Supplemental Appendix: Random Utility with Unobservable Alternatives

Haruki Kono, Kota Saito, Alec Sandroni

January 16, 2026

F Omitted Proofs

F.1 Proof of Remark B.3

Define an extended network by $\mathcal{N}^* = \mathcal{N} \cup \{s^*, t^*\}$ and $\mathcal{A}^* = \mathcal{A} \cup \{(s^*, s)\} \cup \{(t, t^*)\}$. Define a flow $r^* : \mathcal{A}^* \to \mathbf{R}$ by $r^*(t, t^*) = 1$; $r^*(s, s^*) = 1$; and for all $(D, x) \in \mathcal{M}$, $r^*(D \setminus x, D) = K(\rho, D, x)$. Note that r^* satisfies the conservation law, that is for any $D \in \mathcal{N}$, $r^*(\mathcal{N}^*, D) = r^*(D, \mathcal{N}^*)$, where $r^*(\mathcal{N}^*, D)$ is the sum of inflows to D and $r^*(D, \mathcal{N}^*)$ is the sum of outflows from D. Such a flow exists by Lemma C.1. The conservation law implies that for any collection $\mathcal{C} \subseteq \mathcal{N}$

$$r^*(\mathcal{N}^* \setminus \mathcal{C}, \mathcal{C}) = r^*(\mathcal{C}, \mathcal{N}^* \setminus \mathcal{C}), \tag{30}$$

where for all $\mathcal{C} \subseteq \mathcal{N}^*$, $r^*(\mathcal{C}, \mathcal{N}^* \setminus \mathcal{C})$ is the total value of arcs from $\mathcal{C}$ to $\mathcal{N}^* \setminus \mathcal{C}$, and likewise $r^*(\mathcal{N}^* \setminus \mathcal{C}, \mathcal{C})$ is the total value of arcs from $\mathcal{N}^* \setminus \mathcal{C}$ to $\mathcal{C}$.

Fix collection $\mathcal{C} \subseteq \mathcal{N}$ such that $\mathcal{C}$ has no unobservable inflows or outflows. That is, for all arcs $(D \setminus x, D) \in \mathcal{A}^*$ from $\mathcal{C}$ to $\mathcal{N} \setminus \mathcal{C}$ or from to $\mathcal{N} \setminus \mathcal{C}$ to $\mathcal{C}$, $(D \setminus x, D)$ is in $\mathcal{M}$; and thus $r^*(D \setminus x, D) = K(\rho, D, x)$. Therefore,

$$\begin{aligned}
&r^*(\mathcal{C}, \mathcal{N}^* \setminus \mathcal{C}) - r^*(\mathcal{N}^* \setminus \mathcal{C}, \mathcal{C}) \\
&= 1\{t \in \mathcal{C}\} \cdot r^*(t, t^*) + r^*(\mathcal{C}, \mathcal{N} \setminus \mathcal{C}) - [r^*(\mathcal{N} \setminus \mathcal{C}, \mathcal{C}) + 1\{s \in \mathcal{C}\} \cdot r^*(s^*, s)]
\end{aligned}$$

$$= 1\{t \in \mathcal{C}\} + r^*(\mathcal{C}, \mathcal{N} \setminus \mathcal{C}) - [r^*(\mathcal{N} \setminus \mathcal{C}, \mathcal{C}) + 1\{s \in \mathcal{C}\}]$$
$$= \delta_\rho(\mathcal{C}),$$

where the final equality holds by the definition of δ_ρ since all arcs from $\mathcal{C}$ to $\mathcal{N} \setminus \mathcal{C}$ or from $\mathcal{N} \setminus \mathcal{C}$ to $\mathcal{C}$ satisfy $r^*(D \setminus x, D) = K(\rho, D, x)$. Thus $\delta_\rho(\mathcal{C}) = 0$ by (30).

F.2 Details of Proof of Lemma B.5

It suffices to show $\delta_\rho(\mathcal{C} \cup \mathcal{E}) = \delta_\rho(\mathcal{C}) + \delta_\rho(\mathcal{E})$ for any disjoint sets $\mathcal{C}, \mathcal{E} \subseteq 2^X$. We focus on the case of $\emptyset, X \notin \mathcal{C} \cup \mathcal{E}$ for simplicity. Consider the decompositions:

$$\underbrace{\sum_{\substack{(D,x):D\in\mathcal{C},D\cup x\notin\mathcal{C}\\(D\cup x,x)\in\mathcal{M}}} K(\rho, D\cup x, x)}_{\text{flow out of } \mathcal{C}} = \underbrace{\sum_{\substack{(D,x):D\in\mathcal{C},D\cup x\in\mathcal{E}\\(D\cup x,x)\in\mathcal{M}}} K(\rho, D\cup x, x)}_{\text{flow out of } \mathcal{C} \text{ and into } \mathcal{E} \,\equiv\, (1.1)} + \underbrace{\sum_{\substack{(D,x):D\in\mathcal{C},D\cup x\notin\mathcal{C}\cup\mathcal{E}\\(D\cup x,x)\in\mathcal{M}}} K(\rho, D\cup x, x)}_{\text{flow out of } \mathcal{C} \text{ not into } \mathcal{E} \,\equiv\, (1.2)}$$

$$\underbrace{\sum_{\substack{(E,y):E\notin\mathcal{C},E\cup y\in\mathcal{C}\\(E\cup y,y)\in\mathcal{M}}} K(\rho, E\cup y, y)}_{\text{flow into } \mathcal{C}} = \underbrace{\sum_{\substack{(E,y):E\in\mathcal{E},E\cup y\in\mathcal{C}\\(E\cup y,y)\in\mathcal{M}}} K(\rho, E\cup y, y)}_{\text{flow into } \mathcal{C} \text{ from } \mathcal{E} \,\equiv\, (2.1)} + \underbrace{\sum_{\substack{(E,y):E\notin\mathcal{C}\cup\mathcal{E},E\cup y\in\mathcal{C}\\(E\cup y,y)\in\mathcal{M}}} K(\rho, E\cup y, y)}_{\text{flow into } \mathcal{C} \text{ not from } \mathcal{E} \,\equiv\, (2.2)}$$

$$\underbrace{\sum_{\substack{(D,x):D\in\mathcal{E},D\cup x\notin\mathcal{E}\\(D\cup x,x)\in\mathcal{M}}} K(\rho, D\cup x, x)}_{\text{flow out of } \mathcal{E}} = \underbrace{\sum_{\substack{(D,x):D\in\mathcal{E},D\cup x\in\mathcal{C}\\(D\cup x,x)\in\mathcal{M}}} K(\rho, D\cup x, x)}_{\text{flow out of } \mathcal{E} \text{ and into } \mathcal{C} \,\equiv\, (3.1)} + \underbrace{\sum_{\substack{(D,x):D\in\mathcal{E},D\cup x\notin\mathcal{C}\cup\mathcal{E}\\(D\cup x,x)\in\mathcal{M}}} K(\rho, D\cup x, x)}_{\text{flow out of } \mathcal{E} \text{ not into } \mathcal{C} \,\equiv\, (3.2)}$$

$$\underbrace{\sum_{\substack{(E,y):E\notin\mathcal{E},E\cup y\in\mathcal{E}\\(E\cup y,y)\in\mathcal{M}}} K(\rho, E\cup y, y)}_{\text{flow into } \mathcal{E}} = \underbrace{\sum_{\substack{(E,y):E\in\mathcal{C},E\cup y\in\mathcal{E}\\(E\cup y,y)\in\mathcal{M}}} K(\rho, E\cup y, y)}_{\text{flow into } \mathcal{E} \text{ from } \mathcal{C} \,\equiv\, (4.1)} + \underbrace{\sum_{\substack{(E,y):E\notin\mathcal{C}\cup\mathcal{E},E\cup y\in\mathcal{E}\\(E\cup y,y)\in\mathcal{M}}} K(\rho, E\cup y, y)}_{\text{flow into } \mathcal{E} \text{ not from } \mathcal{C} \,\equiv\, (4.2)}$$

Since (1.1) = (4.1) and (2.1) = (3.1), we have

$$\begin{aligned}
\delta_\rho(\mathcal{C}) + \delta_\rho(\mathcal{E}) &= ((1.1) + (1.2)) - ((2.1) + (2.2)) + ((3.1) + (3.2)) - ((4.1) + (4.2)) \\
&= \underbrace{((1.2) + (3.2))}_{\text{flow out of } \mathcal{C}\cup\mathcal{E}} - \underbrace{((2.2) + (4.2))}_{\text{flow into } \mathcal{C}\cup\mathcal{E}} \\
&= \delta_\rho(\mathcal{C} \cup \mathcal{E}),
\end{aligned}$$

where the last equality holds by $\mathcal{C} \cap \mathcal{E} = \emptyset$.

F.3 Detailed Proof of Lemma B.8

In this section, we prove the following statement: $\delta_\rho(\mathcal{C}) \geq 0$ for any complete collection $\mathcal{C} \subseteq \mathcal{N}$ such that $\emptyset \notin \mathcal{C}$ if and only if the condition (22) holds for any $\mathcal{C} \subseteq \mathcal{N}$.

If part: To show the if direction, assume (22) holds for any $\mathcal{C} \subseteq \mathcal{N}$. Fix any complete $\mathcal{C} \subseteq \mathcal{N}$ such that $\emptyset \notin \mathcal{C}$. We will show that $\delta_\rho(\mathcal{C}) \geq 0$ by considering two cases.

Case A: $\mathcal{C}$ does not have unobservable outflows. Note that $\sum_{(D,E)\in\mathcal{C}^c\times\mathcal{C}} l(D,E) = \sum_{(E,y):E\notin\mathcal{C},E\cup y\in\mathcal{C},(E\cup y,y)\in\mathcal{M}} K(\rho, E\cup y, y)$. By the hypothesis of the case, any outflow from $\mathcal{C}$ is observable. Then u does not take the value of $+\infty$. Thus we have $\sum_{(D,E)\in\mathcal{C}\times\mathcal{C}^c} u(D,E) = \sum_{(D,x):D\in\mathcal{C},D\cup x\notin\mathcal{C},(D\cup x,x)\in\mathcal{M}} K(\rho, D\cup x, x)$. It follows that the left-hand side minus the right-hand side of (22) equals to the value of $\delta_\rho(\mathcal{C})$. Therefore, (22) implies $\delta_\rho(\mathcal{C}) \geq 0$.

Case B: $\mathcal{C}$ has unobservable outflows. By disjoint additivity, we have $\delta_\rho(\mathcal{C}) = \delta_\rho(\mathcal{C}\cap\mathcal{D}) + \delta_\rho(\mathcal{C}\setminus\mathcal{D})$. Thus, it suffices to show that $\delta_\rho(\mathcal{C}\cap\mathcal{D}) \geq 0$ and $\delta_\rho(\mathcal{C}\setminus\mathcal{D}) \geq 0$.

Since $\mathcal{D}$ is an upper set and $\mathcal{C}$ is complete, there are no unobservable outflows from $\mathcal{C}\cap\mathcal{D}$. Thus by Case A, $\delta_\rho(\mathcal{C}\cap\mathcal{D}) \geq 0$. Next, notice that $\emptyset \notin \mathcal{C}\setminus\mathcal{D}$, and that $\mathcal{C}\setminus\mathcal{D}$ has no observable inflows. By Remark B.4, $\delta_\rho(\mathcal{C}\setminus\mathcal{D}) \geq 0$ holds.

Only if part: Now, to show the other direction, fix any $C \subseteq \mathcal{N}$. We consider two cases based on whether C has an unobservable outflow.

Case A: C has an unobservable outflow. If there exists an arc $(D, D\cup x)$ with $D \in \mathcal{C}$, $D\cup x \notin \mathcal{C}$, and $(D\cup x, x) \notin \mathcal{M}$, then $u(D, D\cup x) = +\infty$ by the setup in (23). Hence the left-hand side of (22) equals $+\infty$, and condition (22) holds trivially.

Case B: All outflows from $\mathcal{C}$ are observable. In this case, $\mathcal{C}$ is complete by Remark B.2. We further divide B into two subcases.

Case B1: $\emptyset \notin \mathcal{C}$. By the hypothesis of the statement, we have $\delta_\rho(\mathcal{C}) \geq 0$. Since the left-hand side minus the right-hand side of (22) equals to $\delta_\rho(\mathcal{C})$, thus (22) holds.

Case B2: $\emptyset \in \mathcal{C}$. Define $\mathcal{G} \equiv \{D\cup E \mid D \in \mathcal{C}\cap\mathcal{D}^c, E \subseteq X^*\}$. Notice that $\mathcal{G} \subseteq \mathcal{C}$ since if $D \in \mathcal{C}$, then $D\cup E \in \mathcal{C}$ for any $E \subseteq X^*$ by the completeness of $\mathcal{C}$. We will show that $\delta_\rho(\mathcal{G}) = 0$ and $\delta_\rho(\mathcal{C}\setminus\mathcal{G}) \geq 0$, which implies $\delta_\rho(\mathcal{C}) = \delta_\rho(\mathcal{C}\setminus\mathcal{G}) \geq 0$ by disjoint additivity.

Note that $\mathcal{G}$ is complete, and that $\emptyset \in \mathcal{G}$ because $\emptyset \in \mathcal{C}$ by our assumption in Case B2 and $\emptyset \in \mathcal{D}^c$ by definition.

Claim 1: $\mathcal{C} \cap \mathcal{D}^c$ is a lower set.[1]

Proof. Suppose it is not a lower set. Then there exist $E \in \mathcal{C} \cap \mathcal{D}^c$ and $E' \subseteq E$ such that $E' \notin \mathcal{C} \cap \mathcal{D}^c$.

Since $\emptyset \in \mathcal{C} \cap \mathcal{D}^c$, $E \in \mathcal{C} \cap \mathcal{D}^c$, $E' \subseteq E$ and $E' \notin \mathcal{C} \cap \mathcal{D}^c$, somewhere on the path from $\emptyset$ to E' there is an arc coming from a node $D \in \mathcal{C} \cap \mathcal{D}^c$ to a node $D \cup x \notin \mathcal{C} \cap \mathcal{D}^c$. Since $E \in \mathcal{D}^c$, it follows that its subsets E', D, and $D \cup x$ are all elements of $\mathcal{D}^c$. Since $D \cup x \in \mathcal{D}^c$ and $D \cup x \notin \mathcal{C} \cap \mathcal{D}^c$, we have $D \cap x \notin \mathcal{C}$. Thus $(D, D \cup x)$ is an outflow from $\mathcal{C}$, and it is unobservable because $D \cup x \in \mathcal{D}^c$. This contradicts our assumption in Case B that $\mathcal{C}$ does not have unobservable outflows. ■

Claim 2: $\mathcal{G}$ is a lower set.

Proof. Choose an element of $G \in \mathcal{G}$, where $G = D \cup E$ for some $D \in \mathcal{C} \cap \mathcal{D}^c, E \in 2^{X^*}$. Then any subset G' of G can be written as $D' \cup E'$ where $D' \subseteq D$ and $E' \subseteq E$. In particular, set $D' = D \cap G'$ and $E' = E \cap G'$.

Because $\mathcal{C} \cap \mathcal{D}^c$ is a lower set, $D' \in \mathcal{C} \cap \mathcal{D}^c$. Similarly because 2^{X^*} is a lower set $E' \in 2^{X^*}$. It follows that $G' = D' \cup E' \in \mathcal{G}$. ■

Claim 3: $\mathcal{G}$ has no unobservable outflow.

Proof. Suppose towards a contradiction that there exists an unobservable arc $(D, D \cup x)$ such that $D \in \mathcal{G}$, and $D \cup x \notin \mathcal{G}$. Since $\mathcal{G}$ is complete, we have $x \in \tilde{X}$, and thus, $D \cup x \in \mathcal{D}^c$ because $(D, D \cup x)$ is an unobservable arc. Since $D \cup x \notin \mathcal{G}$, we have $D \cup x \notin \mathcal{C}$.[2] But then $(D, D \cup x)$ is an unobservable outflow from $\mathcal{C}$ which does not exist by the assumption in Case B on $\mathcal{C}$. ■

Now we can complete the proof of Case B2. The collection $\mathcal{G}$ has no unobservable inflows (because it is a lower set) and, no unobservable outflows. Thus by Remark B.3, $\delta_\rho(\mathcal{G}) = 0$.

[1] A collection $\mathcal{E}$ is a *lower set* if it satisfies the following: $E \in \mathcal{E}, D \subseteq E \implies D \in \mathcal{E}$. It can be shown that $\mathcal{E}$ is an upper set if and only if $\mathcal{E}^c$ is a lower set.

[2] If $D \cup x \in \mathcal{C}$, then $D \cup x \in \mathcal{C} \cap \mathcal{D}^c$ as $D \cup x \in \mathcal{D}^c$. Thus, $D \cup x \in \mathcal{G}$, which is a contradiction.

Note that $\mathcal{G}^c$ is an upper set (since $\mathcal{G}$ is a lower set), hence $\mathcal{G}^c$ is complete. Therefore $\mathcal{C} \setminus \mathcal{G} = \mathcal{C} \cap \mathcal{G}^c$ is the intersection of two complete sets, and hence is complete. Moreover, $\emptyset \notin \mathcal{C} \setminus \mathcal{G}$ since $\emptyset \in \mathcal{G}$. It follows from the hypothesis of the statement that $\delta_\rho(\mathcal{C} \setminus \mathcal{G}) \geq 0$. Thus, $\delta_\rho(\mathcal{C}) = \delta_\rho(\mathcal{C} \setminus \mathcal{G}) \geq 0$, which implies the inequality (22) for $\mathcal{C}$.

F.4 Proof of Lemma B.7

The only if direction is trivial. We prove the if direction. Define an extended network with lower bound by $\mathcal{N}^* = \mathcal{N} \cup \{s^*, t^*\}$ and $\mathcal{A}^* = \mathcal{A} \cup \{(s^*, s)\} \cup \{(t, t^*)\}$ and $u^*(s^*, s) = 1$, $l^*(s^*, s) = 0$, $u^*(t, t^*) = 1, l^*(t, t^*) = 0$, $u^*(D, E) = u(D, E), l^*(D, E) = l(D, E)$ for all other arcs.

We first define a function e in the augmented network: for any disjoint $\mathcal{C}^*, \mathcal{F}^* \subseteq \mathcal{N}^*$,

$$e(\mathcal{C}^*, \mathcal{F}^*) = \sum_{(D,E) \in \mathcal{A}^* : (D,E) \in \mathcal{C}^* \times \mathcal{F}^*} u^*(D, E) - \sum_{(D,E) \in \mathcal{A}^* : (D,E) \in \mathcal{F}^* \times \mathcal{C}^*} l^*(D, E).$$

If $s^* \in \mathcal{C}^*$ and $t^* \notin \mathcal{C}^*$ then $\mathcal{C}^*$ is called an s^*-t^* *cut*. Furthermore $e(\mathcal{C}^*, \mathcal{N}^* \setminus \mathcal{C}^*)$ is called the *capacity* of the s^*-t^* cut $\mathcal{C}^*$. Then we will prove that $(\mathcal{N}^* \setminus \{t^*\}, t^*)$ is a minimum s^*-t^* cut. Let be any $\mathcal{C}^* \subseteq \mathcal{N}^*$ such that $s^* \in \mathcal{C}^*$ and $t^* \notin \mathcal{C}^*$. That is, $(\mathcal{C}^*, \mathcal{N}^* \setminus \mathcal{C}^*)$ be an arbitrary cut separating s^* and t^*. Let $\mathcal{C} = \mathcal{C}^* \cap \mathcal{N}$. Then by the structure of network,

$$\begin{aligned}
&e(\mathcal{C}^*, \mathcal{N}^* \setminus \mathcal{C}^*) - e(\mathcal{N}^* \setminus \{t^*\}, \{t^*\}) \\
&= e(\{s^*\}, \mathcal{N}^* \setminus \mathcal{C}^*) + e(\mathcal{C}, \mathcal{N}^* \setminus \mathcal{C}^*) - e(\mathcal{N}^* \setminus \{t^*\}, \{t^*\}) \quad (\because s^* \in \mathcal{C}^*) \\
&= e(\{s^*\}, \mathcal{N}^* \setminus \mathcal{C}^*) + e(\mathcal{C}, \mathcal{N} \setminus \mathcal{C}) + e(\mathcal{C}, \{t^*\}) - e(\mathcal{N}^* \setminus \{t^*\}, \{t^*\}) \quad (\because t^* \in \mathcal{N}^* \setminus \mathcal{C}^*) \\
&= e(\{s^*\}, (\mathcal{N} \setminus \mathcal{C}) \cap \{s\}) + e(\mathcal{C}, \mathcal{N} \setminus \mathcal{C}) + e(\mathcal{C} \cap \{t\}, \{t^*\}) - e(\{t\}, \{t^*\}) \\
&= e(\{s^*\}, (\mathcal{N} \setminus \mathcal{C}) \cap \{s\}) + e(\mathcal{C}, \mathcal{N} \setminus \mathcal{C}) - e((\mathcal{N} \setminus \mathcal{C}) \cap \{t\}, \{t^*\}) \\
&= 1(s \notin \mathcal{C}) + e(\mathcal{C}, \mathcal{N} \setminus \mathcal{C}) - 1(t \notin \mathcal{C}) \\
&= 1(s \notin \mathcal{C}) + \left(\sum_{(D,E) \in \mathcal{C} \times (\mathcal{N} \setminus \mathcal{C})} u(D, E) - \sum_{(D,E) \in (\mathcal{N} \setminus \mathcal{C}) \times \mathcal{C}} l(D, E) \right) - 1(t \notin \mathcal{C})
\end{aligned}$$

$$= 1(t \in \mathcal{C}, s \notin \mathcal{C}) + \left(\sum_{(D,E) \in \mathcal{C} \times (\mathcal{N} \setminus \mathcal{C})} u(D,E) - \sum_{(D,E) \in (\mathcal{N} \setminus \mathcal{C}) \times \mathcal{C}} l(D,E) \right) - 1(t \notin \mathcal{C}, s \in \mathcal{C}),$$

which is nonnegative under (22). Thus $e(\mathcal{C}^*, \mathcal{N}^* \setminus \mathcal{C}^*) \geq e(\mathcal{N}^* \setminus \{t^*\}, t^*)$ for any cut $(\mathcal{C}^*, \mathcal{N}^* \setminus \mathcal{C}^*)$ separating s^* and t^* if and only if (22) holds for any $\mathcal{C} \subseteq \mathcal{N}$. For this particular cut, we have $e(\mathcal{N}^* \setminus \{t^*\}, \{t^*\}) = 1$.

Thus assuming (22), by the generalized maximum-flow theorem with lower bounds and upper bounds (See Theorem 6.10 of Ahuja, Magnanti and Orlin (1993)), there exists a feasible flow r^* from s^* to t^* that saturates arc of (t, t^*), that is, $r^*(t, t^*) = 1$. By flow conservation at all nodes, the total flow into t^* must equal the total flow out of s^*. Since (s^*, s) is the only arc leaving s^*, we have $r^*(s^*, s) = 1$. Now define r as a restriction of r^* on $(\mathcal{N}, \mathcal{A})$. Then r satisfies all desired conditions. The capacity constraints follow from $l^* = l$ and $u^* = u$, on $\mathcal{A}$. The flow conservation conditions follow from flow conservation in the augmented network together with $r^*(s^*, s) = 1$ and $r^*(t, t^*) = 1$.

F.5 Proof of Lemma C.1

The if direction of the proof is easy: (i) follows from the fact that $r(X \setminus x, x) = K(\rho^*, X, x) = \rho^*(X, x)$ and summing over $x \in X$ yields $\sum_{x \in X} r(X \setminus \{x\}, X) = 1$. Condition (ii) follows by a direct calculation; equivalently, it can be interpreted as a conservation law for flows at each node D with $1 \leq |D| < |X|$. Condition (iii) follows from the fact that $\sum_{E: E \supseteq D} r(E \setminus x, E) = \rho^*(D, x) \geq 0$.

We show the only if direction as follows. For all (D, x) such that $x \in D \in \mathcal{D}$, define ρ^* by $\rho^*(D, x) = \sum_{E \supseteq D} r(E \setminus x, E)$. By condition (iii), we have $\rho^*(D, x) \geq 0$. Also by the definition and the Möbius inversion, $K(\rho^*, D, x) = r(D \setminus x, D)$. To complete the proof, it suffices to show that $\sum_{x \in D} \rho^*(D, x) = 1$ for all $D \in \mathcal{D}$.

Fix $D \in \mathcal{D}$. We prove by induction on i that for all $0 \leq i \leq |X| - |D|$:

$$\text{(31)} \qquad \sum_{x \in D} \rho^*(D, x) = \sum_{\substack{E \supseteq D \\ |E| = |D| + i}} \sum_{y \in E} r(E \setminus y, E) + \sum_{x \in D} \sum_{\substack{E \supseteq D \\ |E| \geq |D| + i + 1}} r(E \setminus x, E).$$

Base case ($i = 0$): By definition,

$$\sum_{x \in D} \rho^*(D, x) = \sum_{x \in D} \sum_{E \supseteq D} r(E \setminus x, E) \quad = \sum_{x \in D} r(D \setminus x, D) + \sum_{x \in D} \sum_{\substack{E \supseteq D \\ |E| \geq |D|+1}} r(E \setminus x, E),$$

which matches (31) with $i = 0$.

Inductive step from i to $i+1$: Assume (31) holds for some i with $|D|+i < |X|$. Apply condition (ii) to each E with $|E| = |D| + i$:

$$\sum_{y \in E} r(E \setminus y, E) = \sum_{z \notin E} r(E, E \cup z).$$

Substituting this identity into the first term of (31) yields

$$\begin{aligned} &\sum_{\substack{E \supseteq D \\ |E|=|D|+i}} \sum_{y \in E} r(E \setminus y, E) + \sum_{x \in D} \sum_{\substack{E \supseteq D \\ |E| \geq |D|+i+1}} r(E \setminus x, E) \\ &= \sum_{\substack{E \supseteq D \\ |E|=|D|+i}} \sum_{z \notin E} r(E, E \cup z) + \sum_{x \in D} \sum_{\substack{E \supseteq D \\ |E| \geq |D|+i+1}} r(E \setminus x, E). \end{aligned}$$

Reindex the first double sum by letting $F = E \cup z$:

$$\sum_{\substack{E \supseteq D \\ |E|=|D|+i}} \sum_{z \notin E} r(E, E \cup z) = \sum_{\substack{F \supseteq D \\ |F|=|D|+i+1}} \sum_{y \in F \setminus D} r(F \setminus y, F).$$

Next, split the second term:

$$\sum_{x \in D} \sum_{\substack{E \supseteq D \\ |E| \geq |D|+i+1}} r(E \setminus x, E) = \sum_{x \in D} \sum_{\substack{E \supseteq D \\ |E|=|D|+i+1}} r(E \setminus x, E) + \sum_{x \in D} \sum_{\substack{E \supseteq D \\ |E| \geq |D|+i+2}} r(E \setminus x, E).$$

Combining these displays, we obtain

$$\sum_{\substack{F \supseteq D \\ |F|=|D|+i+1}} \sum_{y \in F \setminus D} r(F \setminus y, F) + \sum_{x \in D} \sum_{\substack{E \supseteq D \\ |E|=|D|+i+1}} r(E \setminus x, E) + \sum_{x \in D} \sum_{\substack{E \supseteq D \\ |E| \geq |D|+i+2}} r(E \setminus x, E).$$

By combining the first two terms, we have

$$\sum_{x\in D}\rho^*(D,x) = \sum_{\substack{F\supseteq D\\|F|=|D|+i+1}}\sum_{y\in F} r(F\setminus y, F) + \sum_{x\in D}\sum_{\substack{E\supseteq D\\|E|\geq|D|+i+2}} r(E\setminus x, E),$$

which is (31) with i replaced by $i+1$.

Now, we use (31) with $i = |X| - |D|$. Since $|D| + i = |X|$, the only set $E \supseteq D$ with $|E| = |X|$ is $E = X$. The second sum in (31) is empty (no sets have size $\geq |X| + 1$). Thus:

$$\sum_{x\in D}\rho^*(D,x) = \sum_{y\in X} r(X\setminus y, X) = 1,$$

where the last equality is condition (i).

F.6 Proof of Claim C.4

We first construct a directed path from $\emptyset$ to $\tilde{X}$ that avoids every node in $\mathcal{H}$ by adding the elements of $\tilde{X}$ one at a time. Note that any intermediate node A on this path does not appear in an essential test collection: the only test collection that contains A is $\{A \cup E \mid E \in 2^{X^*}\}$, which is nonessential.

In the same manner, we construct a directed path from $\tilde{X}$ to X that avoids every node in $\mathcal{H}$ by adding the elements of X^* one at a time. Any intermediate node on this path has the form $\tilde{X} \cup E$ for some $E \subseteq X^*$ (i.e., $E \in 2^{X^*}$). Such a node cannot belong to any essential test collection: any test collection that contains it must be of the form $\{\tilde{X} \cup E : E \in \mathcal{E}\}$, which is not essential.

By combining these two directed paths, we obtain a desirable $\emptyset - X$ directed path avoiding any nodes in $\mathcal{H}$.

F.7 Proof of Claim C.5

Construction of r_D^1: Fix $D \in \mathcal{H}$. Let $A = D \cap \tilde{X}$ and $E = D \cap X^*$. Note that $D \neq A$ since $D \in \mathcal{H}$. Consider

- an $\emptyset - D$ directed path Π_1 going through the node E,

- an $A - X$ directed path Π_2 which avoid any nodes in $\mathcal{H}$ (such a path exists because we can take the union of any $A - \tilde{X}$ directed path and any $\tilde{X} - X$ directed path as in Claim C.4.),
- the directed path Π_3 from A to D constructed as follows: Let $(x_1, x_2, \ldots, x_m)$ denote the elements of X^* listed in the order in which they are added along the path Π_2 (as we traverse from A to X). Then Π_3 is the path from A to $D = A \cup E$ that adds precisely the elements of $E = D \cap X^*$, visiting them in the same relative order they appear in the sequence $(x_1, \ldots, x_m)$.[3]
- Π_4 defined in Claim C.4.

Fix $\varepsilon > 0$ and define

$$r_D^1 \equiv (1-\varepsilon) r^{\Pi_4} + \varepsilon r^{\Pi_1} + \varepsilon r^{\Pi_2} - \varepsilon r^{\Pi_3}. \tag{32}$$

See Figure 1 for an illustration of the flow. Note that r_D^1 satisfies the three conditions in Lemma C.1. To verify condition (iii), it suffices to show that every negative term in $\sum_{E \supseteq D} r(E \setminus x, E)$ is offset by a corresponding positive term in the same sum. In our construction, the only negative flows occur on arcs along Π_3. Moreover, each such arc in Π_3 is paired with an arc in Π_2 carrying the same magnitude with the opposite sign, because Π_3 traverses the same sequence of set inclusions as Π_2. Hence all negative contributions cancel, and the sum is nonnegative.

To see $\delta_{r_D^1}(\{D\}) = -\varepsilon$ note that by the definition (24) of δ_r, for $\mathcal{C} = \{D\}$ (so that $\emptyset, X \notin \mathcal{C}$) we have

$$\delta_{r_D^1}(\{D\}) = \sum_{x \in \tilde{X} :\, D \cup x \neq D} r_D^1(D, D \cup x) \; - \sum_{y \in D \cap \tilde{X}} r_D^1(D \setminus y, D),$$

i.e., observable outflow from D minus observable inflow into D. In the construction of r_D^1, $A \neq \emptyset$ implies there exists $y \in A \subseteq \tilde{X}$ and hence an observable arc $(D \setminus y, D)$; r_D^1 increases the flow on this incoming observable arc by ε while leaving all observable arcs leaving D unchanged. Therefore, $\delta_{r_D^1}(\{D\}) = -\varepsilon$.

[3]Formally, if $x, y \in E$ and $x = x_i$, $y = x_j$ with $i < j$, then x is added before y along Π_3.

Moreover, for all other $H \in \mathcal{H}$, $\delta_{r_D^1}(\{H\}) = 0$. To see this remember that δ is non-zero only when observable inflows are not equal to observable outflows.[4]

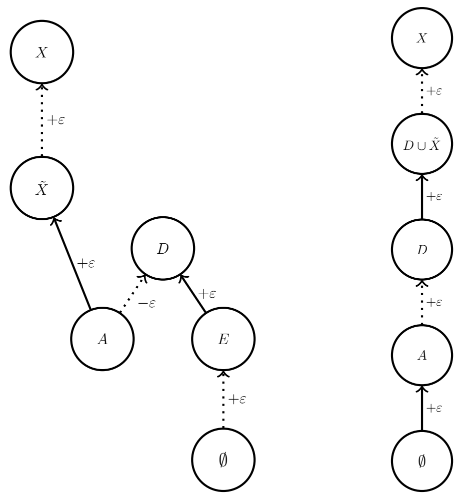

Figure 1: Flow $\varepsilon r^{\Pi_1} + \varepsilon r^{\Pi_2} - \varepsilon r^{\Pi_3}$ in r_D^1 (left); Flow εr^{Π_5} in r_D^2 in Claim C.5.

Construction of r_D^2: Choose any directed path from $\emptyset - X$ that goes through nodes A, D, and $\tilde{X} \cup D$. We denote the path by Π_5. See Figure 1. Define

$$r_D^2 \equiv (1 - \varepsilon) r^{\Pi_4} + \varepsilon r^{\Pi_5}. \tag{33}$$

Note that r_D^2 satisfies the all conditions in Lemma C.1. In the flow, $\delta_{r_D^2}(\{D\}) = \varepsilon$. Moreover, for all other $H \in \mathcal{H} \setminus \{D\}$, $\delta_{r_D^2}(\{H\}) = 0$ by the same reasoning as the previous claim.

F.8 Detailed Proof of Proposition IV.5

In this section, we prove the identity that we used for the proof of Proposition IV.5.

Lemma F.1. *For all $(\hat{D}, x)$ such that $x \in \hat{D} \in \hat{\mathcal{D}}$ and $x \in \tilde{X}$, $\sum_{E \supseteq D} K(\rho, E, x) = \sum_{\hat{E} \supseteq \hat{D}} r(\hat{E}, x)$.*

[4] In Figure 1, this occurs when a dotted line becomes a solid line or vice-versa in the diagram.

Proof. **Case 1:** $x_0 \in \hat{D}$.

$$\sum_{\hat{E}\supseteq\hat{D}} r(\hat{E},x) = \sum_{\hat{E}\supseteq\hat{D}} K(\rho,E,x) = \sum_{E\supseteq D} K(\rho,E,x),$$

where the first equality holds by the definition of r and the second equality holds because $\hat{E} \supseteq \hat{D}$ if and only if $E \supseteq D$.
Case 2: $x_0 \notin \hat{D}$. Note $\hat{D} = D \subseteq \tilde{X}$.

$$\begin{aligned}
\sum_{\hat{E}\supseteq\hat{D}} r(\hat{E},x) &= \sum_{F\subseteq\tilde{X}\setminus D} r(D\cup F,x) + \sum_{F\subseteq\tilde{X}\setminus D} r(D\cup F\cup x_0,x)\\
&= \sum_{F\subseteq\tilde{X}\setminus D}\sum_{G\subsetneq X^*} K(\rho,D\cup F\cup G,x) + \sum_{F\subseteq\tilde{X}\setminus D} K(\rho,D\cup F\cup X^*,x)\\
&= \sum_{F\subseteq\tilde{X}\setminus D}\sum_{G\subseteq X^*} K(\rho,D\cup F\cup G,x)\\
&= \sum_{E\supseteq D} K(\rho,E,x).
\end{aligned}$$

The first equality is just rearranging. The second equality follows by the definition of r. The third equality combines the second term into the inner summation. ∎

G Supplemental Contents

G.1 Interpretation of Condition (ii) in Theorem IV.2

In this section we provide a general interpretation of condition (ii) in Theorem IV.2. Fix any essential test collection $\mathcal{C}$. By the inflow-outflow equality, the total outflows from $\mathcal{C}$ minus the total inflows to $\mathcal{C}$ must be zero. Since $\mathcal{C}$ is an essential test collection, there are no unobservable outflows from $\mathcal{C}$. Thus, the inflow-outflow equality can be written as follows:

$$\left(\sum_{(D,x):D\in\mathcal{C},D\cup x\notin\mathcal{C}} K(\rho,D\cup x,x) - \sum_{(F,y):F\notin\mathcal{C},F\cup y\in\mathcal{C},y\in\tilde{X}} K(\rho,F\cup y,y)\right) \tag{34}$$

$$- \sum_{(E,z):E\notin\mathcal{C},E\cup z\in\mathcal{C},z\in X^*} K(\rho^*, E\cup z, z) = 0, \tag{35}$$

where ρ^* is a complete dataset such that $\rho^* = \rho$ on $\mathcal{M}$. The left hand side of (2) corresponds to the expression (34). Thus, by the equality, the left hand side of (2) equals

$$\sum_{(E,z):E\notin\mathcal{C},E\cup z\in\mathcal{C},z\in X^*} K(\rho^*, E\cup z, z).$$

Assuming ρ is RU rationalizable by μ, applying expression (1) yields

$$\sum_{(E,z):E\notin\mathcal{C},E\cup z\in\mathcal{C},z\in X^*} \mu(\succ\in\mathcal{L}|(E\cup z)^c \succ z \succ E), \tag{36}$$

as desired. If $\mathcal{C}$ is a singleton of $\{D\}$ as in Remark IV.3 then, (36) simplifies to (3).

G.2 Random Utility Polytope

In this section, we provide a geometric intuition for the set of RU rationalizable stochastic choice functions. Let $\mathcal{M}$ be the set of pairs of (D,x) such that $\rho(D,x)$ is observable.

Remark G.1. *For each ranking $\succ\in\mathcal{L}$ and $(D,x)\in\mathcal{M}$, define*

$$\rho^{\succ}(D,x) = \begin{cases} 1 & \textit{if } x \succ y \textit{ for all } y \in D\setminus x; \\ 0 & \textit{otherwise.} \end{cases} \tag{37}$$

The stochastic choice function $\rho^{\succ}$ gives probability one to the best alternative x in a choice set D according to the ranking $\succ$. The set of RU-rationalizable stochastic choice function is a polytope, that is, co.$\{\rho^{\succ} | \succ\in\mathcal{L}\}$, where co. denotes the convex hull. We denote the set by $\mathcal{P}_r$.

The hexagon in Figure 2 illustrates the polytope.[5] The inequality conditions

[5]Although the geometric intuition is useful, it is important to notice that the figure oversimplifies the reality since the number of vertices (i.e., $|X|!$) and the dimension of a random utility function can be very large. To see why the dimension of a random utility function can be very large, notice that it assigns a number for each pair of $(D,x)\in\mathcal{M}$.

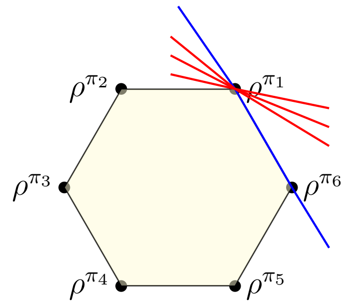

Figure 2: Random utility polytope

in Theorem IV.2 consist of the facet defining inequalities of the polytope, which correspond to the blue hyperplanes. On the other hand, as we will explain in the next section, McFadden and Richter's approach would contain non-facet defining inequalities, which correspond to the red hyperplanes.

G.3 Generalized Axiom of Revealed Stochastic Preference

In this section we discuss two different axioms based on McFadden and Richter (1990)'s approch and their application to our setup. The first, which we refer to as Axiom of Revealed Stochastic Preference (ARSP), is a well-known characterization in econometrics. This is necessary and sufficient for RU rationalizability when the choice frequencies of all alternatives in each available menu are observed. It may fail, however, when some alternatives are unobservable. The second new axiom, which we call Generalized Axiom of Revealed Stochastic Preference (GARSP), is necessary and sufficient even with unobservable alternatives. An equivalent formulation of GARSP may be found in McFadden and Richter (1990). Unlike our characterization (Theorem IV.2) both of these axioms contain an infinite number of redundant inequalities.

We further investigate the relationship between the Theorem IV.2 and GARSP in the next section.

Let $\mathcal{M}$ be the subset of pairs (D, x) such that $x \in D \in 2^X$ and $\rho(D, x)$ is well defined, i.e., $\rho(D, x)$ is observable to the analyst. In the following, we call $\mathcal{M}$ the set of observable pairs. In a typical setup where ARSP is discussed, it is assumed that all alternatives are observable. That is, if $(D, x) \in \mathcal{M}$ then $(D, y) \in \mathcal{M}$ for all $y \in D$. Equivalently, $\mathcal{M} \equiv \{(D, x) \in \mathcal{D} \times X \mid x \in D\}$ for some $\mathcal{D} \subseteq 2^X \setminus \emptyset$. In the main body of the paper, we assumed $\mathcal{M} \equiv \{(D, x) \in \mathcal{D} \times \tilde{X} \mid x \in D\}$.[6] In the following section we will consider arbitrary $\mathcal{M} \subseteq \{(D, x) \mid x \in D \in 2^X\}$. We denote a sequence of length k with each $(D_i, x_i) \in \mathcal{M}$ by $(D_i, x_i)_{i=1}^k \in \mathcal{M}^k$, where $\mathcal{M}^k$ denotes the k-fold Cartesian product of $\mathcal{M}$.

Definition G.2. *A stochastic choice function ρ satisfies the Axiom of Revealed Stochastic Preference (ARSP) if for any positive integer k and any sequence $(D_i, x_i)_{i=1}^k \in \mathcal{M}^k$,*

$$\max_{\succ \in \mathcal{L}(X)} \sum_{i=1}^n \rho^{\succ}(D_i, x_i) \geq \sum_{i=1}^n \rho(D_i, x_i).$$

This is necessary and sufficient when there are no unobservable alternatives. However, this is not sufficient to characterize random utility in our setup. Consider the following example:

Remark G.3. *Let $X = \{a, b, c, d\}$ and $X^* = \{c, d\}$ and $\mathcal{M} = \{(X, a), (X \setminus b, a)\}$. Let ρ be such that $\rho(X, a) = 1$ and $\rho(X \setminus b, a) = 0$. This is obviously not RU-rational as it violates Monotonicity (in particular condition (i) of Theorem IV.2 is violated). However, fix $\succ_0 \in \mathcal{L}(X)$ such that $a \succ_0 b \succ_0 c \succ_0 d$. That is, $\rho^{\succ_0}(X, a) = \rho^{\succ_0}(X \setminus b, a) = 1$. Then since $\rho^{\succ_0}(D, x) \geq \rho(D, x)$ for all $(D, x) \in \mathcal{M}$, we observe that for any k and any sequence $(D_i, x_i)_{i=1}^k$*

$$\sum_{i=1}^k \rho^{\succ_0}(D_i, x_i) \geq \sum_{i=1}^k \rho(D_i, x_i)$$

and thus $\max_{\succ \in \mathcal{L}(X)} \sum_i \rho^{\succ}(D_i, x_i) \geq \sum_i \rho(D_i, x_i)$. Therefore, ρ satisfies ARSP but is not RU rationalizable.

We provide the following characterization that works for arbitrary $\mathcal{M} \subseteq \{(D, x) \in 2^X \times X \mid x \in D\}$. We call it the *Generalized Axiom of Revealed Stochastic Preference (GARSP):*

[6] In Falmagne (1978), $\mathcal{M}$ is assumed to be $\{(D, x) \mid x \in D \in 2^X\}$.

Definition G.4. *A stochastic choice function ρ satisfies GARSP if for any non-negative integers k, l and any sequences $(D_i^+, x_i^+)_{i=1}^k \in \mathcal{M}^k$ and $(D_j^-, x_j^-)_{i=1}^l \in \mathcal{M}^l$,*

(38)
$$\max_{\succ \in \mathcal{L}(X)} \left(\sum_{i=1}^{k} \rho^{\succ}(D_i^+, x_i^+) - \sum_{j=1}^{l} \rho^{\succ}(D_j^-, x_j^-) \right) \geq \sum_{i=1}^{k} \rho(D_i^+, x_i^+) - \sum_{j=1}^{l} \rho(D_j^-, x_j^-).$$

If l is zero, then the GARSP reduces to ARSP. We now show the main characterization result of this section.

Theorem G.5. *Suppose $\mathcal{M} \subseteq \{(D, x) \in 2^X \times X \mid x \in D\}$. A stochastic choice function ρ defined on $\mathcal{M}$ is RU-rationalizable if and only if it satisfies GARSP.*

Proof. First suppose that ρ violates GARSP. That is, there exist sequences $(D_i^+, x_i^+)_{i=1}^k \in \mathcal{M}^k$ and $(D_j^-, x_j^-)_{i=1}^l \in \mathcal{M}^l$ such that

(39)
$$\max_{\succ \in \mathcal{L}(X)} \left(\sum_{i} \rho^{\succ}(D_i^+, x_i^+) - \sum_{j} \rho^{\succ}(D_j^-, x_j^-) \right) < \sum_{i} \rho(D_i^+, x_i^+) - \sum_{j} \rho(D_j^-, x_j^-).$$

For any RU-rationalizable $\hat{\rho}$, there exists $\mu \in \Delta(\mathcal{L})$ such that $\hat{\rho} = \sum_{\succ \in \mathcal{L}} \mu(\succ)\rho^{\succ}$. Thus, (39) implies that $\sum_i \hat{\rho}(D_i^+, x_i^+) - \sum_j \hat{\rho}(D_j^-, x_j^-) < \sum_i \rho(D_i^+, x_i^+) - \sum_j \rho(D_j^-, x_j^-)$. This shows that ρ is not RU rationalizable, as ρ is not equal to any convex combination of $\rho^{\succ}$s.

To show the other direction, suppose that ρ is not RU rationalizable (i.e., $\rho \notin \mathcal{P}_r$, where $\mathcal{P}_r$ is defined in Remark G.1). We will show a violation of GARSP, that is there exist sequences $(D_i^+, x_i^+)_{i=1}^k \in \mathcal{M}^k$ and $(D_j^-, x_j^-)_{i=1}^l \in \mathcal{M}^l$ that satisfies nequality (39).

Since $\mathcal{P}_r$ is compact and convex then by the separating hyperplane theorem there exists $\nu : \mathcal{M} \to \mathbf{R}$ such that $\nu \cdot \hat{\rho} < \nu \cdot \rho$ for all $\hat{\rho} \in \mathcal{P}_r$. In particular,

$$\max_{\succ \in \mathcal{L}(X)} \sum_{(D,x) \in \mathcal{M}} \nu(D, x)\rho^{\succ}(D, x) < \sum_{(D,x) \in \mathcal{M}} \nu(D, x)\rho(D, x). \tag{40}$$

Now by the density of the rationals and since the maximum is over a finite set (i.e.,

$\mathcal{L}(X)$), ν can be taken to be rational valued because the inequality in (40) is strict. Then by multiplying by a large positive integer, ν can also be taken to be integer valued and still satisfy the inequality. Finally, we obtain the desired sequences by letting (D, x) appear $\nu(D, x)$ times in $(D_i^+, x_i^+)_{i=1}^k \in \mathcal{M}^k$ if it is positive and $\nu(D, x)$ times in $(D_i^-, x_i^-)_{j=1}^l \in \mathcal{M}^l$ if it is negative. Then, (40) implies (39) with the defined sequences. ■

This characterization is similar to ARSP, but it allows negative signs in the sum. The fact that we need negative signs is clear geometrically. As Figure 2 shows, we are characterizing the polytope by considering *all* hyperplanes, equivalently *all* linear functionals i.e. ν. Thus, we should consider the negative directions as well as the positive directions.

In McFadden and Richter (1990), the analyst observes choice frequencies from what they call a *weak field* of each menu. A probability distribution over a weak field means that if the probability that an element of $A \subseteq D$ is chosen from D, $\rho(D, A)$, is observable, then $\rho(D, D \setminus A)$ is also observable. In our setup, this is equivalent to the the following condition: whenever we observe the probability $\rho(D, x)$ that x is chosen from D, we also observe the chance that x is not chosen from D, denoted by $\rho(D, D \setminus x)$, which we can obtain by $1 - \rho(D, x)$.

Based on the weak-field assumption above, we replace all $-\rho(D_j^-, x_j^-)$ (respectively, $-\rho^{\succ}(D_j^-, x_j^-)$) with $\rho(D_j^-, D_j^- \setminus x_j^-) \equiv 1 - \rho(D_j^-, x_j^-)$ (respectively, $\rho^{\succ}(D_j^-, D_j^- \setminus x_j^-) \equiv 1 - \rho^{\succ}(D_j^-, x_j^-)$) in the axiom. Notice that this operation yields an equivalent inequality because we are simply adding one l-times to the both sides. This rewriting gives us the original axiom by McFadden and Richter (1990) (page 171)[7]: for any positive integer k and any sequence of pairs $(D_i, A_i)_{i=1}^k$ such that $A_i \subseteq D_i$ and $\rho(D_i, A_i)$ is observable for all i,

$$(41) \qquad \max_{\succ \in \mathcal{L}(X)} \sum_{i=1}^{k} \rho^{\succ}(D_i, A_i) \geq \sum_{i=1}^{k} \rho(D_i, A_i).$$

If the dataset satisfies the following condition, GARSP becomes equivalent to ARSP: if $(D, x) \in \mathcal{M}$, then $(D, y) \in \mathcal{M}$ for all $y \in D$. This means that if we can

[7] McFadden and Richter (1990) call their original axiom as ARSP, not GARSP.

observe $\rho(D,x)$, then we also can observe $\rho(D,y)$ for all $y \in D$. This condition implies:

$$\sum_{x \in D \text{ s.t.} (D,x) \in \mathcal{M}} \rho(D,x) = 1. \tag{42}$$

Under the condition, we can focus on sequences with positive coefficients; thus ARSP becomes equivalent to GARSP.

Theorem G.6. *Let $\mathcal{M} \subseteq \{(D,x) \in 2^X \times X \mid x \in D\}$. Suppose that if $(D,x) \in \mathcal{M}$, then $(D,y) \in \mathcal{M}$ for all $y \in D$. A stochastic choice function ρ defined on $\mathcal{M}$ is RU-rationalizable if and only if it satisfies ARSP.*

Proof. The proof that RU-rationalizability implies GARSP is the same as in the proof of Theorem G.5. For the other direction, we modify the proof as follows. In the proof of Theorem G.5, we showed the existence of integer ν satisfying (40). Let $s = -\min_{(D,x)\in\mathcal{M}} \nu(D,x)$. Then define v^* by $v^*(D,x) = v(D,x) + s$ for all $(D,x) \in \mathcal{M}$. Note that ν^* is a nonnegative vector. Moreover, for any stochastic choice function and any $D \in \mathcal{D}$, $\sum_{x \in D \text{ s.t. } (D,x)\in\mathcal{M}} \nu^*(D,x)\rho(D,x) - \left(\sum_{x \in D \text{ s.t. } (D,x)\in\mathcal{M}} \nu(D,x)\rho(D,x)\right) = s$, where the last equality holds by (42). Thus, (40) holds with ν^* in the place of ν. The rest of the proof is the same. ∎

One drawback of ARSP and GARSP is that these characterizations involve an infinite number of inequalities to test and some of them are redundant, unlike the characterization in Theorem IV.2. In the next section, we show that these redundancies are inherent to the McFadden and Richter (1990) approach.

G.4 Redundancy in the McFadden and Richter (1990) Approach

The conditions of ARSP and GARSP involve an infinite number of sequences and some of the conditions are redundant. To see this notice in Definition G.4, each pair in the sequences $(D_i^+, x_i^+)_{i=1}^k \in \mathcal{M}^k$ and $(D_j^-, x_j^-)_{j=1}^l \in \mathcal{M}^l$ do not have to be distinct.[8] Thus, the number of sequences to be tested in the GARSP is infinite,

[8] For example, it can be that $(D_i^+, x_i^+) = (D_j^+, x_j^+)$ for distinct i and j.

although there are finitely many pairs $(D, x) \in \mathcal{M}$.[9]

In this section, we first show exactly which GARSP inequalities correspond to the conditions in Theorem IV.2. Consequently, we identify all redundant inequalities as those that do not appear in Theorem IV.2.

Remark G.7. *The inequality conditions (i) of Theorem IV.2 can be written as an inequality in GARSP.*

For each (E, y) such that $y \in E \subseteq X$, define $K_{(E,y)} \in \mathbf{R}^{\mathcal{M}}$ as

$$K_{(E,y)}(D, x) = \begin{cases} -1 & \text{if } y = x \text{ and } D \supseteq E \text{ and } |D \setminus E| \text{ is even,} \\ +1 & \text{if } y = x \text{ and } D \supseteq E \text{ and } |D \setminus E| \text{ is odd,} \\ 0 & \text{otherwise.} \end{cases}$$

For any $\rho \in \mathbf{R}^{\mathcal{M}}$, we have $K_{(E,y)} \cdot \rho = -K(\rho, E, y)$. For each (E, y), define sequences $(D_i^+, x_i^+)_{i=1}^k \in \mathcal{M}^k$ and $(D_j^-, x_j^-)_{i=1}^l \in \mathcal{M}^l$ so that each (D, x) appears exactly once in $(D_i^+, x_i^+)_{i=1}^k$ if $K_{(E,y)}(D, x) = 1$ and exactly once in $(D_j^-, x_j^-)_{i=1}^l \in \mathcal{M}^l$ if $K_{(E,y)}(D, x) = -1$. Thus, for all ρ,

$$\sum_i \rho(D_i^+, x_i^+) - \sum_j \rho(D_j^-, x_j^-) = -K(\rho, E, y)$$

Notice also that

$$\begin{aligned} \max_{\succ \in \mathcal{L}(X)} \left(\sum_i \rho^{\succ}(D_i^+, x_i^+) - \sum_j \rho^{\succ}(D_j^-, x_j^-) \right) &= \max_{\succ \in \mathcal{L}(X)} -K(\rho^{\succ}, E, y) \\ &= - \min_{\succ \in \mathcal{L}(X)} K(\rho^{\succ}, E, y) \\ &= 0, \end{aligned}$$

where the last equality holds because $K(\rho^{\succ}, E, y) = 1\{E^c \succ y \succ E \setminus y\}$ for all $\succ \in \mathcal{L}(X)$. Thus, GARSP for this sequence is equivalent to $K(\rho, E, y) \geq 0$. In

[9] McFadden and Richter (1990) discuss the difficulty of providing an upper bound on the number of allowable repetitions needed for their axiom to fully characterize RUM. They prove that sequences containing repetitions must be tested in general.

particular, in the case of complete data, it is sufficient to only consider sequences without repetition.

We can do something similar with condition (ii) in Theorem IV.2.

Remark G.8. *The inequality conditions (ii) of Theorem IV.2 are also implied by GARSP.*

For each essential test collection $\mathcal{C}$ define $\nu_{\mathcal{C}}(F,z)$ as the coefficient on $\rho(F,z)$ in the polynomial below:

$$\delta'_\rho(\mathcal{C}) \equiv \sum_{\substack{(D,x):D\in\mathcal{C},D\cup x\notin\mathcal{C},\\(D\cup x,x)\in\mathcal{M}}} K(\rho,D\cup x,x) - \sum_{\substack{(E,y):E\notin\mathcal{C},E\cup y\in\mathcal{C},\\(E\cup y,y)\in\mathcal{M}}} K(\rho,E\cup y,y). \tag{43}$$

Note that ν_c are integer-valued. Note also that (43) is equivalent to the left hand-side of condition (ii) of Theorem IV.2. The value of (43) coincides with $\delta_\rho(\mathcal{C})$ defined in (17) when $\mathcal{C}$ is an essential test collection.[10]

Define sequences $(D_i^+,x_i^+)_{i=1}^k \in \mathcal{M}^k$ and $(D_j^-,x_j^-)_{i=1}^l \in \mathcal{M}^l$ so that each (D,x) appears exactly $-\nu_{\mathcal{C}}(D,x)$ many times in $(D_i^+,x_i^+)_{i=1}^k$ if $\nu_{\mathcal{C}}(D,x)<0$ and exactly $\nu_{\mathcal{C}}(D,x)$ many times in $(D_j^-,x_j^-)_{i=1}^l \in \mathcal{M}^l$ if $\nu_{\mathcal{C}}(D,x)>0$. Then

$$\sum_i \rho(D_i^+,x_i^+) - \sum_j \rho(D_j^-,x_j^-) = -\delta'_\rho(\mathcal{C}) \tag{44}$$

for all ρ. As shown in the proof of Theorem IV.2, $\delta'_{\rho^\succ}(\mathcal{C}) \geq 0$ for all $\succ$; and furthermore there exists $\succ$ such that $\delta'_{\rho^\succ}(\mathcal{C}) = 0$ by the fact that $\delta'_\rho(\mathcal{C}) = 0$ is a facet defining inequality and the vertices of the polytope consist of $\rho^\succ$s.[11] Thus

$$\max_{\succ\in\mathcal{L}(X)} \sum_i \rho^\succ(D_i^+,x_i^+) - \sum_j \rho^\succ(D_j^-,x_j^-) = \max_{\succ\in\mathcal{L}(X)} \sum_{(D,x)\in\mathcal{M}} -\delta'_{\rho^\succ}(\mathcal{C}) = 0. \tag{45}$$

Plugging these equalities into the GARSP (i.e., (38)) yields $0 \geq -\delta'_\rho(\mathcal{C})$ which is equivalent to condition (ii) of Theorem IV.2 for the test collection $\mathcal{C}$.

[10] When $\mathcal{C}$ is an essential test collection, we have $X \notin \mathcal{C}$ and $\emptyset \notin \mathcal{C}$.

[11] Another way to see this is by Claim C.4. The $\emptyset$-X dipath in Claim C.4 avoids all nodes in $\mathcal{C}$, thus $\delta'_{\rho^{\Pi_4}}(\mathcal{C}) = 0$

Remark G.9. *We have thus shown our conditions can be written as GARSP (by selecting sequences in a very particular way); however the converse is not true. While our conditions are non-redundant, all of the sequences in GARSP apart from those corresponding to our conditions are redundant.*

See Figure 2 for the illustration. Our conditions contain only the blue facet defining hyperplanes, while the GARSP contain all non-facet-defining red hyperplanes.

G.5 Simplification of bounds of unobservable choice probabilities when $\mathcal{D} = 2^X \setminus \emptyset$

In this section, we assume that $\mathcal{D} = 2^X \setminus \emptyset$ and provide a further simplification of bounds of unobservable choice frequencies.

Corollary G.10. *Let $\mathcal{D} = 2^X \setminus \emptyset$. For $(D, x) \notin \mathcal{M}$, write $D = A \cup E$ where $A = D \cap \tilde{X}$ and $E = D \cap X^*$. The bounds for the unobservable choice frequency $\rho^*(D, x)$ are obtained by $[\underline{\rho}(D, x), \bar{\rho}(D, x)]$, where*

$$\bar{\rho}(D,x) = \max_{r \in \mathbf{R}_+^{\{(D\setminus x, D) | x \in D \in 2^X\}}} \sum_{B: A \subseteq B \subseteq \tilde{X}} \sum_{F: E \subseteq F \subseteq X^*} r(B \cup F \setminus x, B \cup F) \tag{46}$$

subject to $r(C \setminus y, C) = K(\rho, C, y)$ for all $(C, y) \in \mathcal{M}$, and

$$\begin{aligned} &\sum_{y \in F} r(B \cup F \setminus y, B \cup F) - \sum_{y \in X^* \setminus F} r(B \cup F, B \cup F \cup y) \\ &= \sum_{x \in \tilde{X} \setminus B} K(\rho, B \cup F \cup x, x) - \sum_{y \in B} K(\rho, B \cup F, y) + \mathbf{1}\{B \cup F = X\} \end{aligned} \tag{47}$$

for all $B \subseteq \tilde{X}$ and $F \subseteq X^*$ such that $B \cup F \neq \emptyset$. The lower bound $\underline{\rho}(D, x)$ solves a similar problem with min *replacing* max.

Corollary G.10 helps us exploit the block structure in the network. To explain this concretely, let $B, B' \subseteq \tilde{X}$, $F, F' \subseteq X^*$, and $y \in F$, $z \in F'$ such that $B \neq B'$. Then, the variables $r(B \cup F \setminus y, B \cup F)$ and $r(B' \cup F' \setminus z, B' \cup F')$ are independent in the sense that neither constrains the other via (47). This is because unobservable arcs connect only nodes differing by a single element of X^*, so nodes with distinct observable parts lie in separate connected components of the unobservable flow network.

The independence in the network allows us to solve the optimization problem for each B separately. That is, instead of solving (46) directly, we can consider the subproblem

$$\max_{r_B \in \mathbf{R}_+^{\{((B\cup F)\setminus y, B\cup F) \mid y\in F\subseteq X^*\}}} \sum_{F:E\subseteq F\subseteq X^*} r_B(B\cup F\setminus x, B\cup F) \quad \text{s.t. (47) for all } F\subseteq X^*$$

for each B with $A\subseteq B\subseteq \tilde{X}$. Given a solution r_B to this problem, we can recover a solution to the original problem (46) as follows:

$$r(B\cup F\setminus y, B\cup F) = \begin{cases} K(\rho, B\cup F, y) & \text{if } y\in B, \\ r_B(B\cup F\setminus y, B\cup F) & \text{if } y \in F, \end{cases}$$

for all $B\subseteq \tilde{X}$, $F\subseteq X^*$, and $y\in B\cup F$.

In general, solving separate linear programs is computationally less demanding than solving the aggregated one. When all menus are available (i.e., $\mathcal{D} = 2^X\setminus\emptyset$), the linear program (46) decomposes as above, and therefore, the computation can become significantly easier.

G.6 Implications to statistical testing of rationality

Theorem IV.2 establishes that following two representations of the set of random utility polytope are equivalent: $\mathcal{P}_r \equiv co.\{\rho^{\succ} \mid \succ\in\mathcal{L}\}$ and

$$\left\{\rho\in\mathbf{R}_+^{\mathcal{M}} \mid \rho \text{ satisfies the inequalities in (i) and (ii) of Theorem IV.2}\right\}.$$

The first representation characterizes an RU-rationalizable dataset as the convex hull of a finite set of points. In convex geometry, this is known as a vertex representation (V-representation). The second representation expresses the polytope via its facet defining inequalities, commonly referred to as an halfspace representation (H-representation).

For testing whether a given incomplete dataset is RU-rationalizable, the H-representation is generally more practical, as it requires only checking for a single violated inequality to reject rationalizability. In contrast, the V-representation involves searching the typically high-dimensional space $\Delta(\mathcal{L})$ for a supporting weight, making it computationally intensive.

A key challenge with the H-representation is that a closed form expression remains unknown for arbitrary patterns of missing data, whereas the V-representation follows directly from the definition of rationalizability. Consequently, empirical studies on testing RU-rationalizability, such as Kitamura and Stoye (2018) and Dean, Ravindran and Stoye (2022), rely on the V-representation, trading computational cost for broader applicability.

Our contribution in Theorem IV.2 is to explicitly enumerate all facet defining inequalities for a specific structure of data incompleteness. Once the H-representation is obtained, rationality testing reduces to verifying inequality constraints. Statistical inference can then proceed using standard techniques for moment inequalities (see Andrews and Soares (2010), Canay, Illanes and Velez (2023)), though assessing the performance of such methods is beyond the scope of this paper.

G.7 Values of BM polynomials of example in Section I

In this section, we provide a table that gives all of values of BM polynomials that can be calculated given ρ in Table 1 of Section I for $\varepsilon = 0$.

$$K(\rho, D, x):$$

$D \diagdown x$	a	b	(c)	(d)
$\{a\}$	0	–	–	–
$\{b\}$	–	0	–	–
$\{a,b\}$	0	1/3	–	–
$\{a,c\}$	1/6	–	?	–
$\{a,d\}$	1/6	–	–	?
$\{b,c\}$	–	1/6	?	–
$\{b,d\}$	–	1/6	–	?
$\{c,d\}$	–	–	?	?
$\{a,b,c\}$	1/6	0	?	–
$\{a,b,d\}$	1/6	0	–	?
$\{a,c,d\}$	1/6	–	?	?
$\{b,c,d\}$	–	1/6	?	?
$\{a,b,c,d\}$	1/6	1/6	?	?

G.8 Details of calibrated dataset

Recall that $\mathcal{L}$ is the set of linear orders on $X = \{0, 1, 2, 3, 4\}$. For a probability distribution μ over $\mathcal{L}$, let $\rho(D, x \mid \mu)$ be the choice probability of x out of D induced by μ, that is,

$$\rho(D, x \mid \mu) \equiv \mu\left(\succ \in \mathcal{L} \mid x \succ y \text{ for all } y \in D \setminus x\right).$$

Given a complete dataset ρ, we shall find an RU-rationalizable dataset that resembles ρ. We solve the following problem:

$$\hat{\mu} \in \operatorname*{argmin}_{\mu \in \Delta(\mathcal{L})} \sum_{x \in D \in \mathcal{D}} \left(\rho(D, x \mid \mu) - \rho(D, x)\right)^2,$$

and we use the calibrated choice data $(\rho(D, x \mid \hat{\mu}))_{x \in D \in \mathcal{D}}$ in the analysis in Section V. Table 1 summarizes the calibrated probabilities for the lottery dataset used in the analysis of Section V.

Table 1: Calibrated choice data ρ

$D \backslash x$	0	1	2	3	4
{0, 1}	0.434996	0.565004	-	-	-
{0, 2}	0.171289	-	0.828711	-	-
{0, 3}	0.119875	-	-	0.880125	-
{0, 4}	0.108795	-	-	-	0.891205
{1, 2}	-	0.444959	0.555041	-	-
{1, 3}	-	0.203263	-	0.796737	-
{1, 4}	-	0.179475	-	-	0.820525
{2, 3}	-	-	0.475196	0.524804	-
{2, 4}	-	-	0.307597	-	0.692403
{3, 4}	-	-	-	0.453832	0.546168
{0, 1, 2}	0.162682	0.291693	0.545626	-	-
{0, 1, 3}	0.119875	0.156422	-	0.723703	-
{0, 1, 4}	0.091033	0.121704	-	-	0.787263
{0, 2, 3}	0.064422	-	0.431887	0.503691	-
{0, 2, 4}	0.066709	-	0.295601	-	0.63769
{0, 3, 4}	0.053548	-	-	0.453832	0.49262
{1, 2, 3}	-	0.198698	0.289444	0.511858	-
{1, 2, 4}	-	0.147176	0.216616	-	0.636208
{1, 3, 4}	-	0.102122	-	0.395562	0.502315
{2, 3, 4}	-	-	0.27787	0.251982	0.470148
{0, 1, 2, 3}	0.064422	0.151858	0.280029	0.503691	-
{0, 1, 2, 4}	0.066709	0.106637	0.215865	-	0.610789
{0, 1, 3, 4}	0.053548	0.08145	-	0.395562	0.469439
{0, 2, 3, 4}	0.046842	-	0.265874	0.251982	0.435302
{1, 2, 3, 4}	-	0.098786	0.192909	0.251982	0.456323
{0, 1, 2, 3, 4}	0.046842	0.078113	0.192158	0.251982	0.430904

References

Ahuja, Ravindra K, Thomas L Magnanti, and James B Orlin. 1993. *Network flows.* Prentice-Hall, Inc.

Andrews, Donald WK, and Gustavo Soares. 2010. "Inference for parameters defined by moment inequalities using generalized moment selection." *Econometrica*, 78(1): 119–157.

Canay, Ivan A, Gastón Illanes, and Amilcar Velez. 2023. "A user's guide for inference in models defined by moment inequalities." *Journal of Econometrics*, 105558.

Dean, Mark, Dilip Ravindran, and Jörg Stoye. 2022. "A Better Test of Choice Overload." *arXiv preprint arXiv:2212.03931.*

Falmagne, Jean-Claude. 1978. "A representation theorem for finite random scale systems." *Journal of Mathematical Psychology*, 18(1): 52–72.

Kitamura, Yuichi, and Jörg Stoye. 2018. "Nonparametric analysis of random utility models." *Econometrica*, 86(6): 1883–1909.

McFadden, D., and M.K. 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.