Supplemental Appendix to "Attention Overload"

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

September 9, 2026

Abstract

This Supplemental Appendix follows the organization of the main paper. It provides additional literature and model comparisons, proofs and extensions for the homogeneous and heterogeneous attention-overload models, econometric proofs and alternative inference procedures, and simulation and computational evidence.

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

Contents

SA-1 Additional Literature and Model Comparisons 1

SA-1.1 Random Attention Models with Homogeneous Preferences . . . 2

SA-1.2 Random Attention Models with Heterogeneous Preferences . . . 5

SA-1.3 Other Related Stochastic Choice Models . . . 6

SA-2 Homogeneous Attention Overload 8

SA-2.1 Proof of Theorem 1 . . . 8

SA-2.2 Proof of Corollary 2 . . . 10

SA-2.3 Proof of Corollary 3 . . . 10

SA-2.4 Proof of Proposition 3 . . . 11

SA-2.5 Proof of Theorem 2 . . . 11

SA-2.6 Proof of Theorem 3 . . . 11

SA-2.7 A Search-Capacity Foundation for Attention Overload . . . 12

SA-2.8 Decision Avoidance as Monotone Engagement . . . 12

SA-2.9 Observable Separation from RAM . . . 14

SA-3 Heterogeneous-Preference List-Based Attention Overload 14

SA-3.1 Proof of Theorem 4 . . . 14

SA-3.2 Proof of Theorem 5 . . . 16

SA-3.3 Standard Random-Utility Implications . . . 17

SA-3.4 Proof of Proposition 4 . . . 18

SA-3.5 Proof of Proposition 5 . . . 19

SA-3.6 Proof of Proposition 6 . . . 19

SA-3.7 Proof of Proposition 7 . . . 19

SA-3.8 Observed-Domain and Sensitivity Analysis . . . 20

SA-4 Econometric Methods 29

SA-4.1 Proof of Theorem 6 . . . 30

SA-4.2 Proof of Theorem 7 . . . 37

SA-4.3 Proof of Corollary 4 . . . 45

SA-4.4 Proof of Theorem 8 . . . 47

SA-4.5 Proof of Theorem 9 . . . 50

SA-4.6 Proof of Corollary 5 . . . 51

SA-4.7 Proof of Corollary 6 . . . 52

SA-4.8 Alternative Inference Procedures for Homogeneous AOM . . . 55

SA-4.9 Alternative Inference Procedures for Heterogeneous H-LAO 61
SA-4.10 Cluster-Robust Inference for Repeated Choices 66

SA-5 Implementation and Computational Scalability **67**
SA-5.1 Homogeneous AOM 67
SA-5.2 Heterogeneous H-LAO 69

SA-6 Simulation Evidence **71**
SA-6.1 Homogeneous AOM 71
SA-6.2 Heterogeneous H-LAO 75

References **83**

SA-1 Additional Literature and Model Comparisons

Evidence relevant to AOM plays three distinct roles. First, limited and stochastic attention has been documented across investment, school choice, job search, grocery consumption, computer purchases, and airport choice (Huberman and Regev, 2001; Laroche, Rosenblatt, and Sinclair, 1984; Rosen, Curran, and Greenlee, 1998; Sheridan, Richards, and Slocum, 1975; Demuynck and Seel, 2018; Goeree, 2008; Başar and Bhat, 2004). The consideration-set literature likewise treats attention as a scarce and variable input into choice (Hauser and Wernerfelt, 1990; Shocker, Ben-Akiva, Boccara, and Nedungadi, 1991).

Second, some evidence speaks directly to menu expansion and random-attention restrictions. Consumers may inspect fewer items as choice sets expand (Reutskaja and Hogarth, 2009; Reutskaja, Nagel, Camerer, and Rangel, 2011), and eye-tracking evidence shows that the probability of inspecting an individual item decreases as the choice set expands (Thomas, Molter, and Krajbich, 2021). Larger choice sets can also increase outside-option choice (Iyengar and Lepper, 2000). Choice-overload experiments document limited and costly search in large menus (Dean, Ravindran, and Stoye, 2025; de Lara and Dean, 2024), while field evidence links larger platform assortments to lower purchase frequency among incumbent consumers through costly attention (Natan, 2024). Experiments that observe consideration directly document widespread stochastic limited consideration and assess restrictions imposed by leading attention models (Ellis, Filiz-Ozbay, and Ozbay, 2025; Chadd, 2023; Khan, Krajbich, Raymond, Sundaresan, and Veiga, 2026). Applying RAM to a travel-mode experiment, Wang and Zhu (2025) find that it accepts the ranking implied by subjects' self-reports while rejecting rankings that place an objectively dominated option first. A health-preference menu experiment similarly finds that the revealed rankings under RAM and RUM are largely consistent (Wang, Sicsic, and Chauvin, 2026). Together, these studies establish that random attention has empirical content and motivate testing attention overload's item-level comparative static in each application.

Third, related evidence helps separate attention overload from other mechanisms. Laboratory evidence on rational inattention shows that subjects adjust information acquisition to incentives while rejecting commonly used linear-Shannon cost specifications (Dean and Neligh, 2023). Abaluck and Gruber (2023) find that reducing the number of offered health-insurance plans can improve choice quality but reject individual choice overload as the explanation. In an experiment with a default option, Chadd, Filiz-Ozbay, and Ozbay (2025) find endogenous selection into costly search and that subjects who reject the default tend to discover higher-ranked alternatives. Recent theory obtains choice overload from still other mechanisms, including menu-dependent discrimination in threshold–Luce choice and minimax-regret search (Caliari and Petri, 2025; Auster and Che, 2025). AOM is distinct: it imposes competition directly on each alternative's marginal attention frequency. These comparisons clarify both the empirical motivation for AOM and why its restriction should be tested in the application at hand.

The evidence above motivates AOM and H-LAO. We compare them with related models of stochastic choice. RUM assumes full attention but allows for preference heterogeneity. Early random limited-attention models typically maintained homogeneous preferences while allowing for attention heterogeneity; more recent work has begun to incorporate both sources of heterogeneity. Several related approaches have appeared since the first circulation of this paper in 2021. We organize the comparison by attention primitives, data requirements, and sources of identifying variation.

After this empirical overview, Section SA-1.1 focuses on random-attention models with homogeneous preferences and compares them with AOM. Section SA-1.2 considers models with heterogeneous preferences and compares them with H-LAO. Section SA-1.3 discusses other stochastic-choice models related to AOM or H-LAO.

SA-1.1 Random Attention Models with Homogeneous Preferences

The main paper compares the attention restrictions of AOM with other random-attention models using the examples of Ann and Ben. Here we compare their observable implications and show that the models can rationalize different choice rules.

We first compare two nonparametric models: the *Random Attention Model* (RAM, Cattaneo, Ma, Masatlioglu, and Suleymanov, 2020) and AOM. Given the flexibility of their attention rules, the central challenge is revealed preference. Table SA-1a has an AOM representation but not a RAM representation, while Table SA-1b has a RAM representation but not an 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 SA-1: Differences in explanatory power: RAM versus AOM

Recall that RAM reveals $a \succ b$ if there exist $a,b \in S$ such that $\pi(a|S \setminus \{b\}) < \pi(a|S)$ (Cattaneo, Ma, Masatlioglu, and Suleymanov, 2020). The choice rule in Table SA-1a therefore has no RAM representation because RAM reveals both $b \succ c$ and $c \succ b$. AOM instead reveals $a \succ c$ and $a \succ b$ and admits either $a \succ b \succ c$ or $a \succ c \succ b$.

There is no 3-alternative example that RAM can explain but AOM cannot: with three alternatives, AOM explains a superset of the choice data explained by RAM. Table SA-1b therefore uses four alternatives. In that example, AOM would require both $a \succ b$ and $b \succ a$, so it cannot rationalize the data. RAM instead reveals $a \succ d$ and $b \succ c$, which is acyclic and rationalizes the data.

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:

$$\pi(a|S) \leq \pi(U_{\succeq}(a)|T) \qquad \text{for } T \subset S.$$

In contrast, RAM imposes a regularity condition when removing "better" alternatives:

$$\pi(a|S) \leq \pi(a|S \setminus \{b\}) \qquad \text{whenever } b \succ a.$$

Hence, the two underlying regularity violations used for identification are conceptually different.

In the realm of nonparametric attention models, we can also compare with the convex hull of deterministic consideration-set mappings. For example, the random competition filter (RCF) is an important special case of AOM. Let $\Gamma_j(\cdot)$ be deterministic consideration-set mappings that satisfy *Competition Filter* (Lleras, Masatlioglu, Nakajima, and Ozbay, 2017). An RCF model is specified by

$$\mu_{\texttt{RCF}}(T|S) = \sum_{j=1}^{J} \alpha_j \mathbb{1}(\Gamma_j(S) = T), \qquad\qquad \phi_{\texttt{RCF}}(a|S) = \sum_{j: a \in \Gamma_j(S)} \alpha_j,$$

where $\alpha_j \geq 0$ for $j = 1, 2, \ldots, J$ and $\sum_{j=1}^{J} \alpha_j = 1$. Random competition filters satisfy attention overload because, for every $a \in T \subseteq S$, $a \in \Gamma_j(S)$ implies $a \in \Gamma_j(T)$.

RCF also nests stochastic versions of bounded rationalization and imprecise narrowing down. In the former, the decision maker randomizes over rationales; in the latter, she randomizes over procedures for forming criteria. Because each realized rationale or procedure induces a deterministic competition filter, a menu-independent mixture over them is an RCF.

Several parametric models can also be compared with AOM. In the *Independent Attention* model of Manzini and Mariotti (2014), each alternative a has a menu-invariant consideration probability $\gamma(a)$, so $\phi_{\texttt{MM}}(a|S) = \gamma(a)$. Equality across menus satisfies attention overload, making independent attention a special case of AOM.

The *Elimination by Aspects* attention model of Aguiar (2017) is also nested. Each category D has a fixed probability $m(D)$; when that category is available, the decision maker selects its best alternative and otherwise chooses the default. If $\mathcal{C}$ denotes the collection of categories, then $\phi_{\texttt{Aguiar}}(a|S) = \sum_{D\in\mathcal{C}:\, a\in D} m(D)$, which is menu invariant and therefore satisfies attention overload.

The *Bounded Processing Capacity Rule* (BPCR) of Marchant and Sen (2023) is another special case. The decision maker has capacity k: she considers the entire menu when $|S| \leq k$ and draws uniformly from the subsets of size k when $|S| > k$. The attention frequency is

$$\phi_{\texttt{MS}}(a|S) = \begin{cases} 1, & |S| \leq k, \\ \dfrac{\binom{|S|-1}{k-1}}{\binom{|S|}{k}}, & |S| > k. \end{cases}$$

The second expression equals $k/|S|$, so attention frequency weakly decreases as the menu expands and BPCR satisfies attention overload.

The *Logit Attention* model (e.g., Brady and Rehbeck, 2016) need not satisfy attention overload. To construct a separating example, let the Luce weight $\ell(\cdot)$ on menus satisfy

$$\begin{aligned} &\ell(abcd) = 10, \quad \ell(bcd) = 10, \quad \ell(bc) = 4, \quad \ell(c) = 2, \quad \ell(ab) = 1, \quad \ell(b) = 1, \\ &\ell(S) = 0 \qquad \text{for all other nonempty } S \subseteq \{a, b, c, d\}. \end{aligned}$$

With $a \succ b \succ c \succ d$, these weights yield the following choice probabilities, listed in menu order:

$$\begin{aligned} \pi(\cdot|abcd) &= (11/28, 15/28, 1/14, 0), \\ \pi(\cdot|abc) &= (1/8, 5/8, 1/4), \\ \pi(\cdot|ab) &= (1/2, 1/2). \end{aligned}$$

Under AOM, $\pi(b|abc) = 5/8 > 1/2 = \pi(b|ab)$ requires $a \succ b$, while $\pi(a|abcd) = 11/28 > 1/8 + 1/4 = \pi(a|abc) + \pi(c|abc)$ requires $b \succ a$. Hence AOM cannot rationalize these choices. Giving the other nonempty subsets sufficiently small positive weights makes the logit rule well defined on every menu and preserves both strict inequalities. Conversely, the choice rule in Table SA-1a is inconsistent with RAM and hence with its logit-attention subclass (Cattaneo, Ma, Masatlioglu, and Suleymanov, 2020), but it is consistent with AOM.

Demirkan and Kimya (2020) allow the independent-consideration parameter $\gamma(a, S)$ to depend on the menu. Because $\gamma(a, S)$ may exceed $\gamma(a, T)$ for $T \subset S$, their model can violate attention overload.

These comparisons locate AOM's contribution precisely. AOM is neither a relabeling of RAM nor another parametric consideration-set model: it places a nonparametric cross-menu restriction on item-level marginal attention. That restriction produces different revealed-preference implications from RAM and accommodates several familiar attention mechanisms while remaining empirically refutable.

SA-1.2 Random Attention Models with Heterogeneous Preferences

A growing literature combines random utility with random limited attention. Within the nonparametric identification paradigm, Kashaev and Aguiar (2022) generalize RAM by incorporating random utility in the set-monotone and stable RAUM. There are two key differences from H-LAO. First, H-LAO targets list-based attention overload, whereas set-monotone and stable RAUM builds on RAM. Second, benchmark H-LAO uses an observed list to discipline sequential attention while allowing an unrestricted marginal preference distribution; under positive reach and full menu data this additional structure point identifies attention and the full-attention stochastic choice rule, whereas their more general model yields partial identification. Two robustness envelopes below first relax benchmark independence between preferences and stopping and then remove Sequential Path Independence itself.

Two subsequent models are especially useful comparisons for the list structure. Masatlioglu and Vu (2024) order heterogeneous consideration sets by an attentiveness index and identify their composition and frequency. Their primitive is an expansion of consideration sets across attention types, while H-LAO imposes cross-menu overload on list reach. Wei (2024) instead identifies preferences and attention from stopping-time variation without requiring menu variation and permits dependence between the two. H-LAO uses menu and observed-list variation; its benchmark independence restriction is relaxed in the dependence-robust identified set.

Several papers instead accommodate heterogeneous preferences and random limited attention through parametric restrictions. For example, Aguiar, Boccardi, Kashaev, and Kim (2023) develops attention-index models, and Dardanoni, Manzini, Mariotti, and Tyson (2020) parameterizes cognitive capacity through consideration-set size. In contrast, H-LAO relies on a nonparametric attention-overload restriction.

A different strand uses enriched choice data to accommodate heterogeneous preferences and random limited attention. For example, Gibbard (2021) studies frame-dependent choice, and Dardanoni, Manzini, Mariotti, Petri, and Tyson (2023) uses mixture choice data. In contrast, AOM uses menu-indexed choice probabilities to identify compatible preferences and bound attention frequencies, while H-LAO combines aggregate choice probabilities with an observed presentation order to identify attention and the full-attention random-utility rule.

The discrete-choice literature has long emphasized preference heterogeneity, and some recent work also allows attention to vary. Barseghyan, Coughlin, Molinari, and Teitelbaum (2021) study partial identification of preferences and consideration-set formation by restricting consideration-set size or the frequency of singleton consideration. Abaluck and Adams (2021) exploit asymmetries in cross-partial derivatives and identify consideration and preferences using rich exogenous covariate variation, while Abaluck, Compiani, and Zhang (2026) use derivative restrictions to identify preferences in the presence of costly search. Barseghyan, Molinari, and Thirkettle (2021) and Barseghyan and Molinari (2023) provide identification results for risk preferences with more restricted covariate variation. In contrast, AOM and H-LAO operate in an abstract choice domain and obtain identification from cross-menu attention restrictions and list structure.

Repeated-choice data provide another source of identification. Aguiar and Kashaev (2025) recover the joint distribution of preferences and latent choice sets when the panel is sufficiently long or the support of choice sets is sparse. Chib and Shimizu (2026) use time-invariant subject-specific consideration sets and a mixture representation to obtain scalable Bayesian estimation with many alternatives. These approaches recover persistent latent choice sets from within-subject choice histories; our methods instead use cross-menu choice restrictions and do not require a panel for the theoretical identification results.

The interpretation of latent consideration also depends on the maintained utility model. Allen (2026) shows that, with bounded utility shifters, an additive random-utility model with exogenous consideration can be observationally equivalent to full and extended additive random utility. That result concerns identification from utility-index variation without the cross-menu restrictions imposed here. It underscores why our attention conclusions derive from attention overload, list structure, and menu variation rather than from an unrestricted random-utility representation alone.

H-LAO's contribution is the recovery of item-specific reach and the latent full-attention stochastic choice rule from aggregate choices under explicit list, domain, and support conditions, together with sharp sensitivity envelopes when preference–stopping independence or Sequential Path Independence is relaxed. Once the full-attention rule has been recovered, the full toolkit of standard random-utility methods becomes available.

SA-1.3 Other Related Stochastic Choice Models

AOM does not impose classical regularity on every alternative, but it contains several regular stochastic-choice models. Standard RUM is a leading example because classical regularity implies $\succ$-Regularity for a compatible ordering.

Once H-LAO recovers the full-attention rule, the remaining identification question is the standard one for finite random utility. Chambers and Turansick (2025) derive tight limits on the support restrictions capable of point identifying a random-utility model, while Caradonna and Turansick (2026) characterize identified sets for a broader class of stochastic-choice models. These results reinforce our distinction between point recovery of the full-attention choice rule and generally set-valued recovery of the complete ranking distribution.

For example, Gul, Natenzon, and Pesendorfer (2014) study attribute rules in which the decision maker draws an attribute and selects an alternative possessing it. Every attribute rule is a RUM and therefore has an AOM representation. Fudenberg, Iijima, and Strzalecki (2015) introduce additive perturbed utility, under which deliberate randomization can make deterministic choice costly. Their choice rules satisfy regularity and hence also admit AOM representations.

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.

H-LAO is also related to the literature on choice over lists. Rubinstein and Salant (2006) develops deterministic choice from lists and random choice induced by randomly appearing lists. Guney (2014) and Yildiz (2016) study successive-elimination procedures, including settings with an unobserved list and recursive random choice. Those papers capture elimination under a binary relation that need not be a preference, whereas H-LAO combines heterogeneous rational preferences with random limited attention. H-LAO is especially close to the *Random Depth Model* of Ishii, Kovach, and Ülkü (2021), where a random depth determines how far down a list the decision maker explores. Their model studies a fixed menu with exogenous list variation, while H-LAO focuses on attention overload across menus.

Recent work also clarifies the consequences of observing coarsened alternatives. Kono, Saito, and Sandroni (2026) study random utility when choices of some alternatives are unobservable and aggregated into an outside option. Liao, Saito, and Sandroni (2026) allow the underlying composition of observed categories to vary across decision makers. H-LAO instead treats the listed alternatives and outside option as observed structural categories and requires the relevant preference comparisons to remain stable across menus. The empirical application states this requirement explicitly because price and travel-time profiles vary within its travel-mode categories.

Honda (2021) studies a *Random Craving Model*, a special case of RUM in which preferences are restricted to temptation orderings. It can therefore be viewed as a full-attention special case of H-LAO with a restricted preference distribution.

SA-2 Homogeneous Attention Overload

This section follows the homogeneous-preference analysis in Section 2 of the main paper. It presents the deferred proofs in the paper's order: Theorem 1, Corollaries 2 and 3, Proposition 3, and Theorems 2 and 3. Corollary 1 follows immediately from Theorem 1, while Propositions 1 and 2 are proved in the main text. The section then develops three additional results concerning a search-capacity foundation, decision avoidance, and observable separation from RAM.

SA-2.1 Proof of Theorem 1

Necessity is established in the main paper, so it remains to prove sufficiency. Fix $S \in \mathcal{D}$ and enumerate it as $x_1 \succ x_2 \succ \cdots \succ x_m$. Because S is fixed, the two objects defined below depend on S; for notational convenience, we suppress this dependence. First, define

$$q_i := \max_{R\in\mathcal{D}\colon R\supseteq S} \pi(x_i|R).$$

It is immediate that $q_1 = \pi(x_1|S)$ by $\succ$-Regularity. Because S is included among its own observed supersets, and from $\succ$-Regularity, we know that the following inequality holds:

$$\begin{aligned}\pi(x_i|S) \le q_i &\le \sum_{j\le i} \pi(x_j|S)\\ \Rightarrow q_i - \pi(x_i|S) &\le \sum_{j<i} \pi(x_j|S)\end{aligned}$$

For $i \geq 2$, define

$$\lambda_i := \begin{cases} \dfrac{q_i - \pi(x_i|S)}{\sum_{j<i}\pi(x_j|S)}, & \text{if } \sum_{j<i}\pi(x_j|S) > 0,\\[2ex] 0, & \text{if } \sum_{j<i}\pi(x_j|S) = 0.\end{cases}$$

The preceding inequalities imply $\lambda_i \in [0,1]$; when the denominator is zero, they also imply $q_i = \pi(x_i|S)$.

Construct the attention rule as follows. For each consideration set $T \subseteq S$, distinguish the best element (according to $\succ$), say x_i, from the rest of the elements in the set, where each inferior alternative, x_j for $j > i$, appears independently with probability λ_j. In particular, suppose the best element of T is x_i. Then, let

$$\tilde{\mu}(T|S) := \pi(x_i|S) \prod_{j>i:\ x_j \in T} \lambda_j \prod_{j>i:\ x_j \notin T} (1-\lambda_j)$$

Therefore, the resulting attention rule rationalizes the observed choices:

$$\sum_{T\subseteq S:\ x_i \text{ is } \succ\text{-best}} \tilde{\mu}(T|S) = \pi(x_i|S) \sum_{T\subseteq S:\ x_i \text{ is } \succ\text{-best}} \left(\prod_{j>i:\ x_j \in T} \lambda_j \prod_{j>i:\ x_j \notin T} (1-\lambda_j) \right) = \pi(x_i|S)$$

Its attention frequency for x_i is

$$\phi_{\tilde{\mu}}(x_i|S) = \pi(x_i|S) + \lambda_i \sum_{j<i} \pi(x_j|S) = q_i.$$

The final equality follows from the definition of λ_i when the sum is positive; when it is zero, the preceding inequalities give $q_i = \pi(x_i|S)$. Finally, if $S, S' \in \mathcal{D}$, $S \subseteq S'$, and $x_i \in S$, then the collection of observed supersets of S' is contained in the collection of observed supersets of S. Hence,

$$\phi_{\tilde{\mu}}(x_i|S) = \max_{R\in\mathcal{D}:\, R\supseteq S} \pi(x_i|R) \geq \max_{R\in\mathcal{D}:\, R\supseteq S'} \pi(x_i|R) = \phi_{\tilde{\mu}}(x_i|S'),$$

so the constructed attention rule satisfies attention overload. Repeating the construction menu by menu proves sufficiency and simultaneously establishes the pessimistic evaluation in Theorem 3.

For completeness, the same construction works for the optimistic evaluation after setting

$$q_i = \min_{T\in\mathcal{D}:\, x_i \in T \subseteq S} \pi(U_{\succeq}(x_i)|T).$$

The inclusion of S and $\succ$-Regularity again imply

$$\pi(x_i|S) \leq q_i \leq \sum_{j\leq i} \pi(x_j|S),$$

and the collection of eligible subsets expands with the menu, so the resulting q_i satisfies attention overload. More generally, the construction rationalizes any attention-frequency selection between the two bounds that itself satisfies attention overload. ■

SA-2.2 Proof of Corollary 2

For any strict ordering $\succ$, set $z_{ba} = 1$ if and only if $b \succ a$. Completeness and asymmetry give $z_{ab} + z_{ba} = 1$. Transitivity gives

$$z_{ab} + z_{bc} - 1 \leq z_{ac}$$

for every three distinct alternatives. Moreover,

$$\pi(U_{\succeq}(a)|T) = \pi(a|T) + \sum_{b \in T \setminus \{a\}} \pi(b|T) z_{ba},$$

so the remaining mixed-integer inequalities are exactly $\succ$-Regularity. Theorem 1 therefore maps every compatible ordering into $\mathcal{Z}_\pi$.

Conversely, take any $z \in \mathcal{Z}_\pi$ and define $b \succ_z a$ when $z_{ba} = 1$. The binary restrictions make $\succ_z$ complete and asymmetric. The three-alternative inequalities make it transitive: if $a \succ_z b$ and $b \succ_z c$, then $z_{ab} = z_{bc} = 1$, so $z_{ac} = 1$ and $a \succ_z c$. Thus $\succ_z$ is a strict ordering. The final set of inequalities says precisely that π satisfies $\succ_z$-Regularity, and Theorem 1 implies $\succ_z \in \mathcal{P}_\pi$. The two maps are inverse, proving the bijection and the model-feasibility claim. Finally, b is revealed preferred to a exactly when every feasible incidence vector has $z_{ba} = 1$, which is equivalent to the stated minimization result. ■

SA-2.3 Proof of Corollary 3

Necessity follows from the argument in the main paper: if $\pi(a|\{a,b\}) > \eta$, the bound on singleton consideration implies positive probability on $\{a,b\}$, so every representation satisfying attentive-at-binaries must have $a \succ b$.

For sufficiency, fix a strict ordering $\succ$ that satisfies $\succ$-Regularity and extends P_B, and use the optimistic construction at the end of the preceding proof. Consider any observed binary menu $\{a,b\} \in \mathcal{D}$ and suppose without loss that $a \succ b$. For the lower-ranked item, every eligible subset in the optimistic minimum gives upper-contour probability one, so $q_b = 1$. If $\pi(a|\{a,b\}) > 0$, the construction therefore sets $\lambda_b = 1$ and $\tilde{\mu}(\{a\}|\{a,b\}) = 0$; if that probability is zero, the same equality holds directly. Moreover,

$$\tilde{\mu}(\{b\}|\{a,b\}) = \pi(b|\{a,b\}) \leq \eta,$$

because a strict reverse inequality would imply $b \ \mathrm{P}_B \ a$, contradicting that $\succ$ extends P_B. Thus both singleton-consideration probabilities satisfy the attentive-at-binaries restriction. The argument applies separately to every observed binary menu and does not require unobserved singletons or binaries. ■

SA-2.4 Proof of Proposition 3

Fix $\succ$. The defining constraints of $\mathcal{M}_\pi(\succ)$ are exactly nonnegativity and adding up, rationalization of the observed choice probabilities, and attention overload. Hence $\mu \in \mathcal{M}_\pi(\succ)$ if and only if $(\mu, \succ)$ is an AOM representation of π. Theorem 1 implies that such a μ exists if and only if $\succ \in \mathcal{P}_\pi$, proving the sharp joint-set formula and nonemptiness claim.

For fixed $\succ$, all constraints are linear and the variables are probabilities in a finite-dimensional product of simplices, so $\mathcal{M}_\pi(\succ)$ is a compact polytope. A linear functional therefore attains its minimum and maximum on each nonempty $\mathcal{M}_\pi(\succ)$. Since X is finite, there are finitely many $\succ \in \mathcal{P}_\pi$; taking the smallest and largest conditional optima gives the sharp extrema over the union. ■

SA-2.5 Proof of Theorem 2

Let $(\succ, \mu)$ be an AOM representing π, and fix $S \in \mathcal{D}$ and $a \in S$. For every $R \in \mathcal{D}$ with $R \supseteq S$, attention overload and the fact that an alternative must be considered before it can be chosen imply

$$\phi_\mu(a|S) \geq \phi_\mu(a|R) \geq \pi(a|R).$$

Taking the maximum over the observed supersets R gives the lower bound. Similarly, for every $T \in \mathcal{D}$ with $a \in T \subseteq S$,

$$\phi_\mu(a|S) \leq \phi_\mu(a|T) \leq \pi(U_\succeq(a)|T).$$

Taking the minimum over T gives the upper bound. ■

SA-2.6 Proof of Theorem 3

The constructive proof of Theorem 1 sets $\phi(a|S) = \max_{R\in\mathcal{D}:\, R\supseteq S} \pi(a|R)$ and therefore establishes the pessimistic evaluation directly. ■

The remaining subsections develop additional behavioral foundations and observable implications of homogeneous AOM.

SA-2.7 A Search-Capacity Foundation for Attention Overload

Attention overload follows from a broad class of deletion-consistent search processes. Let a search type draw a menu-independent priority ordering ρ of X and a capacity $K \in \{0, 1, \ldots, |X|\}$. From menu S, the type considers the first $\min\{K, |S|\}$ available alternatives according to ρ. Both ρ and K may be random and arbitrarily dependent.

Proposition SA-1 (Random priority and capacity imply attention overload)**.** Every mixture of menu-independent priority-and-capacity search types satisfies attention overload. If $K \geq 1$ almost surely, it also satisfies the no-empty-consideration normalization of the homogeneous AOM.

Proof. Fix a search type and observed menus $T, S \in \mathcal{D}$ with $a \in T \subseteq S$. If a is among the first K available alternatives in S, at most $K - 1$ elements of S precede it according to ρ. Removing the alternatives in $S \setminus T$ cannot increase the number of available alternatives preceding a. Hence a is also among the first K alternatives in T. The deterministic consideration indicator therefore satisfies

$$\mathbb{1}\{a \text{ considered in } S\} \leq \mathbb{1}\{a \text{ considered in } T\}.$$

Taking expectations over the joint distribution of (ρ, K) gives $\phi(a|S) \leq \phi(a|T)$. If $K \geq 1$, every nonempty menu generates a nonempty consideration set. ∎

The result includes heterogeneous inspection capacities and heterogeneous salience rankings. It also extends immediately to a menu-independent search sequence with nonnegative inspection costs and a fixed budget: deleting unavailable items weakly reduces the cost incurred before reaching any surviving item.

SA-2.8 Decision Avoidance as Monotone Engagement

To incorporate the outside option in the homogeneous-preference model, consider a two-stage mechanism. The attention-overload restriction here applies conditional on engagement; allowing empty consideration under Assumption 1 of the main paper instead restricts unconditional attention frequencies. For every $S \in \mathcal{D}$, define the observed engagement probability

$$q(S) = 1 - \pi(o^*|S),$$

and suppose $q(S) > 0$. Conditional inside-choice probabilities are

$$p(a|S) = \frac{\pi(a|S)}{q(S)}, \qquad a \in S.$$

A *decision-avoidance AOM* first engages with probability $q(S)$, where engagement is monotone in the sense that $q(T) \geq q(S)$ whenever $T, S \in \mathcal{D}$ and $T \subseteq S$, and chooses o^* otherwise. Conditional on engagement, the decision maker follows a no-empty-consideration AOM with preference $\succ$.

Proposition SA-2 (Decision-avoidance characterization)**.** The observed choice rule has a decision-avoidance AOM representation with preference $\succ$ if and only if

$$q(T) \geq q(S) \qquad \text{for every } T, S \in \mathcal{D} \text{ with } T \subseteq S,$$

and the conditional choice rule p satisfies $\succ$-Regularity. Therefore,

$$\mathcal{P}_\pi^{\mathrm{DA}} = \{\succ : p \text{ satisfies } \succ\text{-Regularity}\}$$

is the sharp identified set of preferences whenever engagement is monotone.

Proof. Necessity of monotone engagement follows from the first stage. Conditional on engagement, the model is exactly the homogeneous AOM, so Theorem 1 of the main paper implies $\succ$-Regularity of p.

Conversely, suppose engagement is monotone and p satisfies $\succ$-Regularity. Theorem 1 of the main paper gives a no-empty attention rule μ^c satisfying attention overload and representing p with $\succ$. Engage with probability $q(S)$ and, conditional on engagement, draw the consideration set from $\mu^c(\cdot|S)$. This reproduces

$$\pi(a|S) = q(S)p(a|S), \qquad \pi(o^*|S) = 1 - q(S).$$

Moreover, the unconditional inside-item attention frequency is

$$\phi(a|S) = q(S)\phi^c(a|S).$$

For $T, S \in \mathcal{D}$ with $a \in T \subseteq S$, both $q(S) \leq q(T)$ and $\phi^c(a|S) \leq \phi^c(a|T)$, so $\phi(a|S) \leq \phi(a|T)$. Thus the overall process also satisfies attention overload. Sharpness of the preference set follows from the conditional AOM characterization. ∎

This extension delivers an exact characterization when outside choice measures nonengagement and makes the choice-overload channel directly testable through $q(S)$. A genuine outside alternative that may be selected after inside items are considered represents a distinct mechanism, which can be separated from inattention with additional information or restrictions.

SA-2.9 Observable Separation from RAM

The examples in Section SA-1.1 can be summarized as a formal comparison of observable choice implications.

Proposition SA-3 (Observable non-nesting of AOM and RAM)**.** Neither the choice implications of AOM nor those of RAM contain the other across finite choice environments. The choice rule in Table SA-1a has an AOM representation but no RAM representation, while the choice rule in Table SA-1b has a RAM representation but no AOM representation. The two models can therefore produce different specification and revealed-preference conclusions from observed choice probabilities, not only from different latent attention rules.

Proof. For Table SA-1a, RAM reveals both $b \succ c$ and $c \succ b$, so no strict preference can rationalize the data under RAM. The AOM characterization instead permits $a \succ b \succ c$ and $a \succ c \succ b$. For Table SA-1b, AOM would require both $a \succ b$ and $b \succ a$, whereas the RAM revealed-preference restrictions $a \succ d$ and $b \succ c$ are acyclic. The constructions and calculations are given in Section SA-1.1. ∎

SA-3 Heterogeneous-Preference List-Based Attention Overload

This section proves the H-LAO results stated in Section 3 of the main paper and then develops the direct observed-domain and robustness formulations. It also records the finite random-utility implications of the recovered full-attention rule and a descriptive measure of agreement with full-attention choice. Throughout this section, if $S = \{a_{s_1}, \ldots, a_{s_{|S|}}\}$ is written in the inherited list order, let $S_{:j} := \{a_{s_1}, \ldots, a_{s_j}\}$ and $S_{j:} := \{a_{s_j}, \ldots, a_{s_{|S|}}\}$ denote its jth prefix and suffix, respectively.

SA-3.1 Proof of Theorem 4

We first prove necessity. Suppose that π has a H-LAO representation (ϕ, τ). The outside-option probability gives $\varphi(a_{s_1}|S) = 1 - \pi(o^*|S) = \phi(a_{s_1}|S)$. Sequential Path Independence and induction on j then yield $\varphi(a_{s_j}|S) = \phi(a_{s_j}|S)$ for every j.

Because the domain is full, every suffix $S_{k:}$ is observed. By the maintained assumption that $\pi(S_{k:}|S_{k:}) >$

0, we have $1-\pi(o^*|S_{k:})>0$ for every k. Sequential Path Independence therefore implies

$$\varphi(a_{s_{|S|}}|S)=\phi(a_{s_{|S|}}|S)=\prod_{k=1}^{|S|}[1-\pi(o^*|S_{k:})]>0.$$

Thus, the recursive definition of f is well-defined.

Next, let $T\subseteq S$ have the same first alternative a. Weak Attention Overload gives $\phi(a|T)\geq\phi(a|S)$. Since $\pi(o^*|R)=1-\phi(a|R)$ whenever a is the first alternative in R, it follows that $\pi(o^*|S)\geq\pi(o^*|T)$. Hence Weak Choice Overload holds. It remains to identify the full-attention choice probabilities.

Claim SA-1. For all $a\in T$, we have

$$f(a|T)=\tau(\{\succ\in\mathcal{P}: a \text{ is } \succ\text{-best in } T\}).$$

Proof. We proceed by strong induction on $|T|$. The result is immediate for a singleton. Suppose it holds for every proper subset of $S=\{a_{s_1},\ldots,a_{s_{|S|}}\}$. For $k=2,\ldots,|S|$, the recursive definition of f, the induction hypothesis, and $\varphi=\phi$ give

$$\begin{aligned}f(a_{s_k}|S)&=\frac{1}{\phi(a_{s_{|S|}}|S)}\left(\pi(a_{s_k}|S)-\sum_{j=k}^{|S|-1}\big(\phi(a_{s_j}|S)-\phi(a_{s_{j+1}}|S)\big)\tau(\{\succ: a_{s_k} \text{ is } \succ\text{-best in } S_{:j}\})\right)\\&=\tau(\{\succ: a_{s_k} \text{ is } \succ\text{-best in } S\}).\end{aligned}$$

The last equality follows by substituting the H-LAO choice equation for $\pi(a_{s_k}|S)$: the terms corresponding to the proper prefixes cancel, leaving $\phi(a_{s_{|S|}}|S)$ times the probability that a_{s_k} is $\succ$-best in S. The denominator is strictly positive by the argument above. Because both $f(\cdot|S)$ and the full-attention choice probabilities add up to one, the result for a_{s_1} follows from the result for $k\geq 2$. ■

The claim implies the BM inequalities, proving necessity.

For sufficiency, suppose that π satisfies Weak Choice Overload and non-negativity of the BM polynomials. Because the domain is full and $\pi(R|R)>0$ for every nonempty menu R, every suffix $S_{k:}$ satisfies $1-\pi(o^*|S_{k:})>0$. Hence the recursive construction of φ gives $\varphi(a_{s_{|S|}}|S)>0$ for every S, so f is well-defined.

Since φ is recursively constructed according to the Sequential Path Independence formula, it satisfies Sequential Path Independence. Moreover, if $T\subseteq S$ have the same first alternative a, Weak Choice Overload gives $\pi(o^*|S)\geq\pi(o^*|T)$, and therefore $\varphi(a|T)=1-\pi(o^*|T)\geq 1-\pi(o^*|S)=\varphi(a|S)$. Thus, φ satisfies Weak Attention Overload.

Set $\tilde{\phi} = \varphi$ and let $\tilde{\mu}$ be the induced list-based attention rule, defined by $\tilde{\mu}(\emptyset|S) = 1 - \varphi(a_{s_1}|S)$ and $\tilde{\mu}(S_{:j}|S) = \varphi(a_{s_j}|S) - \varphi(a_{s_{j+1}}|S)$, with the terminal convention $\varphi(a_{s_{|S|+1}}|S) = 0$. The recursive construction of φ makes it weakly decreasing along the list, so these probabilities are nonnegative and add up to one. Hence $\tilde{\mu}$ is a well-defined list-based attention rule induced by an attention frequency satisfying the required restrictions.

The function f adds up to one by construction, and the BM condition implies that it is random-utility rationalizable. Because X is finite, all nonempty menus are observed, and preferences are strict, the Block–Marschak characterization in Falmagne (1978) implies that there exists a probability measure $\tilde{\tau}$ over $\mathcal{P}$ such that

$$f(a|T) = \tilde{\tau}(\{\succ \in \mathcal{P} : a \text{ is } \succ\text{-best in } T\}).$$

It remains to show that $(\tilde{\mu}, \tilde{\tau})$ explains the data. The outside-option probability is correct by construction. For $k = 2, \ldots, |S|$, the inside choice probability generated by this representation is

$$\begin{aligned}
\tilde{\pi}(a_{s_k}|S) &= \varphi(a_{s_{|S|}}|S)\tilde{\tau}(\{\succ \in \mathcal{P} : a_{s_k} \text{ is } \succ\text{-best in } S\}) \\
&\quad + \sum_{j=k}^{|S|-1} \tilde{\tau}(\{\succ \in \mathcal{P} : a_{s_k} \text{ is } \succ\text{-best in } S_{:j}\})\big(\varphi(a_{s_j}|S) - \varphi(a_{s_{j+1}}|S)\big) \\
&= \varphi(a_{s_{|S|}}|S) f(a_{s_k}|S) + \sum_{j=k}^{|S|-1} f(a_{s_k}|S_{:j})\big(\varphi(a_{s_j}|S) - \varphi(a_{s_{j+1}}|S)\big) \\
&= \pi(a_{s_k}|S),
\end{aligned}$$

where the final equality follows from the definition of f. Finally, both the observed and constructed choice probabilities add up to one over $S \cup \{o^*\}$, so equality for the outside option and for $a_{s_2}, \ldots, a_{s_{|S|}}$ implies equality for a_{s_1}. This proves sufficiency. ■

SA-3.2 Proof of Theorem 5

The equality $\varphi = \phi$ follows directly from the recovery argument. The equality $f(a|T) = \tau(\{\succ \in \mathcal{P} : a \text{ is } \succ\text{-best in } T\})$ was established by the induction argument in the proof of Theorem 4. ■

SA-3.3 Standard Random-Utility Implications

Theorem 5 point identifies f. The remaining identification statements follow from standard finite random-utility theory (Falmagne, 1978; Turansick, 2022). In particular, the sharp set of distributions over rankings that generate the recovered rule is

$$\mathcal{T}(f) = \left\{ \tau \in \Delta(\mathcal{P}) : f(a|S) = \sum_{\succ \in \mathcal{P}} \mathbb{1}\{a \text{ is } \succ\text{-best in } S\}\tau(\succ) \text{ for every } a \in S \subseteq X \right\}.$$

Indeed, every H-LAO representation must generate the recovered f, and combining any $\tau \in \mathcal{T}(f)$ with the identified attention rule reproduces π. The set is a nonempty compact polytope. Consequently, sharp bounds on any preference event are obtained by minimizing and maximizing its τ-mass over $\mathcal{T}(f)$.

Full menu data also identify lower-contour and rank distributions. Let

$$L_{\succ}(a) = \{x \in X : a \succeq x\}$$

be the weak lower contour set of a, and for $a \in T$ define the Block–Marschak polynomial

$$K_f(a, T) = \sum_{S\colon T \subseteq S \subseteq X} (-1)^{|S \setminus T|} f(a|S).$$

Moebius inversion gives

$$K_f(a, T) = \tau(\{\succ\colon L_{\succ}(a) = T\}) \qquad \text{for every } a \in T \subseteq X.$$

Thus, if $|X| = p$ and rank one is best,

$$\mathbb{P}_\tau(\text{rank}_{\succ}(a) = r) = \sum_{\substack{T\colon a \in T \subseteq X \\ |T| = p - r + 1}} K_f(a, T), \qquad r = 1, \ldots, p.$$

Pairwise shares and top-rank shares satisfy

$$\mathbb{P}_\tau(a \succ b) = f(a|\{a, b\}), \qquad \mathbb{P}_\tau(a \text{ is top ranked}) = f(a|X).$$

The complete joint distribution over rankings may nevertheless remain set identified.

For completeness, for every S containing a,

$$f(a|S) = \mathbb{P}_\tau(a \text{ is best in } S) = \mathbb{P}_\tau(S \subseteq L_\succ(a)).$$

Applying Moebius inversion to these containment probabilities yields

$$\mathbb{P}_\tau(L_\succ(a) = T) = \sum_{S:\, T \subseteq S \subseteq X} (-1)^{|S \setminus T|} \mathbb{P}_\tau(S \subseteq L_\succ(a)) = K_f(a, T).$$

The rank formula follows by grouping lower contour sets according to cardinality. Pairwise and top-rank identities follow directly from the random-utility interpretation of f.

For completeness, the probability that observed choice agrees with full-attention choice provides a common descriptive summary for the two models. Under homogeneous AOM and a compatible preference $\succ$, the best item $a^\succ(S)$ is chosen if and only if it is considered, so the agreement probability is $\pi(a^\succ(S)|S) = \phi(a^\succ(S)|S)$. It is therefore point identified when the preference is identified and otherwise takes its sharp values over $\succ \in \mathcal{P}_\pi$. Under benchmark H-LAO, independence of preferences and stopping implies that the agreement probability equals

$$\sum_{\succ \in \mathcal{P}} \tau(\succ)\phi(a^\succ(S)|S) = \sum_{a \in S} \phi(a|S) f(a|S).$$

This probability is point identified wherever ϕ and f are recovered and can otherwise be bounded using the corresponding sharp preference set. It provides a descriptive measure of agreement with full-attention choice and is kept distinct from cardinal welfare; the characterization and econometric results rely directly on the model's choice restrictions.

SA-3.4 Proof of Proposition 4

Let $S = \{a_{s_1}, \ldots, a_{s_{|S|}}\} \in \mathcal{D}$. Since $\mathcal{D}$ is suffix-closed, every suffix $S_{k:}$ is observed. The outside-option probability identifies first-position reach in each suffix as $\phi(a_{s_k}|S_{k:}) = 1 - \pi(o^*|S_{k:})$. Sequential Path Independence therefore gives $\phi(a_{s_j}|S) = \prod_{k=1}^{j}[1 - \pi(o^*|S_{k:})] = \varphi(a_{s_j}|S)$ for every j. Since S was arbitrary, $\phi(a|S)$ is point identified for every observed $S \in \mathcal{D}$ and $a \in S$.

Suppose, in addition, that $\mathcal{D}$ is prefix-closed. By the maintained assumption that $\pi(R|R) > 0$ for every observed menu R, each factor $1 - \pi(o^*|S_{k:})$ is strictly positive, so $\varphi(a_{s_{|S|}}|S) > 0$ for every $S \in \mathcal{D}$. We can therefore recover f recursively. The result is immediate for a singleton. Suppose f has been identified for every observed menu of size strictly smaller than $|S|$. Prefix closure implies that every proper prefix $S_{:j}$ is observed, so the recursive definition of f identifies $f(a_{s_k}|S)$ for $k = 2, \ldots, |S|$. The remaining value $f(a_{s_1}|S)$ is identified by adding up. By the same induction argument used in the proof of Theorem 4, under H-LAO, $f(x|S) = \tau(\{\succ \in \mathcal{P} : x \text{ is } \succ\text{-best in } S\})$. Thus these preference probabilities are point identified for every $x \in S$. ■

SA-3.5 Proof of Proposition 5

Suppose first that a precedes b on the observed list. Sequential Path Independence gives the probability of reaching b in the binary menu as $r_b(a, b) = [1 - \pi(o^*|\{a, b\})][1 - \pi(o^*|\{b\})]$. By the maintained positive-reach assumption, $r_b(a, b) > 0$. Since b can be chosen from $\{a, b\}$ only when attention reaches the full binary menu, independence between preferences and stopping implies $\pi(b|\{a, b\}) = r_b(a, b)\tau(\{\succ : b \succ a\})$. Hence $\tau(\{\succ : b \succ a\}) = \pi(b|\{a, b\})/r_b(a, b)$ and is point identified.

If instead b precedes a, the analogous argument identifies $\tau(\{\succ : a \succ b\})$ from $\{a, b\}$ and $\{a\}$. Since preferences are strict, $\tau(\{\succ : b \succ a\}) = 1 - \tau(\{\succ : a \succ b\})$. Thus $\tau(\{\succ : b \succ a\})$ is point identified in either case. ■

SA-3.6 Proof of Proposition 6

Take any eligible $R \in \mathcal{D}$ with $R \supseteq S$ for which $\varphi(a|R)$ is defined and any eligible $T \in \mathcal{D}$ with $a \in T \subseteq S$ for which $\varphi(a|T)$ is defined. Whenever $\varphi(a|R)$ and $\varphi(a|T)$ are defined from the observed suffixes, Sequential Path Independence gives $\varphi(a|R) = \phi(a|R)$ and $\varphi(a|T) = \phi(a|T)$. Attention overload therefore implies $\varphi(a|R) \leq \phi(a|S) \leq \varphi(a|T)$. Taking the maximum over the eligible supersets R and the minimum over the eligible subsets T proves the result. If either collection is empty, the corresponding bound is omitted. ■

SA-3.7 Proof of Proposition 7

Let $\theta(a, S) := \tau(\{\succ \in \mathcal{P} : a \text{ is } \succ\text{-best in } S\})$. Whenever $f(a|T)$ is recovered from the data, H-LAO implies $f(a|T) = \theta(a, T)$. Because the full-attention choice rule generated by τ satisfies random-utility regularity, for $a \in T \subseteq S \subseteq R$ we have $f(a|T) \geq \theta(a, S) \geq f(a|R)$ whenever the displayed f values are well-defined. Taking the minimum over the well-defined values $f(a|T)$ and the maximum over the well-defined values $f(a|R)$ gives the regularity bounds.

It remains to establish the additional list-based lower bound. Take any $R \supseteq S$ for which

$$\frac{\pi(a|R) - \Psi(a,S,R)}{\varphi(a_{s_{|S|}}|R)}$$

is well-defined. By the definition of $\Psi(a,S,R)$, this requires that S be a prefix of R. Hence the prefixes of R containing a can be partitioned into those strictly contained in S, the prefix S itself, and those strictly containing S. The H-LAO choice equation therefore gives

$$\begin{aligned}\pi(a|R) = &\underbrace{\sum_{\substack{1\le j\le |R|:\\ a\in R_{:j}\subsetneq S}} f(a|R_{:j})\mu(R_{:j}|R) + \theta(a,S)\mu(S|R)}_{\Psi(a,S,R)}\\ &+ \sum_{\substack{1\le j\le |R|:\\ R_{:j}\supsetneq S}} \theta(a,R_{:j})\mu(R_{:j}|R).\end{aligned}$$

For every prefix $R_{:j} \supsetneq S$, random-utility regularity implies $\theta(a,R_{:j}) \le \theta(a,S)$. It follows that

$$\pi(a|R) \le \Psi(a,S,R) + \theta(a,S) \sum_{\substack{1\le j\le |R|:\\ R_{:j}\supseteq S}} \mu(R_{:j}|R).$$

Since S is a prefix of R, the sum on the right is exactly the probability of reaching the last alternative of S in R, which equals $\phi(a_{s_{|S|}}|R) = \varphi(a_{s_{|S|}}|R)$. Hence

$$\theta(a,S) \ge \frac{\pi(a|R) - \Psi(a,S,R)}{\varphi(a_{s_{|S|}}|R)}.$$

This holds for every $R \supseteq S$ for which the displayed expression is well-defined. Taking the maximum over such R and combining it with the regularity bounds yields the stated result. ■

SA-3.8 Observed-Domain and Sensitivity Analysis

The full-domain, positive-engagement, independent H-LAO model remains our benchmark. This subsection builds nested sensitivity envelopes that isolate the identifying content of observed-menu coverage, positive reach, preference–stopping independence, and Sequential Path Independence. Removing these restrictions sequentially makes their contribution to identification transparent while preserving sharpness for each stated specification.

SA-3.8.1 Incomplete domains and zero reach

We first allow $\pi(o^*|S) = 1$ on observed menus, thereby covering zero reach beyond the positive-engagement setting of the recursive identification results. The next result gives a sharp representation on the observed domain $\mathcal{D}$. Full-domain extendibility is a stronger requirement that can be imposed when substantively appropriate.

Recall the observed list $a_1, \ldots, a_{|X|}$. Write every observed menu as $S = \{a_{s_1}, \ldots, a_{s_{|S|}}\}$ in the inherited list order. We call $\mathcal{D}$ *suffix closed* if $S_{j:} \in \mathcal{D}$ for every $S \in \mathcal{D}$ and $j = 1, \ldots, |S|$.

On a suffix-closed domain, define the recovered reach probabilities directly from outside choices:

$$\varphi(a_{s_j}|S) = \prod_{k=1}^{j} \left[1 - \pi(o^*|S_{k:})\right], \qquad j = 1, \ldots, |S|.$$

This product is well defined even when some reach probability is zero. Adopt the terminal convention $\varphi(a_{s_{|S|+1}}|S) = 0$ and let

$$\begin{aligned} m_0(S) &= 1 - \varphi(a_{s_1}|S), \\ m_j(S) &= \varphi(a_{s_j}|S) - \varphi(a_{s_{j+1}}|S), \qquad j = 1, \ldots, |S|. \end{aligned}$$

Thus, $m_j(S)$ is the recovered probability of considering exactly the prefix $S_{:j}$, while $m_0(S) = \pi(o^*|S)$ is the recovered probability of empty consideration.

Because each recovered reach is a product of factors in $[0, 1]$, the sequence is weakly decreasing along every list. Hence $m_j(S) \geq 0$ for every $S \in \mathcal{D}$ and $j = 0, \ldots, |S|$, and these masses add up to one automatically. We say that recovered attention is *valid on* $\mathcal{D}$ if

$$\varphi(a|S) \geq \varphi(a|R) \qquad \text{whenever } a \in S \subseteq R \text{ and } S, R \in \mathcal{D}.$$

This is attention overload on the observed domain. Sequential Path Independence and the validity of the within-menu prefix distribution hold by construction.

For $\tau \in \Delta(\mathcal{P})$, define the choice probabilities generated by the recovered attention rule as

$$G_\tau(a|S) = \sum_{\succ \in \mathcal{P}} \sum_{j=1}^{|S|} m_j(S) \mathbb{1}\{a \in S_{:j},\ a \text{ is } \succ\text{-best in } S_{:j}\} \tau(\succ).$$

The incomplete-domain preference polytope is

$$\mathcal{T}_{\mathcal{D}}(\pi) = \{\tau \in \Delta(\mathcal{P}) : G_\tau(a|S) = \pi(a|S) \text{ for every } S \in \mathcal{D} \text{ and } a \in S\}.$$

Proposition SA-4 (Observed-Domain H-LAO)**.** Let $\mathcal{D}$ be suffix closed. The choice rule π has an H-LAO representation on $\mathcal{D}$ if and only if recovered attention is valid on $\mathcal{D}$ and $\mathcal{T}_{\mathcal{D}}(\pi)$ is nonempty. Moreover, $\mathcal{T}_{\mathcal{D}}(\pi)$ is the sharp identified set for the preference distribution τ. It is a compact polytope, and the sharp bounds on every preference event $A \subseteq \mathcal{P}$ are

$$\min_{\tau \in \mathcal{T}_{\mathcal{D}}(\pi)} \sum_{\succ \in A} \tau(\succ) \qquad \text{and} \qquad \max_{\tau \in \mathcal{T}_{\mathcal{D}}(\pi)} \sum_{\succ \in A} \tau(\succ).$$

The result does not require positive reach.

Proof. Suppose first that (ϕ, τ) is an H-LAO representation on $\mathcal{D}$. For the first item of each observed menu,

$$\phi(a_{s_1}|S) = 1 - \pi(o^*|S).$$

Sequential Path Independence and induction on j give the product formula for $\varphi(a_{s_j}|S)$, including when one of its factors is zero. Hence $\varphi = \phi$ on the observed domain. The list-based attention rule therefore assigns probability $m_j(S)$ to $S_{:j}$ and probability $m_0(S)$ to the empty set. These probabilities are nonnegative, add to one, and satisfy attention overload. Independence between preferences and stopping gives $G_\tau(a|S) = \pi(a|S)$, so $\tau \in \mathcal{T}_{\mathcal{D}}(\pi)$.

Conversely, suppose recovered attention is valid and take $\tau \in \mathcal{T}_{\mathcal{D}}(\pi)$. Assign probability $m_0(S)$ to empty consideration and $m_j(S)$ to prefix $S_{:j}$. Their automatic nonnegativity and adding-up make this a list-based attention rule, while validity supplies attention overload. The product definition implies

$$\varphi(a_{s_j}|S) = \varphi(a_{s_{j-1}}|S)\varphi(a_{s_j}|S_{j:}),$$

so Sequential Path Independence holds. Draw $\succ$ independently according to τ. The defining equations of $\mathcal{T}_{\mathcal{D}}(\pi)$ reproduce every inside choice probability, and $m_0(S) = \pi(o^*|S)$ reproduces outside choice. Thus every τ in the set is attainable, proving sharpness.

The set is the intersection of a finite-dimensional simplex with finitely many linear equalities. It is therefore a compact polytope. Preference-event probabilities are linear in τ, so the displayed extrema are attained and are the sharp bounds. ∎

The observed-domain result complements the closed-form limited-data bounds in the main paper. Those bounds can be computed without enumerating rankings, while the polytope above exhausts the information about τ on any suffix-closed observed domain. Under full menu data and positive reach, it agrees with the standard random-utility polytope $\mathcal{T}(f)$. When terminal reach is zero, the recursive full-attention rule f may be undefined, but the direct choice equations above remain valid and simply leave the unreached preference comparisons unrestricted.

For $p = |X|$, the direct formulation has $p!$ preference-probability variables, one adding-up constraint, and at most $\sum_{S\in\mathcal{D}} |S|$ choice equations. It is practical for small and moderate universes, while event-specific ranking generation extends its reach. The closed-form pairwise result below avoids ranking enumeration entirely and requires only singleton and binary menus. The dependence-robust formulation adds menu-specific joint variables in exchange for its greater flexibility; Section SA-5 reports target-specific implementation strategies and computational benchmarks.

SA-3.8.2 Illustration and implementation

The replication files provide a four-alternative, suffix-closed example with 10 of the 15 nonempty menus and item-specific reach factors $(.86, .73, 0, .64)$ in list order. The zero factor makes the recursive full-attention rule unavailable for some comparisons, while the direct polytope $\mathcal{T}_{\mathcal{D}}(\pi)$ remains well defined. For the event $d \succ a$, the sharp benchmark-independence bounds are $[.545, .780]$ and contain the generating value .674; allowing arbitrary preference–stopping dependence widens them to $[.219, 1]$. The replication script constructs the choice rule and verifies feasibility, containment, and both sets of endpoints. Column generation certifies the two benchmark endpoints using only 5 of the 24 ranking columns in each restricted master problem.

Implementation should follow the target. Pairwise shares use the ranking-free formula in the main paper, and full-data rank probabilities use standard Block–Marschak sums. Ranking optimization is needed only for genuinely joint preference events or sharp incomplete-domain bounds. Section SA-5 gives the implementation hierarchy, solver certificates, and computational benchmarks.

We next relax the benchmark restrictions sequentially. The first envelope retains recovered stopping probabilities but relaxes preference–stopping independence; the second also removes Sequential Path Independence and treats prefix masses as latent.

SA-3.8.3 Preference–Stopping Dependence

The benchmark H-LAO model makes preferences independent of the stopping position. We retain the recovered marginal stopping probabilities but allow arbitrary dependence within each observed menu. This extension is tailored to repeated cross sections: the marginal preference distribution is stable across menus, but the coupling between preferences and stopping may be menu specific. A panel model requiring one persistent joint preference–stopping type across all menus would impose additional restrictions.

For every $S \in \mathcal{D}$, $j = 0, \dots, |S|$, and $\succ \in \mathcal{P}$, let $q_S(\succ, j)$ denote the joint probability of preference $\succ$ and stopping at j, where $j = 0$ denotes empty consideration. Let $\mathcal{Q}_{\mathcal{D}}(\pi)$ be the collection of (τ, q) satisfying

$$
\begin{aligned}
& q_S(\succ, j) \geq 0, \\
& \sum_{\succ \in \mathcal{P}} q_S(\succ, j) = m_j(S), \\
& \sum_{j=0}^{|S|} q_S(\succ, j) = \tau(\succ), \\
& \pi(a|S) = \sum_{j=1}^{|S|} \sum_{\succ \in \mathcal{P}} \mathbb{1}\{a \in S_{:j},\ a \text{ is } \succ\text{-best in } S_{:j}\} q_S(\succ, j)
\end{aligned}
$$

for every relevant S, j, $\succ$, and $a \in S$. Define

$$\mathcal{T}_{\mathcal{D}}^{\text{dep}}(\pi) = \{\tau : (\tau, q) \in \mathcal{Q}_{\mathcal{D}}(\pi) \text{ for some } q\}.$$

Proposition SA-5 (Dependence-Robust H-LAO)**.** Let $\mathcal{D}$ be suffix closed and suppose recovered attention is valid. Then $\mathcal{T}_{\mathcal{D}}^{\text{dep}}(\pi)$ is the sharp identified set for the stable marginal preference distribution when preferences and stopping may be arbitrarily dependent within each observed menu. It is a compact polytope and contains the independent H-LAO identified set:

$$\mathcal{T}_{\mathcal{D}}(\pi) \subseteq \mathcal{T}_{\mathcal{D}}^{\text{dep}}(\pi).$$

Sharp bounds on preference events are obtained by linear programming over $\mathcal{Q}_{\mathcal{D}}(\pi)$. Positive reach is not required.

Proof. Under a dependent representation, $q_S(\succ, j)$ is the joint distribution of the two latent types in menu S. Its stopping marginal must equal the identified prefix mass $m_j(S)$, its preference marginal must equal the stable distribution τ, and utility maximization on each prefix gives the displayed choice equations. Thus every admissible representation belongs to $\mathcal{Q}_{\mathcal{D}}(\pi)$.

Conversely, take any feasible (τ, q). In each observed menu, draw $(\succ, j)$ according to q_S, consider the empty set when $j = 0$ and prefix $S_{:j}$ otherwise, and select the $\succ$-best item in that prefix. The stopping marginal is the recovered list-based attention rule, so it satisfies Sequential Path Independence and attention overload. The preference marginal is τ in every menu, and the final constraints reproduce the observed choices. Hence every projected τ is attainable.

All defining constraints are linear and the variables are probabilities, so $\mathcal{Q}_{\mathcal{D}}(\pi)$ and its projection are compact polytopes. Under independence, $q_S(\succ, j) = \tau(\succ)m_j(S)$ satisfies the marginal constraints and reduces the choice equation to the definition of $\mathcal{T}_{\mathcal{D}}(\pi)$, proving the inclusion. ■

The dependence-robust result gives a direct sensitivity analysis. Comparing event bounds over $\mathcal{T}_{\mathcal{D}}(\pi)$ and $\mathcal{T}_{\mathcal{D}}^{\mathrm{dep}}(\pi)$ isolates the identifying contribution of preference–stopping independence. If the dependence-robust set is empty, the observed list and marginal attention restrictions are rejected even after independence is removed. If it is nonempty while the independent set is empty, the data diagnose independence as the source of misspecification.

SA-3.8.4 Dropping Sequential Path Independence

The preceding results use Sequential Path Independence to recover every prefix mass from outside-option probabilities on suffix menus. To assess this restriction, retain the observed list, prefix consideration, attention overload, and a stable marginal preference distribution, but do not impose Sequential Path Independence. This relaxation also permits arbitrary menu-specific dependence between preferences and stopping.

For every observed menu $S = \{a_{s_1}, \ldots, a_{s_{|S|}}\}$, let $m_j(S)$ be the latent probability of stopping at prefix $S_{:j}$, with $m_0(S)$ denoting empty consideration. For every $\succ \in \mathcal{P}$ and $j = 0, \ldots, |S|$, let $q_S(\succ, j)$ be the joint probability of preference $\succ$ and stopping at j. Define $\mathcal{Q}_{\mathcal{D}}^{\mathrm{noPI}}(\pi)$ by the following linear restrictions:

$$m_j(S) \geq 0, \qquad m_0(S) = \pi(o^*|S), \qquad \sum_{j=0}^{|S|} m_j(S) = 1, \tag{SA-1}$$

$$q_S(\succ, j) \geq 0, \qquad \sum_{\succ \in \mathcal{P}} q_S(\succ, j) = m_j(S), \qquad \sum_{j=0}^{|S|} q_S(\succ, j) = \tau(\succ), \tag{SA-2}$$

$$\pi(a|S) = \sum_{j=1}^{|S|} \sum_{\succ \in \mathcal{P}} \mathbb{1}\{a \in S_{:j},\ a \text{ is } \succ\text{-best in } S_{:j}\} q_S(\succ, j), \tag{SA-3}$$

for all relevant menus, positions, preferences, and inside alternatives. If $a_{s_j} = a_{r_k} \in S \subseteq R$, attention overload requires

$$\sum_{\ell=j}^{|S|} m_\ell(S) \geq \sum_{\ell=k}^{|R|} m_\ell(R). \tag{SA-4}$$

Indeed, the two sums are exactly the latent probabilities of reaching the common alternative in the smaller and larger menus. Let

$$\mathcal{T}_{\mathcal{D}}^{\mathrm{noPI}}(\pi) = \{\tau : (\tau, q, m) \in \mathcal{Q}_{\mathcal{D}}^{\mathrm{noPI}}(\pi) \text{ for some } (q, m)\}.$$

Proposition SA-6 (Path-Independence-Robust H-LAO)**.** For any observed-menu domain $\mathcal{D}$, $\mathcal{T}_{\mathcal{D}}^{\mathrm{noPI}}(\pi)$ is the sharp identified set for the stable marginal preference distribution when consideration sets are list prefixes, preferences and stopping may be arbitrarily dependent within menu, and Sequential Path Independence is not imposed. It is a compact polytope, does not require positive reach or suffix closure, and gives sharp preference-event bounds by linear programming. Whenever $\mathcal{D}$ is suffix closed and recovered attention is valid,

$$\mathcal{T}_{\mathcal{D}}(\pi) \subseteq \mathcal{T}_{\mathcal{D}}^{\mathrm{dep}}(\pi) \subseteq \mathcal{T}_{\mathcal{D}}^{\mathrm{noPI}}(\pi).$$

Proof. Consider first any representation in the stated class. Its menu-specific distribution over prefix stopping positions supplies nonnegative masses $m_j(S)$ satisfying the stopping and adding-up restrictions; the outside-option exclusion gives $m_0(S) = \pi(o^*|S)$. Let $q_S(\succ, j)$ be the joint distribution of preference and stopping in menu S. Stability of the preference marginal and the stopping marginal give the marginal restrictions. Utility maximization on each realized prefix gives the choice restrictions, and attention overload gives the reach inequalities. Every admissible representation therefore projects into $\mathcal{T}_{\mathcal{D}}^{\mathrm{noPI}}(\pi)$.

Conversely, take any feasible (τ, q, m). In each observed menu S, draw $(\succ, j)$ according to q_S, consider the empty set when $j = 0$ and prefix $S_{:j}$ otherwise, and choose the $\succ$-best alternative in a nonempty prefix. The marginal stopping probabilities are $m_j(S)$, so the process is list based and its reach probabilities satisfy attention overload by the displayed reach inequalities. Its preference marginal is the same τ in every menu. The outside and inside choice equations reproduce π. Sequential Path Independence has not been used. Thus every projected τ is attainable, proving sharpness.

All constraints are linear and all variables lie in finite-dimensional probability simplices, so the feasible set and its projection are compact polytopes. Linear optimization therefore gives attained sharp bounds on every preference event. An independent H-LAO representation is a special case of a dependence-robust representation. Every dependence-robust representation satisfying Sequential Path Independence satisfies the preceding stopping, marginal, choice, and attention-overload constraints after its recovered prefix masses are treated as latent variables. This proves both inclusions. ■

The path-independence-robust result converts the unobserved prefix masses into identified-set variables rather than recovering attention pointwise. Comparing the benchmark, dependence-robust, and no-Sequential-Path-Independence bounds therefore measures, in sequence, the identifying content of preference–stopping independence and Sequential Path Independence. The second relaxation naturally widens the bounds while delivering a sharp specification for arbitrary observed-menu domains without requiring suffix observations.

SA-3.8.5 Pairwise bounds under weakened restrictions

The robustness envelopes can remain informative about selected preference events even when the complete ranking distribution is not point identified. The binary case makes this content transparent.

Proposition SA-7 (Robust Pairwise Preference Bounds)**.** Suppose a precedes b and write $\theta = \tau(\{\succ : b \succ a\})$.

1. If $\{a, b\}$ and $\{b\}$ are observed and Sequential Path Independence is maintained while preferences and stopping may be dependent, let $r = r_b(a, b)$ be the recovered probability of reaching b. Using these two menus alone, the sharp identified interval is

$$\pi(b|\{a, b\}) \leq \theta \leq \pi(b|\{a, b\}) + 1 - r.$$

   Its width is $1 - r$; in particular, $r = 1$ point identifies $\theta = \pi(b|\{a, b\})$ without preference–stopping independence.

2. If Sequential Path Independence is also removed and preferences and stopping may be dependent, the binary menu alone gives the sharp interval

$$\pi(b|\{a, b\}) \leq \theta \leq 1.$$

   If an exogenously assigned second list puts b before a, pooling the two list strata gives the sharp interval

$$\pi(b|\{a, b\}) \leq \theta \leq 1 - \pi^{\text{rev}}(a|\{a, b\}).$$

Additional observed menus can only tighten these intervals.

Proof. Let $E = \{\succ : b \succ a\}$ and let $J = 2$ denote attention reaching the full binary prefix. Under Sequential Path Independence, $\mathbb{P}(J = 2) = r$ is recovered. The later item b is chosen exactly on $E \cap \{J = 2\}$, so

$$\pi(b|\{a, b\}) = \mathbb{P}(E \cap \{J = 2\}).$$

Consequently, $\theta = \mathbb{P}(E) \geq \pi(b|\{a, b\})$. Moreover,

$$1 \geq \mathbb{P}(E \cup \{J = 2\}) = \theta + r - \pi(b|\{a, b\}),$$

which gives $\theta \leq \pi(b|\{a, b\}) + 1 - r$. To prove sharpness, fix any θ in this interval and assign masses

$$\mathbb{P}(E, J = 2) = \pi(b|\{a, b\}), \qquad \mathbb{P}(E^c, J = 2) = r - \pi(b|\{a, b\}),$$
$$\mathbb{P}(E, J < 2) = \theta - \pi(b|\{a, b\}), \qquad \mathbb{P}(E^c, J < 2) = 1 - r - \theta + \pi(b|\{a, b\}).$$

All four masses are nonnegative. The two $J < 2$ masses can be distributed across empty consideration and the singleton prefix to match their recovered stopping probabilities; choices at those prefixes do not depend on whether E holds. This construction preserves the stable preference marginal, the recovered stopping marginal, and every choice probability in the two menus.

Without Sequential Path Independence, choosing the later item still implies E, and therefore $\theta \geq \pi(b|\{a, b\})$. Conversely, for any $\theta \in [\pi(b|\{a, b\}), 1]$, assign probability $\pi(b|\{a, b\})$ to the full prefix and couple all of it with E. Assign the remaining preference mass across empty consideration and the singleton prefix, whose combined probability is $1 - \pi(b|\{a, b\})$. This reproduces the binary choice probabilities and proves sharpness using that menu alone. In a reversed exogenous list, choosing the later item a implies E^c, so $1 - \theta \geq \pi^{\text{rev}}(a|\{a, b\})$. Applying the preceding construction separately within each list stratum proves sharpness of the pooled interval. ■

Under benchmark independence and positive reach, the point-identified share $\pi(b|\{a, b\})/r$ lies in the dependence-robust interval because $\pi(b|\{a, b\}) = r\theta$. The result therefore quantifies the identifying contribution of independence: dropping it increases the width from zero to exactly $1-r$. Without Sequential Path Independence, a single ordering still supplies a sharp one-sided revealed-preference share, while exogenous reversal of the list supplies the opposite side.

SA-3.8.6 Exogenous list variation

As a final sensitivity exercise, suppose exogenously assigned list strata $\ell = 1, \ldots, L$ share a common preference distribution τ but may have stratum-specific attention. Under benchmark independence, let $m_{\ell j}(S)$ be the recovered prefix masses and $S^{\ell}_{:j}$ the jth prefix under list ℓ. Stack the inside-choice probabilities in p and define

$$A_{(\ell,S,a),\succ} = \sum_{j=1}^{|S|} m_{\ell j}(S) \mathbb{1}\{a \in S^{\ell}_{:j},\ a \text{ is } \succ\text{-best in } S^{\ell}_{:j}\}.$$

Corollary SA-1 (Multiple-list pooling)**.** The sharp pooled benchmark set is

$$\mathcal{T}^{\text{pool}} = \bigcap_{\ell=1}^{L} \mathcal{T}_{\mathcal{D}_\ell}(\pi_\ell) = \{\tau \geq 0 : \mathbf{1}'\tau = 1,\ A\tau = p\}.$$

The complete ranking distribution is globally point identified if and only if

$$\text{rank}\begin{pmatrix} \mathbf{1}' \\ A \end{pmatrix} = |\mathcal{P}|, \qquad \text{rank}\begin{pmatrix} \mathbf{1}' \\ A \end{pmatrix} \leq (|X| - 2)2^{|X|-1} + 2.$$

Thus, when $|X| \geq 4$, the rank bound rules out global identification of all $|X|!$ ranking probabilities from list variation alone. Each additional list nevertheless weakly tightens preference-event bounds and can identify selected targets. Under dependence or without Sequential Path Independence, the sharp pooled set is the intersection of the corresponding stratum-specific sensitivity sets.

Proof. The pooled set is the intersection because each list stratum must share the same preference marginal and may use its own attention representation. The matrix form follows from the choice equations. Full column rank is equivalent to injectivity. Every row of A is a linear combination of full-attention best-in-subset indicators; within each menu those indicators add to $\mathbf{1}'$, giving the stated rank bound. Adding a list appends rows to A, so it cannot enlarge the identified set. ∎

SA-4 Econometric Methods

This section follows Section 4 of the main paper. It first proves Theorems 6–9 and Corollaries 4–6, then develops additional inference procedures for homogeneous AOM and H-LAO.

SA-4.1 Proof of Theorem 6

We work conditionally on the realized menu assignments, as specified in the main paper. Thus, the menu-specific sample sizes are fixed and observations belonging to distinct menus are independent. We fix a preference $\succ$ throughout this proof. As in the main paper, let

$$\mathcal{I}_{\succ} = \{(a, S, T) : T, S \in \mathcal{D},\ T \subset S,\ a \in T,\ U_{\succeq}(a) \cap T \neq T\}.$$

denote the nonredundant comparison index, and let $\mathfrak{c}_{\succ} = |\mathcal{I}_{\succ}|$. To ease notation, we also assume that alternatives in a choice problem S are ordered such that $a_1 \prec a_2 \prec \cdots \prec a_{|S|}$. Now further fix a choice problem S. Recall that the choice probabilities and their estimates are

$$\pi(a_j|S) = \mathbb{P}[y_i = a_j|Y_i = S], \text{ and } \hat{\pi}(a_j|S) = \frac{1}{N_S}\sum_{i=1}^{n} \mathbb{1}(y_i = a_j, Y_i = S).$$

Standard deviation and standard error of the estimated choice probabilities are $\sigma(a_j|S)$ and $\hat{\sigma}(a_j|S)$, respectively,

$$\sigma(a_j|S) = \sqrt{\frac{\pi(a_j|S)(1 - \pi(a_j|S))}{N_S}}, \text{ and } \hat{\sigma}(a_j|S) = \sqrt{\frac{\hat{\pi}(a_j|S)(1 - \hat{\pi}(a_j|S))}{N_S}}.$$

As our testing procedure involves the probabilities of picking an alternative in an upper contour set, also let

$$\pi(U_{\succeq}(a_j)|S) = \mathbb{P}[y_i \succeq a_j|Y_i = S], \text{ and } \hat{\pi}(U_{\succeq}(a_j)|S) = \frac{1}{N_S}\sum_{i=1}^{n} \mathbb{1}(y_i \succeq a_j, Y_i = S).$$

Collect the preceding population probabilities and their estimates in the column vectors

$$\boldsymbol{\pi}_S = \Big(\pi(a_1|S), \ldots, \pi(a_{|S|}|S), \pi(U_{\succeq}(a_1)|S), \ldots, \pi(U_{\succeq}(a_{|S|})|S)\Big)',$$
$$\hat{\boldsymbol{\pi}}_S = \Big(\hat{\pi}(a_1|S), \ldots, \hat{\pi}(a_{|S|}|S), \hat{\pi}(U_{\succeq}(a_1)|S), \ldots, \hat{\pi}(U_{\succeq}(a_{|S|})|S)\Big)'.$$

The upper-contour entries in $\hat{\boldsymbol{\pi}}_S$ are linear combinations of the individual-alternative probabilities. We retain them because they streamline the proof; implementation can work directly with the individual probabilities.

We first present the following lemma, which provides a strong approximation to the estimated choice probabilities. As in the main paper, every logarithmic factor in the econometric arguments below is understood to be truncated below at one. This convention handles one-dimensional comparison sets, while the maintained restriction $N_S \geq 2$ already excludes zero sample-size logarithms.

Lemma SA-1. In a possibly enlarged probability space, there exists a mean-zero Gaussian vector $\mathfrak{n}_S$ with the same covariance matrix as $\hat{\boldsymbol{\pi}}_S$, such that

$$\mathbb{P}\Big[\frac{N_S}{\log(N_S)}\big\|\hat{\boldsymbol{\pi}}_S - \boldsymbol{\pi}_S - \mathfrak{n}_S\big\|_\infty > 2\Big(\frac{t}{\log(N_S)} + c_1\Big)\Big] \leq c_2 \exp\{-c_3 t\},$$

where c_1–c_3 are universal constants.

Now consider the collection $\mathcal{D}$ of menus that enter at least one comparison for testing $\succ$. These menus occur in nested pairs. Using any fixed ordering of $\mathcal{D}$, define $\boldsymbol{\pi}$ and $\hat{\boldsymbol{\pi}}$ as the column vectors obtained by stacking $\boldsymbol{\pi}_S$ and $\hat{\boldsymbol{\pi}}_S$, respectively, in that common order, and let $\mathfrak{n}$ stack the corresponding $\mathfrak{n}_S$ in the same order. Because every retained menu enters at least one comparison and each comparison uses two menus,

$$\mathfrak{c}_{\mathcal{D}} \leq 2\mathfrak{c}_{\succ}$$

whenever $\mathfrak{c}_{\succ} \geq 1$. This relation lets us replace logarithms of $\mathfrak{c}_{\mathcal{D}}$ by logarithms of $\mathfrak{c}_{\succ}$ in the final rates. We have the following strong approximation result.

Lemma SA-2. Under the setting of Lemma SA-1,

$$\mathbb{P}\Big[\max_{S\in\mathcal{D}} \frac{N_S}{\log(N_S)}\big\|\hat{\boldsymbol{\pi}}_S - \boldsymbol{\pi}_S - \mathfrak{n}_S\big\|_\infty > c_1\Big(\log(\mathfrak{c}_{\mathcal{D}}) + 1\Big)\Big] \leq c_2 \exp\Big\{ - c_3\Big(\log\big(\min_{S\in\mathcal{D}} N_S\big) - 1\Big)\log(\mathfrak{c}_{\mathcal{D}})\Big\},$$

where

$$\mathfrak{c}_{\mathcal{D}} = \#\text{ of choice problems in } \mathcal{D},$$

c_1–c_3 are universal constants, and c_3 could be made arbitrarily large by appropriate choice of c_1.

It is worth mentioning that we retained the sample sizes associated with each choice problem separately, instead of showing a bound on $\|\hat{\boldsymbol{\pi}} - \boldsymbol{\pi} - \mathfrak{n}\|_\infty$ that necessarily depends on the minimum sample size. This turns out to deliver sharper and more informative results.

We denote the nonredundant constraints implied by $\succ$-Regularity by the matrix $\mathbf{R}_{\succ}$, whose rows are indexed by $\mathcal{I}_{\succ}$. The row associated with $(a, S, T) \in \mathcal{I}_{\succ}$ has $+1$ in the coordinate for $\pi(a|S)$, -1 in the coordinate for $\pi(U_{\succeq}(a)|T)$, and zeros elsewhere. Hence,

$$[\mathbf{R}_{\succ}\boldsymbol{\pi}]_{(a,S,T)} = \pi(a|S) - \pi(U_{\succeq}(a)|T) = D(a|S,T).$$

The variances of $\mathbf{R}_{\succ}\hat{\boldsymbol{\pi}}$ are collected in $\boldsymbol{\sigma}^2$, and its estimate is $\hat{\boldsymbol{\sigma}}^2$. Also recall from the main paper that elements of $\boldsymbol{\sigma}^2$ are denoted by $\sigma^2(a|S,T)$:

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

We employ $\boldsymbol{\Omega}$ and $\hat{\boldsymbol{\Omega}}$ for the true and estimated correlation matrices of $\mathbf{R}_{\succ}\hat{\boldsymbol{\pi}}$. Rows and columns of $\hat{\boldsymbol{\Omega}}$ corresponding to zero estimated standard errors are set to zero. Set $\mathbf{Z} = (\mathbf{R}_{\succ}\mathfrak{n}) \oslash \boldsymbol{\sigma}$, where $\oslash$ is the Hadamard (element-wise) division. To state the following result, define

$$\mathfrak{c}_\sigma = \min_{(a,S,T)\in\mathcal{I}_{\succ}} \left\{ \min\left\{\frac{N_S}{\log(N_S)}, \frac{N_T}{\log(N_T)}\right\} \cdot \max\left\{\sigma(a|S),\ \sigma(U_{\succeq}(a)|T)\right\}\right\},$$

which, loosely speaking, corresponds to the root minimum effective sample size.

Lemma SA-3. The centered and normalized constraints $\mathbf{R}_{\succ}(\hat{\boldsymbol{\pi}}-\boldsymbol{\pi})\oslash\boldsymbol{\sigma}$ satisfy

$$\mathbb{P}\left[\left\|\mathbf{R}_{\succ}(\hat{\boldsymbol{\pi}}-\boldsymbol{\pi})\oslash\boldsymbol{\sigma}-\mathbf{Z}\right\|_\infty > c_1\frac{\log(\mathfrak{c}_{\mathcal{D}})}{\mathfrak{c}_\sigma}\right] \le c_2\exp\left\{-c_3\Big(\log\big(\min_{S\in\mathcal{D}} N_S\big)-1\Big)\log(\mathfrak{c}_{\mathcal{D}})\right\},$$

where c_1–c_3 are universal constants, and c_3 could be made arbitrarily large by appropriate choice of c_1.

The next step is to replace the infeasible standard error, for which we employ the lemma below. Division by $\hat{\boldsymbol{\sigma}}$ uses the main paper's coordinatewise convention: $0/0$ is zero and a nonzero numerator divided by zero is signed infinity. The standard-error concentration bound below controls the event that any such denominator vanishes.

Lemma SA-4. Let $\xi_1, \xi_2 > 0$ with $\xi_2 \to 0$. Then,

$$\begin{aligned}\mathbb{P}\Bigg[\Big\|(\mathbf{R}_{\succ}\hat{\boldsymbol{\pi}}-\mathbf{R}_{\succ}\boldsymbol{\pi})\oslash\boldsymbol{\sigma}-(\mathbf{R}_{\succ}\hat{\boldsymbol{\pi}}-\mathbf{R}_{\succ}\boldsymbol{\pi})\oslash\hat{\boldsymbol{\sigma}}\Big\|_\infty &> \Big(\xi_1 + c_4\log^{\frac12}(\mathfrak{c}_{\succ}) + c_1\frac{\log(\mathfrak{c}_{\mathcal{D}})}{\mathfrak{c}_\sigma}\Big)\cdot\xi_2\Bigg]\\ &\le c_2\exp\left\{-c_3\Big(\log\big(\min_{S\in\mathcal{D}}N_S\big)-1\Big)\log(\mathfrak{c}_{\mathcal{D}})\right\} + \exp\left\{-\frac{\xi_1^2}{2}\right\} + c_5\exp\left\{-c_6\mathfrak{c}_\sigma^2\xi_2^2+\log(\mathfrak{c}_{\succ})\right\},\end{aligned}$$

where c_1–c_3 are universal constants defined in Lemma SA-3,

$$\mathfrak{c}_{\succ} = \#\text{ of rows in }\mathbf{R}_{\succ},$$

and c_4–c_6 are universal constants.

As a byproduct of the discussion, we have

Lemma SA-5. Let $\xi_2 > 0$ with $\xi_2 \to 0$. Then,

$$\mathbb{P}\Big[\big\|\,(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}) \oslash \boldsymbol{\sigma}\big\|_\infty > \xi_2\Big] \leq c_5 \exp\Big\{ - c_6 \mathfrak{c}_\sigma^2 \xi_2^2 + \log(\mathfrak{c}_\succ)\Big\},$$

where c_5 and c_6 are universal constants defined in Lemma SA-4.

The proof of the theorem then follows from Lemmas SA-3 and SA-4. Specifically, we have that

$$\begin{aligned}
&\mathbb{P}\Big[\big\|\mathbf{R}_\succ(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}) \oslash \hat{\boldsymbol{\sigma}} - \mathbf{Z}\big\|_\infty > c_1 \frac{\log(\mathfrak{c}_\mathcal{D})}{\mathfrak{c}_\sigma} + \Big(\xi_1 + c_4 \log^{\frac{1}{2}}(\mathfrak{c}_\succ) + c_1 \frac{\log(\mathfrak{c}_\mathcal{D})}{\mathfrak{c}_\sigma}\Big) \cdot \xi_2\Big] \\
\leq& c_2 \exp\Big\{ - c_3 \Big(\log\big(\min_{S\in\mathcal{D}} N_S\big) - 1\Big) \log(\mathfrak{c}_\mathcal{D})\Big\} + \exp\Big\{ - \frac{\xi_1^2}{2}\Big\} + c_5 \exp\Big\{ - c_6 \mathfrak{c}_\sigma^2 \xi_2^2 + \log(\mathfrak{c}_\succ)\Big\}
\end{aligned}$$

by applying the union bound. To simplify, we set

$$\xi_2 \propto \frac{\log^{\frac{1}{2}}(\mathfrak{c}_\succ)}{\mathfrak{c}_\sigma},$$

and let ξ_1 be some large constant. This will imply the probabilistic bound on the Gaussian approximation

$$\big\|\mathbf{R}_\succ(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}) \oslash \hat{\boldsymbol{\sigma}} - \mathbf{Z}\big\|_\infty = O_\mathbb{P}\left(\frac{\log(\mathfrak{c}_\succ)}{\mathfrak{c}_\sigma}\right).$$

SA-4.1.1 Proof of Lemma SA-1

Conditional on the menu assignments, let $u_{S,1}, \dots, u_{S,N_S}$ be i.i.d. uniform random variables on $[0,1]$. Given the choice probabilities $\pi(a_j|S)$, define the disjoint intervals

$$I_j = \left[\sum_{j'<j} \pi(a_{j'}|S),\ \sum_{j'\leq j} \pi(a_{j'}|S)\right),$$

with the right endpoint of the final interval included. Let $I_{j:} = \bigcup_{k\geq j} I_k$. The probability-integral construction gives

$$\pi(a_j|S) = \mathbb{P}(u_{S,i} \in I_j), \qquad \pi(U_\succeq(a_j)|S) = \mathbb{P}(u_{S,i} \in I_{j:}),$$

and the corresponding sample analogues are

$$\hat{\pi}(a_j|S) = \frac{1}{N_S} \sum_{i=1}^{N_S} \mathbb{1}(u_{S,i} \in I_j),$$

$$\hat{\pi}(U_{\succeq}(a_j)|S) = \frac{1}{N_S}\sum_{i=1}^{N_S} \mathbb{1}(u_{S,i} \in I_{j:}).$$

Let $\mathcal{H}_S$ contain the indicator functions of the intervals I_j and $I_{j:}$. Every member has total variation at most two. Indexing this finite class by distinct real numbers and setting the remaining index functions to zero verifies the hypotheses of the bounded-variation KMT corollary in Cattaneo, Feng, and Underwood (2024, Lemma SA26); see also Cattaneo and Yu (2025, Remark 1). That lemma therefore supplies, on an enlarged probability space, a Gaussian process G_S with the covariance of $\sqrt{N_S}(\hat{\boldsymbol{\pi}}_S - \boldsymbol{\pi}_S)$ and universal constants such that

$$\mathbb{P}\left[\left\|\sqrt{N_S}(\hat{\boldsymbol{\pi}}_S - \boldsymbol{\pi}_S) - G_S\right\|_\infty > \frac{2t + c_1\log(N_S)}{\sqrt{N_S}}\right] \leq c_2\exp(-c_3 t).$$

Setting $\mathfrak{n}_S = G_S/\sqrt{N_S}$ and enlarging c_1 if needed gives the displayed bound in the lemma. The covariance of $\mathfrak{n}_S$ is exactly that of $\hat{\boldsymbol{\pi}}_S$.

SA-4.1.2 Proof of Lemma SA-2

Conditional on the menu assignments, choices from distinct menus are independent. We may therefore carry out the preceding construction for every S on one product probability space, with independent auxiliary uniforms and Gaussian vectors across menus. Apply Lemma SA-1 with

$$t_S = C\log(N_S)\log(\mathfrak{c}_{\mathcal{D}})$$

and take a union bound over the $\mathfrak{c}_{\mathcal{D}}$ menus. After increasing the universal constants, the threshold is $c_1\{\log(\mathfrak{c}_{\mathcal{D}}) + 1\}$ and the union probability is bounded by

$$c_2\exp\left[-c_3\left\{\log\left(\min_{S\in\mathcal{D}} N_S\right) - 1\right\}\log(\mathfrak{c}_{\mathcal{D}})\right],$$

as stated.

SA-4.1.3 Proof of Lemma SA-3

To start, note that

$$\mathbf{R}_{\succ}(\hat{\boldsymbol{\pi}}-\boldsymbol{\pi})\oslash\boldsymbol{\sigma}-\mathbf{Z}=\mathbf{R}_{\succ}\big(\hat{\boldsymbol{\pi}}-\boldsymbol{\pi}-\mathfrak{n}\big)\oslash\boldsymbol{\sigma}=\tilde{\mathbf{R}}_{\succ}\begin{bmatrix}\frac{N_S}{\log(N_S)}\big(\hat{\boldsymbol{\pi}}_S-\boldsymbol{\pi}_S-\mathfrak{n}_S\big)\\ \frac{N_T}{\log(N_T)}\big(\hat{\boldsymbol{\pi}}_T-\boldsymbol{\pi}_T-\mathfrak{n}_T\big)\\ \vdots\end{bmatrix}.$$

In other words, we rewrite the normalized constraints as the product of some matrix $\tilde{\mathbf{R}}_{\succ}$ and the scaled choice probabilities. Since each row of $\mathbf{R}_{\succ}$ consists of exactly one $+1$ and one -1, the rows of $\tilde{\mathbf{R}}_{\succ}$ will contain two non-zero entries, taking the form

$$\frac{\log(N_S)}{N_S}\frac{1}{\sigma(a|S,T)}\quad\text{and}\quad-\frac{\log(N_T)}{N_T}\frac{1}{\sigma(a|S,T)}.$$

The next step is to bound the row-sum norm of $\tilde{\mathbf{R}}_{\succ}$ (defined as the maximum of the l_1 norm of each row of a matrix). Specifically, consider a generic row that corresponds to a comparison $(a,S,T)\in\mathcal{I}_{\succ}$. Then the l_1 norm of this row is bounded by

$$2\left[\min\Big(\frac{N_S}{\log(N_S)},\frac{N_T}{\log(N_T)}\Big)\right]^{-1}\left[\max\Big(\frac{\pi(a|S)(1-\pi(a|S))}{N_S},\frac{\pi(U_{\succeq}(a)|T)(1-\pi(U_{\succeq}(a)|T))}{N_T}\Big)\right]^{-\frac{1}{2}}.$$

To summarize, we have

$$\begin{aligned}&\big\|\mathbf{R}_{\succ}(\hat{\boldsymbol{\pi}}-\boldsymbol{\pi})\oslash\boldsymbol{\sigma}-\mathbf{Z}\big\|_\infty\\ &\leq 2\max_{(a,S,T)\in\mathcal{I}_{\succ}}\left[\min\Big(\frac{N_S}{\log(N_S)},\frac{N_T}{\log(N_T)}\Big)\right]^{-1}\left[\max\Big(\sigma(a|S),\ \sigma(U_{\succeq}(a)|T)\Big)\right]^{-1}\\ &\quad\times\max_{S\in\mathcal{D}}\frac{N_S}{\log(N_S)}\big\|\hat{\boldsymbol{\pi}}_S-\boldsymbol{\pi}_S-\mathfrak{n}_S\big\|_\infty,\end{aligned}$$

which implies the claim of the lemma.

SA-4.1.4 Proof of Lemma SA-4

Recall that individual components of $\boldsymbol{\sigma}^2$ are denoted by $\sigma^2(a|S,T)$. On the event that every required $\hat{\sigma}(a|S,T)$ is positive, simple algebra gives the following bound:

$$\Big\|(\mathbf{R}_{\succ}\hat{\boldsymbol{\pi}}-\mathbf{R}_{\succ}\boldsymbol{\pi})\oslash\boldsymbol{\sigma}-(\mathbf{R}_{\succ}\hat{\boldsymbol{\pi}}-\mathbf{R}_{\succ}\boldsymbol{\pi})\oslash\hat{\boldsymbol{\sigma}}\Big\|_\infty\leq\Big\|(\mathbf{R}_{\succ}\hat{\boldsymbol{\pi}}-\mathbf{R}_{\succ}\boldsymbol{\pi})\oslash\boldsymbol{\sigma}\Big\|_\infty\cdot\max_{(a,S,T)\in\mathcal{I}_{\succ}}\left|\frac{\sigma(a|S,T)}{\hat{\sigma}(a|S,T)}-1\right|.$$

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

Let $\xi_1 > 0$ be some constant which can depend on the sample size. Then by the triangle inequality, Lemma SA-3, maximal inequality for sub-Gaussian random variables, and the Borell-Tsirelson-Ibragimov-Sudakov inequality,

$$\begin{aligned}
&\mathbb{P}\left[\left\|(\mathbf{R}_{\succ}\hat{\boldsymbol{\pi}} - \mathbf{R}_{\succ}\boldsymbol{\pi}) \oslash \boldsymbol{\sigma}\right\|_{\infty} > c_4 \log^{\frac{1}{2}}(\mathfrak{c}_{\succ}) + \xi_1 + c_1 \frac{\log(\mathfrak{c}_{\mathcal{D}})}{\mathfrak{c}_{\sigma}}\right] \\
&\qquad \le \mathbb{P}\left[\left\|\mathbf{R}_{\succ}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}) \oslash \boldsymbol{\sigma} - \mathbf{Z}\right\|_{\infty} > c_1 \frac{\log(\mathfrak{c}_{\mathcal{D}})}{\mathfrak{c}_{\sigma}}\right] + \mathbb{P}\Big[\|\mathbf{Z}\|_{\infty} > c_4 \log^{\frac{1}{2}}(\mathfrak{c}_{\succ}) + \xi_1\Big] \\
&\qquad \le c_2 \exp\Big\{ - c_3 \Big(\log\big(\min_{S\in\mathcal{D}} N_S\big) - 1\Big) \log(\mathfrak{c}_{\mathcal{D}})\Big\} + \exp\Big\{ - \frac{\xi_1^2}{2}\Big\},
\end{aligned}$$

where $\mathbf{Z}$ and the constants c_1–c_3 are defined in Lemma SA-3,

$$\mathfrak{c}_{\succ} = \#\text{ of rows in } \mathbf{R}_{\succ},$$

and c_4 is a universal constant.

Next consider the standard error estimator, $\hat{\sigma}(a|S,T)$. We start with

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

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

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

provided that $\xi_2 \to 0$. Using the union bound, we deduce that

$$\mathbb{P}\left[\max_{(a,S,T)\in\mathcal{I}_{\succ}} \frac{\left|\hat{\sigma}(a|S,T)^2 - \sigma(a|S,T)^2\right|}{\sigma(a|S,T)^2} \ge \xi_2\right] \le 4\exp\left\{-\frac{1}{16}\mathfrak{c}_{\sigma}^2\xi_2^2 + \log(\mathfrak{c}_{\succ})\right\}, \qquad \text{provided that } \xi_2 \to 0,$$

which also implies that

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

If any required estimated standard error is zero, the maximum relative error in this display equals at least one. Hence, for $\xi_2 < 1$, the same probability bound also controls the complement of the event on which the preceding division is defined. To control the reciprocal ratio in the first display of the proof, apply the preceding bound with $\xi_2/2$. On the complementary event, and for all sufficiently large sample sizes such that $\xi_2 \leq 1$,

$$\begin{aligned}\left|\frac{\sigma(a|S,T)}{\hat{\sigma}(a|S,T)}-1\right| &= \frac{|\hat{\sigma}(a|S,T)/\sigma(a|S,T)-1|}{\hat{\sigma}(a|S,T)/\sigma(a|S,T)} \\ &\leq \frac{\xi_2/2}{1-\xi_2/2} \leq \xi_2.\end{aligned}$$

Replacing the numerical constants by universal constants c_5 and c_6 gives the probability bound stated in Lemma SA-4. This closes the proof.

SA-4.1.5 Proof of Lemma SA-5

This result follows directly from the proof of Lemma SA-4

SA-4.2 Proof of Theorem 7

Let

$$\mathcal{E}_+ = \left\{\min_{(a,S,T)\in\mathcal{I}_\succ} \hat{\sigma}(a|S,T) > 0\right\}.$$

All algebraic calculations involving division by $\hat{\boldsymbol{\sigma}}$ and the estimated correlation matrix are carried out on $\mathcal{E}_+$. On $\mathcal{E}_+^c$, we use the main paper's coordinatewise convention and bound its contribution to rejection probability by $\mathbb{P}(\mathcal{E}_+^c)$. Moreover, Lemma SA-5 gives, for any $\xi_2 < 1$ used below,

$$\mathbb{P}(\mathcal{E}_+^c) \leq c_5\exp\left\{-c_6\mathfrak{c}_\sigma^2\xi_2^2+\log(\mathfrak{c}_\succ)\right\} = o(1)$$

for the choice of ξ_2 at the end of the proof, with a sufficiently large proportionality constant. This bound is no larger than the covariance-estimation remainder below and can be absorbed by increasing its universal constant. Thus bounds established on $\mathcal{E}_+$ extend to the full sample space without changing their rates.

The proof of Theorem 7 considers a sequence of approximations, and one step involves the comparison between Gaussian vectors $\hat{\mathbf{Z}}$ and $\mathbf{Z}$

$$\hat{\mathbf{Z}} \mid \text{Data} \sim \mathcal{N}(0, \hat{\mathbf{\Omega}}), \quad \mathbf{Z} \sim \mathcal{N}(0, \mathbf{\Omega}).$$

We have the following technical lemma on the estimation error in the correlation matrix.

Lemma SA-6. Let $\xi_2 > 0$ with $\xi_2 \to 0$. Then,

$$\mathbb{P}\Big[\big\|\hat{\mathbf{\Omega}} - \mathbf{\Omega}\big\|_\infty > \xi_2\Big] \leq c_5 \exp\Big\{ - c_6 \mathfrak{c}_\sigma^2 \xi_2^2 + 2\log(\mathfrak{c}_\succ)\Big\},$$

where c_5 and c_6 are universal constants.

We remark that the constants in the above result may be different from those in Lemmas SA-4 and SA-5. Since they are universal constants and do not depend on the underlying data-generating process, however, we decide to keep the same notation.

The following technical lemma helps establish distributional proximity in terms of the estimation error in correlation/covariance matrices.

Lemma SA-7 (Proposition 2.1 in Chernozhukov, Chetverikov, Kato, and Koike 2022)**.** Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathfrak{c}_\succ}$ 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. Then,

$$\sup_{\substack{A \subseteq \mathbb{R}^{\mathfrak{c}_\succ} \\ A \text{ rectangular}}} \Big|\mathbb{P}[\mathbf{x} \in A] - \mathbb{P}[\mathbf{y} \in A]\Big| \leq c_7 \sqrt{\|\mathbf{\Sigma}^{\mathbf{x}} - \mathbf{\Sigma}^{\mathbf{y}}\|_\infty} \log(\mathfrak{c}_\succ),$$

where c_7 is a universal constant.

Now we are ready to state the following result, which provides a feasible Gaussian approximation to $\mathbf{Z}$.

Lemma SA-8. Take $\xi_2 > 0$ such that $\xi_2 \to 0$. Then

$$\mathbb{P}\left[\sup_{\substack{A \subseteq \mathbb{R}^{\mathfrak{c}_\succ} \\ A \text{ rectangular}}} \Big|\mathbb{P}[\mathbf{Z} \in A] - \mathbb{P}[\hat{\mathbf{Z}} \in A | \text{Data}]\Big| \leq c_7 \xi_2^{\frac{1}{2}} \log(\mathfrak{c}_\succ)\right] \geq 1 - c_5 \exp\Big\{ - c_6 \mathfrak{c}_\sigma^2 \xi_2^2 + 2\log(\mathfrak{c}_\succ)\Big\},$$

where the universal constants c_5–c_7 are defined in Lemmas SA-6 and SA-7.

The results we have established so far allow making connections between the critical values. Write

$\mathbf{Z} = (Z_1, \ldots, Z_{\mathfrak{c}_\succ})'$ and $\hat{\mathbf{Z}} = (\hat{Z}_1, \ldots, \hat{Z}_{\mathfrak{c}_\succ})'$. Define

$$\hat{\mathsf{T}}^{\mathsf{G}}(\succ) = \max\left\{0, \max_{1\le\ell\le\mathfrak{c}_\succ} \hat{Z}_\ell\right\}, \quad \mathsf{T}^{\mathsf{G}}(\succ) = \max\left\{0, \max_{1\le\ell\le\mathfrak{c}_\succ} Z_\ell\right\},$$

and the corresponding critical values

$$\hat{\mathrm{cv}}(\alpha, \succ) = \inf\{t \ge 0 : \ \mathbb{P}[\hat{\mathsf{T}}^{\mathsf{G}}(\succ) \le t|\mathrm{Data}] \ge 1-\alpha\}, \text{ and } \mathrm{cv}(\alpha, \succ) = \inf\{t \ge 0 : \ \mathbb{P}[\mathsf{T}^{\mathsf{G}}(\succ) \le t] \ge 1-\alpha\}.$$

Lemma SA-9. Take $\xi_2 > 0$ such that $\xi_2 \to 0$, let $\delta_2 = c_7\xi_2^{1/2}\log(\mathfrak{c}_\succ)$, and suppose

$$0 < \alpha - \delta_2 < \alpha + \delta_2 < \frac{1}{2}.$$

Then the critical values $\hat{\mathrm{cv}}(\alpha, \succ)$ and $\mathrm{cv}(\alpha, \succ)$ satisfy

$$\mathbb{P}\Big[\mathrm{cv}\big(\alpha+\delta_2, \succ\big) \le \hat{\mathrm{cv}}(\alpha, \succ) \le \mathrm{cv}\big(\alpha-\delta_2, \succ\big)\Big] \ge 1 - c_5\exp\Big\{-c_6\mathfrak{c}_\sigma^2\xi_2^2 + 2\log(\mathfrak{c}_\succ)\Big\},$$

where the universal constants c_5–c_7 are defined in Lemmas SA-6 and SA-7.

To wrap up the proof of Theorem 7, we rely on a sequence of error bounds. Let

$$\begin{aligned}
\mathsf{T}(\succ) &= \max\left\{0, \max_{1\le\ell\le\mathfrak{c}_\succ} [(\mathbf{R}_\succ\hat{\boldsymbol{\pi}}) \oslash \hat{\boldsymbol{\sigma}}]_\ell\right\}, \\
\mathsf{T}^\circ(\succ) &= \max\left\{0, \max_{1\le\ell\le\mathfrak{c}_\succ} [\mathbf{R}_\succ(\hat{\boldsymbol{\pi}}-\boldsymbol{\pi}) \oslash \hat{\boldsymbol{\sigma}}]_\ell\right\}, \\
\tilde{\mathsf{T}}^\circ(\succ) &= \max\left\{0, \max_{1\le\ell\le\mathfrak{c}_\succ} [\mathbf{R}_\succ(\hat{\boldsymbol{\pi}}-\boldsymbol{\pi}) \oslash \boldsymbol{\sigma}]_\ell\right\}.
\end{aligned}$$

To start, under H_0 and after enlarging c_5 to absorb $\mathbb{P}(\mathcal{E}_+^c)$, we have

$$\begin{aligned}
\mathbb{P}\Big[\mathsf{T}(\succ) > \hat{\mathrm{cv}}(\alpha, \succ)\Big] &\le \mathbb{P}\Big[\mathsf{T}^\circ(\succ) > \hat{\mathrm{cv}}(\alpha, \succ), \ \mathcal{E}_+\Big] + \mathbb{P}(\mathcal{E}_+^c) \\
&\le \mathbb{P}\Big[\mathsf{T}^\circ(\succ) > \mathrm{cv}\big(\alpha + c_7\xi_2^{\frac{1}{2}}\log(\mathfrak{c}_\succ), \succ\big)\Big] + c_5\exp\Big\{-c_6\mathfrak{c}_\sigma^2\xi_2^2 + 2\log(\mathfrak{c}_\succ)\Big\}.
\end{aligned}$$

Next, we apply Lemma SA-4 and obtain that

$$\begin{aligned}
&\mathbb{P}\Big[\mathsf{T}^\circ(\succ) > \mathrm{cv}\big(\alpha + c_7\xi_2^{\frac{1}{2}}\log(\mathfrak{c}_\succ), \succ\big)\Big] \\
&\le \mathbb{P}\Big[\tilde{\mathsf{T}}^\circ(\succ) > \mathrm{cv}\big(\alpha + c_7\xi_2^{\frac{1}{2}}\log(\mathfrak{c}_\succ), \succ\big) - \Big(\xi_1 + c_4\log^{\frac{1}{2}}(\mathfrak{c}_\succ) + c_1\frac{\log(\mathfrak{c}_{\mathcal{D}})}{\mathfrak{c}_\sigma}\Big)\cdot\xi_2\Big] \\
&+ c_2\exp\Big\{-c_3\Big(\log\big(\min_{S\in\mathcal{D}} N_S\big) - 1\Big)\log(\mathfrak{c}_{\mathcal{D}})\Big\} + \exp\Big\{-\frac{\xi_1^2}{2}\Big\} + c_5\exp\Big\{-c_6\mathfrak{c}_\sigma^2\xi_2^2 + \log(\mathfrak{c}_\succ)\Big\}.
\end{aligned}$$

The final error bound in our analysis is due to Lemma SA-3, which implies that

$$\begin{aligned}
&\mathbb{P}\Big[\tilde{\mathsf{T}}^{\circ}(\succ) > \mathrm{cv}\big(\alpha + c_7\xi_2^{\frac{1}{2}}\log(\mathfrak{c}_{\succ}), \succ\big) - \Big(\xi_1 + c_4\log^{\frac{1}{2}}(\mathfrak{c}_{\succ}) + c_1\frac{\log(\mathfrak{c}_{\mathcal{D}})}{\mathfrak{c}_{\sigma}}\Big)\cdot\xi_2\Big] \\
&\leq \mathbb{P}\Big[\mathsf{T}^{\mathsf{G}}(\succ) > \mathrm{cv}\big(\alpha + c_7\xi_2^{\frac{1}{2}}\log(\mathfrak{c}_{\succ}), \succ\big) - \Big(\xi_1 + c_4\log^{\frac{1}{2}}(\mathfrak{c}_{\succ}) + c_1\frac{\log(\mathfrak{c}_{\mathcal{D}})}{\mathfrak{c}_{\sigma}}\Big)\cdot\xi_2 - c_1\frac{\log(\mathfrak{c}_{\mathcal{D}})}{\mathfrak{c}_{\sigma}}\Big] \\
&\qquad + c_2\exp\Big\{-c_3\Big(\log\big(\min_{S\in\mathcal{D}} N_S\big) - 1\Big)\log(\mathfrak{c}_{\mathcal{D}})\Big\}.
\end{aligned}$$

Collecting all pieces, the size distortion of $\mathbb{P}[\mathsf{T}(\succ) > \hat{\mathrm{cv}}(\alpha, \succ)]$ is bounded as

$$\begin{aligned}
&\mathbb{P}[\mathsf{T}(\succ) > \hat{\mathrm{cv}}(\alpha, \succ)] \\
&\leq \mathbb{P}\Big[\mathsf{T}^{\mathsf{G}}(\succ) > \mathrm{cv}\big(\alpha + c_7\xi_2^{\frac{1}{2}}\log(\mathfrak{c}_{\succ}), \succ\big) - \Big(\xi_1 + c_4\log^{\frac{1}{2}}(\mathfrak{c}_{\succ}) + c_1\frac{\log(\mathfrak{c}_{\mathcal{D}})}{\mathfrak{c}_{\sigma}}\Big)\cdot\xi_2 - c_1\frac{\log(\mathfrak{c}_{\mathcal{D}})}{\mathfrak{c}_{\sigma}}\Big] \\
&\quad + c_2\exp\Big\{-c_3\Big(\log\big(\min_{S\in\mathcal{D}} N_S\big) - 1\Big)\log(\mathfrak{c}_{\mathcal{D}})\Big\} + \exp\Big\{-\frac{\xi_1^2}{2}\Big\} + c_5\exp\Big\{-c_6\mathfrak{c}_{\sigma}^2\xi_2^2 + 2\log(\mathfrak{c}_{\succ})\Big\}.
\end{aligned}$$

In the last step, we have absorbed a multiplicative constant 2 into the universal constant c_2 and c_5.

To proceed, we employ the following anti-concentration result for normal random vectors. The restriction $\alpha + \delta_2 < 1/2$ also resolves the truncation at zero: since any coordinate of $\mathbf{Z}$ is standard normal,

$$\mathbb{P}\left[\max_{1\leq\ell\leq\mathfrak{c}_{\succ}} Z_\ell \leq 0\right] \leq \frac{1}{2},$$

so every oracle critical value used below is strictly positive. In a neighborhood of such a critical value, $\max\{0, \max_\ell Z_\ell\}$ and $\max_\ell Z_\ell$ coincide.

Lemma SA-10 (Lemma D.4 in the supplemental appendix to Chernozhukov, Chetverikov, and Kato 2019)**.** Let $\mathbf{Z} \in \mathbb{R}^{\mathfrak{c}_{\succ}}$ be a mean-zero normal random vector such that $\mathbb{V}[Z_\ell] = 1$ for all $1 \leq \ell \leq \mathfrak{c}_{\succ}$. Then, for any $t \in \mathbb{R}$ and any $\varepsilon > 0$,

$$\mathbb{P}\left[\left|\max_{1\leq\ell\leq\mathfrak{c}_{\succ}} Z_\ell - t\right| \leq \varepsilon\right] \leq 4\varepsilon\Big(\sqrt{2\log(\mathfrak{c}_{\succ})} + 1\Big).$$

For brevity, define

$$\eta_2 = \left(\xi_1 + c_4\log^{1/2}(\mathfrak{c}_{\succ}) + c_1\frac{\log(\mathfrak{c}_{\mathcal{D}})}{\mathfrak{c}_{\sigma}}\right)\xi_2 + c_1\frac{\log(\mathfrak{c}_{\mathcal{D}})}{\mathfrak{c}_{\sigma}}.$$

The lemma above immediately implies that

$$\mathrm{cv}\Big(\alpha+\delta_2+4\eta_2\{\sqrt{2\log(\mathfrak{c}_{\succ})}+1\},\succ\Big)\le \mathrm{cv}(\alpha+\delta_2,\succ)-\eta_2.$$

A direct consequence is that

$$\mathbb{P}\left[\mathsf{T}^{\mathsf{G}}(\succ)>\mathrm{cv}(\alpha+\delta_2,\succ)-\eta_2\right]\le\alpha+\delta_2+4\eta_2\{\sqrt{2\log(\mathfrak{c}_{\succ})}+1\}.$$

Therefore, by possibly enlarging the constants,

$$\begin{aligned}
&\mathbb{P}[\mathsf{T}(\succ)>\hat{\mathrm{cv}}(\alpha,\succ)]\le\alpha\\
&\quad+c_2\exp\Big\{-c_3\Big(\log\big(\min_{S\in\mathcal{D}}N_S\big)-1\Big)\log(\mathfrak{c}_{\mathcal{D}})\Big\}+\exp\Big\{-\frac{\xi_1^2}{2}\Big\}+c_5\exp\Big\{-c_6\mathfrak{c}_\sigma^2\xi_2^2+2\log(\mathfrak{c}_{\succ})\Big\}\\
&\quad+c_7\xi_2^{\frac{1}{2}}\log(\mathfrak{c}_{\succ})+c_1\xi_2\Big(\xi_1+\log^{\frac{1}{2}}(\mathfrak{c}_{\succ})\Big)\log^{\frac{1}{2}}(\mathfrak{c}_{\succ})\\
&\quad+c_1\frac{\log(\mathfrak{c}_{\mathcal{D}})\log^{1/2}(\mathfrak{c}_{\succ})}{\mathfrak{c}_\sigma}.
\end{aligned}$$

Under the growth conditions of Theorem 7, all probability-level shifts above are smaller than $1/2-\alpha$ for sufficiently large samples. Now set

$$\xi_1\propto\log^{\frac{1}{2}}(\mathfrak{c}_\sigma),\quad\text{and}\quad\xi_2\propto\frac{\log^{\frac{1}{2}}(\mathfrak{c}_{\succ})\log^{\frac{1}{2}}(\mathfrak{c}_\sigma)}{\mathfrak{c}_\sigma},$$

Using $\mathfrak{c}_{\mathcal{D}}\le 2\mathfrak{c}_{\succ}$, the last term in the preceding display is of smaller order than the rate below under the stated growth conditions. Thus the chosen values give the desired bound

$$\mathbb{P}[\mathsf{T}(\succ)>\hat{\mathrm{cv}}(\alpha,\succ)]\le\alpha+c\frac{\log^{\frac{5}{4}}(\mathfrak{c}_{\succ})\log^{\frac{1}{4}}(\mathfrak{c}_\sigma)}{\mathfrak{c}_\sigma^{\frac{1}{2}}}$$

for some constant c.

SA-4.2.1 Proof of Lemma SA-6

Each restriction compares probabilities from two menus. Two restrictions, indexed by ℓ and ℓ', therefore involve at most four menus, denoted by $T\subset S$ and $T'\subset S'$. If $\{S,T\}\cap\{S',T'\}=\emptyset$, conditional independence across menus makes both the population covariance and its estimate zero. It remains to control pairs of restrictions that share at least one menu.

To start, we first decompose the difference as

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

Therefore,

$$\begin{aligned}
&\max_{(a,S,T)\in\mathcal{I}_\succ}\max_{(a',S',T')\in\mathcal{I}_\succ}\left|\frac{\hat\sigma(a|S,T;a'|S',T')}{\hat\sigma(a|S,T)\hat\sigma(a'|S',T')}-\frac{\sigma(a|S,T;a'|S',T')}{\sigma(a|S,T)\sigma(a'|S',T')}\right|\\
&\leq\max_{(a,S,T)\in\mathcal{I}_\succ}\max_{(a',S',T')\in\mathcal{I}_\succ}\left|\frac{\hat\sigma(a|S,T;a'|S',T')-\sigma(a|S,T;a'|S',T')}{\sigma(a|S,T)\sigma(a'|S',T')}\right|\\
&\qquad+2\max_{(a,S,T)\in\mathcal{I}_\succ}\left|\frac{\hat\sigma(a|S,T)}{\sigma(a|S,T)}-1\right|+\max_{(a,S,T)\in\mathcal{I}_\succ}\left|\frac{\hat\sigma(a|S,T)}{\sigma(a|S,T)}-1\right|^2.
\end{aligned}$$

We already have an error bound for the last two terms from Lemma SA-5, and hence it suffices to study the first term in the above display.

Here $\sigma(a|S,T;a'|S',T')$ denotes a covariance, whereas $\sigma(a|S,T)$ denotes a standard deviation. Up to interchanging the two restrictions, there are four overlap configurations: (i) $S=S'$ and $T\neq T'$; (ii) $T=T'$ and $S\neq S'$; (iii) $T=S'$; and (iv) $S=S'$ and $T=T'$. Menus not identified by these equalities are distinct.

Case (i). The covariance $\sigma(a|S,T;a'|S,T')$ can take two forms: $\frac{1}{N_S}\pi(a|S)(1-\pi(a|S))$ or $-\frac{1}{N_S}\pi(a|S)\pi(a'|S)$. Nevertheless, the following bound applies

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

Next we apply Bernstein's inequality, which gives

$$\begin{aligned}
&\mathbb{P}\left[\frac{1}{N_S\sigma(a|S,T)\sigma(a'|S,T')}\left|\hat\pi(a|S)-\pi(a|S)\right|\geq\xi_2\right]\\
&\leq 2\exp\left\{-\frac{1}{2}\frac{N_S^4\sigma(a|S,T)^2\sigma(a'|S,T')^2\xi_2^2}{N_S\pi(a|S)(1-\pi(a|S))+\frac{1}{3}N_S^2\sigma(a|S,T)\sigma(a'|S,T')\xi_2}\right\}\\
&=2\exp\left\{-\frac{1}{2}\frac{N_S^2\sigma(a|S,T)^2\sigma(a'|S,T')^2\xi_2^2}{\frac{1}{N_S}\pi(a|S)(1-\pi(a|S))+\frac{1}{3}\sigma(a|S,T)\sigma(a'|S,T')\xi_2}\right\}
\end{aligned}$$

$$
\begin{aligned}
&\leq 2\exp\left\{-\frac{1}{2}\frac{N_S^2\sigma(a|S,T)^2\sigma(a'|S,T')^2\xi_2^2}{\sigma(a|S,T)^2+\frac{1}{3}\sigma(a|S,T)\sigma(a'|S,T')\xi_2}\right\}\\
&\leq 2\exp\left\{-\frac{1}{2}\frac{N_S^2\sigma(a|S,T)^2\sigma(a'|S,T')^2\xi_2^2}{(\sigma(a|S,T)\vee\sigma(a'|S,T'))^2+\frac{1}{3}(\sigma(a|S,T)\vee\sigma(a'|S,T'))^2\xi_2}\right\}\\
&= 2\exp\left\{-\frac{1}{2}\frac{N_S^2(\sigma(a|S,T)\wedge\sigma(a'|S,T'))^2\xi_2^2}{1+\frac{1}{3}\xi_2}\right\}\\
&\leq 2\exp\left\{-\frac{1}{4}N_S^2(\sigma(a|S,T)\wedge\sigma(a'|S,T'))^2\xi_2^2\right\}\leq 2\exp\left\{-\frac{1}{4}\mathfrak{c}_\sigma^2\xi_2^2\right\}.
\end{aligned}
$$

provided that $\xi_2 \to 0$. Therefore, we have

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

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

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

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

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

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

$$
\underbrace{\frac{1}{N_T}\pi(U_\succeq(a)|T)\pi(a'|T)}_{\text{if } a'\notin U_\succeq(a)}\qquad\text{or}\qquad\underbrace{-\frac{1}{N_T}\Big(\pi(a'|T)-\pi(a'|T)\pi(U_\succeq(a)|T)\Big)}_{\text{if } a'\in U_\succeq(a)}.
$$

In either case, we have

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

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

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

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

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

Then,

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

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

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

Combining the four cases and taking a union bound over at most $\mathfrak{c}_\succ^2$ pairs gives, after changing universal constants,

$$\mathbb{P}\left[\max_{\ell,\ell'}\frac{|\hat{\sigma}_{\ell\ell'} - \sigma_{\ell\ell'}|}{\sigma_\ell\sigma_{\ell'}} > \xi_2\right] \leq c_5\exp\left\{-c_6\mathfrak{c}_\sigma^2\xi_2^2 + 2\log(\mathfrak{c}_\succ)\right\}.$$

Applying Lemma SA-5 to the remaining standard-error terms in the decomposition above, and absorbing their constants, proves the stated bound for $\|\hat{\mathbf{\Omega}} - \mathbf{\Omega}\|_\infty$.

SA-4.2.2 Proof of Lemma SA-8

This follows directly from Lemmas SA-6 and SA-7.

SA-4.2.3 Proof of Lemma SA-9

Work on the event in Lemma SA-8. The level restrictions in Lemma SA-9 imply that the critical values in this argument are positive. At nonnegative cutoffs, the distribution functions of the truncated maxima correspond to rectangular events for the untruncated Gaussian vectors, so the uniform comparison applies. Therefore,

$$\begin{aligned}\mathbb{P}\Big[\hat{\mathsf{T}}^{\mathtt{G}}(\succ) > \mathrm{cv}\big(\alpha - c_7\xi_2^{\frac{1}{2}}\log(\mathfrak{c}_\succ), \succ\big)\Big|\mathrm{Data}\Big] &\le \mathbb{P}\Big[\mathsf{T}^{\mathtt{G}}(\succ) > \mathrm{cv}\big(\alpha - c_7\xi_2^{\frac{1}{2}}\log(\mathfrak{c}_\succ), \succ\big)\Big] + c_7\xi_2^{\frac{1}{2}}\log(\mathfrak{c}_\succ)\\ &\le \alpha - c_7\xi_2^{\frac{1}{2}}\log(\mathfrak{c}_\succ) + c_7\xi_2^{\frac{1}{2}}\log(\mathfrak{c}_\succ) = \alpha.\end{aligned}$$

Therefore, on this event,

$$\hat{\mathrm{cv}}(\alpha, \succ) \le \mathrm{cv}\big(\alpha - c_7\xi_2^{\frac{1}{2}}\log(\mathfrak{c}_\succ), \succ\big).$$

Similarly, for every $0 \le t < \mathrm{cv}(\alpha + \delta_2, \succ)$, the same uniform comparison gives

$$\mathbb{P}\big[\hat{\mathsf{T}}^{\mathtt{G}}(\succ) \le t \mid \mathrm{Data}\big] \le \mathbb{P}\big[\mathsf{T}^{\mathtt{G}}(\succ) \le t\big] + \delta_2 < 1 - \alpha.$$

Here the oracle probability is marginal, not conditional on the data. By the definition of the feasible critical value, this implies

$$\hat{\mathrm{cv}}(\alpha, \succ) \ge \mathrm{cv}\big(\alpha + c_7\xi_2^{\frac{1}{2}}\log(\mathfrak{c}_\succ), \succ\big)$$

under the same event.

SA-4.3 Proof of Corollary 4

Corollary 4 restricts the comparison index in Theorem 7 to $(a, S, \{a, b\})$ with $S \in \mathcal{D}_{ab}$. The standardized statistic and its feasible Gaussian analogue are therefore coordinate restrictions of the corresponding objects in Theorem 7. The coupling, covariance-estimation, and Gaussian-comparison arguments apply unchanged with $\mathfrak{c}_\succ = |\mathcal{D}_{ab}|$ and with $\mathfrak{c}_\sigma$ specialized to the binary-screening comparisons, yielding the first displayed size-distortion bound.

It remains to prove the sharper fixed-pair statement. Let $\mathbf{Z}_{ab} \sim \mathcal{N}(0, \mathbf{\Omega}_{ab})$ and $\widehat{\mathbf{Z}}_{ab} \mid \mathrm{Data} \sim \mathcal{N}(0, \widehat{\mathbf{\Omega}}_{ab})$ be the corresponding oracle and feasible Gaussian vectors for the binary-screening comparisons indexed by

$S \in \mathcal{D}_{ab}$. Set

$$\rho_{ab} = \lambda_{\min}(\boldsymbol{\Omega}_{ab}), \qquad \Delta_{ab} = \|\widehat{\boldsymbol{\Omega}}_{ab} - \boldsymbol{\Omega}_{ab}\|_\infty.$$

Lemma SA-15 and Corollary SA-2 below imply that the balance condition gives $\rho_{ab} \geq 1/(1+C^2)$. The correlation-estimation argument in Lemma SA-6, restricted to these $|\mathcal{D}_{ab}|$ coordinates, gives

$$\mathbb{P}(\Delta_{ab} > \xi_2) \leq c_5 \exp\{-c_6 \mathfrak{c}_\sigma^2 \xi_2^2 + 2\log(|\mathcal{D}_{ab}|)\}.$$

On the event $\Delta_{ab} \leq \xi_2 < e^{-1}$, Theorem 2.3 of Lopes (2022) and the eigenvalue bound yield

$$\sup_{A \text{ rectangular}} \Big|\mathbb{P}(\mathbf{Z}_{ab} \in A) - \mathbb{P}(\widehat{\mathbf{Z}}_{ab} \in A \mid \text{Data})\Big| \leq c(1+C^2)\log(|\mathcal{D}_{ab}|)\log(1/\xi_2)\xi_2.$$

The result permits $\widehat{\boldsymbol{\Omega}}_{ab}$ to be singular; only the oracle correlation matrix must satisfy the displayed eigenvalue bound.

Let $L_{ab} = \max\{\log(|\mathcal{D}_{ab}|\mathfrak{c}_\sigma), 1\}$ and choose $\xi_1 = AL_{ab}^{1/2}$ and $\xi_2 = AL_{ab}^{1/2}/\mathfrak{c}_\sigma$ for a sufficiently large universal constant A. Repeat the proof of Theorem 7 with the preceding Lopes bound in place of Lemma SA-7. Because the menus entering a fixed-pair screen consist of the binary menu and at most one larger menu per comparison, $\mathfrak{c}_{\mathcal{D}} \leq |\mathcal{D}_{ab}| + 1$. The Gaussian-comparison contribution is bounded by

$$c(1+C^2)\frac{L_{ab}^{5/2}}{\mathfrak{c}_\sigma},$$

while the studentization and anti-concentration contribution and the strong-coupling contribution are each of order at most $L_{ab}^{3/2}/\mathfrak{c}_\sigma$. With A sufficiently large, the exponential remainder probabilities are also of smaller order. The stated condition makes the resulting level shift vanish and gives

$$\mathbb{P}\big[\mathsf{T}(ab) > \widehat{\text{cv}}(\alpha, ab)\big] \leq \alpha + c(1+C^2)\frac{\log^{5/2}(|\mathcal{D}_{ab}|\mathfrak{c}_\sigma)}{\mathfrak{c}_\sigma}.$$

This argument concerns a fixed ordered pair (a, b). Applying the same result to a statistic that stacks multiple pairs would additionally require a lower bound for the minimum eigenvalue of the joint correlation matrix. ■

SA-4.4 Proof of Theorem 8

If an estimated standard error is zero, the lower confidence bound is zero by convention and coverage is immediate. We therefore work on the complementary event. We start with three lemmas, the first two giving a Gaussian coupling and the third handling uncertainty in the standard error.

Lemma SA-11. In a possibly enlarged probability space, there exists a mean-zero Gaussian vector $\mathfrak{n}_R$ for each $R \supseteq S$ with the same covariance matrix as $\hat{\boldsymbol{\pi}}_R$, such that

$$\mathbb{P}\left[\max_{\substack{R\supseteq S\\ R\in\mathcal{D}}} \frac{N_R}{\log(N_R)}\|\hat{\boldsymbol{\pi}}_R - \boldsymbol{\pi}_R - \mathfrak{n}_R\|_\infty > c_1\{\log(\mathfrak{c}_{\supseteq S}) + 1\}\right]$$
$$\leq c_2 \exp\left[-c_3\left\{\log\left(\min_{\substack{R\supseteq S\\ R\in\mathcal{D}}} N_R\right) - 1\right\}\log(\mathfrak{c}_{\supseteq S})\right],$$

where

$$\mathfrak{c}_{\supseteq S} = \#\text{ of supersets of } S,$$

c_1–c_3 are universal constants, and c_3 could be made arbitrarily large by appropriate choice of c_1.

Lemma SA-12. Under the setting of Lemma SA-11, the centered and normalized choice probabilities satisfy

$$\mathbb{P}\Big[\max_{R\supseteq S, R\in\mathcal{D}}\Big|\frac{\hat{\pi}(a|R)-\pi(a|R)}{\sigma(a|R)} - \mathsf{Z}_{a,S}(R)\Big| > c_1\frac{\log(\mathfrak{c}_{\supseteq S})}{\mathfrak{c}_\sigma}\Big]$$
$$\leq c_2\exp\Big\{-c_3\Big(\log\big(\min_{R\supseteq S, R\in\mathcal{D}} N_R\big) - 1\Big)\log(\mathfrak{c}_{\supseteq S})\Big\},$$

where $\mathsf{Z}_{a,S}(R)$ are independent standard Gaussian random variables,

$$\mathfrak{c}_\sigma = \min_{R\supseteq S, R\in\mathcal{D}} \frac{\sqrt{N_R\pi(a|R)(1-\pi(a|R))}}{\log(N_R)},$$

and c_1–c_3 are universal constants such that c_3 could be made arbitrarily large by appropriate choice of c_1.

Lemma SA-13. Let $\xi_1, \xi_2 > 0$ with $\xi_2 \to 0$. Then,

$$\mathbb{P}\left[\max_{R\supseteq S, R\in\mathcal{D}}\Big|\frac{\hat{\pi}(a|R)-\pi(a|R)}{\sigma(a|R)} - \frac{\hat{\pi}(a|R)-\pi(a|R)}{\hat{\sigma}(a|R)}\Big| > \Big(\xi_1 + c_4\log^{\frac{1}{2}}(\mathfrak{c}_{\supseteq S}) + c_1\frac{\log(\mathfrak{c}_{\supseteq S})}{\mathfrak{c}_\sigma}\Big)\cdot\xi_2\right]$$
$$\leq c_2\exp\Big\{-c_3\Big(\log\big(\min_{R\supseteq S, R\in\mathcal{D}} N_R\big) - 1\Big)\log(\mathfrak{c}_{\supseteq S})\Big\} + \exp\Big\{-\frac{\xi_1^2}{2}\Big\} + c_5\exp\Big\{-c_6\mathfrak{c}_\sigma^2\xi_2^2 + \log(\mathfrak{c}_{\supseteq S})\Big\},$$

where c_1–c_3 are universal constants defined in Lemma SA-12, and c_4–c_6 are additional universal constants.

The rest of the proof then follows from the observation that $\phi(a|S) \geq \max_{R\supseteq S, R\in\mathcal{D}} \pi(a|R)$, and therefore

$$\begin{aligned}
&\mathbb{P}\Big[\phi(a|S) < \underline{\widehat{\phi}}(a|S)\Big] \\
&\leq \mathbb{P}\Big[\exists R \supseteq S,\ R\in\mathcal{D} : \hat{\pi}(a|R) - \mathrm{cv}(\alpha, \underline{\phi}(a|S)) \cdot \hat{\sigma}(a|R) > \pi(a|R)\Big] \\
&= \mathbb{P}\Big[\max_{R\supseteq S, R\in\mathcal{D}} \frac{\hat{\pi}(a|R) - \pi(a|R)}{\hat{\sigma}(a|R)} > \mathrm{cv}(\alpha, \underline{\phi}(a|S))\Big] \\
&\leq \mathbb{P}\Bigg[\max_{\substack{R\supseteq S\\ R\in\mathcal{D}}} \mathsf{Z}_{a,S}(R) > \mathrm{cv}(\alpha, \underline{\phi}(a|S)) - c_1\frac{\log(\mathfrak{c}_{\supseteq S})}{\mathfrak{c}_\sigma} \\
&\qquad\qquad - \Big\{\xi_1 + c_4\log^{1/2}(\mathfrak{c}_{\supseteq S}) + c_1\frac{\log(\mathfrak{c}_{\supseteq S})}{\mathfrak{c}_\sigma}\Big\}\xi_2\Bigg] \\
&\quad + c_2\exp\Bigg[-c_3\Bigg\{\log\Bigg(\min_{\substack{R\supseteq S\\ R\in\mathcal{D}}} N_R\Bigg) - 1\Bigg\}\log(\mathfrak{c}_{\supseteq S})\Bigg] \\
&\quad + \exp(-\xi_1^2/2) + c_5\exp\{-c_6\mathfrak{c}_\sigma^2\xi_2^2 + \log(\mathfrak{c}_{\supseteq S})\}.
\end{aligned}$$

Apply the anti-concentration result from Lemma SA-10, which gives

$$\begin{aligned}
&\mathbb{P}\Big[\max_{R\supseteq S, R\in\mathcal{D}} \mathsf{Z}_{a,S}(R) > \mathrm{cv}(\alpha, \underline{\phi}(a|S)) - c_1\frac{\log(\mathfrak{c}_{\supseteq S})}{\mathfrak{c}_\sigma} - \Big(\xi_1 + c_4\log^{\frac{1}{2}}(\mathfrak{c}_{\supseteq S}) + c_1\frac{\log(\mathfrak{c}_{\supseteq S})}{\mathfrak{c}_\sigma}\Big)\cdot\xi_2\Big] \\
&\leq \alpha + 4\Big(\xi_1 + c_4\log^{\frac{1}{2}}(\mathfrak{c}_{\supseteq S}) + c_1\frac{\log(\mathfrak{c}_{\supseteq S})}{\mathfrak{c}_\sigma}\Big)\xi_2\Big(\sqrt{2\log(\mathfrak{c}_{\supseteq S})} + 1\Big) + 4c_1\frac{\log(\mathfrak{c}_{\supseteq S})}{\mathfrak{c}_\sigma}\Big(\sqrt{2\log(\mathfrak{c}_{\supseteq S})} + 1\Big).
\end{aligned}$$

Collecting pieces, we have

$$\begin{aligned}
&\mathbb{P}\Big[\phi(a|S) < \underline{\widehat{\phi}}(a|S)\Big] \\
&\leq \alpha + 4\Big(\xi_1 + c_4\log^{\frac{1}{2}}(\mathfrak{c}_{\supseteq S}) + c_1\frac{\log(\mathfrak{c}_{\supseteq S})}{\mathfrak{c}_\sigma}\Big)\xi_2\Big(\sqrt{2\log(\mathfrak{c}_{\supseteq S})} + 1\Big) + 4c_1\frac{\log(\mathfrak{c}_{\supseteq S})}{\mathfrak{c}_\sigma}\Big(\sqrt{2\log(\mathfrak{c}_{\supseteq S})} + 1\Big) \\
&\quad + c_2\exp\Big\{-c_3\Big(\log\big(\min_{R\supseteq S, R\in\mathcal{D}} N_R\big) - 1\Big)\log(\mathfrak{c}_{\supseteq S})\Big\} + \exp\Big\{-\frac{\xi_1^2}{2}\Big\} + c_5\exp\Big\{-c_6\mathfrak{c}_\sigma^2\xi_2^2 + \log(\mathfrak{c}_{\supseteq S})\Big\}.
\end{aligned}$$

Therefore, by possibly enlarging the constants,

$$\begin{aligned}
&\mathbb{P}\Big[\phi(a|S) < \underline{\widehat{\phi}}(a|S)\Big] \\
&\leq \alpha + c_1\Big(\xi_1 + \log^{\frac{1}{2}}(\mathfrak{c}_{\supseteq S}) + \frac{\log(\mathfrak{c}_{\supseteq S})}{\mathfrak{c}_\sigma}\Big)\xi_2\log^{\frac{1}{2}}(\mathfrak{c}_{\supseteq S}) \\
&\quad + c_1\frac{\log^{3/2}(\mathfrak{c}_{\supseteq S})}{\mathfrak{c}_\sigma} \\
&\quad + c_2\exp\Big\{-c_3\Big(\log\big(\min_{R\supseteq S, R\in\mathcal{D}} N_R\big) - 1\Big)\log(\mathfrak{c}_{\supseteq S})\Big\} + \exp\Big\{-\frac{\xi_1^2}{2}\Big\} + c_5\exp\Big\{-c_6\mathfrak{c}_\sigma^2\xi_2^2 + \log(\mathfrak{c}_{\supseteq S})\Big\}.
\end{aligned}$$

Now set

$$\xi_1 \propto \log^{\frac{1}{2}}(\mathfrak{c}_\sigma), \quad \text{and} \quad \xi_2 \propto \frac{\log^{\frac{1}{2}}(\mathfrak{c}_{\supseteq S})\log^{\frac{1}{2}}(\mathfrak{c}_\sigma)}{\mathfrak{c}_\sigma},$$

The added strong-coupling term is dominated by the displayed rate because $\log(\mathfrak{c}_\sigma)$ diverges. Thus these choices give the desired bound

$$\mathbb{P}\Big[\phi(a|S) < \underline{\widehat{\phi}}(a|S)\Big] \leq \alpha + c\frac{\log^{\frac{3}{2}}(\mathfrak{c}_{\supseteq S})\log(\mathfrak{c}_\sigma)}{\mathfrak{c}_\sigma}$$

for some constant c.

SA-4.4.1 Proof of Lemma SA-11

Apply Lemma SA-1 to the collection

$$\mathcal{D}_S = \{R \in \mathcal{D} : R \supseteq S\}.$$

Conditional on the menu assignments, the samples indexed by distinct R are independent, so the menu-level couplings can be placed on one product probability space. Taking

$$t_R = C\log(N_R)\log(\mathfrak{c}_{\supseteq S})$$

and applying a union bound over the $\mathfrak{c}_{\supseteq S}$ menus yields exactly the probability and threshold displayed in the lemma, by the same calculation as in the proof of Lemma SA-2.

SA-4.4.2 Proof of Lemma SA-12

Let $\mathfrak{n}_R(a)$ denote the coordinate of $\mathfrak{n}_R$ corresponding to $\hat{\pi}(a|R)$ and set

$$\mathsf{Z}_{a,S}(R) = \frac{\mathfrak{n}_R(a)}{\sigma(a|R)}.$$

The variables $\mathsf{Z}_{a,S}(R)$ are independent standard normal variables because the menu-level Gaussian vectors are independent and $\mathbb{V}(\mathfrak{n}_R(a)) = \sigma^2(a|R)$. On the coupling event from the preceding lemma,

$$\max_{R\in\mathcal{D}_S}\left|\frac{\hat{\pi}(a|R) - \pi(a|R)}{\sigma(a|R)} - \mathsf{Z}_{a,S}(R)\right|$$

$$
\begin{aligned}
&\leq \max_{R\in\mathcal{D}_S} \frac{\log(N_R)}{N_R\sigma(a|R)} \max_{R\in\mathcal{D}_S} \frac{N_R}{\log(N_R)} \|\hat{\boldsymbol{\pi}}_R - \boldsymbol{\pi}_R - \mathfrak{n}_R\|_\infty \\
&\leq \frac{c_1 \log(\mathfrak{c}_{\supseteq S})}{\mathfrak{c}_\sigma},
\end{aligned}
$$

where $N_R\sigma(a|R) = \sqrt{N_R\pi(a|R)(1-\pi(a|R))}$. The probability bound is inherited from the preceding lemma.

SA-4.4.3 Proof of Lemma SA-13

For every $R \in \mathcal{D}_S$,

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

Bernstein's inequality and a union bound therefore give

$$
\mathbb{P}\left[\max_{R\in\mathcal{D}_S} \left|\frac{\hat{\sigma}(a|R)}{\sigma(a|R)} - 1\right| > \xi_2\right] \leq c_5 \exp\left\{-c_6\mathfrak{c}_\sigma^2\xi_2^2 + \log(\mathfrak{c}_{\supseteq S})\right\}.
$$

On the complementary event, for ξ_2 sufficiently small,

$$
\left|\frac{\hat{\pi}(a|R) - \pi(a|R)}{\sigma(a|R)} - \frac{\hat{\pi}(a|R) - \pi(a|R)}{\hat{\sigma}(a|R)}\right| \leq 2\xi_2 \left|\frac{\hat{\pi}(a|R) - \pi(a|R)}{\sigma(a|R)}\right|.
$$

Lemma SA-12 and the Gaussian maximal inequality imply, except on events with probabilities given by the first two terms in the lemma's bound,

$$
\max_{R\in\mathcal{D}_S} \left|\frac{\hat{\pi}(a|R) - \pi(a|R)}{\sigma(a|R)}\right| \leq \xi_1 + c_4 \log^{1/2}(\mathfrak{c}_{\supseteq S}) + c_1 \frac{\log(\mathfrak{c}_{\supseteq S})}{\mathfrak{c}_\sigma}.
$$

Combining the three events and absorbing the factor two into the universal constants proves the result.

SA-4.5 Proof of Theorem 9

The homogeneous-AOM procedures test moment inequalities associated with a candidate preference. H-LAO instead generally produces an identified set for a distribution of preferences. For the benchmark and dependence-robust models imposing Sequential Path Independence, $\mathcal{D}$ is suffix closed, as required by Proposition SA-4. Proposition SA-6 does not require this closure. We use projection: construct a simultaneous confidence region for the primitive choice rule, retain every latent structure compatible with some choice rule in that region, and then project onto the preference or attention object of interest. This approach remains valid when reach is weak because it never divides by an estimated reach probability.

Let p_0 be the true choice rule and let $\mathcal{C}_{1-\alpha}$ be any confidence region satisfying

$$\mathbb{P}[p_0 \in \mathcal{C}_{1-\alpha}] \geq 1 - \alpha.$$

For a candidate choice rule p, use the convention that $\mathcal{T}_{\mathcal{D}}(p)$ and $\mathcal{T}_{\mathcal{D}}^{\text{dep}}(p)$ are empty when recovered attention is invalid. Define

$$\begin{aligned}
\widehat{\mathcal{T}}_{1-\alpha}^{\text{ind}} &= \bigcup_{p \in \mathcal{C}_{1-\alpha}} \mathcal{T}_{\mathcal{D}}(p), \\
\widehat{\mathcal{T}}_{1-\alpha}^{\text{dep}} &= \bigcup_{p \in \mathcal{C}_{1-\alpha}} \mathcal{T}_{\mathcal{D}}^{\text{dep}}(p), \\
\widehat{\mathcal{T}}_{1-\alpha}^{\text{noPI}} &= \bigcup_{p \in \mathcal{C}_{1-\alpha}} \mathcal{T}_{\mathcal{D}}^{\text{noPI}}(p).
\end{aligned}$$

On the event $p_0 \in \mathcal{C}_{1-\alpha}$, sharp identification gives

$$\tau_0 \in \mathcal{T}_{\mathcal{D}}(p_0) \subseteq \widehat{\mathcal{T}}_{1-\alpha}^{\text{ind}}$$

under independence, and the same argument applies to the dependence-robust and no-SPI sharp sets. Because inclusion of τ_0 is one common event, every projection of τ_0 is covered simultaneously. Lemma SA-14 supplies the finite-sample probability bound for the stated choice of $\mathcal{C}_{1-\alpha}$.

The same construction gives confidence regions for attention. For every candidate p, the product formula above recovers all reach probabilities without division. Taking their union over $p \in \mathcal{C}_{1-\alpha}$ therefore covers the true attention frequencies. On portions of the parameter space with positive terminal reach, one may also recursively compute f for every candidate p and project the resulting collection. Near zero reach, retaining the undivided choice equations is preferable because f is then weakly or partially identified. ■

SA-4.6 Proof of Corollary 5

For a pair in which a precedes b, abbreviate

$$\begin{aligned}
\underline{y}_{ba} &= \underline{p}(b|\{a,b\}), & \overline{y}_{ba} &= \overline{p}(b|\{a,b\}), \\
\underline{r}_b(a,b) &= [1 - \overline{p}(o^*|\{a,b\})]\,[1 - \overline{p}(o^*|\{b\})], & \overline{r}_b(a,b) &= \left[1 - \underline{p}(o^*|\{a,b\})\right]\left[1 - \underline{p}(o^*|\{b\})\right].
\end{aligned}$$

These are the exact extrema of the terminal reach over the rectangular cell bands. Define

$$\widehat{C}^{\mathrm{H}}_{ba} = \left\{\theta \in [0,1] : [\underline{r}_b(a,b)\theta, \overline{r}_b(a,b)\theta] \cap [\underline{y}_{ba}, \overline{y}_{ba}] \neq \emptyset\right\}.$$

On the simultaneous event in Lemma SA-14, the true inside-choice probability belongs to $[\underline{y}_{ba}, \overline{y}_{ba}]$ and the true terminal reach belongs to $[\underline{r}_b(a,b), \overline{r}_b(a,b)]$ for every pair. The Pairwise Preference Shares corollary in the main paper gives the undivided moment

$$\pi(b|\{a,b\}) = r_b(a,b)\mathbb{P}_\tau(b \succ a),$$

including when $r_b(a,b) = 0$. Hence every true pairwise share belongs to its displayed confidence set on the same event. The interval formula follows because $[\underline{r}\theta, \overline{r}\theta]$ intersects $[\underline{y}, \overline{y}]$ if and only if $\overline{r}\theta \geq \underline{y}$ and $\underline{r}\theta \leq \overline{y}$, with the stated zero-reach conventions.

Corollary 5 of the main paper is an Anderson–Rubin-type inversion: it tests the undivided structural equation rather than first estimating a ratio. If reach is bounded away from zero, both reach endpoints are eventually positive and the interval contracts at the rate of the primitive probability bands. If reach is zero, the data correctly leave the pairwise preference share unrestricted. Simultaneous correlated-Gaussian or bootstrap bands for the primitive probabilities can replace the Hoeffding bands to improve power while preserving the same inversion logic. ∎

SA-4.7 Proof of Corollary 6

The finite-sample interval above obtains its uniform coverage from rectangular cell bands. Studentizing the same undivided structural equation provides a sharper complementary procedure under positive or weak but estimable reach, without introducing the unstable ratio estimator as the object being approximated.

Fix a pair in which a precedes b and abbreviate

$$y_{ba} = \pi(b|\{a,b\}), \qquad e_{ab} = \pi(o^*|\{a,b\}), \qquad e_b = \pi(o^*|\{b\}).$$

For a candidate pairwise share $\theta \in [0,1]$, define

$$g_{ba}(\theta) = y_{ba} - \theta(1 - e_{ab})(1 - e_b).$$

Under independent H-LAO, the true share $\theta_{ba} = \mathbb{P}_\tau(b \succ a)$ satisfies $g_{ba}(\theta_{ba}) = 0$, including when terminal reach is zero. Let $N_{ab} = N_{\{a,b\}}$ and $N_b = N_{\{b\}}$. Conditional independence of menu subsamples and the multinomial covariance between the b and outside-option cells in the binary menu give the delta-method variance

$$\begin{aligned} V_{ba}(\theta) = &\frac{1}{N_{ab}}\Big[y_{ba}(1-y_{ba}) + \theta^2(1-e_b)^2 e_{ab}(1-e_{ab}) - 2\theta(1-e_b)y_{ba}e_{ab}\Big] \\ &+ \frac{\theta^2(1-e_{ab})^2 e_b(1-e_b)}{N_b}. \end{aligned}$$

Let $\widehat{g}_{ba}(\theta)$ and $\widehat{V}_{ba}(\theta)$ replace the three probabilities by their empirical counterparts. For a finite target collection $\mathcal{B}$ of ordered pairs, set

$$c_{\mathcal{B}}(\alpha) = \Phi^{-1}\left(1 - \frac{\alpha}{2|\mathcal{B}|}\right)$$

and define

$$\widehat{C}^{\mathrm{S}}_{ba} = \left\{\theta \in [0,1] : \widehat{g}_{ba}(\theta)^2 \le c_{\mathcal{B}}(\alpha)^2 \widehat{V}_{ba}(\theta)\right\}.$$

If $\widehat{V}_{ba}(\theta) = 0$ for every $\theta \in [0,1]$, set $\widehat{C}^{\mathrm{S}}_{ba} = [0,1]$ by convention. Otherwise, the quadratic inequality includes any zero-variance candidate at which $\widehat{g}_{ba}(\theta) = 0$. This is the quadratic inversion used in the software.

For a generic choice W_{ab} from $\{a,b\}$ and W_b from $\{b\}$, define the centered influence terms

$$\begin{aligned} \psi_{ab}(W_{ab};\theta) &= \mathbb{1}\{W_{ab} = b\} - y_{ba} + \theta(1-e_b)\big(\mathbb{1}\{W_{ab} = o^*\} - e_{ab}\big), \\ \psi_b(W_b;\theta) &= \theta(1-e_{ab})\big(\mathbb{1}\{W_b = o^*\} - e_b\big). \end{aligned}$$

Their menu-sample averages have total variance $V_{ba}(\theta)$. The exact expansion around the population probabilities is

$$\begin{aligned} \widehat{g}_{ba}(\theta) - g_{ba}(\theta) = &\frac{1}{N_{ab}} \sum_{i:Y_i=\{a,b\}} \psi_{ab}(y_i;\theta) + \frac{1}{N_b}\sum_{i:Y_i=\{b\}} \psi_b(y_i;\theta) \\ &- \theta(\widehat{e}_{ab} - e_{ab})(\widehat{e}_b - e_b). \end{aligned}$$

For each $(b,a) \in \mathcal{B}$, Corollary 6 of the main paper assumes

$$N_{ab}, N_b \to \infty, \qquad V_{ba}(\theta_{ba}) > 0, \qquad \frac{\widehat{V}_{ba}(\theta_{ba})}{V_{ba}(\theta_{ba})} \to_{\mathbb{P}} 1,$$

as well as

$$\frac{\mathbb{E}|\psi_{ab}(W_{ab};\theta_{ba})|^3/N_{ab}^2+\mathbb{E}|\psi_b(W_b;\theta_{ba})|^3/N_b^2}{V_{ba}(\theta_{ba})^{3/2}}\longrightarrow 0,$$

$$\frac{\sqrt{e_{ab}(1-e_{ab})e_b(1-e_b)/(N_{ab}N_b)}}{\sqrt{V_{ba}(\theta_{ba})}}\longrightarrow 0.$$

At the true share, $g_{ba}(\theta_{ba}) = 0$. The first displayed condition is Lyapunov's condition for the two independent menu-sample sums in the exact expansion, so their sum divided by $\sqrt{V_{ba}(\theta_{ba})}$ converges to a standard normal distribution. The second condition and the independence of the two menu samples make the product remainder negligible relative to the same standard deviation. Variance consistency and Slutsky's theorem therefore imply

$$\frac{\widehat{g}_{ba}(\theta_{ba})}{\sqrt{\widehat{V}_{ba}(\theta_{ba})}}\rightsquigarrow \mathcal{N}(0,1)$$

for every pair. Variance consistency makes $\widehat{V}_{ba}(\theta_{ba})$ positive with probability tending to one; on this event, the quadratic inequality is equivalent to the studentized inequality. Hence the probability that a given true share is excluded is at most $\alpha/|\mathcal{B}|+o(1)$. The union bound over the fixed collection proves the simultaneous result. No division by terminal reach occurs.

The displayed variance follows from the gradient

$$(1,\ \theta(1-e_b),\ \theta(1-e_{ab}))$$

with respect to (y_{ba}, e_{ab}, e_b), together with

$$\mathbb{C}\mathrm{ov}\big(\widehat{y}_{ba},\widehat{e}_{ab}\big) = -\frac{y_{ba}e_{ab}}{N_{ab}}.$$

Primitive sufficient conditions include diverging effective counts in the relevant binary and singleton cells and the absence of asymptotic cancellation in $V_{ba}(\theta_{ba})$. These conditions allow reach to shrink, but they exclude a boundary at which every contribution to the moment variance vanishes.

The conditions of Corollary 6 of the main paper permit terminal reach to drift toward zero provided the undivided moment remains studentizable. At an exactly degenerate boundary, the finite-sample set $\widehat{C}_{ba}^{\mathrm{H}}$ remains valid. Bonferroni calibration avoids estimating cross-pair correlations and is the procedure covered by that corollary and used in the simulations. ∎

SA-4.8 Alternative Inference Procedures for Homogeneous AOM

This subsection develops valid alternatives to the feasible correlated-Gaussian procedure in the main paper. We first give a finite-sample projection method based on simultaneous confidence bands for the primitive choice probabilities. We then verify nondegeneracy for binary screening, develop a universal critical value that avoids covariance estimation, and explain how generalized moment selection and an alpha-split rule can be combined with these procedures. Throughout, we retain the notation and conditional sampling framework used above.

SA-4.8.1 Finite-Sample Projection from Choice-Probability Bands

The Gaussian procedures exploit covariance information and are therefore our preferred methods when power is important. A useful comparison-free benchmark follows from concentration of the menu-level empirical probabilities themselves. Let $\mathcal{Y}(S)$ denote the observed outcome set: $\mathcal{Y}(S) = S$ for the homogeneous-AOM data and $\mathcal{Y}(S) = S \cup \{o^*\}$ when the H-LAO outside option is observed. Thus, for H-LAO, the choice-data framework in the main paper is extended only by allowing the observed outcome to equal o^*. For $\alpha \in (0, 1)$, let

$$d_{\mathcal{D}} = \sum_{S \in \mathcal{D}} |\mathcal{Y}(S)|$$

be the number of menu–outcome cells. For $N_S \geq 1$, define

$$\varepsilon_S(\alpha) = \sqrt{\frac{\log(2d_{\mathcal{D}}/\alpha)}{2N_S}},$$

and, for every $a \in \mathcal{Y}(S)$,

$$\begin{aligned}\underline{p}(a|S) &= \max\{0, \widehat{\pi}(a|S) - \varepsilon_S(\alpha)\},\\ \overline{p}(a|S) &= \min\{1, \widehat{\pi}(a|S) + \varepsilon_S(\alpha)\}.\end{aligned}$$

Let $\mathcal{C}^{\mathrm{H}}_{1-\alpha}$ contain all candidate choice rules p satisfying

$$\begin{aligned}\underline{p}(a|S) \leq p(a|S) &\leq \overline{p}(a|S),\\ \sum_{a \in \mathcal{Y}(S)} p(a|S) &= 1\end{aligned}$$

for every observed menu and outcome. The superscript H refers to the Hoeffding construction, not to preference heterogeneity.

Lemma SA-14 (Simultaneous menu-probability region)**.** Under the choice-data sampling framework of the main paper, conditionally on the realized menu assignments,

$$\mathbb{P}\big[\pi \in \mathcal{C}^{\mathrm{H}}_{1-\alpha}\big] \geq 1-\alpha.$$

The result is finite sample, allows zero cell probabilities, and imposes no covariance nondegeneracy condition.

Proof. For each menu–outcome cell, Hoeffding's inequality gives

$$\mathbb{P}\left(|\widehat{\pi}(a|S) - \pi(a|S)| > \varepsilon_S(\alpha)\right) \leq 2\exp\{-2N_S\varepsilon_S^2(\alpha)\} = \frac{\alpha}{d_{\mathcal{D}}}.$$

A union bound over the $d_{\mathcal{D}}$ cells proves simultaneous coverage. The true probabilities satisfy the adding-up restrictions automatically. ■

For homogeneous AOM, define the finite-sample projection confidence set

$$\widehat{\mathcal{P}}^{\mathrm{H}}_{1-\alpha} = \left\{\succ : p \text{ satisfies } \succ\text{-Regularity for some } p \in \mathcal{C}^{\mathrm{H}}_{1-\alpha}\right\}.$$

Proposition SA-8 (Finite-sample inference for homogeneous AOM)**.** Let $\mathcal{P}_\pi$ be the sharp identified set of preferences under homogeneous AOM. Then

$$\mathbb{P}\left[\mathcal{P}_\pi \subseteq \widehat{\mathcal{P}}^{\mathrm{H}}_{1-\alpha}\right] \geq 1-\alpha.$$

For a fixed $\succ$, membership in $\widehat{\mathcal{P}}^{\mathrm{H}}_{1-\alpha}$ is a linear feasibility problem. Projecting the same feasible set through the lower and upper formulas in Theorem 2 of the main paper gives simultaneous finite-sample-valid outer intervals for every attention frequency associated with $\succ$.

Proof. On the event $\pi \in \mathcal{C}^{\mathrm{H}}_{1-\alpha}$, every $\succ \in \mathcal{P}_\pi$ is feasible in the definition of $\widehat{\mathcal{P}}^{\mathrm{H}}_{1-\alpha}$ by taking $p = \pi$. Lemma SA-14 gives the first claim. The simplex, band, and $\succ$-Regularity restrictions are linear in p. On the same simultaneous event, Theorem 2 of the main paper places every rationalizing attention frequency between its lower and upper formulas evaluated at the true π; optimizing those formulas over the feasible candidate probabilities therefore preserves coverage for all coordinates. ■

This construction provides a degeneracy-robust, finite-sample complement to the high-dimensional Gaussian method and avoids both Gaussian approximation and covariance comparison. Rectangular concentration bands prioritize coverage over power, making them especially useful as a finite-sample benchmark and as the common primitive confidence region for H-LAO inference below.

SA-4.8.2 Feasible Correlated-Gaussian Critical Values

The default procedure in the main paper uses the feasible correlated-Gaussian critical value $\widehat{\mathrm{cv}}(\alpha, \succ)$. This critical value simulates the distribution of

$$\widehat{\mathsf{T}}^{\mathsf{G}}(\succ) = \max\left\{0, \max_{1\le \ell \le \mathfrak{c}_{\succ}} \hat{Z}_\ell\right\},$$

where $\hat{\mathbf{Z}}|\text{Data} \sim \mathcal{N}(0, \hat{\mathbf{\Omega}})$. Its advantage is power: it uses the estimated dependence structure among the inequality comparisons, and hence can be much less conservative than a universal bound based only on the number of inequalities.

The proof of Theorem 7 passes from $\mathbf{Z} \sim \mathcal{N}(0, \mathbf{\Omega})$ to $\hat{\mathbf{Z}} \sim \mathcal{N}(0, \hat{\mathbf{\Omega}})$ using Lemma SA-7. Its fully general treatment accommodates singular or nearly singular covariance matrices and produces the extra square-root term in the final size-distortion bound. The refinements below sharpen that rate under nondegeneracy or avoid the comparison step altogether.

SA-4.8.3 Binary-Comparison Nondegeneracy

The binary-comparison statistic has a special covariance structure. Fix two alternatives a and b, and consider the comparisons

$$D(a|S, b) = \pi(a|S) - \pi(a|b), \qquad S \in \mathcal{D}_{ab}.$$

The only shared sampling component across different S is the binary comparison $\widehat{\pi}(a|b)$. This produces a diagonal-plus-rank-one correlation matrix.

Lemma SA-15 (Binary-comparison covariance)**.** Let $\mathbf{\Omega}_{ab}$ be the correlation matrix of

$$\left\{\frac{\widehat{D}(a|S, b) - D(a|S, b)}{\sigma(a|S, b)} :\ S \in \mathcal{D}_{ab}\right\}.$$

Then

$$\boldsymbol{\Omega}_{ab} = \operatorname{diag}\left(\frac{\sigma^2(a|S)}{\sigma^2(a|S,b)}\right)_{S\in\mathcal{D}_{ab}} + vv', \qquad v_S = \frac{\sigma(a|b)}{\sigma(a|S,b)}.$$

Consequently,

$$\lambda_{\min}(\boldsymbol{\Omega}_{ab}) \geq \min_{S\in\mathcal{D}_{ab}} \frac{\sigma^2(a|S)}{\sigma^2(a|S,b)}.$$

Proof. Conditional on the menu sample sizes, the estimators from different choice problems are independent. Let Z_S and Z_b be mean-zero Gaussian random variables with variances $\sigma^2(a|S)$ and $\sigma^2(a|b)$, respectively, with Z_b independent of all Z_S and the Z_S mutually independent across S. The oracle Gaussian fluctuation for the comparison $D(a|S,b)$ can be represented as $Z_S - Z_b$.

For $S \neq T$,

$$\mathbb{C}\mathrm{ov}(Z_S - Z_b, Z_T - Z_b) = \sigma^2(a|b),$$

while

$$\mathbb{V}(Z_S - Z_b) = \sigma^2(a|S) + \sigma^2(a|b) = \sigma^2(a|S,b).$$

After normalizing by $\sigma(a|S,b)$, the stated diagonal-plus-rank-one representation follows. Since vv' is positive semidefinite, Weyl's inequality gives the lower bound on the minimum eigenvalue. ∎

Corollary SA-2 (Balanced standard errors imply nondegeneracy)**.** If there exists a constant $C < \infty$ such that

$$C^{-1} \leq \frac{\sigma(a|S)}{\sigma(a|b)} \leq C \qquad \text{for all } S \in \mathcal{D}_{ab},$$

then

$$\lambda_{\min}(\boldsymbol{\Omega}_{ab}) \geq \frac{1}{1+C^2}.$$

Proof. The ratio

$$\frac{\sigma^2(a|S)}{\sigma^2(a|S,b)} = \frac{\sigma^2(a|S)}{\sigma^2(a|S) + \sigma^2(a|b)}$$

is bounded below by $1/(1+C^2)$ under the stated balance condition. The result follows from Lemma SA-15. ■

This result verifies the key nondegeneracy condition needed for sharper Gaussian comparison in binary screening. Under the balance condition, one may replace Lemma SA-7 with Theorem 2.3 of Lopes (2022), whose covariance-error dependence is nearly linear. The sharper fixed-pair bound in Corollary 4 makes this improvement formal. It does not assert the same eigenvalue bound for a joint statistic stacking multiple ordered pairs.

SA-4.8.4 Universal Critical Values Without Gaussian Comparison

A comparison-free alternative avoids estimating the correlation matrix altogether. Define

$$c_B(\alpha, \mathfrak{c}_\succ) = \Phi^{-1}(1-\alpha/\mathfrak{c}_\succ).$$

The universal test rejects H_0 when

$$\mathsf{T}(\succ) > c_B(\alpha, \mathfrak{c}_\succ).$$

The critical value depends only on the number of comparisons. It is typically more conservative than $\widehat{\mathrm{cv}}(\alpha, \succ)$, but it does not require the feasible Gaussian vector $\hat{\mathbf{Z}}$ or Lemma SA-7.

Lemma SA-16 (Universal critical values)**.** Fix $\alpha \in (0, 1/2)$ and a preference $\succ$ with $\mathfrak{c}_\succ \geq 1$. Suppose that, under H_0,

$$\mathbb{P}\left[\|\mathbf{R}_\succ(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}) \oslash \hat{\boldsymbol{\sigma}} - \mathbf{Z}\|_\infty > \eta\right] \leq \rho,$$

where $\mathbf{Z}$ is the oracle Gaussian vector with unit variances. Then

$$\mathbb{P}\left[\mathsf{T}(\succ) > c_B(\alpha, \mathfrak{c}_\succ)\right] \leq \alpha + \rho + 4\eta\left(\sqrt{2\log(\mathfrak{c}_\succ)} + 1\right).$$

Proof. Under H_0, the population component $\mathbf{R}_\succ \boldsymbol{\pi} \oslash \hat{\boldsymbol{\sigma}}$ is weakly negative coordinatewise. Therefore, on the event

$$\|\mathbf{R}_\succ(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}) \oslash \hat{\boldsymbol{\sigma}} - \mathbf{Z}\|_\infty \leq \eta,$$

the rejection event implies

$$\max_{1\le \ell \le \mathfrak{c}_{\succ}} Z_\ell > c_B(\alpha, \mathfrak{c}_{\succ}) - \eta.$$

Thus,

$$\mathbb{P}\left[\mathsf{T}(\succ) > c_B(\alpha, \mathfrak{c}_{\succ})\right] \le \mathbb{P}\left[\max_{1\le \ell \le \mathfrak{c}_{\succ}} Z_\ell > c_B(\alpha, \mathfrak{c}_{\succ}) - \eta\right] + \rho.$$

The first term is bounded by

$$\mathbb{P}\left[\max_{1\le \ell \le \mathfrak{c}_{\succ}} Z_\ell > c_B(\alpha, \mathfrak{c}_{\succ})\right] + \mathbb{P}\left[\left|\max_{1\le \ell \le \mathfrak{c}_{\succ}} Z_\ell - c_B(\alpha, \mathfrak{c}_{\succ})\right| \le \eta\right].$$

By the union bound,

$$\mathbb{P}\left[\max_{1\le \ell \le \mathfrak{c}_{\succ}} Z_\ell > c_B(\alpha, \mathfrak{c}_{\succ})\right] \le \mathfrak{c}_{\succ}\{1 - \Phi(c_B(\alpha, \mathfrak{c}_{\succ}))\} = \alpha.$$

The anti-concentration bound in Lemma SA-10 controls the second term and gives the stated result. ∎

Taking η and ρ from the strong approximation proof of Theorem 6 yields a valid test whose excess size is governed by the strong-coupling error, up to the anti-concentration factor. The improvement is conceptual and theoretical: the extra square-root loss from estimated-covariance Gaussian comparison is optional. The tradeoff is power, because $c_B(\alpha, \mathfrak{c}_{\succ})$ protects against the worst case using only the number of inequalities.

SA-4.8.5 Generalized Moment Selection

The statistic $\mathsf{T}(\succ)$ can involve many inequalities that are far inside the null. These slack inequalities increase the dimension of the max statistic and can make the critical value conservative. Generalized moment selection addresses this by using the data to focus the critical-value calculation on inequalities that are close to binding.

The follow-up simulation pipeline records two critical values from the same Gaussian draws. The first is the zero-centered least-favorable critical value, labeled `LF` in the software, which retains every inequality and directly implements Theorem 7 of the main paper. The second is the generalized moment-selection (GMS) refinement in equation (4.4) of Andrews and Soares (2010). If N is the total sample size, the implementation sets `MNRatioGMS`$= 1/\log(N)$ and multiplies each estimated moment entering the Gaussian recentering by $\sqrt{\texttt{MNRatioGMS}} = 1/\sqrt{\log(N)}$, truncating positive recentering terms at zero. Thus, binding and local-to-binding inequalities remain protected while deeply slack inequalities receive a negative recentering.

Moment selection can reduce the effective contribution of slack inequalities to the max statistic and thereby improve power. It does not remove the Gaussian-comparison step. The all-inequalities `LF` procedure is the formal Theorem 7 baseline; GMS is reported as an additional numerical refinement whose validity relies on the safeguards in Andrews and Soares (2010).

SA-4.8.6 Hybrid Critical Values

The feasible correlated-Gaussian and universal critical values can also be combined by splitting the nominal level.

Lemma SA-17 (Alpha-split hybrid critical values)**.** Suppose two critical values c_1 and c_2 satisfy, under H_0,

$$\mathbb{P}[\mathsf{T}(\succ) > c_k] \leq \alpha_k + \delta_k, \qquad k = 1, 2.$$

Then

$$\mathbb{P}[\mathsf{T}(\succ) > \min\{c_1, c_2\}] \leq \alpha_1 + \alpha_2 + \delta_1 + \delta_2.$$

Proof. The event $\{\mathsf{T}(\succ) > \min\{c_1, c_2\}\}$ is contained in

$$\{\mathsf{T}(\succ) > c_1\} \cup \{\mathsf{T}(\succ) > c_2\}.$$

The result follows from the union bound. ∎

For example, take $c_1 = \widehat{\mathrm{cv}}(\alpha_G, \succ)$ and $c_2 = c_B(\alpha_B, \mathfrak{c}_\succ)$ with $\alpha_G + \alpha_B = \alpha$. The hybrid rule rejects when

$$\mathsf{T}(\succ) > \min\{\widehat{\mathrm{cv}}(\alpha_G, \succ), c_B(\alpha_B, \mathfrak{c}_\succ)\}.$$

This rule is valid up to the sum of the two approximation errors. It allows the correlated-Gaussian test to carry most of the power while preserving a comparison-free route to rejection.

SA-4.9 Alternative Inference Procedures for Heterogeneous H-LAO

This subsection develops additional H-LAO inference procedures beyond those stated in the main paper. It first gives finite-sample linear-programming procedures under preference–stopping dependence and without Sequential Path Independence, and then discusses how projection and Gaussian procedures can be combined.

The exact independent projection preserves Sequential Path Independence and preference–stopping independence through polynomial joint feasibility equations. The dependence-robust projection is likewise nonlinear when the exact product formula for reach is imposed while p varies. The linear outer construction below delivers transparent computation and finite-sample coverage; its relaxation disappears as the primitive probability bands contract.

SA-4.9.1 A Dependence-Robust Linear-Programming Procedure

For larger preference events, an LP implementation is available after forming simultaneous outer bands for the recovered reach probabilities. For $S = \{a_{s_1}, \ldots, a_{s_{|S|}}\}$, write $S_{k:} = \{a_{s_k}, \ldots, a_{s_{|S|}}\}$ and define

$$\underline{\varphi}_j(S) = \prod_{k=1}^{j} \left[1 - \overline{p}(o^*|S_{k:})\right],$$
$$\overline{\varphi}_j(S) = \prod_{k=1}^{j} \left[1 - \underline{p}(o^*|S_{k:})\right].$$

On the event in Lemma SA-14, every true recovered reach probability lies in the corresponding interval.

Consider variables $v_j(S)$, $m_j(S)$, $p(a|S)$, $\tau(\succ)$, and $q_S(\succ, j)$. Let $\widehat{\mathcal{Q}}^{\mathrm{LP}}_{1-\alpha}$ be defined by

$$\begin{aligned}
&\underline{\varphi}_j(S) \le v_j(S) \le \overline{\varphi}_j(S), \qquad v_{|S|+1}(S) = 0,\\
&m_0(S) = 1 - v_1(S), \qquad m_j(S) = v_j(S) - v_{j+1}(S) \ge 0,\\
&v_j(S) \ge v_k(R) \quad \text{whenever } a_{s_j} = a_{r_k} \in S \subseteq R,\\
&q_S(\succ, j) \ge 0, \qquad \sum_{\succ \in \mathcal{P}} q_S(\succ, j) = m_j(S),\\
&\sum_{j=0}^{|S|} q_S(\succ, j) = \tau(\succ), \qquad \sum_{\succ \in \mathcal{P}} \tau(\succ) = 1,\\
&\underline{p}(a|S) \le p(a|S) \le \overline{p}(a|S),\\
&p(a|S) = \sum_{j=1}^{|S|} \sum_{\succ \in \mathcal{P}} \mathbb{1}\{a \in T_j(S),\ a \text{ is } \succ\text{-best in } T_j(S)\} q_S(\succ, j)
\end{aligned}$$

for all relevant menus, positions, preferences, and inside alternatives. The third line imposes attention overload, matching the same alternative across the two inherited list positions.

Proposition SA-9 (Finite-sample LP inference for dependence-robust H-LAO)**.** Suppose the dependence-robust H-LAO model in Proposition SA-5 holds with preference distribution τ_0. Then, with probability at least $1 - \alpha$, the true latent structure is feasible in $\widehat{\mathcal{Q}}^{\mathrm{LP}}_{1-\alpha}$. Consequently, for every collection of preference events A, the intervals obtained by minimizing and maximizing

$$\sum_{\succ \in A} \tau(\succ)$$

over $\widehat{\mathcal{Q}}^{\mathrm{LP}}_{1-\alpha}$ cover all corresponding true event probabilities simultaneously. Each endpoint is the solution to a linear program. The conclusion also covers the independent H-LAO model, viewed as a subclass of the dependence-robust model.

Proof. On the simultaneous probability-band event, set $v_j(S)$ equal to the true recovered reach, $m_j(S)$ equal to the true prefix-stopping probability, $p = \pi$, and (τ, q) equal to the true marginal preference distribution and menu-specific preference–stopping couplings. The product bounds contain every $v_j(S)$. Attention validity gives the monotonicity and attention-overload constraints, while Proposition SA-5 gives the remaining marginal and choice equations. Thus the true structure is feasible. Lemma SA-14 bounds the probability of this common event, and linear projection gives simultaneous coverage of all event probabilities. ■

The LP uses interval restrictions on reach rather than imposing the estimated nonlinear product equations exactly. It is therefore a finite-sample outer procedure. As every N_S diverges, the primitive bands and the induced reach bands contract, so this relaxation disappears. Under benchmark independence, adding $q_S(\succ, j) = \tau(\succ)m_j(S)$ recovers the exact restriction but makes the program bilinear. The explicit pairwise procedure above avoids that difficulty for the most empirically useful preference margins; exact projection or global polynomial optimization remains available for more general independent-H-LAO events.

SA-4.9.2 A Linear Projection Procedure without Sequential Path Independence

The robustness envelope in Proposition SA-6 permits a simpler exact projection over the primitive probability region because its prefix masses are latent and need not satisfy nonlinear product equalities. On any observed-menu domain $\mathcal{D}$, let $\widehat{\mathcal{Q}}^{\mathrm{noPI}}_{1-\alpha}$ contain variables $p(a|S)$ for $a \in S \cup \{o^*\}$, $m_j(S)$, $q_S(\succ, j)$, and $\tau(\succ)$ satisfying

$$\underline{p}(a|S) \le p(a|S) \le \overline{p}(a|S), \qquad \sum_{a \in S \cup \{o^*\}} p(a|S) = 1,$$

$$m_j(S) \ge 0, \qquad m_0(S) = p(o^*|S), \qquad \sum_{j=0}^{|S|} m_j(S) = 1,$$

$$q_S(\succ, j) \ge 0, \qquad \sum_{\succ \in \mathcal{P}} q_S(\succ, j) = m_j(S), \qquad \sum_{j=0}^{|S|} q_S(\succ, j) = \tau(\succ),$$

$$p(a|S) = \sum_{j=1}^{|S|} \sum_{\succ \in \mathcal{P}} \mathbb{1}\{a \in T_j(S),\ a \text{ is } \succ\text{-best in } T_j(S)\} q_S(\succ, j),$$

for all relevant indices, together with

$$\sum_{\ell=j}^{|S|} m_\ell(S) \geq \sum_{\ell=k}^{|R|} m_\ell(R) \qquad \text{whenever } a_{s_j} = a_{r_k} \in S \subseteq R.$$

Proposition SA-10 (Finite-sample LP inference without Sequential Path Independence)**.** Suppose the model in Proposition SA-6 holds with marginal preference distribution τ_0. Then, with probability at least $1-\alpha$, its true choice rule and latent structure are feasible in $\widehat{\mathcal{Q}}^{\text{noPI}}_{1-\alpha}$. Consequently, minimizing and maximizing any collection of preference-event probabilities over this set gives simultaneous finite-sample confidence bounds. Every endpoint is a linear program, no suffix-closure or positive-reach condition is required, and the procedure exactly projects the rectangular cell region with menu-specific adding-up onto the path-independence-robust identified set.

Proof. On the simultaneous-band event, set $p = \pi$ and set (τ, q, m) equal to the true stable preference marginal, menu-specific preference–stopping couplings, and prefix masses. The primitive bands and adding-up restrictions contain p. Proposition SA-6 gives all remaining equalities and inequalities, so the true structure is feasible. The event has probability at least $1 - \alpha$ by Lemma SA-14. Because the displayed constraints are exactly those defining $\mathcal{T}^{\text{noPI}}_{\mathcal{D}}(p)$ for some p in the primitive region, optimizing a linear event probability performs the stated projection without further relaxation. ■

SA-4.9.3 Relation to the Gaussian Procedures

We implement a hybrid primitive-probability region as follows. Let $K = d_{\mathcal{D}}$ be the number of observed menu–outcome cells and set the screening threshold to $b = 5$. For a row-i.i.d. cell (a, S), write $X_{aS} = N_S \widehat{\pi}(a|S)$ and let $\widehat{s}_{aS}$ denote its estimated standard error. The cell is regular when $X_{aS} \geq b$, $N_S - X_{aS} \geq b$, and $\widehat{s}_{aS} > \sqrt{\varepsilon_{\text{mach}}}$; it is boundary or degenerate when $\widehat{s}_{aS} \leq \sqrt{\varepsilon_{\text{mach}}}$, and sparse otherwise. Let $\widehat{\mathcal{R}}$ and $\widehat{\mathcal{F}}$ collect the regular and fallback cells. If both sets are nonempty, assign $\alpha_{\text{G}} = \alpha_{\text{F}} = \alpha/2$; if only one is nonempty, it receives the full error budget α.

For $k \in \widehat{\mathcal{R}}$, let $\widehat{c}_{\text{G}}$ be the conditional $(1 - \alpha_{\text{G}})$ quantile of

$$\max_{k \in \widehat{\mathcal{R}}} \frac{|Z_k|}{\widehat{s}_k}, \qquad Z \mid \text{data} \sim \mathcal{N}(0, \widehat{V}),$$

where $\widehat{V}$ is the plug-in block-multinomial covariance matrix. The regular-cell interval is $[\widehat{\pi}_k - \widehat{c}_{\mathrm{G}}\widehat{s}_k, \widehat{\pi}_k + \widehat{c}_{\mathrm{G}}\widehat{s}_k]$, truncated to $[0,1]$. For $k = (a,S) \in \widehat{\mathcal{F}}$, put $\eta = \alpha_{\mathrm{F}}/K$. Write the two-sided Clopper–Pearson interval as $I^{\mathrm{CP}}_{aS} = [L_{aS}, U_{aS}]$, where its lower endpoint is

$$L_{aS} = \begin{cases} 0, & X_{aS} = 0, \\ B^{-1}_{X_{aS},\, N_S - X_{aS}+1}(\eta/2), & X_{aS} > 0, \end{cases}$$

and its upper endpoint is

$$U_{aS} = \begin{cases} 1, & X_{aS} = N_S, \\ B^{-1}_{X_{aS}+1,\, N_S - X_{aS}}(1-\eta/2), & X_{aS} < N_S, \end{cases}$$

where $B^{-1}_{u,v}$ is the quantile function of a $\mathrm{Beta}(u,v)$ variable. The hybrid region contains every vector whose cells lie in their assigned intervals and whose probabilities add to one within each menu.

The fallback component has simultaneous finite-sample coverage at least $1 - \alpha_{\mathrm{F}}$ by a union bound. The Gaussian component, and hence a hybrid region containing regular cells, is an asymptotic procedure whose coverage is subject to Gaussian-approximation, covariance-estimation, and simulation errors. If no Gaussian cell is used, the exact-binomial region is finite-sample valid. This coverage distinction concerns the primitive region; projection through the model-feasible sets preserves the coverage status of the region supplied to it.

The finite-sample bands emphasize robustness, whereas the correlated-Gaussian machinery in the main paper emphasizes power in high-dimensional problems. They can be combined. Under terminal reach bounded away from zero, the pairwise ratio is a smooth function of $\pi(b|\{a,b\})$, $\pi(o^*|\{a,b\})$, and $\pi(o^*|\{b\})$. Under weaker reach, Corollary 6 of the main paper applies the Gaussian approximation to the undivided moment instead. In both cases, simultaneous Gaussian or multiplier-bootstrap inference can use the dependence across pairs induced by shared menus. At degenerate zero reach, finite-sample Hoeffding or exact-binomial projection remains valid. Thus, the support condition determines the most informative implementation, not the validity of the underlying identified-set analysis.

The studentized moments also suggest a covariance-aware Gaussian-max calibration: stack the moments, estimate their candidate-dependent correlation matrix from the menu-level multinomial covariance blocks, and simulate the corresponding quantile. A fully uniform theory additionally entails joint approximation, correlation consistency, quantile continuity, simulation-error control, and uniform inversion over the candidate set. Our formal result and simulations therefore use the theorem-backed Bonferroni calibration, while the projection method already exploits covariance-aware Gaussian bands for the primitive probabilities.

SA-4.10 Cluster-Robust Inference for Repeated Choices

The preceding procedures extend to repeated-choice data in which observations are independent across clusters but may be arbitrarily dependent within a cluster. Let $g = 1, \ldots, G$ index clusters, let n_{gS} be cluster g's number of observations from menu S, and let $N_S = \sum_g n_{gS}$. For observation i, write S_i for the menu and y_i for the chosen alternative. For every observed menu–outcome cell (a, S), define the cluster influence contribution

$$\widehat{\psi}_{g,aS} = \frac{1}{N_S} \sum_{i: g_i = g,\, S_i = S} \{\mathbb{1}(y_i = a) - \widehat{\pi}(a|S)\}\,.$$

After stacking all cells, the cluster covariance estimator and its multiplier analogue are

$$\widehat{V}_{\mathrm{cl}} = \frac{G}{G-1} \sum_{g=1}^{G} \widehat{\psi}_g \widehat{\psi}_g', \qquad \widehat{Z}^* = \sqrt{\frac{G}{G-1}} \sum_{g=1}^{G} \xi_g \widehat{\psi}_g, \qquad \xi_g \overset{\text{iid}}{\sim} \mathcal{N}(0,1).$$

Under the cluster-level analogues of the moment, covariance-consistency, and Gaussian-approximation conditions used above, the common multiplier maximum yields asymptotically valid simultaneous primitive-probability bands as the number of independent clusters grows. The same draws can instead be applied directly to the AOM inequalities and regular H-LAO diagnostic moments. Additional subject-level statistics, such as elicited preference reports, can be stacked with the choice cells when constructing confidence intervals for differences. Thus one joint cluster process supports the AOM ranking tests, H-LAO specification diagnostics, pairwise preference intervals, and projection or linear-programming bounds.

For clustered data, let $X_{g,aS}$ be cluster g's number of choices of a from S, and define

$$G_{aS}^{+} = \sum_{g=1}^{G} \mathbb{1}\{X_{g,aS} > 0\}, \qquad G_{aS}^{-} = \sum_{g=1}^{G} \mathbb{1}\{X_{g,aS} < n_{gS}\}.$$

A cell is regular when $G_{aS}^{+} \geq b$, $G_{aS}^{-} \geq b$, and its cluster standard error exceeds $\sqrt{\varepsilon_{\mathrm{mach}}}$; it is boundary or degenerate when that standard error does not exceed the cutoff, and sparse otherwise. For regular cells, the multiplier analogue $\widehat{Z}^*$ replaces Z in the Gaussian maximum above. Fallback cells use the cluster-Hoeffding half-width

$$\left\{ \frac{\log(2K/\alpha_{\mathrm{F}})}{2} \sum_{g=1}^{G} \left(\frac{n_{gS}}{N_S} \right)^2 \right\}^{1/2}.$$

This bound treats clusters as independent bounded contributions and remains valid under unrestricted dependence among observations within a cluster. The same error-budget rule applies: multiplier and cluster-Hoeffding cells each receive $\alpha/2$ when both occur, and the sole component receives α otherwise. Accordingly, the cluster-Hoeffding component is finite-sample valid, while a region containing multiplier cells has the asymptotic coverage status of the multiplier approximation. The resulting joint primitive region is propagated through the same AOM and H-LAO transformations and optimization problems as in the independent-row analysis.

SA-5 Implementation and Computational Scalability

This section collects the implementation details that are common across empirical designs. The numerical benchmarks vary the size of the universe and menu domain while holding the population choice system fixed and treating it as observed. They therefore measure the cost of constructing and solving the paper's optimization problems, rather than statistical performance under a particular simulation design. Section SA-6 separately reports finite-sample behavior and execution times for the Monte Carlo designs used in the paper.

SA-5.1 Homogeneous AOM

For a fixed candidate preference, the econometric procedures in Section 4 of the main paper construct and evaluate the $\mathfrak{c}_{\succ}$ observed $\succ$-Regularity inequalities. For each observed pair $T \subset S$ with $|T| = \ell \geq 2$, the comparison associated with the $\succ$-worst alternative in T reduces to $\pi(a|S) \leq 1$ and is omitted; singleton T contributes no nonredundant comparison. Hence, with complete menu data,

$$\mathfrak{c}_{\succ} = \sum_{k=3}^{|X|} \binom{|X|}{k} \sum_{\ell=2}^{k-1} \binom{k}{\ell} (\ell - 1).$$

For $|X| = 6$, this formula gives 664 comparisons, whereas retaining every $a \in T \subset S$ gives 1,266 regularity rows. The subset relations depend only on the menu domain and should be constructed once, stored sparsely, and reused across candidate preferences, bootstrap or Gaussian draws, and specification tests. On incomplete domains, only observed pairs $T \subset S$ enter, so the relevant count can be much smaller than the complete-data expression.

When the objective is to recover the entire compatible-preference set, Corollary 2 of the main paper replaces explicit enumeration of $|X|!$ rankings with a mixed-integer linear program. Its directed variables z_{ba} indicate $b \succ a$. Pairwise totality and transitivity make z a strict linear order, and the remaining

Table SA-2: Deterministic computational benchmark for homogeneous AOM

$\lvert X\rvert$	$\lvert\mathcal{D}\rvert$	$\mathfrak{c}_{\succ}$	$\lvert X\rvert!$	Binary	Constraints	Build	Feasibility	All pairs
3	7	3	6	6	24	< 0.01	< 0.01	< 0.01
4	15	26	24	12	106	< 0.01	< 0.01	0.01
5	31	145	120	20	395	0.02	0.02	0.03
6	63	664	720	30	1,401	0.09	0.09	0.13
7	127	2,723	5,040	42	4,886	0.43	0.43	0.60

Notes: Times are median elapsed seconds over 3 runs. The column $\mathfrak{c}_{\succ}$ counts the nonredundant fixed-preference inequalities used for inference. Constraints counts the direct mixed-integer rows retained by the software, including totality, transitivity, and algebraically redundant regularity rows. Build is the time spent validating the population and constructing this system in the feasibility-only call. Feasibility includes system construction and one solver call. All pairs includes system construction, the feasibility solve, and one reverse-comparison query for every unordered pair. The deterministic choice system is generated by full attention and the ordering $1 \succ 2 \succ \cdots \succ |X|$ on the complete menu domain.

constraints impose $\succ$-Regularity. The formulation uses $|X|(|X|-1)$ directed binary variables, of which $\binom{|X|}{2}$ are independent after totality, cubically many ordering constraints, and one row for each observed regularity comparison. Because the ranking is an optimization variable, the direct mixed-integer formulation retains a row for every $a \in T \subset S$. Rows that become vacuous under the selected ranking do not change the feasible set, but retaining them avoids having to determine the worst alternative before solving. This is why its total constraint count exceeds $\mathfrak{c}_{\succ}$. Binary and nonbinary revealed-preference screens can fix coordinates before optimization. Model compatibility requires one feasibility solve; for each unresolved pair, a solve imposing the reverse comparison determines whether that comparison occurs in any compatible preference. Thus the number of mixed-integer queries grows quadratically rather than factorially in $|X|$.

The implementation can be tightened further without changing the identified set. One direction of each totality equality can be eliminated, and transitivity rows can be added lazily whenever an integer solution contains a directed cycle. The number of nested-menu inequalities remains the main complete-data bottleneck, so screening and sparse observed domains are complementary to the mixed-integer representation rather than substitutes for it. Solver status is part of the reported result: status-checked infeasibility establishes incompatibility, while an unresolved numerical status is not interpreted as rejection of the model.

The accompanying replication benchmark uses the public `aomIdentify` interface on deterministic, model-compatible choice systems. For each universe size it records the number of observed menus, $\mathfrak{c}_{\succ}$, the factorial ranking count, binary variables, constraints, mixed-integer solves, construction time, solve time, and total elapsed time. Separate feasibility-only and all-pairs calls isolate the cost of the initial model check from the additional revealed-preference queries. This design replaces machine-specific timing anecdotes with output that can be regenerated under the reported package version, solver, hardware, and numerical tolerance.

Table SA-2 shows the combinatorial growth of the complete-data restriction system without a factorial search over rankings. From three to seven alternatives, $\mathfrak{c}_{\succ}$ rises from 3 to 2,723 and the mixed-integer system from 24 to 4,886 rows. On the reported machine, constructing and solving the seven-alternative feasibility problem takes about half a second, and adding all $\binom{7}{2}$ reverse-comparison queries raises the total to less than one second. Because performance depends on instance difficulty and the observed domain, we report the package version, solver, hardware, tolerance, benchmark design, and scaling quantities needed to reproduce and interpret these timings.

SA-5.2 Heterogeneous H-LAO

H-LAO has a target-specific computational hierarchy. Pairwise preference shares use Proposition 5 of the main paper and require only singleton and binary menus. Alternative-specific rank probabilities use Block–Marschak sums once f is recovered, and full-attention agreement is computed directly as $\sum_{a\in S}\phi(a|S)f(a|S)$. None of these calculations enumerates rankings. General joint preference events and sharp incomplete-domain bounds use the preference polytope in Proposition SA-4. Under benchmark independence, its direct formulation has $|X|!$ ranking-probability variables, one adding-up equation, and at most $\sum_{S\in\mathcal{D}}|S|$ choice equations. The dependence-robust and no-SPI formulations are larger because they add menu-specific preference–stopping variables or latent prefix masses.

For a direct benchmark solve, one column is constructed for each ranking. Its entries equal one in the adding-up row and, in each observed choice row, the recovered probability of the prefixes on which that ranking selects the corresponding alternative. Columns with identical choice-equation entries and the same target value are observationally equivalent and can be collapsed before optimization. This compression can be substantial on sparse menu domains even though the formal ranking count remains $|X|!$.

When the ranking matrix is too large to materialize, the benchmark-independent event LP admits column generation. A Phase-I restricted master with artificial variables first produces a feasible collection of rankings. The master dual assigns a shadow value to the adding-up and observed-choice equations. Mixed-integer pricing then minimizes the reduced cost over strict linear orders, using pairwise-order and prefix-winner variables with transitivity constraints. A negatively priced ranking is added to the master; otherwise the current endpoint is optimal up to the reported tolerance. The same procedure with the sign of the target reversed yields the other endpoint. Generated rankings are reused across endpoints and related events.

Column generation avoids storing irrelevant ranking columns but does not remove worst-case factorial

Table SA-3: Deterministic computational benchmark for H-LAO

$\lvert X\rvert$	$\lvert\mathcal{D}\rvert$	Choice eq.	$\lvert X\rvert!$	Enumerated	Active	Pricing binary	Enumeration	Column generation	Certified
3	3	6	6	6	5	16	0.01	0.01	Yes
4	4	10	24	24	9	32	0.01	0.02	Yes
5	5	15	120	120	15	55	0.01	0.05	Yes
6	6	21	720	720	18	86	0.03	0.11	Yes
7	7	28	5,040	5,040	28	126	0.14	0.35	Yes

Notes: Times are median elapsed seconds over 3 runs. The deterministic suffix-domain system has outside-option probability 0.2 and selects the first listed alternative with probability 0.8. Both methods bound the event that alternative 1 is top ranked. Active is the number of ranking columns retained by column generation. Certified requires successful master and pricing statuses and all residual, reduced-cost, and gap checks at the reported tolerance.

complexity: pricing itself is a linear-ordering problem. A numerical endpoint is therefore reported as certified only when the master and pricing solvers terminate successfully, primal and dual residuals satisfy the requested tolerance, the final reduced cost is no smaller than minus that tolerance, and the resulting optimality-gap bound is acceptable. The `ramchoice` diagnostics record these quantities, the number of generated columns, pricing dimensions, iteration count, and solver statuses. Exhaustive four-alternative tests reproduce both endpoints. In a six-alternative software test, a cap of 100 forces column generation and only 18 of the possible 720 ranking columns are needed. Dependence-robust and no-SPI event bounds currently use direct enumeration because introducing a ranking creates a block of menu-specific variables rather than one master column.

The replication benchmark compares direct enumeration and certified column generation on the same deterministic H-LAO choice systems. It records $|X|!$, observed menus and choice equations, enumerated and active ranking columns, pricing variables and constraints, iterations, certification diagnostics, and elapsed time. These general scalability results are distinct from the H-LAO Monte Carlo runtimes in Section SA-6, which include estimation and inference for specific data-generating processes.

Table SA-3 illustrates the tradeoff. At seven alternatives, direct enumeration materializes all 5,040 ranking columns, whereas column generation certifies the same event bounds using 28. Direct enumeration is faster at the small sizes in this benchmark because repeated mixed-integer pricing has a fixed cost. Column generation's advantage is instead that memory and master-problem dimensions depend on the rankings that matter for the target rather than on $|X|!$, which becomes decisive once the full ranking matrix is too large to construct or store.

SA-6 Simulation Evidence

This section provides the simulation designs and complete results for the procedures developed in the paper. The first subsection studies homogeneous AOM, reporting finite-sample size and power under alternative menu supports, size at a boundary null, and design-specific computation time. The second evaluates H-LAO estimation, confidence sets, specification tests, robustness to departures from the baseline assumptions, subject-clustered inference, and design-specific computation time.

SA-6.1 Homogeneous AOM

The universe contains six alternatives. The data-generating preference is $a_1 \succ a_2 \succ \cdots \succ a_6$. We use the logit attention rule of Brady and Rehbeck (2016),

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

where $|T|$ is the cardinality of T. We set $\varsigma = 2$. Direct calculation verifies attention overload: attention frequency depends only on menu size and equals 0.833, 0.750, 0.700, 0.667, and 0.643 for $|S| = 2, 3, 4, 5, 6$, respectively. We compute the implied choice rule from Definition 2 of the main paper and generate independent menu-level samples with $N_S \in \{50, 100, 200, 500\}$ observations per menu.

We test the four hypothesized preferences

$$\succ_1:\ a_1 \succ a_2 \succ a_3 \succ a_4 \succ a_5 \succ a_6, \qquad \succ_2:\ a_2 \succ a_3 \succ a_4 \succ a_5 \succ a_6 \succ a_1,$$

$$\succ_3:\ a_1 \succ a_2 \succ a_6 \succ a_5 \succ a_4 \succ a_3, \qquad \succ_4:\ a_1 \succ a_6 \succ a_5 \succ a_4 \succ a_3 \succ a_2.$$

The data-generating ordering $\succ_1$ is compatible with AOM, so its rejection probability measures size. The other three orderings are incompatible under complete data, although some remain compatible on restricted menu domains. Table SA-4 reports rejection probabilities from 2,000 Monte Carlo replications and 2,000 critical-value draws per replication at nominal level 0.05. The first four result columns use the all-inequalities zero-centered critical value covered directly by Theorem 7 of the main paper; the last four apply the GMS refinement described in Section SA-4.8.5.

The complete-data design uses all 57 menus of size at least two and contains 664 nonredundant inequalities. None is violated by the data-generating ordering. The other three orderings violate 90, 6, and 23 inequalities, respectively. Their rejection probabilities constitute a fixed-alternative power analysis. Power is highest against $\succ_2$, which places the true top alternative last and violates many inequalities. The other alternatives are closer to the identified set and are correspondingly harder to reject.

The remaining menu supports omit selected menu sizes. The number of inequalities consequently ranges from 664 with complete data to 15 with only binary menus and the grand menu. A dagger in the tables marks an ordering that satisfies every population inequality on the indicated domain. Both procedures are conservative for these compatible orderings. When the retained menus expose a violation, rejection probabilities generally increase with sample size; when they do not, the corresponding ordering is correctly retained.

The least-favorable procedure provides substantial size protection in this baseline design: its largest rejection probability for a compatible ordering is 0.009. GMS reduces the contribution of deeply slack inequalities to the effective Gaussian maximum and raises power while keeping rejection probabilities at most 0.035 for compatible orderings. With complete data and $N_S = 200$, GMS raises rejection of the most separated false ordering from 0.208 to 0.478; at $N_S = 500$, the corresponding probabilities are 0.680 and 0.910. With only binary menus and the grand menu, GMS raises rejection of the sole incompatible ordering from 0.179 to 0.523 at $N_S = 200$ and from 0.555 to 0.872 at $N_S = 500$. These comparisons quantify the practical power gain from moment selection alongside the theorem-backed least-favorable baseline.

The baseline null lies in the interior of the population inequality region. To evaluate size when many inequalities bind, we use a second data-generating process based on a nondegenerate capacity-one mixture of search priorities. All four reported orderings satisfy the population inequalities with maximum slack zero; the numbers of binding inequalities for $\succ_1$ through $\succ_4$ are 65, 115, 186, and 301. Thus every entry in Table SA-5 is a boundary-null rejection probability.

At the boundary, the least-favorable procedure rejects with probability between 0.003 and 0.031, while GMS rejection probabilities range from 0.045 to 0.071. GMS therefore concentrates rejection closer to the nominal level, modestly exceeding it in some cells while delivering the power gains documented above. The experiment makes this size–power tradeoff transparent: the all-inequalities procedure remains the theorem-backed primary method, and GMS provides a useful power refinement when greater boundary sensitivity is acceptable.

Table SA-6 reports end-to-end execution time at $N_S = 500$. Average time per replication ranges from 0.17 seconds with 24 inequalities to 2.31 seconds with all 664 inequalities. LF and GMS take nearly identical time because both evaluate the same restriction system and use the same number of Gaussian draws. Absolute times are machine dependent, but the comparison shows that the reported power refinement has essentially no additional computational cost in these designs.

Table SA-4: Finite-sample size and power for homogeneous AOM

			All inequalities (LF)				GMS refinement			
$\{\lvert S\rvert : S \in \mathcal{D}\}$	N_S	$\mathfrak{c}_{\succ}$	$\succ_1$	$\succ_2$	$\succ_3$	$\succ_4$	$\succ_1$	$\succ_2$	$\succ_3$	$\succ_4$
2–6	50	664	$0.001^{\dagger}$	0.029	0.009	0.029	$0.017^{\dagger}$	0.191	0.035	0.070
2–6	100	664	$0.002^{\dagger}$	0.087	0.017	0.054	$0.011^{\dagger}$	0.279	0.055	0.139
2–6	200	664	$0.000^{\dagger}$	0.208	0.030	0.120	$0.003^{\dagger}$	0.478	0.128	0.351
2–6	500	664	$0.000^{\dagger}$	0.680	0.197	0.569	$0.002^{\dagger}$	0.910	0.533	0.880
3–6	50	439	$0.001^{\dagger}$	0.015	0.002	0.003	$0.021^{\dagger}$	0.119	0.020	0.019
3–6	100	439	$0.002^{\dagger}$	0.038	0.003	0.011	$0.006^{\dagger}$	0.140	0.011	0.033
3–6	200	439	$0.001^{\dagger}$	0.043	0.005	0.029	$0.005^{\dagger}$	0.175	0.023	0.089
3–6	500	439	$0.000^{\dagger}$	0.104	0.019	0.159	$0.001^{\dagger}$	0.304	0.070	0.400
4–6	50	159	$0.002^{\dagger}$	0.018	$0.008^{\dagger}$	0.007	$0.035^{\dagger}$	0.095	$0.032^{\dagger}$	0.026
4–6	100	159	$0.003^{\dagger}$	0.017	$0.005^{\dagger}$	0.006	$0.025^{\dagger}$	0.094	$0.024^{\dagger}$	0.021
4–6	200	159	$0.001^{\dagger}$	0.028	$0.001^{\dagger}$	0.001	$0.009^{\dagger}$	0.111	$0.009^{\dagger}$	0.015
4–6	500	159	$0.000^{\dagger}$	0.034	$0.000^{\dagger}$	0.005	$0.003^{\dagger}$	0.136	$0.003^{\dagger}$	0.042
5–6	50	24	$0.004^{\dagger}$	0.010	$0.008^{\dagger}$	$0.009^{\dagger}$	$0.032^{\dagger}$	0.057	$0.032^{\dagger}$	$0.029^{\dagger}$
5–6	100	24	$0.002^{\dagger}$	0.019	$0.004^{\dagger}$	$0.004^{\dagger}$	$0.026^{\dagger}$	0.083	$0.026^{\dagger}$	$0.022^{\dagger}$
5–6	200	24	$0.002^{\dagger}$	0.015	$0.002^{\dagger}$	$0.002^{\dagger}$	$0.013^{\dagger}$	0.067	$0.013^{\dagger}$	$0.011^{\dagger}$
5–6	500	24	$0.001^{\dagger}$	0.024	$0.001^{\dagger}$	$0.001^{\dagger}$	$0.005^{\dagger}$	0.082	$0.005^{\dagger}$	$0.004^{\dagger}$
2, 3, 4, 6	50	370	$0.000^{\dagger}$	0.026	0.011	0.032	$0.011^{\dagger}$	0.162	0.044	0.088
2, 3, 4, 6	100	370	$0.000^{\dagger}$	0.086	0.019	0.057	$0.003^{\dagger}$	0.275	0.073	0.178
2, 3, 4, 6	200	370	$0.000^{\dagger}$	0.215	0.050	0.148	$0.002^{\dagger}$	0.507	0.175	0.404
2, 3, 4, 6	500	370	$0.000^{\dagger}$	0.665	0.206	0.549	$0.001^{\dagger}$	0.909	0.584	0.881
2, 3, 6	50	115	$0.001^{\dagger}$	0.040	0.013	0.035	$0.006^{\dagger}$	0.139	0.051	0.094
2, 3, 6	100	115	$0.000^{\dagger}$	0.085	0.026	0.058	$0.000^{\dagger}$	0.233	0.097	0.182
2, 3, 6	200	115	$0.000^{\dagger}$	0.174	0.049	0.111	$0.001^{\dagger}$	0.447	0.225	0.374
2, 3, 6	500	115	$0.000^{\dagger}$	0.545	0.205	0.409	$0.000^{\dagger}$	0.865	0.621	0.821
2, 6	50	15	$0.000^{\dagger}$	0.040	$0.000^{\dagger}$	$0.000^{\dagger}$	$0.000^{\dagger}$	0.180	$0.000^{\dagger}$	$0.001^{\dagger}$
2, 6	100	15	$0.000^{\dagger}$	0.087	$0.000^{\dagger}$	$0.000^{\dagger}$	$0.000^{\dagger}$	0.302	$0.000^{\dagger}$	$0.000^{\dagger}$
2, 6	200	15	$0.000^{\dagger}$	0.179	$0.000^{\dagger}$	$0.000^{\dagger}$	$0.000^{\dagger}$	0.523	$0.000^{\dagger}$	$0.000^{\dagger}$
2, 6	500	15	$0.000^{\dagger}$	0.555	$0.000^{\dagger}$	$0.000^{\dagger}$	$0.000^{\dagger}$	0.872	$0.000^{\dagger}$	$0.000^{\dagger}$

Notes: Entries are Monte Carlo rejection probabilities at the five-percent level based on 2,000 replications and 2,000 critical-value draws. LF is the all-inequalities feasible Gaussian test corresponding to Theorem 7; GMS is the Andrews–Soares refinement with `MNRatioGMS`$= 1/\log(N)$. A dagger marks a preference ordering that satisfies the population inequalities on the indicated menu support. Daggered entries measure size; the remaining entries measure power against fixed alternatives. The quantity $\mathfrak{c}_{\succ}$ is the number of $\succ$-Regularity inequalities. All menus of the listed sizes are observed. The four rankings are defined in the text. Each observed menu has the reported sample size.

Table SA-5: Boundary-null size for homogeneous AOM

	All inequalities (LF)				GMS refinement			
N_S	$\succ_1$	$\succ_2$	$\succ_3$	$\succ_4$	$\succ_1$	$\succ_2$	$\succ_3$	$\succ_4$
50	0.011	0.010	0.018	0.022	0.060	0.045	0.066	0.067
100	0.009	0.009	0.020	0.031	0.065	0.059	0.067	0.068
200	0.009	0.012	0.018	0.029	0.068	0.064	0.067	0.071
500	0.003	0.009	0.013	0.021	0.065	0.071	0.060	0.067

Notes: Entries are rejection probabilities at the five-percent level based on 2,000 replications and 2,000 critical-value draws. The DGP is a nondegenerate capacity-one priority mixture for which all four reported orderings satisfy the population inequalities with maximum slack zero. The numbers of binding inequalities for $\succ_1$ through $\succ_4$ are 65, 115, 186, 301 respectively. Thus every entry measures size at a boundary null rather than at the strictly interior null used in the baseline size-and-power design.

Table SA-6: Average homogeneous-AOM execution time at $N_S = 500$

$\{\lvert S\rvert : S \in \mathcal{D}\}$	$\mathfrak{c}_\succ$	LF	GMS
2–6	664	2.31	2.30
3–6	439	1.55	1.53
4–6	159	0.51	0.51
5–6	24	0.17	0.17
2, 3, 4, 6	370	2.21	2.19
2, 3, 6	115	0.86	0.86
2, 6	15	0.25	0.25

Notes: Entries are average wall-clock seconds per Monte Carlo replication. Each method is run end to end and jointly tests the four reported rankings using the same simulated sample and common Gaussian draws. A call includes data processing, construction and evaluation of the restrictions, and 2,000 critical-value draws. LF and GMS are run consecutively on the same compute node within each replication; absolute times remain machine dependent.

Table SA-7: H-LAO Monte Carlo configurations

Design	$\lvert X \rvert$	Preference distribution	Stopping rule	Reach	Menu support	Dependence	SPI
H01	4	Diffuse	Geometric	High	Full	Independent	Yes
H02	4	Concentrated	Geometric	High	Full	Independent	Yes
H03	4	Diffuse	Rank dependent	High	Prefix rich	Independent	Yes
H04	4	Diffuse	Alternative dependent	High	Prefix rich	Independent	Yes
H05	4	Diffuse	Geometric	Low	Full	Independent	Yes
H06	4	Diffuse	Rank dependent	Low	Prefix rich	Independent	Yes
H07	4	Concentrated	Alternative dependent	Low	Prefix rich	Dependent	Yes
H08	4	Diffuse	Rank dependent	Zero	Prefix rich	Independent	Yes
H09	6	Diffuse	Geometric	High	Sparse pairwise	Independent	Yes
H10	6	Concentrated	Rank dependent	Low	Sparse pairwise	Independent	Yes
H11	4	Late concentrated	Menu dependent	Moderate	Full	Independent	No

Notes: Full support contains every nonempty menu. Prefix-rich support contains the menus required by the recursive reach construction for the selected lists. Sparse-pairwise support contains singleton and binary menus and targets the ranking-free pairwise procedure. In H07, preferences and stopping are coupled while their marginal distributions are held fixed. H11 satisfies prefix consideration and attention overload but violates Sequential Path Independence (SPI).

SA-6.2 Heterogeneous H-LAO

We use a separate Monte Carlo experiment to evaluate the H-LAO estimators, confidence sets, specification diagnostics, and robustness envelopes. The experiment varies the dimensions that govern identification and finite-sample performance: the marginal preference distribution, the stopping rule, terminal reach, observed menu support, dependence between preferences and stopping, and Sequential Path Independence. Designs H01–H08 and H11 have four alternatives; H09 and H10 have six. For each configuration, we generate independent menu-level samples with $N_S \in \{50, 100, 200, 500\}$ observations per observed menu. Table SA-7 records the eleven configurations.

The diffuse preference distribution assigns positive mass to every strict ordering, while the concentrated distribution places three quarters of its mass on one ordering and spreads the remainder over all orderings. Geometric stopping uses a common continuation probability. Rank-dependent stopping varies continuation with list position, and alternative-dependent stopping varies it with the identity of the item reached. The high-, low-, and zero-reach designs respectively evaluate stable recursion, weak-denominator inference, and identified-set inference at prefixes where point recovery is unavailable. H07 couples preferences and stopping while preserving their marginal distributions, thereby separating the dependence-robust analysis from the benchmark independence interpretation.

H11 isolates Sequential Path Independence. It places 0.95 of preference mass on the reverse-list ordering and distributes the remaining 0.05 according to the diffuse design. For the global list $1, 2, 3, 4$, let $h = (0.92, 0.82, 0.72, 0.62)$ and $d_k = 1 - 0.10(k-1)$. If alternative a appears in menu S, its latent reach probability is

$$\phi(a|S) = h_a d_{|S|}.$$

The resulting reach probabilities decrease along every observed list and weakly decrease when the menu expands, so the implied prefix masses are nonnegative and attention overload holds. Sequential Path Independence fails already at $S = \{1,2\}$:

$$\phi(2|\{1,2\}) = 0.82(0.90) = 0.738 \neq \phi(1|\{1,2\})\phi(2|\{2\}) = 0.828(0.82) = 0.67896.$$

The failure is observable through a particularly simple population certificate. The SPI recovery assigns masses 0.172, 0.14904, and 0.67896 to $\emptyset$, $\{1\}$, and $\{1,2\}$, respectively. Every decision maker who stops at $\{1\}$ chooses alternative 1, regardless of any dependence between preference and stopping, so either SPI polytope requires

$$\pi(1|\{1,2\}) \geq 0.14904.$$

The H11 population instead has $\pi(1|\{1,2\}) = 0.1071558$, a violation of about 0.0418842. Equivalently, the independent SPI recursion produces $f(1|\{1,2\}) = -0.0616888$. This one inequality certifies that both the independent and dependence-robust SPI polytopes are empty. By contrast, the true no-SPI prefix masses 0.172, 0.09, and 0.738, coupled independently with the stated preference distribution, reproduce the population choices and furnish an explicit feasible point of the no-SPI polytope.

For point estimation, we report bias and root mean squared error for recovered reach probabilities and, when the observed support and SPI permit the recursion, the full-attention rule f. For inference under independence and SPI, we report familywise coverage—the probability that every valid pairwise target is covered in a replication—and average width for three simultaneous pairwise procedures: finite-sample Hoeffding projection, hybrid Gaussian/exact-binomial projection, and Bonferroni studentization of the undivided moments. The last method uses the exact quadratic inversion implemented in the software; its reported width is the total length if the Fieller set is disconnected, and it returns $[0,1]$ at an exactly degenerate moment.

For four-alternative designs, the preference event is that alternative 4 is preferred to alternative 1. We compute dependence-robust projection intervals under SPI and no-SPI projection intervals for the targeted H01, H05, H07, and H11 configurations. Their population widths quantify how relaxing preference–stopping independence and then SPI changes identification. H07 is generated outside the independent benchmark but inside both robust models. H11 is generated outside both SPI models but inside the no-SPI model. All projection procedures begin from simultaneous menu-level multinomial bands. We compare finite-sample Hoeffding bands with the hybrid region defined in Section SA-4.10. The latter uses Gaussian calibration with the plug-in menu-level multinomial covariance structure for regular cells and exact Clopper–Pearson bounds for boundary and sparse cells, with the common error-budget allocation stated above.

The specification experiment uses four alternatives, complete menu data, high reach, and a preference distribution on the boundary of a Block–Marschak inequality. It contains a correctly specified null and local or fixed violations of either Recovered Attention Overload or Block–Marschak non-negativity at $N_S \in \{100, 200, 500\}$. The local attention-overload and Block–Marschak perturbations are $0.75/\sqrt{N_S}$ and $3/\sqrt{N_S}$, respectively; the fixed perturbations are 0.08 and 0.30. We compare three simultaneous diagnostics: a finite-sample Hoeffding outer test, a hybrid Gaussian/exact-binomial outer test, and the direct delta-Gaussian test available in this regular positive-reach design. The reported rejection probabilities use the targeted restriction under each alternative and the omnibus test under the null. The experiment uses 2,000 Monte Carlo replications and 2,000 critical-value draws at nominal level 0.05. The complete design has 44 menu-iid H-LAO estimation and inference tasks, four subject-clustered inference tasks, and 15 specification-diagnostic tasks.

Table SA-8 shows that reach and full-attention estimates are essentially unbiased in every design where the target is identified. Reach RMSE ranges from 0.019 to 0.046 at $N_S = 100$ and from 0.008 to 0.020 at $N_S = 500$. Full-attention recovery is less precise under weak reach, as expected: in H05 its RMSE falls from 0.260 to 0.111, compared with 0.059 to 0.026 in the otherwise similar high-reach H01 design.

Table SA-9 shows exceptionally reliable familywise coverage: it equals one for Hoeffding projection in every reported design and ranges from 0.994 to one for Gaussian projection. The accompanying intervals are wider, especially under weak reach. The studentized undivided-moment procedure is much shorter but relies on regular-case approximation; its familywise coverage ranges from 0.150 to 0.960 across the positive-reach designs. H10, which combines concentrated preferences, sparse pairwise support, and low reach, is the most demanding configuration: coverage rises from 0.150 at $N_S = 50$ to 0.866 at $N_S = 500$. At the exactly degenerate H08 boundary, every procedure correctly returns $[0, 1]$, making visible the distinction between finite-sample validity and informativeness at zero reach. These results strongly support projection intervals as the primary procedure and studentized intervals as a shorter regular-case diagnostic.

The subject-clustered experiment is calibrated to the repeated-choice structure and numbers of independent subjects in the empirical application. To isolate within-subject dependence, it uses a balanced

Table SA-8: H-LAO point estimation

Design	N_S	Reach probabilities Bias	RMSE	Full-attention rule Bias	RMSE
H01	100	-0.000	0.036	-0.000	0.059
H01	500	0.000	0.016	-0.000	0.026
H02	100	-0.000	0.036	-0.000	0.036
H02	500	-0.000	0.016	-0.000	0.016
H03	100	-0.000	0.034	–	–
H03	500	-0.000	0.015	–	–
H04	100	0.000	0.036	–	–
H04	500	-0.000	0.016	–	–
H05	100	-0.000	0.041	-0.000	0.260
H05	500	0.000	0.018	-0.000	0.111
H06	100	-0.001	0.043	–	–
H06	500	0.000	0.019	–	–
H07	100	-0.000	0.042	–	–
H07	500	0.000	0.019	–	–
H08	100	0.000	0.019	–	–
H08	500	0.000	0.008	–	–
H09	100	-0.000	0.034	–	–
H09	500	0.000	0.015	–	–
H10	100	0.000	0.046	–	–
H10	500	0.000	0.020	–	–

Notes: Bias and RMSE are averaged across replications and the estimands available in each design. Dashes indicate that the menu support does not identify the recursive full-attention rule.

full-menu design: each subject draws one preference ranking that remains fixed across the full 15-menu domain and makes three choices per menu, while stopping draws vary across choices. We use $G \in \{21, 37\}$ subjects, so $N_S = 3G \in \{63, 111\}$, under either diffuse preferences (C01) or concentrated preferences (C02). The cluster-Hoeffding and cluster-multiplier procedures propagate simultaneous cluster-robust primitive-probability bands; the cluster-studentized procedure applies the shorter undivided-moment inversion with cluster calibration.

Table SA-10 shows familywise coverage equal to one for cluster-Hoeffding projection and between 0.9995 and one for cluster-multiplier projection. Cluster-studentized coverage is 0.902 and 0.921 under diffuse preferences, but only 0.703 and 0.797 under concentrated preferences. Its intervals are shorter, but the loss of simultaneous coverage is material at precisely the cluster counts used in the application. This evidence reinforces the use of cluster projection intervals as the primary empirical results.

Table SA-11 shows that the sensitivity designs reproduce the theoretical nesting of the identified sets. In H01, for example, the population width rises from zero under independence to 0.190 under dependence and 0.590 without Sequential Path Independence. The corresponding widths in weak-reach H05 are 0, 0.750, and 0.873. The robust projection regions cover in every reported design, and their average widths decrease toward the corresponding population identified-set widths as sample size increases.

Table SA-12 shows that the direct delta-Gaussian diagnostic combines conservative null rejection probabilities of 0.011–0.018 with power of 0.986 and 0.977 against the fixed attention-overload and Block–Marschak violations at $N_S = 500$. Both outer tests make no false rejections under correct specification in these designs. Among them, the Gaussian outer test reaches power of 0.874 against the fixed attention-overload violation, while power against the Block–Marschak alternative is limited; the Hoeffding outer test prioritizes finite-sample validity and remains highly conservative at these sample sizes. The results distinguish the roles of the procedures: outer tests provide finite-sample or boundary-robust safeguards, while the direct regular-case diagnostic is highly informative when positive reach and differentiability are credible.

For H01–H10 at $N_S = 500$, Figure SA-1 reports average end-to-end time per replication below 0.65 seconds for both projection procedures. Runtime is highest in the full-menu designs and lower with prefix-rich or sparse pairwise support. Correlated-Gaussian projection is slightly slower because it estimates and simulates from the joint covariance structure, but the difference is small throughout.

Table SA-9: H-LAO pairwise confidence sets

		Hoeffding projection		Gaussian projection		Studentized moment	
Design	N_S	Familywise	Width	Familywise	Width	Familywise	Width
H01	50	1.000	0.871	1.000	0.821	0.926	0.456
H01	100	1.000	0.789	1.000	0.626	0.936	0.321
H01	200	1.000	0.656	1.000	0.434	0.945	0.226
H01	500	1.000	0.416	1.000	0.272	0.955	0.143
H02	50	1.000	0.924	1.000	0.522	0.787	0.256
H02	100	1.000	0.595	1.000	0.314	0.866	0.195
H02	200	1.000	0.409	0.994	0.211	0.903	0.138
H02	500	1.000	0.268	0.998	0.132	0.932	0.088
H03	50	1.000	0.836	0.999	0.780	0.938	0.404
H03	100	1.000	0.755	1.000	0.589	0.941	0.285
H03	200	1.000	0.607	1.000	0.385	0.938	0.201
H03	500	1.000	0.380	1.000	0.241	0.960	0.127
H04	50	1.000	0.894	1.000	0.834	0.933	0.500
H04	100	1.000	0.814	1.000	0.688	0.941	0.351
H04	200	1.000	0.694	1.000	0.481	0.952	0.247
H04	500	1.000	0.444	1.000	0.300	0.946	0.155
H05	50	1.000	1.000	1.000	0.995	0.887	0.876
H05	100	1.000	0.999	1.000	0.964	0.921	0.705
H05	200	1.000	0.988	1.000	0.877	0.938	0.510
H05	500	1.000	0.890	1.000	0.687	0.951	0.317
H06	50	1.000	0.998	1.000	0.978	0.926	0.751
H06	100	1.000	0.986	1.000	0.906	0.945	0.556
H06	200	1.000	0.922	1.000	0.793	0.941	0.387
H06	500	1.000	0.784	1.000	0.557	0.951	0.241
H08	50	1.000	1.000	1.000	1.000	1.000	1.000
H08	100	1.000	1.000	1.000	1.000	1.000	1.000
H08	200	1.000	1.000	1.000	1.000	1.000	1.000
H08	500	1.000	1.000	1.000	1.000	1.000	1.000
H09	50	1.000	0.872	1.000	0.822	0.914	0.510
H09	100	1.000	0.791	1.000	0.639	0.938	0.359
H09	200	1.000	0.668	1.000	0.440	0.939	0.252
H09	500	1.000	0.422	1.000	0.276	0.952	0.159
H10	50	1.000	1.000	1.000	0.996	0.150	0.407
H10	100	1.000	1.000	1.000	0.809	0.566	0.312
H10	200	1.000	0.915	1.000	0.426	0.740	0.243
H10	500	1.000	0.538	0.998	0.244	0.866	0.156

Notes: Familywise coverage is the probability that all valid pairwise targets are covered in a replication; width is averaged over pairwise targets and replications. H07 is omitted because pairwise point identification requires preference–stopping independence. The studentized procedure inverts the undivided moment with Bonferroni calibration; width is the total length when the confidence set is disconnected. At an exactly degenerate moment it reports $[0, 1]$.

Table SA-10: Subject-clustered H-LAO pairwise confidence sets

Design	Preference heterogeneity	G	N_S	Cluster Hoeffding Familywise	Cluster Hoeffding Width	Cluster multiplier Familywise	Cluster multiplier Width	Cluster studentized Familywise	Cluster studentized Width
C01	Diffuse	21	63	1.000	0.970	1.000	0.898	0.902	0.619
C01	Diffuse	37	111	1.000	0.910	1.000	0.751	0.921	0.469
C02	Concentrated	21	63	1.000	1.000	1.000	0.947	0.703	0.312
C02	Concentrated	37	111	1.000	0.998	1.000	0.552	0.797	0.270

Notes: Each subject draws one preference ranking that is held fixed across the full 15-menu domain and makes three choices per menu; stopping draws vary across choices. Hence $N_S = 3G$ and observations are dependent within subject. The values $G = 21$ and $G = 37$ mirror the complete-design and dominance-consistent empirical samples. Familywise coverage requires simultaneous coverage of all six pairwise preference shares in a replication; width is averaged over the six targets and replications.

Table SA-11: Sensitivity to preference–stopping independence and Sequential Path Independence

Design	N_S	Population identified-set width: Independent	Population identified-set width: Dependent	Population identified-set width: No SPI	Dependence robust: Coverage	Dependence robust: Width	No SPI: Coverage	No SPI: Width
H01	100	0.000	0.190	0.590	1.000	0.675	1.000	0.749
H01	500	0.000	0.190	0.590	1.000	0.408	1.000	0.661
H05	100	0.000	0.750	0.873	1.000	0.983	1.000	0.983
H05	500	0.000	0.750	0.873	1.000	0.909	1.000	0.921
H07	100	–	0.914	0.914	1.000	0.992	1.000	0.992
H07	500	–	0.914	0.914	1.000	0.951	1.000	0.951
H11	100	–	–	0.456	–	–	1.000	0.608
H11	500	–	–	0.456	–	–	1.000	0.527

Notes: Population columns report the width of the sharp identified interval for the designated preference event. Sampling columns use hybrid Gaussian/exact-binomial probability bands. H07 violates preference–stopping independence; H11 violates Sequential Path Independence while preserving prefix consideration, attention overload, and a stable preference marginal. Dashes mark a misspecified model or an inapplicable coverage calculation.

Table SA-12: H-LAO specification diagnostics

Design	N_S	Violation	Hoeffding outer	Gaussian outer	Direct delta
Null	100	0.000	0.000	0.000	0.018
Null	200	0.000	0.000	0.000	0.013
Null	500	0.000	0.000	0.000	0.011
AO local	100	0.075	0.000	0.000	0.114
AO local	200	0.053	0.000	0.005	0.102
AO local	500	0.034	0.000	0.005	0.092
AO fixed	100	0.080	0.000	0.000	0.144
AO fixed	200	0.080	0.000	0.020	0.593
AO fixed	500	0.080	0.000	0.874	0.986
BM local	100	0.300	0.000	0.000	0.160
BM local	200	0.212	0.000	0.000	0.126
BM local	500	0.134	0.000	0.000	0.142
BM fixed	100	0.300	0.000	0.000	0.171
BM fixed	200	0.300	0.000	0.000	0.502
BM fixed	500	0.300	0.000	0.034	0.977

Notes: Entries in the final three columns are rejection probabilities at the five-percent level. The null row uses the omnibus diagnostic; AO and BM alternatives use their corresponding targeted restrictions. Local violations shrink at the parametric rate, while fixed violations do not depend on sample size.

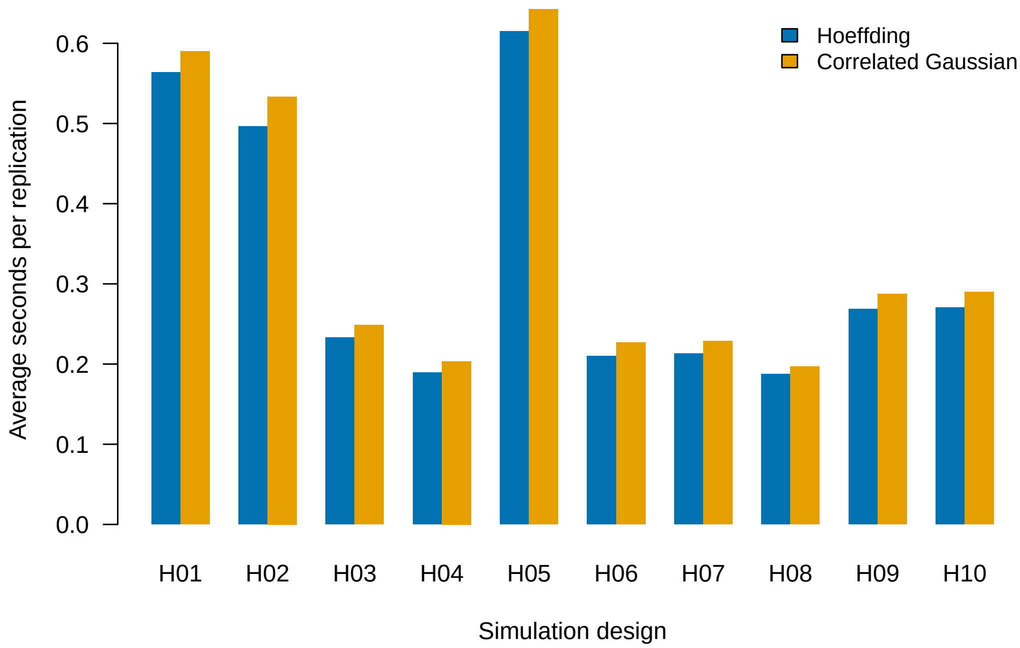

Figure SA-1: Computation time for H-LAO projection inference at $N_S = 500$.

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.

Abaluck, J., G. Compiani, and F. Zhang (2026): "A Method to Estimate Discrete Choice Models That Is Robust to Consumer Search," *Journal of Political Economy*, 134(7), 1967–2022.

Abaluck, J., and J. Gruber (2023): "When Less Is More: Improving Choices in Health Insurance Markets," *Review of Economic Studies*, 90(3), 1011–1040.

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.

Aguiar, V. H., and N. Kashaev (2025): "Identification and Estimation of Discrete Choice Models with Unobserved Choice Sets," *Journal of Business & Economic Statistics*, 43(1), 204–215.

Allen, R. (2026): "Exogenous Consideration and Extended Random Utility," *Journal of Econometrics*, 253, 106166.

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

Auster, S., and Y.-K. Che (2025): "Pandora's Box Reopened: Robust Search and Choice Overload," Working paper, arXiv:2512.21192.

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*, 41(4), 1011–1029.

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

Başar, G., and C. Bhat (2004): "A Parameterized Consideration Set Model for Airport Choice: An Application to the San Francisco Bay Area," *Transportation Research Part B: Methodological*, 38(10), 889–904.

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

Caliari, D., and H. Petri (2025): "The Luce Model, Regularity, and Choice Overload," Working paper, arXiv:2502.21063.

Caradonna, P., and C. Turansick (2026): "Identification in Stochastic Choice," Working paper, arXiv:2602.19950.

Cattaneo, M. D., Y. Feng, and W. G. Underwood (2024): "Uniform Inference for Kernel Density Estimators with Dyadic Data," *Journal of the American Statistical Association*, 119(548), 2695–2708.

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

Cattaneo, M. D., and R. R. Yu (2025): "Strong Approximations for Empirical Processes Indexed by Lipschitz Functions," *The Annals of Statistics*, 53(3), 1203–1229.

Chadd, I. (2023): "Random Network Consideration: Theory and Experiment," *Journal of Economic Behavior & Organization*, 211, 251–269.

Chadd, I., E. Filiz-Ozbay, and E. Y. Ozbay (2025): "Choosing to Search: Choice with a Default Option," *Experimental Economics*, 28(2), 237–253.

Chambers, C. P., and C. Turansick (2025): "The Limits of Identification in Discrete Choice," *Games and Economic Behavior*, 150, 537–551.

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 (2022): "Improved Central Limit Theorem and Bootstrap Approximations in High Dimensions," *Annals of Statistics*, 50(5), 2562–2586.

Chib, S., and K. Shimizu (2026): "Scalable Estimation of Multinomial Response Models with Random Consideration Sets," *Journal of the American Statistical Association*.

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.

de Lara, L., and M. Dean (2024): "Rational Choice Overload," Working paper.

Dean, M., and N. Neligh (2023): "Experimental Tests of Rational Inattention," *Journal of Political Economy*, 131(12), 3415–3461.

Dean, M., D. Ravindran, and J. Stoye (2025): "A Better Test of Choice Overload," Working paper.

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

Demuynck, T., and C. Seel (2018): "Revealed Peference with Limited Consideration," *American Economic Journal: Microeconomics*, 10(1), 102–131.

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.

Ellis, K., E. Filiz-Ozbay, and E. Y. Ozbay (2025): "Consistency and Heterogeneity in Consideration and Choice," Working paper.

Falmagne, J.-C. (1978): "A Representation Theorem for Finite Random Scale Systems," *Journal of Mathematical Psychology*, 18(1), 52–72.

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.

Goeree, M. S. (2008): "Limited Information and Advertising in the US Personal Computer Industry," *Econometrica*, 76(5), 1017–1074.

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.

Hauser, J. R., and B. Wernerfelt (1990): "An Evaluation Cost Model of Consideration Sets," *Journal of Consumer Research*, 16(4), 393–408.

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

Huberman, G., and T. Regev (2001): "Contagious Speculation and a Cure for Cancer: A Nonevent That Made Stock Prices Soar," *Journal of Finance*, 56(1), 387–396.

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

Iyengar, S. S., and M. R. Lepper (2000): "When choice is demotivating: Can one desire too much of a good thing?," *Journal of Personality and Social Psychology*, 79(6), 995–1006.

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

Khan, I., I. Krajbich, C. Raymond, S. Sundaresan, and A. Veiga (2026): "Measuring Consideration: The Attentional Origins of Stochastic Choice," Working paper.

Kono, H., K. Saito, and A. Sandroni (2026): "Random Utility with Unobservable Alternatives," *American Economic Review*, Forthcoming.

Laroche, M., J. Rosenblatt, and I. Sinclair (1984): "Brand Categorization Strategies in an Extensive Problem Solving Situation: A Study of University Choice," *ACR North American Advances*, NA-11.

Liao, Y., K. Saito, and A. Sandroni (2026): "Random Utility with Aggregation," Working paper, arXiv:2506.00372, revised 2026.

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.

Lopes, M. E. (2022): "Central Limit Theorem and Bootstrap Approximation in High Dimensions: Near $1/\sqrt{n}$ Rates via Implicit Smoothing," *The Annals of Statistics*, 50(5), 2492–2513.

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.

Masatlioglu, Y., and T. P. Vu (2024): "Growing Attention," *Journal of Economic Theory*, 222, 105929.

Natan, O. R. (2024): "Choice Frictions in Large Assortments," *Marketing Science*, 44(3), 593–625.

Reutskaja, E., and R. M. Hogarth (2009): "Satisfaction in Choice as a Function of the Number of Alternatives: When "Goods Satiate"," *Psychology and Marketing*, 26(3), 197–203.

Reutskaja, E., R. Nagel, C. F. Camerer, and A. Rangel (2011): "Search Dynamics in Consumer Choice under Time Pressure: An Eye-Tracking Study," *American Economic Review*, 101(2), 900–926.

Rosen, D. E., J. M. Curran, and T. B. Greenlee (1998): "College Choice in a Brand Elimination Framework: The High School Students Perspective," *Journal of Marketing for Higher Education*, 8(3), 73–92.

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

Sheridan, J. E., M. D. Richards, and J. W. Slocum (1975): "Comparative Analysis of Expectancy and Heuristic Models of Decision Behavior," *Journal of Applied Psychology*, 60(3), 361–368.

Shocker, A. D., M. Ben-Akiva, B. Boccara, and P. Nedungadi (1991): "Consideration Set Influences on Consumer Decision-making and Choice: Issues, Models, and Suggestions," *Marketing Letters*, 2(3), 181–197.

Thomas, A. W., F. Molter, and I. Krajbich (2021): "Uncovering the Computational Mechanisms Underlying Many-Alternative Choice," *eLife*, 10, e57012.

Turansick, C. (2022): "Identification in the Random Utility Model," *Journal of Economic Theory*, p. 105489.

Wang, Z., J. Sicsic, and P. Chauvin (2026): "From Random Utility to Random Attention: Validating the Random Attention Model in Health Preference Analysis," Working paper.

Wang, Z., and D. Zhu (2025): "Identification of Daily Travel Mode Preferences under Random Attention Allocation: Evidence from a Lab Experiment and Paris," Working paper.

Wei, D. (2024): "Random Attention Span," Working paper, arXiv:2405.11578.

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