons. To begin with, the model overcomes the limitations of the logit model. The mixed logit model allows various substitution patterns across the alternatives. Moreover, despite its specific formula, the model is flexible. In fact, McFadden and Train (2000) show that any random utility function can be approximated by a mixed logit function.

In an empirical analysis, an alternative x can be identified by the real vector of explanatory variables of x.[1] With the vector x of explanatory variables, an empirical researcher usually uses a special case of a mixed logit function in which u takes the linear form of $u(x) = x \cdot \beta$. I call a logit function with such a linear u a *linear logit function*. I call the special case of a mixed logit function a *mixed linear logit function*.

I provide several axiomatizations of the mixed logit model. Each axiom by itself is necessary and sufficient for the mixed logit model. To motivate the first axiomatization, consider an expected-utility maximizer who chooses an alternative from a choice set without knowing his true utility function. The choice set will be randomly chosen and the agent has a subjective belief over the choice sets. One simple strategy of the agent is to pick a deterministic strict preference relation and to maximize the strict preference relation. This strategy is *naive* because it ignores the possibility that the agent's utility could be different across the choice sets.

The first axiom requires that for any subjective belief over the choice sets and for any nonconstant realization of utility function, the agent's random choice should give a higher expected utility than the worst naive strategy. Notice that the requirement of the axiom is weak in that the axiom does *not* require that the agent's random choice dominate the naive strategies; the axiom only requires that the agent's ran-

[1] For example, in Berry et al. (1995), an alternative is a car available in the market. Each car is identified by the car's price, weight, size, fuel efficiency, and other attributes.

dom choice should be better than the *worst* naive strategy. In Theorem 1, I show that *a random choice function satisfies the axiom if and only if it is a mixed logit function.*

The second axiomatization is based on the Block-Marschak polynomials. Falmagne (1978) has shown that the nonnegativity of the Block-Marschak polynomials characterizes the random utility model. In Theorem 2, I show that *the positivity of the Block-Marschak polynomials characterizes the mixed logit model.* The number of the Block-Marschak polynomials is finite. Thus it is easy to test this second axiom, although the behavioral meaning of the second axiom may be less clear than the meaning of the first axiom. McFadden and Richter (1990) and Clark (1996) have provided other axiomatizations of the random utility model. By modifying their axioms, I obtain alternative axiomatizations of the mixed logit model in the appendix.

Moreover, I provide the axiomatizations of the mixed *linear* logit model. As mentioned earlier, empirical researchers usually use the mixed linear logit model, not the mixed logit model.[2] I show that the same axioms described above respectively characterize the mixed *linear* logit model if the set of explanatory variables of the alternatives is affinely independent.

By the way of proving the axiomatizations described above, I have obtained several remarks. Remark 1 states that if the set of explanatory variables of the alternatives is affinely independent, then (i) any interior random utility function can be *represented* as a convex combination of linear logit functions; (ii) any noninterior random utility function can be *approximated* by a convex combination of linear logit functions.

Remark 1 is related with Theorem 1 of McFadden and Train (2000). As mentioned earlier, their result has contributed to the popularity of the mixed logit model. There is, however, one limitation in Theorem 1 of McFadden and Train (2000). They say "One limitation of Theorem 1 is that it provides no practical indication of how to choose parsimonious mixing families, or how many terms are needed to obtain acceptable approximations..." (p. 452)

Remark 1 overcomes this limitation, although the set up of McFadden and Train (2000) is more general than mine. They do not state how one can construct the vectors of polynomials, which can contain arbitrarily higher degree terms. In contrast, in Remark 1, it is not necessary to construct the polynomials; instead it is enough

[2] In fact, in the empirical literature, the mixed linear loigit model is often called the mixed logit model.

to construct a convex combination of linear logit functions. The construction of the convex combination is simple as shown in Remark 2. Furthermore, statement (i) of Remark 1 claims the exact equality, not an approximation, for the case of an interior random utility function.

In the next section, I introduce the models formally. In section 3, I show the axiomatizations of the mixed logit model. Then in section 4, I show the axiomatizations of the mixed linear logit model. In section 5, I state the remarks to conclude the paper.

2 Model

Let X be a finite set. X is the set of outcomes. Let $\mathcal{D} \equiv 2^X \setminus \{\emptyset\}$.

Definition 1. *A function $\rho : \mathcal{D} \times X \to [0,1]$ is called a* random choice function *if $\sum_{x \in D} \rho(D,x) = 1$ and $\rho(D,x) = 0$ for any $x \notin D$. The set of random choice functions is denoted by $\mathcal{P}$.*

For each $(D,x) \in \mathcal{D} \times X$, the number $\rho(D,x)$ is the probability that an alternative x is chosen from a choice set D. A random choice function ρ is an element of $\mathbf{R}^{\mathcal{D} \times X}$.

Let Π be the set of bijections between $X \to \{1, \dots, |X|\}$, where $|X|$ is the number of elements of X. If $\pi(x) = k$, I interpret x to be the $|X| + 1 - k$-th best element of X with respect to π. If $\pi(x) > \pi(y)$, then x is better than y with respect to π. An element π of Π is called a *strict preference ranking* (or simply, a *ranking*) over X.

For all $(D,x) \in \mathcal{D} \times X$, if $\pi(x) > \pi(y)$ for all $y \in D \setminus \{x\}$, then I often write $\pi(x) \geq \pi(D)$. There are $|X|!$ elements in Π. I denote the set of probability measures over Π by $\Delta(\Pi)$. Since Π is finite, $\Delta(\Pi) = \left\{(\nu_1, \dots, \nu_{|\Pi|}) \in \mathbf{R}_+^{|\Pi|} \middle| \sum_{i=1}^{|\Pi|} \nu_i = 1\right\}$, where $\mathbf{R}_+$ is the set of nonnegative real numbers.

Definition 2. *A random choice function ρ is called a* random utility function *if there exists a probability measure $\nu \in \Delta(\Pi)$ such that for all $(D,x) \in \mathcal{D} \times X$,*

$$\rho(D,x) = \nu(\pi \in \Pi | \pi(x) \geq \pi(D)).$$

The probability measure ν is said to rationalize *ρ. The set of random utility functions is denoted by $\mathcal{P}_r$.*

A random utility function is a probability distribution over the strict preference rankings over X.[3]

Definition 3. *A random choice function ρ is called a* logit function *if there exists a function $u : X \to \mathbf{R}$ such that for all $(D, x) \in \mathcal{D} \times X$,*

$$\rho(D, x) = \frac{\exp(u(x))}{\sum_{y \in D} \exp(u(y))}.$$

The set of logit functions is denoted by $\mathcal{P}_l$.

In a logit function, u is an element of $\mathbf{R}^{|X|}$. Let $\mathcal{B}^{|X|}$ denote the Borel algebra of $\mathbf{R}^{|X|}$ and consider a measurable space $(\mathbf{R}^{|X|}, \mathcal{B}^{|X|})$. I denote the set of probability measures over $\mathbf{R}^{|X|}$ by $\Delta(\mathbf{R}^{|X|})$.

Definition 4. *A random choice function ρ is called a* mixed logit function *if there exists a probability measure $m \in \Delta(\mathbf{R}^{|X|})$ such that for all $(D, x) \in \mathcal{D} \times X$,*

$$\rho(D, x) = \int \frac{\exp(u(x))}{\sum_{y \in D} \exp(u(y))} dm(u). \tag{2}$$

The set of logit functions is denoted by $\mathcal{P}_{ml}$.

The integral is well defined because $f^{(D,x)}(u) \equiv \exp(u(x)) / \sum_{y \in D} \exp(u(y))$ is measurable with respect to $\mathcal{B}^{|X|}$ for each $(D, x) \in \mathcal{D} \times X$.[4]

In the empirical literature, for each alternative x of X, there is a vector of explanatory variables for the alternative x. For example, as mentioned earlier, in Berry et al. (1995), X consists of cars available on the market. Then each car $x \in X$ is described by its price, weight, size, fuel efficiency, and other attributes. The vectors of explanatory variables are usually different across the alternatives. So one can identify each alternative x by the vector of explanatory variables for x. Proceeding in this way, in some parts of this paper I assume that the set X is a finite subset of k-dimensional real space (where k is the number of explanatory variables).

[3] While the function above is often called a random ranking function, a random utility function is often defined differently–by using the existence of a probability measure μ over utilities such that for all $(D, x) \in \mathcal{D} \times X$, $\rho(D, x) = \mu(u \in \mathbf{R}^{|X|} | u(x) \geq u(D))$. Block and Marschak (1960)(Theorem 3.1) state that the two definitions are equivalent.

[4] The formula can be written as $\int f^{(D,x)}(u) dm(u)$. Since the function $f^{(D,x)}$ is continuous in u, the function $f^{(D,x)}$ is measurable with respect to $\mathcal{B}^{|X|}$. Moreover, since $f^{(D,x)}(u) \in (0, 1)$, the function $f^{(D,x)}$ is bounded and nonnegative and hence integrable.

Definition 5. *Let X be a finite subset of $\mathbf{R}^k$. A random choice function ρ is called a* linear logit function *if there exists $\beta \in \mathbf{R}^k$ such that for all $(D, x) \in \mathcal{D} \times X$,*

$$\rho(D, x) = \frac{\exp(\beta \cdot x)}{\sum_{y \in D} \exp(\beta \cdot y)}.$$

The set of linear logit functions is denoted by $\mathcal{P}_{ll}$.

The next model is a special case of the mixed logit model. To define the model, let $\mathcal{B}^k$ be the product Borel algebra of $\mathbf{R}^k$ and consider a measurable space $(\mathbf{R}^k, \mathcal{B}^k)$. I denote the set of probability measures over $\mathbf{R}^k$ by $\Delta(\mathbf{R}^k)$.

Definition 6. *Let X be a finite subset of $\mathbf{R}^k$. A random choice function ρ is called a* mixed linear logit function *if there exists a probability measure $m \in \Delta(\mathbf{R}^k)$ such that for all $(D, x) \in \mathcal{D} \times X$,*

$$\rho(D, x) = \int \frac{\exp(\beta \cdot x)}{\sum_{y \in D} \exp(\beta \cdot y)} dm(\beta). \tag{3}$$

The set of mixed linear logit functions is denoted by $\mathcal{P}_{mll}$.

A mixed linear logit function is sometimes called a *latent class function* if m has a finite support. A latent class function is a convex combination of linear logit functions.

In the empirical literature, the mixed linear logit model and the latent class model are sometimes treated as competing models. For example, Greene and Hensher (2003) claim that the performance of the latent class model is better than that of the mixed logit model.[5] The following proposition (statement (ii)) states, however, that the two models are equivalent.

Proposition 1. *For any random choice function ρ,*
(i) the function ρ is a mixed logit function if and only if ρ is a convex combination of logit functions (i.e., $\mathcal{P}_{ml} = co.\mathcal{P}_l$),
(ii) the function ρ is a mixed linear logit function if and only if ρ is a convex combination of linear logit functions (i.e., $\mathcal{P}_{mll} = co.\mathcal{P}_{ll}$).

[5] Greene and Hensher (2003) (p.698) state "Which model is superior on all behavioral measures of performance is inconclusive despite stronger statistical support overall for the latent class model (on this occasion). The inconclusiveness is an encouraging result since it motivates further research involving more than one specification of the choice process."

Statement (i) implies that for any mixed logit function, one can find an observationally equivalent convex combination of logit functions. Thus to axiomatize the mixed logit model it is necessary and sufficient to axiomatize the convex hull of logit functions. Statement (ii) implies that the same observations hold for a mixed *linear* logit function.

The mixed logit model has been known for a long time, but has become popular relatively recently since the development of simulation method. This is because calculating the integration used to be difficult. The proposition states that focusing on a convex combination of logit functions entails no loss of generality. Hence, the calculation of the integration is not necessary.[6]

Finally, I introduce essential mathematical concepts. A *polyhedron* is an intersection of finitely many closed half spaces. A *polytope* is a bounded polyhedron. Equivalently, a polytope is a convex hull of finitely many points.

The convex hull of a set C is denoted by co.C. The closure of a set C is denoted by cl.C. The *affine hull* of a set C is the smallest affine set that contains C; and it is denoted by aff.C.

The *relative interior* of a convex set C is an interior of C in the relative topology with respect to aff.C. The relative interior of C is denoted by rint.C. If C is not empty, then (i) rint.C is not empty, and (ii) rint.$C = \{x \in C | \text{for all } y \in C \text{ there exists } \alpha > 1 \text{ such that } \alpha x + (1-\alpha)y \in C\}$. (See Theorem 6.4 in Rockafellar (2015) for the proof.)

3 Axiomatization of the Mixed Logit Model

In this section, I provide two axiomatizations of the mixed logit model. First, I prove two propositions which are necessary for the axiomatization. The first proposition proves that the mixed logit model is the interior of the random utility model.

Proposition 2. *The set of mixed logit functions is the relative interior of the set of random utility functions. That is,* $\mathcal{P}_{ml} = rint.\mathcal{P}_r$.

[6]The nested logit model also can be seen as a convex combination of the logit model when the nests do not overlap. Gul et al. (2014) axiomatize a model called *the complete attribute rule*, which is similar to the nested logit model. Neither the complete attribute rule nor the mixed logit model is more general than the other. The intersection between the two models is the (degenerate) logit model. See appendix B for details.

The next proposition characterizes the affine hull of the set $\mathcal{P}_r$ of random utility functions.

Proposition 3. *The affine hull of $\mathcal{P}_r$ is*

$$\Big\{p \in \mathbf{R}^{\mathcal{D}\times X} \Big| \textit{(i)} \sum_{x\in D} p(D,x) = 1 \textit{ for any } D \in \mathcal{D},\ \textit{(ii)}\ p(D,x) = 0 \textit{ for any } D \in \mathcal{D}, x \notin D\Big\}.$$

Hence, $\dim \mathcal{P}_r = (|X| - 2)2^{|X|-1} + 1$*, where $|X|$ is the number of elements in X.*

The first statement of Proposition 3 implies that the set of random choice function is contained by the affine hull of the set of random utility functions (i.e., $\mathcal{P} \subset \text{aff.}\mathcal{P}_r$). I will use this implication to obtain the axiomatizations below.[7] The second statement of Proposition 3 on the dimension of $\mathcal{P}_r$ will be used to discuss the identification of the mixed logit model in section 5.

3.1 Axiomatization based on Expected Utility

For each strict preference ranking $\pi \in \Pi$, define

$$\rho^\pi(D,x) = \begin{cases} 1 & \text{if } \pi(x) \geq \pi(D), \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

The function ρ^π is a deterministic random choice function, which gives probability one to the best alternative x in a choice set D according to the strict preference ranking π.

To motivate the first axiomatization, consider an agent who chooses an element from a choice set $D \in \mathcal{D}$ without knowing his true utility function. The choice set will be randomly chosen, and let $q(D)$ be the agent's subjective probability that his choice set will be D. Let $u(D,x)$ be the utility when the agent chooses x from D. If the agent's choice is described by a random choice function ρ, then his expected utility is

$$E(\rho : q, u) = \sum_{D\in\mathcal{D}} q(D) \sum_{x\in D} \rho(D,x)u(D,x).$$

[7] Another nontrivial implication of the result is that for any random choice function ρ, there exist a real number α and a pair (ρ_1, ρ_2) of random utility functions such that $\rho = \alpha\rho_1 + (1-\alpha)\rho_2$. To see the implication, notice that for any $\rho \in \mathcal{P}$, there exist $\{\lambda_i\}_{i=1}^n \subset \mathbf{R}$ and $\{\rho'_i\}_{i=1}^n \subset \mathcal{P}_r$ such that $\rho = \sum_{i=1}^n \lambda_i \rho'_i$ and $\sum_{i=1}^n \lambda_i = 1$. Define $\alpha = \sum_{i:\lambda_i>0} \lambda_i$ and $\beta = \sum_{i:\lambda_i<0} \lambda_i$. Then, $\alpha + \beta = 1$. Define $\rho_1 = \sum_{i:\lambda_i>0}(\lambda_i/\alpha)\rho'_i$ and $\rho_2 = \sum_{i:\lambda_i<0}(-\lambda_i/-\beta)\rho'_i$. Then, $\rho_1, \rho_2 \in \mathcal{P}_r$. It follows that $\rho = \sum_{i=1}^n \lambda_i\rho'_i = \alpha\rho_1 + \beta\rho_2 = \alpha\rho_1 + (1-\alpha)\rho_2$. I wish to acknowledge Jay Lu for the discussion that led to this remark.

One simple strategy of the agent is to pick a deterministic strict preference ranking π arbitrarily and maximize the strict preference ranking. Then his choice is described by ρ^π, as defined by (4). This strategy is *naive* because it ignores the possibility that the agent's utility could be different across the choice sets.

The following axiom requires that for any subjective belief q over the choice sets and for any (nonconstant) realization $u(D,\cdot)$ of the utility function, the agent's random choice should give a higher expected utility than the worst naive strategy. As mentioned earlier, the requirement of the axiom is weak in that the axiom does *not* require that the agent's random choice dominate the naive strategies; the axiom only requires that the agent's random choice should be better than the *worst* naive strategy.

Axiom 1. *(Quasi-Stochastic Rationality) For any $q \in \Delta(\mathcal{D})$ and any $u(D,\cdot) \in \mathbf{R}^D$ for each $D \in \mathcal{D}$, if $u(D,\cdot)$ is not constant for some D with $q(D) > 0$, then*

$$E(\rho : q, u) > \min_{\pi \in \Pi} E(\rho^\pi : q, u). \tag{5}$$

In the axiom, notice that the set Π is finite, so $\min_{\pi\in\Pi} E(\rho^\pi : q, u)$ exists for any u, q, and $\pi \in \Pi$. Notice also that if $u(D,\cdot)$ is constant for all D with $q(D) > 0$, then the expected utility is also constant for any random choice function.

Theorem 1. *A random choice function ρ satisfies Quasi-Stochastic Rationality if and only if ρ is a mixed logit function.*

The sufficiency part of the proof can be sketched as follows. It can be shown that the set $\mathcal{P}_r$ of random utility functions is a polytope. That is, $\mathcal{P}_r = \text{co.}\{\rho^\pi | \pi \in \Pi\}$. Moreover, it follows that there exist a set $\{t_i\}_{i=1}^n \subset \mathbf{R}^{\mathcal{D}\times X} \setminus \{0\}$ and a set $\{\alpha_i\}_{i=1}^n \subset \mathbf{R}$ such that

$$\mathcal{P}_r = \cap_{i=1}^n \{p \in \mathbf{R}^{\mathcal{D}\times X} | p \cdot t_i \geq \alpha_i\} \cap \text{aff.}\mathcal{P}_r. \tag{6}$$

As mentioned earlier, Proposition 3 implies that $\mathcal{P}_r \subset \mathcal{P} \subset \text{aff.}\mathcal{P}_r$. This implication and (6) show that $\mathcal{P}_r = \cap_{i=1}^n \{\rho \in \mathcal{P} | \rho \cdot t_i \geq \alpha_i\}$. It follows that $\text{rint.}\mathcal{P}_r = \cap_{i=1}^n \{\rho \in \mathcal{P} | \rho \cdot t_i > \alpha_i\}$. Since Proposition 2 states that $\mathcal{P}_{ml} = \text{rint.}\mathcal{P}_r$, I obtain $\mathcal{P}_{ml} = \cap_{i=1}^n \{\rho \in \mathcal{P} | \rho \cdot t_i > \alpha_i\}$.

For each $i \in \{1, \dots, n\}$, I can find a utility vector u_i and a belief q_i such that $\rho \cdot t_i > \alpha_i$ if and only if $E(\rho : q_i, u_i) > \alpha_i / |\mathcal{D}|$. Therefore, $\mathcal{P}_r = \cap_{i=1}^n \{\rho \in \mathcal{P} | E(\rho : q_i, u_i) \geq \alpha_i / |\mathcal{D}|\}$ and $\mathcal{P}_{ml} = \cap_{i=1}^n \{\rho \in \mathcal{P} | E(\rho : q_i, u_i) > \alpha_i / |\mathcal{D}|\}$. Since $\rho^\pi \in \mathcal{P}_r$ for any $\pi \in \Pi$, it follows that $E(\rho^\pi : q_i, u_i) \geq \alpha_i/|\mathcal{D}|$ for all $i \in \{1, \ldots, n\}$. Hence, Quasi-Stochastic Rationality implies that $E(\rho : q_i, u_i) > \alpha_i/|\mathcal{D}|$ for all $i \in \{1, \ldots, n\}$. So, $\rho \in \cap_{i=1}^n \{\rho \in \mathcal{P} | E(\rho : q_i, u_i) > \alpha_i/|\mathcal{D}|\} = \mathcal{P}_{ml}$. See the appendix for the concrete proof.[8]

3.2 Axiomatization by the Block-Marschak Polynomials

In this section, I provide an alternative axiomatization of the mixed logit model based on a finite number of polynomials called the Block-Marschak polynomials.

Definition 7. *(Block-Marschak polynomials) For any random choice function ρ and $(D, x) \in \mathcal{D} \times X$ such that $x \in D$, define*

$$K(\rho, D, x) = \sum_{E : D \subset E} (-1)^{|E \setminus D|} \rho(E, x).$$

Block and Marschak (1960) have shown that if ρ is a random utility function, then $K(\rho, D, x) \geq 0$ for any $\rho \in \mathcal{P}$ and any $(D, x) \in \mathcal{D} \times X$ such that $x \in D$. Falmagne (1978) has shown the converse.

The next theorem states that the positivity of the Block-Marschak polynomials characterizes the mixed logit model.

Theorem 2. *A random choice function ρ is a mixed logit function if and only if $K(\rho, D, x) > 0$ for any $(D, x) \in \mathcal{D} \times X$ such that $x \in D$.*

Notice that there are only finitely many pairs $(D, x) \in \mathcal{D} \times X$ such that $x \in D$. So it is easy to test this axiom. This is the benefit of this second axiomatization, although the behavioral meaning of this axiom may not be clear.

The sufficiency part of the proof can be sketched as follows. Fix a random choice function ρ and assume that the Block-Marschak polynomials of ρ are strictly positive. I will show that ρ belongs to the set $\mathcal{P}_{ml}$ of mixed logit functions. Since Proposition 2 states $\mathcal{P}_{ml} = \text{rint.}\mathcal{P}_r$, it suffices to show that $\rho \in \text{rint.}\mathcal{P}_r$, equivalently there exists a (relative) neighborhood of ρ such that any element of the neighborhood belongs to the set $\mathcal{P}_r$ of random utility functions.

[8] In a similar way, I can be prove that a weaker version of Quasi-Stochastic Rationality, which allows the equality in (5), characterizes the random utility model.

Since the Block-Marschak polynomials of ρ are strictly positive, it follows from the continuity of K in ρ that the Block-Marschak polynomials are nonnegative in a small neighborhood of ρ. Moreover, it is possible to make the neighborhood small enough to be contained by the set $\mathcal{P}$ of the random choice functions. Thus, any element of the neighborhood is a random choice function whose Block-Marschak polynomials are nonnegative. Therefore, by the axiomatization of Falmagne (1978), any element of the neighborhood belongs to $\mathcal{P}_r$. It follows that $\rho \in \text{rint}.\mathcal{P}_r$. See the concrete proof in the appendix

Besides the axiomatization by Falmagne (1978), McFadden and Richter (1990) and Clark (1996) have proposed other axiomatizations of the random utility model. I obtain alternative axiomatizations of the mixed logit model by modifying the axioms of McFadden and Richter (1990) and Clark (1996). However, the ways I need to modify the axioms are not as simple the way I modified the axiom of Falmagne (1978) in this section. Moreover, the meaning of the axioms may be not so clear. For these reasons, the alternative axiomatizations appear in the appendix.

4 Axiomatization of the Mixed Linear Logit Model

In an empirical analysis, as mentioned before Definition 5, an alternative $x \in X$ can be identified by the vector of explanatory variables of x. Therefore, in this section, I assume that X is a finite subset of $\mathbf{R}^k$ for some natural number k (where k is the number of the explanatory variables). Then, I show that if X is affinely independent, then the same results obtained in Theorems 1 and 2 for the mixed logit model also hold for the mixed *linear* logit model. To show this result, I first prove the two preliminary propositions.

Definition 8. *A strict preference ranking $\pi \in \Pi$ is linearly representable if there exists $\beta \in \mathbf{R}^k$ such that for all $x, y \in X$,*

$$\pi(x) > \pi(y) \Longleftrightarrow \beta \cdot x > \beta \cdot y.$$

To motivate the first preliminary proposition, notice that, depending on the structure of X, there may be a ranking π which is not linearly representable. For example, let $X = \{x, y, z\}$ and $y = 1/2x + 1/2z$. Then for any $\beta \in \mathbf{R}^k$, it is the case that either $\beta \cdot x \geq \beta \cdot y \geq \beta \cdot z$ or $\beta \cdot z \geq \beta \cdot y \geq \beta \cdot x$. Hence, the ranking in which y is the strictly best element is not linearly representable. This is the crucial difference between this section and the previous section. The following proposition implies that the difference "disappears" when and only when X is affinely independent.

Proposition 4. *Let X be a finite subset of $\mathbf{R}^k$. The set X is affinely independent if and only if any ranking $\pi \in \Pi$ is linearly representable.*

To understand this proposition graphically, see Figure 1 and Figure 2. In the figures, I assume that $k = 2$. So $X = \{x, y, z\}$ in Figure 1 is affinely independent and $X = \{x, y, z, w\}$ in Figure 2 is affinely dependent.

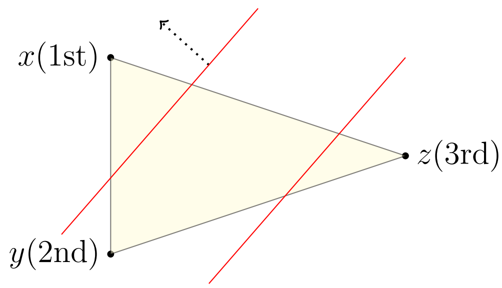

Figure 1: The set $X = \{x, y, z\}$ is affinely independent. Any ranking is linearly representable with some $\beta \in \mathbf{R}^2$. For example, the ranking $\pi(x) > \pi(y) > \pi(z)$ is linearly representable with $\beta \in \mathbf{R}^2$, which defines the parallel hyperplanes.

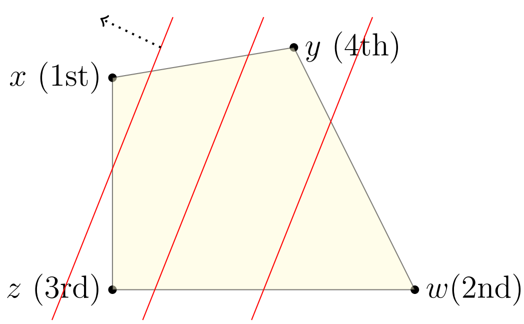

Figure 2: The set $X = \{x, y, z, w\}$ is affinely dependent. The ranking $\pi(x) > \pi(w) > \pi(y) > \pi(z)$ is not linearly representable. As the figure shows, no matter how one chooses $\beta \in \mathbf{R}^2$ and draws parallel hyperplanes, it is impossible to have $\beta \cdot x > \beta \cdot w > \beta \cdot z > \beta \cdot y$.

The condition that X is affinely independent could be easily satisfied in an empirical analysis. An empirical researcher may want to include a constant term (i.e.,

1) in the vector x of explanatory variables. (In that case, one needs to use $(x, 1)$.) The relevant condition for that case is that $\{(x, 1)|x \in X\}$ is linearly independent.[9] Given Proposition 4, I can prove the same result obtained in Proposition 2 for the mixed logit model also holds for the mixed *linear* logit model.

Proposition 5. *Let X be a finite subset of $\mathbf{R}^k$. The set of mixed linear logit functions is the relative interior of the set of random utility functions (i.e., $\mathcal{P}_{mll} = rint.\mathcal{P}_r$) if and only if X is affinely independent.*

Given Proposition 5, I can prove that if X is affinely independent, then the same results obtained in Theorems 1 and 2 for the mixed logit model also hold for the mixed *linear* logit model.

Theorem 3. *Let X be an affinely independent finite subset of $\mathbf{R}^k$. For any random choice function ρ, the following statements are equivalent:*

(i) the function ρ is a mixed linear logit function,

(ii) the function ρ satisfies Quasi-Stochastic Rationality,

(iii) $K(\rho, D, x) > 0$ for any $(D, x) \in \mathcal{D} \times X$ such that $x \in D$.

To see intuitively how Theorem 3 holds, notice that the sketch of proofs of Theorems 1 and 2 depends on the use of the mixed logit functions only because of Propotion 2 (i.e., $\mathcal{P}_{ml} = \text{rint}.\mathcal{P}_r$). Proposition 5 proves that the same result holds for the mixed *linear* logit functions (i.e., $\mathcal{P}_{mll} = \text{rint}.\mathcal{P}_r$). Hence Theorem 3 holds. See appendix for the concrete proof.

5 Concluding Remarks

I conclude the paper with some remarks, most of which are implied by the results in the previous sections. Remarks 1, 2, 3 involve the approximation of a random utility function by a mixed logit function. Remark 4 concerns the identification of the mixed logit model. Remarks 5 provides a representation result of a random utility function. Finally, in Remark 6, I mention the alternative axiomatizations of the mixed logit model.

Proposition 1 (ii) and Proposition 5 immediately imply Remark 1.

[9] X is affinely independent if and only if $\{(x, 1)|x \in X\}$ is linearly independent.

Remark 1. *Let X be a finite subset of $\mathbf{R}^k$.*
(i) If X is affinely independent, then (a) any interior random utility function can be represented as a convex combination of linear logit functions; (b) any noninterior random utility function can be approximated by a convex combination of linear logit functions.
(ii) If X is not affinely independent, then there is a random utility function which cannot be approximated by a convex combination of linear logit functions.

Remark 1 is related with Theorem 1 of McFadden and Train (2000). In their Theorem 1, McFadden and Train (2000) state that under some technical conditions, if $\rho(\cdot)$ is a random utility function, then for any positive number ε, there exist (i) a vector $p(x)$ of polynomials of x for each $x \in X$; and (ii) a mixed logit function ρ' defined by the equation (7) below such that the distance between $\rho'(D, x)$ and $\rho(D, x)$ is less than ε for any $x \in D$ and any finite subset D of X, where the function ρ' is defined with the vectors $\{p(x)\}_{x \in X}$ of polynomials as follows:

$$\rho'(D, x) = \int \frac{\exp(p(x) \cdot \beta(x))}{\sum_{y \in D} \exp(p(y) \cdot \beta(y))} dm(\beta). \tag{7}$$

Theorem 1 of McFadden and Train (2000) implies the generality of the mixed logit model; the generality is one of the essential reasons why the mixed logit model has been popular. As mentioned earlier, however, there is one limitation of Theorem 1 of McFadden and Train (2000). They say "One limitation of Theorem 1 is that it provides no practical indication of how to choose parsimonious mixing families, or how many terms are needed to obtain acceptable approximations..." (p. 452)

Remark 1 overcomes the limitation. To see this notice that in McFadden and Train (2000), each logit function is linear in the vector $p(x)$ of polynomials but not in x. The authors do not specify how one can construct the vector $p(x)$ or even the dimension of the vector. Depending on the bound ε, the vector of polynomials can be arbitrarily long by including higher degree terms. In contrast, in Remark 1, one can focus on the mixed *linear* logit model. In other words, one can assume $p(x) = x$ for any $x \in X$. In an empirical analysis, researchers often use this linear model, so Remark 1 provides direct support for this model.

There are three additional advantages to Remark 1 in comparison with Theorem 1 of McFadden and Train (2000). First, the result by McFadden and Train (2000) guarantees only an approximation, while result (ia) in Remark 1 guarantees the exact equality for the case of interior random utility functions. Second, to achieve the exact equality, Remark 1 states that it is enough to use a convex combination of linear logit functions. Third, part (ii) of the remark shows that if X is not affinely independent, then the set of mixed linear logit functions is *not* large enough to approximate any random utility function.

The setup of McFadden and Train (2000) is more general than mine in that they allow X to be infinite. McFadden and Train (2000) also allow that for a random choice function to be dependent on the observed attributes of agents. To make the discussion above clearer, I assumed that the set of the agents is homogeneous. However, I can easily include the set of the observed attributes in my model by allowing a primitive random choice function to be dependent on the agents' observed attributes.

In the next remark, I describe how one can construct a convex combination of logit functions that is arbitrarily close to a random utility function.

Remark 2. *Let X be an affinely independent finite subset of $\mathbf{R}^k$. Let ρ be a random utility function. Then there exists a set $\{\lambda_\pi\}_{\pi\in\Pi}$ of nonnegative numbers such that $\rho = \sum_{\pi\in\Pi}\lambda_\pi\rho^\pi$ and $\sum_{\pi\in\Pi}\lambda_\pi = 1$.*[10] *Fix any $\pi \in \Pi$. By Proposition 4, there exists $\beta \in \mathbf{R}^k$ such that $\pi(x) > \pi(y)$ if and only if $\beta\cdot x > \beta\cdot y$ for any $x, y \in X$.*[11] *For any positive integer n and any $(D,x) \in \mathcal{D}\times X$ such that $x \in D$, define*

$$\rho^\pi_{n\beta}(D,x) \equiv \frac{\exp(n\beta\cdot x)}{\sum_{y\in D}\exp(n\beta\cdot y)}.$$

An easy calculation shows that $\rho^\pi_{n\beta} \to \rho^\pi$ as $n\to\infty$. For each $\pi\in\Pi$, such a sequence $\{\rho^\pi_{n\beta}\}_{n=1}^\infty$ exists. For each positive integer n, define $\rho_n \equiv \sum_{\pi\in\Pi}\lambda_\pi\rho^\pi_{n\beta}$. Hence $\rho_n \to \sum_{\pi\in\Pi}\lambda_\pi\rho^\pi \equiv \rho$ as $n\to\infty$.

The remarks above involve logit functions. As the next remark implies, similar results can be proved for some other classes of random utility functions.

Remark 3. *Let $\mathcal{Q}$ be a nonempty subset of the set $\mathcal{P}_r$ of random utility functions. Suppose that for any ranking $\pi\in\Pi$, there exists a sequence $\{\rho_n\}_{n=1}^\infty$ of $\mathcal{Q}$ such that $\rho_n \to \rho^\pi$ as $n\to\infty$. Then, (a) any interior random utility function can be represented as a convex combination of elements of $\mathcal{Q}$; (b) any noninterior random utility function can be approximated by a convex combination of elements of $\mathcal{Q}$.*

[10] To see this, remember that $\mathcal{P}_r = \text{co.}\{\rho^\pi|\pi\in\Pi\}$. The set $\{\lambda\}_{\pi\in\Pi}$ can be easily obtained by a computer as a solution of linear inequalities.

[11] Such β can be easily obtained by a computer as a solution of linear inequalities.

This remark is implied by Lemma 4 in the appendix.[12] The conditions of Remark 3 are satisfied when $\mathcal{Q}$ is the set $\mathcal{P}_l$ of logit functions. (See the proof of Proposition 2.) The conditions of Remark 3 can also be satisfied by some other classes of random utility functions. For instance, the set of probit functions satisfies these conditions. Therefore, (a) any interior random utility function can be represented as a convex combination of probit functions; (b) any noninterior random utility function can be approximated by a convex combination of probit functions.

Remark 4 concerns the identification of the mixed logit model. Empirical researchers have intensively studied the identification of the random coefficients model including the mixed logit model.[13] Although the identification problem is not the main topic of this paper, Propositions 1, 2, and 3 imply the following remark concerning the identification of the mixed loigt model.

Remark 4. *Statement (i) of Proposition 1 implies that for any mixed logit function defined with a probability measure whose support is infinite, one can find an observationally equivalent convex combination of logit functions. In the same way, statement (ii) implies the nonuniqueness of the representation of a mixed linear logit function.*

Even a convex combination of logit functions may be represented in multiple ways. To see this, notice that it follows from Propositions 1, 2, and 3 that $\dim co.\mathcal{P}_l = \dim rint.\mathcal{P}_r = \dim \mathcal{P}_r = (|X|-2)2^{|X|-1}+1$.[14] *On the other hand, there are infinitely many logit functions when $|X| \geq 2$. Hence, an element of $co.\mathcal{P}_l$ may be represented in multiple ways.*[15] *(Moreover, it follows from Caratheodory's theorem that an element of $co.\mathcal{P}_l$ is represented as a convex combination of at most $(|X|-2)2^{|X|-1}+2$ logit functions.) If X is affinely independent, the same arguments above hold for a convex combination of linear logit functions.*

Fox et al. (2012) have studied the identification of a special case of a mixed linear logit function defined with a probability measure whose support is compact.

[12] Under the supposition of the remark, Lemma 4 implies that $\text{rint}.\mathcal{P}_r = \text{co}.\mathcal{Q}$. This means statement (a) in Remark 3. Moreover, since $\mathcal{P}_r$ is closed, it follows that $\mathcal{P}_r = \text{cl}.\mathcal{P}_r = \text{cl.rint}.\mathcal{P}_r = \text{cl.co}.\mathcal{Q}$. This means statement (b) in Remark 3.

[13] See Berry and Haile (2009), Fox et al. (2012), and Fox and Gandhi (2016) for examples.

[14] The second equality holds by Theorem 2.1.3 of Hiriart-Urruty and Lemaréchal (2012).

[15] Remember that (i) the maximal number of affinely independent points in a set C is $\dim C + 1$; (ii) a set C is affinely independent if and only if for any $y \in \text{co}.C$, there exists a unique set of nonnegative numbers $\{\lambda_x\}_{x \in C}$ such that $y = \sum_{x \in C} \lambda_x x$ and $\sum_{x \in C} \lambda_x = 1$.

Fox et al. (2012) show that the identification is possible if the set X of alternatives contains a nonempty open set and all elements of x are continuous. This result by Fox et al. (2012) is consistent with Remark 4 because X is finite in this paper.

Proposition 4 immediately implies Remark 5 on a representation of a random utility function.

Remark 5. *For any random utility function ρ, there exists $\mu \in \Delta(\mathbf{R}^k)$ such that*

$$\rho(D, x) = \mu(\{\beta \in \mathbf{R}^k | \beta \cdot x \geq \beta \cdot y \text{ for all } y \in D\})$$

if and only if X is affinely independent.

In the empirical literature of the random-coefficients model, researchers have analyzed various ways to introduce the randomness of coefficients (i.e., β). In the literature, assuming the linear model is sometimes considered to be restrictive. Remark 5 states, however, that one can focus on the linear model with no loss of generality if and only if X is affinely independent.[16]

In Remark 6, I mention the alternative axiomatizations of the mixed logit model. McFadden and Richter (1990) characterize the random utility model by the *Axiom of Revealed Stochastic Preference*. Clark (1996) characterizes the random utility model by the axiom of *Coherency*. I modify these two axioms to obtain the *Strict Axiom of Revealed Stochastic Preference* (Definition 12) and the axiom of *Strict Coherency* (Definition 15). Then in Theorems 4 and 5, I characterize the mixed logit model by each axiom. However, the ways I modify the two axioms are not as simple as the way I modified the axiom of Falmagne (1978) in section 3.2. So these alternative axiomatizations appear in the appendix.

Remark 6 summarizes all the axiomatizations in this paper including those in the appendix as follows:

Remark 6. *For any random utility function ρ, the following five statements are equivalent: (i) ρ is a mixed logit function; (ii) ρ satisfies Quasi-Stochastic Rationality (Axiom 1); (iii) the Block-Marschak polynomials of ρ are strictly positive; (iv)*

[16]Remark 5 is consistent with the axiomatization of the random expected utility model by Gul and Pesendorfer (2006). They show that ρ satisfies the axioms of mixture continuity, linearity, extremeness, and regularity if and only if ρ is a random expected-utility function. In my setup, all of the axioms except regularity are satisfied vacuously when X is affinely independent. Regularity is satisfied by the random utility model.

ρ satisfies the Strict Axiom of Revealed Stochastic Preference; and (v) ρ is Strictly Coherent.

Moreover, if X is an affinely independent subset of $\mathbf{R}^k$, then statements (i)–(v) are also equivalent to this statement: (vi) ρ is a mixed linear logit function.

A Proofs

A.1 Proof of Proposition 1

To show the proposition, I will show the following general result as a lemma. The lemma is trivial when the set C is closed. I used the lemma with $C = \mathcal{P}_l$, where the set $\mathcal{P}_l$ is not closed.

Let n be a positive integer. For any $x \in \mathbf{R}^n$, x_i denotes the i-th element of x for any $i \in \{1, \ldots, n\}$.

Lemma 1. *For any set $C \subset \mathbf{R}^n$, let $\Delta(C)$ denote the set of probability measures over C.[17] Then, $co.C = \left\{ \int xdm(x) | m \in \Delta(C) \right\}$, where $\int xdm(x)$ denotes n-dimensional vector whose i-th element is $\int x_i dm(x)$ for any $i \in \{1, \ldots, n\}$.*

Proof. By definition, I immediately obtain $\text{co.}C \subset \{\int xdm(x)|m \in \Delta(C)\}$. In the following, I will show that

$$\left\{ \int xdm(x)|m \in \Delta(C) \right\} \subset \text{co.}C. \tag{8}$$

First I will show that

$$\left\{ \int xdm(x)|m \in \Delta(C) \right\} \subset \text{cl.co.}C. \tag{9}$$

To prove this statement, suppose by way of contradiction that $\int xdm(x) \notin \text{cl.co.}C$ for some $m \in \Delta(C)$. Then by the strict separating hyperplane theorem (Corollary 11.4.2 of Rockafellar (2015)), there exist $t \in \mathbf{R}^n \setminus \{0\}$ and $\alpha \in \mathbf{R}$ such that $(\int xdm(x)) \cdot t = \alpha > x \cdot t$ for any $x \in \text{cl.co.}C$. This is a contradiction because $\alpha = (\int xdm(x)) \cdot t = \int (x \cdot t)dm(x) < \int \alpha dm(x) = \alpha$.

I now will show (8) by the induction on the dimension of co.C.

Induction Base: If $\dim \text{co.}C = 1$, then (8) holds obviously. If $\dim \text{co.}C = 2$, then there must exist y, z such that co.C is the line segment between y and z.

[17] The Borel algebra here is the smallest sigma algebra that contains all open set relative to the set C.

In the following, I assume that the line segment does not contain both y and z but the proof for the other cases are similar. Then for any $x \in \text{co}.C$, there exists unique $\alpha(x) \in (0,1)$ such that $x = \alpha(x)y + (1-\alpha(x))z$. Notice that the function α is continuous in x and hence measurable. Moreover, the function α is integrable because α is bounded and nonnegative. Choose any $m \in \Delta(C)$. Then $\int \alpha(x)dm(x)$ exists. Moreover, since $0 < \alpha(x) < 1$, it follows from the monotonicity of integral that $0 < \int \alpha(x)dm(x) < 1$. Denote the value of the integral by $\beta \in (0,1)$. Then, $\int xdm(x) = \int \alpha(x)y + (1-\alpha(x))zdm(x) = \beta y + (1-\beta)z \in \text{co}.C$, as desired.

Choose an integer $k \geq 3$.

Induction Hypothesis: Now suppose that (8) holds for any C such that $\dim C \leq k$.

Induction Step: For any C such that $\dim C = k+1$, (8) holds. To prove the step, choose any $m \in \Delta(C)$. By (9), I have $\int xdm(x) \in \text{cl.co}.C$.

First consider the case where $\int xdm(x) \in \text{rint.cl.co}.C$. Then since $\text{rint.cl.co}.C = \text{rint.co}.C$ (by Theorem 6.3 of Rockafellar (2015)), so $\int xdm(x) \in \text{co}.C$, as desired.

Next consider the case where $\int xdm(x) \notin \text{rint.cl.co}.C$. Then, $\int xdm(x) \in \partial\text{cl.co}.C \equiv \text{cl.co}.C \setminus \text{rint.co}.C$. There exists a supporting hyperplane H of cl.co.C at $\int xdm(x)$. Then, there exist $t \in \mathbf{R}^n \setminus \{0\}$ and $\alpha \in \mathbf{R}$ such that $H = \{x | x \cdot t = \alpha\}$ and $\int xdm(x) \cdot t = \alpha > x \cdot t$ for any $x \in \text{cl.co}.C \cap H^c$. This implies that $m(H) = 1$. Hence, $m(H \cap C) = 1$. Since H is a supporting hyperplane and $\text{cl.co}.C \not\subset H$, I obtain $\dim(H \cap \text{aff}.C) \leq k$. Hence, $\dim(H \cap C) \leq k$. Therefore, the induction hypothesis shows that $\int xdm(x) \in \text{co}.(H \cap C) \subset \text{co}.C$, as desired. □

The result is not true in an infinite dimensional space.[18] The lemma immediately implies the two statements in Proposition 1.

A.2 Lemmas

I prove three more lemmas that I use in the rest of the appendix.

Lemma 2. *The set $\mathcal{P}_r$ of random utility functions is a polytope. Moreover, $\mathcal{P}_r = co.\{\rho^\pi | \pi \in \Pi\}$, and there exist hyperplanes $\{H_i\}_{i=1}^n$ in $\mathbf{R}^{\mathcal{D}\times X}$ such that aff.$\mathcal{P}_r \not\subset H_i^-$*

[18] Let $\{e_i\}_{i=1}^\infty$ be the base of the infinite dimensional real space. Define $C = \{e_i\}_{i=1}^\infty$. Define a measure m on C such that $m(e_i) = (1/2)^i$ for each i. Then, $\sum_{i=1}^\infty m(e_i) = 1$, so that m is a probability measure on C. $\int xdm$ cannot be represented as any convex combination of elements of C. For any $y \in \text{co}.C$, there exists i such that $y(e_i) = 0$.

and $\mathcal{P}_r = (\cap_{i=1}^n H_i^-) \cap \text{aff.}\mathcal{P}_r$, *where* H_i^- *is the closed lower-half space of* H_i *for each* $i \in \{1, \ldots, n\}$.

Proof. Choose any $\rho \in \mathcal{P}_r$ to show $\rho \in \text{co.}\{\rho^\pi | \pi \in \Pi\}$. There exists $\nu \in \Delta(\Pi)$ that rationalizes ρ. Define $\lambda_\pi = \nu(\pi)$ for each $\pi \in \Pi$. Define $\rho' = \sum_{\pi\in\Pi} \lambda_\pi \rho^\pi$ to show $\rho = \rho'$. For each $(D, x) \in \mathcal{D} \times X$, $\rho(D, x) = \nu(\pi \in \Pi | \pi(x) \geq \pi(D)) = \sum_{\pi\in\Pi} \nu(\pi) 1(\pi(x) \geq \pi(D)) = \rho'(D, x)$. Then $\rho = \rho' \in \text{co.}\{\rho^\pi | \pi \in \Pi\}$. So $\mathcal{P}_r \subset \text{co.}\{\rho^\pi | \pi \in \Pi\}$. The argument can be reversed to obtain the converse. By the definition of polytope and Theorem 9.4 of Soltan (2015), the desired hyperplanes exist. □

The next lemma says that any convex combination of logit functions is a *full-support* random utility function.[19]

Lemma 3. *For any* $\rho \in \text{co.}\mathcal{P}_l$, *there exists* $\nu \in \Delta(\Pi)$ *such that (i)* ρ *is rationalized by* ν*; (ii)* $\nu(\pi) > 0$ *for all* $\pi \in \Pi$.

Proof. I show the following two statements: (i) For any $\rho \in \mathcal{P}_l$, there exists $\nu \in \Delta(\Pi)$ such that ρ is rationalized by ν. Moreover $\nu(\pi) > 0$ for all $\pi \in \Pi$; (ii) For any $\alpha \in [0, 1]$, if logit functions ρ and ρ' are respectively rationalized by ν and ν', then $\alpha\rho + (1 - \alpha)\rho'$ is rationalized by $\alpha\nu + (1 - \alpha)\nu'$.

To show (i), remember that for any $\rho \in \mathcal{P}_l$, there exists $u \in \mathbf{R}_{++}^{|X|}$ such that $\rho(D, x) = u(x)/\sum_{y\in D} u(y)$ and $\sum_{x\in X} u(x) = 1$, where $\mathbf{R}_{++}$ is the set of all positive real numbers. By Block and Marschak (1960), $\rho \in \mathcal{P}_r$, so there exists $\nu \in \Delta(\Pi)$ such that ν rationalizes ρ. Moreover, in their construction of ν, they obtain that for any $\pi \in \Pi$,

$$\nu(\pi) = \prod_{k=1}^{|X|} \frac{u(x_k)}{\sum_{l=k}^{|X|} u(x_k)},$$

where $X = \{x_1, x_2, \ldots, x_{|X|}\}$ and $\pi(x_1) > \pi(x_2) > \cdots > \pi(x_{|X|})$. Since $u > 0$, I have $\nu(\pi) > 0$.

Statement (ii) can be proved as follows: $(\alpha\rho + (1 - \alpha)\rho')(D, x) = \alpha\rho(D, x) + (1 - \alpha)\rho'(D, x) = \alpha\nu(\{\pi \in \Pi | \pi(x) \geq \pi(D)\}) + (1 - \alpha)\nu'(\{\pi \in \Pi | \pi(x) \geq \pi(D)\}) = \alpha \sum_{\pi\in\Pi:\pi(x)\geq\pi(D)} \nu(\pi) + (1-\alpha) \sum_{\pi\in\Pi:\pi(x)\geq\pi(D)} \nu'(\pi) = \sum_{\pi\in\Pi:\pi(x)\geq\pi(D)} \alpha\nu(\pi) + (1-\alpha)\nu'(\pi) = (\alpha\nu + (1 - \alpha)\nu')(\{\pi \in \Pi | \pi(x) \geq \pi(D)\})$. □

[19] Block and Marschak (1960) show that any logit function is a full-support random utility function, although they do not state this explicitly.

Lemma 4 is used to prove Propositions 2 and 5. Moreover, Lemma 4 implies Remark 3.

Lemma 4. *Let $\mathcal{Q}$ be a nonempty subset of the set $\mathcal{P}_r$ of random utility functions. Suppose that for any $\pi \in \Pi$, there exists a sequence $\{\rho_n\}_{n=1}^{\infty}$ of $\mathcal{Q}$ such that $\rho_n \to \rho^\pi$ as $n \to \infty$. Then, rint.$\mathcal{P}_r \subset$ co.$\mathcal{Q}$.*

Proof. Suppose by way of contradiction that there exists $\rho \in \text{rint}.\mathcal{P}_r \setminus \text{co}.\mathcal{Q}$. Because $\text{co}.\mathcal{Q} \neq \emptyset$, I obtain $\text{rint.co}.\mathcal{Q} \neq \emptyset$. Since $\rho \notin \text{co}.\mathcal{Q}$, then by the proper separating hyperplane theorem (Theorem 11.3 of Rockafellar (2015)), there exist $t \in \mathbf{R}^{\mathcal{D}\times X} \setminus \{0\}$ and $a \in \mathbf{R}$ such that $\rho \cdot t \geq a \geq \rho' \cdot t$ for any $\rho' \in \text{co}.\mathcal{Q}$, and $a > \rho'' \cdot t$ for some $\rho'' \in \text{co}.\mathcal{Q}$.

I obtain a contradiction by two steps. Define $\hat{\mathcal{P}}_r = \{\hat{\rho} \in \mathcal{P}_r | t \cdot \hat{\rho} > t \cdot \rho\}$.

Step 1: $\hat{\mathcal{P}}_r \neq \emptyset$. To prove the step, remember that there exists $\rho'' \in \text{co}.\mathcal{Q}$ such that $\rho'' \cdot t < \rho \cdot t$. Moreover, since $\mathcal{Q} \subset \mathcal{P}_r$ and the set $\mathcal{P}_r$ is convex, it follows that $\rho'' \in \text{co}.\mathcal{P}_l \subset \mathcal{P}_r$. Since $\rho \in \text{rint}.\mathcal{P}_r$, there exists $\lambda > 1$ such that $\lambda\rho + (1-\lambda)\rho'' \in \mathcal{P}_r$. Moreover, $(\lambda\rho + (1-\lambda)\rho'') \cdot t = \lambda\rho \cdot t + (1-\lambda)\rho'' \cdot t = \rho \cdot t + (\lambda - 1)(\rho \cdot t - \rho'' \cdot t) > \rho \cdot t$, where the last inequality holds because $\lambda > 1$ and $\rho'' \cdot t < \rho \cdot t$. So $\lambda\rho + (1-\lambda)\rho'' \in \hat{\mathcal{P}}_r$, and $\hat{\mathcal{P}}_r \neq \emptyset$.

Step 2: There exists $\rho' \in \text{co}.\mathcal{Q}$ such that $\rho' \cdot t > \rho \cdot t$. To prove the step, choose any $\hat{\rho} \in \hat{\mathcal{P}}_r$. By Lemma 2, there exist nonnegative numbers $\{\hat{\lambda}_\pi\}_{\pi\in\Pi}$ such that $\hat{\rho} = \sum_{\pi\in\Pi} \hat{\lambda}_\pi \rho^\pi$ and $\sum_{\pi\in\Pi} \hat{\lambda}_\pi = 1$.

By the supposition, for any $\pi \in \Pi$, there exists a sequence $\{\rho'_n\}_{n=1}^{\infty}$ of $\mathcal{Q}$ such that $\rho'_n \to \rho^\pi$ as $n \to \infty$. Therefore, for any $\pi \in \Pi$ and any positive number ε, there exists $\rho'_\pi \in \{\rho'_n\}_{n=1}^{\infty}$ such that $\|\rho'_\pi - \rho^\pi\| < \varepsilon$. Define $\rho' = \sum_{\pi\in\Pi} \hat{\lambda}_\pi \rho'_\pi$. Then $\rho' \in \text{co}.\mathcal{Q}$ and $\|\rho' - \hat{\rho}\| = \left\| \sum_{\pi\in\Pi} \hat{\lambda}_\pi (\rho'_\pi - \rho^\pi) \right\| \leq \sum_{\pi\in\Pi} \hat{\lambda}_\pi \|\rho'_\pi - \rho^\pi\| \leq \sum_{\pi\in\Pi} \hat{\lambda}_\pi \varepsilon = \varepsilon$. Therefore, $|t \cdot \rho' - t \cdot \hat{\rho}| \leq \|t\| \|\rho' - \hat{\rho}\| \leq \|t\|\varepsilon$. Since $t \cdot \hat{\rho} > t \cdot \rho$, then by choosing ε small enough, I obtain $t \cdot \rho' > t \cdot \rho$. □

A.3 Proof of Proposition 2

By Proposition 1, it suffices to show that $\text{co}.\mathcal{P}_l = \text{rint}.\mathcal{P}_r$.

First, I show that $\text{co}.\mathcal{P}_l \subset \text{rint}.\mathcal{P}_r$. By Lemma 3, for any $\rho \in \text{co}.\mathcal{P}_l$, there exists $\lambda_\pi > 0$ for any $\pi \in \Pi$ such that $\rho = \sum_{\pi\in\Pi} \lambda_\pi \rho^\pi$ and $\sum_{\pi\in\Pi} \lambda_\pi = 1$. Therefore, by Theorem 6.9 in Rockafellar (2015), $\rho \in \text{rint.co}.\{\rho^\pi | \pi \in \Pi\} = \text{rint}.\mathcal{P}_r$, where the last equality holds by Lemma 2.

Next, I show that rint.$\mathcal{P}_r \subset$ co.$\mathcal{P}_l$. I apply Lemma 4 with $\mathcal{Q} = \mathcal{P}_l$. To see the conditions of Lemma 4 are satisfied remember that, by Lemma 3 $\mathcal{P}_l$ is a nonempty subset of $\mathcal{P}_r$. Moreover, by Fact 5 in appendix A of Gul et al. (2014), for any $\pi \in \Pi$, there exists a sequence $\{\rho_n\}_{n=1}^{\infty}$ of $\mathcal{P}_l$ such that $\rho_n \to \rho^\pi$ as $n \to \infty$. It follows that rint.$\mathcal{P}_r \subset$ co.$\mathcal{P}_l$.[20]

A.4 Proof of Proposition 3

To prove Proposition 3, I prove one more lemma.

Lemma 5. *(i) For any $q \in \Delta(\mathcal{D})$ and any $u(D,\cdot) \in \mathbf{R}^D$ for each $D \in \mathcal{D}$, $E(\rho^\pi : q,u) \neq E(\rho^{\pi'} : q,u)$ for some $\pi, \pi' \in \Pi$ if and only if $u(D,\cdot)$ is not constant for some D with $q(D) > 0$.*
(ii) For any $t \in \mathbf{R}^{\mathcal{D}\times X}$, $\rho^\pi \cdot t = \rho^{\pi'} \cdot t$ for all $\pi, \pi' \in \Pi$ if and only if $t(D,x) = t(D,y)$ for all $D \in \mathcal{D}$ and $x, y \in D$.

Proof. First I will show statement (i) by assuming statement (ii). Fix any $q \in \Delta(\mathcal{D})$ and any $u(D,\cdot) \in \mathbf{R}^D$ for each $D \in \mathcal{D}$. For each $(D,x) \in \mathcal{D} \times X$ such that $x \in D$, define $t(D,x) = q(D)u(D,x)$. For each $(D,x) \in \mathcal{D} \times X$ such that $x \notin D$, define $t(D,x) = 0$. Then $t \in \mathbf{R}^{\mathcal{D}\times X}$. Remember that for any $\rho \in \mathcal{P}$, $\rho(D,x) = 0$ for any $x \notin D$. Hence, $\rho \cdot t = \sum_{(D,x)\in\mathcal{D}\times X} q(D)u(D,x)\rho(D,x) \equiv E(\rho : q,u)$. Then

$$
\begin{aligned}
& E(\rho^\pi : q,u) \neq E(\rho^{\pi'} : q,u) \text{ for some } \pi,\pi' \in \Pi \\
\iff\ & \rho^\pi \cdot t \neq \rho^{\pi'} \cdot t \text{ for some } \pi,\pi' \in \Pi \\
\iff\ & t(D,x) \neq t(D,y) \text{ for some } D \in \mathcal{D} \text{ and } x,y \in D && (\because \text{(ii)}) \\
\iff\ & q(D)u(D,x) \neq q(D)u(D,y) \text{ for some } D \in \mathcal{D} \text{ and } x,y \in D && (\because \text{the definition of } t) \\
\iff\ & u(D,x) \neq u(D,y) \text{ for some } D \in \mathcal{D} \text{ with } q(D) > 0 \text{ and } x,y \in D.
\end{aligned}
$$

So statement (i) holds.

In the following, I will show statement (ii). For notational convenience, for any $\pi \in \Pi$ and $D \in \mathcal{D}$ with $D = \{x_1, \ldots, x_{|D|}\}$, I write $\rho^\pi(D) = \big(\rho^\pi(D,x_1), \ldots, \rho^\pi(D,x_{|D|})\big)$.

The if part of the statement (ii) is easy to prove. Assume $t(D,x) = t(D,y)$ for all

[20] For completeness, I describe here how Gul et al. (2014) construct the sequence $\{\rho_n\}_{n=1}^{\infty}$ of $\mathcal{P}_l$. For each natural number n, each $\pi \in \Pi$, and each $x \in X$, define $u_\pi^n(x) \equiv (1/n)^{|X|-\pi(x)}$. For each $(D,x) \in \mathcal{D} \times X$ such that $x \in D$, define $\rho_n(D,x) \equiv \frac{u_\pi^n(x)}{\sum_{y\in D} u_\pi^n(y)} = \frac{1}{\sum_{y\in D:\pi(y)>\pi(x)}(1/n)^{\pi(x)-\pi(y)} + 1 + \sum_{y\in D:\pi(y)<\pi(x)}(1/n)^{\pi(x)-\pi(y)}}$. For each $(D,x) \in \mathcal{D} \times X$ such that $x \notin D$, define $\rho_n(D,x) \equiv 0$. Then $\rho_n(D,x) \to \rho^\pi(D,x)$ as $n \to \infty$ for each $(D,x) \in \mathcal{D} \times X$.

$D \in \mathcal{D}$ and $x, y \in D$. Define $t(D) = t(D, x)$ for any $x \in D$. Then for any $\pi \in \Pi$, $\rho^\pi \cdot t = \sum_{D \in \mathcal{D}} \sum_{x \in D} \rho^\pi(D, x) t(D, x) = \sum_{D \in \mathcal{D}} t(D) \sum_{x \in D} \rho^\pi(D, x) = \sum_{D \in \mathcal{D}} t(D)$.

I now prove the only if part of the statement (ii) by the induction on $|D|$.

Induction Base: When $|D| = 1$. Then $x = y$, so $t(D, x) = t(D, y)$. When $|D| = 2$. Then $D = \{x, y\}$. Consider $\pi, \pi' \in \Pi$ over X such that for any $z \in X \setminus \{x, y\}$, $\pi(z) = \pi'(z)$, $\pi(z) > \pi(x) > \pi(y)$, and $\pi'(z) > \pi'(y) > \pi'(x)$. Then for any $E \in \mathcal{D}$ such that $E \neq \{x, y\}$, $\rho^\pi(E) = \rho^{\pi'}(E)$. Moreover, $\rho^\pi(\{x, y\}, x) = 1 = \rho^{\pi'}(\{x, y\}, y)$ and $\rho^\pi(\{x, y\}, y) = 0 = \rho^{\pi'}(\{x, y\}, x)$. Since $t \cdot \rho^\pi = t \cdot \rho^{\pi'}$,

$$0 = \sum_{E \in \mathcal{D}} \sum_{x \in X} t(E, x)(\rho^\pi(E, x) - \rho^{\pi'}(E, x)) = t(\{x, y\}, x) - t(\{x, y\}, y).$$

So $t(\{x, y\}, x) = t(\{x, y\}, y)$. This provides the induction base.

Choose a positive integer $k \geq 2$.

Induction Hypothesis: For any $D \in \mathcal{D}$ such that $|D| \leq k$, $t(D, x) = t(D, y)$ for any $x, y \in D$.

Induction Step: For any $D \in \mathcal{D}$ such that $|D| = k + 1$ and any $x, y \in D$, $t(D, x) = t(D, y)$. To prove the step, denote D by $\{x, y, w_1, \ldots, w_{k-1}\}$. Choose any $\pi, \pi' \in \Pi$ such that for any $z \in X \setminus \{x, y, w_1, \ldots, w_{k-1}\}$ and any $i \in \{1, \ldots, k-1\}$, $\pi(z) = \pi'(z)$, $\pi(z) > \pi(x) > \pi(y) > \pi(w_i)$, $\pi'(z) > \pi'(y) > \pi'(x) > \pi'(w_i)$, and $\pi(w_i) = \pi'(w_i)$.

To show the induction step, I will show the following two facts: (a) For any $E \in \mathcal{D}$, $\{x, y\} \subset E$ and $\pi(x) \geq \pi(E)$ if and only if $\rho^\pi(E) \neq \rho^{\pi'}(E)$; (b) If $E \in \mathcal{D}$, $\{x, y\} \subset E$ and $\pi(x) \geq \pi(E)$, then $\rho^\pi(E, x) = 1$, $\rho^\pi(E, z) = 0$ for any $z \in D \setminus \{x\}$ and $\rho^{\pi'}(E, y) = 1$, $\rho^{\pi'}(E, z) = 0$ for any $z \in E \setminus \{y\}$.

It is easy to see statement (b) and the only if part of statement (a). To show the if part of statement (a), assume $\{x, y\} \not\subset E$ or $\pi(x) < \pi(z)$ for some $z \in E$. First consider the case where $\{x, y\} \not\subset E$. If both x, y do not belong to E, then $\rho^\pi(E) = \rho^{\pi'}(E)$ because the ranking over $X \setminus \{x, y\}$ is the same for π and π'. If only one of them, say x, belongs to E, then $\rho^\pi(E) = \rho^{\pi'}(E)$ because the ranking over $X \setminus \{y\}$ is the same for π and π'.

Next consider the case where $\pi(x) < \pi(z)$ for some $z \in E$. Then by the definition of π, I obtain $z \in X \setminus \{x, y, w_1, \ldots, w_{k-1}\}$. Therefore, $\pi'(y) < \pi'(z)$. Hence, $\rho^\pi(E, z) = 1 = \rho^{\pi'}(E, z)$ and $\rho^\pi(E, z') = 0 = \rho^{\pi'}(E, z')$ for all $z' \in E \setminus \{z\}$.

Now, I will prove the induction step. Since $t \cdot \rho^{\pi} = t \cdot \rho^{\pi'}$,

$$\begin{aligned}
0 &= \textstyle\sum_{(E,z)\in\mathcal{D}\times X} t(E,z)(\rho^{\pi}(E,z) - \rho^{\pi'}(E,z)) \\
&= \textstyle\sum_{(E,z)\in\mathcal{D}\times X:\{x,y\}\subset E, \pi(x)\geq\pi(E)} t(E,z)(\rho^{\pi}(E,z) - \rho^{\pi'}(E,z)) && (\because \text{(a)}) \\
&= \textstyle\sum_{E\in\mathcal{D}:\pi(x)\geq\pi(E),\{x,y\}\subset E} t(E,x) - t(E,y) && (\because \text{(b)}) \\
&= t(D,x) - t(D,y) + \textstyle\sum_{E\in\mathcal{D}:\pi(x)\geq\pi(E),\{x,y\}\subset E,|E|\leq k}(t(E,x) - t(E,y)).
\end{aligned}$$

Moreover by the Induction Hypothesis, the second term is zero. So $t(D,x) = t(D,y)$. □

Now I will prove Proposition 3.

The set $\{p \in \mathbf{R}^{\mathcal{D}\times X} | \text{(i) and (ii)}\}$ is affine. So it suffices to show that for any affine set A, if $\mathcal{P}_r \subset A$, then $\{p \in \mathbf{R}^{\mathcal{D}\times X} | \text{(i) and (ii)}\} \subset A$. Since the set is affine, then by Rockafellar (2015), there exist a positive integer L, $L \times (|\mathcal{D}| \times |X|)$ matrix B, and $L \times 1$ vector b such that $A = \{p \in \mathbf{R}^{\mathcal{D}\times X} | Bp = b\}$. For any $l \in \{1, \ldots, L\}$, $B_l(D,x)$ denotes $(l,(D,x))$ entry of B. (Remember that B has a column vector for each $(D,x) \in \mathcal{D} \times X$.) So $Bp = b$ means that for any $l \in \{1, \ldots, L\}$,

$$\sum_{D\in\mathcal{D}} \sum_{x\in X} B_l(D,x)p(D,x) = b_l. \tag{10}$$

By assuming $\mathcal{P}_r \subset \{p \in \mathbf{R}^{\mathcal{D}\times X} | Bp = b\}$, I will show that if ρ satisfies (i) and (ii), then (10) holds for any $l \in \{1, \ldots, L\}$.

Step 1: $B_l(D,x) = B_l(D,y)$ for any $l \in \{1, \ldots, L\}$, $D \in \mathcal{D}$, and $x, y \in D$. To prove step 1, fix any l. For any $\pi \in \Pi$, $\rho^{\pi} \in \mathcal{P}_r \subset \{p \in \mathbf{R}^{\mathcal{D}\times X} | Bp = b\}$. Hence, (10) holds with $p = \rho^{\pi}$ for any $\pi \in \Pi$. Thus $\rho^{\pi} \cdot B_l = \rho^{\pi'} \cdot B_l$ for any $\pi, \pi' \in \Pi$. By Lemma 5 (ii), this implies that $B_l(D,x) = B_l(D,y)$ for any $D \in \mathcal{D}$, and $x, y \in D$.

By Step 1, I can define $B_l(D) = B_l(D,x)$ for any $x \in D$.

Step 2: If p satisfies (i) and (ii), then $Bp = b$, or $\sum_{D\in\mathcal{D}} \sum_{x\in X} B_l(D,x)p(D,x) = b_l$ for any $l \in \{1, \ldots, L\}$. To prove step 2, choose any $\pi \in \Pi$ and $l \in \{1, \ldots, L\}$. Since $\rho^{\pi} \in \mathcal{P}_r \subset \{p \in \mathbf{R}^{\mathcal{D}\times X} | Bp = b\}$, then by (10),

$$b_l = \sum_{D\in\mathcal{D}} \sum_{x\in X} B_l(D,x)\rho^{\pi}(D,x) = \sum_{D\in\mathcal{D}} B_l(D), \tag{11}$$

where the second equality holds by $\rho^{\pi}(D,z) = 1$ if $\pi(z) \geq \pi(D)$ and $\rho^{\pi}(D,z) = 0$ otherwise.

Finally by using these equalities, for each $l \in \{1, \ldots, L\}$, I obtain the following equations:

$$\begin{aligned}
\sum_{D\in\mathcal{D}}\sum_{z\in X} B_l(D,x)p(D,z) &= \sum_{D\in\mathcal{D}}\sum_{z\in D} B_l(D,x)p(D,z) && (\because\ \text{(ii)})\\
&= \sum_{D\in\mathcal{D}}\sum_{z\in D} B_l(D)p(D,z) && (\because \text{Step 1})\\
&= \sum_{D\in\mathcal{D}} B_l(D)\sum_{z\in D} p(D,z) &&\\
&= \sum_{D\in\mathcal{D}} B_l(D) && (\because\ \text{(i)})\\
&= b_l. && (\because (11))
\end{aligned}$$

This establishes that aff.$\mathcal{P}_r = \{p \in \mathbf{R}^{\mathcal{D}\times X} | \text{(i) and (ii)}\}$. The equalities in (i) and (ii) are independent. So the dimension of $\mathcal{P}_r$ is $|\mathcal{D}| \times |X|$ minus the number of equalities of (i) and (ii). The number of equalities of (i) is the number of $D \in \mathcal{D}$, which is $2^n - 1$. The number of equalities of (ii) is the number of $(D, x) \in \mathcal{D} \times X$ such that $x \notin D$, which is $n2^{n-1} - n$. To see this notice that for each $x \in X$ (there are n of them), the number of $D \neq \emptyset$ such that $x \notin D$ is $2^{n-1} - 1$. Since $|\mathcal{D}| \times |X| = (2^n - 1)n$, $\dim \mathcal{P}_r = (2^n - 1)n - (2^n - 1) - (n2^{n-1} - n) = (n-2)2^{n-1} + 1$.

A.5 Proof of Theorem 1

To show the necessity of Quasi-Stochastic Rationality, fix any $q \in \Delta(\mathcal{D})$ and any $u(D, \cdot) \in \mathbf{R}^D$ for each $D \in \mathcal{D}$ such that $u(D, \cdot)$ is not constant for some $D \in \mathcal{D}$ with $q(D) > 0$. By Lemma 5 (i), if $u(D, \cdot)$ is not constant for some $D \in \mathcal{D}$ with $q(D) > 0$, then $E(\rho^\pi : q, u) \neq E(\rho^{\pi'} : q, u)$ for some $\pi, \pi' \in \Pi$. By Proposition 1 and Lemma 3, any $\rho \in \mathcal{P}_{ml}$ is rationalized by full support $\nu \in \Delta(\Pi)$. Then, $E(\rho : q, u) = \sum_{\pi\in\Pi} \nu(\pi)E(\rho^\pi : q, u) > \min_{\pi\in\Pi} E(\rho^\pi : q, u)$.

To show the sufficiency of Quasi-Stochastic Rationality, assume that ρ satisfies Quasi-Stochastic Rationality. I will show that $\rho \in \mathcal{P}_{ml}$. By Lemma 2, there exist a set $\{t_i\}_{i=1}^n \subset \mathbf{R}^{\mathcal{D}\times X} \setminus \{0\}$ and a set $\{\alpha_i\}_{i=1}^n \subset \mathbf{R}$ such that $\mathcal{P}_r = \cap_{i=1}^n \{p \in \mathbf{R}^{\mathcal{D}\times X} | p \cdot t_i \geq \alpha_i\} \cap \text{aff}.\mathcal{P}_r$ and aff.$\mathcal{P}_r \not\subset \{p \in \mathbf{R}^{\mathcal{D}\times X} | p \cdot t_i \geq \alpha_i\}$ for all $i \in \{1, \ldots, n\}$. Since rint.$\mathcal{P}_r \neq \emptyset$, then by Theorem 6.5 of Rockafellar (2015), rint.$\mathcal{P}_r = \cap_{i=1}^n \text{rint}.\{p \in \mathbf{R}^{\mathcal{D}\times X} | p \cdot t_i \geq \alpha_i\} \cap \text{aff}.P_r = \cap_{i=1}^n \{p \in \mathbf{R}^{\mathcal{D}\times X} | p \cdot t_i > \alpha_i\} \cap \text{aff}.P_r$. By Proposition 3, $\mathcal{P}_r \subset \mathcal{P} \subset \text{aff}.\mathcal{P}_r$. Thus

$$\begin{aligned}
\mathcal{P}_r &= \mathcal{P}_r \cap \mathcal{P} && (\because \mathcal{P}_r \subset \mathcal{P})\\
&= \cap_{i=1}^n \text{rint}.\{p \in \mathbf{R}^{\mathcal{D}\times X} | p \cdot t_i \geq \alpha_i\} \cap \text{aff}.P_r \cap \mathcal{P} &&\\
&= \cap_{i=1}^n \text{rint}.\{p \in \mathbf{R}^{\mathcal{D}\times X} | p \cdot t_i \geq \alpha_i\} \cap \mathcal{P} && (\because \mathcal{P} \subset \text{aff}.\mathcal{P})\\
&= \cap_{i=1}^n \{\rho \in \mathcal{P} | \rho \cdot t_i \geq \alpha_i\}.
\end{aligned}$$

Hence

$$\mathcal{P}_r = \cap_{i=1}^n \{\rho \in \mathcal{P} | \rho \cdot t_i \geq \alpha_i\}. \tag{12}$$

This implies that rint.$\mathcal{P}_r = \cap_{i=1}^n \{\rho \in \mathcal{P} | \rho \cdot t_i > \alpha_i\}$. Since Proposition 2 sates $\mathcal{P}_{ml} = \text{rint.}\mathcal{P}_r$,

$$\mathcal{P}_{ml} = \cap_{i=1}^n \{\rho \in \mathcal{P} | \rho \cdot t_i > \alpha_i\}. \tag{13}$$

Fix any $i \in \{1, \ldots, n\}$. I will show that there exist $\pi, \pi' \in \Pi$ such that $\rho^\pi \cdot t_i \neq \rho^{\pi'} \cdot t_i$. Suppose, by way of contradiction, that for all $\pi, \pi' \in \Pi$, $\rho^\pi \cdot t_i = \rho^{\pi'} \cdot t_i$. Let $\alpha_i' \equiv \rho^\pi \cdot t_i$ for some $\pi \in \Pi$. Since $\rho^\pi \in \mathcal{P}_r$ and (12) holds, I have $\alpha_i' \geq \alpha_i$. Then, aff.$\mathcal{P}_r$ = aff.co.$\{\rho^\pi | \pi \in \Pi\}$ = aff.$\{\rho^\pi | \pi \in \Pi\} \subset \{p \in \mathbf{R}^{\mathcal{D} \times X} | p \cdot t_i = \alpha_i'\} \subset \{p \in \mathbf{R}^{\mathcal{D} \times X} | p \cdot t_i \geq \alpha_i\}$. This is a contradiction.

By Lemma 5 (ii), the existence of $\pi, \pi' \in \Pi$ such that $\rho^\pi \cdot t_i \neq \rho^{\pi'} \cdot t_i$ implies that $t_i(D, \cdot)$ is nonconstant for some $D \in \mathcal{D}$. For any $(D, x) \in \mathcal{D} \times X$ such that $x \in D$, define $u_i(D, x) = t_i(D, x)$, so that $u_i(D, \cdot) \in \mathbf{R}^D$. Note also that $u_i(D, \cdot)$ is nonconstant for some $D \in \mathcal{D}$ with $q(D) > 0$. In addition, by Proposition 3, for any $p \in$ aff.$\mathcal{P}_r$, $p(D, x) = 0$ for any $D \in \mathcal{D}$ and $x \notin D$. Therefore, for any $p \in$ aff.$\mathcal{P}_r$,

$$\sum_{D \in \mathcal{D}} \sum_{x \in D} u_i(D, x) p(D, x) = p \cdot t_i. \tag{14}$$

Define $q \in \Delta(\mathcal{D})$ by $q(D) = 1/|\mathcal{D}|$ for any $D \in \mathcal{D}$. Since $\rho^\pi \in \mathcal{P}_r$, then by (12), $\rho^\pi \cdot t_i \geq \alpha_i$ for any $\pi \in \Pi$. Hence, for any $\pi \in \Pi$

$$\begin{aligned} E(\rho^\pi : q, u_i) &= \textstyle\sum_{D \in \mathcal{D}} q(D) \sum_{x \in D} u_i(D, x) \rho^\pi(D, x) \\ &= \left(\textstyle\sum_{D \in \mathcal{D}} \sum_{x \in D} u_i(D, x) \rho^\pi(D, x) \right) / |\mathcal{D}| \\ &= (\rho^\pi \cdot t_i)/|\mathcal{D}| \qquad\qquad (\because (14)) \\ &\geq \alpha_i/|\mathcal{D}|. \end{aligned}$$

Hence, $\min_{\pi \in \Pi} E(\rho^\pi : q, u_i) \geq \alpha_i/|\mathcal{D}|$ for all $i \in \{1, \ldots, n\}$. Moreover, by Quasi-Stochastic Rationality, $E(\rho : q, u_i) > \min_{\pi \in \Pi} E(\rho^\pi : q, u_i)$, so that $E(\rho : q, u_i) > \alpha_i/|\mathcal{D}|$ for all $i \in \{1, \ldots, n\}$. By (14) $\rho \cdot t_i = \sum_{D \in \mathcal{D}} \sum_{x \in D} u_i(D, x) \rho(D, x) \equiv |\mathcal{D}| E(\rho : q, u_i) > |\mathcal{D}| \alpha_i/|\mathcal{D}| = \alpha_i$ for all $i \in \{1, \ldots, n\}$. Therefore, $\rho \in \cap_{i=1}^n \{\rho \in \mathcal{P} | \rho \cdot t_i > \alpha_i\} = \mathcal{P}_{ml}$ by (13).

A.6 Proof of Theorem 2

First I will show the necessity of the positivity of the Block-Marschak polynomials. I show that if $\rho \in \mathcal{P}_{ml}$, then $K(\rho, D, x) > 0$ for any $(D, x) \in \mathcal{D} \times X$ such that $x \in D$.

By Proposition 1 (i), $\rho \in \text{co}.\mathcal{P}_l$. Since $K(\alpha\rho + (1-\alpha)\rho', D, x) = \alpha K(\rho, D, x) + (1-\alpha)K(\rho', D, x)$, it suffices to show that $K(\rho, D, x) > 0$ for any $\rho \in \mathcal{P}_l$ and any $(D, x) \in \mathcal{D} \times X$ such that $x \in D$. Fix $\rho \in \mathcal{P}_l$ and $(D, x) \in \mathcal{D} \times X$ such that $x \in D$. By Theorem 2.1 in Barberá and Pattanaik (1986), $K(\rho, D, x) = \nu(\{\pi \in \Pi | \pi(D^c) > \pi(x) \geq \pi(D)\})$. Then by Lemma 3, there exists $\nu \in \Delta(\Pi)$ such that ν rationalizes ρ and $\nu(\pi) > 0$ for all $\pi \in \Pi$. Since $x \in D$, the set $\{\pi \in \Pi | \pi(D^c) > \pi(x) \geq \pi(D)\}$ is nonempty. Hence, $K(\rho, D, x) = \nu(\{\pi \in \Pi | \pi(D) > \pi(x) \geq \pi(D^c)\}) > 0$.

Next I will show the sufficiency of the positivity of the Block-Marschak polynomials. Fix $\rho \in \mathcal{P}$ and assume that $K(\rho, D, x) > 0$ for any $(D, x) \in \mathcal{D} \times X$ such that $x \in D$. By the axiomatization of Falmagne (1978), $\rho \in \mathcal{P}_r$. Since Proposition 2 states that $\mathcal{P}_{ml} = \text{rint}.\mathcal{P}_r$, it suffices to show that $\rho \in \text{rint}.\mathcal{P}_r$.

Choose any $\rho' \in \mathcal{P}_r$ to show that there exists $\alpha > 1$ such that $\alpha\rho + (1-\alpha)\rho' \in \mathcal{P}_r$ by the following three steps. (Remember that the existence of such α means that $\rho \in \text{rint}.\mathcal{P}_r$.)

Step 1: $\rho(D, x) > 0$ for any $(D, x) \in \mathcal{D} \times X$ such that $x \in D$. Suppose by way of contradiction that $\rho(D, x) = 0$ for some $(D, x) \in \mathcal{D} \times X$ such that $x \in D$. Then for any $E \supset D$, $\rho(E, x) \leq \rho(D, x) = 0$ because $\rho \in \mathcal{P}_r$.[21] Then by definition, $K(\rho, D, x) = 0$. This is a contradiction.

Step 2: There exists $\overline{\alpha} > 1$ such that, for any $\alpha \in (1, \overline{\alpha})$, $\alpha\rho + (1-\alpha)\rho' \in \mathcal{P}$. To prove the step, fix $(D, x) \in \mathcal{D} \times X$ such that $x \in D$. Since Step 1 has shown that $\rho(D, x) > 0$, there exists $\overline{\alpha}(D, x) > 1$ such that, for any $\alpha \in (1, \overline{\alpha}(D, x))$, $(\alpha\rho + (1-\alpha)\rho')(D, x) = \rho(D, x) + (\alpha - 1)(\rho(D, x) - \rho'(D, x)) > 0$. Define $\overline{\alpha} \equiv \min_{(D,x) \in \mathcal{D} \times X : x \in D} \overline{\alpha}(D, x)$. Since there are finitely many pairs (D, x) such that $x \in D$, such $\overline{\alpha}$ exists. The definition of $\overline{\alpha}$ shows that $\overline{\alpha} > 1$ and $\overline{\alpha}$ satisfies the desired property.

Step 3: There exists $\hat{\alpha} > 1$ such that, for any $\alpha \in (1, \hat{\alpha})$, $K(\alpha\rho + (1-\alpha)\rho', D, x) > 0$ for any $(D, x) \in \mathcal{D} \times X$ such that $x \in D$. To prove this step, fix any $(D, x) \in \mathcal{D} \times X$ such that $x \in D$. Since $K(\rho, D, x) > 0$, there exists $\hat{\alpha}(D, x) > 1$ such that, for any $\alpha \in (1, \hat{\alpha}(D, x))$, $K(\alpha\rho + (1-\alpha)\rho', D, x) = K(\rho, D, x) + (\alpha - 1)(K(\rho, D, x) - K(\rho', D, x)) > 0$. Define $\hat{\alpha} \equiv \min_{(D,x) \in \mathcal{D} \times X : x \in D} \hat{\alpha}(D, x)$. Since there are finitely many pairs (D, x) such that $x \in D$, such $\hat{\alpha}$ exists. The definition of $\hat{\alpha}$ shows that $\hat{\alpha} > 1$ and $\hat{\alpha}$ satisfies the desired property.

Now choose α such that $1 < \alpha < \min\{\overline{\alpha}, \hat{\alpha}\}$. Then, by Steps 2 and 3, $\alpha\rho + (1 -$

[21] A random utility function $\rho \in \mathcal{P}_r$ satisfies the following property: if $x \in D \subset E$, then $\rho(E, x) \leq \rho(D, x)$ .This property is called *regularity*, or *monotonicity*.

$\alpha)\rho' \in \mathcal{P}$ and $K(\alpha\rho + (1-\alpha)\rho')(D, x) > 0$ for any $(D, x) \in \mathcal{D} \times X$ such that $x \in D$. Then, by the axiomatization of Falmagne (1978), $\alpha\rho + (1-\alpha)\rho' \in \mathcal{P}_r$.

A.7 Proof of Proposition 4

Let $n \equiv |X|$ and $X = \{x_1, \ldots, x_n\}$. For any ranking $\pi \in \Pi$, consider the following condition: if $\sum_{i=1}^{n-1} \lambda_i(\pi^{-1}(n+1-i) - \pi^{-1}(n-i)) = 0$ and $\lambda_i \geq 0$ for all $i \in \{1, \ldots, n-1\}$, then $\lambda_i = 0$ for all $i \in \{1, \ldots, n-1\}$. I call this condition as Condition $(*)$.

Step 1: For each $\pi \in \Pi$, Condition $(*)$ holds if and only if there exists $\beta \in \mathbf{R}^k$ such that for any $x, y \in X$, $\pi(x) > \pi(y) \Longleftrightarrow \beta \cdot x > \beta \cdot y$.

Proof. Fix $\pi \in \Pi$.

$$
\begin{aligned}
& \exists \beta \in \mathbf{R}^k \;\; \beta \cdot \pi^{-1}(n) > \beta \cdot \pi^{-1}(n-1) > \cdots > \beta \cdot \pi^{-1}(2) > \beta \cdot \pi^{-1}(1) \\
\Longleftrightarrow \;\; & \exists \beta \in \mathbf{R}^k \;\; \beta \cdot (\pi^{-1}(n) - \pi^{-1}(n-1)) > 0, \ldots, \beta \cdot (\pi^{-1}(2) - \pi^{-1}(1)) > 0 \\
\Longleftrightarrow \;\; & \nexists \lambda \in \mathbf{R}^{n-1} \;\; \textstyle\sum_{i=1}^{n-1} \lambda_i(\pi^{-1}(n+1-i) - \pi^{-1}(n-i)) = 0, \lambda \geq 0, \text{ and } \lambda \neq 0 \\
\Longleftrightarrow \;\; & \text{Condition}(*),
\end{aligned}
$$

where the second to the last equivalence is by Lamme 9 with $\mathcal{F} = \mathbf{R}$ in section A.10. □

Step 2: X is affinely independent if and only if Condition $(*)$ holds for any $\pi \in \Pi$.

Proof. I first show that if X is affinely independent then Condition $(*)$ holds for any ranking $\pi \in \Pi$. Fix any $\pi \in \Pi$. Without loss of generality assume that $\pi(x_i) = n + 1 - i$ for all $i \in \{1, \ldots, n\}$. Suppose that $\sum_{i=1}^{n-1} \lambda_i(\pi^{-1}(n+1-i) - \pi^{-1}(n-i)) \equiv \sum_{i=1}^{n-1} \lambda_i(x_i - x_{i+1}) = 0$ and $\lambda_i \geq 0$ for all i. Define $\mu_1 = \lambda_1$, $\mu_i = \lambda_i - \lambda_{i-1}$ for all $i \in \{2, \ldots, n-1\}$, and $\mu_n = -\lambda_{n-1}$. Then $\sum_{i=1}^{n-1} \lambda_i(x_i - x_{i+1}) = \lambda_1 x_1 + \sum_{i=2}^{n-1}(\lambda_i - \lambda_{i-1})x_i + (-\lambda_{n-1})x_n = \mu_1 x_1 + \sum_{i=2}^{n-1} \mu_i x_i + \mu_n x_n = \sum_{i=1}^{n} \mu_i x_i$. Since $\sum_{i=1}^{n-1} \lambda_i(x_i - x_{i+1}) = 0$, I have $\sum_{i=1}^{n} \mu_i x_i = 0$. Moreover, $\sum_{i=1}^{n} \mu_i = \lambda_1 + \sum_{i=2}^{n-1}(\lambda_i - \lambda_{i-1}) + (-\lambda_{n-1}) = 0$. If X is affinely independent, then $\mu_i = 0$ for all $i \in \{1, \ldots, n\}$. Hence, $\lambda_i = 0$ for all $i \in \{1, \ldots, n-1\}$.

Next I will show that if Condition $(*)$ holds for any $\pi \in \Pi$ then X is affinely independent. Choose any real numbers $\{\mu_i\}_{i=1}^n$ such that $\sum_{i=1}^{n} \mu_i x_i = 0$ and $\sum_{i=1}^{n} \mu_i = 0$ to show $\mu_i = 0$ for all $i \in \{1, \ldots, n\}$. Order μ_i by its value. Without loss of generality assume that $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n$. If $\mu = 0$, then the proof is finished. If $\mu \neq 0$ then $\mu_1 > 0$. For each $x_i \in X$, define $\pi(x_i) = n + 1 - i$. Then $\pi \in \Pi$.

Define $\lambda_1 = \mu_1$ and $\lambda_i = \sum_{j=1}^{i} \mu_j$ for all $i \in \{2, \ldots, n-1\}$. Then $\lambda \neq 0$ because $\mu_1 > 0$. I will show that $\lambda_i \geq 0$ for all $i \in \{1, \ldots, n-1\}$. Suppose by way of contradiction that $\lambda_i < 0$ for some i. Then $\mu_i < 0$ because $\mu_1 \geq \cdots \geq \mu_i$. Since $0 > \mu_i \geq \mu_j$ for all $j \geq i$, I have $\sum_{j=i+1}^{n} \mu_j < 0$. It follows that $\sum_{j=1}^{n} \mu_j = \lambda_i + \sum_{j=i+1}^{n} \mu_j < 0$. This contradicts that $\sum_{i=1}^{n} \mu_i = 0$. Therefore, $\lambda_i \geq 0$ for all $i \in \{1, \ldots, n-1\}$. Moreover $\sum_{i=1}^{n-1} \lambda_i(\pi^{-1}(n+1-i) - \pi^{-1}(n-i)) = \sum_{i=1}^{n-1} \lambda_i(x_i - x_{i+1}) = \lambda_1 x_1 + \sum_{i=2}^{n-1}(\lambda_i - \lambda_{i-1})x_i + (-\lambda_{n-1})x_n = \mu_1 x_1 + \sum_{i=2}^{n-1} \mu_i x_i + (-\sum_{i=1}^{n-1} \mu_i)x_n = \sum_{i=1}^{n} \mu_i x_i = 0$, where the second to the last equality holds because $\sum_{i=1}^{n} \mu_i = 0$.Therefore, by Condition $(*)$, $\lambda_i = 0$ for all $i \in \{1, \ldots, n-1\}$. Hence, $\mu_i = 0$ for all $i \in \{1, \ldots, n\}$. □

A.8 Proof of Proposition 5

To prove Proposition 5, I prove two lemmas. To simplify the notation, define Π^* as the set of linearly representable rankings. Notice that Theorem 4 states $\Pi = \Pi^*$ if and only if X is affinely independent.

Lemma 6. *Let X be a finite subset of $\mathbf{R}^k$. For any $\pi \in \Pi$, $\pi \in \Pi^*$ if and only if there exists a sequence $\{\beta_n\}_{n=1}^{\infty} \subset \mathbf{R}^k$ such that $\rho_{\beta_n} \to \rho^\pi$ as $n \to \infty$.*

Proof. Choose any $\pi \in \Pi^*$. Without loss of generality, assume that $X = \{x_1, \ldots, x_{|X|}\}$ and $\pi(x_1) > \pi(x_2) > \cdots > \pi(x_{|X|})$. Since $\pi \in \Pi^*$, there exists $\beta \in \mathbf{R}^k$ such that $\beta \cdot x_1 > \beta \cdot x_2 > \cdots > \beta \cdot x_{|X|}$. For any positive integer k and any $(D, x) \in \mathcal{D} \times X$ such that $x \in D$,

$$\begin{aligned} \rho_{k\beta}(D, x) &\equiv \frac{\exp(k\beta \cdot x)}{\sum_{y \in D} \exp(k\beta \cdot y)} \\ &= \frac{1}{\sum_{y \in D: \pi(y) > \pi(x)} \exp(k\beta \cdot (y - x)) + 1 + \sum_{y \in D: \pi(y) < \pi(x)} \exp(k\beta \cdot (y - x))}. \end{aligned}$$

For any $y \in D$, $\pi(y) > \pi(x)$ if and only if $\beta \cdot (y - x) > 0$. Therefore, as $k \to \infty$, if $\pi(x) \geq \pi(D)$, then $\rho_{k\beta}(D, x) \to 1$; if $\pi(x) < \pi(D)$, then $\rho_{k\beta}(D, x) \to 0$. Hence, $\rho_{k\beta} \to \rho^\pi$ as $k \to \infty$.

To show the converse, fix a sequence $\{\beta_n\}_{n=1}^{\infty}$ such that $\rho_{\beta_n} \to \rho^\pi$ as $n \to \infty$. For any $D \in \mathcal{D}$ and $x \in D$, notice that

$$\rho_{\beta_n}(D, x) = \frac{1}{1 + \sum_{y \in D \setminus x} \exp(\beta_n \cdot (y - x))}.$$

Let $\pi(x) \geq \pi(D)$. Since $\rho_{\beta_n} \to \rho^\pi$ as $n \to \infty$, it must hold that $\beta_n \cdot (y - x) \to -\infty$ as $n \to \infty$ for all $y \in D \setminus \{x\}$. Therefore, for each $D \in \mathcal{D}$ there exists $\overline{n}(D)$ such that for all $n > \overline{n}(D)$ and all $y \in D \setminus \{x\}$, $\beta_n \cdot x > \beta_n \cdot y$, where $\pi(x) \geq \pi(D)$.

Without loss of generality assume that $X = \{x_1, \ldots, x_{|X|}\}$ and $\pi(x_1) > \pi(x_2) > \cdots > \pi(x_{|X|})$. Let $n > \max\{\overline{n}(X), \overline{n}(\{x_i\}_{i=2}^{|X|}), \ldots, \overline{n}(\{x_i\}_{i=|X|-1}^{|X|})\}$. Then, $\beta_n \cdot x_1 > \beta_n \cdot x_2 > \cdots > \beta_n \cdot x_{|X|-1} > \beta_n \cdot x_{|X|}$. Therefore, $\pi \in \Pi^*$. □

Lemma 7. *For any $\pi \in \Pi$, if there exist strictly positive numbers $\{\lambda_i\}_{i=1}^m$ and a sequence $\{\beta_n^i\} \subset \mathbf{R}^k$ for all $i \in \{1, \ldots, m\}$ such that $\sum_{i=1}^m \lambda_i = 1$ and $\sum_{i=1}^m \lambda_i \rho_{\beta_n^i} \to \rho^\pi$ as $n \to \infty$, then $\rho_{\beta_n^i} \to \rho^\pi$ as $n \to \infty$ for all $i \in \{1, \ldots, m\}$.*

Proof. As in the proof of Lemma 6,

$$\sum_{i=1}^m \lambda_i \rho_{\beta_n^i}(D, x) = \sum_{i=1}^m \frac{\lambda_i}{1 + \sum_{y \in D \setminus x} \exp(\beta_n^i \cdot (y - x))}.$$

Let $\pi(x) > \pi(y)$ for all $y \in D \setminus \{x\}$. Since $\sum_{i=1}^m \lambda_i \rho_{\beta_n^i} \to \rho^\pi$ as $n \to \infty$ and $\lambda_i > 0$ for all $i \in \{1, \ldots, m\}$, it must hold that $\beta_n^i \cdot (y - x) \to -\infty$ as $n \to \infty$ for all $i \in \{1, \ldots, m\}$. Therefore, $\rho_{\beta_n^i} \to \rho^\pi$ as $n \to \infty$ for all $i \in \{1, \ldots, m\}$. □

In the following, I prove Proposition 5. First I will show that if X is affinely independent, then $\mathcal{P}_{mll} = \text{rint}.\mathcal{P}_r$. Since Proposition 1 (ii) states $\mathcal{P}_{mll} = \text{co}.\mathcal{P}_{ll}$, it suffices to show that $\text{co}.\mathcal{P}_{ll} = \text{rint}.\mathcal{P}_r$ assuming X is affinely independent.

Since $\mathcal{P}_{ll} \subset \mathcal{P}_l$ and Proposition 2 states $\text{co}.\mathcal{P}_l \subset \text{rint}.\mathcal{P}_r$, it follows that $\text{co}.\mathcal{P}_{ll} \subset \text{rint}.\mathcal{P}_r$. I now show that $\text{rint}.\mathcal{P}_r \subset \text{co}.\mathcal{P}_{ll}$ by applying Lemma 4 with $\mathcal{Q} = \mathcal{P}_{ll}$. The conditions of Lemma 4 are satisfied because of Proposition 4 and Lemma 6. In fact, these two results jointly show that for any ranking $\pi \in \Pi$, there exists a sequence $\{\rho_{\beta_n}\}_{n=1}^\infty$ of linear logit functions such that $\rho_{\beta_n} \to \rho^\pi$ as $n \to \infty$. Therefore, $\text{rint}.\mathcal{P}_r \subset \text{co}.\mathcal{P}_{ll}$.

To show the converse, assume now that X is not affinely independent. Suppose by way of contradiction that $\text{rint}.\mathcal{P}_r = \text{co}.\mathcal{P}_{ll}$. Then

$$\mathcal{P}_r = \text{cl}.\mathcal{P}_r = \text{cl}.\text{rint}.\mathcal{P}_r = \text{cl}.\text{co}.\mathcal{P}_{ll} = \text{co}.\text{cl}.\mathcal{P}_{ll}, \tag{15}$$

where the first equality holds because $\mathcal{P}_r$ is closed, the second equality holds by Theorem 6.3 of Rockafellar (2015), and the last equality holds because $\mathcal{P}_{ll}$ is bounded and by Theorem 17.2 of Rockafellar (2015).

Since X is not affinely independent, then by Proposition 4, there exists $\pi \in \Pi \setminus \Pi^*$. Moreover, by (15), $\rho^\pi \in \mathcal{P}_r = \text{co}.\text{cl}.\mathcal{P}_{ll}$. Then, there exist positive numbers

$\{\lambda_i\}_{i=1}^m$ such that $\sum_{i=1}^m \lambda_i = 1$ and sequences $\{\beta_n^i\}_{n=1}^\infty$ for each $i \in \{1,\dots,m\}$ such that $\sum_{i=1}^m \lambda_i \rho_{\beta_n^i} \to \rho^\pi$ as $n \to \infty$. It follows from Lemma 7 that $\rho_{\beta_n^i} \to \rho^\pi$ as $n \to \infty$ for all $i \in \{1,\dots,m\}$. Then, by Lemma 6, $\pi \in \Pi^*$, which is a contradiction.

A.9 Proof of Theorem 3

The proofs of Theorems 1 and 2 depend on the use of the mixed logit functions only because the set of mixed logit functions is the relative interior of the set of random utility functions (i.e., $\mathcal{P}_{ml} = \text{rint}.\mathcal{P}_r$).

Proposition 5 shows that X is affinely independent if and only if the set of mixed *linear* logit functions is the relative interior of the set of random utility functions (i.e., $\mathcal{P}_{mll} = \text{rint}.\mathcal{P}_r$). Hence, Theorem 1 and Proposition 5 prove the equivalence between (i) and (ii). Moreover, Theorem 2 and Proposition 5 prove the equivalence between (i) and (iii).

A.10 Theorems of Alternatives

In Theorem 3.2, Fishburn (2015) states the following result.

Lemma 8. *Let A be an $r \times n$ matrix, B be an $l \times n$ matrix, and E be an $m \times n$ matrix. Suppose that the entries of the matrices A, B, and E are rational numbers. Exactly one of the following alternatives is true.*

1. *There is $u \in \mathbf{R}^n$ such that $A \cdot u = 0$, $B \cdot u \geq 0$, and $E \cdot u \gg 0$.*
2. *There is $\theta \in \mathbf{Z}^r$, $\eta \in \mathbf{Z}^l$, and $\pi \in \mathbf{Z}^m$ such that $\theta \cdot A + \eta \cdot B + \pi \cdot E = 0$; $\pi > 0$ and $\eta \geq 0$.*

In Theorem 1.6.1, Stoer and Witzgall (2012) show the following result.

Lemma 9. *Let $\mathcal{F}$ be a field. Let A be an $r \times n$ matrix, B be an $l \times n$ matrix, and E be an $m \times n$ matrix. Suppose that the entries of the matrices A, B, and E belong to a commutative ordered field $\mathcal{F}$. Exactly one of the following alternatives is true.*

1. *There is $u \in \mathcal{F}^n$ such that $A \cdot u = 0$, $B \cdot u \geq 0$, $E \cdot u \gg 0$.*
2. *There is $\theta \in \mathcal{F}^r$, $\eta \in \mathcal{F}^l$, and $\pi \in \mathcal{F}^m$ such that $\theta \cdot A + \eta \cdot B + \pi \cdot E = 0$; $\pi > 0$ and $\eta \geq 0$.*

By Lemmas 8 and 9, I prove the following lemma.

Lemma 10. *Let A be an $r \times n$ matrix, B be an $l \times n$ matrix, and E be an $m \times n$ matrix. Suppose that the entries of the matrices A, B, and E are rational numbers. The followings are equivalent*

1. *There is $u \in \mathbf{R}^n$ such that $A \cdot u = 0$, $B \cdot u \geq 0$, and $E \cdot u \gg 0$.*
2. *There is $u \in \mathbf{Z}^n$ such that $A \cdot u = 0$, $B \cdot u \geq 0$, and $E \cdot u \gg 0$.*

Proof. By the supposition, the entries of the matrices A, B, and E are rational numbers. Then

$$\begin{aligned}
& \exists u \in \mathbf{R}^n \ [A \cdot u = 0, B \cdot u \geq 0, E \cdot u \gg 0.] \\
\iff & \neg\big[\exists \theta \in \mathbf{Z}^r, \eta \in \mathbf{Z}^l, \pi \in \mathbf{Z}^m \ [\theta \cdot A + \eta \cdot B + \pi \cdot E = 0; \pi > 0; \eta \geq 0.]\big] \\
& (\because \text{Lemma 8}) \\
\iff & \neg\big[\exists \theta \in \mathbf{Q}^r, \eta \in \mathbf{Q}^l, \pi \in \mathbf{Q}^m \ [\theta \cdot A + \eta \cdot B + \pi \cdot E = 0; \pi > 0; \eta \geq 0.]\big] \\
\iff & \exists u \in \mathbf{Q}^n \ A \cdot u = 0, B \cdot u \geq 0, E \cdot u \gg 0. \quad (\because \text{Lemma 9 with } \mathcal{F} = \mathbf{Q}) \\
\iff & \exists u \in \mathbf{Z}^n \ A \cdot u = 0, B \cdot u \geq 0, E \cdot u \gg 0,
\end{aligned}$$

where I obtain the second equivalence by dividing by a positive integer; and the last equivalence by multiplying by a positive integer. □

B Relationship with Gul et al. (2014)

Gul et al. (2014) axiomatize the complete attribute rule under strong richness assumption. Neither the complete attribute rule nor the mixed logit model is more general than the other. The intersection between the two models is the (degenerate) logit model.

Definition 9. *A random choice function ρ is called an* attribute rule *if there exists a set A of attributes, a function $w : A \to \mathbf{R}_{++}$ and $\eta : A \times X \to \mathbf{N} \cup \{0\}$ such that*

$$\rho(D, x) = \sum_{a \in A^x} \frac{w(a)}{\sum_{b \in A^D} w(b)} \frac{\eta_a(x)}{\sum_{y \in D} \eta_a(y)},$$

where $A^x = \{a \in A | \eta_a(x) > 0\}$ and $A^D = \bigcup_{x \in D} A^x$.

An element $x \in X$ is called an arhetype *for $a \in A$ if $A^x = \{a\}$ and $\eta_a(x) = 1$. An attribute rule is called* complete *if every attributes has at least two arhetypes.*

An attribute rule can be a convex combination of logit functions if for any $x, y \in X$, $A^x = A^y$. To see this define $A^* = A^x$. For any $(D, x) \in \mathcal{D} \times X$ and any

$a \in A$, define $\rho^a(D,x) = \eta_a(x)/(\sum_{y\in D}\eta_a(y))$. For any $a \in A$, ρ^a is a logit function. Moreover, if $A^x = A^*$ for any $x \in X$, we can define a probability measure m on $\{\rho^a\}_{a\in A^*}$ by $m(\rho^a) = w(a)/(\sum_{b\in A^*} w(b))$.

However, the assumption that $A^x = A^y$ for any $x, y \in X$ is compatible with their completeness assumption only when there is only one attribute (i.e., $A^x = A^y = \{a\}$ for any $x, y \in X$). This corresponds to the degenerate logit model.

Moreover, even besides the assumption of the completeness, since η can take only nonnegative integers, the set of attribute rules may not include the convex hull of the set of logit functions.

C Axiomatization by the Strict Axiom of Revealed Stochastic Preference

In this section, I provide an additional axiomatization of the mixed logit model by modifying the axiom provided by McFadden and Richter (1990).

Definition 10. *For any $\rho \in \mathcal{P}$ and any sequence $(D_i, x_i)_{i=1}^n \subset \mathcal{D} \times X$, define*

$$B((D_i,x_i)_{i=1}^n, \rho) = \max_{\pi\in\Pi} \sum_{i=1}^n 1(\pi(x_i) \geq \pi(D_i)) - \sum_{i=1}^n \rho(D_i, x_i).$$

McFadden and Richter (1990) show that a random choice function ρ is a random utility function if and only if $B((D_i,x_i)_{i=1}^n, \rho) \geq 0$ for any sequence $(D_i, x_i)_{i=1}^n$.

Given Theorem 2, one might suspect that by simply changing the weak inequality to the strict inequality, one could characterize the mixed logit model. This is false, because the resulting axiom is too strong. Instead, the sequence needs to be restricted in a certain way that excludes *redundant sequences.*

Definition 11. *A sequence $(D_i, x_i)_{i=1}^n$ of elements of $\mathcal{D} \times X$ is called* redundant *if there exists $D \in \{D_i\}_{i=1}^n$ such that for any $x, y \in D$, $|\{i \in \{1, \ldots, n\} | (D_i, x_i) = (D, x)\}| = |\{i \in \{1, \ldots, n\} | (D_i, x_i) = (D, y)\}|$. Otherwise, a sequence is called* nonredundant.

If a sequence $(D_i, x_i)_{i=1}^n$ is redundant, there exists $D \in \{D_i\}_{i=1}^n$ such that all of the elements in D must appear the same number of times in the sequence.

Definition 12. *A random choice function ρ is said to satisfy the* Strict Axiom of Revealed Stochastic Preference *if $B((D_i,x_i)_{i=1}^n, \rho) > 0$ for any nonredundant sequence $(D_i, x_i)_{i=1}^n$.*

Theorem 4.

(i) A random choice function ρ is a mixed logit function if and only if ρ satisfies the Strict Axiom of Revealed Stochastic Preference.

(ii) Let X be an affinely independent finite subset of $\mathbf{R}^k$. A random choice function ρ is a mixed linear logit function if and only if ρ satisfies the Strict Axiom of Revealed Stochastic Preference.

C.1 Proof of the Necessity of the Axiom

Let ρ be a mixed logit function. By Proposition 1 (i), $\rho \in \text{co}.\mathcal{P}_l$. Then by Lemma 3, there exists full support $\nu^* \in \Delta(\Pi)$ such that ν^* rationalizes ρ. Then, $\sum_{i=1}^n \rho(D_i, x_i) = \sum_{i=1}^n \nu^*(\{\pi \in \Pi | \pi(x_i) \geq \pi(D_i)\})$. Also, $\max_{\nu \in \Delta(\Pi)} \sum_{i=1}^n \nu(\{\pi \in \Pi | \pi(x_i) \geq \pi(D_i)\}) = \max_{\pi \in \Pi} \sum_{i=1}^n 1(\pi(x_i) \geq \pi(D_i))$ because the objective function is linear in ν and $\Delta(\Pi)$ is compact. Therefore, $B((D_i, x_i)_{i=1}^n, \rho) = \max_{\nu \in \Delta(\Pi)} \sum_{i=1}^n \nu(\{\pi \in \Pi | \pi(x_i) \geq \pi(D_i)\}) - \sum_{i=1}^n \nu^*(\{\pi \in \Pi | \pi(x_i) \geq \pi(D_i)\})$.

So to complete the proof I will show that $\nu^* \notin \arg\max_{\nu \in \Delta(\Pi)} \sum_{i=1}^n \nu(\{\pi \in \Pi | \pi(x_i) \geq \pi(D_i)\})$. Since ν^* is full support, it suffices to show that for any nonredundant sequence $(D_i, x_i)_{i=1}^n$, there exist $\pi, \pi' \in \Pi$ such that $\sum_{i=1}^n 1(\pi(x_i) \geq \pi(D_i)) \neq \sum_{i=1}^n 1(\pi'(x_i) \geq \pi'(D_i))$.

By way of contradiction suppose that there exist a nonredundant sequence $(D_i, x_i)_{i=1}^n$ and $\alpha \in \mathbf{R}$ such that $\sum_{i=1}^n 1(\pi(x_i) \geq \pi(D_i)) = \alpha$ for any $\pi \in \Pi$. For each $(D, x) \in \mathcal{D} \times X$ define $t(D, x) = |\{i \in \{1, \ldots, n\} | (D_i, x_i) = (D, x)\}|$. Then, $t \in \mathbf{R}^{\mathcal{D} \times X}$ and for each $\pi \in \Pi$, $t \cdot \rho^\pi = \sum_{i=1}^n \rho^\pi(D_i, x_i) = \sum_{i=1}^n 1(\pi(x_i) \geq \pi(D_i)) = \alpha$. Then by Lemma 5 (ii), $t(D, x) = t(D, y)$ for any $D \in \mathcal{D}$ and $x, y \in D$. This contradicts with the definition of the nonredundancy of $(D_i, x_i)_{i=1}^n$.

C.2 Proof of the Sufficiency of the Axiom

To show the result, I show three lemmas.

Lemma 11. *For any sequence $(D_i, x_i)_{i=1}^n$, if $B((D_i, x_i)_{i=1}^n, \rho) > 0$ for some $\rho \in \mathcal{P}$, then there exists a nonredundant subsequence $(D_j, x_j)_{i=1}^m$ of $(D_i, x_i)_{i=1}^n$ such that $B((D_i, x_i)_{i=1}^n, \rho) = B((D_j, x_j)_{j=1}^m, \rho)$ for all $\rho \in \mathcal{P}$.*

Proof. Fix a sequence $(D_i, x_i)_{i=1}^n$. Denote the sequence by $\mathcal{S}$. If the sequence is nonredundant, then I obtain the desired result by letting $(D_j, x_j)_{i=1}^m = (D_i, x_i)_{i=1}^n$.

If the sequence $\mathcal{S}$ is redundant, then exists $D' \in \{D_i\}_{i=1}^n$ such that for any $x, y \in D'$, $|\{i \in \{1, \ldots, n\} | (D_i, x_i) = (D', x)\}| = |\{i \in \{1, \ldots, n\} | (D_i, x_i) = (D', y)\}|$. Denote the set of such D' by $\mathcal{D}'$. For each $D' \in \mathcal{D}'$, construct subsequence $(D_j, x_j)_{j=1}^m$ of $(D_i, x_i)_{i=1}^n$ such that $D_j = D'$ for all $j \in \{1, \ldots, m\}$. Denote the subsequence by $\mathcal{S}(D')$. I obtain the subsequence $\mathcal{S}^*$ by removing all subsequences of $\{\mathcal{S}(D') | D' \in \mathcal{D}'\}$ from $\mathcal{S}$. If $\mathcal{S}^*$ is not empty, then $\mathcal{S}^*$ is a nonredundant sequence.

In the following I will show that $B(\mathcal{S}, \rho) = B(\mathcal{S}^*, \rho)$ for all $\rho \in \mathcal{P}$ and that $\mathcal{S}^*$ is a nonredundant sequence. By the definition of $\mathcal{D}'$, for any $D' \in \mathcal{D}'$ all elements of D' must appear the same number of times. Say it is $K(D')$ times. Since $\sum_{x \in D'} \rho(D', x) = 1$, I have $\sum_{(D_i, x_i) \in \mathcal{S}(\mathcal{D}')} \rho(D_i, x_i) = K(D')$. Moreover, $\sum_{(D_i, x_i) \in \mathcal{S}(D')} 1(\pi(x_i) \geq \pi(D_i)) = K(D')$. This is because for any $\pi \in \Pi$, $1(\pi(x) \geq \pi(D))$ is one if x is the best element and zero otherwise. Therefore

$$\begin{aligned} \sum_{(D_i, x_i) \in \mathcal{S}} \rho(D_i, x_i) &= \sum_{D' \in \mathcal{D}'} \sum_{(D_i, x_i) \in \mathcal{S}(\mathcal{D}')} \rho(D_i, x_i) + \sum_{(D_i, x_i) \in \mathcal{S}^*} \rho(D_i, x_i) \\ &= \sum_{D' \in \mathcal{D}'} K(D') + \sum_{(D_i, x_i) \in \mathcal{S}^*} \rho(D_i, x_i) \end{aligned}$$

and

$$\begin{aligned} & \max_{\pi \in \Pi} \sum_{(D_i, x_i) \in \mathcal{S}} 1(\pi(x_i) \geq \pi(D_i)) \\ = \; & \max_{\pi \in \Pi} \sum_{D' \in \mathcal{D}'} \sum_{(D_i, x_i) \in \mathcal{S}(\mathcal{D}')} 1(\pi(x_i) \geq \pi(D_i)) + \sum_{(D_i, x_i) \in \mathcal{S}^*} 1(\pi(x_i) \geq \pi(D_i)) \\ = \; & \sum_{D' \in \mathcal{D}'} K(D') + \max_{\pi \in \Pi} \sum_{(D_i, x_i) \in \mathcal{S}^*} 1(\pi(x_i) \geq \pi(D_i)). \end{aligned}$$

Hence $B(\mathcal{S}, \rho) = \max_{\pi \in \Pi} \sum_{(D_i, x_i) \in \mathcal{S}^*} 1(\pi(x_i) \geq \pi(D_i)) - \sum_{(D_i, x_i) \in \mathcal{S}^*} \rho(D_i, x_i) = B(\mathcal{S}^*, \rho)$. Since $B(\mathcal{S}, \rho) > 0$ for some $\rho \in \mathcal{P}$, the subsequence $\mathcal{S}^*$ is not empty. Thus the subsequence $\mathcal{S}^*$ is a desired nonredundant sequence. □

Lemma 12. *If a random choice function ρ satisfies the Strict Axiom of Revealed Stochastic Preference, then ρ satisfies the Axiom of Revealed Stochastic Preference.*

Proof. Fix a sequence $(D_i, x_i)_{i=1}^n$. Denote the sequence by $\mathcal{S}$. By the same argument as in the proof of Lemma 11, I obtain a subsequence $\mathcal{S}^*$ of $\mathcal{S}$ such that $B(\mathcal{S}, \rho) = B(\mathcal{S}^*, \rho)$ for any $\rho \in \mathcal{P}$. If $\mathcal{S}^*$ is empty, then $B(\mathcal{S}, \rho) = 0$ for any $\rho \in \mathcal{P}$. Moreover, if $\mathcal{S}^*$ is not empty, then it is a nonredundant sequence. Then, by the Strict Axiom of Revealed Stochastic Preference, $B(\mathcal{S}^*, \rho) > 0$, hence $B(\mathcal{S}, \rho) > 0$ for any $\rho \in \mathcal{P}$. Therefore, ρ satisfies the Axiom of Revealed Stochastic Preference. □

Lemma 13. *For any $s \in \mathbf{Z}^{\mathcal{D} \times X}$ and $\beta \in \mathbf{R}$, there exist $t \in \mathbf{Z}_+^{\mathcal{D} \times X}$ and $\alpha \in \mathbf{R}$ such that $\rho \cdot s < \beta$ if and only if $\rho \cdot t < \alpha$, where $\mathbf{Z}$ is the set of integers and $\mathbf{Z}_+$ is the set of nonnegative integers.*

Proof. I will construct a nonnegative integer t from s and a number α from β. To do this, set $t = s$ initially. If $s(D, y) < 0$ for some (D, y), then add $-s(D, y)$ to $t(D, x)$ for all $x \in X$. This transformation changes only the constant because $\sum_{x \in X} \rho(D, x) = 1$. Formally, for each (D, x) define

$$\begin{aligned} t(D,x) &= s(D,x) + \textstyle\sum_{y \in X : s(D,y)<0} (-s(D,y)) \\ \alpha &= \beta - \textstyle\sum_{(D,y) \in \mathcal{D} \times X : s(D,y) < 0} s(D,y). \end{aligned}$$

Then, t is a nonnegative integer vector. For any $\rho \in \mathcal{P}$,

$$\begin{aligned} \rho \cdot s &= \rho \cdot t - \textstyle\sum_{D \in \mathcal{D}} \sum_{x \in X} \rho(D,x) \sum_{y \in X : s(D,y) < 0} (-s(D,y)) \\ &= \rho \cdot t - \textstyle\sum_{D \in \mathcal{D}} \sum_{y \in X : s(D,y)<0} (-s(D,y)) \quad (\because \textstyle\sum_{x \in X} \rho(D,x) = 1) \\ &= \rho \cdot t + \textstyle\sum_{(D,y) \in \mathcal{D} \times X : s(D,y) < 0} s(D,y). \end{aligned}$$

Hence, $\rho \cdot s < \beta$ if and only if $\rho \cdot t < \alpha$. □

Lemma 14. *For any hyperplane H in $\mathbf{R}^{\mathcal{D} \times X}$ such that $\mathcal{P}_r \subset H^-$, there exist $t \in \mathbf{Z}_+^{\mathcal{D} \times X} \setminus \{0\}$ and $\alpha \in \mathbf{R}$ such that $H \cap \mathcal{P}_r = \{p \in \mathbf{R}^{\mathcal{D} \times X} | p \cdot t = \alpha\} \cap \mathcal{P}_r$ and rint.$H^- \cap \mathcal{P}_r = \{p \in \mathbf{R}^{\mathcal{D} \times X} | p \cdot t < \alpha\} \cap \mathcal{P}_r$.*

Proof. Since H is a hyperplane, there exist $s \in \mathbf{R}^{\mathcal{D} \times X} \setminus \{0\}$ and $\beta \in \mathbf{R}$ such that $H = \{p \in \mathbf{R}^{\mathcal{D} \times X} | p \cdot s = \beta\}$ and rint.$H^- = \{p \in \mathbf{R}^{\mathcal{D} \times X} | p \cdot s < \beta\}$. Since $\mathcal{P}_r$ is a polytope, $\mathcal{P}_r \cap H$ is also a polytope if it is not empty. There exist a (possibly empty) subset Π' of Π such that co.$\{\rho^\pi | \pi \in \Pi'\} = \{p \in \mathbf{R}^{\mathcal{D} \times X} | p \cdot s = \beta\} \cap \mathcal{P}_r$ and $\rho^\pi \cdot s < \beta$ for any $\pi \in \Pi \setminus \Pi'$.

Therefore, $\rho^\pi \cdot s = \beta$ for any $\pi \in \Pi'$ and $\rho^\pi \cdot s < \beta$ for any $\pi \in \Pi \setminus \Pi'$. I shall define matrices A and E such that the above inequalities hold if and only if $A \cdot (s, \beta)^T = 0$ and $E \cdot (s, \beta)^T \gg 0$, where $(s, \beta)^T$ denotes the transpose of (s, β).

The matrix A has one row for each $\pi \in \Pi'$; one column for each $(D, x) \in \mathcal{D} \times X$; and one last column. In the row corresponding to $\pi \in \Pi$, A has $\rho^\pi(D, x)$ at the column of $(D, x) \in \mathcal{D} \times X$. The entries of the last column are all -1. The matrix E has one row for each $\pi \in \Pi \setminus \Pi'$; one column for each $(D, x) \in \mathcal{D} \times X$; and one last column. In the row corresponding to $\pi \in \Pi \setminus \Pi'$, E has $-\rho^\pi(D, x)$ at the column of $(D, x) \in \mathcal{D} \times X$. The entries of the last column are all $+1$.

Then, $A \cdot (s, \beta)^T = 0$ and $E \cdot (s, \beta)^T \gg 0$. Moreover, since $\rho^\pi(\cdot) \in \{0, 1\}$ for any $\pi \in \Pi$, the entries of the matrices A and E are rational numbers. It follows from Lemma 10 that there exists $(t, \alpha) \in \mathbf{Z}^{\mathcal{D} \times X} \times \mathbf{R}$ such that $A \cdot (t, \alpha)^T = 0$ and $E \cdot (t, \alpha)^T \gg 0$. So $\rho^\pi \cdot t = \alpha$ for any $\pi \in \Pi'$ and $\rho^\pi \cdot t < \alpha$ for any $\pi \in \Pi \setminus \Pi'$.

Now I will show $\{p \in \mathbf{R}^{\mathcal{D}\times X} | p \cdot s < \beta\} \cap \mathcal{P}_r = \{p \in \mathbf{R}^{\mathcal{D}\times X} | p \cdot t < \alpha\} \cap \mathcal{P}_r$. Choose any $\rho \in \mathcal{P}_r$ such that $\rho \cdot s < \beta$. Then, there exists $\{\lambda_\pi\}_{\pi \in \Pi} \subset \mathbf{R}_+$ such that $\sum_{\pi\in\Pi} \lambda_\pi = 1$ and $\rho = \sum_{\pi\in\Pi} \lambda_\pi \rho^\pi$. Since $\rho \cdot s < \beta$, $\lambda_{\pi^*} > 0$ for some $\pi^* \in \Pi \setminus \Pi'$. Since $\rho^\pi \cdot t \leq \alpha$ for all $\pi \in \Pi$ and $\rho^{\pi^*} \cdot t < \alpha$, then $\rho \cdot t = \sum_{\pi\in\Pi} \lambda_\pi(\rho^\pi \cdot t) < \alpha$. This establishes $\{p \in \mathbf{R}^{\mathcal{D}\times X} | p \cdot s < \beta\} \cap \mathcal{P}_r \subset \{p \in \mathbf{R}^{\mathcal{D}\times X} | p \cdot t < \alpha\} \cap \mathcal{P}_r$. Since the argument can be reversed to obtain the other inclusion, $\{p \in \mathbf{R}^{\mathcal{D}\times X} | p\cdot s < \beta\}\cap\mathcal{P}_r = \{p \in \mathbf{R}^{\mathcal{D}\times X} | p\cdot t < \alpha\}\cap\mathcal{P}_r$. In a similar way, I can obtain $\{p \in \mathbf{R}^{\mathcal{D}\times X} | p\cdot s = \beta\}\cap\mathcal{P}_r = \{p \in \mathbf{R}^{\mathcal{D}\times X} | p \cdot t = \alpha\} \cap \mathcal{P}_r$.

Therefore $H \cap \mathcal{P}_r = \{p \in \mathbf{R}^{\mathcal{D}\times X} | p \cdot s = \beta\} \cap \mathcal{P}_r = \{p \in \mathbf{R}^{\mathcal{D}\times X} | p \cdot t = \alpha\} \cap \mathcal{P}_r$ and $\text{rint}.H^- \cap \mathcal{P}_r = \{p \in \mathbf{R}^{\mathcal{D}\times X} | p \cdot s < \beta\} \cap \mathcal{P}_r = \{p \in \mathbf{R}^{\mathcal{D}\times X} | p \cdot t < \alpha\} \cap \mathcal{P}_r$. □

By using the lemmas above, I will show the sufficiency of the axiom. By Lemma 2, there exist hyperplanes $\{H_i\}_{i=1}^n$ in $\mathbf{R}^{\mathcal{D}\times X}$ such that $\text{aff}.\mathcal{P}_r \not\subset H_i^-$ for each $i \in \{1, \dots, n\}$ and $\mathcal{P}_r = (\cap_{i=1}^n H_i^-) \cap \text{aff}.\mathcal{P}_r$. By Theorem 6.5 of Rockafellar (2015),

$$\text{rint}.\mathcal{P}_r = (\cap_{i=1}^n \text{rint}.H_i^-) \cap \text{aff}.\mathcal{P}_r. \tag{16}$$

For each hyperplane H_i, since $\text{aff}.\mathcal{P}_r \not\subset H_i^-$, I have $\mathcal{P}_r \not\subset H_i$. Since $\mathcal{P}_r = \text{co}.\{\rho^\pi | \pi \in \Pi\}$, there exists $\Pi'_i \subsetneq \Pi$ such that $\{\rho^\pi | \pi \in \Pi'_i\} \subset H_i \cap \mathcal{P}_r$ and $\{\rho^\pi | \pi \in \Pi \setminus \Pi'_i\} \subset \text{rint}.H_i^- \cap \mathcal{P}_r$.

By Lemma 14, for each hyperplane H_i, there exist $t_i \in \mathbf{Z}_+^{\mathcal{D}\times X} \setminus \{0\}$ and $\alpha_i \in \mathbf{R}$ such that $H_i \cap \mathcal{P}_r = \{p \in \mathbf{R}^{\mathcal{D}\times X} | p \cdot t_i = \alpha_i\} \cap \mathcal{P}_r$ and

$$\text{rint}.H_i^- \cap \mathcal{P}_r = \{p \in \mathbf{R}^{\mathcal{D}\times X} | p \cdot t_i < \alpha_i\} \cap \mathcal{P}_r. \tag{17}$$

This implies that for any $\pi' \in \Pi'_i$ and $\pi \in \Pi \setminus \Pi'_i$,

$$\rho^{\pi'} \cdot t_i = \alpha_i > \rho^\pi \cdot t_i. \tag{18}$$

For each hyperplane H_i, consider a sequence $(D_j, x_j)_{j=1}^{n_i}$ such that each (D, x) appears $t_i(D, x)$ times, where $n_i \equiv \sum_{(D,x)\in\mathcal{D}\times X} t_i(D, x)$. (The order of the pair in the sequence does not matter.) Then for each $\rho \in \mathcal{P}$, $\sum_{j=1}^{n_i} \rho(D_j, x_j) = \rho \cdot t_i$. By (18), $\max_{\pi\in\Pi} \sum_{j=1}^{n_i} 1(\pi(x_j) \geq \pi(D_j)) = \max_{\pi\in\Pi} \sum_{j=1}^{n_i} \rho^\pi(D_j, x_j) = \max_{\pi\in\Pi} \rho^\pi \cdot t_i = \alpha_i$. For any $\rho \in \mathcal{P}$, $B((D_j, x_j)_{j=1}^{n_i}, \rho) = \max_{\pi\in\Pi} \sum_{j=1}^{n_i} 1(\pi(x_j) \geq \pi(D_j)) - \sum_{j=1}^{n_i} \rho(D_j, x_j) = \alpha_i - \rho\cdot t_i$. Moreover, there exists $\pi \in \Pi\setminus\Pi'_i$ such that $B((D_j, x_j)_{j=1}^{n_i}, \rho^\pi) = \alpha_i - \rho^\pi \cdot t_i > 0$. Then by Lemma 11, I obtain a nonredundant sequence $(D'_j, x'_j)_{j=1}^{n'_i}$ such that $B((D'_j, x'_j)_{j=1}^{n'_i}, \rho) = B((D_j, x_j)_{j=1}^{n_i}, \rho)$ for all $\rho \in \mathcal{P}$. Hence, $B((D'_j, x'_j)_{j=1}^{n'_i}, \rho) > 0$ if and only if $\rho \cdot t_i < \alpha_i$ for all $\rho \in \mathcal{P}$.

Since $\mathcal{P}_r \subset \mathcal{P}$, for all $i \in \{1, \dots, n\}$

$$\{p \in \mathbf{R}^{\mathcal{D}\times X} | p \cdot t_i < \alpha_i\} \cap \mathcal{P}_r = \{\rho \in \mathcal{P} | B((D'_j, x'_j)_{j=1}^{n'_i}, \rho) > 0\} \cap \mathcal{P}_r. \qquad (19)$$

Suppose that a random choice function ρ satisfies the Strict Axiom of Revealed Stochastic Preference. So $B((D'_j, x'_j)_{j=1}^{n'_i}, \rho) > 0$ for all $i \in \{1, \dots, n\}$. Then by Lemma 12, ρ satisfies the Axiom of Revealed Stochastic Preference. By the result of McFadden and Richter (1990), $\rho \in \mathcal{P}_r$. Therefore,

$$\begin{aligned} \rho &\in \cap_{i=1}^n \{\rho \in \mathcal{P} | B((D'_j, x'_j)_{j=1}^{n'_i}, \rho) > 0\} \cap \mathcal{P}_r \\ &= \cap_{i=1}^n \{p \in \mathbf{R}^{\mathcal{D}\times X} | p \cdot t_i < \alpha_i\} \cap \mathcal{P}_r && (\because (19)) \\ &= \cap_{i=1}^n \text{rint}.H_i^- \cap \mathcal{P}_r && (\because (17)) \\ &\subset \cap_{i=1}^n \text{rint}.H_i^- \cap \text{aff}.\mathcal{P}_r \\ &= \text{rint}.\mathcal{P}_r && (\because (16)) \end{aligned}$$

So $\rho \in \text{rint}.\mathcal{P}_r$. It follows from Propositions 2 and 5 that statements (i) and (ii) hold.

D Axiomatization by Strict Coherency

Besides the axiomatizations by Falmagne (1978) and McFadden and Richter (1990), there is another axiomatization for the random utility model proposed by Clark (1996). The axiomatization by Clark (1996) is based on DeFinetti's *Coherency* condition. DeFinetti shows that if a function defined on a set of subsets satisfies Coherency then the function can be extended to a finitely additive probability measure on the smallest algebra that contains the subsets.

To introduce Coherency, for $\Pi' \subset \Pi$, let $I_{\Pi'}$ denote the indicator function on the set Π'. For any $f : \Pi \to \mathbf{R}$, $f \geq 0$ means that $f(\pi) \geq 0$ for all $\pi \in \Pi$.

Definition 13. *A random choice function ρ is* Coherent *if, for every sequence $\{(D_i, x_i)\}_{i=1}^m$ of $\mathcal{D} \times X$ such that $x_i \in D_i$ for all $i \in \{1, \dots, m\}$, and for every finite sequence of real numbers $\{\lambda_i\}_{i=1}^m$,*

$$\sum_{i=1}^m \lambda_i I_{\{\pi \in \Pi | \pi(x_i) \geq \pi(D_i)\}} \geq 0 \implies \sum_{i=1}^m \lambda_i \rho(D_i, x_i) \geq 0.$$

Based on the result of DeFinetti, Clark (1996) shows that a random choice function ρ is Coherent if and only if ρ is a random utility function.

To axiomatize the mixed logit model, I need to modify the axiom of Coherency. As in the previous section, changing the weak inequality to the strict inequality is not enough to characterize the mixed logit model. (The resulting axiom is too strong). I need to restrict the sequences.

Definition 14. *A sequence* $\{(D_i, x_i, \lambda_i)\}_{i=1}^m$ *of* $\mathcal{D} \times X \times \mathbf{R}$ *such that* $x_i \in D_i$ *is said to be* balanced *if, for every* $D \in \{D_i\}_{i=1}^m$ *and for every* $x, y \in D$,

$$\sum_{j \in \{i \in \{1,\ldots,m\} | (D_i, x_i) = (D, x)\}} \lambda_j = \sum_{j \in \{i \in \{1,\ldots,m\} | (D_i, x_i) = (D, y)\}} \lambda_j.$$

Otherwise, a sequence is called unbalanced.

Definition 15. *A random choice function* ρ *is* Strictly Coherent *if for every sequence* $\{(D_i, x_i)\}_{i=1}^m$ *of* $\mathcal{D} \times X$ *such that* $x_i \in D_i$*, and for every sequence of real numbers* $\{\lambda_i\}_{i=1}^m$ *such that the sequence* $\{(D_i, x_i, \lambda_i)\}_{i=1}^m$ *is unbalanced,*

$$\sum_{i=1}^m \lambda_i I_{\{\pi \in \Pi | \pi(x_i) \geq \pi(D_i)\}} \geq 0 \implies \sum_{i=1}^m \lambda_i \rho(D_i, x_i) > 0.$$

Theorem 5.
(i) A random choice function ρ *is Strictly Coherent if and only if* ρ *is a mixed logit function.*
(ii) Let X *be an affinely independent finite subset of* $\mathbf{R}^k$*. A random choice function* ρ *is Strictly Coherent if and only if* ρ *is a mixed linear logit function.*

D.1 Proof of the Necessity of Strict Coherency

By Theorems 1 and 3, it suffices to show that if ρ satisfies Quasi-Stochastic Rationality, then ρ is Strictly Coherent.

Choose a sequence $\{(D_i, x_i)\}_{i=1}^m$ of $\mathcal{D} \times X$ such that $x_i \in D_i$ and a sequence of real numbers $\{\lambda_i\}_{i=1}^m$ such that the sequence $\{(D_i, x_i, \lambda_i)\}_{i=1}^m$ is unbalanced. Suppose that $\sum_{i=1}^m \lambda_i I_{\{\pi \in \Pi | \pi(x_i) \geq \pi(D_i)\}} \geq 0$ to show $\sum_{i=1}^m \lambda_i \rho(D_i, x_i) > 0$.

For each $(D, x) \in \mathcal{D} \times X$ such that $x \in D$, define

$$u(D, x) = \sum_{j \in \{i \in \{1,\ldots,m\} | (D_i, x_i) = (D, x)\}} \lambda_j.$$

If (D, x) does not appear in the sequence, then $u(D, x) = 0$. Since $\{(D_i, x_i, \lambda_i)\}_{i=1}^m$ is unbalanced, u is not constant for some $D \in \mathcal{D}$. Define $q \in \Delta(\mathcal{D})$ by $q(D) = 1/|\mathcal{D}|$ for each $D \in \mathcal{D}$.

Now notice that for each $\pi \in \Pi$, if $\pi(x_i) \geq \pi(D_i)$, then $I_{\{\pi\in\Pi|\pi(x_i)\geq\pi(D_i)\}}(\pi) = 1 = \rho^\pi(D_i, x_i)$. If $\pi(x_i) < \pi(D_i)$, then $I_{\{\pi\in\Pi|\pi(x_i)\geq\pi(D_i)\}}(\pi) = 0 = \rho^\pi(D_i, x_i)$. Therefore, for each $\pi \in \Pi$, $I_{\{\pi\in\Pi|\pi(x_i)\geq\pi(D_i)\}}(\pi) = \rho^\pi(D_i, x_i)$.

Hence, for each $\pi \in \Pi$,

$$\begin{aligned} \sum_{i=1}^m \lambda_i I_{\{\pi\in\Pi|\pi(x_i)\geq\pi(D_i)\}}(\pi) &= |\mathcal{D}| \sum_{D\in\mathcal{D}} q(D) \sum_{x\in D} u(D,x)\rho^\pi(D,x) \\ &= |\mathcal{D}| E(\rho^\pi : q, u). \end{aligned} \tag{20}$$

Since $\sum_{i=1}^m \lambda_i I_{\{\pi\in\Pi|\pi(x_i)\geq\pi(D_i)\}} \geq 0$, I have $E(\rho^\pi : q, u) \geq 0$ for all $\pi \in \Pi$. By Quasi-Stochastic Rationality, $E(\rho^\pi : q, u) > 0$. Hence

$$\sum_{i=1}^m \lambda_i \rho(D_i, x_i) = |\mathcal{D}| \sum_{D\in\mathcal{D}} q(D) \sum_{x\in D} u(D,x)\rho^\pi(D,x) = |\mathcal{D}| E(\rho^\pi : q, u) > 0.$$

Therefore, ρ is Strictly Coherent.

D.2 Proof of the Sufficiency of Strict Coherency

By Theorems 1 and 3, it suffices to show that if ρ is Strictly Coherent then ρ satisfies Quasi-Stochastic Rationality. Choose any $q \in \Delta(D)$ and any $u(D, \cdot) \in \mathbf{R}^D$ such that $u(D, \cdot)$ is not constant for some D with $q(D) > 0$. Let $\alpha = \min_{\pi\in\Pi} E(\rho^\pi : q, u)$. Choose any $D' \in \mathcal{D}$ such that $q(D') > 0$. For any $x \in D'$, define $v(D', x) = u(D', x) - (\alpha/q(D'))$. For any $(D, x) \in (\mathcal{D} \setminus \{D'\}) \times X$ such that $x \in D$, define $v(D, x) = u(D, x)$. Since $u(D, \cdot)$ is not constant for some D with $q(D) > 0$, $v(D, \cdot)$ is not constant for some D with $q(D) > 0$. Moreover, $E(\rho^\pi : q, u) = E(\rho^\pi : q, v) + \alpha$ for any $\pi \in \Pi$. Therefore, $\min_{\pi\in\Pi} E(\rho^\pi : q, v) = 0$ and $E(\rho^\pi : q, v) \geq 0$ for any $\pi \in \Pi$.

Define sequences $\{(D_i, x_i)\}_{i=1}^m$ and $\{\lambda_i\}_{i=1}^m$ as follows. For each $(D, x) \in \mathcal{D} \times X$ such that $x \in D$, if $v(D, x) \neq 0$, then include (D, x) in the sequence. Since the number of a pair (D, x) such that $x \in D$ is finite, I obtain a sequence $\{(D_i, x_i)\}_{i=1}^m$. For each (D_i, x_i) in the sequence, define $\lambda_i = q(D_i)v(D_i, x_i)$ for each $i \in \{1, \ldots, m\}$. Then for any $\rho \in \mathcal{P}$,

$$E(\rho : q, v) \equiv \sum_{D\in\mathcal{D}} q(D) \sum_{x\in D} v(D,x)\rho(D,x) = \sum_{i=1}^m \lambda_i \rho(D_i, x_i). \tag{21}$$

Since $v(D, \cdot)$ is not constant for some D with $q(D) > 0$, there exist $x, y \in D$ such that $q(D)v(D, x) \neq q(D)v(D, y)$. Hence, $\sum_{j\in\{i\in\{1,\ldots,m\}|(D_i,x_i)=(D,x)\}} \lambda_j = q(D)v(D, x) \neq q(D)v(D, y) = \sum_{j\in\{i\in\{1,\ldots,m\}|(D_i,x_i)=(D,y)\}} \lambda_j$, where the equalities hold because by the definition of the sequence. Therefore, $\{(D_i, x_i, \lambda_i)\}_{i=1}^m$ is unbalanced.

As in the proof of the necessity, for each $\pi \in \Pi$, $I_{\{\pi\in\Pi|\pi(x_i)\geq\pi(D_i)\}}(\pi) = \rho^\pi(D_i, x_i)$. Therefore, for each $\pi \in \Pi$,

$$\begin{aligned}\sum_{i=1}^m \lambda_i I_{\{\pi\in\Pi|\pi(x_i)\geq\pi(D_i)\}}(\pi) &= \sum_{i=1}^m q(D_i)v(D_i, x_i)\rho^\pi(D_i, x_i) \\ &= \sum_{D\in\mathcal{D}} q(D) \sum_{x\in D} v(D, x)\rho^\pi(D, x) \\ &= E(\rho^\pi : q, v).\end{aligned}$$

Since $\min_{\pi\in\Pi} E(\rho^\pi : q, v) \geq 0$, this implies that $\sum_{i=1}^m \lambda_i I_{\{\pi\in\Pi|\pi(x_i)\geq\pi(D_i)\}} \geq 0$. By Strict Coherency, $\sum_{i=1}^m \lambda_i \rho(D_i, x_i) > 0$. Therefore,

$$E(\rho : q, u) = \alpha + E(\rho : q, v) = \alpha + \sum_{i=1}^m \lambda_i \rho(D_i, x_i) > \alpha = \min_{\pi\in\Pi} E(\rho^\pi : q, u),$$

where the second equality holds by (21). Therefore, ρ satisfies Quasi-Stochastic Rationality.

References

Barberá, S. and P. K. Pattanaik (1986): "Falmagne and the rationalizability of stochastic choices in terms of random orderings," *Econometrica*, 707–715.

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

Berry, S. T. and P. A. Haile (2009): "Nonparametric identification of multinomial choice demand models with heterogeneous consumers," Tech. rep., National Bureau of Economic Research.

Block, H. D. and J. Marschak (1960): "Random orderings and stochastic theories of responses," *Contributions to probability and statistics*, 2, 97–132.

Clark, S. A. (1996): "The random utility model with an infinite choice space," *Economic Theory*, 7, 179–189.

Falmagne, J.-C. (1978): "A representation theorem for finite random scale systems," *Journal of Mathematical Psychology*, 18, 52–72.

Fishburn, P. C. (2015): *The theory of social choice*, Princeton University Press.

Fox, J. T. and A. Gandhi (2016): "Nonparametric identification and estimation of random coefficients in multinomial choice models," *The RAND Journal of Economics*, 47, 118–139.

Fox, J. T., K. il Kim, S. P. Ryan, and P. Bajari (2012): "The random coefficients logit model is identified," *Journal of Econometrics*, 166, 204–212.

Greene, W. H. and D. A. Hensher (2003): "A latent class model for discrete choice analysis: contrasts with mixed logit," *Transportation Research Part B: Methodological*, 37, 681–698.

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

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

Hiriart-Urruty, J.-B. and C. Lemaréchal (2012): *Fundamentals of convex analysis*, Springer Science & Business Media.

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

McFadden, D. and K. Train (2000): "Mixed MNL models for discrete response," *Journal of applied Econometrics*, 447–470.

Rockafellar, R. T. (2015): *Convex analysis*, Princeton university press.

Soltan, V. (2015): *Lectures on convex sets*, World Scientific.

Stoer, J. and C. Witzgall (2012): *Convexity and optimization in finite dimensions I*, vol. 163, Springer Science & Business Media.