# Online Supplemental Appendix to “Attention Overload”

Matias D. Cattaneo[*] Paul Cheung[†] Xinwei Ma[‡] Yusufcan Masatlioglu[§]

October 19, 2021

**Abstract**

This Supplemental Appendix contains omitted technical details in the proofs of the main results.

## Contents



[*]Department of Operations Research and Financial Engineering, Princeton University.

[†]Department of Economics, University of Maryland.

[‡]Department of Economics, UC San Diego.

[§]Department of Economics, University of Maryland.

# I Omitted Details in Appendix A.2

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 the AC 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, $\mathcal{S} = \{\{a, b, c\}, \{a, b, c, d, e\}\}$. 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 $\mathcal{S}$ 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, $\mathcal{S} = \{\{a, b, c\}, \{a, b, c, d, e\}\}$ 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 $\mathbf{R}_\succ \mathbb{V}[\hat{\boldsymbol{\pi}}]\mathbf{R}_\succ^\top$.

## I.1 Proof of Lemma A.1

*Proof.* We adopt the following result.

**Lemma SA.1** (Theorem 5.1 in Chernozhukov, Chetverikov, Kato, and Koike 2021)**.** 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

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

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

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

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

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

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 $\mathfrak{c}_2$. ■

## I.2 Proof of Lemma A.2

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

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

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

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

Let $\xi_1$ be some generic constant which can depend on the sample size. Then by the triangle inequality,

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

Note that $\check{\mathbf{z}}$ is defined in Lemma A.1.

Term (I) has the bound

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

by Markov's inequality, and $c$ is an absolute constant. By Lemma A.1, term (II) is simply bounded by

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

which holds for any $\xi_1$. Again $c$ is an absolute constant in the above.

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

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

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

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

Using the union bound, we deduce that

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

which also implies that

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

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.

## I.3 Proof of Lemma A.3

See the proof of Lemma A.2.

## I.4 Proof of Lemma A.4

Recall that each restriction involves comparing choice probabilities across two choice problems. For two restrictions, $\ell$ and $\ell'$, 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,

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

We already have an error bound for the last two terms from Lemma A.3, 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

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

Next we apply Bernstein's inequality, which gives

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

Therefore, we have

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

**Case (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

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

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

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

In either case, we have

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

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

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

**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

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

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

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

## I.5 Proof of Lemma A.5

This follows directly from Lemma A.4 and the result below.

**Lemma SA.2 (Corollary 5.1 in Chernozhukov, Chetverikov, Kato, and Koike 2021).** 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

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

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

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

where $c$ is an absolute constant.

### I.6 Proof of Lemma A.6

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

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

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.5.

Next, note that by conditioning on the following event,

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

one has

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

which implies that

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

Similarly,

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

which implies

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

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

## II Validity of Two-step Critical Values

We follow Chernozhukov, Chetverikov, and Kato (2019) to establish the validity of the two-step moment selection procedure, using our novel results above. Recall from the proof of Theorem 5 that

$$\check{\text{cv}}(\alpha, \succ) = \inf_t \left\{t \ge 0 :\ \mathbb{P}\left[\check{\mathcal{T}}^{\mathtt{G}}(\succ) \le t\right] \ge 1 - \alpha\right\}, \quad \check{\mathcal{T}}^{\mathtt{G}}(\succ) = \max\left\{\check{\mathbf{z}},\ 0\right\},$$

which is the critical value generated from the infeasible normal vector. Define the following infeasible selection event

$$\check{\psi}_{\succ}(a|S,T) = \begin{cases} 0 & \text{if } D(a|S,T)/\sigma(a|S,T) \ge -\check{\text{cv}}(\mathfrak{c}_3 + c\xi_2^{\frac{1}{2}} \log \mathfrak{c}_2, \succ) \\ -\infty & \text{otherwise,} \end{cases}$$

where, as before, we set

$$\xi_2 = \frac{\sqrt{2c\log \mathfrak{c}_1 + \frac{c}{2}\log \mathfrak{c}_2}}{\mathfrak{c}_2}.$$

We will first establish two preliminary results.

**Lemma SA.3.** Assume the Assumptions of Theorem 5 hold. Then

$$\mathbb{P}\left[\max_{\substack{a\in T\subset S,\ \check{\psi}_{\succ}(a|S,T)=-\infty\\ T,S\in\mathcal{S}}} \hat{D}(a|S,T)\le 0\right] \ge 1-\mathfrak{c}_3 - c\left(\frac{\log^5(n\mathfrak{c}_1)}{\mathfrak{c}_2^2}\right)^{\frac{1}{4}},$$

and

$$\mathbb{P}\left[\min_{\substack{a\in T\subset S,\ \check{\psi}_{\succ}(a|S,T)=0\\ T,S\in\mathcal{S}}} \hat{\psi}_{\succ}(a|S,T)= 0\right] \ge 1-\mathfrak{c}_3 - c\left(\frac{\log^5(n\mathfrak{c}_1)}{\mathfrak{c}_2^2}\right)^{\frac{1}{4}}.$$

To demonstrate the validity of the two-step selection method, we first note that the second event in the above lemma corresponds to $\hat{\psi}_{\succ}(a|S,T)\ge\check{\psi}_{\succ}(a|S,T)$. Let $\mathrm{cv}_{\check{\psi}}(\alpha,\succ)$ be the critical value with $\hat{\psi}_{\succ}(a|S,T)$ replaced by $\check{\psi}_{\succ}(a|S,T)$, then

$$\mathbb{P}\left[\mathcal{T}(\succ)>\mathrm{cv}(\alpha,\succ)\right]\le \mathbb{P}\left[\mathcal{T}(\succ)>\mathrm{cv}_{\check{\psi}}(\alpha,\succ)\right]+\mathfrak{c}_3+c\left(\frac{\log^5(n\mathfrak{c}_1)}{\mathfrak{c}_2^2}\right)^{\frac{1}{4}}.$$

Next, consider the first event in Lemma SA.3. Under this event, we have $\check{\phi}(a|S,T)=-\infty$ implies $\hat{D}(a|S,T)\le 0$, meaning that the inequality constraint is not violated in the sample. As a result, we will be able to further bound the above as

$$\le \mathbb{P}\left[\mathcal{T}_{\check{\psi}}(\succ)>\mathrm{cv}_{\check{\psi}}(\alpha,\succ)\right]+2\mathfrak{c}_3+c\left(\frac{\log^5(n\mathfrak{c}_1)}{\mathfrak{c}_2^2}\right)^{\frac{1}{4}},$$

where the statistic $\mathcal{T}_{\check{\psi}}(\succ)$ is defined as

$$\mathcal{T}_{\check{\psi}}(\succ)=\max\left\{\max_{\substack{a\in T\subset S,\check{\psi}(a|S,T)=0\\ T,S\in\mathcal{S}}}\frac{\widehat{D}(a|S,T)}{\widehat{\sigma}(a|S,T)}\ ,\ 0\right\}.$$

The rest of the proof follows from that of Theorem 5, with the only change that we now focus on the subset of inequality constraints with $\check{\phi}(a|S,T)=0$.

## II.1 Proof of Lemma SA.3

To show the first claim,

$$\mathbb{P}\left[\max_{\substack{a\in T\subset S,\ \check{\psi}_{\succ}(a|S,T)=-\infty\\ T,S\in\mathcal{S}}} \hat{D}(a|S,T)\le 0\right]$$

$$
\begin{aligned}
&= \mathbb{P}\left[\max_{\substack{a\in T\subset S,\ \check{\psi}_{\succ}(a|S,T)=-\infty\\ T,S\in\mathcal{S}}} \left(\frac{\hat{D}(a|S,T)-D(a|S,T)}{\sigma(a|S,T)} + \frac{D(a|S,T)}{\sigma(a|S,T)}\right) \leq 0\right]\\
&\geq \mathbb{P}\left[\max_{\substack{a\in T\subset S,\ \check{\psi}_{\succ}(a|S,T)=-\infty\\ T,S\in\mathcal{S}}} \left(\frac{\hat{D}(a|S,T)-D(a|S,T)}{\sigma(a|S,T)}\right) \leq \check{\mathrm{cv}}(\mathfrak{c}_3 + c\xi_2^{\frac{1}{2}}\log\mathfrak{c}_2, \succ)\right]\\
&\geq \mathbb{P}\left[\max_{\substack{a\in T\subset S\\ T,S\in\mathcal{S}}} \left(\frac{\hat{D}(a|S,T)-D(a|S,T)}{\sigma(a|S,T)}\right) \leq \check{\mathrm{cv}}(\mathfrak{c}_3 + c\xi_2^{\frac{1}{2}}\log\mathfrak{c}_2, \succ)\right]\\
&= 1 - \mathfrak{c}_3 - c\xi_2^{\frac{1}{2}}\log\mathfrak{c}_2 - c\left(\frac{\log^5(n\mathfrak{c}_1)}{\mathfrak{c}_2^2}\right)^{\frac{1}{4}},
\end{aligned}
$$

where the last line follows from the normal approximation result of Lemma A.1.

To show the second claim, we consider the following sequence of probability bounds

$$
\begin{aligned}
&\mathbb{P}\left[\min_{\substack{a\in T\subset S,\ \check{\psi}_{\succ}(a|S,T)=0\\ T,S\in\mathcal{S}}} \hat{\psi}_{\succ}(a|S,T)=0\right] = \mathbb{P}\left[\min_{\substack{a\in T\subset S,\ \check{\psi}_{\succ}(a|S,T)=0\\ T,S\in\mathcal{S}}} \frac{\hat{D}(a|S,T)}{\hat{\sigma}(a|S,T)} \geq -2\mathrm{cv}_{\mathtt{LF}}(\mathfrak{c}_3, \succ)\right]\\
&\geq \mathbb{P}\left[\min_{\substack{a\in T\subset S,\ \check{\psi}_{\succ}(a|S,T)=0\\ T,S\in\mathcal{S}}} \frac{\hat{D}(a|S,T)}{\hat{\sigma}(a|S,T)} \geq -2\check{\mathrm{cv}}(\mathfrak{c}_3 + c\xi_2^{\frac{1}{2}}\log\mathfrak{c}_2, \succ)\right] - c\exp\left\{-\frac{1}{c}\mathfrak{c}_2^2\xi_2^2 + 2\log\mathfrak{c}_1\right\},
\end{aligned}
$$

where the last line follows from Lemma A.5. To proceed, we apply Lemma A.2, which further bounds the above by

$$
\begin{aligned}
&\geq \mathbb{P}\left[\min_{\substack{a\in T\subset S,\ \check{\psi}_{\succ}(a|S,T)=0\\ T,S\in\mathcal{S}}} \frac{\hat{D}(a|S,T)}{\sigma(a|S,T)} \geq -2\check{\mathrm{cv}}(\mathfrak{c}_3 + c\xi_2^{\frac{1}{2}}\log\mathfrak{c}_2, \succ) + \xi_1\xi_2\right]\\
&\qquad - c\xi_1^{-1}\sqrt{\log\mathfrak{c}_1} - c\left(\frac{\log^5(n\mathfrak{c}_1)}{\mathfrak{c}_2^2}\right)^{\frac{1}{4}} - c\exp\left\{-\frac{1}{c}\mathfrak{c}_2^2\xi_2^2 + \log\mathfrak{c}_1\right\}.
\end{aligned}
$$

Then,

$$
\begin{aligned}
&\mathbb{P}\left[\min_{\substack{a\in T\subset S,\ \check{\psi}_{\succ}(a|S,T)=0\\ T,S\in\mathcal{S}}} \frac{\hat{D}(a|S,T)}{\sigma(a|S,T)} \geq -2\check{\mathrm{cv}}(\mathfrak{c}_3 + c\xi_2^{\frac{1}{2}}\log\mathfrak{c}_2, \succ) + \xi_1\xi_2\right]\\
&= \mathbb{P}\left[\min_{\substack{a\in T\subset S,\ \check{\psi}_{\succ}(a|S,T)=0\\ T,S\in\mathcal{S}}} \left(\frac{\hat{D}(a|S,T)-D(a|S,T)}{\sigma(a|S,T)} - \frac{D(a|S,T)}{\sigma(a|S,T)}\right) \geq -2\check{\mathrm{cv}}(\mathfrak{c}_3 + c\xi_2^{\frac{1}{2}}\log\mathfrak{c}_2, \succ) + \xi_1\xi_2\right]\\
&\geq \mathbb{P}\left[\min_{\substack{a\in T\subset S,\ \check{\psi}_{\succ}(a|S,T)=0\\ T,S\in\mathcal{S}}} \frac{\hat{D}(a|S,T)-D(a|S,T)}{\sigma(a|S,T)} \geq -\check{\mathrm{cv}}(\mathfrak{c}_3 + c\xi_2^{\frac{1}{2}}\log\mathfrak{c}_2, \succ) + \xi_1\xi_2\right]\\
&\geq 1 - \mathfrak{c}_3 - c\xi_2^{\frac{1}{2}}\log\mathfrak{c}_2 - 4\xi_1\xi_2(\sqrt{2\log\mathfrak{c}_1}+1),
\end{aligned}
$$

where the last line follows from Lemmas A.1 and A.7. Finally, we set

$$\xi_1^{-2} = \xi_2 = \frac{\sqrt{2c\log\mathfrak{c}_1 + \frac{c}{2}\log\mathfrak{c}_2}}{\mathfrak{c}_2}.$$

# III Omitted Details in Appendix A.3

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 $\mathbf{z}_{i,\phi_{\mathtt{L}}(a|S)} = \mathbf{R}_{\phi_{\mathtt{L}}(a|S)}\mathbf{z}_i$, and the row in $\mathbf{z}_{i,\phi_{\mathtt{L}}(a|S)}$ for estimating $\hat{\pi}(a|R)$ is represented by $z_i(a|R)$.

## III.1 Proof of Lemma A.8

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

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

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

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

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

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

which closes the proof.

## III.2 Proof of Lemma A.9

To begin with,

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

which means

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

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

Let $\xi_1$ be some generic constant which can depend on the sample size. Then by Lemma A.8,

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

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

$$\mathbb{P}\left[\frac{1}{N_R\sigma(a|R)^2}\left|\hat{\pi}(a|R) - \pi(a|R)\right| \geq \xi_2\right] \leq 2\exp\left\{-\frac{1}{2}\frac{N_R^4\sigma(a|R)^4\xi_2^2}{N_R\pi(a|R)(1-\pi(a|R)) + \frac{1}{3}N_R^2\sigma(a|R)^2\xi_2}\right\}$$
$$= 2\exp\left\{-\frac{1}{2}\frac{N_R^2\sigma(a|R)^4\xi_2^2}{\frac{1}{N_R}\pi(a|R)(1-\pi(a|R)) + \frac{1}{3}\sigma(a|R)^2\xi_2}\right\}$$
$$\leq 2\exp\left\{-\frac{1}{2}\frac{N_R^2\sigma(a|R)^4\xi_2^2}{\sigma(a|R)^2 + \frac{1}{3}\sigma(a|R)^2\xi_2}\right\}$$
$$= 2\exp\left\{-\frac{1}{2}\frac{N_R^2\sigma(a|R)^2\xi_2^2}{1 + \frac{1}{3}\xi_2}\right\}$$

$$
\begin{aligned}
&\leq 2\exp\left\{-\frac{1}{4}N_R^2\sigma(a|R)^2\xi_2^2\right\}, \qquad \text{provided that } \xi_2 \to 0 \\
&\leq 2\exp\left\{-\frac{1}{4}\mathfrak{c}_2^2\xi_2^2\right\}.
\end{aligned}
$$

Using the union bound, we deduce that

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

which also implies that

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

This closes the proof.

## References


Chernozhukov, V., D. Chetverikov, and K. Kato (2019): “Inference on Causal and Structural Parameters using Many Moment Inequalities,” *Review of Economic Studies*, 86(5), 1867–1900.

Chernozhukov, V., D. Chetverikov, K. Kato, and Y. Koike (2021): “Improved Central Limit Theorem and Bootstrap Approximations in High Dimensions,” *Working paper*.