Supplemental Appendix to "Attention Overload"

Matias D. Cattaneo* Paul H.Y. Cheung† Xinwei Ma‡ Yusufcan Masatlioglu§

May 28, 2023

Abstract

This Supplemental Appendix contains an in-depth discussion of related literature, simulation evidence, and omitted technical details in the proofs of the main results.

Contents

*Department of Operations Research and Financial Engineering, Princeton University.
†Jindal School of Management, University of Texas at Dallas
‡Department of Economics, UC San Diego.
§Department of Economics, University of Maryland.

SA.1 Related Literature

We provide an overview of the related decision theory literature with the goal of comparing and contrasting our AOM and HAOM$_{\triangleright}$ with other choice models. The Random Utility Model (RUM) assumes attention homogeneity (i.e., full attention) but allows for preference heterogeneity. In contrast, early random limited attention models started with a focus on preference homogeneity while allowing for (parametric or nonparametric) attention rule heterogeneity, but only recently they have began to incorporate preference heterogeneity (via parametric assumptions).

Section SA.1.1 focuses on random attention models with homogeneous preferences, and thus compares them to AOM. Section SA.1.2 considers random attention models that intend to capture multiple preferences, and thus compares them to HAOM$_{\triangleright}$. Lastly, Section SA.1.3 discusses other stochastic choices models that are related to AOM and/or HAOM$_{\triangleright}$.

SA.1.1 Random Limited Attention Models with Homogeneous Preferences

Our AOM is a missing piece of the puzzle in the random limited consideration choice literature. In the main text, we explored the two extreme cases of Ann and Ben, and showed how other random limited attention models perform in terms of attention allocation behaviors under attention overload. In this section, we also illustrate key differences in (observable) behavioral implications of different random attention models with homogeneous preferences: we explore whether the models explain different sets of choice data.

We first compare nonparametric models: *Random Attention Model* (RAM, Cattaneo, Ma, Masatlioglu, and Suleymanov, 2020) and AOM. Due to the flexibility of the nonparametric attention rule in both models, the important challenge is to understand revealed preference. Consider the two choice data examples in the following table: Table 1a has an AOM representation but not a RAM representation, while Table 1b has an RAM representation but not a AOM representation.

$\pi(\cdot\|S)$	a	b	c
$\{a,b,c\}$	0.4	0.3	0.3
$\{a,b\}$	0.8	0.2	
$\{a,c\}$	0.8		0.2
$\{b,c\}$		0.5	0.5

(a)

$\pi(\cdot\|S)$	a	b	c	d
$\{a,b,c,d\}$	1/2	1/2	0	0
$\{a,b,c\}$	0	2/3	1/3	
$\{a,b\}$	1/2	1/2		

(b)

Table 1: Examples for Differences in Explanatory Power: RAM vs. AOM

Recall that RAM predicts that a is preferred to b if there exists $a,b \in S$ such that $\pi(a|S \setminus b) < \pi(a|S)$ (Cattaneo, Ma, Masatlioglu, and Suleymanov, 2020). Thus, the example in Table 1a does not have a RAM representation because RAM predicts that $b \succ c$ and $c \succ b$, but a cyclic preference cannot be explained by RAM. On the other hand, AOM implies $a \succ c$ and $a \succ b$, which can be rationalized through two different preferences: $a \succ b \succ c$ or $a \succ c \succ b$. Therefore, it has two AOM representations.

Interestingly, there does not exist a 3-alternative example that RAM can explain but AOM cannot. In other words, AOM explains more choice data than RAM in settings with three alternatives. Therefore, our second example in Table 1b resorts to 4-alternative case to illustrate a case that can be explained by RAM but not by AOM. In that example, AOM predicts that $a \succ b$ and $b \succ a$, and hence it could not be explained by AOM. On the other hand, revealed preference under RAM implies $a \succ d$ and $b \succ c$, and thus the data can be rationalized under RAM.

More generally, while AOM and RAM provide testable restrictions (inequalities) in terms of the choice rule, they can take very different forms. To compare, first recall from our main characterization result that a preference ordering has an AOM representation if and only if $\succ$-Regularity holds: $\pi(a|S) \leq \pi(U_{\succeq}(a)|T)$ for $T \subset S$. In contrast, RAM imposes a regularity condition when removing "better" alternatives: it requires $\pi(a|S) \leq \pi(a|S \setminus b)$ whenever $b \succ a$. As a consequence, the two underlying regularity violations used for identification are conceptually different.

In the realm of nonparametric attention models, we can also compare to the convex hull of deterministic consideration set mappings. For example, the random competition filter (RCF) is a important special case of AOM. Let $\Gamma_j(\cdot)$ be deterministic consideration set mappings that satisfy *Competition Filter* (Lleras, Masatlioglu, Nakajima, and Ozbay, 2017). A RCF model is specified by an attention rule $\mu_{\texttt{RCF}}(T|S) = \sum_{j=1}^{J} \alpha_j \mathbb{1}(\Gamma_j(S) = T)$ with the respective attention frequency, $\phi_{\texttt{RCF}}(a|S) = \sum_{j:a\in\Gamma_j(S)} \alpha_j$ with $\alpha_j \geq 0$, $j = 1, 2, \ldots, J$, and $\sum_{j=1}^{J} \alpha_j = 1$. Random competition filters satisfy attention overload because $a \in \Gamma_j(T)$ for all $T \subseteq S$ if $a \in \Gamma_j(S)$.

The RCF model nests two other choice models of interest: *bounded rationalization* and *imprecise narrowing down*. Bounded rationalization is a generalization of Cherepanov, Feddersen, and Sandroni (2013). It states that the decision maker does not always stick to the same set of rationale given the same choice set. Hence, it is as-if the decision maker assigns a probability distribution over the power set on the set of rationale. Since Cherepanov, Feddersen, and Sandroni (2013) is a special case of Lleras, Masatlioglu, Nakajima, and Ozbay (2017), it follows that the bounded rationalization model is a special case of RCF. Imprecise narrowing down shares a similar idea. Given the same choice set, the decision maker does not necessarily follow the same procedure on setting up criteria. Thus, it is as-if the decision maker assigns a probability distribution over the set of all possible procedures. It makes imprecise narrow down again a special case of RCF.

Last but not least, we consider several parametric models and compare them to AOM in terms of explanatory power. First, the *Independent Attention* model of Manzini and Mariotti (2014) is a special case of AOM. They assume that each alternative a has a fixed probability, $\gamma(a)$, to be considered. In other words, the attention frequency, $\phi_{\texttt{MM}}(a|S)$, is held fixed across different choice problems in their model ($\phi_{\texttt{MM}}(a|S) = \gamma(a)$). Since attention overload requires only weak inequality, the model falls into AOM. While full encapsulating the variability in attention rule, any choice rule that is explained through the independent attention model can be explained through AOM.

Second, the *Elimination by Aspects* attention model of Aguiar (2017) is also a special of AOM: the author assumes that each category D has a fixed probability $m(D)$. If the category is available, the decision maker picks the best alternative out of it. If not, she chooses the default option. Let the set of all categories be $\mathcal{D}$. Then the attention frequency is given by $\phi_{\texttt{Aguiar}}(a|S) = \sum_{a\in D\in\mathcal{D}} m(D)$. It is straightforward to see that this attention rule satisfies attention overload, and thus AOM nests all the explanatory power of that model.

Third, the *Bounded Processing Capacity Rule* (BPCR) of Marchant and Sen (2023) is another special case of AOM. The model assumes that the DM has a fixed capacity, k, for considering alternatives. In particular, when the size of menu is less than k, the DM considers everything; when the size of menu exceeds k, the DM considers all menus of size k with equal probability. The attention frequency is then given by

$\phi_{\texttt{MS}}(a|S) = 1$ if $|S| \leq k$, and $\phi_{\texttt{MS}}(a|S) = \frac{\binom{|S|-1}{k-1}}{\binom{|S|}{k}}$ if $|S| > k$. It can be shown that attention frequency satisfies attention overload. Therefore, BPCR necessarily falls under the scope of AOM regarding choice behaviors.

Fourth, as illustrated in the previous discussion, the attention rule in the *Logit Attention* model (e.g., Brady and Rehbeck, 2016) may fall outside of attention overload (depending on the parameter specification). Indeed, the Logit Attention model would capture certain behaviors that AOM would not capture. To see this, we generate a choice from the Logit Attention model. Consider a Luce weight, say $\ell(.)$, as a function on menu so that $\ell(abcd) = \frac{1}{2}$, $\ell(bcd) = \frac{1}{2}$, $\ell(bc) = \frac{2}{3}$, $\ell(c) = \frac{1}{2}$, $\ell(ab) = \frac{1}{2}$, $\ell(b) = \frac{1}{2}$ and for all other $S \subseteq \{a, b, c, d\}$, we let $\ell(S) = 0$. (For ease of illustration, we relax the assumptions for ℓ to be strictly between 0 and 1; but it is possible to construct an example with a similar result without this relaxation.) We also let $a \succ b \succ c \succ d$. This will produce the choice data in Table 1b, and hence it cannot be explained by AOM. On the other hand, it is also true that AOM can explain choice data and the Logit Attention model cannot explain. For example, recall that the choice data in Table 1a cannot be explained by RAM, but the Logit Attention model is a special case of RAM (Cattaneo, Ma, Masatlioglu, and Suleymanov, 2020), which implies that the choice data cannot be explained by the Logit attention model either while it can be rationalized by AOM.

Finally, Demirkan and Kimya (2020) considers a version of independent consideration from Manzini and Mariotti (2014) with a menu-dependent attraction parameter. In other words, they specify a parameter, say $\gamma(a, S)$, which depends on both the alternative and the menu. Since it is now possible that $\gamma(a, S) > \gamma(a, T)$, their model can produce choice behavior that is outside of AOM. On the other hand, due to the parametric nature of the model, there exists choice behaviors that can be captured by AOM but not their model.

SA.1.2 Random Limited Attention Models with Heterogeneous Preferences

There exist only a few papers combining random utility (heterogeneous preferences) and random limited attention. Within the paradigm of nonparametric identification, Kashaev and Aguiar (2022) attempts to generalize RAM by incorporating random utility (set-monotone and stable RAUM). There are two key differences between that model and $\text{HAOM}_{\triangleright}$. First, we aim at the property of list-based attention overload, which can better capture attention allocation behavior under attention scarcity, while set-monotone and stable RAUM focuses on the RAM model. Second, in order to provide sharp identifications of preference and attention, we limit the variability of attention and preference through a list $\triangleright$, while Kashaev and Aguiar (2022) can only obtain partial identification results because full preference and attention heterogeneity (in terms of RAM) are allowed in their model.

Outside of the nonparametric identification paradigm, there are several papers allowing for heterogeneous preferences and random limited attention via parametric restrictions. For example, Aguiar, Boccardi, Kashaev, and Kim (2023) assumes a family of the parametric models with the attention index model, Gibbard (2021) utilize a version of independence consideration model as discussed in Horan (2019), and Dardanoni, Manzini, Mariotti, and Tyson (2020) models parameterize cognitive capacities with a fixated size of consideration sets. In contrast, $\text{HAOM}_{\triangleright}$ relies on a nonparametric assumption on attention which captures attention overload.

A different strand of the literature relies on enriched type of choice data to allow for heterogeous preferences and random limited attention. For example, Gibbard (2021) assumes a frame-dependent choice function, and Dardanoni, Manzini, Mariotti, Petri, and Tyson (2023) utilizes mixture choice data. In contrast, AOM and HAOM$_\triangleright$ only requires standard choice data for identificationa, estimation and inference of heterogeneous preference and attention frequencies.

Finally, the literature on discrete choice has always prioritized heterogeneous preference, and a fraction of that literature has also factored in variations in attention. We discuss a few recent contributions to this literature, and explain how they differ from AOM and HAOM$_\triangleright$. Barseghyan, Coughlin, Molinari, and Teitelbaum (2021) study partial identification of preference and consideration set formation. Unlike AOM, which imposes monotonicity of the attention frequency across nested choice problems, they restrict the size of consideration sets (or, alternatively, assume that the decision maker cannot pay attention to singleton sets too often). Their nonparametric identifying assumption is thus similar in spirit to our (additional) attentive at binary assumption in AOM and the full attention assumption over binary sets in HAOM$_\triangleright$, both of which aiming to increase identification power. Abaluck and Adams (2021) exploit asymmetries in cross-partial derivatives and show that consideration set formation and preference distribution can be separately identified from observed choices when there is rich exogenous variation in observed covariates. Conceptually, HAOM$_\triangleright$ shares the feature of richness by assuming choices from all menus are observed (but not AOM) for characterization. Nonetheless, identification of preference in HAOM$_\triangleright$ requires only binary menus. Barseghyan, Molinari, and Thirkettle (2021) and Barseghyan and Molinari (2023), on the other hand, provide identification results for risk preference when exogenous variation in observed covariates is more restricted. They also demonstrate the tradeoff between the exclusion restrictions and the assumptions on choice set formation. In contrast, AOM, or the non-parametric attention literature in general, remains in the abstract domain where the object of choice can be arbitrary. It will be interesting to see how the other side of the spectrum of the attention literature can also benefit in terms of identification power when restricted to risky choices.

SA.1.3 Other Related Stochastic Choice Models

Since AOM is a general model that does not require regularity, it is natural to ask whether other stochastic choice models would fall within AOM in terms of explanatory power. Indeed, there are a number of them: the random utility model (RUM) is a prime example. Because $\succ$-Regularity is automatically satisfied when a model satisfies regularity, a standard RUM can be represented by AOM.

There are several other models nested in AOM. For example, Gul, Natenzon, and Pesendorfer (2014) consider an attribute rule in which the decision maker first draws an attribute and then picks an alternative which contains such attribute. They show that every attribute rule is a RUM; hence, every attribute rule can be represented by AOM. Fudenberg, Iijima, and Strzalecki (2015) introduce the additive perturbed utility model where the decision maker intentionally randomizes as deterministic choices can be costly. Since the choices in their model always satisfy regularity, any choice rule in the additive perturbed utility model has an AOM representation.

Moreover, there are several stochastic choice models that allow for regularity violations. Intriguingly, we can show that some of them are AOM by directly checking $\succ$-Regularity. Important examples include Echenique, Saito, and Tserenjigmid (2018) and Echenique and Saito (2019). In addition, Filiz-Ozbay and Masatlioglu (2023) introduces the *Less-is-more Progressive Random Choice* model relying on the less-is-more choice function from Lleras, Masatlioglu, Nakajima, and Ozbay (2017). Since each less-is-more choice function can be mapped back to a competition filter, the model is essentially a special case of RCF and thus it can be rationalized by AOM.

HAOM$_\triangleright$ is of course related to the literature on choice over a list. We focus on the differences between HAOM$_\triangleright$ and models on choices over lists in terms of modeling ideas and approaches. Rubinstein and Salant (2006) consider a deterministic model of choice function from lists and, building on their deterministic model, they consider random choice rules defined over menus in which the decision maker chooses from a randomly appearing list. On the other hand, Guney (2014) and Yildiz (2016) consider models of choice procedure of successive elimination over a list. Similar to the framework of Rubinstein and Salant (2006), Guney (2014) explicitly considers a model building on choice function over lists, while Yildiz (2016) assumes the list is unobservable and considers a nested and recursive structure in the random choice function. The focus of this strand of literature is different from HAOM$_\triangleright$: they intend to capture specific elimination procedures under a binary relation (which need not be a preference), while HAOM$_\triangleright$ looks at decisions made under a set of well-defined heterogenours preferences with random limited attention. Interestingly, HAOM$_\triangleright$ is closely related to the *Random Depth Model* (RDM) of Ishii, Kovach, and Ülkü (2021). The RDM assumes that the decision maker can have a random preference and a random depth parameter when making their choices over a list. In particular, a depth parameter, says k, determines how far down the list she will explore, whereas our list-based attention overload assumes the decision makers will consider items up to certain alternative down a list. RDM considers a fixed choice set and exogenous variation in list, while HAOM$_\triangleright$ allows for endogenous lists focusing on attention overload across choice sets.

Finally, outside of the literature on choice over list, Honda (2021) considers a *Random Craving Model* (RCM). RCM is a special case of RUM where the set of preferences is restricted to the temptation preferences. In particular, for a true underlying preference $\succ$, the collection of temptation preference is defined by improving the ranking of an element in the preference to the best position. In our notation, for m alternatives, $\{\succ_{i1}\}_{i=1}^m$ is a collection of temptation. Therefore, RCM can be seen as a special case of HAOM$_\triangleright$ where full attention is assumed and the reference point can only be moved to the top.

SA.2 Simulation Evidence

This section presents simulation evidence on the finite sample properties of our econometric methods.

SA.2.1 Homogeneous Preferences

The grand set contains six alternatives, and without loss of generality, we assume the preference ordering $a_1 \succ a_2 \succ \cdots \succ a_6$. We employ the logit attention rule of Brady and Rehbeck (2016), which takes the form

$$\mu(T|S) = \frac{|T|^\varsigma}{\sum_{T' \subseteq S} |T'|^\varsigma},$$

where recall that $|T|$ denotes the size (cardinality) of a set. For specificity, we set $\varsigma = 2$ in our simulation study. Explicit calculation shows that this logit attention rule satisfies attention overload (Assumption 1).[1] With the preference and the attention rule introduced above, we are able to find the choice rule according to Definition 2. We then generate the choice data. In our simulation studies, we assume each choice problem has the same effective sample size $N_S \in \{50, 100, 200\}$.

[1] With the logit attention rule, the attention frequency only depends on the size of the choice problem but not the alternatives. For $|S| = 2, 3, 4, 5, 6$, the attention frequencies are 0.833, 0.750, 0.700, 0.667 and 0.643, respectively.

We implement our test (Theorem 1) against four hypothesized preferences. The first preference, $a_1 \succ a_2 \succ a_3 \succ a_4 \succ a_5 \succ a_6$, is compatible with our AOM, and hence we do not expect rejecting the null hypothesis very often. In other words, the rejection probability in this case corresponds to the size of our test. The other three preferences, however, cannot be represented by the AOM, and therefore the rejection probabilities will shed light on the power of the test. In each simulation setting, we conduct 2,000 Monte Carlo repetitions and the average rejection probabilities are reported in Table 2 (row "**rej prob**"). The nominal size is set to be 0.05. Our implementation employs the generalized moment selection method (Andrews and Soares, 2010, equation 4.4).

We focus on the first column, which reports simulation evidence for the complete data scenario (i.e., all 57 choice problems are available in the simulated dataset). There are 664 inequality constraints in total (row "**# restrictions**"). As the first preference ordering satisfies our AOM, none of these constraints will be violated. On the other hand, for the other three preferences, 90, 6 and 23 out of the 664 inequalities are strictly positive (row "**# violations**"). We also show the largest inequality constraint (row "**max inequality**"). A larger number indicates that the preference is further away from the null space, and hence it should be easier for our test to detect. To be more precise, for the specific logit attention rule that we consider, the AOM is best at eliciting the "best" alternative, that is, it is most powerful against mistakes regarding the most preferred alternative, a_1. As we can see from the table (row "**rej prob**"), among the three preferences in the alternative space, the AOM has the highest power for testing the second preference, $a_2 \succ a_3 \succ a_4 \succ a_5 \succ a_6 \succ a_1$. This is because by placing a_1 in the last position, lots of the inequalities will be violated according to our AOM. As a comparison, although the third and fourth preferences are also rejected by our AOM, they are relatively close to the null space due to the small fraction of violated inequality constraints, and hence the power of our test is not very high.

In practice, not all choice problems are available. Fortunately, our Theorem 1 adapts to incomplete data. As a sensitively analysis, we also conduct simulation studies where only a subset of choice problems are used to test against each of the four preferences. The results are reported in columns 2 to 7. For example, the second column of Table 2 contains rejection probabilities when the data consists of choice problems of size 3, 4, 5, and 6. It is not surprising that the number of inequalities constraints varies significantly: while there are 664 inequalities in the complete data scenario (column 1), there are only 15 constraints for the setting of Column 6 (choice problems of size 2 and 6; that is, only the grand set and binary comparisons are available). Overall, we see that our test demonstrates satisfactory size and power properties.

As our testing procedure involves estimating and evaluating a large number of inequality constraints, it can be conservative. For theoretical comparison, we also show the size adjusted rejection probabilities (row "**rej prob (size adj)**"). These are obtained by employing the (infeasible) critical values which are simulated from the correctly centered multivariate normal distribution. It should not come as a surprise that the empirical rejection probabilities are much closer to the nominal size for the first preference, and that the power of the test becomes much higher.

Table 2: Empirical Rejection Probabilities.

		Choice Problem Size $\lvert S\rvert$						
		$2,\dots,6$	$3,\dots,6$	$4,5,6$	$5,6$	$2,3,4,6$	$2,3,6$	$2,6$
# restrictions		664	439	159	24	370	115	15
$a_1 \succ a_2 \succ a_3 \succ a_4 \succ a_5 \succ a_6$								
# violations		0	0	0	0	0	0	0
max inequality		−0.024	−0.024	−0.024	−0.024	−0.050	−0.083	−0.190
rej prob	$N_S = 50$	0.014	0.016	0.033	0.030	0.012	0.004	0.000
	$N_S = 100$	0.006	0.008	0.018	0.024	0.005	0.002	0.000
	$N_S = 200$	0.004	0.003	0.010	0.020	0.002	0.000	0.000
rej prob (size adj)	$N_S = 50$	0.074	0.069	0.064	0.054	0.060	0.062	0.062
	$N_S = 100$	0.051	0.063	0.052	0.054	0.048	0.058	0.056
	$N_S = 200$	0.056	0.054	0.048	0.050	0.049	0.045	0.068
$a_2 \succ a_3 \succ a_4 \succ a_5 \succ a_6 \succ a_1$								
# violations		90	64	25	4	51	15	1
max inequality		0.071	0.030	0.013	0.005	0.071	0.071	0.071
rej prob	$N_S = 50$	0.159	0.103	0.088	0.054	0.162	0.142	0.180
	$N_S = 100$	0.260	0.120	0.088	0.077	0.267	0.253	0.322
	$N_S = 200$	0.468	0.168	0.092	0.065	0.464	0.457	0.512
rej prob (size adj)	$N_S = 50$	0.182	0.116	0.084	0.058	0.190	0.202	0.240
	$N_S = 100$	0.280	0.124	0.080	0.076	0.292	0.306	0.366
	$N_S = 200$	0.478	0.170	0.085	0.064	0.486	0.490	0.545
$a_1 \succ a_2 \succ a_6 \succ a_5 \succ a_4 \succ a_3$								
# violations		6	1	0	0	6	4	0
max inequality		0.058	0.021	−0.008	−0.024	0.058	0.042	−0.140
rej prob	$N_S = 50$	0.038	0.014	0.032	0.030	0.036	0.052	0.000
	$N_S = 100$	0.064	0.010	0.016	0.024	0.081	0.089	0.000
	$N_S = 200$	0.136	0.020	0.009	0.020	0.168	0.214	0.000
rej prob (size adj)	$N_S = 50$	0.113	0.068	0.063	0.054	0.116	0.160	0.051
	$N_S = 100$	0.186	0.076	0.050	0.054	0.228	0.241	0.052
	$N_S = 200$	0.350	0.116	0.046	0.050	0.409	0.381	0.062
$a_1 \succ a_6 \succ a_5 \succ a_4 \succ a_3 \succ a_2$								
# violations		23	7	1	0	20	10	0
max inequality		0.067	0.033	0.008	−0.003	0.058	0.042	−0.083
rej prob	$N_S = 50$	0.080	0.017	0.025	0.025	0.083	0.106	0.001
	$N_S = 100$	0.139	0.038	0.014	0.020	0.156	0.168	0.000
	$N_S = 200$	0.344	0.091	0.016	0.017	0.377	0.366	0.000
rej prob (size adj)	$N_S = 50$	0.164	0.054	0.049	0.052	0.161	0.200	0.059
	$N_S = 100$	0.271	0.124	0.040	0.050	0.305	0.268	0.053
	$N_S = 200$	0.548	0.248	0.066	0.043	0.578	0.453	0.053

Note. Shown in the table are the empirical rejection probabilities of our test for four preference orderings and different sets of inequality constraints. The results are based on 2,000 Monte Carlo simulations with nominal size 0.05. The effective sample size for each choice problem is 50, 100, or 200. Across the columns, we vary the set of choice problems available. For example, Column 1 contains simulations results for the complete data scenario ($\lvert S\rvert = 2, 3, \dots, 6$), while Column 7 only employs the grand set and binary comparisons ($\lvert S\rvert = 2, 6$).

Implementing Theorem 1 can be numerically challenging when there are many alternatives to choose from. For example, $\succ$-Regularity implies $\sum_{k=3}^{|X|}\binom{|X|}{k}\sum_{\ell=2}^{k-1}\binom{k}{\ell}(\ell-1)$ inequalities in a complete data scenario. Before closing this subsection, we illustrate the computing time needed by our numerical procedure (on a standard MacBook). We first abstract away from econometric implementation aspects; that is, we assume that the choice rule π is observed in the discussion below.

Testing $\succ$-Regularity involves two steps: organizing choice problems into $T \subset S$ pairs, and comparing choice probabilities according to $\succ$-Regularity. It turns out that the first step is much more time consuming. Once subset relationships have been established, comparing choice probabilities across choice problems is quite straightforward. For example, we form $T \subset S$ pairs given a collection of choice problems using for-loops. With 10 alternatives in the grand set X, this takes about seven seconds (only needs to be done once, since subset relationship is not specific to any preference ordering). Once we form subset pairs, checking $\succ$-Regularity only takes about 0.007 seconds for any preference. Even if we increase the number of alternatives to $|X| = 15$ (which is a very large choice problem), forming subset pairs will take about 30 minutes, after which each preference can be tested in about three seconds. Econometric implementation of Theorem 1 requires an additional step: computing the critical values. This requires simulating normal random vectors and is usually quite fast.

Finally, we illustrate numerically the computation gains from using Proposition 1 relative to Theorem 1. As we demonstrated in the main paper (Example 1), employing Proposition 1 first can drastically reduce the number of preference orderings to be tested. This potentially leads to substantial time save and makes our numerical procedure more scalable. In fact, employing Proposition 1 first has another advantage: it is also numerically much faster to implement than checking $\succ$-Regularity. This is because one does not need to form all subset relations. With 15 alternatives in the grand set, checking Proposition 1 for one pair ($a \succ b$) only takes about 0.02 seconds. Even checking all pairs of alternatives will take less than two seconds. If we increase $|X|$ to 25, numerically testing the inequalities in Proposition 1 takes about 30 seconds for one pair of alternatives. Therefore, we believe that Proposition 1 and Theorem 1 are complimentary. It is much easier to first use Proposition 1 to screen out certain preference orderings, and then test the remaining with Theorem 1.

As a final remark, the computing time reported above reflects the worst case scenario. First, our algorithm does not assume any particular structure on the input, and only requires a collection of choice problems and choice probabilities. As such, the algorithm has to loop over choice problems and alternatives. Second, when there are many alternatives in the grand set X, it is unlikely that one has complete data: many choice problems will not be present. This will decrease the number of subset pairs and the number of inequalities to compute/compare.

SA.2.2 Heterogeneous Preferences

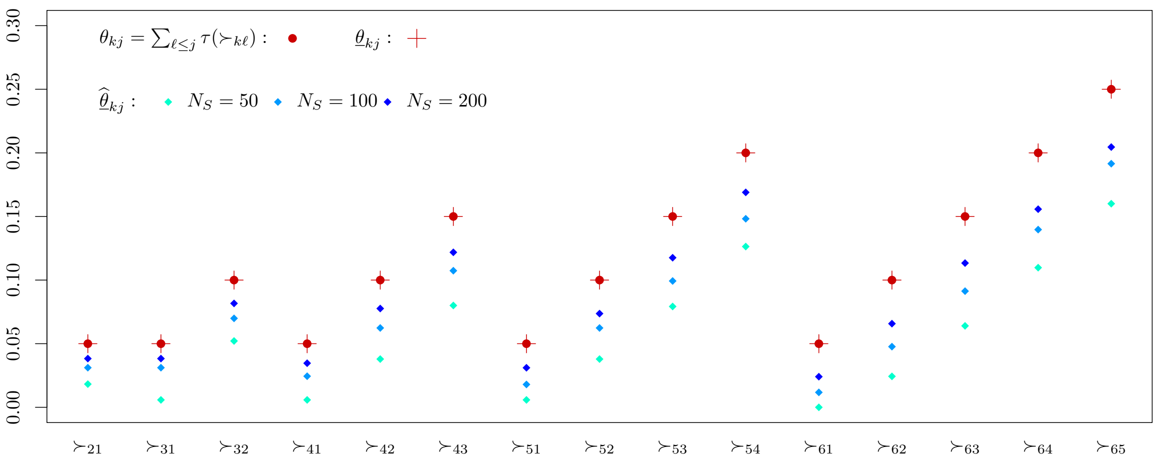
We also provide simulation evidence on the identification of heterogeneous preferences when alternatives are ordered in a list. Recall from Section 3 of the main paper that the list is represented by $\langle a_1, a_2, \ldots, a_m\rangle$. Given a choice problem S, we also recall that its elements are labeled by $s_1, s_2, \ldots, s_{|S|}$. We continue employing the logit attention rule of Brady and Rehbeck (2016). In addition, the first two alternatives in each choice problem, a_{s_1} and a_{s_2}, always receive attention. For this simulation study, we assume there are six alternatives in the grand set, that is, $m = 6$. The preferences are distributed as $\tau(\succ_{kj}) = 0.05$ for $k > j$, and $\tau(\succ_{11}) = 0.25$. This leads to

$$\theta_{kj} = \sum_{\ell \leq j} \tau(\succ_{k\ell}) = 0.05j.$$

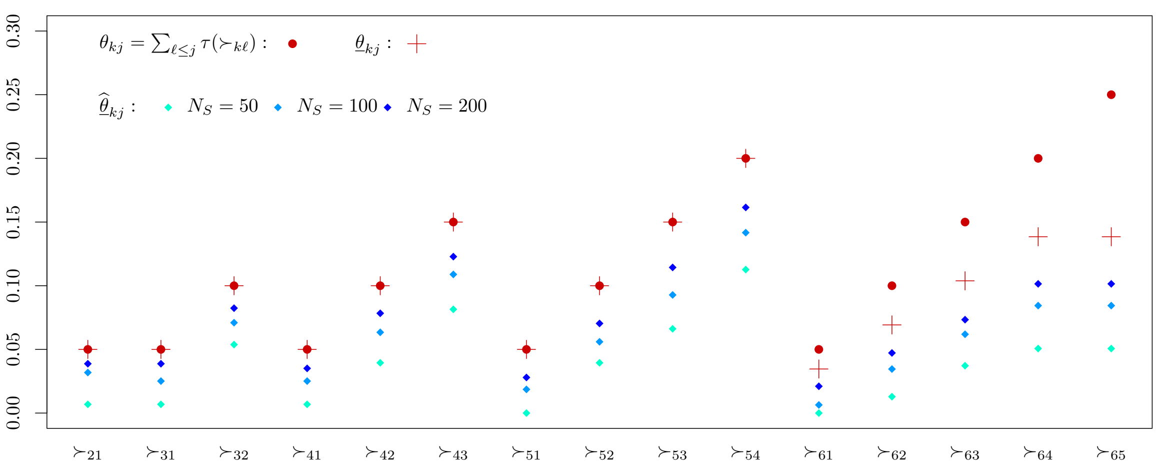
We further assume the random attention is independent of the heterogeneous preference for simplicity. It is worth noting that our theory and econometric implementation do not require such independence.

In Figure 1, we first show the true preference distribution (red dots). However, depending on which choice problems are available, the lower bound implied by Proposition 3 (red crosses) may not coincide with the true preference distribution. For example, a_6 never receives full attention unless binary comparisons are available. Therefore, we see from panel (b)–(d) that the the lower bounds involving a_6 (i.e., $\underline{\theta}_{6j}$ for $j = 1, 2, \ldots, 5$) always sit below the true preference distributions (θ_{6j} for $j = 1, 2, \ldots, 5$).

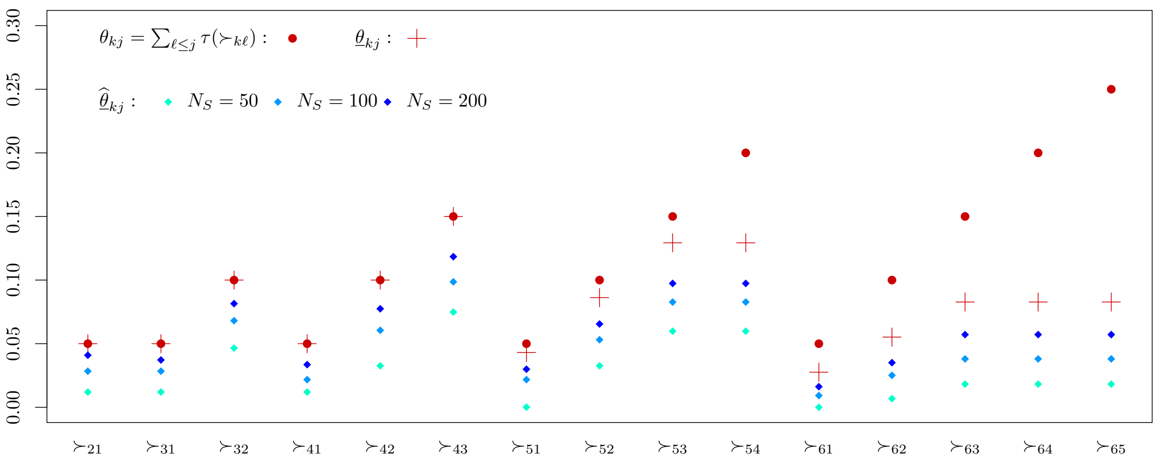
We also conducted 2,000 Monte Carlo simulations to investigate the performance of our econometric procedure for bounding the preference distribution. In each simulation iteration, we estimate the lower bound using the methods proposed in Section 3.5. We set $\alpha = 0.05$, which means the estimates should not cross the theoretical bounds more than 5%. In the same figure, we plot the 95th percentile of the estimated lower bounds for different effective sample sizes (blue diamonds). Indeed, they are always below the theoretical lower bounds.



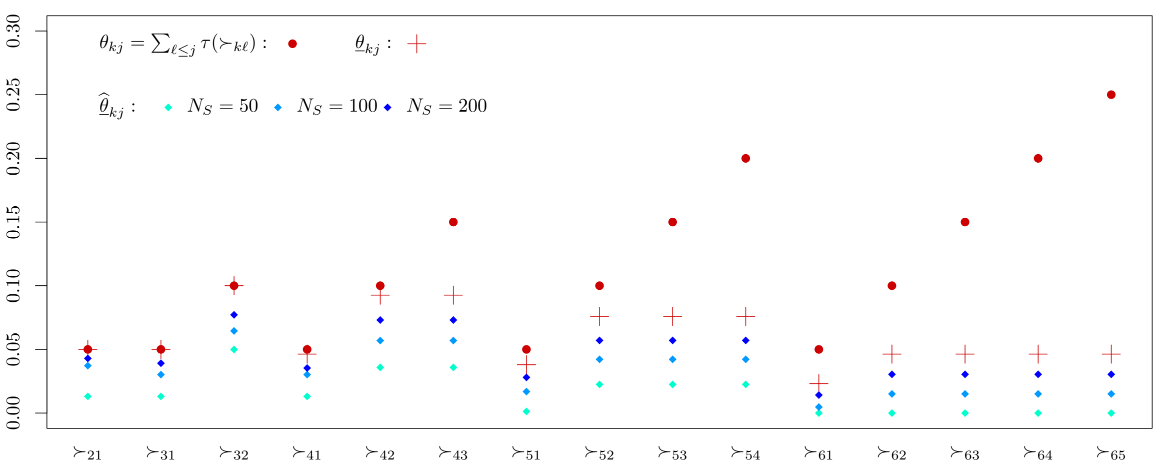
(a) $|S| = 2, \ldots, 6$



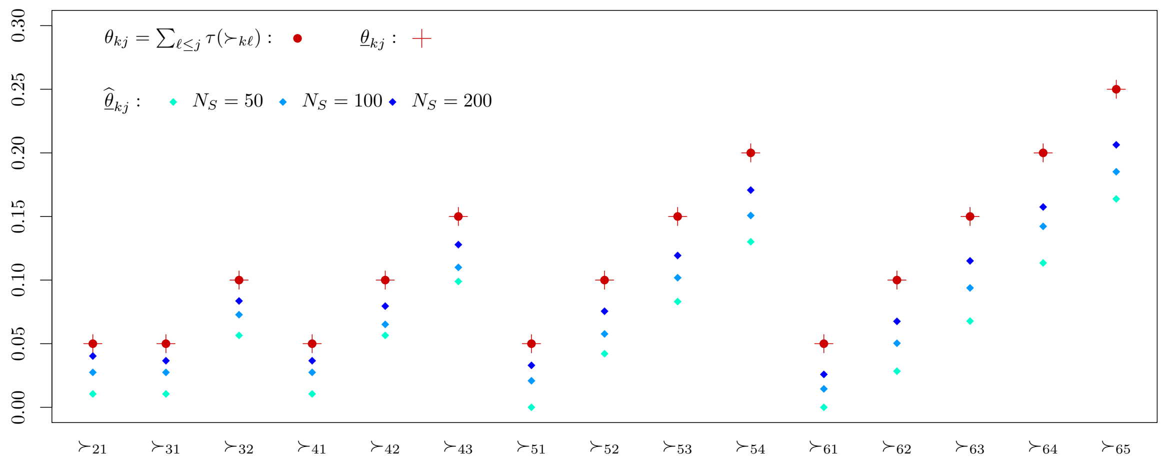
(b) $|S| = 3, \ldots, 6$


$\theta_{kj} = \sum_{\ell \leq j} \tau(\succ_{k\ell})$:
$\underline{\theta}_{kj}$:
$\widehat{\underline{\theta}}_{kj}$: $N_S = 50$ $N_S = 100$ $N_S = 200$
0.00
0.05
0.10
0.15
0.20
0.25
0.30
$\succ_{21}$ $\succ_{31}$ $\succ_{32}$ $\succ_{41}$ $\succ_{42}$ $\succ_{43}$ $\succ_{51}$ $\succ_{52}$ $\succ_{53}$ $\succ_{54}$ $\succ_{61}$ $\succ_{62}$ $\succ_{63}$ $\succ_{64}$ $\succ_{65}$

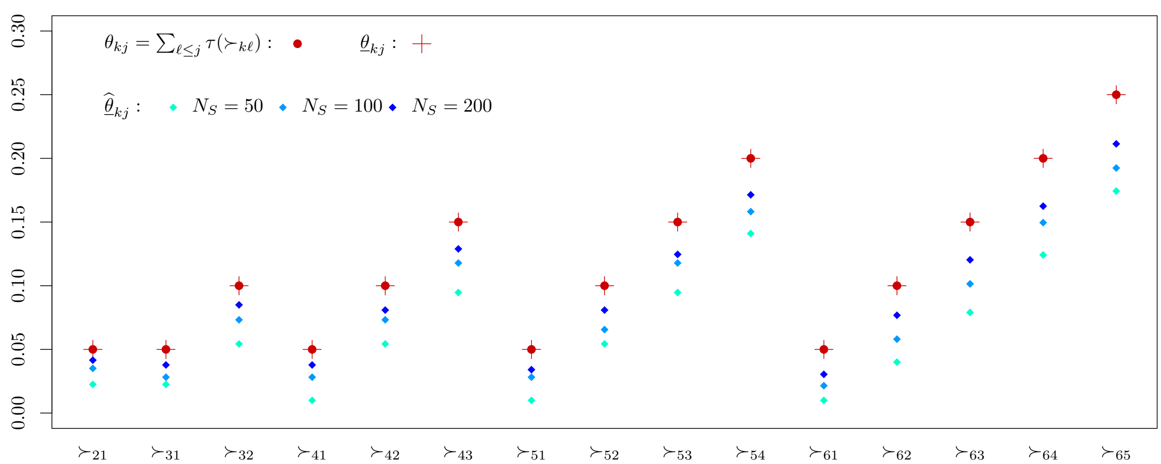
(c) $|S| = 4, 5, 6$


$\theta_{kj} = \sum_{\ell \leq j} \tau(\succ_{k\ell})$:
$\underline{\theta}_{kj}$:
$\widehat{\underline{\theta}}_{kj}$: $N_S = 50$ $N_S = 100$ $N_S = 200$
0.00
0.05
0.10
0.15
0.20
0.25
0.30
$\succ_{21}$ $\succ_{31}$ $\succ_{32}$ $\succ_{41}$ $\succ_{42}$ $\succ_{43}$ $\succ_{51}$ $\succ_{52}$ $\succ_{53}$ $\succ_{54}$ $\succ_{61}$ $\succ_{62}$ $\succ_{63}$ $\succ_{64}$ $\succ_{65}$

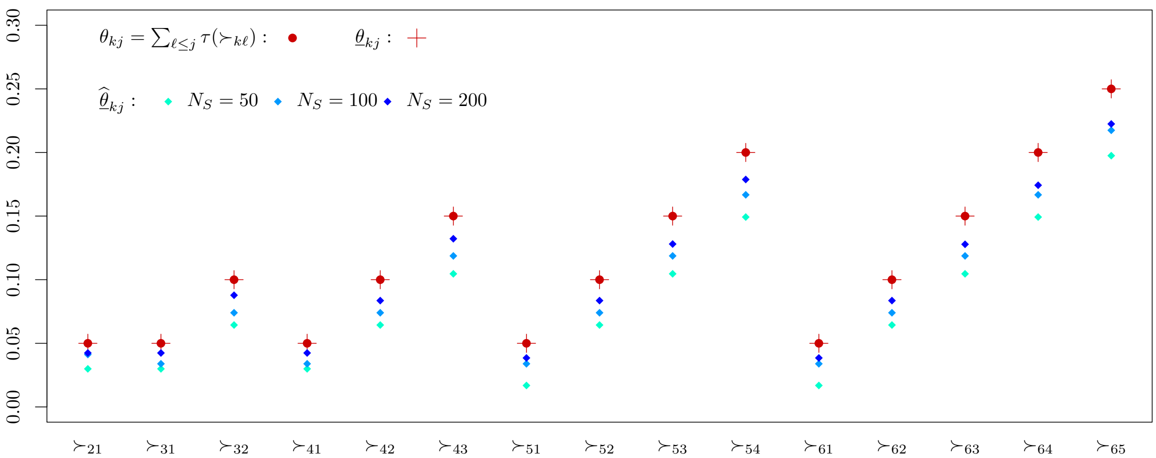
(d) $|S| = 5, 6$



(e) $|S| = 2, 3, 4, 6$



(f) $|S| = 2, 3, 6$



(g) $|S| = 2, 6$

Figure 1: Theoretical and empirical preference distribution.

Note. Shown in the figure are theoretical preference distributions (θ_{kj}, red dots), their theoretical lower bounds suggested by Proposition 3 (red crosses), and the empirically constructed lower bounds (blue diamonds). In each simulation repetition, we set the effective sample size to $50, 100$, and 200. We also set $\alpha = 0.05$ for simulating the critical values. Across the panels (a)–(g), we change what choice problems are available in the data. For example, panel (a) represents a complete data scenario, in which choice problems of all sizes ($|S| = 2, 3, \ldots, 6$) are available. In panel (c), our simulated data only contains choice problems of sizes $|S| = 4, 5, 6$.

SA.3 Omitted Details in Appendix A.4

For ease of presentation, we will first write the choice probability as a vector, which is denoted by $\boldsymbol{\pi}$. This will also allow us to collect all constraints implied by $\succ$-Regularity into a matrix, as the following example demonstrates.

Example SA.1. Assume there are five alternatives in the grand set: $X = \{a, b, c, d, e\}$, and we observe choice probabilities for two choice problems, $\mathcal{D} = \{\{a, b, c\}, \{a, b, c, d, e\}\}$. First consider the preference $a \succ b \succ c \succ d \succ e$, which leads to three constrains:

$$\begin{aligned}
&\pi(a|\{a,b,c,d,e\}) \leq \pi(a|\{a,b,c\}) \\
&\pi(b|\{a,b,c,d,e\}) \leq \pi(a|\{a,b,c\}) + \pi(b|\{a,b,c\}) \\
&\pi(c|\{a,b,c,d,e\}) \leq \pi(a|\{a,b,c\}) + \pi(b|\{a,b,c\}) + \pi(c|\{a,b,c\}).
\end{aligned}$$

If we rewrite the choice probabilities as a long vector, such as

$$\boldsymbol{\pi} = \Big(\underbrace{\pi(a|\cdot), \pi(b|\cdot), \pi(c|\cdot)}_{\{a,b,c\}}, \underbrace{\pi(a|\cdot), \pi(b|\cdot), \pi(c|\cdot), \pi(d|\cdot), \pi(e|\cdot)}_{\{a,b,c,d,e\}}\Big)^{\top},$$

then the three constrains will correspond to:

$$\mathbf{R}_{a\succ b\succ c\succ d\succ e}\boldsymbol{\pi} \leq 0, \quad \text{where } \mathbf{R}_{a\succ b\succ c\succ d\succ e} = \begin{bmatrix} -1 & 0 & 0 & +1 & 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & +1 & 0 & 0 & 0 \\ -1 & -1 & -1 & 0 & 0 & +1 & 0 & 0 \end{bmatrix}.$$

Note that the matrix, $\mathbf{R}_{a\succ b\succ c\succ d\succ e}$, only depends on the preference ordering and how the choice probabilities are ordered in the long vector $\boldsymbol{\pi}$. Importantly, one does not need the choice probabilities to construct this matrix, and hence it is nonrandom.

Constructing the matrix of inequality constraints for other preference orderings is also straightforward. For example, the following corresponds to the preference $c \succ d \succ e \succ b \succ a$

$$\mathbf{R}_{c\succ d\succ e\succ b\succ a} = \begin{bmatrix} 0 & 0 & -1 & 0 & 0 & +1 & 0 & 0 \\ 0 & -1 & -1 & 0 & +1 & 0 & 0 & 0 \\ -1 & -1 & -1 & +1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

■

From the previous example, it should be clear that the set of inequality constraints depends not only on the specific preference ordering under consideration but also the collection of choice problems available to the researcher. As an instance, consider the same setting but assume now that $\mathcal{D}$ contains $\{a, b, c\}$, $\{a, b, c, d\}$ and $\{a, b, c, d, e\}$. Then it is not difficult to see that now there will be $3 + 3 + 4 = 10$ inequality constraints for the preference $a \succ b \succ c \succ d \succ e$.

Now recall from the main paper that the choice probabilities are estimated by the sub-sample averages

$$\hat{\pi}(a|S) = \frac{1}{N_S}\sum_{i=1}^{n} \mathbb{1}(y_i = a, Y_i = S),$$

where $N_S = \sum_{i=1}^{n} \mathbb{1}(Y_i = S)$ is the effective sample size for the choice problem S. For developing econometric methods and establish their formal statistical properties, it is more convenient to also write the vector of choice probabilities as an average. Consider the following example.

Example SA.2. Consider the same setting as in Example SA.1, where the grand set is $X = \{a, b, c, d, e\}$, and two choice problems, $\mathcal{D} = \{\{a, b, c\}, \{a, b, c, d, e\}\}$ are observed in the date. We will label the two choice problems by S_1 and S_2 respectively, and N_{S_1} and N_{S_2} their effective sample sizes. To estimate the vector of choice probabilities, we rewrite the original data as

$$\Big(\underbrace{\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_{N_{S_1}}}_{S_1=\{a,b,c\}}, \underbrace{\mathbf{z}_{N_{S_1}+1}, \mathbf{z}_{N_{S_1}+2}, \cdots, \mathbf{z}_{N_{S_1}+N_{S_2}}}_{S_2=\{a,b,c,d,e\}}\Big) = \begin{bmatrix} 0 & \frac{n}{N_{S_1}} & \cdots & \frac{n}{N_{S_1}} & & & & \\ 0 & 0 & \cdots & 0 & & & & \\ \frac{n}{N_{S_1}} & 0 & \cdots & 0 & & & & \\ & & & & \frac{n}{N_{S_2}} & 0 & \cdots & 0 \\ & & & & 0 & \frac{n}{N_{S_2}} & \cdots & 0 \\ & & & & 0 & 0 & \cdots & 0 \\ & & & & 0 & 0 & \cdots & 0 \\ & & & & 0 & 0 & \cdots & \frac{n}{N_{S_2}} \end{bmatrix}.$$

It should be clear that the estimated choice probabilities can be obtained by the sample average of $\mathbf{z}_i$: $\hat{\boldsymbol{\pi}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i$. ■

Albeit being cumbersome, expressing the original choice data into the vector form has two advantages. First, the estimated choice probabilities can be obtained by the sample average of $\mathbf{z}_i$, $\hat{\boldsymbol{\pi}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i$, as demonstrated in the previous example. More importantly, the new variables, $\mathbf{z}_i$, are mutually independent if N_S are nonrandom or if we condition on the realizations of the choice problems. However, it should be clear that $\mathbf{z}_i$ do not have the same distribution. The same conclusions holds even after pre-multiplying by the constraint matrix, $\mathbf{R}_{\succ}$.

$$\mathbf{R}_{\succ}\hat{\boldsymbol{\pi}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{R}_{\succ}\mathbf{z}_i = \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_{i,\succ}.$$

In fact, for each $\mathbf{z}_i$, pre-multiplying by a row of $\mathbf{R}_{\succ}$ can either leave $\mathbf{z}_i$ unchanged, alter its sign, or lead to a zero vector. The last scenario arises if a data point is not relevant for a specific constraint.

Recall from the main paper that each inequality restriction corresponds to a pair $T \subset S$ and an alternative $a \in T$. Therefore, we the corresponding row vector in $\mathbf{R}_{\succ}$ will be denoted by $\mathbf{r}_{\succ}(a|S,T)^{\top}$. We also define

$$z_{i,\succ}(a|S,T) = \mathbf{r}_{\succ}(a|S,T)^{\top}\mathbf{z}_i,$$

which is simply one element in $\mathbf{z}_{i,\succ}$. Also recall that the standard deviation of $\mathbf{r}_{\succ}(a|S,T)^{\top}\hat{\boldsymbol{\pi}}$ is denoted by $\sigma(a|S,T)$. We collect the individual standard deviations into the vector $\boldsymbol{\sigma}$ in a conformable way. In other words, $\boldsymbol{\sigma}^2$ contains the diagonal elements in the covariance matrix $\mathbf{R}_{\succ}\mathbb{V}[\hat{\boldsymbol{\pi}}]\mathbf{R}_{\succ}^{\top}$.

SA.3.1 Proof of Lemma A.4

Proof. We adopt the following result.

Lemma SA.1 (Theorem 2.1 in Chernozhukov, Chetverikov, Kato, and Koike 2022)**.** Let $\{\mathbf{x}_i, 1 \le i \le n\}$ be mean-zero independent random vectors of dimension $\mathfrak{c}_1$, and $\{\tilde{\mathbf{x}}_i, 1 \le i \le n\}$ be mean-zero independent normal random vectors such that $\mathbf{x}_i$ and $\tilde{\mathbf{x}}_i$ have the same covariance matrix. Assume the following holds.

(i) For some fixed constants $C, C' > 0$,

$$C \le \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[x_{i,\ell}^2] \le C', \quad \forall 1 \le \ell \le \mathfrak{c}_1.$$

(ii) For some fixed constant $C' > 0$ and some sequence $\mathfrak{c}_2 > 0$ which can depend on the sample size,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[|x_{i,\ell}|^4] \le \mathfrak{c}_2^2 C', \quad \forall 1 \le \ell \le \mathfrak{c}_1,\ \varepsilon = 1, 2,$$

and

$$\mathbb{E}\left[\exp\left(\frac{|x_{i,\ell}|^2}{\mathfrak{c}_2^2}\right)\right] \le 2, \forall 1 \le \ell \le \mathfrak{c}_1,\ 1 \le i \le n.$$

Then

$$\sup_{\substack{A \subseteq \mathbb{R}^{\mathfrak{c}_1} \\ A \text{ rectangular}}} \left| \mathbb{P}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{x}_i \in A\right] - \mathbb{P}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\mathbf{x}}_i \in A\right]\right| \le c\left(\frac{\mathfrak{c}_2^2\log^5(n\mathfrak{c}_1)}{n}\right)^{\frac{1}{4}},$$

where the constant c only depends on C and C' in conditions (i) and (ii).

We will ultimately apply Lemma SA.1 to

$$x_i(a|S,T) = \frac{z_{i,\succ}(a|S,T) - \mathbb{E}[z_{i,\succ}(a|S,T)]}{\sqrt{n}\sigma(a|S,T)}.$$

Condition (i) there is trivially satisfied with $C = C' = 1$. To verify condition (ii), we will need to take a closer look at the individual summands along each coordinate, $x_i(a|S,T)$. We first consider $z_{i,\succ}(a|S,T)$.

From the previous discuss, it should be clear that each constraint will involve comparing choice probabilities across two choice problems. This means that $z_{i,\succ}(a|S,T)$ is nonzero for at most $N_S + N_T$ observations (recall that N_S is the effective sample size for the choice problem S in the data). For those observations such that $z_{i,\succ}(a|S,T)$ is nonzero, we have

$$\begin{aligned}
|z_{i,\succ}(a|S,T) - \mathbb{E}[z_{i,\succ}(a|S,T)]| &\le \frac{n}{N_S} \text{ or } \frac{n}{N_T} \\
\mathbb{E}[|z_{i,\succ}(a|S,T) - \mathbb{E}[z_{i,\succ}(a|S,T)]|^{2+\varepsilon}] &\le 2\left(\frac{n}{N_S}\right)^{2+\varepsilon}\pi(a|S)(1-\pi(a|S)) \\
&\quad \text{or } \le\ 2\left(\frac{n}{N_T}\right)^{2+\varepsilon}\pi(U_{\succeq}(a)|T)(1-\pi(U_{\succeq}(a)|T)),
\end{aligned}$$

As a result,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[|z_{i,\succ}(a|S,T) - \mathbb{E}[z_{i,\succ}(a|S,T)]|^{2+\varepsilon}]$$

$$\leq 2\left[\left(\frac{n}{N_S}\right)^{1+\varepsilon}\pi(a|S)(1-\pi(a|S))+\left(\frac{n}{N_T}\right)^{1+\varepsilon}\pi(U_{\succeq}(a)|T)(1-\pi(U_{\succeq}(a)|T))\right].$$

Now consider $\sqrt{n}\sigma(a|S,T)$, which takes the form

$$\sqrt{n}\sigma(a|S,T)=\sqrt{\frac{n}{N_S}\pi(a|S)(1-\pi(a|S))+\frac{n}{N_T}\pi(U_{\succeq}(a)|T)(1-\pi(U_{\succeq}(a)|T))}.$$

Combining previous results, we have

$$\begin{aligned}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[|x_i(a|S,T)|^{2+\varepsilon}]\right)^{\frac{1}{\varepsilon}} &\leq \left(2\frac{\left(\frac{n}{N_S}\right)^{1+\varepsilon}\pi(a|S)(1-\pi(a|S))+\left(\frac{n}{N_T}\right)^{1+\varepsilon}\pi(U_{\succeq}(a)|T)(1-\pi(U_{\succeq}(a)|T))}{\left[\frac{n}{N_S}\pi(a|S)(1-\pi(a|S))+\frac{n}{N_T}\pi(U_{\succeq}(a)|T)(1-\pi(U_{\succeq}(a)|T))\right]^{1+\frac{\varepsilon}{2}}}\right)^{\frac{1}{\varepsilon}}\\ &\leq 2\left(\frac{n}{N_S}\vee\frac{n}{N_T}\right)\frac{1}{\sqrt{n}\sigma(a|S,T)}.\end{aligned}$$

To apply Lemma SA.1, set $\varepsilon=2$ and

$$\mathfrak{c}_2=2\sqrt{n}\left[\min_{S\in\mathcal{D}}N_S\right]^{-1}\left[\min_{\substack{a\in T\subset S\\ T,S\in\mathcal{D}}}\sigma(a|S,T)\right]^{-1}.$$

It is easy to see that the above choice satisfies the first part of condition (ii) in Lemma SA.1. For the second part in condition (ii), we note that

$$\begin{aligned}\left|\frac{x_i(a|S,T)}{\mathfrak{c}_2}\right| &= \left|\frac{z_{i,\succ}(a|S,T)-\mathbb{E}[z_{i,\succ}(a|S,T)]}{\sqrt{n}\sigma(a|S,T)}\right|\frac{\min_{S\in\mathcal{D}}N_S}{2\sqrt{n}}\left[\min_{\substack{a\in T\subset S\\ T,S\in\mathcal{D}}}\sigma(a|S,T)\right]\\ &\leq \left|z_{i,\succ}(a|S,T)-\mathbb{E}[z_{i,\succ}(a|S,T)]\right|\frac{\min_{S\in\mathcal{D}}N_S}{2n}.\end{aligned}$$

Next observe that $z_{i,\succ}(a|S,T)$ is simply a centered Bernoulli random variable scaled by n/N_S for some S, which means the second part of condition (ii) also holds with the above choice of $\mathfrak{c}_2$. ■

SA.3.2 Proof of Lemma A.5

For now, let $\hat{\boldsymbol{\sigma}}$ be some estimator, then

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\frac{z_{i,\succ}(a|S,T)-\mathbb{E}[z_{i,\succ}(a|S,T)]}{\hat{\sigma}(a|S,T)} &= \frac{1}{n}\sum_{i=1}^{n}\frac{z_{i,\succ}(a|S,T)-\mathbb{E}[z_{i,\succ}(a|S,T)]}{\sigma(a|S,T)}\\ &\quad+\left(\frac{1}{n}\sum_{i=1}^{n}\frac{z_{i,\succ}(a|S,T)-\mathbb{E}[z_{i,\succ}(a|S,T)]}{\sigma(a|S,T)}\right)\left(\frac{\sigma(a|S,T)}{\hat{\sigma}(a|S,T)}-1\right),\end{aligned}$$

which means

$$\max_{\substack{a\in T\subset S\\ T,S\in\mathcal{D}}}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{z_{i,\succ}(a|S,T)-\mathbb{E}[z_{i,\succ}(a|S,T)]}{\hat{\sigma}(a|S,T)}-\frac{1}{n}\sum_{i=1}^{n}\frac{z_{i,\succ}(a|S,T)-\mathbb{E}[z_{i,\succ}(a|S,T)]}{\sigma(a|S,T)}\right|$$

$$\leq \max_{\substack{a\in T\subset S\\ T,S\in\mathcal{D}}} \left| \frac{1}{n}\sum_{i=1}^{n} \frac{z_{i,\succ}(a|S,T) - \mathbb{E}[z_{i,\succ}(a|S,T)]}{\sigma(a|S,T)} \right| \cdot \max_{\substack{a\in T\subset S\\ T,S\in\mathcal{D}}} \left| \frac{\sigma(a|S,T)}{\hat\sigma(a|S,T)} - 1 \right|.$$

We will control the two terms on the right-hand side separately.

Let ξ_1 be some generic constant which can depend on the sample size. Then by the triangle inequality,

$$\begin{aligned}
\mathbb{P}\left[\max_{\substack{a\in T\subset S\\ T,S\in\mathcal{D}}} \left| \frac{1}{n}\sum_{i=1}^{n} \frac{z_{i,\succ}(a|S,T) - \mathbb{E}[z_{i,\succ}(a|S,T)]}{\sigma(a|S,T)} \right| \geq \xi_1 \right] &\leq \underbrace{\mathbb{P}\left[\|\check{\mathbf{z}}\|_\infty \geq \xi_1\right]}_{\text{(I)}} \\
&+ \underbrace{\left| \mathbb{P}\left[\max_{\substack{a\in T\subset S\\ T,S\in\mathcal{D}}} \left| \frac{1}{n}\sum_{i=1}^{n} \frac{z_{i,\succ}(a|S,T) - \mathbb{E}[z_{i,\succ}(a|S,T)]}{\sigma(a|S,T)} \right| \geq \xi_1 \right] - \mathbb{P}\left[\|\check{\mathbf{z}}\|_\infty \geq \xi_1\right] \right|}_{\text{(II)}}.
\end{aligned}$$

Note that $\check{\mathbf{z}}$ is defined in Lemma A.4.

Term (I) has the bound

$$\text{(I)} \leq c\xi_1^{-1}\sqrt{\log \mathfrak{c}_1}$$

by Markov's inequality, and c is an absolute constant. By Lemma A.4, term (II) is simply bounded by

$$\text{(II)} \leq c\left(\frac{\log^5(n\mathfrak{c}_1)}{\mathfrak{c}_2^2}\right)^{\frac{1}{4}},$$

which holds for any ξ_1. Again c is an absolute constant in the above.

Next consider the standard error estimator, $\hat\sigma(a|S,T)$. Then

$$\left|\hat\sigma(a|S,T)^2 - \sigma(a|S,T)^2\right| \leq \frac{1}{N_S}\left|\hat\pi(a|S) - \pi(a|S)\right| + \frac{1}{N_T}\left|\hat\pi(U_{\succeq}(a)|T) - \pi(U_{\succeq}(a)|T)\right|.$$

Consider, for example, the first term on the right-hand side in the above. Using Bernstein's inequality, one has

$$\begin{aligned}
\mathbb{P}\left[\frac{1}{N_S\sigma(a|S,T)^2}\left|\hat\pi(a|S) - \pi(a|S)\right| \geq \xi_2\right] &\leq 2\exp\left\{-\frac{1}{2}\frac{N_S^4\sigma(a|S,T)^4\xi_2^2}{N_S\pi(a|S)(1-\pi(a|S)) + \frac{1}{3}N_S^2\sigma(a|S,T)^2\xi_2}\right\} \\
&= 2\exp\left\{-\frac{1}{2}\frac{N_S^2\sigma(a|S,T)^4\xi_2^2}{\frac{1}{N_S}\pi(a|S)(1-\pi(a|S)) + \frac{1}{3}\sigma(a|S,T)^2\xi_2}\right\} \\
&\leq 2\exp\left\{-\frac{1}{2}\frac{N_S^2\sigma(a|S,T)^4\xi_2^2}{\sigma(a|S,T)^2 + \frac{1}{3}\sigma(a|S,T)^2\xi_2}\right\} \\
&= 2\exp\left\{-\frac{1}{2}\frac{N_S^2\sigma(a|S,T)^2\xi_2^2}{1+\frac{1}{3}\xi_2}\right\} \\
&\leq 2\exp\left\{-\frac{1}{4}N_S^2\sigma(a|S,T)^2\xi_2^2\right\}, \qquad \text{provided that } \xi_2 \to 0 \\
&\leq 2\exp\left\{-\frac{1}{4}\mathfrak{c}_2^2\xi_2^2\right\}.
\end{aligned}$$

Using the union bound, we deduce that

$$\mathbb{P}\left[\max_{\substack{a\in T\subset S\\ T,S\in\mathcal{D}}} \frac{\left|\hat{\sigma}(a|S,T)^2-\sigma(a|S,T)^2\right|}{\sigma(a|S,T)^2}\geq \xi_2\right]\leq 4\exp\left\{-\frac{1}{16}\mathfrak{c}_2^2\xi_2^2+\log\mathfrak{c}_1\right\},\qquad \text{provided that } \xi_2\to 0,$$

which also implies that

$$\mathbb{P}\left[\max_{\substack{a\in T\subset S\\ T,S\in\mathcal{D}}} \frac{\left|\hat{\sigma}(a|S,T)-\sigma(a|S,T)\right|}{\sigma(a|S,T)}\geq \xi_2\right]\leq 4\exp\left\{-\frac{1}{16}\mathfrak{c}_2^2\xi_2^2+\log\mathfrak{c}_1\right\},\qquad \text{provided that } \xi_2\to 0.$$

This closes the proof of the first claim.

To show the second claim, we note that given any two vectors, $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{R}^{\mathfrak{c}_1}$, one has

$$\begin{aligned}\max_{1\leq\ell\leq\mathfrak{c}_1} a_\ell-\max_{1\leq\ell\leq\mathfrak{c}_1} b_\ell &= \max_{1\leq\ell\leq\mathfrak{c}_1}(a_\ell-b_\ell+b_\ell)-\max_{1\leq\ell\leq\mathfrak{c}_1} b_\ell\\ &\leq \max_{1\leq\ell\leq\mathfrak{c}_1}(a_\ell-b_\ell)+\max_{1\leq\ell\leq\mathfrak{c}_1}+b_\ell-\max_{1\leq\ell\leq\mathfrak{c}_1} b_\ell\\ &\leq \max_{1\leq\ell\leq\mathfrak{c}_1}|a_\ell-b_\ell|.\end{aligned}$$

Similarly, one has the other direction because

$$\begin{aligned}\max_{1\leq\ell\leq\mathfrak{c}_1} a_\ell-\max_{1\leq\ell\leq\mathfrak{c}_1} b_\ell &= \max_{1\leq\ell\leq\mathfrak{c}_1} a_\ell-\max_{1\leq\ell\leq\mathfrak{c}_1}(b_\ell-a_\ell+a_\ell)\\ &\geq \max_{1\leq\ell\leq\mathfrak{c}_1} a_\ell-\max_{1\leq\ell\leq\mathfrak{c}_1}(b_\ell-a_\ell)-\max_{1\leq\ell\leq\mathfrak{c}_1} a_\ell\\ &\geq -\max_{1\leq\ell\leq\mathfrak{c}_1}|b_\ell-a_\ell|.\end{aligned}$$

Then taking

$$a_\ell=\frac{1}{n}\sum_{i=1}^{n}\frac{z_{i,\succ}(a|S,T)-\mathbb{E}[z_{i,\succ}(a|S,T)]}{\hat{\sigma}(a|S,T)},\qquad b_\ell=\frac{1}{n}\sum_{i=1}^{n}\frac{z_{i,\succ}(a|S,T)-\mathbb{E}[z_{i,\succ}(a|S,T)]}{\sigma(a|S,T)}$$

will lead to the desired result.

SA.3.3 Proof of Lemma A.6

See the proof of Lemma A.5.

SA.3.4 Proof of Lemma A.7

Recall that each restriction involves comparing choice probabilities across two choice problems. For two restrictions, ℓ and ℓ', there will be at most four choice problems involved, which we denote by $T\subset S$ and $T'\subset S'$. Now consider the case where the two restrictions are non-overlapping, meaning that $T\neq T'$ or S' and $S\neq T'$ or S'. Then both the population covariance/correlation and its estimate will be zero. (The reason that the estimated covariance/correlation is because the "middle matrix," $\hat{\mathbb{V}}[\hat{\boldsymbol{\pi}}]$, is block diagonal.) As a result, the estimation error of the correlation matrix is trivially 0 in this special case.

Given the previous discussions, we will consider the estimation error when the two restrictions involve overlapping choice problems.

To start, we first decompose the difference as

$$
\begin{aligned}
&\left|\frac{\hat{\sigma}(a|S,T;a'|S',T')}{\hat{\sigma}(a|S,T)\hat{\sigma}(a'|S',T')} - \frac{\sigma(a|S,T;a'|S',T')}{\sigma(a|S,T)\sigma(a'|S',T')}\right| \\
&\le \left|\frac{\hat{\sigma}(a|S,T;a'|S',T') - \sigma(a|S,T;a'|S',T')}{\sigma(a|S,T)\sigma(a'|S',T')}\right| + \left|\frac{\hat{\sigma}(a|S,T;a'|S',T')}{\hat{\sigma}(a|S,T)\hat{\sigma}(a'|S',T')}\right| \cdot \left|\frac{\hat{\sigma}(a|S,T)\hat{\sigma}(a'|S',T')}{\sigma(a|S,T)\sigma(a'|S',T')} - 1\right| \\
&\le \left|\frac{\hat{\sigma}(a|S,T;a'|S',T') - \sigma(a|S,T;a'|S',T')}{\sigma(a|S,T)\sigma(a'|S',T')}\right| + \left|\frac{\hat{\sigma}(a|S,T)\hat{\sigma}(a'|S',T')}{\sigma(a|S,T)\sigma(a'|S',T')} - 1\right| \\
&\le \left|\frac{\hat{\sigma}(a|S,T;a'|S',T') - \sigma(a|S,T;a'|S',T')}{\sigma(a|S,T)\sigma(a'|S',T')}\right| + 2\left|\frac{\hat{\sigma}(a|S,T)}{\sigma(a|S,T)} - 1\right| \vee \left|\frac{\hat{\sigma}(a'|S',T')}{\sigma(a'|S',T')} - 1\right| \\
&\qquad\qquad + \left|\frac{\hat{\sigma}(a|S,T)}{\sigma(a|S,T)} - 1\right| \cdot \left|\frac{\hat{\sigma}(a'|S',T')}{\sigma(a'|S',T')} - 1\right|.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\max_{\substack{a\in T\subset S\\ T,S\in\mathcal{D}}}\ \max_{\substack{a'\in T'\subset S'\\ T',S'\in\mathcal{D}}} \left|\frac{\hat{\sigma}(a|S,T;a'|S',T')}{\hat{\sigma}(a|S,T)\hat{\sigma}(a'|S',T')} - \frac{\sigma(a|S,T;a'|S',T')}{\sigma(a|S,T)\sigma(a'|S',T')}\right| \\
&\le \max_{\substack{a\in T\subset S\\ T,S\in\mathcal{D}}}\ \max_{\substack{a'\in T'\subset S'\\ T',S'\in\mathcal{D}}} \left|\frac{\hat{\sigma}(a|S,T;a'|S',T') - \sigma(a|S,T;a'|S',T')}{\sigma(a|S,T)\sigma(a'|S',T')}\right| + 2\max_{\substack{a\in T\subset S\\ T,S\in\mathcal{D}}}\left|\frac{\hat{\sigma}(a|S,T)}{\sigma(a|S,T)} - 1\right| + \max_{\substack{a\in T\subset S\\ T,S\in\mathcal{D}}}\left|\frac{\hat{\sigma}(a|S,T)}{\sigma(a|S,T)} - 1\right|^2.
\end{aligned}
$$

We already have an error bound for the last two terms from Lemma A.6, and hence it suffices to study the first term in the above display.

To proceed, consider four possible scenarios.

(i) $S = S' \supset T \neq T'$.

(ii) $S \neq S' \supset T = T'$.

(iii) $S \supset T = S' \supset T'$.

(iv) $S = S' \supset T = T'$.

Case (i). The covariance $\sigma(a|S,T;a'|S',T')$ can take two forms,

$$
\sigma(a|S,T;a'|S',T') = \underbrace{\frac{1}{N_S}\pi(a|S)(1-\pi(a|S))}_{\text{if } a = a'} \qquad \text{or} \qquad \underbrace{-\frac{1}{N_S}\pi(a|S)\pi(a'|S)}_{\text{if } a \neq a'}.
$$

Nevertheless, the following bound applies

$$
|\hat{\sigma}(a|S,T;a'|S',T') - \sigma(a|S,T;a'|S',T')| \le \frac{1}{N_S}\left|\hat{\pi}(a|S) - \pi(a|S)\right| + \frac{1}{N_S}\left|\hat{\pi}(a'|S) - \pi(a'|S)\right|.
$$

Next we apply Bernstein's inequality, which gives

$$\begin{aligned}
&\mathbb{P}\left[\frac{1}{N_S\sigma(a|S,T)\sigma(a'|S',T')}\,|\hat{\pi}(a|S)-\pi(a|S)|\geq\xi_2\right]\\
&\leq 2\exp\left\{-\frac{1}{2}\frac{N_S^4\sigma(a|S,T)^2\sigma(a'|S',T')^2\xi_2^2}{N_S\pi(a|S)(1-\pi(a|S))+\frac{1}{3}N_S^2\sigma(a|S,T)\sigma(a'|S',T')\xi_2}\right\}\\
&=2\exp\left\{-\frac{1}{2}\frac{N_S^2\sigma(a|S,T)^2\sigma(a'|S',T')^2\xi_2^2}{\frac{1}{N_S}\pi(a|S)(1-\pi(a|S))+\frac{1}{3}\sigma(a|S,T)\sigma(a'|S',T')\xi_2}\right\}\\
&\leq 2\exp\left\{-\frac{1}{2}\frac{N_S^2\sigma(a|S,T)^2\sigma(a'|S',T')^2\xi_2^2}{\sigma(a|S,T)^2+\frac{1}{3}\sigma(a|S,T)\sigma(a'|S',T')\xi_2}\right\}\\
&\leq 2\exp\left\{-\frac{1}{2}\frac{N_S^2\sigma(a|S,T)^2\sigma(a'|S',T')^2\xi_2^2}{(\sigma(a|S,T)\vee\sigma(a'|S',T'))^2+\frac{1}{3}(\sigma(a|S,T)\vee\sigma(a'|S',T'))^2\xi_2}\right\}\\
&=2\exp\left\{-\frac{1}{2}\frac{N_S^2(\sigma(a|S,T)\wedge\sigma(a'|S',T'))^2\xi_2^2}{1+\frac{1}{3}\xi_2}\right\}\\
&\leq 2\exp\left\{-\frac{1}{4}N_S^2(\sigma(a|S,T)\wedge\sigma(a'|S',T'))^2\xi_2^2\right\},\qquad\text{provided that }\xi_2\to 0\\
&\leq 2\exp\left\{-\frac{1}{4}\mathfrak{c}_2^2\xi_2^2\right\}.
\end{aligned}$$

Therefore, we have

$$\mathbb{P}\left[\frac{1}{\sigma(a|S,T)\sigma(a'|S',T')}\,|\hat{\sigma}(a|S,T;a'|S',T')-\sigma(a|S,T;a'|S',T')|\geq\xi_2\right]\leq 4\exp\left\{-\frac{1}{16}\mathfrak{c}_2^2\xi_2^2\right\}.$$

Case (ii). The covariance $\sigma(a|S,T;a'|S',T')$ can be conveniently written as

$$\sigma(a|S,T;a'|S',T')=\frac{1}{N_T}\Big(\pi(U_{\succeq}(a)|T)\wedge\pi(U_{\succeq}(a')|T)-\pi(U_{\succeq}(a)|T)\pi(U_{\succeq}(a')|T)\Big).$$

Then we have the bound

$$\begin{aligned}
|\hat{\sigma}(a|S,T;a'|S',T')-\sigma(a|S,T;a'|S',T')|&\leq\frac{1}{N_T}\,|\hat{\pi}(U_{\succeq}(a)|T)-\pi(U_{\succeq}(a)|T)|+\frac{1}{N_{T'}}\,|\hat{\pi}(U_{\succeq}(a')|T')-\pi(U_{\succeq}(a')|T')|\\
&=\frac{1}{N_T}\,|\hat{\pi}(U_{\succeq}(a)|T)-\pi(U_{\succeq}(a)|T)|+\frac{1}{N_T}\,|\hat{\pi}(U_{\succeq}(a')|T)-\pi(U_{\succeq}(a')|T)|\,.
\end{aligned}$$

Again by using the union bound and Bernstein's inequality, one has

$$\mathbb{P}\left[\frac{1}{\sigma(a|S,T)\sigma(a'|S',T')}\,|\hat{\sigma}(a|S,T;a'|S',T')-\sigma(a|S,T;a'|S',T')|\geq\xi_2\right]\leq 4\exp\left\{-\frac{1}{16}\mathfrak{c}_2^2\xi_2^2\right\}.$$

Case (iii). The covariance $\sigma(a|S,T;a'|S',T')$ can take two forms,

$$\sigma(a|S,T;a'|S',T')=\underbrace{-\frac{1}{N_T}\pi(U_{\succeq}(a)|T)\pi(a'|S')}_{\text{if }a'\notin U_{\succeq}(a)}\quad\text{or}\quad\underbrace{\frac{1}{N_S}\pi(a'|S')-\pi(a'|S')\pi(U_{\succeq}(a)|T)}_{\text{if }a'\in U_{\succeq}(a)}.$$

In either case, we have

$$\begin{aligned}|\hat{\sigma}(a|S,T;a'|S',T') - \sigma(a|S,T;a'|S',T')| &\leq \frac{1}{N_T}\left|\hat{\pi}(U_{\succeq}(a)|T) - \pi(U_{\succeq}(a)|T)\right| + \frac{1}{N_{S'}}\left|\hat{\pi}(a'|S') - \pi(a'|S')\right| \\ &= \frac{1}{N_T}\left|\hat{\pi}(U_{\succeq}(a)|T) - \pi(U_{\succeq}(a)|T)\right| + \frac{1}{N_T}\left|\hat{\pi}(a|T) - \pi(a|T)\right|.\end{aligned}$$

Again by using the union bound and Bernstein's inequality, one has

$$\mathbb{P}\left[\frac{1}{\sigma(a|S,T)\sigma(a'|S',T')}\left|\hat{\sigma}(a|S,T;a'|S',T') - \sigma(a|S,T;a'|S',T')\right| \geq \xi_2\right] \leq 4\exp\left\{-\frac{1}{16}\mathfrak{c}_2^2\xi_2^2\right\}.$$

Case (iv). First note that $j(\ell) = j(\ell')$ and $j'(\ell) = j'(\ell')$ in this case. The covariance $\sigma(a|S,T;a'|S',T')$ can take two forms,

$$\begin{aligned}\sigma(a|S,T;a'|S',T') &= \underbrace{\frac{1}{N_S}\pi(a|S)(1-\pi(a|S)) + \frac{1}{N_T}\Big(\pi(U_{\succeq}(a)|T) \wedge \pi(U_{\succ}(a')|T) - \pi(U_{\succeq}(a)|T)\pi(U_{\succ}(a')|T)\Big)}_{\text{if } a = a'} \\ &\text{or} \quad \underbrace{-\frac{1}{N_S}\pi(a|S)\pi(a'|S) + \frac{1}{N_T}\Big(\pi(U_{\succeq}(a)|T) \wedge \pi(U_{\succ}(a')|T) - \pi(U_{\succeq}(a)|T)\pi(U_{\succ}(a')|T)\Big)}_{\text{if } a \neq a'}.\end{aligned}$$

Then we have

$$\begin{aligned}|\hat{\sigma}(a|S,T;a'|S',T') - \sigma(a|S,T;a'|S',T')| &\leq \frac{1}{N_S}\left|\hat{\pi}(a|S) - \pi(a|S)\right| + \frac{1}{N_S}\left|\hat{\pi}(a'|S) - \pi(a'|S)\right| \\ &\quad + \frac{1}{N_T}\left|\hat{\pi}(U_{\succeq}(a)|T) - \pi(U_{\succeq}(a)|T)\right| + \frac{1}{N_T}\left|\hat{\pi}(U_{\succ}(a')|T) - \pi(U_{\succ}(a')|T)\right|.\end{aligned}$$

Again by using the union bound and Bernstein's inequality, one has

$$\mathbb{P}\left[\frac{1}{\sigma(a|S,T)\sigma(a'|S',T')}\left|\hat{\sigma}(a|S,T;a'|S',T') - \sigma(a|S,T;a'|S',T')\right| \geq \xi_2\right] \leq 8\exp\left\{-\frac{1}{64}\mathfrak{c}_2^2\xi_2^2\right\}.$$

SA.3.5 Proof of Lemma A.8

This follows directly from Lemma A.7 and the result below.

Lemma SA.2 (Corollary 5.1 in Chernozhukov, Chetverikov, Kato, and Koike 2022)**.** Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathfrak{c}_1}$ be two mean-zero Gaussian random vectors with covariance matrices $\mathbf{\Sigma}^{\mathbf{x}}$ and $\mathbf{\Sigma}^{\mathbf{y}}$, respectively. Further assume that the diagonal elements in $\mathbf{\Sigma}_{\mathbf{x}}$ are all one. Define

$$\xi_3 = \max_{1 \leq \ell,\ell' \leq \mathfrak{c}_1} |\Sigma^{\mathbf{x}}_{\ell,\ell'} - \Sigma^{\mathbf{y}}_{\ell,\ell'}|,$$

where $\Sigma^{\cdot}_{\ell,\ell'}$ denotes the (ℓ,ℓ')th element in the matrix $\mathbf{\Sigma}^{\cdot}$. Then

$$\sup_{\substack{A \subseteq \mathbb{R}^{\mathfrak{c}_1} \\ A \text{ rectangular}}} \left|\mathbb{P}\left[\mathbf{x} \in A\right] - \mathbb{P}\left[\mathbf{y} \in A\right]\right| \leq c\xi_3^{\frac{1}{2}} \log \mathfrak{c}_1,$$

where c is an absolute constant.

SA.3.6 Proof of Lemma A.9

Take $\mathsf{T}^{\mathsf{G}}(\succ)$ as an example. It is easy to see that

$$\mathbb{P}\left[\check{\mathsf{T}}^{\mathsf{G}}(\succ) \le t\right] = \mathbb{P}\left[\max(\check{\mathbf{z}}) \le t\right],$$

where the set $\cdot \le t$ in the second probability above is a rectangular region. And hence the first part of this lemma follows from Lemma A.8.

Next, note that by conditioning on the following event,

$$\sup_t \left|\mathbb{P}\left[\check{\mathsf{T}}^{\mathsf{G}}(\succ) \le t\right] - \mathbb{P}\left[\mathsf{T}^{\mathsf{G}}(\succ) \le t \middle| \mathrm{Data}\right]\right| \le c\xi_3^{\frac{1}{3}} \log^{\frac{2}{3}} \mathfrak{c}_1,$$

one has

$$\begin{aligned}\mathbb{P}\left[\mathsf{T}^{\mathsf{G}}(\succ) > \check{\mathrm{cv}}\left(\alpha - c\xi_3^{\frac{1}{3}}\log^{\frac{2}{3}}\mathfrak{c}_1, \succ\right)\middle| \mathrm{Data}\right] &\le \mathbb{P}\left[\check{\mathsf{T}}^{\mathsf{G}}(\succ) > \check{\mathrm{cv}}\left(\alpha - c\xi_3^{\frac{1}{3}}\log^{\frac{2}{3}}\mathfrak{c}_1, \succ\right)\right] + c\xi_3^{\frac{1}{3}}\log^{\frac{2}{3}}\mathfrak{c}_1 \\ &\le \alpha - c\xi_3^{\frac{1}{3}}\log^{\frac{2}{3}}\mathfrak{c}_1 + c\xi_3^{\frac{1}{3}}\log^{\frac{2}{3}}\mathfrak{c}_1 = \alpha,\end{aligned}$$

which implies that

$$\mathrm{cv}(\alpha, \succ) \le \check{\mathrm{cv}}\left(\alpha - c\xi_3^{\frac{1}{3}}\log^{\frac{2}{3}}\mathfrak{c}_1, \succ\right).$$

Similarly,

$$\begin{aligned}\mathbb{P}\left[\check{\mathsf{T}}^{\mathsf{G}}(\succ) > \mathrm{cv}\left(\alpha, \succ\right)\middle| \mathrm{Data}\right] &\le \mathbb{P}\left[\mathsf{T}^{\mathsf{G}}(\succ) > \mathrm{cv}\left(\alpha, \succ\right)\middle| \mathrm{Data}\right] + c\xi_3^{\frac{1}{3}}\log^{\frac{2}{3}}\mathfrak{c}_1 \\ &\le \alpha + c\xi_3^{\frac{1}{3}}\log^{\frac{2}{3}}\mathfrak{c}_1,\end{aligned}$$

which implies

$$\mathrm{cv}(\alpha, \succ) \ge \check{\mathrm{cv}}\left(\alpha + c\xi_3^{\frac{1}{3}}\log^{\frac{2}{3}}\mathfrak{c}_1, \succ\right).$$

This concludes our proof of the second part of this lemma.

SA.4 Omitted Details in Appendix A.5

For simplicity, we continue using the notation $\mathbf{z}_i$, so that the average $\sum_{i=1}^n \mathbf{z}_i/n = \hat{\boldsymbol{\pi}}$ estimates the vector of choice probabilities. We denote by $\mathbf{z}_{i,\underline{\phi}(a|S)} = \mathbf{R}_{\underline{\phi}(a|S)}\mathbf{z}_i$, and the row in $\mathbf{z}_{i,\underline{\phi}(a|S)}$ for estimating $\hat{\pi}(a|R)$ is represented by $z_i(a|R)$.

SA.4.1 Proof of Lemma A.11

We again adopt the normal approximation result in Lemma SA.1, and consider

$$x_i(a|R) = \frac{\frac{n}{N_R}(\mathbb{1}(y_i = a, Y_i = R) - \pi(a|R)\mathbb{1}(Y_i = R))}{\sqrt{n}\sigma(a|R)}.$$

Condition (i) there is trivially satisfied with $C = C' = 1$. To verify condition (ii), we will need to take a closer look at the individual summands along each coordinate, $x_i(a|R)$. We observe that

$$\left|\frac{n}{N_R}(\mathbb{1}(y_i = a, Y_i = R) - \pi(a|R)\mathbb{1}(Y_i = R))\right| \leq \frac{n}{N_R}$$
$$\mathbb{E}\left[\left|\frac{n}{N_R}(\mathbb{1}(y_i = a, Y_i = R) - \pi(a|R)\mathbb{1}(Y_i = R))\right|^{2+\varepsilon}\right] \leq 2\left(\frac{n}{N_R}\right)^{2+\varepsilon}\pi(a|R)(1-\pi(a|R)).$$

In addition, the summands are nonzero for at most N_R observations. As a result,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\left|\frac{n}{N_R}(\mathbb{1}(y_i = a, Y_i = R) - \pi(a|R)\mathbb{1}(Y_i = R))\right|^{2+\varepsilon}\right] \leq 2\left(\frac{n}{N_R}\right)^{1+\varepsilon}\pi(a|R)(1-\pi(a|R)).$$

Now consider $\sqrt{n}\sigma(a|R)$, which takes the form

$$\sqrt{n}\sigma_(a|R) = \sqrt{\frac{n}{N_R}\pi(a|R)(1-\pi(a|R))}.$$

Combining previous results, we have

$$\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[|x_i(a|R)|^{2+\varepsilon}]\right)^{\frac{1}{\varepsilon}} \leq 2\left(\frac{n}{N_R}\right)\frac{1}{\sqrt{n}\sigma(a|R)}.$$

To apply Lemma SA.1, set $\varepsilon = 2$ and

$$\mathfrak{c}_2 = 2\sqrt{n}\left[\min_{R\supseteq S, R\in\mathcal{D}} N_R\right]^{-1}\left[\min_{R\supseteq S, R\in\mathcal{D}}\sigma(a|R)\right]^{-1}.$$

It is easy to see that the above choice satisfies the first part of condition (ii) in Lemma SA.1. For the second part in condition (ii), we note that

$$\begin{aligned}\left|\frac{x_i(a|R)}{\mathfrak{c}_2}\right| &= \frac{\frac{n}{N_R}(\mathbb{1}(y_i = a, Y_i = R) - \pi(a|R)\mathbb{1}(Y_i = R))}{\sqrt{n}\sigma(a|R)}\frac{\min_{R\supseteq S, R\in\mathcal{D}} N_R}{2\sqrt{n}}\left[\min_{R\supseteq S, R\in\mathcal{D}}\sigma(a|R)\right]\\ &\leq \left|\frac{n}{N_R}(\mathbb{1}(y_i = a, Y_i = R) - \pi(a|R)\mathbb{1}(Y_i = R))\right|\frac{\min_{R\supseteq S, R\in\mathcal{D}} N_R}{2n},\end{aligned}$$

which closes the proof.

SA.4.2 Proof of Lemma A.12

To begin with,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\frac{z_i(a|R) - \mathbb{E}[z_i(a|R)]}{\hat{\sigma}(a|R)} &= \frac{1}{n}\sum_{i=1}^{n}\frac{z_i(a|R) - \mathbb{E}[z_i(a|R)]}{\sigma(a|R)}\\ &\quad + \left(\frac{1}{n}\sum_{i=1}^{n}\frac{z_i(a|R) - \mathbb{E}[z_i(a|R)]}{\sigma(a|R)}\right)\left(\frac{\sigma(a|R)}{\hat{\sigma}(a|R)} - 1\right),\end{aligned}$$

which means

$$\max_{R\supseteq S, R\in\mathcal{D}} \left| \frac{1}{n}\sum_{i=1}^{n} \frac{z_i(a|R) - \mathbb{E}[z_i(a|R)]}{\hat{\sigma}(a|R)} - \frac{1}{n}\sum_{i=1}^{n} \frac{z_i(a|R) - \mathbb{E}[z_i(a|R)]}{\sigma(a|R)} \right|$$
$$\leq \max_{R\supseteq S, R\in\mathcal{D}} \left| \frac{1}{n}\sum_{i=1}^{n} \frac{z_i(a|R) - \mathbb{E}[z_i(a|R)]}{\sigma(a|R)} \right| \cdot \max_{R\supseteq S, R\in\mathcal{D}} \left| \frac{\sigma(a|R)}{\hat{\sigma}(a|R)} - 1 \right|.$$

We will control the two terms on the right-hand side separately.

Let ξ_1 be some generic constant which can depend on the sample size. Then by Lemma A.11,

$$\mathbb{P}\left[\max_{R\supseteq S, R\in\mathcal{D}} \left| \frac{1}{n}\sum_{i=1}^{n} \frac{z_i(a|R) - \mathbb{E}[z_i(a|R)]}{\sigma(a|R)} \right| \geq \xi_1 \right] \leq c\xi_1^{-1}\sqrt{\log \mathfrak{c}_1} + c\left(\frac{\log^5(n\mathfrak{c}_1)}{\mathfrak{c}_2^2}\right)^{\frac{1}{4}}.$$

Next consider the standard error estimator, $\hat{\sigma}(a|R)$.

$$\left|\hat{\sigma}(a|R)^2 - \sigma(a|R)^2\right| \leq \frac{1}{N_R}\left|\hat{\pi}(a|R) - \pi(a|R)\right|.$$

Using Bernstein's inequality, one has

$$\begin{aligned}
\mathbb{P}\left[\frac{1}{N_R\sigma(a|R)^2}\left|\hat{\pi}(a|R) - \pi(a|R)\right| \geq \xi_2\right] &\leq 2\exp\left\{-\frac{1}{2}\frac{N_R^4\sigma(a|R)^4\xi_2^2}{N_R\pi(a|R)(1-\pi(a|R)) + \frac{1}{3}N_R^2\sigma(a|R)^2\xi_2}\right\} \\
&= 2\exp\left\{-\frac{1}{2}\frac{N_R^2\sigma(a|R)^4\xi_2^2}{\frac{1}{N_R}\pi(a|R)(1-\pi(a|R)) + \frac{1}{3}\sigma(a|R)^2\xi_2}\right\} \\
&\leq 2\exp\left\{-\frac{1}{2}\frac{N_R^2\sigma(a|R)^4\xi_2^2}{\sigma(a|R)^2 + \frac{1}{3}\sigma(a|R)^2\xi_2}\right\} \\
&= 2\exp\left\{-\frac{1}{2}\frac{N_R^2\sigma(a|R)^2\xi_2^2}{1 + \frac{1}{3}\xi_2}\right\} \\
&\leq 2\exp\left\{-\frac{1}{4}N_R^2\sigma(a|R)^2\xi_2^2\right\}, \qquad \text{provided that } \xi_2 \to 0 \\
&\leq 2\exp\left\{-\frac{1}{4}\mathfrak{c}_2^2\xi_2^2\right\}.
\end{aligned}$$

Using the union bound, we deduce that

$$\mathbb{P}\left[\max_{R\supseteq S, R\in\mathcal{D}} \frac{\left|\hat{\sigma}(a|R)^2 - \sigma(a|R)^2\right|}{\sigma(a|R)^2} \geq \xi_2\right] \leq 4\exp\left\{-\frac{1}{16}\mathfrak{c}_2^2\xi_2^2 + \log \mathfrak{c}_1\right\}, \qquad \text{provided that } \xi_2 \to 0,$$

which also implies that

$$\mathbb{P}\left[\max_{R\supseteq S, R\in\mathcal{D}} \frac{\left|\hat{\sigma}(a|R) - \sigma(a|R)\right|}{\sigma(a|R)} \geq \xi_2\right] \leq 4\exp\left\{-\frac{1}{16}\mathfrak{c}_2^2\xi_2^2 + \log \mathfrak{c}_1\right\}, \qquad \text{provided that } \xi_2 \to 0.$$

This closes the proof.

References

Abaluck, J., and A. Adams (2021): "What Do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses," *Quarterly Journal of Economics*, 136(3), 1611–1663.

Aguiar, V. H. (2017): "Random Categorization and Bounded Rationality," *Economics Letters*, 159, 46–52.

Aguiar, V. H., M. J. Boccardi, N. Kashaev, and J. Kim (2023): "Random utility and limited consideration," *Quantitative Economics*, 14(1), 71–116.

Andrews, D. W., and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78(1), 119–157.

Barseghyan, L., M. Coughlin, F. Molinari, and J. C. Teitelbaum (2021): "Heterogeneous Choice Sets and Preferences," *Econometrica*, 89(5), 2015–2048.

Barseghyan, L., and F. Molinari (2023): "Risk Preference Types, Limited Consideration, and Welfare," *Journal of Business & Economic Statistics*, forthcoming.

Barseghyan, L., F. Molinari, and M. Thirkettle (2021): "Discrete Choice under Risk with Limited Consideration," *American Economic Review*, 111(6), 1972–2006.

Brady, R. L., and J. Rehbeck (2016): "Menu-Dependent Stochastic Feasibility," *Econometrica*, 84(3), 1203–1223.

Cattaneo, M. D., X. Ma, Y. Masatlioglu, and E. Suleymanov (2020): "A Random Attention Model," *Journal of Political Economy*, 128(7), 2796–2836.

Cherepanov, V., T. Feddersen, and A. Sandroni (2013): "Rationalization," *Theoretical Economics*, 8(3), 775–800.

Chernozhukov, V., D. Chetverikov, K. Kato, and Y. Koike (2022): "Improved Central Limit Theorem and Bootstrap Approximations in High Dimensions," *Annals of Statistics*, 50(5), 2562–2586.

Dardanoni, V., P. Manzini, M. Mariotti, H. Petri, and C. J. Tyson (2023): "Mixture Choice Data: Revealing Preferences and Cognition," *Journal of Political Economy*, 131(3), 687–715.

Dardanoni, V., P. Manzini, M. Mariotti, and C. J. Tyson (2020): "Inferring Cognitive Heterogeneity From Aggregate Choices," *Econometrica*, 88(3), 1269–1296.

Demirkan, Y., and M. Kimya (2020): "Hazard Rate, Stochastic Choice and Consideration Sets," *Journal of Mathematical Economics*, 87, 142–150.

Echenique, F., and K. Saito (2019): "General Luce Model," *Economic Theory*, 68, 811–826.

Echenique, F., K. Saito, and G. Tserenjigmid (2018): "The Perception-Adjusted Luce Model," *Mathematical Social Sciences*, 93, 67–76.

Filiz-Ozbay, E., and Y. Masatlioglu (2023): "Progressive Random Choice," *Journal of Political Economy*, 131(3), 716–750.

Fudenberg, D., R. Iijima, and T. Strzalecki (2015): “Stochastic Choice and Revealed Perturbed Utility,” *Econometrica*, 83(6), 2371–2409.

Gibbard, P. (2021): “Disentangling preferences and limited attention: Random-utility models with consideration sets,” *Journal of Mathematical Economics*, 94, 102468.

Gul, F., P. Natenzon, and W. Pesendorfer (2014): “Random Choice as Behavioral Optimization,” *Econometrica*, 82(5), 1873–1912.

Guney, B. (2014): “A theory of iterative choice in lists,” *Journal of Mathematical Economics*, 53, 26–32, Special Section: Economic Theory of Bubbles (I).

Honda, E. (2021): “A Model of Random Cravings,” *Working Paper*.

Horan, S. (2019): “Random Consideration and Choice: A Case Study of “Default” Options,” *Mathematical Social Sciences*, 102, 73–84.

Ishii, Y., M. Kovach, and L. Ülkü (2021): “A model of stochastic choice from lists,” *Journal of Mathematical Economics*, 96, 102509.

Kashaev, N., and V. H. Aguiar (2022): “A random attention and utility model,” *Journal of Economic Theory*, 204, 105487.

Lleras, J. S., Y. Masatlioglu, D. Nakajima, and E. Y. Ozbay (2017): “When More Is Less: Limited Consideration,” *Journal of Economic Theory*, 170, 70–85.

Manzini, P., and M. Mariotti (2014): “Stochastic Choice and Consideration Sets,” *Econometrica*, 82(3), 1153–1176.

Marchant, T., and A. Sen (2023): “Stochastic choice with bounded processing capacity,” *Journal of Mathematical Psychology*, 114, 102771.

Rubinstein, A., and Y. Salant (2006): “A model of choice from lists,” *Theoretical Economics*, 1(1), 3–17.

Yildiz, K. (2016): “List-rationalizable choice,” *Theoretical Economics*, 11(2), 587–599.