Sharp ordinal benefit bounds for survivor compliers under selection and noncompliance

CausalSmith*

September 9, 2026

Abstract

This paper characterizes sharp bounds on the survivor-complier strict-benefit probability in a finite ordered-outcome instrumental-variables model with treatment-induced selection. The target is the probability θ, the survivor-complier benefit probability, that treatment raises the ordered outcome among units whose treatment receipt is shifted by the instrument and whose outcome would be observed under either treatment state. Under conditional instrument independence, consistency and exclusion through received treatment, instrument overlap, treatment monotonicity, covariate-specific weak selection monotonicity, and positive aggregate survivor-complier mass, observed instrument contrasts identify two cellwise selected-complier outcome capacities. Their smaller total is the survivor-complier mass, including cells with zero selected-complier capacity gap. The sharp identified set is the interval obtained by solving a branch-free exact-mass partial-transport problem in each covariate cell and aggregating the resulting threshold-cut values. The lower and upper endpoints have closed finite formulas, and endpoint-attaining full latent laws establish sharpness. The same construction gives sparse endpoint witnesses computable in $O(|\mathcal{X}|K)$ operations for finite covariate support $\mathcal{X}$ and K ordered outcome levels, with $O(|\mathcal{X}|K^2)$ work when full coupling matrices are materialized. A three-level synthetic witness yields the sharp interval $[0, 0.7]$. For i.i.d. samples from the fixed finite-slate structural class with fixed finite support, prescribed screening thresholds η_n satisfying $\eta_n \to 0$ and $\sqrt{n}\,\eta_n \to \infty$, a common instrument-overlap bound, and a uniform positive lower bound on aggregate survivor-complier mass, a projected screened plug-in estimator has a fixed-law Hadamard directional limit and a deterministic guarded confidence interval contains the full sharp identified interval uniformly.

1 Introduction

Many empirical questions concern a treatment's probability of making a unit better off on an ordered scale. In a training-program setting, for example, a randomized offer can shift actual program receipt, employment determines whether wages are observed, and the outcome of interest may be a prespecified ordered wage category. A natural principal-stratum target is then the probability that program receipt raises the wage category among offer compliers who would be employed under either receipt state. This paper gives sharp bounds and guarded

*Machine-generated by the CausalSmith research pipeline. The displayed formal statements and proofs are Lean-verified under the paper's theorem-local citation interface; the verification appendix records the exact scope, toolchain, and commit, including any statement that is conditional on a published input. Cited work otherwise supplies framing and positioning.

inference for that same-unit ordinal benefit probability in a finite ordered-outcome instrumental-variables model with treatment-induced selection.

The model combines a binary instrument Z, treatment receipt D, selection S, covariates X, and an ordered outcome $Y \in \{0, \ldots, K-1\}$ observed among selected units. The complier event C is the group with $D_1 > D_0$, where D_0, D_1 are potential treatment receipts under the two instrument states. The survivor-complier target conditions further on $S_0 = S_1 = 1$, where S_0, S_1 are potential selection indicators under the two treatment states. The estimand is

$$\theta = P(Y_1 > Y_0 \mid S_1 = S_0 = 1, C),$$

the probability that the treatment-one potential ordered outcome exceeds the treatment-zero potential ordered outcome among survivor compliers. In the Job Corps-style interpretation, Z is offer assignment, D is program receipt, S is employment, and Y is an ordered wage category; the numerical example below is a synthetic compatible law.

The identifying restrictions follow the local-average-treatment-effect and principal-stratification traditions [Imbens and Angrist, 1994, Angrist et al., 1996, Frangakis and Rubin, 2002]. Conditional on covariates, the instrument is independent of potential treatments, potential selections, and potential ordered outcomes. Observed treatment equals treatment under the realized instrument, observed selection equals selection under the realized treatment, and observed outcomes equal potential outcomes under the realized treatment on selected units. Treatment receipt is monotone in the instrument, and instrument propensities satisfy cellwise overlap. Selection is governed by a covariate-specific weak monotonicity direction: within each complier cell, either treatment weakly raises selection or treatment weakly lowers selection. These conditions match the selected-complier logic used in IV sample-selection work [Chen and Flores, 2015, Kennedy et al., 2019, Dong and Heiler, 2026].

The first result identifies the observable capacities that enter the benefit problem. For each covariate cell x, the lower-arm capacities $\ell_i(x)$ and upper-arm capacities $h_j(x)$ are observed selected-complier outcome submeasures for $Y_0 = i$ and $Y_1 = j$, respectively. Their totals $q_0(x)$ and $q_1(x)$ can differ because treatment changes selection. Weak selection monotonicity makes

$$m(x) = \min\{q_0(x), q_1(x)\}$$

the survivor-complier mass in cell x. Thus the observed law supplies two possibly unequal nonnegative capacities and an exact amount of mass that must be paired between them.

The construction follows a four-step route. Observed conditional instrument contrasts first produce the lower- and upper-arm selected-complier capacities. Their totals determine the cellwise exact survivor-complier mass. The threshold cuts then solve the exact-mass transport problem in each cell. Finally, the cellwise cut values are averaged with covariate-cell weights and normalized by aggregate survivor-complier mass to obtain the sharp interval.

The sharp-bound problem is a finite exact-mass partial-transport problem. In each cell, a feasible coupling γ_x places mass on ordered pairs (i, j), respects the row capacities $\ell_i(x)$, respects the column capacities $h_j(x)$, and has total mass $m(x)$. The cellwise benefit functional

$$b_x(\gamma_x) = \sum_{i<j} \gamma_{ij,x}$$

counts strict ordinal improvement. The paper proves that the feasible projections of full latent IV-selection laws are exactly these exact-mass transport polytopes, and that the sharp lower and upper values are the threshold-cut quantities $B_L(x)$ and $B_U(x)$. Aggregating by covariate-cell probabilities p_x gives

$$\Theta_I(P_{\text{obs}}) = \left[\frac{\sum_x p_x B_L(x)}{\sum_x p_x m(x)}, \frac{\sum_x p_x B_U(x)}{\sum_x p_x m(x)}\right].$$

Endpoint-attaining full latent laws show that the interval is sharp under the maintained structural law class.

The threshold formulas connect this construction to existing ordinal benefit bounds. Makarov-type and transport bounds characterize feasible joint distributions from marginal information [Makarov, 1982, Williamson and Downs, 1990, Fan and Park, 2010, Villani, 2009]. For complete ordered marginals, Lu et al. [2018] give closed-form sharp bounds for the strict-benefit probability. The present exact-mass formulation recovers that complete-marginal formula on the no-selection face where survivor-complier status coincides with complier status and the two capacity totals agree. Away from that face, the same threshold logic applies to selected-complier submeasures generated by noncompliance and treatment-induced selection. Related ordinal-benefit and probability-of-causation analyses study graphical, interventional, or unconfoundedness-based restrictions [Tian and Pearl, 2000, Sachs et al., 2023, Duarte et al., 2023, Gabriel et al., 2024, de Aguas et al., 2025, Chen and Li, 2026]; the contribution here is the sharp survivor-complier IV-selection interval with endpoint-attaining latent laws.

The constructive proof is also an algorithm. The strict-benefit graph $i < j$ and its complement have nested threshold structure on a finite ordered support. A sparse threshold-flow routine computes all cellwise endpoint values and positive allocation traces in $O(|\mathcal{X}|K)$ operations. Writing full lower and upper $K \times K$ coupling matrices adds the expected $O(|\mathcal{X}|K^2)$ materialization cost. A three-level witness makes the calculation explicit: one covariate cell, three ordered outcomes, full compliance, and treatment-increased selection produce capacities $(0.075, 0.10, 0.075)$ and $(0.10, 0.25, 0.15)$, exact survivor-complier mass $1/4$, and the sharp interval $[0, 0.7]$.

The estimation and inference results treat the sharp interval as a finite-dimensional functional of the observed-data law. The estimator replaces observed probabilities with empirical analogues, projects raw capacities onto the nonnegative cone, screens cells by empirical survivor-complier mass, and evaluates the same threshold formulas. At a fixed observed law, under independent sampling, overlap, the structural restrictions, positive aggregate survivor-complier mass, and a screening threshold η_n with $\eta_n \to 0$ and $\sqrt{n}\,\eta_n \to \infty$, the retained support is recovered with probability tending to one, the plug-in endpoints are consistent, and the scaled endpoint error converges to the Hadamard directional derivative of the reduced-support endpoint map applied to the centered multinomial Gaussian limit. This accommodates simultaneous active threshold cuts and zero selected-complier capacity gaps through the directional derivative framework for nonsmooth maps [van der Vaart, 1998, Fang and Santos, 2019, Hong and Li, 2018].

For finite-sample containment, the paper gives a deterministic guard. Over a finite event collection determined by $\mathcal{X}$ and K, a union-bound deviation threshold yields an explicit capacity-and-screening envelope. Under fixed instrument overlap and a uniform lower bound $M \geq m_\star > 0$ on aggregate survivor-complier mass, padding the screened plug-in endpoints by this envelope produces a guarded interval that contains the entire sharp identified interval with probability at least $1 - \alpha$, uniformly over the indexed finite-slate law family and every sample size. The guard radius is $O(n^{-1/2} + \eta_n)$, and the choice $\eta_n = n^{-1/2} L_n$ with $L_n \to \infty$ and $L_n = o(\sqrt{n})$ gives radius $O(n^{-1/2} L_n)$. The technical appendix records supporting derivations and cited inputs.

The paper proceeds as follows. Section 2 discusses the related econometric literature. Section 3 introduces the potential-outcome setup, the structural restrictions, and the observable capacity identities. Section 4 states the exact-mass partial-transport representation, the threshold-cut endpoint formulas, the endpoint-attaining law construction, and the no-selection reduction. Section 5 gives the sparse threshold-flow algorithm and the three-level witness. Section 6 develops the projected screened estimator, the fixed-law directional limit, and the deterministic guarded confidence interval. Section 7 records limitations and future work, and the technical appendix collects supporting derivations.

2 Related work

This paper works in a potential-outcome setting with a binary instrument, treatment-induced selection, and a finite ordered outcome. Its target combines three strands of econometric work: instrumental variables and principal strata, sample selection bounds, and sharp bounds for ordinal benefit probabilities.

The instrumental-variables component follows the local-treatment-effect tradition in which a binary instrument shifts treatment take-up and identifies causal information for compliers [Imbens and Angrist, 1994, Angrist et al., 1996]. Principal stratification gives the language for conditioning on joint potential statuses, including compliance and post-treatment selection strata [Frangakis and Rubin, 2002]. Partial-identification results for IV models, including sharp bounds under latent response restrictions, provide the surrounding identification perspective [Balke and Pearl, 1997, Manski, 1990, 1997, Manski and Pepper, 2000, Manski, 2003]. The present analysis uses that tradition to study survivor compliers, a principal stratum defined jointly by compliance behavior and selection under treatment states, and asks for sharp bounds on the probability that treatment strictly improves an ordered outcome among that group.

Selection is central because the ordered outcome is observed on selected units. Classical sample-selection models motivate the econometric concern that outcome comparisons among selected units mix treatment effects with selection effects [Heckman, 1979]. Lee-style trimming bounds give a transparent solution under monotone selection [Lee, 2009], and subsequent work develops this approach for instrumental variables and principal strata in settings with attrition or post-treatment selection [Chen and Flores, 2015, Blanco et al., 2013, Chen et al., 2018, Kennedy et al., 2019]. The closest selection-bounds comparison is Dong and Heiler [2026], who study covariate-specific weak selection monotonicity and selected-complier outcome distributions. Closest on the benefit side under selection is Possebom and Riva [2025], who bound the probability of causation for a *binary* outcome when selection determines whether the outcome is recorded, and who identify their target on the always-observed stratum. The present paper differs on both axes that generate its mathematics: the outcome is ordered with $K \geq 3$ levels, so the target is a same-unit comparison $Y_1 > Y_0$ whose sharp value is a coupling problem rather than a pair of marginal probabilities; and the conditioning stratum is the survivor compliers, whose two arms carry *unequal* capacity totals when treatment shifts selection. That inequality is what forces the exact-mass partial-transport formulation below: the two arms must be coupled subject to a common exact-mass constraint, so the sharp value solves a transport problem rather than following from a pair of marginal probabilities. The present paper uses the same type of covariate-specific selection direction as an input to a sharp ordinal-benefit analysis for survivor compliers.

The ordinal outcome component connects to work on distribution-free bounds for functions of paired potential outcomes. Makarov-type bounds characterize feasible joint distributions given marginal distributions [Makarov, 1982, Williamson and Downs, 1990], and optimal-transport formulations give a useful finite-dimensional view of the same coupling problem [Villani, 2009, Fan and Park, 2010]. For ordered categorical outcomes, the benefit probability is a functional of the unobserved joint distribution of two potential ordered outcomes; complete marginal information yields closed-form formulas for this functional [Lu et al., 2018, 2020]. This paper adapts that coupling logic to the selected-complier capacities generated by an IV model with treatment-induced selection: within each covariate cell, the relevant margins are selected-complier subprobability capacities with potentially unequal totals.

Recent work also develops sharp ordinal benefit bounds under other observable restrictions. Gabriel et al. [2024] study sharp ordinal-benefit bounds including response-function IV restrictions, and de Aguas et al. [2025] establish sharp bounds for tiered benefit under monotonicity with inference. Related causal-graphical analyses of probability-of-causation and benefit-type estimands include Tian and Pearl [2000], Sachs et al. [2023], Duarte et al. [2023]. Chen and Li [2026] study principal generalized causal effects under treatment unconfoundedness and ignorability conditions. The contribution here is a sharp identified interval for a survivor-complier strict-benefit probability under the paper's IV, consistency, overlap, no-defiers, weak-selection-monotonicity, and positive survivor-mass conditions, together with a finite-dimensional selected-complier capacity and partial-transport representation used for computation and inference.

The inference part of the paper belongs to the econometric literature on partially identified parameters and directionally nonregular maps. Confidence procedures for identified sets motivate coverage statements for an interval-valued target [Imbens and Manski, 2004, Chernozhukov et al., 2007, Romano and Shaikh, 2010, Tamer, 2010, Molinari, 2020, Chernozhukov et al., 2013]. Directional differentiability and bootstrap validity for nonsmooth functionals provide the local asymptotic framework [van der Vaart, 1998, Fang and Santos, 2019, Hong and Li, 2018]. Recent contributions on inference for bounds and partially identified causal parameters, including Russell [2021], Levis et al. [2025], Ben-Michael [2025], Khandamiryan and Semenova [2026], Noack [2026], and Lee and Liu [2024], frame the inferential goal: to report uncertainty for a set identified by observable distributions and maintained structural restrictions. Here that goal is implemented through a guarded finite-sample construction tailored to the threshold-cut representation of the survivor-complier benefit interval.

3 Setup and assumptions

Throughout the paper, covariate support is finite and sums over covariate cells range over $\mathcal{X}$. Outcome levels are ordered as $0, \ldots, K-1$, and indices i, j, t range over this ordered support unless stated otherwise. Generic constants are positive and finite, Euclidean norms are used for finite-dimensional vectors, and asymptotic notation is always along the sampling sequence introduced in the inference section.

We begin with the potential-outcome structure. The observed system has covariates X, a binary instrument Z, treatment receipt D, selection S, and an ordered outcome Y observed on selected units. The model also fixes the latent probability law P, the finite covariate carrier $\mathcal{X}$, the outcome-support size K, and the ordered outcome set $\mathcal{Y}$.

The section proceeds from variables to restrictions to capacities: the slate fixes the variables, the consistency and exclusion restrictions connect observed cells to potential outcomes, monotonicity identifies the complier and survivor-complier components, and the observable-capacity definition packages the two selected-complier submeasures used by the transport problem.

Definition 1. [synth_13] (POSlate system structure) *A POSlate system over a potential-outcome system P, a measurable covariate carrier $\mathcal{X}$, and an outcome-support size K consists of:*

- ***(Latent space.)*** *The latent probability space Ω of P is standard Borel.*
- ***(Outcome support.)*** *The integer K satisfies $K \geq 3$, and the ordered outcome support is $\{0, \ldots, K-1\}$.*
- ***(Selected nodes.)*** *Five pairwise distinct nodes of P, written X, Z, D, S, Y.*
- ***(Value spaces.)*** *The node X is measurably identified with $\mathcal{X}$, while Z, D, and S are measurably identified with $\{0, 1\}$.*
- ***(Outcome node.)*** *The node Y is measurably identified with the finite ordered support $\{0, \ldots, K-1\}$.*

Here Z is the binary instrument, X is the covariate, D is treatment receipt, S is selection, and Y is the finite ordered outcome.

Potential treatments, selections, and outcomes are indexed by the intervention that defines them. The treatment potentials D_0, D_1 vary the instrument, while the selection potentials S_0, S_1 and outcome potentials Y_0, Y_1 vary treatment receipt.

Definition 2. [synth_11] (Control potential outcome) *The symbol Y_0, the potential ordered outcome under treatment zero, is the map from the sample space to the K ordered outcome levels obtained by setting treatment to zero:*

$$Y_0(\omega) = Y^{D \leftarrow 0}(\omega) \in \{0, \ldots, K-1\}.$$

Definition 3. [synth_10] (Selection under treatment one) *The potential selection indicator under treatment one is the map $S_1 : \Omega \to \{0, 1\}$ defined by*

$$S_1(\omega) := S(\omega; D \leftarrow 1).$$

Definition 4. [synth_9] (Treatment-zero selection) *Define the potential selection indicator under treatment zero by*

$$S_0(\omega) := S^{D \leftarrow 0}(\omega).$$

Equivalently, S_0 is the selection variable evaluated counterfactually when the treatment variable is set to 0.

Definition 5. [synth_8] (Potential treatment under instrument one) *Let D_1, the potential treatment under instrument state one, be the Boolean unit-level variable on Ω given by the treatment value that would be realized under the intervention $Z = 1$. Equivalently, for each unit $\omega \in \Omega$,*

$$D_1(\omega) = D_{Z \leftarrow 1}(\omega).$$

Definition 6. [synth_7] (Potential treatment under zero) *The potential treatment under instrument value zero is the binary variable D_0 defined by*

$$D_0(\omega) = D(\omega; Z = 0), \qquad \omega \in \Omega.$$

Definition 7. [synth_12] (Treatment-one outcome) *The potential ordered outcome under treatment one is the map*

$$Y_1 : \Omega \to \{0, \ldots, K-1\}$$

given by evaluating the ordered outcome counterfactual under the intervention that sets treatment receipt to 1*.*

The first group of identifying restrictions is the standard IV structure with treatment-mediated selection and outcomes. It separates instrument assignment from latent response types within covariate cells, and it connects the observed variables to their potential-outcome counterparts.

Assumption 1. [ass:iv-independence] (Instrument independence) *The potential treatments D_0, D_1, potential selection indicators S_0, S_1, and potential ordered outcomes Y_0, Y_1 are conditionally independent of the binary instrument Z given the covariate cell X:*

$$(D_0, D_1, S_0, S_1, Y_0, Y_1) \perp Z \mid X.$$

Assumption 1 is the conditional instrument independence condition: after conditioning on X, variation in Z is as good as randomly assigned for the potential treatments, selections, and ordered outcomes [Angrist et al., 1996].

Assumption 2. [ass:treatment-consistency] (Treatment consistency) *Observed treatment receipt equals the potential treatment under the realized instrument state:*

$$D = D_Z$$

almost surely.

Assumption 2 is the treatment consistency condition, which links realized receipt to the potential treatment indexed by the realized instrument [Angrist et al., 1996]. Together with independence, it gives observed first-stage contrasts their usual complier interpretation.

Assumption 3. [ass:selection-exclusion-consistency] (Selection consistency) *Observed selection equals the potential selection indicator under the realized treatment:*

$$S = S_D$$

almost surely.

Assumption 3 is the selection exclusion through received treatment condition: instrument assignment affects selection through the treatment actually received [Kennedy et al., 2019].

Assumption 4. [ass:outcome-exclusion-consistency] (Outcome consistency) *Among selected units, the observed outcome equals the potential ordered outcome under the realized treatment:*

$$Y = Y_D$$

on $\{S = 1\}$ almost surely.

Assumption 4 is the outcome exclusion through received treatment condition on selected units [Kennedy et al., 2019]. Because outcomes are observed when $S = 1$, this condition fixes the outcome content of selected observed cells.

The observed-data law records the variables available to the analyst, with the outcome marked missing outside the selected sample. We write the observed tuple as O and its law as P_{obs}.

Definition 8. [synth_17] (Observed finite data law) *The observed-data map sends each latent unit ω to the finite observed cell*

$$O(\omega) = \big(X(\omega), Z(\omega), D(\omega), S(\omega), \widetilde{Y}(\omega)\big), \qquad \widetilde{Y}(\omega) = \begin{cases} Y(\omega), & S(\omega) = 1, \\ \varnothing, & S(\omega) = 0. \end{cases}$$

The observed-data law P_{obs} is the pushforward of the latent law P through this observed-data map:

$$P_{\text{obs}} = P \circ O^{-1}.$$

Cell weights enter every aggregate target below. The covariate-cell probability p_x is defined before the complier and survivor-complier quantities are formed.

Definition 9. [synth_14] (Covariate cell mass) *For each covariate cell x, the cell probability is*

$$p_x := P\{\omega : X(\omega) = x\}.$$

Definition 10. [synth_1] (Complier event definition) *The complier event C is the set of units whose potential treatment equals 0 when $Z = 0$ and equals 1 when $Z = 1$:*

$$C := \{\omega : D_0(\omega) = 0 \wedge D_1(\omega) = 1\}.$$

The complier event C is the latent stratum whose treatment receipt is shifted by the instrument. The benefit parameter studied later is restricted further to compliers who would be selected under both treatment states.

For compact reference, the following definition collects the finite slate, the observed law, the monotonicity structure, and the cellwise capacity objects used throughout the rest of the paper.

Definition 11. [synth_22] (Slate-benefit partial-transport system) *A slate-benefit partial-transport system is a POSlate potential-outcome structure with finite discrete covariate support $\mathcal{X}$, ordered outcome support $\mathcal{Y} = \{0, 1, \dots, K-1\}$ for $K \geq 3$, binary instrument Z, binary treatment D, selection indicator S, potential treatments D_0, D_1, potential selections S_0, S_1, and potential outcomes Y_0, Y_1. The observed variables satisfy $D = D_Z$, $S = S_D$, and $Y = Y_D$ on $\{S = 1\}$, with observed element $O = (X, Z, D, S, SY)$ and observed law $P_{\text{obs}} = \mathcal{L}(O)$. The system carries a latent law P under which $(D_0, D_1, S_0, S_1, Y_0, Y_1) \perp Z \mid X$, $D_1 \geq D_0$ almost surely, and for each cell with $P(C \mid X = x) > 0$, where $C = \{D_1 > D_0\}$, a direction $d(x) \in \{-1, +1\}$ satisfies $P\{d(x)(S_1 - S_0) \geq 0 \mid C, X = x\} = 1$. Its cellwise observable capacities are $\ell_i(x)$ and $h_j(x)$, with $m(x) = \min\{\sum_i \ell_i(x), \sum_j h_j(x)\}$, aggregate survivor-complier mass $M = \sum_x p_x m(x) > 0$, and survivor-complier outcome subcouplings constrained by the exact-mass partial-transport sets Γ_x^*.*

The remaining assumptions impose overlap and monotonicity. The instrument propensity $\pi(x)$ and the overlap constant ε_Z determine where conditional instrument contrasts are well defined.

Definition 12. [synth_16] (Instrument propensity by cell) *For each covariate cell x, the conditional instrument propensity $\pi(x)$ is the real-valued conditional probability of $Z = 1$ in that cell:*

$$\pi(x) = \begin{cases} \dfrac{\mathbb{P}(Z = 1,\ X = x)}{\mathbb{P}(X = x)}, & \text{if } \mathbb{P}(X = x) > 0, \\ 0, & \text{if } \mathbb{P}(X = x) = 0. \end{cases}$$

Assumption 5. [ass:instrument-overlap] (Instrument propensity overlap) *For a slate-benefit partial-transport system, the instrument-overlap condition at the instrument-overlap constant ε_Z consists of:*

- ***(Positivity.)*** $0 < \varepsilon_Z$.
- ***(Half-overlap.)*** $\varepsilon_Z < 1/2$.
- ***(Cellwise overlap.)*** *For every covariate cell $x \in \mathcal{X}$ with covariate-cell probability $p_x > 0$, the conditional instrument propensity $\pi(x)$ satisfies*

$$\varepsilon_Z \leq \pi(x) \leq 1 - \varepsilon_Z.$$

Assumption 5 is the conditional instrument overlap condition [Kennedy et al., 2019]. It keeps both instrument states present in every supported covariate cell, so the observed conditional contrasts used to form capacities are defined cell by cell.

Assumption 6. [ass:no-defiers] (Treatment monotonicity) *Potential treatment receipt is monotone in the instrument:*

$$D_1 \geq D_0$$

almost surely.

Assumption 6 is the instrument monotonicity condition [Angrist et al., 1996]. It orients the first stage so that the shifted latent group is the complier stratum in Definition 10.

Assumption 7. [ass:weak-selection-monotonicity] (Weak selection monotonicity) *For every covariate cell x with $P(C \mid X = x) > 0$, the covariate-specific direction $d(x) \in \{-1, +1\}$ satisfies*

$$P\{d(x)(S_1 - S_0) \geq 0 \mid C, X = x\} = 1.$$

Assumption 7 is the covariate-specific weak sample-selection monotonicity condition of Dong and Heiler [2026, Assumption 3.2]. The direction $d(x)$ is allowed to vary across covariate cells, and the condition makes the sign of the selected-complier selection contrast interpretable within cells that contain compliers.

Assumption 8. [ass:positive-aggregate-survivors] (Positive aggregate survivors) *The aggregate survivor-complier mass satisfies*

$$M > 0.$$

Assumption 8 is the positive target-stratum probability condition [Kennedy et al., 2019]. It ensures that the aggregate survivor-complier mass M can normalize the benefit probability studied in the next section.

The maintained structural law class $\mathcal{M}$ is the collection of latent laws satisfying the restrictions just stated.

Definition 13. [`def:structural-law-class`] (Structural law class $\mathcal{M}$) *The structural law class is*

$$\mathcal{M} = \{P : P \textit{ satisfies Assumptions } 1 \textit{to } 8\}.$$

This is the finite-ordered, finite-covariate specialization of the corresponding Dong–Heiler cellwise restrictions on cells with $P(C \mid X = x) > 0$. Cells with $P(C \mid X = x) = 0$ have zero capacities and are inert for the target.

The observable selected-complier capacities are the bridge from the observed law to the partial-transport problem. For each cell, $\ell_i(x)$ records lower-arm selected-complier mass at outcome level i, while $h_j(x)$ records upper-arm selected-complier mass at level j. Their totals $q_0(x)$ and $q_1(x)$, their difference $\Delta q(x)$, and their exact survivor-complier mass $m(x)$ determine the feasible cellwise coupling sets introduced in the next section.

Definition 14. [`def:observable-capacities`] (Observable capacities $\ell_i(x), h_j(x)$) *For each $x \in \mathcal{X}$ and $i, j \in \mathcal{Y}$, let*

$$\ell_i(x)$$

and

$$h_j(x)$$

be the lower-arm and upper-arm selected-complier outcome capacities obtained from the corresponding conditional instrument contrasts. Define the capacity totals, the capacity-total gap, and the exact survivor-complier mass by

$$q_0(x) = \sum_{i \in \mathcal{Y}} \ell_i(x), \qquad q_1(x) = \sum_{j \in \mathcal{Y}} h_j(x),$$

$$\Delta q(x) = q_1(x) - q_0(x), \qquad m(x) = \min\{q_0(x), q_1(x)\}.$$

The section closes with the identification statement for these capacities. It converts the observed IV contrasts into latent selected-complier submeasures and identifies the exact survivor-complier mass cell by cell.

Theorem 1. [`prop:capacity-identification`] (Capacity identification) *Let S be a five-node potential-outcome slate with covariate X, binary Z, D, S, ordered outcome support $\mathcal{Y} = \{0, \ldots, K-1\}$ with $K \geq 3$, and standard-Borel latent probability space. Let ε_Z be the instrument-overlap constant, and let $d : \mathcal{X} \to \{+1, -1\}$ be the covariate-specific weak selection-monotonicity direction. Suppose that:*

- ***(IV independence.)*** *Assumption 1 holds.*
- ***(Treatment consistency.)*** *Assumption 2 holds.*
- ***(Selection exclusion and consistency.)*** *Assumption 3 holds.*
- ***(Outcome exclusion and consistency.)*** *Assumption 4 holds.*
- ***(Instrument overlap.)*** *Assumption 5 holds with constant ε_Z.*
- ***(No defiers.)*** *Assumption 6 holds.*
- ***(Weak selection monotonicity.)*** *Assumption 7 holds with direction d.*

Let $\ell_i(x)$, $h_j(x)$, $\Delta q(x)$, and $m(x)$ be the observable capacity contrasts formed from the observed-data law. Then the observable capacities are nonnegative in every covariate cell,

$$\ell_i(x) \geq 0 \quad \text{and} \quad h_j(x) \geq 0 \qquad \text{for every } x \in \mathcal{X} \text{ and every } i, j \in \mathcal{Y},$$

all covariate-cell probabilities satisfy $p_x \geq 0$, and, for every supported cell x with $p_x > 0$,

$$\ell_i(x) = P(Y_0 = i, S_0 = 1, C \mid X = x)$$

for every $i \in \mathcal{Y}$,

$$h_j(x) = P(Y_1 = j, S_1 = 1, C \mid X = x)$$

for every $j \in \mathcal{Y}$,

$$\Delta q(x) = P(S_1 = 1, C \mid X = x) - P(S_0 = 1, C \mid X = x),$$

and

$$m(x) = P(S_0 = 1, S_1 = 1, C \mid X = x).$$

Moreover, for every supported x, $\Delta q(x) > 0$ implies $d(x) = +1$, $\Delta q(x) < 0$ implies $d(x) = -1$, and $\Delta q(x) = 0$ makes both directions observationally admissible under the observed-data law.

The intuition is that the IV independence and consistency restrictions express selected outcome contrasts by latent complier components, while treatment monotonicity fixes which treatment response type contributes to the contrast. Weak selection monotonicity then aligns the selection contrast so that the smaller selected-complier total is exactly the always-selected complier mass in that cell. Consequently, Theorem 1 supplies the two nonnegative marginal submeasures that feed the exact-mass partial-transport construction below.

4 Sharp benefit bounds

The capacity identification result in Theorem 1 reduces the target parameter to a finite collection of cellwise coupling problems. In each covariate cell, the lower-arm and upper-arm selected-complier outcome capacities provide two submarginals, while the exact survivor-complier mass fixes how much of these capacities must be paired. The resulting object is the exact-mass partial-transport polytope Γ_x^*, whose elements are cellwise survivor-complier coupling arrays γ_x. This is the finite ordinal analogue of a partial transport problem with capacity constraints [Villani, 2009].

Definition 15. [`def:partial-transport-polytope`] (Partial-transport polytopes Γ_x^*, Γ_x) *For each $x \in \mathcal{X}$, define*

$$\Gamma_x^* = \left\{ \gamma_x \geq 0 : \sum_{j \in \mathcal{Y}} \gamma_{ij,x} \leq \ell_i(x), \ \sum_{i \in \mathcal{Y}} \gamma_{ij,x} \leq h_j(x), \ \sum_{i,j \in \mathcal{Y}} \gamma_{ij,x} = m(x) \right\}.$$

The comparison polytopes are

$$\Gamma_x^+ = \left\{ \gamma_x \geq 0 : \sum_{j \in \mathcal{Y}} \gamma_{ij,x} = \ell_i(x), \ \sum_{i \in \mathcal{Y}} \gamma_{ij,x} \leq h_j(x) \right\},$$

$$\Gamma_x^- = \left\{ \gamma_x \geq 0 : \sum_{j \in \mathcal{Y}} \gamma_{ij,x} \leq \ell_i(x), \ \sum_{i \in \mathcal{Y}} \gamma_{ij,x} = h_j(x) \right\},$$

and

$$\Gamma_x^0 = \left\{ \gamma_x \geq 0 : \sum_{j \in \mathcal{Y}} \gamma_{ij,x} = \ell_i(x), \ \sum_{i \in \mathcal{Y}} \gamma_{ij,x} = h_j(x) \right\}.$$

The cellwise survivor-complier coupling polytope is

$$\Gamma_x = \Gamma_x^*.$$

The comparison polytopes in Definition 15 make explicit which side of the two selected-complier capacity totals binds. When the upper-arm total exceeds the lower-arm total, the row constraints can bind; when the lower-arm total exceeds the upper-arm total, the column constraints can bind. The branch-free notation Γ_x records that the paper's identified coupling set is always the exact-mass set, with the operative branch determined directly by the observed capacity totals.

The endpoint calculations use ordinal threshold cuts. The lower endpoint is obtained by forcing enough mass onto weakly adverse pairs to limit strict benefit, while the upper endpoint packs as much mass as possible onto pairs with treatment-one outcome above treatment-zero outcome. The finite support lets these constraints be summarized by prefixes and tails, in the spirit of sharp distributional bounds for ordinal and discrete outcomes [Makarov, 1982, Williamson and Downs, 1990, Fan and Park, 2010].

Definition 16. [`def:threshold-cuts`] (Threshold cuts $B_L(x), B_U(x)$) *For every $x \in \mathcal{X}$, define the prefix and tail sums*

$$\ell_{\leq t}(x) = \sum_{i \leq t} \ell_i(x), \qquad h_{\leq t}(x) = \sum_{j \leq t} h_j(x),$$

$$\ell_{<t}(x) = \sum_{i < t} \ell_i(x), \qquad h_{>t}(x) = \sum_{j > t} h_j(x).$$

The lower and upper threshold-cut values are

$$B_L(x) = \max\left\{0, \ \max_{t \in \mathcal{Y}} \left[\ell_{\leq t}(x) - h_{\leq t}(x)\right] + \min\{\Delta q(x), 0\}\right\}$$

and

$$B_U(x) = \min\left\{m(x), \ \min_{t \in \mathcal{Y}} \left[\ell_{<t}(x) + h_{>t}(x)\right]\right\}.$$

These formulas apply on positive-survivor zero-gap cells and on zero-survivor cells.

The lower threshold-cut value $B_L(x)$ and upper threshold-cut value $B_U(x)$ are unnormalized benefit-mass bounds within a cell. Their formulas combine the ordinal cut constraints with the exact survivor-complier mass, including zero-gap cells where the two capacity totals coincide.

Definition 17. [`def:identified-interval`] (Identified interval $\Theta_I(P_{\text{obs}})$) *The endpoint map is*

$$\Psi(P_{\text{obs}}) = (\theta_L, \theta_U) = \left(\frac{\sum_{x \in \mathcal{X}} p_x B_L(x)}{M}, \frac{\sum_{x \in \mathcal{X}} p_x B_U(x)}{M} \right).$$

The sharp identified interval for the benefit probability is

$$\Theta_I(P_{\text{obs}}) = [\theta_L, \theta_U].$$

Definition 17 aggregates the cellwise cut values into the endpoint map $\Psi(P_{\text{obs}})$. The lower endpoint θ_L, upper endpoint θ_U, and interval $\Theta_I(P_{\text{obs}})$ are normalized by the aggregate survivor-complier mass from Assumption 8, so the estimand is the fraction benefiting among survivor compliers.

Definition 18. [`synth_4`] (Threshold-cut capacity aggregates) *For each covariate cell $x \in \mathcal{X}$ and threshold $t \in \mathcal{Y} = \{0, \dots, K-1\}$, define the lower-arm prefix, strict-prefix, and upper-tail capacities by $\ell_{\le t}(x) = \sum_{i \le t} \ell_i(x)$, $\ell_{<t}(x) = \sum_{i<t} \ell_i(x)$, and $\ell_{>t}(x) = \sum_{i>t} \ell_i(x)$. Define the corresponding higher-arm prefix and upper-tail capacities by $h_{\le t}(x) = \sum_{j \le t} h_j(x)$ and $h_{>t}(x) = \sum_{j>t} h_j(x)$.*

The aggregate notation in Definition 18 records the prefix and tail capacities used by the threshold formulas. The additional upper tail $\ell_{>t}(x)$, together with $\ell_{\le t}(x)$, $\ell_{<t}(x)$, $h_{\le t}(x)$, and $h_{>t}(x)$, gives the algebraic identities needed to compare the exact-mass problem with complete-marginal ordinal bounds.

Theorem 2. [`prop:tie-face-collapse`] (Tie face collapse) *Let P be a five-node potential-outcome system as in Definition 13, with ordered outcome support of size $K \ge 3$. Fix an instrument-overlap constant ε_Z and a covariate-specific weak selection-monotonicity direction $d(x)$. Assume:*

- ***(Structural restrictions.)*** *The law satisfies Assumptions 1 to 8, with instrument-overlap constant ε_Z and weak selection-monotonicity direction $d(x)$.*
- ***(Observable capacities.)*** *Let c be the observable capacity vector formed from the observed-data law P_{obs} as in Definition 14, and let its validity be the validity delivered by Theorem 1.*

Then, for every covariate cell x with $p_x > 0$,

$$\left(\Gamma_x^+ = \Gamma_x^- \text{ and } \Gamma_x^- = \Gamma_x^0\right) \quad \Longleftrightarrow \quad \Delta q(x) = 0.$$

For every such cell x, if $\Delta q(x) = 0$, then

$$P(S_0 = S_1,\ C \mid X = x) = P(C \mid X = x).$$

Under the same zero-gap condition,

$$\Gamma_x = \Gamma_x^+, \qquad \Gamma_x = \Gamma_x^-, \qquad \Gamma_x = \Gamma_x^0.$$

Finally, for every such cell x, if $\Delta q(x) = 0$, then for every threshold t,

$$\ell_{\le t}(x) - h_{\le t}(x) = h_{>t}(x) - \ell_{>t}(x).$$

Theorem 2 characterizes the zero-gap face of the partial-transport problem. At $\Delta q(x) = 0$, the exact survivor-complier mass equals both selected-complier capacity totals, so the row-exact, column-exact, and doubly exact comparison polytopes coincide with the maintained coupling polytope. The threshold identity then expresses the same balance through prefixes and tails, which lets implementations evaluate zero-gap cells using the branch-free formulas in Definition 16.

The next construction turns the threshold formulas into full latent laws. Cellwise, it first builds extremal couplings for the strict-benefit and nonbenefit graphs; globally, it embeds those couplings into potential-outcome laws that retain the observed law and the structural restrictions.

Algorithm 1 (Threshold-flow construction P^L, P^U)

The inputs are the observable cellwise capacities $\ell_i(x)$ and $h_j(x)$, the capacity totals $q_0(x)$ and $q_1(x)$, and the threshold-cut values $B_L(x)$ and $B_U(x)$.

1. For each $x \in \mathcal{X}$, compute the exact mass and the threshold sums:

$$m(x) = \min\{q_0(x), q_1(x)\},$$

$$\ell_{\le t}(x) = \sum_{i \le t} \ell_i(x), \qquad h_{\le t}(x) = \sum_{j \le t} h_j(x),$$

$$\ell_{<t}(x) = \sum_{i<t} \ell_i(x), \qquad h_{>t}(x) = \sum_{j>t} h_j(x).$$

2. Compute the threshold-cut values

$$B_L(x) = \max\left\{0,\ \max_{t \in \mathcal{Y}} \big[\ell_{\le t}(x) - h_{\le t}(x)\big] + \min\{q_1(x) - q_0(x), 0\}\right\}$$

and

$$B_U(x) = \min\left\{m(x),\ \min_{t \in \mathcal{Y}} \big[\ell_{<t}(x) + h_{>t}(x)\big]\right\}.$$

3. For the upper endpoint, construct a maximum subflow on the nested benefit graph $i < j$. Extend it to total mass $m(x)$ by allocating the residual mass on the complete bipartite graph.

4. For the lower endpoint, construct a maximum subflow on the nested nonbenefit graph $j \le i$. Extend it to total mass $m(x)$ by allocating the residual mass on the complete bipartite graph.

5. The resulting coupling in each cell satisfies

$$\gamma_x \in \Gamma_x^*.$$

6. Allocate row or column residual mass to the one-sided selected-complier stratum determined by the sign of

$$q_1(x) - q_0(x).$$

Place the remaining complier mass in the never-selected stratum and adjoin the unchanged never-taker and always-taker components of a baseline compatible law.

7. The construction outputs endpoint-attaining latent laws P^L and P^U with endpoint map

$$\Psi(P_{\mathrm{obs}}) = \left(\frac{\sum_{x\in\mathcal{X}} p_x B_L(x)}{M}, \frac{\sum_{x\in\mathcal{X}} p_x B_U(x)}{M} \right).$$

The lower endpoint law P^L and upper endpoint law P^U are constructed cell by cell and then completed with the compliance and selection strata. The residual-mass step aligns the partial-transport coupling with the weak selection-monotonicity direction from Assumption 7, while the unchanged never-taker and always-taker components preserve the observed IV law.

Theorem 3. [`thm:full-law-endpoint-attainment`] (Endpoint law attainment) *Let P_{obs} be the observed-data law induced by a five-node potential-outcome slate system with covariates X, binary instrument Z, treatment D, selection S, and ordered outcome $Y \in \{0, \dots, K-1\}$, where $K \geq 3$. Let $d : \mathcal{X} \to \{-1, +1\}$ encode the covariate-specific weak selection-monotonicity direction, with $d(x) = +1$ corresponding to $S_0 \leq S_1$ and $d(x) = -1$ corresponding to $S_1 \leq S_0$. Let ε_Z be the instrument-overlap constant. Suppose that:*

- ***(Structural restrictions.)*** *The latent slate system satisfies Assumption 1, Assumption 6, and Assumption 7.*
- ***(Tie-safe survivor model.)*** *The same law satisfies Assumption 2, Assumption 3, Assumption 4, Assumption 5 at ε_Z, Assumption 6, and Assumption 7 in direction d.*
- ***(Positive aggregate mass.)*** *For the observable capacity vector c formed from P_{obs} as in Definition 14, the cell probabilities p_x, and the survivor-complier masses $m(x)$,*

$$M = \sum_x p_x m(x) > 0.$$

Then there exists a baseline law compatible with P_{obs} and c: it induces P_{obs}, has observable capacities c, has conditional complier mass in every cell at least $\max\{q_0(x), q_1(x)\}$, and carries the same never-taker and always-taker components used by the latent completion. For this baseline, the cellwise threshold-flow construction in Algorithm 1 returns two full laws P^L and P^U over $(Y_0, Y_1, S_0, S_1, D_0, D_1, X, Z)$. These laws are endpoint witnesses for the endpoint map $\Psi(P_{\mathrm{obs}}) = (\theta_L, \theta_U)$: each induces exactly P_{obs}, satisfies the structural restrictions with instrument-overlap constant ε_Z and direction d, and attains

$$P^L(Y_1 > Y_0 \mid S_1 = S_0 = 1, C) = \theta_L, \qquad P^U(Y_1 > Y_0 \mid S_1 = S_0 = 1, C) = \theta_U.$$

Moreover, for every covariate cell x, P^L realizes the lower threshold latent completion generated from the baseline at x, and P^U realizes the upper threshold latent completion generated from the baseline at x.

Theorem 3 establishes that the cellwise extremal couplings are attainable by full potential-outcome laws. The construction first preserves the observed capacities identified in Theorem 1, then assigns the unmatched selected-complier capacity to the one-sided selected stratum allowed by weak selection monotonicity. Because the theorem attains the two endpoint probabilities under the maintained structural restrictions, the threshold formulas are realized by feasible full latent laws.

The exact-mass formulation also contains the familiar complete-marginal ordinal benefit problem on the no-selection face. In that submodel, survivor-complier status coincides with complier status in every relevant cell, so the two selected-complier capacity totals agree and the partial-transport polytope becomes a doubly exact coupling set.

Theorem 4. [`thm:no-selection-reduction`] (No-selection reduction)[1] *Let the ordered outcome support have $K \geq 3$ levels, and let the observed capacities c be formed from the observed-data law P_{obs} as in Theorem 1. Suppose that:*

- ***(Structural law.)*** *The latent law satisfies the structural restrictions in Definition 13, including Assumption 1 and Assumption 6, with instrument-overlap constant ε_Z and weak selection-monotonicity direction $d(x)$.*
- ***(No selection submodel.)*** *For every covariate cell x with $P(C \mid X = x) > 0$,*

$$P(S_0 = S_1 = 1 \mid C, X = x) = 1.$$

Then, for every covariate cell x with $p_x > 0$,

$$\Delta q(x) = 0 \quad \text{and} \quad \Gamma_x = \Gamma_x^0.$$

Moreover, if $m(x) > 0$, then the normalized threshold-cut endpoints satisfy

$$\frac{B_L(x)}{m(x)} = \max\left\{0, \max_{t \in \mathcal{Y}} \frac{\ell_{\leq t}(x) - h_{\leq t}(x)}{m(x)}\right\}$$

and

$$\frac{B_U(x)}{m(x)} = \min\left\{1, \min_{t \in \mathcal{Y}} \frac{\ell_{<t}(x) + h_{>t}(x)}{m(x)}\right\}.$$

The normalized formulas in Theorem 4 match the closed-form strict-benefit bounds for two complete ordinal marginals in Lu et al. [2018, Proposition 2, equation (6), p. 546]. The theorem identifies the precise face of the present model where that comparison applies: every relevant complier is always selected, the capacity gap is zero, and the exact-mass polytope has both margins fixed.

It remains to name the cellwise objective optimized over the coupling set.

Definition 19. [`synth_5`] (Cellwise benefit-mass functional) *For a survivor-complier outcome subcoupling $\gamma_x \in \mathbb{R}_+^{K \times K}$ in cell x, define the unnormalized benefit mass by $b_x(\gamma_x) = \sum_{0 \leq i < j \leq K-1} \gamma_{ij,x}$. Thus b_x counts exactly the mass assigned to outcome pairs whose treatment-one outcome level strictly exceeds the treatment-zero outcome level.*

The functional $b_x(\gamma_x)$ is linear in the coupling and counts strict ordinal improvement. With this objective, the threshold cuts in Definition 16 become the exact lower and upper values of the cellwise optimization problem.

[1] **Formalization scope.** This result uses the published conclusion [Lu et al., 2018] (Proposition 2, equation (6), page 546). The cited source proof is not formalized here: Lean verifies its use and all remaining steps. If that cited conclusion is false, the portions of this result that depend on it are not certified.

Theorem 5. [thm:sharp-exact-mass-threshold-interval] (Sharp exact-mass interval) *Let P_{obs} be the observed-data law generated by a five-node potential-outcome slate with ordered outcome support of size $K \geq 3$. Suppose the latent law satisfies the structural restrictions in Definition 13 for the instrument-overlap constant ε_Z and weak-selection direction $d(x)$. Let c be the observable capacity vector from Definition 14, identified from P_{obs} as in Theorem 1, and suppose that*

$$M = \sum_{x \in \mathcal{X}} p_x m(x) > 0.$$

Then the following statements hold.

- ***(Sharp cellwise projection.)*** *For every cell x with $p_x > 0$, the set of survivor-complier couplings generated by feasible full laws with observed law P_{obs} is exactly*

$$\Gamma_x^* = \left\{ \gamma_x \geq 0 : \sum_j \gamma_{ij,x} \leq \ell_i(x),\ \sum_i \gamma_{ij,x} \leq h_j(x),\ \sum_{i,j} \gamma_{ij,x} = m(x) \right\}.$$

- ***(Exact-mass branch.)*** *For every cell x with $p_x > 0$,*

$$\Delta q(x) > 0 \Rightarrow \Gamma_x^* = \Gamma_x^+, \qquad \Delta q(x) < 0 \Rightarrow \Gamma_x^* = \Gamma_x^-, \qquad \Delta q(x) = 0 \Rightarrow \Gamma_x^* = \Gamma_x^0.$$

- ***(Threshold endpoints.)*** *For every cell x with $p_x > 0$,*

$$\inf_{\gamma_x \in \Gamma_x^*} b_x(\gamma_x) = B_L(x), \qquad \sup_{\gamma_x \in \Gamma_x^*} b_x(\gamma_x) = B_U(x).$$

  Writing γ_x^L and γ_x^U for the lower and upper cellwise threshold-flow couplings from Algorithm 1,

$$\gamma_x^L \in \Gamma_x^*, \qquad b_x(\gamma_x^L) = B_L(x), \qquad \gamma_x^U \in \Gamma_x^*, \qquad b_x(\gamma_x^U) = B_U(x).$$

- ***(Zero-mass cells.)*** *For every cell x with $m(x) = 0$,*

$$\Gamma_x^* = \{0\}.$$

Consequently, the sharp identified interval for the benefit probability over feasible full laws is

$$\Theta_I(P_{\text{obs}}) = \left[\frac{\sum_{x \in \mathcal{X}} p_x B_L(x)}{M}, \frac{\sum_{x \in \mathcal{X}} p_x B_U(x)}{M} \right].$$

Theorem 5 is the section's sharpness result. It characterizes the projection of feasible full laws onto the exact-mass partial-transport polytope, identifies the sign-specific comparison polytope selected by the capacity gap, and evaluates the lower and upper values of the strict-benefit objective by the threshold cuts. Together with endpoint law attainment in Theorem 3, these statements deliver the sharp identified interval for the survivor-complier benefit probability θ under the structural law class in Definition 13.

5 Computation and witness

The threshold-cut formulas in Definition 16 and the endpoint construction in Algorithm 1 have two practical consequences. First, the construction can be implemented cell by cell by sorting the ordered outcome levels once and moving through nested lower and upper threshold sets. Second, the same calculation produces explicit sparse couplings, so the numerical endpoint is accompanied by a witness allocation in the partial-transport polytope.

We begin with a three-level law that makes the objects in Definition 14 concrete. The example has one covariate cell, three ordered outcomes, and full compliance. Selection differs across treatment states, so the lower- and upper-arm selected-complier capacities have unequal totals before the exact survivor-complier mass is imposed.

Definition 20. [`def:three-level-witness`] (Three-level witness $P^\star_{\text{obs}}$ and $P^\star$) *The observed law $P^\star_{\text{obs}}$ and latent law $P^\star$ are defined as follows. The covariate satisfies $X = x^\star$ almost surely, the binary instrument Z satisfies*

$$P^\star_{\text{obs}}(Z=1) = P^\star_{\text{obs}}(Z=0) = \frac{1}{2},$$

and the ordered outcome has $K = 3$ levels. Conditional on $Z = 0$,

$$P^\star_{\text{obs}}(D,S,Y) = \begin{cases} 0.075, & (D,S,Y) = (0,1,0), \\ 0.10, & (D,S,Y) = (0,1,1), \\ 0.075, & (D,S,Y) = (0,1,2), \\ 0.75, & (D,S,Y) = (0,0,\star), \end{cases}$$

with all other conditional cells assigned probability zero. Conditional on $Z = 1$,

$$P^\star_{\text{obs}}(D,S,Y) = \begin{cases} 0.10, & (D,S,Y) = (1,1,0), \\ 0.25, & (D,S,Y) = (1,1,1), \\ 0.15, & (D,S,Y) = (1,1,2), \\ 0.50, & (D,S,Y) = (1,0,\star), \end{cases}$$

with all other conditional cells assigned probability zero.

The latent law $P^\star$ assigns every unit potential treatments $D_0 = 0$ and $D_1 = 1$. It allocates masses $0.075, 0.10, 0.075$ to units with potential selection indicators $S_0 = S_1 = 1$ and diagonal potential-outcome pairs

$$(Y_0, Y_1) = (0,0), (1,1), (2,2),$$

respectively. It allocates mass 0.25 to units with $S_0 = 0$ and $S_1 = 1$, with residual Y_1 masses

$$(0.025, 0.15, 0.075)$$

on outcome levels $0, 1, 2$. It allocates mass 0.50 to units with $S_0 = S_1 = 0$. The instrument Z is independent of these potential variables with probability $1/2$, and latent outcomes are assigned arbitrarily wherever selection makes them unobserved.

The observed witness law $P^\star_{\text{obs}}$ displays all observed cells needed to recover the capacity vector, while the latent witness law $P^\star$ supplies a compatible potential-outcome system. The selected always-survivor mass is diagonal in the ordered outcome, and the additional selected mass under treatment one raises the upper-arm capacities. The resulting calculation is small enough to audit directly but still contains the capacity imbalance that drives a nondegenerate sharp interval.

Theorem 6. [`prop:three-level-witness`] (Three-level witness sharpness) *For the single-cell three-level witness, let $P^\star_{\text{obs}}$ be the observed-data law assigning mass*

$$\frac{3}{80}, \frac{1}{20}, \frac{3}{80}, \frac{3}{8}, \frac{1}{20}, \frac{1}{8}, \frac{3}{40}, \frac{1}{4}$$

to the eight displayed observed cells, in order. Let c be the observable capacity vector obtained from $P^\star_{\text{obs}}$. Then the following assertions hold:

- ***(Full-law realization.)*** *There exists a latent potential-outcome system $P^\star$ with a three-level slate system such that the survivor model is tie-safe at instrument overlap $1/4$ with the single covariate cell retained, its observed law is $P^\star_{\text{obs}}$, its encoded full law is the displayed full witness law, $D_0 = 0$ and $D_1 = 1$ almost surely, and $S_0 \le S_1$ almost surely.*
- ***(Observable capacities.)*** *At the single covariate cell $x^\star$,*

$$(\ell_0(x^\star), \ell_1(x^\star), \ell_2(x^\star)) = \left(\frac{3}{40}, \frac{1}{10}, \frac{3}{40}\right), \qquad (h_0(x^\star), h_1(x^\star), h_2(x^\star)) = \left(\frac{1}{10}, \frac{1}{4}, \frac{3}{20}\right).$$

- ***(Derived totals and cuts.)*** *The same capacity vector satisfies*

$$q_0(x^\star) = \frac{1}{4}, \qquad m(x^\star) = \frac{1}{4}, \qquad q_1(x^\star) = \frac{1}{2}, \qquad \Delta q(x^\star) = \frac{1}{4},$$

  and

$$B_L(x^\star) = 0, \qquad B_U(x^\star) = \frac{7}{40}.$$

- ***(Identified interval.)*** *With the unit cell weight $p_{x^\star} = 1$, the nonnegative-capacity validity certificate, positive aggregate mass, and the observed-law-domain certificate for $P^\star_{\text{obs}}$, the sharp identified interval is*

$$\Theta_I(P^\star_{\text{obs}}) = \left[0, \frac{7}{10}\right].$$

- ***(Endpoint attainment.)*** *There exist latent laws P^L and P^U, each with a three-level slate system, satisfying the endpoint-witness conditions for $P^\star_{\text{obs}}$, overlap $1/4$, and the retained single covariate cell, such that P^L attains 0 and P^U attains $7/10$.*

Theorem 6 shows how the observed margins translate into the interval $[0, 0.7]$. In the witness, the exact survivor-complier mass is $1/4$, while the upper threshold cut is $7/40$. Dividing the upper cut by the aggregate survivor-complier mass gives $7/10$, and the lower cut gives the lower endpoint 0. Endpoint-attaining latent laws therefore realize both ends of the interval within the same IV-selection model.

The computational routine implements the same threshold logic used in the sharpness theorem. It computes the two cut values, records the positive allocations that attain them, and separates the sparse work needed for endpoint evaluation from the additional work required to materialize full coupling matrices.

Definition 21. [synth_20] (Costed threshold-flow routines) *For finite $\mathcal{X}$, $\mathcal{Y} = \{0, \ldots, K-1\}$, and nonnegative capacity arrays $\ell_i(x)$ and $h_j(x)$, the costed sparse threshold-flow routine is the cellwise algorithm that computes $q_0(x)$, $q_1(x)$, $\Delta q(x)$, $m(x)$, the prefix and tail sums, and the threshold values $B_L(x)$ and $B_U(x)$; constructs an upper-attaining subflow greedily on the nested graph $i < j$ and a lower-attaining subflow greedily on the nested graph $j \leq i$; and completes each subflow to total mass $m(x)$ by a two-pointer allocation on the residual complete bipartite graph. Its output is $(B_L(x), B_U(x))_{x \in \mathcal{X}}$ together with sparse allocation traces listing exactly the positive row–column allocations produced for the lower- and upper-attaining couplings in Γ_x^*. The actual sparse operation count is the number of arithmetic operations, comparisons, pointer advances, and trace writes executed by this sparse implementation. The charged sparse operation count assigns one charge to each prefix or tail update, threshold comparison, pointer advance, residual-capacity update, and positive-allocation trace append. The costed dense threshold-flow routine performs the same threshold-flow construction and then materializes each lower- and upper-attaining $K \times K$ coupling matrix, writing zero entries as explicit matrix entries. Its actual dense operation count is the number of arithmetic operations, comparisons, pointer advances, residual-capacity updates, and matrix writes executed by the dense implementation. Its charged dense operation count is the sparse charge plus one charge for each materialized matrix entry.*

Definition 22. [synth_21] (Sparse threshold-flow operation counts) *For a covariate cell x, let $\mathrm{Ops}_{\mathrm{sparse,actual}}(x)$ be the number of arithmetic operations and comparisons actually performed by the costed sparse threshold-flow routine of Algorithm 1 in cell x, including the computation of $m(x)$, the prefix and tail sums, the threshold values $B_L(x)$ and $B_U(x)$, and the sparse lower- and upper-attaining couplings. Let $\mathrm{Ops}_{\mathrm{sparse,charged}}(x)$ be the operation count assigned by the routine's charging scheme in cell x, where each pass over an outcome level and each pointer advance or allocation step in the greedy nested-graph subflow and residual two-pointer completion is charged to the corresponding sparse routine step.*

These definitions count the operations at the level used by the constructive proof of Theorem 5. The sparse trace is the economically relevant certificate: it lists the row–column masses that attain the endpoint. Dense matrices are useful when an implementation or diagnostic display requires a complete $K \times K$ array, and their cost includes the explicit zero writes.

Theorem 7. [thm:linear-sparse-threshold-flow] (Sparse threshold flow) *There exist positive integer constants C_{sparse} and C_{dense} such that the following holds. Let $\mathcal{X}$ be a finite covariate support, let $K \geq 3$, and let $\ell_y(x)$ and $h_y(x)$, the lower- and upper-arm capacity arrays, be indexed by $x \in \mathcal{X}$ and $y \in \{0, \ldots, K-1\}$. Suppose*

$$\ell_y(x) \geq 0, \qquad h_y(x) \geq 0$$

for every $x \in \mathcal{X}$ and every $y \in \{0, \ldots, K-1\}$. For the threshold-flow construction in Algorithm 1, write $B_L(x)$ and $B_U(x)$ for the lower and upper threshold-cut values in Definition 16. Then:

- ***(Sparse operation bounds.)*** *The actual sparse operation count*

$$\sum_{x \in \mathcal{X}} \mathrm{Ops}_{\mathrm{sparse,actual}}(x)$$

and the charged sparse operation count

$$\sum_{x \in \mathcal{X}} \mathrm{Ops}_{\mathrm{sparse,charged}}(x)$$

are each bounded above by

$$C_{\text{sparse}}\,|\mathcal{X}|\,K.$$

- ***(Dense operation bounds.)*** *The actual dense operation count*

$$\sum_{x\in\mathcal{X}} \mathrm{Ops}_{\text{sparse,actual}}(x) + 2|\mathcal{X}|K^2$$

and the charged dense operation count are each bounded above by

$$C_{\text{dense}}\,|\mathcal{X}|\,K^2.$$

- ***(Magnitude-invariant charged schedules.)*** *For any other nonnegative capacity arrays on the same $\mathcal{X}$ and the same K, the charged sparse operation counts are equal and the charged dense operation counts are equal.*

- ***(Cellwise output.)*** *For every $x \in \mathcal{X}$, the costed sparse threshold-flow routine returns the threshold-cut pair $(B_L(x), B_U(x))$ and the sparse lower and upper allocation traces from Algorithm 1. The induced lower and upper couplings γ_x^L and γ_x^U satisfy*

$$|\{(i,j) : \gamma_x^L(i,j) > 0\}| \le 2K-1, \qquad |\{(i,j) : \gamma_x^U(i,j) > 0\}| \le 2K-1,$$

belong to the branch-free partial-transport polytope in Definition 15, and attain the threshold-cut benefit masses

$$b_x(\gamma_x^L) = B_L(x), \qquad b_x(\gamma_x^U) = B_U(x).$$

Theorem 7 establishes that exact endpoint computation scales linearly in the number of outcome levels when the output is kept sparse, uniformly across covariate cells. The sparse bound follows the nested-threshold structure: each pass updates a prefix or tail quantity, and each pointer movement is charged once. A quadratic term enters precisely when the implementation writes full $K \times K$ lower and upper matrices. The magnitude-invariance clause is useful for repeated evaluation because the charged schedule depends on the support size and outcome grid, while the returned cut values and allocation masses depend on the realized capacities.

6 Estimation and guarded inference

The sharp interval in Definition 17 is a finite-dimensional functional of the observed-data law. Estimation therefore proceeds by replacing the observable probabilities in Definition 14 with empirical analogues, projecting the resulting capacities onto their nonnegative range, and evaluating the same threshold-cut formula cell by cell. The inferential problem is the usual one for nonsmooth partially identified endpoints: max, min, screening, and quotient operations can all determine the local behavior of the endpoint map [Imbens and Manski, 2004, Chernozhukov et al., 2007, Tamer, 2010, Molinari, 2020, Fang and Santos, 2019]. The construction below keeps these operations explicit.

The finite-sample guard uses two sampling conditions. The first keeps the target stratum uniformly away from zero in aggregate, so that the endpoint quotients remain stable. The second fixes the sampling scheme for the empirical observed law.

Assumption 9. [ass:uniform-aggregate-survivor-bound] (Uniform aggregate survivor bound) *The uniform aggregate lower bound $m_\star$ is strictly positive and bounds the aggregate survivor-complier mass M in Definition 17 from below:*

$$0 < m_\star \quad \textit{and} \quad m_\star \leq M.$$

The requirement in Assumption 9 is a uniform positive target-stratum probability condition: the aggregate survivor-complier mass entering the denominator of the identified endpoints is bounded below by $m_\star$. This is the standard positivity requirement for inference on survivor-type causal targets [Kennedy et al., 2019].

Assumption 10. [ass:iid-sampling] (I.i.d. sampling) *The observations $O_1, \ldots, O_n$ are independent draws from the observed-data law P_{obs}.*

Assumption 10 is the independent sampling condition. It supplies the empirical-process input for the finite observed support and is the conventional baseline for the delta-method and multiplier arguments used below [van der Vaart, 1998].

Definition 23. [def:uniform-law-class] (Uniform law class $\mathcal{P}_n$) *The uniform law class at index n is*

$$\mathcal{P}_n = \{P \in \mathcal{M} : P \textit{ satisfies the uniform aggregate-survivor condition and the independent-sampling condition at}$$

The class Definition 23 collects the laws for which the structural restrictions in Definition 13, the lower aggregate-mass condition, and independent sampling hold at the sample-size index. The notation $\mathcal{P}_n$ will be used for uniform probability statements indexed by n.

Definition 24. [synth_15] (Empirical cell probability) *For observations $O_0, O_1, \ldots$ taking values in observed-data records with covariate cell component X, realization ω, sample size $n \in \mathbb{N}$, and covariate cell x, the empirical covariate-cell probability is*

$$\widehat{p}_x = \begin{cases} \frac{1}{n} \sum_{r=0}^{n-1} \mathbf{1}\{X(O_r(\omega)) = x\}, & n \geq 1, \\ 0, & n = 0. \end{cases}$$

The empirical covariate-cell probability $\widehat{p}_x$ is the sample analogue of the cell weight p_x in Definition 17. It is the outer weight in the plug-in sums and also enters the screening score.

Algorithm 2 (Plug-in endpoint estimator $\widehat{\Psi}_n$)

Given observations $O_1, \ldots, O_n$, a finite covariate support $\mathcal{X}$, outcome support $\mathcal{Y} = \{0, \ldots, K-1\}$, and screening threshold η_n, define the projected screened plug-in endpoint estimator $\widehat{\Psi}_n$ by the following steps.

1. For each empirical conditional probability given $(X = x, Z = z)$, use the empirical ratio when the empirical denominator is positive and use the fixed value zero when that denominator is zero. These empirical conditional probabilities form the raw capacity vector

$$\widetilde{c}_n = (\widetilde{\ell}, \widetilde{h}).$$

2. Define the nonnegative capacity cone and the projected empirical capacity vector by

$$\mathfrak{C} = \prod_{x \in \mathcal{X}} \mathbb{R}_+^{2K}, \qquad \widehat{c}_n = \Pi_{\mathfrak{C}}(\widetilde{c}_n) = \arg\min_{c \in \mathfrak{C}} \|c - \widetilde{c}_n\|_2^2.$$

With the fixed Euclidean metric, the minimizer is unique, and $\Pi_{\mathfrak{C}}$ is the coordinatewise positive-part map. Write

$$\widehat{c}_n = (\widehat{\ell}, \widehat{h}).$$

3. For each $x \in \mathcal{X}$, define the empirical capacity totals, capacity-total gap, and empirical exact mass by

$$\widehat{q}_0(x) = \sum_i \widehat{\ell}_i(x), \qquad \widehat{q}_1(x) = \sum_j \widehat{h}_j(x),$$

$$\widehat{\Delta q}(x) = \widehat{q}_1(x) - \widehat{q}_0(x), \qquad \widehat{m}(x) = \min\{\widehat{q}_0(x), \widehat{q}_1(x)\}.$$

4. For each $x \in \mathcal{X}$, define the empirical lower and upper threshold-cut values by

$$\widehat{B}_L(x) = \max\left\{0,\ \max_{t \in \mathcal{Y}} \left[\sum_{i \le t} \widehat{\ell}_i(x) - \sum_{j \le t} \widehat{h}_j(x)\right] + \min\{\widehat{\Delta q}(x), 0\}\right\}$$

and

$$\widehat{B}_U(x) = \min\left\{\widehat{m}(x),\ \min_{t \in \mathcal{Y}} \left[\sum_{i < t} \widehat{\ell}_i(x) + \sum_{j > t} \widehat{h}_j(x)\right]\right\}.$$

5. Define the empirical retained-cell mass score and retained-cell indicator by

$$\widehat{r}_x = \widehat{p}_x \widehat{m}(x), \qquad \widehat{I}_x = \mathbf{1}\{\widehat{r}_x > \eta_n\}.$$

6. Define the screened empirical aggregate mass and screened empirical endpoint numerators by

$$\widehat{M}_n = \sum_x \widehat{I}_x\, \widehat{p}_x \widehat{m}(x),$$

$$\widehat{N}_{L,n} = \sum_x \widehat{I}_x\, \widehat{p}_x \widehat{B}_L(x), \qquad \widehat{N}_{U,n} = \sum_x \widehat{I}_x\, \widehat{p}_x \widehat{B}_U(x).$$

7. Output the projected screened plug-in endpoint estimator

$$\widehat{\Psi}_n = \begin{cases} \left(\widehat{N}_{L,n}/\widehat{M}_n,\ \widehat{N}_{U,n}/\widehat{M}_n\right), & \widehat{M}_n > 0, \\ (0,1), & \widehat{M}_n = 0. \end{cases}$$

Algorithm 2 mirrors the population construction in Definition 17. The raw empirical capacity vector $\widetilde{c}_n$ is projected to the nonnegative capacity cone, giving the projected empirical capacity vector $\widehat{c}_n$. The retained-cell score $\widehat{r}_x$ and screening threshold η_n restrict the quotient to cells with empirical survivor-complier mass large enough to enter stably. The resulting estimator $\widehat{\Psi}_n$ is the sample analogue of the endpoint map.

For diagnostics and finite-sample containment, the guard works with a finite collection of observable events and a multiplier process over observed cells. The operational event collection consists of cell events $\{X = x\}$, arm events $\{X = x, Z = z\}$, and selected-outcome events $\{X = x, Z = z, D = d, S = 1, \widetilde{Y} = k\}$, indexed by $x \in \mathcal{X}$, $z, d \in \{0, 1\}$, and $k \in \{0, \ldots, K-1\}$. Hence

$$|\mathcal{E}| = |\mathcal{X}| + 2|\mathcal{X}| + 4K|\mathcal{X}| = |\mathcal{X}|(4K + 3).$$

Definition 25. [`synth_18`] (Guard event collection) *For a finite covariate-cell support $\mathcal{X}$ and a nonnegative integer K, define $\mathcal{E} = \mathcal{E}(\mathcal{X}, K)$ to be the full finite collection of guard events:*

$$\mathcal{E}(\mathcal{X}, K) = \{\textit{all guard-event indices for } (\mathcal{X}, K)\}.$$

Definition 26. [`synth_2`] (Empirical event frequency) *For observed-data records $O_0, O_1, \ldots$, each containing a covariate cell, binary instrument, treatment, selection indicator, and optional observed outcome, the empirical frequency $\widehat{P}_n$ of any observable event E at sample size n and state ω is*

$$\widehat{P}_n(E; \omega) = \frac{1}{n} \sum_{r=0}^{n-1} \mathbf{1}\{E(O_r(\omega))\}.$$

Definition 27. [`synth_3`] (Multiplier process) *For observations $O_1, \ldots, O_n$, multiplier inputs $\xi_1, \ldots, \xi_n$, sample point ω, and an observed-data cell $o = (x, z, d, s, y)$, with $x \in \mathcal{X}$, $z, d, s \in \{0, 1\}$, and $y \in \{0, \ldots, K-1\}$ when selected and $y = \varnothing$ when unobserved, the multiplier process $\mathbb{G}_n^{\xi}$ is*

$$\mathbb{G}_n^{\xi}(o) = n^{-1/2} \sum_{r=1}^{n} \xi_r(\omega) \left\{ \mathbf{1}\{O_r(\omega) = o\} - \widehat{P}_n(o) \right\}, \qquad \widehat{P}_n(o) = n^{-1} \sum_{r=1}^{n} \mathbf{1}\{O_r(\omega) = o\}.$$

The finite guard-event index collection consists exactly of the cell events indexed by $x \in \mathcal{X}$, the arm events indexed by $(x, z) \in \mathcal{X} \times \{0, 1\}$, and the selected-outcome events indexed by $(x, z, d, k) \in \mathcal{X} \times \{0, 1\} \times \{0, 1\} \times \{0, \ldots, K-1\}$.

The event collection $\mathcal{E}$ is finite because both the covariate and outcome supports are finite. Empirical frequencies $\widehat{P}_n$ control every sample input to the capacity construction, while the multiplier process $\mathbb{G}_n^{\xi}$ records the local face diagnostics associated with simultaneous max, min, and projection operations.

The construction below names three objects that play different roles, and it is worth separating them before reading it. The first is the *reported interval*: it is implementable from $\widehat{\Psi}_n$, η_n, α, ε_Z, $m_\star$, and the support dimensions $(|\mathcal{X}|, K)$, which between them determine $|\mathcal{E}|$ and $b_{n,\alpha}$. Everything an analyst computes is in that list. The second is δ_n, the maximal deviation of the empirical frequencies from P_{obs} over $\mathcal{E}$. It depends on the unknown P_{obs}, so it belongs to the analysis rather than to the computation: it is the quantity the concentration argument bounds, while the padding radius actually applied is computed from $b_{n,\alpha}$ alone. The third is the multiplier face envelope, a *diagnostic* for the local nonsmooth behaviour of the endpoint map: it reports which threshold cuts and projection faces are near-active, and it is read on its own, alongside the reported interval. The coverage guarantee rests on the deterministic padding; uniform multiplier calibration under changing survivor support remains open, and is posed in Remark 1.

Algorithm 3 (Guarded confidence set $\mathcal{C}_{1-\alpha,n}$)

Given $\widehat{\Psi}_n$, η_n, confidence level $1-\alpha$, instrument-overlap constant ε_Z, aggregate survivor-complier lower bound $m_\star$, and the finite observable event collection $\mathcal{E}$, define the guarded confidence set $\mathcal{C}_{1-\alpha,n}$ by the following steps.

1. Use the retained-cell rule
$$\widehat{I}_x = \mathbf{1}\{\widehat{r}_x > \eta_n\}.$$

2. Define the near-active-face localization tolerance by
$$a_n = (n\eta_n)^{-1/4}.$$
The near-active lower and upper threshold cuts, the active face of $\min\{\widehat{\Delta q}(x), 0\}$, the active face of $\min\{\widehat{q}_0(x), \widehat{q}_1(x)\}$, and the coordinatewise nonnegativity faces generated by $\Pi_{\mathfrak{C}}$ determine a computable finite diagnostic face envelope for the multiplier process $\mathbb{G}_n^{\xi}$ at tolerance a_n.

3. Let δ_n be the maximal empirical deviation over $\mathcal{E}$:
$$\delta_n = \max_{E\in\mathcal{E}} \left| \widehat{P}_n(E) - P_{\text{obs}}(E) \right|.$$

4. Define the union-bound empirical-deviation threshold by
$$b_{n,\alpha} = \sqrt{\frac{|\mathcal{E}|}{4n\alpha}}.$$

5. Define the capacity-and-screening deterministic error envelope and deterministic endpoint-padding guard by
$$A_{n,\alpha} = \frac{64K|\mathcal{X}|}{\varepsilon_Z^2} b_{n,\alpha} + |\mathcal{X}|\eta_n,$$
$$g_{n,\alpha} = \min\left\{1, \frac{4A_{n,\alpha}}{m_\star}\right\}.$$

6. Output the conservative guarded confidence set
$$\mathcal{C}_{1-\alpha,n} = \left[\max\{0, \widehat{\Psi}_{n,L} - g_{n,\alpha}\},\ \min\{1, \widehat{\Psi}_{n,U} + g_{n,\alpha}\}\right].$$

The confidence set in Algorithm 3 uses the deterministic deviation threshold $b_{n,\alpha}$ to form a capacity-and-screening envelope $A_{n,\alpha}$. Dividing by the aggregate-mass lower bound yields the endpoint guard $g_{n,\alpha}$, and padding the plug-in endpoints gives the interval $\mathcal{C}_{1-\alpha,n}$. The multiplier component supplies a face-aware diagnostic for nonsmooth local behavior; the stated confidence set is the deterministic guarded interval.

The first large-sample result fixes one observed-data law and characterizes the local endpoint limit on the recovered positive-survivor support.

Theorem 8. [`thm:branch-free-pointwise-directional-limit`] (Branch-free endpoint limit) *Fix a slate potential-outcome system with finite covariate support and ordered outcome support of size $K \geq 3$. Let P_{obs} be its observed-data law, let $O_1, \ldots, O_n$ denote the observed cells, let $d(x) \in \{-1, +1\}$ be the weak selection-monotonicity direction, let ε_Z be the instrument-overlap constant, and let η_n be the screening threshold from Algorithm 2. Suppose that:*

- ***(Sampling.)*** *For every n, the first n observations satisfy Assumption 10 with common law P_{obs}.*
- ***(Overlap.)*** *The instrument propensity satisfies Assumption 5 with constant ε_Z.*
- ***(Structural restrictions.)*** *The latent law satisfies Assumptions 1 to 4 and 6 to 8, with weak selection monotonicity in direction d.*
- ***(Screening rate.)*** *For every n, $\eta_n > 0$, with $\eta_n \to 0$ and $\sqrt{n}\,\eta_n \to \infty$.*
- ***(Limit inputs.)*** *The finite multinomial central limit theorem applies to the empirical observed-cell vector, and the Hadamard directional delta method applies to tangentially directionally differentiable maps.*

Let c be the observable capacity vector from Definition 14. Let

$$M = \sum_x p_x m(x),$$

where $m(x)$ is the exact survivor-complier mass in cell x, and let $\mathcal{X}_+(P) = \{x : p_x m(x) > 0\}$. Let $\Psi(P_{\text{obs}}) = (\theta_L, \theta_U)$ be the endpoint map identified by Theorem 1, and let $\widehat{\Psi}_n$ be the projected screened plug-in estimator from Algorithm 2. Then

$$P(\{x : \widehat{r}_x > \eta_n\} = \mathcal{X}_+(P)) \to 1$$

and

$$\widehat{\Psi}_n \to_P \Psi(P_{\text{obs}}).$$

Moreover, there exist a Gaussian law $\mathbb{Z}_{P_{\text{obs}}}$ on observed-cell score functions and a map $D\Psi$ such that

$$\int z \, d\mathbb{Z}_{P_{\text{obs}}}(z) = 0$$

and, for observed cells a, b,

$$\int z(a)z(b)\, d\mathbb{Z}_{P_{\text{obs}}}(z) = \begin{cases} P_{\text{obs}}(a)\{1 - P_{\text{obs}}(a)\}, & a = b, \\ -P_{\text{obs}}(a)P_{\text{obs}}(b), & a \neq b. \end{cases}$$

The map $D\Psi$ is the Hadamard directional derivative, at the observed-law mass vector and on the recovered support $\mathcal{X}_+(P)$, of the reduced-support endpoint map. With this derivative,

$$\sqrt{n}\{\widehat{\Psi}_n - \Psi(P_{\text{obs}})\} \rightsquigarrow D\Psi(\mathbb{Z}_{P_{\text{obs}}}).$$

Theorem 8 establishes consistency of the screened support and of the endpoint estimator, then identifies the fixed-law directional limit. The Gaussian input $\mathbb{Z}_{P_{\text{obs}}}$ is the centered multinomial limit for the observed cells, and the derivative is taken after the positive-survivor support

$\mathcal{X}_+(P)$ has been recovered. The branch-free threshold representation handles simultaneous max and min faces through a Hadamard directional derivative; this derivative is the econometric object governing nonsmooth endpoint asymptotics [Fang and Santos, 2019, Hong and Li, 2018].

The next three algebraic facts record how the cellwise masses and threshold cuts scale with nonnegative cell weights. They justify treating covariate-cell probabilities as outer weights in the aggregate endpoint formula and in the empirical analogue.

Lemma 1. [`lem:weighted-capacity-mass-homogeneity`] (Weighted mass homogeneity) *Let $c = (\ell, h)$ be a capacity array as in Definition 14, and let $w = (w_x)$ be real cell weights satisfying $w_x \geq 0$ for every covariate cell x. Define the weighted capacity array c^w cellwise by*

$$\ell_i^w(x) = w_x \ell_i(x), \qquad h_j^w(x) = w_x h_j(x).$$

If $m^w(x)$ denotes the exact survivor-complier mass computed from c^w by the formulas in Definition 14, then, for every covariate cell x,

$$m^w(x) = w_x m(x).$$

The proof is deferred to Section B.

Lemma 2. [`lem:weighted-capacity-lower-cut-homogeneity`] (Weighted lower cut homogeneity) *Let $c = (\ell, h)$ be a capacity array, and let $w = (w_x)_x$ be a cell weight function satisfying $w_x \geq 0$ for every cell x. Let c^w be the cellwise weighted capacity array with lower and upper coordinates*

$$\ell_i^w(x) = w_x \ell_i(x), \qquad h_j^w(x) = w_x h_j(x).$$

If $B_L^w(x)$ denotes the lower threshold-cut value of Definition 16 computed from c^w, and $B_L(x)$ denotes the same lower threshold-cut value computed from c, then, for every cell x,

$$B_L^w(x) = w_x B_L(x).$$

The proof is deferred to Section B.

Lemma 3. [`lem:weighted-capacity-upper-cut-homogeneity`] (Upper cut homogeneity) *Let $c = (\ell, h)$ be a capacity array, and let $w : \mathcal{X} \to \mathbb{R}$ satisfy $w_x \geq 0$ for every $x \in \mathcal{X}$. Form the cellwise weighted capacity array c^w by*

$$\ell_i^w(x) = w_x \ell_i(x), \qquad h_j^w(x) = w_x h_j(x).$$

Writing B_U^w for the upper threshold cut of Definition 16 computed from c^w, and B_U for the same cut computed from c, one has, for every $x \in \mathcal{X}$,

$$B_U^w(x) = w_x B_U(x).$$

The proof is deferred to Section B.

Lemmas 1 to 3 show that multiplying all capacities in a cell by the same nonnegative weight multiplies the exact mass and both benefit cut values by that weight. Consequently, aggregation over cells preserves the threshold-cut form of the endpoint numerators and denominator.

For the uniform statement, laws are indexed by λ. The empirical law and auxiliary multipliers are defined law by law, while the deterministic guard keeps the same finite-event form.

Definition 28. [`synth_19`] (Empirical observed-data law for the λ-indexed sample) *For each law $\lambda \in \Lambda$ and sample $(O_{\lambda,1}, \dots, O_{\lambda,n})$, the empirical observed-data law is $\widehat{P}_{n,\lambda} = n^{-1} \sum_{r=1}^{n} \delta_{O_{\lambda,r}}$.*

Definition 29. [`synth_6`] (Auxiliary multiplier process) *For each law $\lambda \in \Lambda$, let $\xi_\lambda = (\xi_{\lambda,r})_{r \geq 1}$ denote auxiliary multiplier variables, conditionally independent given the observed sample and satisfying conditional mean zero and conditional variance one. They generate the diagnostic multiplier process $\mathbb{G}^{\xi}_{n,\lambda} = n^{-1/2} \sum_{r=1}^{n} \xi_{\lambda,r} \{\delta_{O_{\lambda,r}} - \widehat{P}_{n,\lambda}\}$ used to form the finite face diagnostic. The guarded confidence interval itself is constructed from the screened endpoint estimator and the deterministic padding $g_{n,\alpha}$.*

The empirical law $\widehat{P}_{n,\lambda}$ and auxiliary process in Definitions 28 and 29 provide the indexed-sample notation for applying the same construction uniformly over Λ. The multiplier process is diagnostic for the finite face envelope, while containment is delivered by the deterministic padding rule in Algorithm 3.

Theorem 9. [`thm:uniform-deterministic-guard`] (Uniform deterministic guard) *Fix a family of finite-slate potential-outcome systems indexed by $\lambda \in \Lambda$, with nonempty finite covariate support $\mathcal{X}$, ordered outcome support of size K, and $K \geq 3$. For each λ, let $P_{\mathrm{obs},\lambda}$ be the observed-data law of $(X, Z, D, S, \widetilde{Y})$, where $\widetilde{Y} = Y$ on selected units and is missing off selection. Let $O_{\lambda,1}, O_{\lambda,2}, \dots$ be observations taking values in the observed-data cells, let ξ_λ be the auxiliary process used by the guarded interval in Algorithm 3, let $d_\lambda(x)$ be the weak selection-monotonicity direction, and let μ_λ be the sampling law.*

Assume:

- ***(Nominal level.)*** *The nominal error level satisfies $0 < \alpha < 1$.*
- ***(Overlap and mass constants.)*** *The instrument-overlap constant and aggregate survivor-complier mass lower bound satisfy*
$$0 < \varepsilon_Z < \frac{1}{2}, \qquad m_\star > 0.$$
- ***(Screening sequence.)*** *The screening thresholds satisfy $\eta_n > 0$ for every n,*
$$\eta_n \to 0, \qquad \sqrt{n}\,\eta_n \to \infty.$$
- ***(Uniform law class.)*** *For every law λ and every $n \geq 1$, the system belongs to the uniform finite-slate law class of Definition 23 at index n, with overlap constant ε_Z, aggregate survivor-complier mass bound $m_\star$, direction d_λ, sampling law μ_λ, and observations $O_{\lambda,1}, \dots, O_{\lambda,n}$.*

Let $\Theta_I(P_{\mathrm{obs},\lambda})$ be the sharp identified interval for the benefit probability from Definition 17, and let
$$\Psi(P_{\mathrm{obs},\lambda}) = (\theta_{L,\lambda}, \theta_{U,\lambda})$$
be its endpoint map. Let $\widehat{\Psi}_{n,\lambda}(\omega)$ be the projected screened plug-in endpoint estimator from Algorithm 2. Let $\mathcal{C}_{1-\alpha,n,\lambda}(\omega)$ and $g_{n,\alpha}$ be the guarded confidence interval and deterministic endpoint-padding guard from Algorithm 3.

Then, for every $n \geq 1$,
$$\inf_{\lambda \in \Lambda} \mu_\lambda\{\omega : \Theta_I(P_{\mathrm{obs},\lambda}) \subseteq \mathcal{C}_{1-\alpha,n,\lambda}(\omega)\} \geq 1 - \alpha.$$

Moreover,

$$\sup_{\lambda\in\Lambda} \mu_\lambda \Big\{ \omega : g_{n,\alpha} < \max\Big\{ \Big|\widehat{\theta}_{L,n,\lambda}(\omega) - \theta_{L,\lambda}\Big|, \Big|\widehat{\theta}_{U,n,\lambda}(\omega) - \theta_{U,\lambda}\Big| \Big\} \Big\} \longrightarrow 0,$$

where $\widehat{\Psi}_{n,\lambda} = (\widehat{\theta}_{L,n,\lambda}, \widehat{\theta}_{U,n,\lambda})$, *and*

$$g_{n,\alpha} = O\Big(n^{-1/2} + \eta_n\Big).$$

Finally, for every deterministic sequence $L_n > 0$ *such that*

$$L_n \to \infty, \qquad L_n = o(\sqrt{n}),$$

the screening choice

$$\eta_n = n^{-1/2} L_n$$

satisfies, for every $n \geq 1$*,*

$$\eta_n > 0, \qquad \eta_n \to 0, \qquad \sqrt{n}\,\eta_n \to \infty,$$

and the corresponding deterministic guard radius obeys

$$g_{n,\alpha} = O\Big(n^{-1/2} L_n\Big).$$

Theorem 9 gives the advertised guarded inference guarantee. For every sample size, the interval from Algorithm 3 contains the full sharp identified interval with probability at least $1 - \alpha$, uniformly over the indexed law family. The same result establishes uniform endpoint consistency under the deterministic guard, with radius $O(n^{-1/2} + \eta_n)$; choosing $\eta_n = n^{-1/2} L_n$ for any $L_n \to \infty$ and $L_n = o(\sqrt{n})$ yields a guard of order $O(n^{-1/2} L_n)$. The construction is conservative in the sense of finite-sample containment, and its rate statement shows that the extra screening term can vanish arbitrarily close to the root-n scale.

7 Limitations and future work

The most important practical limitation is the width of the deterministic guard. The guarantee in Theorem 9 is a union bound over the finite event collection $\mathcal{E}$, and it buys uniformity at the price of large constants. Taking the smallest configuration the model allows — $K = 3$, $|\mathcal{X}| = 1$, $\varepsilon_Z = 1/4$, $\alpha = 0.05$, and $m_\star = 1/4$ — gives $|\mathcal{E}| = 15$, so

$$b_{n,\alpha} \approx 8.7\, n^{-1/2}, \qquad A_{n,\alpha} \approx 2.7 \times 10^4\, n^{-1/2}, \qquad g_{n,\alpha} \approx 4.3 \times 10^5\, n^{-1/2}$$

before the screening term is added. Since $g_{n,\alpha}$ is capped at 1, the reported interval equals the whole of $[0, 1]$ for every n below roughly 10^{11}, and its half-width $g_{n,\alpha}$ first falls below 0.1 — a reported width of about 0.2 plus the width of the identified interval itself — at n of order 10^{13}. At any sample size an applied researcher will encounter, the guarded interval is therefore uninformative: it is a proof that a finite-sample, uniformly valid procedure exists in this model, not a procedure we recommend reporting on its own. The rate $O(n^{-1/2} + \eta_n)$ is the honest content of the result; the constants are not sharp, and we have not attempted to optimize them. Reducing them — by replacing the union bound over $\mathcal{E}$ with a self-normalized or empirical-process concentration argument, and by exploiting that the capacity map is 1-Lipschitz in far fewer directions than $|\mathcal{E}|$ suggests — is the most valuable next step for making the guarantee usable, and a simulation study establishing the achievable widths is a prerequisite for any empirical recommendation. We make no claim about practical performance here.

The sharp bounds and guarded confidence set above also leave one calibrated-inference problem for future work. The issue arises when threshold-cut faces intersect, selection-direction signs are close to changing, and the retained covariate support varies with the screening threshold. A useful strengthening of the inference theory would combine the finite deterministic guard in Theorem 9 with a first-order calibration based on the near-active faces recorded by Algorithm 3. The additional sign separation needed for such a direction-stable analysis is the following standard strong variant of the covariate-specific selection-direction condition.

Assumption 11. [`ass:direction-margin`] (Direction margin) *For every $x \in \mathcal{X}$ with $p_x > 0$ and $m(x) > 0$,*

$$|\Delta q(x)| \geq \kappa_\sigma.$$

Assumption 11 is imposed for a *strictly positive* constant $\kappa_\sigma > 0$; read that way it is a sign-separation condition on the covariate-specific selected-complier capacity gap, requiring the selection direction to stay at least κ_σ away from zero on cells with positive survivor-complier mass. The displayed inequality alone carries no content when $\kappa_\sigma \leq 0$, since $|\Delta q(x)| \geq \kappa_\sigma$ then holds automatically; positivity of the margin is what the direction-stable analysis below uses, and it is not implied by the inequality as displayed. This is the analogue, for the capacity representation used here, of the strong direction conditions used to stabilize sign-sensitive inference under sample selection [Semenova, 2025]. It gives the multiplier program a fixed local direction in cells that contribute to the survivor-complier target, while retaining the threshold-cut ties that make the endpoint map nonsmooth.

Remark 1. [`oeq:uniform-face-multiplier`] (Uniform face multiplier calibration) *A natural next question is whether the near-active-face directional-multiplier handle admits a calibration such that, uniformly over direction-separated triangular arrays $\mathcal{P}_n$,*

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}_n} P\{\Theta_I(P_{\text{obs}}) \subseteq \mathcal{C}_{1-\alpha,n}\} \geq 1-\alpha,$$

while allowing arbitrary simultaneous threshold-cut ties and covariate cells whose survivor mass approaches zero and is omitted at the unnormalized threshold η_n.

Remark 1 formulates the corresponding open calibration target. The desired confidence set would cover $\Theta_I(P_{\text{obs}})$, the sharp identified interval for the survivor-complier benefit probability, uniformly over direction-separated triangular arrays, using $\mathcal{C}_{1-\alpha,n}$, the guarded interval constructed from the finite face diagnostic. The challenge is to preserve first-order calibration when several threshold cuts bind at once and when cells enter or leave the retained support as their empirical survivor mass crosses η_n. This problem is closely connected to modern inference for partially identified treatment effects and selected samples, where uniformity at boundary and contact points is central [Lee and Liu, 2024, Bartalotti et al., 2023].

Several empirical and modeling extensions are also natural. The Job Corps application that motivates the ordinal wage target can be revisited with the sharp survivor-complier bounds and the guarded interval reported jointly. Continuous outcomes can be handled through discretizations or through an analogue of the threshold-transport construction on richer ordered supports, with the estimand and support conditions stated directly for that setting. Sensitivity analysis for weak selection monotonicity would index departures from Assumption 7 and trace their effect on the endpoint formulas in Definition 17. Additional concordance restrictions, such as restrictions tying potential wage ranks across treatment states, would sharpen the cellwise transport polytopes in Definition 15 and produce a different identified interval whose sharpness would need to be established under the enriched model.

Appendices

A Technical derivations and verification note

The appendix collects the supporting derivations for the sharp survivor-complier benefit bounds, the computational construction, and the guarded inference result. The first part works through the capacity algebra behind Definition 16 and the residual completion used by Algorithm 1. The second part records the comparison argument behind Theorem 4. The final technical part gives the concentration calculations that enter the deterministic guard in Algorithm 3 and Theorem 9.

A.1 Threshold-cut algebra and residual completion

The threshold-cut calculations organize the cellwise problem in Definition 15 by ordered lower and upper outcome levels. For the lower endpoint, the relevant comparisons aggregate lower-arm mass weakly below a threshold and upper-arm mass weakly below the same threshold; for the upper endpoint, they compare lower-arm mass below a threshold with upper-arm mass above it. These finite ordered-support identities are the algebraic content behind Definition 16 and explain why the endpoint formulas in Definition 17 depend only on observed capacities and cell probabilities.

The residual completion step fills the remaining rows and columns after the threshold mass has been assigned. Its role is to turn the cut values into feasible cellwise couplings in the exact-mass transport set, preserving the row and column inequalities while attaining the desired strict-benefit mass. This is the constructive bridge from the formulas in Definition 16 to the endpoint-attaining laws in Theorem 3. The same finite-dimensional structure also underlies the sparse implementation summarized by Theorem 7.

A.2 No-selection reduction

The no-selection reduction specializes the general selected-survivor construction to the setting in which the selection constraint is inactive. Under the hypotheses of Theorem 4, the cellwise capacity arrays collapse to the ordinary complier outcome margins, and the same threshold formulas recover the established ordinal benefit bounds. This comparison fixes the relationship between the survivor-complier estimand studied here, sharp complete-marginal ordinal-outcome benefit bounds [Lu et al., 2018], and selection-bounds analyses of intensive-margin effects under endogenous treatment [Dong and Heiler, 2026].

A.3 Guard calculations

The inference appendix derives the deterministic inequalities used by the guarded confidence set in Algorithm 3. The calculations combine finite-support empirical deviations over the event collection in that construction with the screening rule in Algorithm 2. Standard empirical-process and delta-method arguments for finite-dimensional laws supply the pointwise Gaussian approximation underlying Theorem 8 [van der Vaart, 1998]. The guard in Theorem 9 then converts uniform control of the observed event probabilities into endpoint padding for the full identified interval.

The face-aware diagnostic records the threshold cuts that are near active at the empirical law. This finite diagnostic is useful because the endpoint map is piecewise linear in the capacity vector, with different linear pieces selected by the active threshold comparisons. The deterministic guard separates two tasks: it bounds capacity and screening errors by an explicit envelope, and it expands the resulting endpoint interval by a finite-sample padding term. This organization follows the econometric logic of inference for nonsmooth and partially identified parameters, where contact sets and binding inequalities determine the local approximation [Fang and Santos, 2019].

A.4 Verification note

The formal assumptions, definitions, propositions, algorithms, theorem statements, and proof dependencies in the paper are machine-checked in Lean 4 for the finite ordered-outcome model stated in the main text. The checking covers the capacity-identification formulas, the threshold-cut endpoint characterization, endpoint attainment by the threshold-flow construction, the no-selection reduction, the three-level witness, the sparse threshold-flow calculation, the pointwise directional limit statement, and the deterministic guarded-inference guarantee.

Scope, toolchain, and commit. The development is checked with Lean 4, toolchain `leanprover/lean4:v4.33.0`, at commit `8f179a26a191`, in the module tree `CausalSmith/PartialID/PID_Slate`
This paper displays 56 numbered formal objects. Of these, 48 are matched to the Lean development in that tree, 7 are presentation-level restatements carrying no Lean declaration of their own (listed below), and 1 — the direction margin of Assumption 11 — is stated only to pose the open calibration problem and is used by no verified theorem. Every declaration certifying a matched object is free of `sorry` and of added axioms beyond the ambient Lean and Mathlib development.

Objects with no matching declaration. The seven displayed definitions above are presentation-level restatements introduced for readability rather than separate Lean declarations: the slate-benefit partial-transport system (Definition 11), the threshold-cut capacity aggregates (Definition 18), the cellwise benefit-mass functional (Definition 19), the costed threshold-flow routines (Definition 21), the sparse threshold-flow operation counts (Definition 22), the empirical observed-data law for the indexed sample (Definition 28), and the auxiliary multiplier process (Definition 29). Each names or aggregates quantities that the matched declarations define; none carries a verified claim of its own, and no theorem in this paper is certified through one of them. Readers should treat the verification claim as attaching to the matched statements and their proofs, not to these expository restatements.

Cited inputs. Published econometric and probability results invoked through citations are bibliographic dependencies; their source proofs are used in the ordinary way through the cited literature. Where a displayed result depends on such an input, the result carries a footnote naming the cited conclusion and stating that the portions depending on it are certified conditional on that conclusion.

B Proofs of the main results

Proof of Theorem 1. Throughout, conditional probabilities on a zero-mass conditioning event are interpreted by the totalized convention as zero; on positive-mass conditioning events they are the usual ratios. Write

$$\mathsf{Cell}_x = \{X = x\}, \qquad \mathsf{Arm}_{x,z} = \{X = x, Z = z\}.$$

Cell probabilities are nonnegative because $p_x = P(\mathsf{Cell}_x)$. Now fix x with $p_x > 0$. For the conditional instrument propensity $\pi(x)$, the totalized convention is the usual ratio on this cell, so

$$\pi(x) = \frac{P(\mathsf{Arm}_{x,1})}{P(\mathsf{Cell}_x)}, \qquad P(\mathsf{Arm}_{x,1}) = \pi(x)p_x.$$

By Assumption 5,

$$0 < \varepsilon_Z \le \pi(x) \le 1 - \varepsilon_Z < 1.$$

Hence $P(\mathsf{Arm}_{x,1}) > 0$. Since the two instrument fibers partition the cell,

$$p_x = P(\mathsf{Cell}_x) = P(\mathsf{Arm}_{x,0}) + P(\mathsf{Arm}_{x,1}),$$

and therefore

$$P(\mathsf{Arm}_{x,0}) = p_x - P(\mathsf{Arm}_{x,1}) = (1 - \pi(x))p_x > 0.$$

Thus $P(\mathsf{Arm}_{x,z}) > 0$ for both $z \in \{0, 1\}$.

For supported x, the observed lower-arm contrast pulls back along the observed-data law to

$$\ell_i(x) = P(D = 0, S = 1, Y = i \mid X = x, Z = 0) - P(D = 0, S = 1, Y = i \mid X = x, Z = 1).$$

Similarly,

$$h_j(x) = P(D = 1, S = 1, Y = j \mid X = x, Z = 1) - P(D = 1, S = 1, Y = j \mid X = x, Z = 0).$$

The observed-data law is the pushforward of the latent law by the observed datum map, whose coordinate events are measurable; hence these are exactly the corresponding conditional probabilities under the latent probability space.

Using Assumptions 1 to 4, for each $z \in \{0, 1\}$,

$$P(D = 0, S = 1, Y = i \mid X = x, Z = z) = P(D_z = 0, S_0 = 1, Y_0 = i \mid X = x),$$

and

$$P(D = 1, S = 1, Y = j \mid X = x, Z = z) = P(D_z = 1, S_1 = 1, Y_1 = j \mid X = x).$$

Substitution gives

$$\ell_i(x) = P(D_0 = 0, S_0 = 1, Y_0 = i \mid X = x) - P(D_1 = 0, S_0 = 1, Y_0 = i \mid X = x),$$

and

$$h_j(x) = P(D_1 = 1, S_1 = 1, Y_1 = j \mid X = x) - P(D_0 = 1, S_1 = 1, Y_1 = j \mid X = x).$$

By Assumption 6, $D_1 \geq D_0$, so

$$\{D_0 = 0, S_0 = 1, Y_0 = i\} = \{D_1 = 0, S_0 = 1, Y_0 = i\} \,\dot{\cup}\, \{C, S_0 = 1, Y_0 = i\}$$

up to null sets. Therefore

$$\ell_i(x) = P(Y_0 = i, S_0 = 1, C \mid X = x).$$

The same monotonicity gives

$$\{D_1 = 1, S_1 = 1, Y_1 = j\} = \{D_0 = 1, S_1 = 1, Y_1 = j\} \,\dot{\cup}\, \{C, S_1 = 1, Y_1 = j\},$$

and hence

$$h_j(x) = P(Y_1 = j, S_1 = 1, C \mid X = x).$$

These identities also show nonnegativity in supported cells. If $p_x = 0$, then $P(\mathsf{Arm}_{x,z}) = 0$ for both z, so every conditional term in the observable contrasts is zero and the corresponding capacities are nonnegative. Thus the observed-data law is a probability law and its observable capacity contrasts are valid, so it is compatible with the maintained capacity domain.

Because $\mathcal{Y} = \{0, \ldots, K-1\}$ is finite, summing the lower selected-complier identities over outcome levels gives

$$q_0(x) = \sum_{i \in \mathcal{Y}} \ell_i(x) = P(S_0 = 1, C \mid X = x).$$

Applying the same finite partition to the upper selected-complier identities gives

$$q_1(x) = \sum_{j \in \mathcal{Y}} h_j(x) = P(S_1 = 1, C \mid X = x).$$

By Definition 14,

$$\Delta q(x) = q_1(x) - q_0(x) = P(S_1 = 1, C \mid X = x) - P(S_0 = 1, C \mid X = x).$$

It remains to identify $m(x)$. Define the finite restricted measure ν_x by

$$\nu_x(A) = P(A \cap C \cap \mathsf{Cell}_x)$$

for every measurable event A. If $\nu_x(\Omega) = 0$, all selected-complier masses in the cell are zero. Otherwise, Assumption 7 applies on this restricted cell. If $d(x) = +1$, then $S_0 \leq S_1$ ν_x-almost surely, so

$$P(S_0 = 1, S_1 = 1, C \mid X = x) = P(S_0 = 1, C \mid X = x),$$

and

$$P(S_0 = 1, C \mid X = x) \leq P(S_1 = 1, C \mid X = x).$$

If $d(x) = -1$, then $S_1 \leq S_0$ ν_x-almost surely, so

$$P(S_0 = 1, S_1 = 1, C \mid X = x) = P(S_1 = 1, C \mid X = x),$$

and

$$P(S_1 = 1, C \mid X = x) \le P(S_0 = 1, C \mid X = x).$$

Combining these alternatives with $m(x) = \min\{q_0(x), q_1(x)\}$ from Definition 14 yields

$$m(x) = P(S_0 = 1, S_1 = 1, C \mid X = x).$$

The same two alternatives determine every strict sign of the gap. If $d(x) = -1$, then $q_1(x) \le q_0(x)$, so $\Delta q(x) \le 0$. Hence $\Delta q(x) > 0$ forces $d(x) = +1$. If $d(x) = +1$, then $q_0(x) \le q_1(x)$, so $\Delta q(x) \ge 0$. Hence $\Delta q(x) < 0$ forces $d(x) = -1$.

Finally suppose $p_x > 0$ and $\Delta q(x) = 0$. The gap identity gives

$$P(S_1 = 1, C \mid X = x) = P(S_0 = 1, C \mid X = x).$$

Equivalently, $\nu_x(\{S_1 = 1\}) = \nu_x(\{S_0 = 1\})$. If $\nu_x(\Omega) = 0$, both direction inequalities are vacuous in the complier cell. If $\nu_x(\Omega) > 0$, Assumption 7 gives one almost-sure inclusion between $\{S_0 = 1\}$ and $\{S_1 = 1\}$ under ν_x; equality of their finite measures upgrades that inclusion to

$$\nu_x(\{S_0 \ne S_1\}) = 0.$$

Thus, in either case, both inequalities $S_0 \le S_1$ and $S_1 \le S_0$ hold ν_x-almost surely in the cell. Define the updated direction maps by

$$d^{+x}(u) = \begin{cases} +1, & u = x, \\ d(u), & u \ne x, \end{cases} \qquad d^{-x}(u) = \begin{cases} -1, & u = x, \\ d(u), & u \ne x. \end{cases}$$

Away from x, the original weak-monotonicity condition is unchanged; at x, the preceding alternatives supply either direction. Therefore the same latent law, with direction map d^{+x}, witnesses observational admissibility of $+1$, and the same latent law, with direction map d^{-x}, witnesses observational admissibility of -1. The observed-data law is unchanged in both witnesses. □

Proof of Theorem 2. Fix a cell x with $p_x > 0$, and let c be the observable capacity vector from P_{obs}. By Theorem 1, c is valid, so

$$\ell_i(x) \ge 0, \qquad h_j(x) \ge 0$$

for all $i, j \in \mathcal{Y}$, and

$$\Delta q(x) = P(S_1 = 1, C \mid X = x) - P(S_0 = 1, C \mid X = x).$$

Also, by Definition 14,

$$q_0(x) = \sum_i \ell_i(x), \qquad q_1(x) = \sum_j h_j(x), \qquad \Delta q(x) = q_1(x) - q_0(x), \qquad m(x) = \min\{q_0(x), q_1(x)\}.$$

For a coupling array γ_x, write

$$\text{row}_i(\gamma_x) = \sum_j \gamma_{ij,x}, \qquad \text{col}_j(\gamma_x) = \sum_i \gamma_{ij,x}, \qquad \text{tot}(\gamma_x) = \sum_i \sum_j \gamma_{ij,x}.$$

These totals satisfy

$$\sum_i \text{row}_i(\gamma_x) = \text{tot}(\gamma_x), \qquad \sum_j \text{col}_j(\gamma_x) = \text{tot}(\gamma_x).$$

1. We first prove the face equivalence. If $\gamma_x \in \Gamma_x^*$, then Definition 15 gives

$$\mathrm{row}_i(\gamma_x) \le \ell_i(x), \qquad \mathrm{col}_j(\gamma_x) \le h_j(x), \qquad \mathrm{tot}(\gamma_x) = m(x).$$

When $q_0(x) \le q_1(x)$,

$$\sum_i \{\ell_i(x) - \mathrm{row}_i(\gamma_x)\} = q_0(x) - m(x) = 0.$$

Each summand is nonnegative, hence

$$\mathrm{row}_i(\gamma_x) = \ell_i(x) \qquad \text{for every } i.$$

Similarly, when $q_1(x) \le q_0(x)$,

$$\mathrm{col}_j(\gamma_x) = h_j(x) \qquad \text{for every } j.$$

By the cellwise output of Algorithm 1, choose a feasible exact-mass coupling

$$\gamma_x^{\mathrm{L}} \in \Gamma_x^*.$$

Assume

$$\Gamma_x^+ = \Gamma_x^-, \qquad \Gamma_x^- = \Gamma_x^0.$$

If $q_0(x) \le q_1(x)$, the preceding row-exactness argument puts γ_x^{L} in Γ_x^+. The displayed face equalities then put γ_x^{L} in Γ_x^0. Therefore

$$q_0(x) = \mathrm{tot}(\gamma_x^{\mathrm{L}}) = q_1(x),$$

using exact rows for the first equality and exact columns for the second. If $q_1(x) \le q_0(x)$, the same argument with columns first puts γ_x^{L} in Γ_x^-, then in Γ_x^0, and again gives

$$q_0(x) = \mathrm{tot}(\gamma_x^{\mathrm{L}}) = q_1(x).$$

Thus $\Delta q(x) = 0$.

Conversely, suppose $\Delta q(x) = 0$. Then

$$q_0(x) = q_1(x) = m(x).$$

If $\gamma_x \in \Gamma_x^+$, then its exact row constraints give

$$\mathrm{tot}(\gamma_x) = \sum_i \mathrm{row}_i(\gamma_x) = \sum_i \ell_i(x) = q_0(x) = m(x),$$

so $\gamma_x \in \Gamma_x^*$. Since $q_1(x) \le q_0(x)$, the column-exactness argument above gives

$$\mathrm{col}_j(\gamma_x) = h_j(x) \qquad \text{for every } j,$$

and hence $\gamma_x \in \Gamma_x^-$. The reverse inclusion follows symmetrically from exact columns and row-exactness. Therefore

$$\Gamma_x^+ = \Gamma_x^-.$$

The same reasoning sends every element of Γ_x^- to Γ_x^0, while $\Gamma_x^0 \subseteq \Gamma_x^-$ is immediate from the defining constraints. Hence

$$\Gamma_x^- = \Gamma_x^0.$$

This proves

$$\left(\Gamma_x^+ = \Gamma_x^- \text{ and } \Gamma_x^- = \Gamma_x^0\right) \iff \Delta q(x) = 0.$$

2. Now assume $\Delta q(x) = 0$. Since $p_x > 0$, conditional probabilities given $X = x$ use the denominator p_x. The displayed identity from Theorem 1 yields

$$P(S_1 = 1, C \mid X = x) = P(S_0 = 1, C \mid X = x).$$

If $P(C \mid X = x) = 0$, then both

$$P(S_0 = S_1, C \mid X = x) \quad \text{and} \quad P(C \mid X = x)$$

are zero. If $P(C \mid X = x) > 0$, then Assumption 7 gives one of the two almost-sure orders on the complier cell:

$$1\{S_0 = 1\} \le 1\{S_1 = 1\} \quad \text{or} \quad 1\{S_1 = 1\} \le 1\{S_0 = 1\}.$$

The two indicators have equal conditional integrals by the preceding display, so the ordered indicators are equal almost surely on $\{C, X = x\}$. Therefore

$$P(S_0 = S_1,\ C \mid X = x) = P(C \mid X = x).$$

3. Under the same zero-gap condition, the argument in the first step also gives

$$\Gamma_x^* = \Gamma_x^0.$$

Indeed, every $\gamma_x \in \Gamma_x^*$ has exact rows and exact columns because $q_0(x) = q_1(x) = m(x)$, and every $\gamma_x \in \Gamma_x^0$ has bounded rows, bounded columns, and total mass $m(x)$. Combining this equality with the already proved

$$\Gamma_x^+ = \Gamma_x^-, \qquad \Gamma_x^- = \Gamma_x^0,$$

and with $\Gamma_x = \Gamma_x^*$ from Definition 15, gives

$$\Gamma_x = \Gamma_x^+, \qquad \Gamma_x = \Gamma_x^-, \qquad \Gamma_x = \Gamma_x^0.$$

4. Finally fix a threshold $t \in \mathcal{Y}$. The prefix sums $\ell_{\le t}(x)$ and $h_{\le t}(x)$, and the upper tail $h_{>t}(x)$, are those of Definition 16. The lower tail appearing in the statement is

$$\ell_{>t}(x) := \sum_{\substack{i \in \mathcal{Y} \\ t < i}} \ell_i(x).$$

Because $\mathcal{Y}$ is finite and each index lies in exactly one of the two sets $\{i : i \le t\}$ and $\{i : t < i\}$,

$$\ell_{\le t}(x) + \ell_{>t}(x) = \sum_{\substack{i \in \mathcal{Y} \\ i \le t}} \ell_i(x) + \sum_{\substack{i \in \mathcal{Y} \\ t < i}} \ell_i(x) = \sum_{i \in \mathcal{Y}} \ell_i(x) = q_0(x).$$

The same finite partition gives

$$h_{\le t}(x) + h_{>t}(x) = \sum_{\substack{j \in \mathcal{Y} \\ j \le t}} h_j(x) + \sum_{\substack{j \in \mathcal{Y} \\ t < j}} h_j(x) = \sum_{j \in \mathcal{Y}} h_j(x) = q_1(x).$$

When $\Delta q(x) = 0$, $q_0(x) = q_1(x)$, and hence

$$\ell_{\le t}(x) - h_{\le t}(x) = \big(q_0(x) - \ell_{>t}(x)\big) - \big(q_1(x) - h_{>t}(x)\big) = h_{>t}(x) - \ell_{>t}(x).$$

The four displayed conclusions are exactly the asserted cellwise conclusions for every x with $p_x > 0$. $\square$

Proof of Theorem 3. Let c be the observable capacity vector from Definition 14, and let

$$p_x = P_{\text{obs}}(X = x).$$

By Theorem 1, the capacity vector c is valid, the weights satisfy $p_x \geq 0$, and, for every supported cell $p_x > 0$,

$$\ell_i(x) = P(Y_0 = i, S_0 = 1, C \mid X = x), \qquad h_j(x) = P(Y_1 = j, S_1 = 1, C \mid X = x),$$

$$\Delta q(x) = P(S_1 = 1, C \mid X = x) - P(S_0 = 1, C \mid X = x), \qquad m(x) = P(S_0 = S_1 = 1, C \mid X = x).$$

It also gives the direction implications

$$\Delta q(x) > 0 \Rightarrow d(x) = +1, \qquad \Delta q(x) < 0 \Rightarrow d(x) = -1.$$

The capacity-identification step applies on supported cells. On cells with $p_x = 0$, the observed cell event has zero probability. Hence, for each instrument arm z,

$$P_{\text{obs}}(X = x, Z = z) = 0.$$

With the paper's zero-denominator convention for conditional probabilities, the two conditional probabilities entering every lower-arm contrast and every upper-arm contrast are therefore zero. Thus

$$\ell_i(x) = 0, \qquad h_j(x) = 0$$

for every i, j. Since

$$q_0(x) = \sum_i \ell_i(x), \qquad q_1(x) = \sum_j h_j(x), \qquad m(x) = \min\{q_0(x), q_1(x)\},$$

it follows that

$$q_0(x) = q_1(x) = m(x) = 0.$$

Choose as baseline the original compatible law generating P_{obs}. The following cell quantities use the finite-table totalization convention: if $p_x > 0$, they are the ordinary conditional probabilities given $X = x$; if $p_x = 0$, they are the corresponding zero-cell baseline table entries. Write the baseline complier mass as

$$\bar{c}_C(x) = P(C \mid X = x),$$

and write its never-taker and always-taker components as

$$\mathsf{N}_x(s_0, s_1, y_0, y_1) = P(D_0 = D_1 = 0, S_0 = s_0, S_1 = s_1, Y_0 = y_0, Y_1 = y_1 \mid X = x),$$

$$\mathsf{A}_x(s_0, s_1, y_0, y_1) = P(D_0 = D_1 = 1, S_0 = s_0, S_1 = s_1, Y_0 = y_0, Y_1 = y_1 \mid X = x).$$

For $p_x > 0$, the selected-complier masses are bounded by the complier mass, hence

$$q_0(x) \leq \bar{c}_C(x), \qquad q_1(x) \leq \bar{c}_C(x), \qquad \max\{q_0(x), q_1(x)\} \leq \bar{c}_C(x).$$

For $p_x = 0$, both capacity totals are zero, so the same inequality holds. Thus the baseline is compatible with P_{obs} and c, has conditional complier mass at least $\max\{q_0(x), q_1(x)\}$ in every cell, and carries the displayed never-taker and always-taker components.

Since $K \geq 3$, fix the reference outcome

$$\bar{y} = 0 \in \mathcal{Y}.$$

For each cell, apply the lower and upper threshold-flow constructions from Algorithm 1. Let

$$\gamma_x^L(i,j) \quad \text{and} \quad \gamma_x^U(i,j)$$

be the two survivor-complier couplings. They satisfy, for $\bullet \in \{L, U\}$,

$$\gamma_x^\bullet(i,j) \geq 0, \qquad \sum_j \gamma_x^\bullet(i,j) \leq \ell_i(x), \qquad \sum_i \gamma_x^\bullet(i,j) \leq h_j(x), \qquad \sum_{i,j} \gamma_x^\bullet(i,j) = m(x),$$

and their benefit masses are

$$\sum_{i<j} \gamma_x^L(i,j) = B_L(x), \qquad \sum_{i<j} \gamma_x^U(i,j) = B_U(x).$$

For each $\bullet \in \{L, U\}$, define the one-sided selected-complier residuals by

$$\rho_{x,i}^{0,\bullet} = \begin{cases} \ell_i(x) - \sum_j \gamma_x^\bullet(i,j), & \Delta q(x) < 0, \\ 0, & \Delta q(x) \geq 0, \end{cases}$$

$$\rho_{x,j}^{1,\bullet} = \begin{cases} h_j(x) - \sum_i \gamma_x^\bullet(i,j), & \Delta q(x) > 0, \\ 0, & \Delta q(x) \leq 0, \end{cases}$$

and the never-selected complier mass by

$$\nu_x = \bar{c}_C(x) - \max\{q_0(x), q_1(x)\}.$$

Then, for every cell and every $\bullet \in \{L, U\}$,

$$\sum_j \gamma_x^\bullet(i,j) + \rho_{x,i}^{0,\bullet} = \ell_i(x), \qquad \sum_i \gamma_x^\bullet(i,j) + \rho_{x,j}^{1,\bullet} = h_j(x).$$

Indeed, when $\Delta q(x) < 0$ or $\Delta q(x) > 0$, the corresponding identity is the definition of the residual. When $\Delta q(x) \geq 0$, the row upper bounds have total $q_0(x) = m(x)$, so every row bound binds; when $\Delta q(x) \leq 0$, the column upper bounds have total $q_1(x) = m(x)$, so every column bound binds.

Build, for each $\bullet \in \{L, U\}$, a cell table

$$\mathsf{t}_x^\bullet(d_0, d_1, s_0, s_1, y_0, y_1), \qquad (d_0, d_1, s_0, s_1) \in \{0,1\}^4, \quad y_0, y_1 \in \mathcal{Y}.$$

For $p_x > 0$, set

$$\mathsf{t}_x^\bullet(d_0, d_1, s_0, s_1, y_0, y_1) = \begin{cases} \mathsf{N}_x(s_0, s_1, y_0, y_1), & d_0 = d_1 = 0, \\ \gamma_x^\bullet(y_0, y_1), & d_0 = 0,\ d_1 = 1,\ s_0 = s_1 = 1, \\ \rho_{x,y_0}^{0,\bullet}, & d_0 = 0,\ d_1 = 1,\ s_0 = 1,\ s_1 = 0,\ y_1 = \bar{y}, \\ \rho_{x,y_1}^{1,\bullet}, & d_0 = 0,\ d_1 = 1,\ s_0 = 0,\ s_1 = 1,\ y_0 = \bar{y}, \\ \nu_x, & d_0 = 0,\ d_1 = 1,\ s_0 = s_1 = 0,\ y_0 = y_1 = \bar{y}, \\ \mathsf{A}_x(s_0, s_1, y_0, y_1), & d_0 = d_1 = 1, \\ 0, & \text{otherwise.} \end{cases}$$

For $p_x = 0$, set

$$\mathfrak{t}_x^\bullet(d_0, d_1, s_0, s_1, y_0, y_1) = \begin{cases} 1, & d_0 = d_1 = s_0 = s_1 = 0,\ y_0 = y_1 = \bar{y}, \\ 0, & \text{otherwise.} \end{cases}$$

The nonnegative entries sum to one in every cell: in supported cells the survivor, one-sided, and never-selected complier entries sum to $\bar{c}_C(x)$, while Assumption 6 sets the defier stratum $\{D_0 = 1, D_1 = 0\}$ to zero, so the never-taker, complier, and always-taker components exhaust the principal-stratum partition; in unsupported cells the table is a point mass.

Let $P^\bullet$ be the law obtained by drawing X with probabilities p_x, drawing Z conditionally on $X = x$ with the same instrument propensity $\pi(x)$ as P_{obs}, and drawing $(D_0, D_1, S_0, S_1, Y_0, Y_1)$ conditionally on $X = x$ according to $\mathfrak{t}_x^\bullet$, independently of Z given X. The associated slate sets

$$D = D_Z, \qquad S = S_D, \qquad Y = Y_D \text{ on } \{S = 1\}.$$

The conditional product construction gives conditional instrument independence. The displayed definitions of D, S, and Y give treatment consistency, selection consistency, and outcome consistency. The entry with $D_0 = 1, D_1 = 0$ has zero mass in every cell of $\mathfrak{t}_x^\bullet$; hence the normalized cell table has no defier mass, and the induced slate satisfies treatment monotonicity.

The weak selection direction is also preserved. If $p_x > 0$ and $d(x) = +1$, the implication above gives $\Delta q(x) < 0$ as impossible, hence

$$\rho_{x,i}^{0,\bullet} = 0 \quad \text{for all } i,$$

so the complier stratum with $(S_0, S_1) = (1, 0)$ has zero mass. If $p_x > 0$ and $d(x) = -1$, then $\Delta q(x) > 0$ is excluded and

$$\rho_{x,j}^{1,\bullet} = 0 \quad \text{for all } j,$$

so the complier stratum with $(S_0, S_1) = (0, 1)$ has zero mass. Unsupported cells have $p_x = 0$, so they contribute no conditional positive-mass case.

The observed law is unchanged. For supported cells, the selected-complier margins displayed above give exactly the observable capacities $\ell_i(x)$ and $h_j(x)$; the never-taker and always-taker margins are the baseline ones; and the complier total is

$$\sum_{i,j} \gamma_x^\bullet(i, j) + \sum_i \rho_{x,i}^{0,\bullet} + \sum_j \rho_{x,j}^{1,\bullet} + \nu_x = \bar{c}_C(x).$$

For $z \in \{0, 1\}$, put

$$\varpi_z(x) = \begin{cases} 1 - \pi(x), & z = 0, \\ \pi(x), & z = 1. \end{cases}$$

Assumption 5 gives $\varpi_z(x) > 0$ whenever $p_x > 0$. For treatment state $\mathfrak{d} \in \{0, 1\}$, selection state $\mathfrak{s} \in \{0, 1\}$, and reported outcome $\upsilon \in \{\varnothing\} \cup \mathcal{Y}$, define

$$R_{\mathfrak{d}} = \begin{cases} \varnothing, & S_{\mathfrak{d}} = 0, \\ Y_{\mathfrak{d}}, & S_{\mathfrak{d}} = 1, \end{cases}$$

and define the original latent tail margin

$$\lambda_x(z, \mathfrak{d}, \mathfrak{s}, \upsilon) = P(D_z = \mathfrak{d},\ S_{\mathfrak{d}} = \mathfrak{s},\ R_{\mathfrak{d}} = \upsilon \mid X = x).$$

For the constructed table, define the decoded observed margin

$$\Lambda_x^{\bullet}(z,\mathfrak{d},\mathfrak{s},\upsilon)=\sum_{d_0,d_1,s_0,s_1,y_0,y_1}\mathbf{1}\left\{\begin{array}{c} d_z=\mathfrak{d},\ \ s_{\mathfrak{d}}=\mathfrak{s}, \\ \upsilon=\varnothing \text{ if } s_{\mathfrak{d}}=0, \quad \upsilon=y_{\mathfrak{d}} \text{ if } s_{\mathfrak{d}}=1 \end{array}\right\}\mathsf{t}_x^{\bullet}(d_0,d_1,s_0,s_1,y_0,y_1).$$

For selected reported outcomes, the table definitions and the displayed selected-complier margins give

$$\Lambda_x^{\bullet}(0,0,1,i)=\sum_{s_1,y_1}\mathsf{N}_x(1,s_1,i,y_1)+\ell_i(x),$$

$$\Lambda_x^{\bullet}(0,1,1,j)=\sum_{s_0,y_0}\mathsf{A}_x(s_0,1,y_0,j),$$

$$\Lambda_x^{\bullet}(1,0,1,i)=\sum_{s_1,y_1}\mathsf{N}_x(1,s_1,i,y_1),$$

and

$$\Lambda_x^{\bullet}(1,1,1,j)=\sum_{s_0,y_0}\mathsf{A}_x(s_0,1,y_0,j)+h_j(x).$$

The corresponding original latent margins have the same four values: Theorem 1 identifies $\ell_i(x)$ and $h_j(x)$ with the selected complier outcome masses on supported cells, while Assumption 6 makes the defier terms in the mixed treatment strata equal to zero. Hence

$$\Lambda_x^{\bullet}(z,\mathfrak{d},1,i)=\lambda_x(z,\mathfrak{d},1,i)$$

for every selected reported outcome i. The invalid observed-outcome margins also agree:

$$\Lambda_x^{\bullet}(z,\mathfrak{d},0,i)=\lambda_x(z,\mathfrak{d},0,i)=0, \qquad \Lambda_x^{\bullet}(z,\mathfrak{d},1,\varnothing)=\lambda_x(z,\mathfrak{d},1,\varnothing)=0.$$

It remains only the unselected reported outcome. The constructed table has the same conditional mass for each principal treatment event as the baseline, so

$$\Lambda_x^{\bullet}(z,\mathfrak{d},0,\varnothing)+\sum_i\Lambda_x^{\bullet}(z,\mathfrak{d},1,i)=P(D_z=\mathfrak{d}\mid X=x).$$

The original latent tails form the same finite partition,

$$\lambda_x(z,\mathfrak{d},0,\varnothing)+\sum_i\lambda_x(z,\mathfrak{d},1,i)=P(D_z=\mathfrak{d}\mid X=x).$$

Together with the selected-margin identities, this gives

$$\Lambda_x^{\bullet}(z,\mathfrak{d},0,\varnothing)=\lambda_x(z,\mathfrak{d},0,\varnothing).$$

Thus, for every supported cell and every observable singleton,

$$P^{\bullet}(X=x,Z=z,D=\mathfrak{d},S=\mathfrak{s},Y=\upsilon)=p_x\varpi_z(x)\Lambda_x^{\bullet}(z,\mathfrak{d},\mathfrak{s},\upsilon)=p_x\varpi_z(x)\lambda_x(z,\mathfrak{d},\mathfrak{s},\upsilon)=P(X=x,Z=z,D=\mathfrak{d}$$

The first equality is the decoded-table construction; the last equality uses Assumptions 1 to 4 and the positive arm mass supplied by Assumption 5. Cells with $p_x=0$ have zero observed mass under both laws. Hence

$$P_{\text{obs}}^{L}=P_{\text{obs}}, \qquad P_{\text{obs}}^{U}=P_{\text{obs}}.$$

Because the observed law is the same, the inequalities in Assumption 5 and the mass condition in Assumption 8 transfer unchanged.

It remains to compute the target value. In supported cells, the survivor coupling of $P^\bullet$ is $\gamma_x^\bullet$; in unsupported cells its contribution is multiplied by $p_x = 0$. Hence

$$\sum_x p_x \sum_{i,j} \gamma_x^\bullet(i,j) = \sum_x p_x m(x) = M.$$

The benefit numerators are

$$\sum_x p_x \sum_{i<j} \gamma_x^L(i,j) = \sum_x p_x B_L(x), \qquad \sum_x p_x \sum_{i<j} \gamma_x^U(i,j) = \sum_x p_x B_U(x).$$

Since $M > 0$, division by M and the endpoint formula in Algorithm 1 give

$$P^L(Y_1 > Y_0 \mid S_1 = S_0 = 1, C) = \theta_L, \qquad P^U(Y_1 > Y_0 \mid S_1 = S_0 = 1, C) = \theta_U.$$

Thus P^L and P^U are endpoint witnesses.

Finally, the realization statement follows cell by cell. If $p_x > 0$, the conditional table of P^L equals the lower completion components just defined, and the conditional table of P^U equals the upper completion components. If $p_x = 0$, the zero-cell baseline convention gives

$$\ell_i(x) = 0, \qquad h_j(x) = 0, \qquad \bar{c}_C(x) = 0,$$

and all baseline never-taker and always-taker component masses in that cell are zero. Therefore the lower and upper threshold-flow completions are both the zero completion,

$$\gamma_x^L(i,j) = \gamma_x^U(i,j) = 0, \qquad \rho_{x,i}^{0,L} = \rho_{x,i}^{0,U} = 0, \qquad \rho_{x,j}^{1,L} = \rho_{x,j}^{1,U} = 0, \qquad \nu_x = 0,$$

for all i, j. The realization identities in a zero-probability cell are evaluated under the same zero conditioning-cell mass convention: every conditional baseline component used by the completion is zero, and the zero-flow rule sets the lower and upper threshold completions to the zero array. Thus the survivor, one-sided selected-complier, never-selected, never-taker, and always-taker component functionals of P^L and P^U are the zero functionals displayed above. Hence P^L realizes the lower threshold latent completion and P^U realizes the upper threshold latent completion for every covariate cell. □

Proof of Theorem 4. 1. Fix a covariate cell x with $p_x > 0$, and form the observable capacities c from P_{obs}. By Theorem 1, these capacities are valid and, in this supported cell,

$$\Delta q(x) = P(S_1 = 1, C \mid X = x) - P(S_0 = 1, C \mid X = x).$$

The no-selection submodel implies the two selected-complier masses equal the complier mass:

$$P(S_0 = 1, C \mid X = x) = P(C \mid X = x), \qquad P(S_1 = 1, C \mid X = x) = P(C \mid X = x).$$

Indeed, if $P(C \mid X = x) > 0$, then

$$P(S_0 = S_1 = 1 \mid C, X = x) = 1,$$

and the event $\{S_0 = S_1 = 1, C\}$ is contained in each of $\{S_0 = 1, C\}$ and $\{S_1 = 1, C\}$, while both are contained in C. Thus both conditional probabilities are squeezed to $P(C \mid X = x)$. If $P(C \mid X = x) = 0$, the same containment in C gives both selected-complier masses equal to zero. Substituting these two identities in the displayed formula for $\Delta q(x)$ gives

$$\Delta q(x) = 0.$$

2. It remains to identify the cellwise feasible set. Since $\Delta q(x) = q_1(x) - q_0(x)$, the zero-gap identity gives

$$q_0(x) = q_1(x).$$

The exact survivor-complier mass is

$$m(x) = \min\{q_0(x), q_1(x)\} = q_0(x) = q_1(x).$$

For any $\gamma_x \in \Gamma_x$, Definition 15 gives

$$\sum_j \gamma_{ij,x} \le \ell_i(x), \qquad \sum_i \gamma_{ij,x} \le h_j(x), \qquad \sum_{i,j} \gamma_{ij,x} = m(x).$$

Summing the row inequalities gives

$$m(x) = \sum_{i,j} \gamma_{ij,x} = \sum_i \sum_j \gamma_{ij,x} \le \sum_i \ell_i(x) = q_0(x) = m(x),$$

so every row inequality binds. The same argument with columns gives

$$m(x) = \sum_{i,j} \gamma_{ij,x} = \sum_j \sum_i \gamma_{ij,x} \le \sum_j h_j(x) = q_1(x) = m(x),$$

so every column inequality binds. Hence $\gamma_x \in \Gamma_x^0$. Conversely, every $\gamma_x \in \Gamma_x^0$ has both exact margins, and therefore has total mass $q_0(x) = q_1(x) = m(x)$ and satisfies the defining inequalities of Γ_x. Thus

$$\Gamma_x = \Gamma_x^0.$$

3. Now assume $m(x) > 0$. By Definition 16,

$$B_L(x) = \max\left\{0,\ \max_{t \in \mathcal{Y}} \big[\ell_{\le t}(x) - h_{\le t}(x)\big] + \min\{\Delta q(x), 0\}\right\}.$$

Using $\Delta q(x) = 0$, this becomes

$$B_L(x) = \max\left\{0,\ \max_{t \in \mathcal{Y}} \big[\ell_{\le t}(x) - h_{\le t}(x)\big]\right\}.$$

Division by the positive number $m(x)$ commutes with the finite maximum: for any finite nonempty set of real numbers $\{a_r\}$,

$$\frac{\max_r a_r}{m(x)} = \max_r \frac{a_r}{m(x)}.$$

Applying this to the displayed maximum, with the zero candidate included as $0/m(x) = 0$, gives

$$\frac{B_L(x)}{m(x)} = \max\left\{0,\ \max_{t \in \mathcal{Y}} \frac{\ell_{\le t}(x) - h_{\le t}(x)}{m(x)}\right\}.$$

4. Similarly, Definition 16 gives

$$B_U(x) = \min\left\{m(x),\ \min_{t\in\mathcal{Y}}\left[\ell_{<t}(x) + h_{>t}(x)\right]\right\}.$$

Division by $m(x) > 0$ commutes with the finite minimum:

$$\frac{\min_r a_r}{m(x)} = \min_r \frac{a_r}{m(x)}.$$

The cap contributes $m(x)/m(x) = 1$, and the remaining candidates scale termwise. Therefore

$$\frac{B_U(x)}{m(x)} = \min\left\{1,\ \min_{t\in\mathcal{Y}} \frac{\ell_{<t}(x) + h_{>t}(x)}{m(x)}\right\}.$$

These two normalized expressions are the fixed-marginal ordinal strict-benefit formulas recorded by [Lu et al., 2018, Proposition 2, equation (6), p. 546].

□

Proof of Theorem 5. Let

$$c = \{\ell_i(x), h_j(x), q_0(x), q_1(x), \Delta q(x), m(x)\}_{x,i,j}$$

be the observable capacity vector associated with P_{obs}. By Theorem 1, c is a valid capacity vector, $p_x \geq 0$ for every x, and the positive aggregate survivor condition gives

$$M = \sum_{x\in\mathcal{X}} p_x m(x) > 0.$$

For a coupling γ_x, write

$$b_x(\gamma_x) = \sum_{i\in\mathcal{Y}} \sum_{j\in\mathcal{Y}:\, i<j} \gamma_{ij,x}.$$

1. Fix a cell x with $p_x > 0$. If W is any feasible full law with observed law P_{obs}, its survivor-complier coupling

$$\gamma_{ij,x} = P_W(Y_0 = i, Y_1 = j, S_0 = S_1 = 1, C \mid X = x)$$

is nonnegative. Its row sums are bounded by the lower-arm selected-complier capacities, its column sums are bounded by the upper-arm selected-complier capacities, and its total mass is the exact survivor-complier mass:

$$\sum_j \gamma_{ij,x} \leq \ell_i(x), \qquad \sum_i \gamma_{ij,x} \leq h_j(x), \qquad \sum_{i,j} \gamma_{ij,x} = m(x).$$

Thus every feasible full law projects into Γ_x^*.

Conversely, take any $\gamma_x \in \Gamma_x^*$. Define a cellwise family of couplings by

$$\widetilde{\gamma}_{x'} = \begin{cases} \gamma_x, & x' = x, \\ \gamma_{x'}^L, & x' \neq x, \end{cases}$$

where $\gamma^L_{x'}$ is the lower threshold-flow coupling from Algorithm 1. By Theorem 7,

$$\gamma^L_{x'} \in \Gamma^*_{x'}$$

for every x', and the chosen γ_x lies in Γ^*_x by hypothesis. Hence

$$\widetilde{\gamma}_{x'} \in \Gamma^*_{x'} \qquad \text{for every } x' \in \mathcal{X}.$$

We record the realization step as a proof-local realization claim. Let $\bar{\gamma} = \{\bar{\gamma}_{x'} : x' \in \mathcal{X}\}$ be any family satisfying

$$\bar{\gamma}_{x'} \in \Gamma^*_{x'} \qquad \text{for every } x' \in \mathcal{X}.$$

Choose one compatible generating law $\bar{P}$ for P_{obs}. For each supported cell x', put

$$\mu_C(x') = \bar{P}(C \mid X = x')$$

and choose a fixed outcome level $y_\star \in \mathcal{Y}$. Define the one-sided residual masses

$$\rho^0_i(x') = \begin{cases} \ell_i(x') - \sum_j \bar{\gamma}_{ij,x'}, & \Delta q(x') < 0, \\ 0, & \Delta q(x') \geq 0, \end{cases}$$

$$\rho^1_j(x') = \begin{cases} h_j(x') - \sum_i \bar{\gamma}_{ij,x'}, & \Delta q(x') > 0, \\ 0, & \Delta q(x') \leq 0, \end{cases}$$

and

$$\rho^{\varnothing}(x') = \mu_C(x') - \max\{q_0(x'), q_1(x')\}.$$

These quantities are nonnegative on every supported cell. The nonnegativity of ρ^0_i and ρ^1_j follows from the row and column inequalities defining $\Gamma^*_{x'}$. For $\rho^{\varnothing}$, Theorem 1 and Definition 14 give

$$q_0(x') = \sum_i \ell_i(x') = \sum_i \bar{P}(Y_0 = i, S_0 = 1, C \mid X = x') = \bar{P}(S_0 = 1, C \mid X = x') \leq \bar{P}(C \mid X = x') = \mu_C(x')$$

and similarly

$$q_1(x') = \sum_j h_j(x') = \sum_j \bar{P}(Y_1 = j, S_1 = 1, C \mid X = x') = \bar{P}(S_1 = 1, C \mid X = x') \leq \bar{P}(C \mid X = x') = \mu_C(x')$$

The equalities summing over outcomes use the finite ordered outcome support, and the inequalities use inclusion of each selected-complier event in the complier event. Hence $\max\{q_0(x'), q_1(x')\} \leq \mu_C(x')$. The residual formulas are used on supported cells; cells with $p_{x'} = 0$ are assigned a zero-weight normalized table below.

The residuals complete $\bar{\gamma}_{x'}$ to the selected-complier margins. If $\Delta q(x') \geq 0$, then $m(x') = q_0(x')$, and the row inequalities in $\Gamma^*_{x'}$ have total slack zero; if $\Delta q(x') < 0$, the definition of ρ^0_i fills the row slack. Thus, in all cases,

$$\sum_j \bar{\gamma}_{ij,x'} + \rho^0_i(x') = \ell_i(x') \qquad \text{for every } i.$$

Similarly, if $\Delta q(x') \le 0$, then $m(x') = q_1(x')$, and the column inequalities have total slack zero; if $\Delta q(x') > 0$, the definition of ρ_j^1 fills the column slack. Hence

$$\sum_i \bar{\gamma}_{ij,x'} + \rho_j^1(x') = h_j(x') \qquad \text{for every } j.$$

The complier masses also add up correctly:

$$\sum_{i,j} \bar{\gamma}_{ij,x'} + \sum_i \rho_i^0(x') + \sum_j \rho_j^1(x') + \rho^{\varnothing}(x') = \mu_C(x').$$

Indeed, the left side equals $m(x')+q_0(x')-m(x')+\mu_C(x')-q_0(x')$ when $\Delta q(x') < 0$, equals $m(x')+q_1(x')-m(x')+\mu_C(x')-q_1(x')$ when $\Delta q(x') > 0$, and equals $m(x')+\mu_C(x')-m(x')$ when $\Delta q(x') = 0$.

Construct a latent table in cell x' by assigning $\bar{\gamma}_{ij,x'}$ to compliers with $(S_0, S_1) = (1, 1)$ and $(Y_0, Y_1) = (i, j)$, assigning $\rho_i^0(x')$ to compliers with $(S_0, S_1) = (1, 0)$, $Y_0 = i$, and $Y_1 = y_\star$, assigning $\rho_j^1(x')$ to compliers with $(S_0, S_1) = (0, 1)$, $Y_0 = y_\star$, and $Y_1 = j$, and assigning $\rho^{\varnothing}(x')$ to compliers with $(S_0, S_1) = (0, 0)$ and $(Y_0, Y_1) = (y_\star, y_\star)$. Keep the never-taker and always-taker conditional latent components from $\bar{P}$. On cells with $p_{x'} = 0$, choose any normalized latent table, since those cells have zero joint weight.

Combining these conditional tables with the original cell weights $p_{x'}$ and instrument propensities gives a full law $W^{\bar{\gamma}}$. The preceding displays show that the selected-complier outcome margins are $\ell_i(x')$ and $h_j(x')$, the complier total is $\mu_C(x')$, and the noncomplier components are unchanged; therefore the observed law is P_{obs}. The construction makes the instrument conditionally independent given X, and the observed variables are generated from the potential variables by the consistency and exclusion equations. Treatment monotonicity is inherited from the retained principal strata. The one-sided residuals respect the weak-selection direction because Theorem 1 gives the sign implications for $\Delta q(x')$. Thus $W^{\bar{\gamma}}$ is feasible, and for every supported cell

$$P_{W^{\bar{\gamma}}}(Y_0 = i, Y_1 = j, S_0 = S_1 = 1, C \mid X = x') = \bar{\gamma}_{ij,x'}.$$

Moreover, since $\sum_{i,j} \bar{\gamma}_{ij,x'} = m(x')$, the same cell decomposition gives

$$\Pr_{W^{\bar{\gamma}}}(Y_1 > Y_0 \mid S_0 = S_1 = 1, C) = \frac{\sum_{x'} p_{x'} b_{x'}(\bar{\gamma}_{x'})}{M}.$$

Applying this proof-local realization claim to $\widetilde{\gamma}$, the realized coupling in cell x is γ_x. Therefore the projection set of feasible full laws is exactly Γ_x^*.

2. Again fix x with $p_x > 0$. If $\Delta q(x) > 0$, then $q_0(x) \le q_1(x)$. For any $\gamma_x \in \Gamma_x^*$,

$$\sum_i \sum_j \gamma_{ij,x} = m(x) = q_0(x) = \sum_i \ell_i(x),$$

while each row satisfies $\sum_j \gamma_{ij,x} \le \ell_i(x)$. Since all row deficits are nonnegative and their sum is zero,

$$\sum_j \gamma_{ij,x} = \ell_i(x) \qquad \text{for every } i.$$

Thus $\Gamma_x^* \subseteq \Gamma_x^+$, and the reverse inclusion follows because exact rows imply total mass $q_0(x) = m(x)$ while retaining nonnegativity and the column bounds. Hence

$$\Gamma_x^* = \Gamma_x^+.$$

If $\Delta q(x) < 0$, then $q_1(x) \le q_0(x)$. The same argument with columns gives

$$\sum_i \gamma_{ij,x} = h_j(x) \qquad \text{for every } j$$

for every $\gamma_x \in \Gamma_x^*$, and exact columns conversely give total mass $q_1(x) = m(x)$. Hence

$$\Gamma_x^* = \Gamma_x^-.$$

If $\Delta q(x) = 0$, then $q_0(x) = q_1(x) = m(x)$, so both the row and column deficit sums vanish. Consequently

$$\sum_j \gamma_{ij,x} = \ell_i(x) \quad \text{and} \quad \sum_i \gamma_{ij,x} = h_j(x)$$

for every $\gamma_x \in \Gamma_x^*$, and the converse is immediate from the same total-mass identity. Therefore

$$\Gamma_x^* = \Gamma_x^0.$$

3. For every $\gamma_x \in \Gamma_x^*$, the threshold-cut inequalities are

$$B_L(x) \le b_x(\gamma_x) \le B_U(x).$$

Indeed, the upper endpoint inequality follows from

$$b_x(\gamma_x) \le m(x)$$

and the threshold cut cover. Fix t. For every benefit pair $i < j$, at least one of $i < t$ or $t < j$ holds; otherwise $t \le i < j \le t$, a contradiction. Since γ_x is entrywise nonnegative,

$$b_x(\gamma_x) = \sum_i \sum_{j:\, i<j} \gamma_{ij,x} \le \sum_{i<t} \sum_j \gamma_{ij,x} + \sum_{j>t} \sum_i \gamma_{ij,x}.$$

The row and column bounds in Γ_x^* then give

$$\sum_{i<t} \sum_j \gamma_{ij,x} + \sum_{j>t} \sum_i \gamma_{ij,x} \le \ell_{<t}(x) + h_{>t}(x).$$

Together with the bound by $m(x)$, this gives $b_x(\gamma_x) \le B_U(x)$. For the lower bound, nonnegativity gives $0 \le b_x(\gamma_x)$. Fix t. Entry by entry,

$$\big(\mathbf{1}\{i \le t\} - \mathbf{1}\{j \le t\}\big)\gamma_{ij,x} \le \mathbf{1}\{i < j\}\gamma_{ij,x},$$

where $\mathbf{1}\{\cdot\}$ denotes an indicator. If $i \le t < j$, the two sides are equal because $i < j$; if $i \le t$ and $j \le t$, the left side is zero; if $t < i$ and $j \le t$, the left side is nonpositive; and if

$t < i$ and $t < j$, the left side is zero. Nonnegativity of $\gamma_{ij,x}$ handles the nonpositive and zero cases. Summing this entrywise comparison gives

$$\sum_{i \le t} \sum_{j} \gamma_{ij,x} - \sum_{j \le t} \sum_{i} \gamma_{ij,x} \le b_x(\gamma_x).$$

When $\Delta q(x) \ge 0$, the row sums bind by the preceding step and $\sum_{j \le t} \sum_i \gamma_{ij,x} \le h_{\le t}(x)$, so

$$\ell_{\le t}(x) - h_{\le t}(x) \le b_x(\gamma_x).$$

When $\Delta q(x) < 0$, the column sums bind and the total row slack is

$$\sum_i \left(\ell_i(x) - \sum_j \gamma_{ij,x} \right) = q_0(x) - m(x) = -\Delta q(x).$$

The prefix row slack is at most this total slack, and therefore

$$\ell_{\le t}(x) - h_{\le t}(x) + \Delta q(x) \le b_x(\gamma_x).$$

Combining the two cases gives

$$\ell_{\le t}(x) - h_{\le t}(x) + \min\{\Delta q(x), 0\} \le b_x(\gamma_x)$$

for every t, and hence $B_L(x) \le b_x(\gamma_x)$.

For the lower and upper threshold-flow couplings from Algorithm 1, Theorem 7 gives

$$\gamma_x^L \in \Gamma_x^*, \qquad b_x(\gamma_x^L) = B_L(x), \qquad \gamma_x^U \in \Gamma_x^*, \qquad b_x(\gamma_x^U) = B_U(x).$$

Thus

$$\inf_{\gamma_x \in \Gamma_x^*} b_x(\gamma_x) = B_L(x), \qquad \sup_{\gamma_x \in \Gamma_x^*} b_x(\gamma_x) = B_U(x).$$

4. If $m(x) = 0$, then any $\gamma_x \in \Gamma_x^*$ has

$$\sum_{i,j} \gamma_{ij,x} = 0$$

and all entries are nonnegative. Hence every entry is zero, so $\Gamma_x^* \subseteq \{0\}$. Conversely, the zero coupling is nonnegative, has all row and column sums equal to zero, and has total mass zero. Since the capacity vector is valid, $\ell_i(x) \ge 0$ and $h_j(x) \ge 0$ for every i, j, so the row and column bounds hold; the total-mass constraint holds because $m(x) = 0$. Thus $0 \in \Gamma_x^*$, and therefore

$$\Gamma_x^* = \{0\}.$$

5. Define

$$\theta_L = \frac{\sum_x p_x B_L(x)}{M}, \qquad \theta_U = \frac{\sum_x p_x B_U(x)}{M}.$$

The cellwise endpoint inequality $B_L(x) \le B_U(x)$ follows by applying the preceding bounds to γ_x^L, and $p_x \ge 0$ gives

$$\sum_x p_x B_L(x) \le \sum_x p_x B_U(x), \qquad \theta_L \le \theta_U.$$

Let W be any feasible full law with observed law P_{obs}. Equality of observed laws transfers the cell weights:

$$p_x^W = p_x.$$

For supported cells, the projected coupling γ_x^W lies in Γ_x^*; for unsupported cells, $p_x = 0$. Therefore

$$\sum_x p_x B_L(x) \le \sum_x p_x^W b_x(\gamma_x^W) \le \sum_x p_x B_U(x),$$

and the denominator of the benefit probability is

$$\sum_x p_x^W \sum_{i,j} \gamma_{ij,x}^W = \sum_x p_x m(x) = M.$$

The aggregate identity for the benefit probability gives

$$\Pr_W(Y_1 > Y_0 \mid S_0 = S_1 = 1, C) = \frac{\sum_x p_x^W b_x(\gamma_x^W)}{M},$$

so every feasible value belongs to $[\theta_L, \theta_U]$.

For the reverse inclusion, take a real number $\vartheta \in [\theta_L, \theta_U]$. If $\theta_L = \theta_U$, choose the family $\bar{\gamma}_x = \gamma_x^L$. The proof-local realization lemma gives a feasible law W^L with

$$\Pr_{W^L}(Y_1 > Y_0 \mid S_0 = S_1 = 1, C) = \frac{\sum_x p_x b_x(\gamma_x^L)}{M} = \theta_L = \vartheta.$$

If $\theta_L < \theta_U$, set

$$\lambda = \frac{\vartheta - \theta_L}{\theta_U - \theta_L}, \qquad 0 \le \lambda \le 1,$$

and define

$$\gamma_x^\lambda = (1-\lambda)\gamma_x^L + \lambda\gamma_x^U.$$

Since nonnegativity, the row bounds, the column bounds, and the exact total-mass identity are preserved by convex combinations,

$$\gamma_x^\lambda \in \Gamma_x^* \qquad \text{for every } x.$$

Linearity of b_x gives

$$b_x(\gamma_x^\lambda) = (1-\lambda)B_L(x) + \lambda B_U(x).$$

The proof-local realization lemma therefore yields a feasible law W^λ with

$$\Pr_{W^\lambda}(Y_1 > Y_0 \mid S_0 = S_1 = 1, C) = \frac{\sum_x p_x\{(1-\lambda)B_L(x) + \lambda B_U(x)\}}{M} = (1-\lambda)\theta_L + \lambda\theta_U = \vartheta.$$

Hence the feasible values are exactly $[\theta_L, \theta_U]$, which is the interval in Definition 17.

□

Proof of Theorem 6. Let c be the observable capacity vector obtained from $P^\star_{\text{obs}}$ on the singleton covariate space $\{x^\star\}$ and outcome levels $\{0, 1, 2\}$.

1. Consider the latent allocation from Definition 20, written with the canonical hidden-outcome completion

(D_0, D_1)	(S_0, S_1)	(Y_0, Y_1)	mass
$(0, 1)$	$(1, 1)$	$(0, 0)$	$3/40$
$(0, 1)$	$(1, 1)$	$(1, 1)$	$1/10$
$(0, 1)$	$(1, 1)$	$(2, 2)$	$3/40$
$(0, 1)$	$(0, 1)$	$(0, 0)$	$1/40$
$(0, 1)$	$(0, 1)$	$(0, 1)$	$3/20$
$(0, 1)$	$(0, 1)$	$(0, 2)$	$3/40$
$(0, 1)$	$(0, 0)$	$(0, 0)$	$1/2.$

The masses are nonnegative and sum to

$$\frac{3}{40} + \frac{1}{10} + \frac{3}{40} + \frac{1}{40} + \frac{3}{20} + \frac{3}{40} + \frac{1}{2} = 1.$$

Let Z be independent of these potential variables with $P(Z = 0) = P(Z = 1) = 1/2$, and set $D = D_Z$, $S = S_D$, and $Y = Y_D$ on selected units. Every latent atom has $D_0 = 0$, $D_1 = 1$, and $S_0 \leq S_1$. The instrument propensity in the only cell is $1/2$, so the overlap inequalities at $\varepsilon_Z = 1/4$ are

$$0 < \frac{1}{4} < \frac{1}{2}, \qquad \frac{1}{4} \leq \frac{1}{2} \leq \frac{3}{4}.$$

The always-selected complier mass is

$$\frac{3}{40} + \frac{1}{10} + \frac{3}{40} = \frac{1}{4} > 0.$$

Thus the law satisfies the structural requirements collected in Definition 13, with the single cell retained and weak selection direction $S_0 \leq S_1$.

Under $Z = 0$, the selected outcome masses are $3/40, 1/10, 3/40$, and the unselected mass is $3/4$. Multiplying by $P(Z = 0) = 1/2$ gives the first four observed masses

$$\frac{3}{80}, \quad \frac{1}{20}, \quad \frac{3}{80}, \quad \frac{3}{8}.$$

Under $Z = 1$, the selected outcome masses are

$$\left(\frac{3}{40}, \frac{1}{10}, \frac{3}{40}\right) + \left(\frac{1}{40}, \frac{3}{20}, \frac{3}{40}\right) = \left(\frac{1}{10}, \frac{1}{4}, \frac{3}{20}\right),$$

and the unselected mass is $1/2$. Multiplying by $P(Z = 1) = 1/2$ gives the last four observed masses

$$\frac{1}{20}, \quad \frac{1}{8}, \quad \frac{3}{40}, \quad \frac{1}{4}.$$

Hence this three-level slate system realizes $P^{\star}_{\text{obs}}$, has the displayed encoded full law, and satisfies $D_0 = 0$, $D_1 = 1$, and $S_0 \leq S_1$ almost surely.

2. By Definition 14, for $i \in \{0, 1, 2\}$,

$$\ell_i(x^{\star}) = P(D = 0, S = 1, Y = i \mid Z = 0, X = x^{\star}) - P(D = 0, S = 1, Y = i \mid Z = 1, X = x^{\star}).$$

The second conditional probability is 0 for each i, and the $Z = 0$ denominator is 1/2. Therefore

$$(\ell_0(x^\star), \ell_1(x^\star), \ell_2(x^\star)) = \left(\frac{(3/80)}{1/2}, \frac{(1/20)}{1/2}, \frac{(3/80)}{1/2}\right) = \left(\frac{3}{40}, \frac{1}{10}, \frac{3}{40}\right).$$

3. Similarly, for $j \in \{0, 1, 2\}$,

$$h_j(x^\star) = P(D = 1, S = 1, Y = j \mid Z = 1, X = x^\star) - P(D = 1, S = 1, Y = j \mid Z = 0, X = x^\star).$$

The second conditional probability is 0 for each j, and the $Z = 1$ denominator is 1/2. Hence

$$(h_0(x^\star), h_1(x^\star), h_2(x^\star)) = \left(\frac{(1/20)}{1/2}, \frac{(1/8)}{1/2}, \frac{(3/40)}{1/2}\right) = \left(\frac{1}{10}, \frac{1}{4}, \frac{3}{20}\right).$$

4. Summing the displayed capacities gives

$$q_0(x^\star) = \frac{3}{40} + \frac{1}{10} + \frac{3}{40} = \frac{1}{4}, \qquad q_1(x^\star) = \frac{1}{10} + \frac{1}{4} + \frac{3}{20} = \frac{1}{2}.$$

Thus

$$m(x^\star) = \min\{q_0(x^\star), q_1(x^\star)\} = \frac{1}{4}, \qquad \Delta q(x^\star) = q_1(x^\star) - q_0(x^\star) = \frac{1}{4}.$$

Using the threshold-cut formulas in Definition 16, the lower-cut correction is

$$\min\{\Delta q(x^\star), 0\} = 0.$$

The three prefix differences are

$$\begin{aligned}
\ell_{\le 0}(x^\star) - h_{\le 0}(x^\star) &= -\frac{1}{40}, \\
\ell_{\le 1}(x^\star) - h_{\le 1}(x^\star) &= -\frac{7}{40}, \\
\ell_{\le 2}(x^\star) - h_{\le 2}(x^\star) &= -\frac{1}{4}.
\end{aligned}$$

The maximum with 0 is therefore

$$B_L(x^\star) = 0.$$

For the upper cut,

$$m(x^\star) = \frac{10}{40},$$

and the strict-prefix/tail sums are

$$\begin{aligned}
\ell_{<0}(x^\star) + h_{>0}(x^\star) &= \frac{2}{5}, \\
\ell_{<1}(x^\star) + h_{>1}(x^\star) &= \frac{9}{40}, \\
\ell_{<2}(x^\star) + h_{>2}(x^\star) &= \frac{7}{40}.
\end{aligned}$$

Taking the minimum gives

$$B_U(x^\star) = \min\left\{\frac{10}{40}, \frac{2}{5}, \frac{9}{40}, \frac{7}{40}\right\} = \frac{7}{40}.$$

5. The capacity vector is valid because each displayed coordinate of ℓ and h is nonnegative. The singleton cell has observed weight

$$p_{x^\star} = 1 \geq 0,$$

and the aggregate survivor-complier mass is

$$M = \sum_x p_x m(x) = 1 \cdot \frac{1}{4} = \frac{1}{4} > 0.$$

Since the observed law is induced by the latent slate system constructed above, it lies in the observed-law domain. By Definition 17,

$$\Theta_I(P^\star_{\text{obs}}) = \left[\frac{p_{x^\star} B_L(x^\star)}{M}, \frac{p_{x^\star} B_U(x^\star)}{M}\right] = \left[\frac{0}{1/4}, \frac{(7/40)}{1/4}\right] = \left[0, \frac{7}{10}\right].$$

6. The hypotheses of Theorem 3 hold for the realized three-level slate system: the structural restrictions and tie-safe survivor model hold by the first step, and $M = 1/4 > 0$ by the preceding step. Applying that result to $P^\star_{\text{obs}}$, overlap $1/4$, and the retained singleton cell gives full laws P^L and P^U, with three-level slate systems, that induce $P^\star_{\text{obs}}$ and attain the two endpoint values

$$0 \qquad \text{and} \qquad \frac{7}{10}.$$

Combining the six displayed conclusions gives all assertions of Theorem 6. □

Proof of Theorem 7. Choose

$$C_{\text{sparse}} = 75, \qquad C_{\text{dense}} = 77.$$

Both constants are positive. Fix a finite covariate support $\mathcal{X}$, an integer $K \geq 3$, and nonnegative capacity arrays $\ell_i(x)$ and $h_j(x)$. Write $N_{\mathcal{X}} = |\mathcal{X}|$. For one cell x, let

$$s_K = 64K + 32$$

be the fixed sparse cell budget charged by the implementation. Since $K \geq 3$,

$$s_K = 64K + 32 \leq 64K + 11K = 75K.$$

1. The actual sparse work in cell x is the sum of the threshold-cut scan, the lower nested-graph pass, the upper nested-graph pass, and the two completion passes. The threshold-cut scan uses

$$12K + 8$$

operations. If Rows and Cols are the current row and column lists, each nested pass performs at most

$$5\big(|\mathsf{Rows}| + |\mathsf{Cols}|\big) + 1$$

operations, and each completion pass performs at most

$$4\big(|\mathsf{Rows}| + |\mathsf{Cols}|\big) + 1$$

operations. Initially the row and column lists have lengths K and K. After either nested pass, the residual row and column lists have total length at most $2K$. Hence, for the actual cell count a_x,

$$\begin{aligned} a_x &\le (12K+8) + \{5(2K)+1\} + \{5(2K)+1\} + \{4(2K)+1\} + \{4(2K)+1\} \\ &= 48K + 12 \le 64K + 32 = s_K. \end{aligned}$$

Consequently

$$\sum_{x\in\mathcal{X}} a_x \le \sum_{x\in\mathcal{X}} s_K \le \sum_{x\in\mathcal{X}} 75K = 75\, N_{\mathcal{X}} K.$$

This proves the actual sparse bound. The charged sparse count is exactly

$$\sum_{x\in\mathcal{X}} s_K,$$

so the same calculation gives

$$\sum_{x\in\mathcal{X}} s_K \le 75\, N_{\mathcal{X}} K.$$

2. Since $K \ge 3$, in particular $1 \le K$, and therefore

$$K \le K^2.$$

Multiplying the sparse bounds by the nonnegative factor $75N_{\mathcal{X}}$ gives

$$75\, N_{\mathcal{X}} K \le 75\, N_{\mathcal{X}} K^2.$$

The dense actual implementation first performs the sparse computation and then materializes the two dense $K \times K$ cell matrices, so its actual count is

$$\sum_{x\in\mathcal{X}} a_x + 2N_{\mathcal{X}} K^2.$$

Using the preceding sparse actual bound,

$$\sum_{x\in\mathcal{X}} a_x + 2N_{\mathcal{X}} K^2 \le 75 N_{\mathcal{X}} K^2 + 2N_{\mathcal{X}} K^2 = 77 N_{\mathcal{X}} K^2.$$

The charged dense count is

$$\sum_{x\in\mathcal{X}} s_K + 2N_{\mathcal{X}} K^2,$$

and the same argument gives

$$\sum_{x\in\mathcal{X}} s_K + 2N_{\mathcal{X}} K^2 \le 77 N_{\mathcal{X}} K^2.$$

3. The charged sparse count is

$$\sum_{x\in\mathcal{X}} s_K = \sum_{x\in\mathcal{X}} (64K+32),$$

which depends only on $|\mathcal{X}|$ and K. For any other nonnegative capacity arrays on the same $\mathcal{X}$ and K, the charged sparse count is the same sum. The charged dense count adds the same deterministic materialization term:

$$\sum_{x\in\mathcal{X}} s_K + 2|\mathcal{X}|K^2.$$

Thus the charged sparse counts are equal across such capacity arrays, and the charged dense counts are equal across such capacity arrays.

4. Fix a cell x. Put $\mathcal{Y}$ for the K-point ordered support indexing the row and column capacities in the statement; all sums, extrema, and inequalities among indices below use this finite ordered set. With these arrays, Definition 14 gives

$$q_0(x) = \sum_{i\in\mathcal{Y}} \ell_i(x), \qquad q_1(x) = \sum_{j\in\mathcal{Y}} h_j(x),$$

$$\Delta q(x) = q_1(x) - q_0(x), \qquad m(x) = \min\{q_0(x), q_1(x)\}.$$

Denote the deterministic lower and upper scan outputs by $B_L^{\text{scan}}(x)$ and $B_U^{\text{scan}}(x)$, defined by

$$B_L^{\text{scan}}(x) = \max\left\{0, \max_{t\in\mathcal{Y}} [\ell_{\le t}(x) - h_{\le t}(x) + \min\{\Delta q(x), 0\}]\right\}$$

and

$$B_U^{\text{scan}}(x) = \min\left\{m(x), \min_{t\in\mathcal{Y}} [\ell_{<t}(x) + h_{>t}(x)]\right\}.$$

The costed sparse routine stores as its value

$$\big((B_L^{\text{scan}}(x), B_U^{\text{scan}}(x)), (\tau_x^L, \tau_x^U)\big),$$

where (τ_x^L, τ_x^U) are the lower and upper sparse allocation traces. By Definition 16, these scan outputs are the threshold-cut values:

$$(B_L^{\text{scan}}(x), B_U^{\text{scan}}(x)) = (B_L(x), B_U(x)).$$

Moreover

$$\tau_x^L = \text{the sparse lower allocation trace}, \qquad \tau_x^U = \text{the sparse upper allocation trace}.$$

5. Let γ_x^L and γ_x^U be the matrices obtained by materializing the sparse traces τ_x^L and τ_x^U: for $\bullet \in \{L, U\}$,

$$\gamma_x^\bullet(i,j) = \sum_{e\in\tau_x^\bullet:\ \text{row}(e)=i,\ \text{col}(e)=j} \text{mass}(e).$$

For any cellwise coupling γ, the benefit-mass functional used below is the strict-benefit sum

$$b_x(\gamma) = \sum_{i\in\mathcal{Y}} \sum_{\substack{j\in\mathcal{Y}\\ i<j}} \gamma(i,j).$$

For either the lower nonbenefit primary pass or the upper benefit primary pass, the input row and column lists have lengths K and K. If $\mathsf{prim}_x^\bullet$ is the primary allocation list and $\mathsf{row}_x^{\bullet,\mathrm{res}}$, $\mathsf{col}_x^{\bullet,\mathrm{res}}$ are the two residual lists, the recursive pass gives

$$|\mathsf{prim}_x^\bullet| + |\mathsf{row}_x^{\bullet,\mathrm{res}}| + |\mathsf{col}_x^{\bullet,\mathrm{res}}| \le 2K, \qquad 0 < |\mathsf{row}_x^{\bullet,\mathrm{res}}| + |\mathsf{col}_x^{\bullet,\mathrm{res}}|.$$

The complete bipartite residual pass consumes one residual row or one residual column at each emitted allocation and stops when one residual side is exhausted, so

$$|\operatorname{complete}(\mathsf{row}_x^{\bullet,\mathrm{res}}, \mathsf{col}_x^{\bullet,\mathrm{res}})| < |\mathsf{row}_x^{\bullet,\mathrm{res}}| + |\mathsf{col}_x^{\bullet,\mathrm{res}}|.$$

Thus

$$|\tau_x^\bullet| = |\mathsf{prim}_x^\bullet| + |\operatorname{complete}(\mathsf{row}_x^{\bullet,\mathrm{res}}, \mathsf{col}_x^{\bullet,\mathrm{res}})| < 2K,$$

and, since $|\tau_x^\bullet|$ is an integer,

$$|\tau_x^L| \le 2K-1, \qquad |\tau_x^U| \le 2K-1.$$

Materializing a sparse trace can only put positive mass on a listed row-column pair, hence

$$|\{(i,j) : \gamma_x^L(i,j) > 0\}| \le |\tau_x^L| \le 2K-1, \qquad |\{(i,j) : \gamma_x^U(i,j) > 0\}| \le |\tau_x^U| \le 2K-1.$$

Every emitted mass in the two primary passes and in the residual completion pass is the minimum of two nonnegative residual capacities, and each residual update subtracts that minimum. Therefore

$$\gamma_x^L(i,j) \ge 0, \qquad \gamma_x^U(i,j) \ge 0$$

for all i,j. Let $\gamma_{x,\mathrm{prim}}^\bullet$ and $\gamma_{x,\mathrm{comp}}^\bullet$ be the materialized primary and completion parts, so $\gamma_x^\bullet = \gamma_{x,\mathrm{prim}}^\bullet + \gamma_{x,\mathrm{comp}}^\bullet$. For the row and column residual lists left by the primary pass, define the residual capacities

$$\rho_{x,i}^\bullet = \sum_{(i',u)\in\mathsf{row}_x^{\bullet,\mathrm{res}}:\ i'=i} u, \qquad \chi_{x,j}^\bullet = \sum_{(j',v)\in\mathsf{col}_x^{\bullet,\mathrm{res}}:\ j'=j} v.$$

The primary pass conserves each original capacity after its residual is included, while the completion pass uses at most the residual capacities:

$$\sum_j \gamma_{x,\mathrm{prim}}^\bullet(i,j) + \rho_{x,i}^\bullet = \ell_i(x), \qquad \sum_j \gamma_{x,\mathrm{comp}}^\bullet(i,j) \le \rho_{x,i}^\bullet,$$

and

$$\sum_i \gamma_{x,\mathrm{prim}}^\bullet(i,j) + \chi_{x,j}^\bullet = h_j(x), \qquad \sum_i \gamma_{x,\mathrm{comp}}^\bullet(i,j) \le \chi_{x,j}^\bullet.$$

Consequently

$$\sum_j \gamma_x^\bullet(i,j) \le \ell_i(x), \qquad \sum_i \gamma_x^\bullet(i,j) \le h_j(x).$$

The completion transports the smaller of the two residual totals, and the primary pass conserves the original row and column totals. Hence

$$\sum_{i,j} \gamma_x^\bullet(i,j) = \sum_{i,j} \gamma_{x,\mathrm{prim}}^\bullet(i,j) + \min\left\{\sum_i \rho_{x,i}^\bullet, \sum_j \chi_{x,j}^\bullet\right\} = \min\{q_0(x), q_1(x)\} = m(x).$$

By the definition of Γ_x^* in Definition 15, these displays give

$$\gamma_x^L \in \Gamma_x^*, \qquad \gamma_x^U \in \Gamma_x^*.$$

It remains to identify the two benefit masses. For the lower pass, let ν_x^L be the total mass emitted by the primary nonbenefit pass. Every primary lower allocation satisfies $j \le i$, and every residual row-column pair left for completion satisfies $i < j$. Hence the primary part contributes zero benefit mass and the completion part contributes all of its transported mass to benefit edges. Since the completed lower trace transports total mass $m(x)$,

$$b_x(\gamma_x^L) = m(x) - \nu_x^L.$$

For the upper pass, let β_x^U be the total mass emitted by the primary benefit pass. Every primary upper allocation satisfies $i < j$, and every residual row-column pair left for completion satisfies $j \le i$. Thus the residual completion contributes zero benefit mass and

$$b_x(\gamma_x^U) = \beta_x^U.$$

Define the finite lower and upper cut families on $\{\varnothing\} \cup \mathcal{Y}$ by

$$A_x^L(\varnothing) = m(x), \qquad A_x^L(t) = \ell_{>t}(x) + h_{\le t}(x),$$

$$A_x^U(\varnothing) = m(x), \qquad A_x^U(t) = \ell_{<t}(x) + h_{>t}(x),$$

where $\ell_{>t}(x) = \sum_{i>t} \ell_i(x)$. For the lower primary pass, the empty cut bound is

$$\nu_x^L \le q_0(x), \qquad \nu_x^L \le q_1(x), \qquad \nu_x^L \le m(x) = A_x^L(\varnothing).$$

For an ordinary threshold $t \in \mathcal{Y}$, each nonbenefit primary allocation is counted by the row side $i > t$ or by the column side $j \le t$, so

$$\nu_x^L \le \sum_{i>t} \sum_j \gamma_{x,\mathrm{prim}}^L(i,j) + \sum_{j \le t} \sum_i \gamma_{x,\mathrm{prim}}^L(i,j).$$

Using the primary conservation identities and the nonnegative residual capacities,

$$\sum_{i>t} \sum_j \gamma_{x,\mathrm{prim}}^L(i,j) \le \ell_{>t}(x), \qquad \sum_{j \le t} \sum_i \gamma_{x,\mathrm{prim}}^L(i,j) \le h_{\le t}(x),$$

and therefore

$$\nu_x^L \le A_x^L(t) \qquad (t \in \mathcal{Y}).$$

Thus

$$\nu_x^L \le \inf_{t \in \{\varnothing\} \cup \mathcal{Y}} A_x^L(t).$$

The reverse inequality is obtained from the boundary invariant of the same ascending pass. The input row and column lists are ordered increasingly. By induction on the sum of their lengths, the recursive pass maintains

every emitted primary allocation (i, j, u) satisfies $j \le i$,

every residual row (i, u) and residual column (j, v) satisfy $i < j$, and, for every residual row (r,u) and emitted allocation (i,j,v),

$$\neg(r < i \text{ and } j \le r).$$

The empty-list cases are immediate. In an eligible step $j \le i$, the pass emits $(i,j,\min\{a,b\})$ and recurses after reducing or deleting the exhausted endpoint; the reduced lists preserve the increasing order, and any later residual row has index at least the emitted row index, so the new emitted allocation cannot cross a later residual-row boundary. In an ineligible step $i < j$, the row i is moved to the residual list; every column that can remain residual has index at least j, and every later emitted allocation uses such a column, giving both the residual separation and the displayed noncrossing condition for this new residual row.

If the lower primary pass leaves no row residual, row-total conservation gives $\nu_x^L = q_0(x)$, and the column bound above gives $\nu_x^L \le q_1(x)$. Thus, using Definition 14,

$$\nu_x^L = \min\{q_0(x), q_1(x)\} = m(x) = A_x^L(\varnothing),$$

so the infimum is at most ν_x^L. If a row residual remains, let t_L be the largest row index among the lower residual rows. The finite maximum transfers the preceding invariants to the cut:

$$i' \le t_L \quad \text{for every } (i',u) \in \mathsf{row}_x^{L,\mathrm{res}}, \qquad t_L < j' \quad \text{for every } (j',v) \in \mathsf{col}_x^{L,\mathrm{res}},$$

and no primary lower allocation with positive mass has both $t_L < i$ and $j \le t_L$. Hence

$$\sum_{i>t_L} \rho_{x,i}^L = 0, \qquad \sum_{j \le t_L} \chi_{x,j}^L = 0.$$

Since every primary lower allocation satisfies $j \le i$ and no such allocation crosses the cut t_L, its mass is counted exactly once by the row side $i > t_L$ or the column side $j \le t_L$:

$$\nu_x^L = \sum_{i>t_L} \sum_j \gamma_{x,\mathrm{prim}}^L(i,j) + \sum_{j \le t_L} \sum_i \gamma_{x,\mathrm{prim}}^L(i,j).$$

Adding the zero residual sums and applying row and column conservation gives

$$\nu_x^L = \ell_{>t_L}(x) + h_{\le t_L}(x) = A_x^L(t_L),$$

so again $\inf_{t \in \{\varnothing\} \cup \mathcal{Y}} A_x^L(t) \le \nu_x^L$. Consequently

$$\nu_x^L = \inf_{t \in \{\varnothing\} \cup \mathcal{Y}} A_x^L(t).$$

For the upper primary pass the cut bounds use the strict-benefit cuts. The empty cut gives

$$\beta_x^U \le q_0(x), \qquad \beta_x^U \le q_1(x), \qquad \beta_x^U \le m(x) = A_x^U(\varnothing),$$

and, for every $t \in \mathcal{Y}$, each benefit primary allocation is counted by the row side $i < t$ or by the column side $j > t$:

$$\beta_x^U \le \sum_{i<t} \sum_j \gamma_{x,\mathrm{prim}}^U(i,j) + \sum_{j>t} \sum_i \gamma_{x,\mathrm{prim}}^U(i,j) \le \ell_{<t}(x) + h_{>t}(x) = A_x^U(t).$$

Thus $\beta_x^U \le \inf_{t\in\{\varnothing\}\cup\mathcal{Y}} A_x^U(t)$. For the reverse inequality, use the symmetric boundary invariant for the descending benefit pass. Its input row and column lists are ordered decreasingly, and induction on the same recursion gives

$$\text{every emitted primary allocation } (i,j,u) \text{ satisfies } i<j,$$

$$\text{every residual row } (i,u) \text{ and residual column } (j,v) \text{ satisfy } j\le i,$$

and, for every residual row (r,u) and emitted allocation (i,j,v),

$$\neg(i<r \text{ and } r<j).$$

The eligible branch $i<j$ emits $(i,j,\min\{a,b\})$ and recurses on decreasing lists after the exhausted endpoint is reduced or deleted; any later residual row has index at most the emitted row index, so the new emitted allocation cannot cross its residual-row boundary. In the ineligible branch $j\le i$, the row i is moved to the residual list; all remaining columns have index at most j, and all later emitted allocations use such columns, giving the residual separation and noncrossing condition for the new residual row.

If the upper primary pass leaves no row residual, then $\beta_x^U = q_0(x)$ and the column bound above gives $\beta_x^U \le q_1(x)$, so, by Definition 14,

$$\beta_x^U = \min\{q_0(x), q_1(x)\} = m(x) = A_x^U(\varnothing).$$

If a row residual remains, let t_U be the smallest row index among the upper residual rows. The finite minimum transfers the preceding invariants to the cut:

$$t_U \le i' \quad \text{for every } (i',u)\in \mathsf{row}_x^{U,\mathrm{res}}, \qquad j'\le t_U \quad \text{for every } (j',v)\in \mathsf{col}_x^{U,\mathrm{res}},$$

and no primary upper allocation with positive mass has both $i<t_U$ and $t_U<j$. Hence

$$\sum_{i<t_U} \rho_{x,i}^U = 0, \qquad \sum_{j>t_U} \chi_{x,j}^U = 0.$$

Since every primary upper allocation satisfies $i<j$ and no such allocation crosses the cut t_U, its mass is counted exactly once by the row side $i<t_U$ or the column side $j>t_U$:

$$\beta_x^U = \sum_{i<t_U}\sum_j \gamma_{x,\mathrm{prim}}^U(i,j) + \sum_{j>t_U}\sum_i \gamma_{x,\mathrm{prim}}^U(i,j).$$

Conservation then yields

$$\beta_x^U = \ell_{<t_U}(x) + h_{>t_U}(x) = A_x^U(t_U).$$

Therefore

$$\beta_x^U = \inf_{t\in\{\varnothing\}\cup\mathcal{Y}} A_x^U(t).$$

For the lower endpoint, the partition identity $\ell_{\le t}(x) + \ell_{>t}(x) = q_0(x)$, together with $m(x) = q_0(x) + \min\{\Delta q(x), 0\}$ from Definition 14, implies

$$m(x) - A_x^L(\varnothing) = 0, \qquad m(x) - A_x^L(t) = \ell_{\le t}(x) - h_{\le t}(x) + \min\{\Delta q(x), 0\}.$$

Therefore, by Definition 16,

$$b_x(\gamma_x^L) = m(x) - \inf_t A_x^L(t) = \sup_t \{m(x) - A_x^L(t)\} = B_L(x).$$

For the upper endpoint, the displayed formula for A_x^U is exactly the upper threshold-cut family in Definition 16, and hence

$$b_x(\gamma_x^U) = \inf_t A_x^U(t) = B_U(x).$$

Since x was arbitrary, the asserted cellwise output properties hold in every covariate cell.

□

Proof of Theorem 8. 1. Let $\mathcal{O}$ be the finite set of observed cells and write

$$\vartheta_o = P_{\text{obs}}(\{o\}), \qquad \widehat{\vartheta}_{n,o} = \frac{1}{n}\sum_{r=1}^{n} \mathbf{1}\{O_r = o\}, \qquad o \in \mathcal{O}.$$

For an event $E \subseteq \mathcal{O}$, set

$$u(E) = \sum_{o \in E} u_o.$$

The finite multinomial central limit input applied under Assumption 10 gives a Gaussian law $\mathbb{Z}_{P_{\text{obs}}}$ on $\mathbb{R}^{\mathcal{O}}$ such that

$$\sqrt{n}(\widehat{\vartheta}_n - \vartheta) \rightsquigarrow \mathbb{Z}_{P_{\text{obs}}},$$
$$\int z \, d\mathbb{Z}_{P_{\text{obs}}}(z) = 0,$$

and, for observed cells a, b,

$$\int z(a)z(b)\, d\mathbb{Z}_{P_{\text{obs}}}(z) = \begin{cases} P_{\text{obs}}(a)\{1 - P_{\text{obs}}(a)\}, & a = b, \\ -P_{\text{obs}}(a)P_{\text{obs}}(b), & a \neq b. \end{cases}$$

Here the displayed empirical vector is the atom vector of the sample empirical law, because its o-coordinate is exactly the centered empirical frequency of $\{o\}$.

2. Put

$$\mathcal{X}_+(P) = \{x : p_x m(x) > 0\}.$$

For an arbitrary vector $u \in \mathbb{R}^{\mathcal{O}}$, define

$$\mathsf{C}_u(A \mid B) = \begin{cases} u(A \cap B)/u(B), & u(B) > 0, \\ 0, & u(B) \leq 0, \end{cases}$$

and form the raw contrasts

$$\ell_i(u,x) = \mathsf{C}_u(X = x, D = 0, S = 1, Y = i \mid X = x, Z = 0) - \mathsf{C}_u(X = x, D = 0, S = 1, Y = i \mid X = x, Z = 1),$$

$$h_j(u,x) = \mathsf{C}_u(X = x, D = 1, S = 1, Y = j \mid X = x, Z = 1) - \mathsf{C}_u(X = x, D = 1, S = 1, Y = j \mid X = x, Z = 0)$$

Let $\ell_i^+(u,x) = \max\{\ell_i(u,x), 0\}$ and $h_j^+(u,x) = \max\{h_j(u,x), 0\}$. From these projected capacities define $q_0(u,x), q_1(u,x), m(u,x), B_L(u,x), B_U(u,x)$ by the formulas in Definitions 14 and 16, and define the fixed-support endpoint map

$$\Phi(u) = \left(\frac{\sum_{x\in\mathcal{X}_+(P)} u(X=x)B_L(u,x)}{\sum_{x\in\mathcal{X}_+(P)} u(X=x)m(u,x)}, \frac{\sum_{x\in\mathcal{X}_+(P)} u(X=x)B_U(u,x)}{\sum_{x\in\mathcal{X}_+(P)} u(X=x)m(u,x)} \right),$$

with the quotient coordinates totalized by the value 0 when the common denominator is zero.

3. Theorem 1 gives compatibility of the observable capacities, $p_x \geq 0$, and the structural identification of $m(x)$. Since Assumption 8 gives

$$M = \sum_x p_x m(x) > 0,$$

the recovered support is nonempty after restriction to positive products. If $x \in \mathcal{X}_+(P)$, then $p_x m(x) > 0$, hence $p_x > 0$. Writing $\pi(x) = P_{\text{obs}}(Z = 1 \mid X = x)$, Assumption 5 gives $0 < \varepsilon_Z$, $\varepsilon_Z \leq \pi(x)$, and $\pi(x) \leq 1 - \varepsilon_Z$. Therefore

$$P_{\text{obs}}(X = x, Z = 1) = p_x\pi(x) \geq p_x\varepsilon_Z > 0$$

and

$$P_{\text{obs}}(X = x, Z = 0) = p_x\{1 - \pi(x)\} \geq p_x\varepsilon_Z > 0.$$

Thus all conditional denominators used by Φ on $\mathcal{X}_+(P)$ are positive at ϑ. Moreover,

$$\ell_i(\vartheta, x) = \ell_i(x), \qquad h_j(\vartheta, x) = h_j(x),$$

because $u(E)$ at $u = \vartheta$ is $P_{\text{obs}}(E)$. The nonnegativity supplied by Theorem 1 then yields

$$\ell_i^+(\vartheta, x) = \ell_i(x), \qquad h_j^+(\vartheta, x) = h_j(x),$$

and

$$\vartheta(X = x) = p_x, \qquad m(\vartheta, x) = m(x).$$

Capacity validity gives $\ell_i(x) \geq 0$ and $h_j(x) \geq 0$, so

$$q_0(x) = \sum_i \ell_i(x) \geq 0, \qquad q_1(x) = \sum_j h_j(x) \geq 0, \qquad m(x) = \min\{q_0(x), q_1(x)\} \geq 0.$$

Consequently

$$\sum_{x\in\mathcal{X}_+(P)} \vartheta(X = x)m(\vartheta, x) = \sum_{x\in\mathcal{X}_+(P)} p_x m(x) = M > 0.$$

For the last equality, if $x \notin \mathcal{X}_+(P)$, then $\neg(0 < p_x m(x))$; since Theorem 1 gives $p_x \geq 0$ and the preceding display gives $m(x) \geq 0$, we have $0 \leq p_x m(x)$, hence $p_x m(x) = 0$.

4. The map $u \mapsto \Phi(u)$ is Hadamard directionally differentiable at ϑ along all directions in $\mathbb{R}^{\mathcal{O}}$. Indeed, on the neighborhood where the arm denominators in the preceding step are positive, the raw conditional contrasts are smooth quotient maps. The coordinatewise positive-part map has scalar directional derivative

$$\dot{a} \mapsto \begin{cases} \dot{a}, & a > 0, \\ \max\{\dot{a}, 0\}, & a = 0, \end{cases}$$

at every nonnegative coordinate a. Finite sums preserve directional differentiability, the scalar minimum gives the derivative of $m(u,x) = \min\{q_0(u,x), q_1(u,x)\}$, and the finite maximum and minimum defining $B_L(u,x)$ and $B_U(u,x)$ give the active-face directional derivatives of the threshold cuts. Therefore the denominator and the two numerators in Φ are continuously Hadamard directionally differentiable at ϑ. Since the denominator is positive at ϑ, the quotient rule gives a continuous directional derivative $D\Psi$ satisfying

$$\frac{\Phi(\vartheta + t_n g_n) - \Phi(\vartheta)}{t_n} \longrightarrow D\Psi(g)$$

whenever $t_n \downarrow 0$ and $g_n \to g$ in $\mathbb{R}^{\mathcal{O}}$. Applying the directional delta-method input to Φ gives

$$\sqrt{n}\{\Phi(\widehat{\vartheta}_n) - \Phi(\vartheta)\} \rightsquigarrow D\Psi(\mathbb{Z}_{P_{\mathrm{obs}}}).$$

5. The population value of the fixed-support map equals the identified endpoint map. For every array b_x satisfying

$$0 \le b_x \le m(x) \qquad \text{for all } x,$$

one has

$$\sum_{x \in \mathcal{X}_+(P)} p_x b_x = \sum_x p_x b_x,$$

because $x \notin \mathcal{X}_+(P)$ implies $p_x m(x) = 0$ and hence $p_x b_x = 0$. The capacity validity supplied by Theorem 1 gives

$$\ell_i(x) \ge 0, \qquad h_j(x) \ge 0$$

for every i, j, x. Hence

$$q_0(x) = \sum_i \ell_i(x) \ge 0, \qquad q_1(x) = \sum_j h_j(x) \ge 0,$$

so

$$m(x) = \min\{q_0(x), q_1(x)\} \ge 0.$$

The lower formula in Definition 16 includes the candidate 0, and therefore

$$0 \le B_L(x).$$

To prove the upper bound for $B_L(x)$, it is enough to bound every lower-threshold candidate by $m(x)$. The zero candidate is bounded by $m(x)$ by the preceding display. For a threshold t, put

$$\ell_{\le t}(x) = \sum_{i \le t} \ell_i(x), \qquad h_{\le t}(x) = \sum_{j \le t} h_j(x),$$

so that $0 \leq h_{\leq t}(x)$ and $\ell_{\leq t}(x) \leq q_0(x)$. If $q_0(x) \leq q_1(x)$, then $m(x) = q_0(x)$ and

$$\ell_{\leq t}(x) - h_{\leq t}(x) + \min\{\Delta q(x), 0\} \leq \ell_{\leq t}(x) \leq q_0(x) = m(x),$$

because $-h_{\leq t}(x) \leq 0$ and $\min\{\Delta q(x), 0\} \leq 0$. If $q_1(x) \leq q_0(x)$, then $m(x) = q_1(x)$, $\Delta q(x) = q_1(x) - q_0(x) \leq 0$, and

$$\ell_{\leq t}(x) - h_{\leq t}(x) + \min\{\Delta q(x), 0\} = \ell_{\leq t}(x) - h_{\leq t}(x) + q_1(x) - q_0(x) \leq q_1(x) = m(x).$$

Thus $B_L(x) \leq m(x)$. For the upper threshold cut, Definition 16 writes $B_U(x)$ as the minimum of $m(x)$ and the threshold sums $\ell_{<t}(x) + h_{>t}(x)$. The candidate $m(x)$ gives $B_U(x) \leq m(x)$, while all candidates are nonnegative by capacity validity, so $0 \leq B_U(x)$. Consequently

$$0 \leq B_L(x) \leq m(x), \qquad 0 \leq B_U(x) \leq m(x).$$

Using the restriction identity with $b_x = m(x), B_L(x), B_U(x)$ gives

$$\Phi(\vartheta) = \left(\frac{\sum_x p_x B_L(x)}{M}, \frac{\sum_x p_x B_U(x)}{M} \right) = \Psi(P_{\text{obs}}),$$

where the final equality is the endpoint formula in Definition 17.

6. For each cell x, define the empirical screening score

$$\widehat{r}_{n,x} = \widehat{p}_{n,x} \widehat{m}_n(x) = \widehat{\vartheta}_n(X = x) m(\widehat{\vartheta}_n, x).$$

The screening rule in Algorithm 2 is exactly

$$x \in \{y : \widehat{r}_{n,y} > \eta_n\} \quad \Longleftrightarrow \quad \eta_n < \widehat{r}_{n,x}.$$

If $p_x > 0$, the same positive-denominator argument as above and the directional delta method applied to the scalar score map give tightness of the scaled score error: for every $\epsilon_{\text{cell}} > 0$, there is $R_x \in \mathbb{R}$ such that, for all n,

$$P\big(\max\{R_x, 0\} < \sqrt{n}\, |\widehat{r}_{n,x} - p_x m(x)|\big) \leq \epsilon_{\text{cell}}.$$

If $p_x = 0$, then $P_{\text{obs}}(X = x) = 0$, and independent sampling gives

$$P(\widehat{p}_{n,x} = 0) = 1, \qquad P(\widehat{r}_{n,x} = 0) = 1.$$

7. Fix a cell x and $\epsilon_{\text{cell}} > 0$. If $p_x = 0$, then the preceding display gives exact agreement of the screened and population membership decisions. Suppose $p_x > 0$. If $x \in \mathcal{X}_+(P)$, write

$$r_x = p_x m(x) > 0.$$

For all sufficiently large n,

$$\eta_n < r_x/2, \qquad \frac{2\max\{R_x, 0\}}{r_x} < \sqrt{n}.$$

On the event that x is not screened, $\widehat{r}_{n,x} \leq \eta_n$, and hence

$$\sqrt{n}\, |\widehat{r}_{n,x} - r_x| = \sqrt{n}\, (r_x - \widehat{r}_{n,x}) \geq \sqrt{n}\, (r_x - \eta_n) > \max\{R_x, 0\}.$$

If $x \notin \mathcal{X}_+(P)$, then $p_x m(x) = 0$. For all sufficiently large n,

$$\max\{R_x, 0\} < \sqrt{n}\,\eta_n.$$

On the event that x is screened, $\widehat{r}_{n,x} > \eta_n$, and therefore

$$\sqrt{n}\,|\widehat{r}_{n,x} - 0| = \sqrt{n}\,\widehat{r}_{n,x} > \max\{R_x, 0\}.$$

Thus, for every x,

$$P(\{x \in \{y : \widehat{r}_{n,y} > \eta_n\}\} \neq \{x \in \mathcal{X}_+(P)\}) \leq \epsilon_{\text{cell}}$$

eventually. A finite union over $x \in \mathcal{X}$ yields

$$P(\{x : \widehat{r}_{n,x} > \eta_n\} = \mathcal{X}_+(P)) \to 1.$$

8. On the event

$$\{x : \widehat{r}_{n,x} > \eta_n\} = \mathcal{X}_+(P),$$

the screened sums in $\widehat{\Psi}_n$ are exactly the fixed-support sums defining $\Phi(\widehat{\vartheta}_n)$. The support is nonempty and each retained score exceeds $\eta_n > 0$, so the screened denominator is positive and the fallback branch in Algorithm 2 is inactive on this event. Hence

$$P\Big(\widehat{\Psi}_n = \Phi(\widehat{\vartheta}_n)\Big) \to 1.$$

Together with $\Phi(\vartheta) = \Psi(P_{\text{obs}})$, equality with probability tending to one transfers the fixed-support weak limit to the screened estimator:

$$\sqrt{n}\{\widehat{\Psi}_n - \Psi(P_{\text{obs}})\} \rightsquigarrow D\Psi(\mathbb{Z}_{P_{\text{obs}}}).$$

9. It remains to record consistency. The weak convergence just proved implies tightness of the scaled endpoint error: for every $\delta > 0$ there is $R_\Psi \in \mathbb{R}$ such that, for all n,

$$P\Big(R_\Psi < \sqrt{n}\,\|\widehat{\Psi}_n - \Psi(P_{\text{obs}})\|\Big) \leq \delta/2.$$

Given $\epsilon > 0$, choose n large enough that

$$\frac{\max\{R_\Psi, 0\}}{\epsilon} + 1 \leq \sqrt{n}.$$

Then

$$\{\|\widehat{\Psi}_n - \Psi(P_{\text{obs}})\| > \epsilon\} \subseteq \Big\{R_\Psi < \sqrt{n}\,\|\widehat{\Psi}_n - \Psi(P_{\text{obs}})\|\Big\},$$

and therefore

$$P\{\|\widehat{\Psi}_n - \Psi(P_{\text{obs}})\| > \epsilon\} \leq \delta/2$$

eventually. Since δ is arbitrary,

$$\widehat{\Psi}_n \to_P \Psi(P_{\text{obs}}).$$

The displayed Gaussian law, covariance identities, support recovery, consistency, and directional weak limit are precisely the asserted conclusions.

□

Proof of Lemma 1. Fix a covariate cell x. By Definition 14, the capacity totals formed from the weighted array c^w are

$$q_0^w(x) = \sum_{i \in \mathcal{Y}} \ell_i^w(x), \qquad q_1^w(x) = \sum_{j \in \mathcal{Y}} h_j^w(x), \qquad m^w(x) = \min\{q_0^w(x), q_1^w(x)\}.$$

1. Substituting the definition of c^w into the first weighted total gives

   $$q_0^w(x) = \sum_{i \in \mathcal{Y}} w_x \ell_i(x) = w_x \sum_{i \in \mathcal{Y}} \ell_i(x) = w_x q_0(x),$$

   where the middle equality is finite-sum linearity.

2. The same computation for the upper-arm total gives

   $$q_1^w(x) = \sum_{j \in \mathcal{Y}} w_x h_j(x) = w_x \sum_{j \in \mathcal{Y}} h_j(x) = w_x q_1(x).$$

   This finite-sum identity also covers the degenerate case in which the outcome index set is empty, since both sums are then zero.

3. Since the weight satisfies $w_x \geq 0$, multiplication by w_x commutes with the binary minimum:

   $$\min\{w_x q_0(x), w_x q_1(x)\} = w_x \min\{q_0(x), q_1(x)\}.$$

   Combining this identity with the two displayed total identities yields

   $$m^w(x) = \min\{q_0^w(x), q_1^w(x)\} = \min\{w_x q_0(x), w_x q_1(x)\} = w_x m(x),$$

   again using the definition of $m(x)$ in Definition 14.

□

Proof of Lemma 2. Fix a cell x. Write $\omega = w_x$, so $\omega \geq 0$. By Definition 16, the lower cut can be written as the finite maximum over the dummy alternative and the threshold alternatives:

$$B_L(x) = \max\Big(\{0\} \cup \big\{\ell_{\leq t}(x) - h_{\leq t}(x) + \min\{\Delta q(x), 0\} : t \in \mathcal{Y}\big\}\Big).$$

The same formula applied to c^w gives $B_L^w(x)$. Since multiplication by $\omega \geq 0$ commutes with a finite maximum,

$$\omega B_L(x) = \max\Big(\{0\} \cup \big\{\omega\big(\ell_{\leq t}(x) - h_{\leq t}(x) + \min\{\Delta q(x), 0\}\big) : t \in \mathcal{Y}\big\}\Big).$$

Thus it remains to compare the dummy term and each threshold term.

For the dummy alternative the equality is immediate, because the term is 0 on both sides:

$$0 = \omega \cdot 0.$$

Now fix $t \in \mathcal{Y}$. The weighted prefix sums satisfy

$$\ell_{\leq t}^w(x) = \sum_{i \leq t} \omega \ell_i(x) = \omega \sum_{i \leq t} \ell_i(x) = \omega \ell_{\leq t}(x),$$

and similarly

$$h^w_{\le t}(x) = \omega h_{\le t}(x).$$

The weighted capacity-total gap is

$$\Delta q^w(x) = \sum_{j \in \mathcal{Y}} \omega h_j(x) - \sum_{i \in \mathcal{Y}} \omega \ell_i(x) = \omega\big(q_1(x) - q_0(x)\big) = \omega \Delta q(x).$$

Because $\omega \ge 0$,

$$\min\{\Delta q^w(x), 0\} = \min\{\omega \Delta q(x), 0\} = \min\{\omega \Delta q(x), \omega \cdot 0\} = \omega \min\{\Delta q(x), 0\}.$$

Combining these identities,

$$\ell^w_{\le t}(x) - h^w_{\le t}(x) + \min\{\Delta q^w(x), 0\} = \omega\big(\ell_{\le t}(x) - h_{\le t}(x) + \min\{\Delta q(x), 0\}\big).$$

The dummy term and all threshold terms therefore agree with the corresponding ω-scaled unweighted terms inside the same finite maximum. Hence

$$B^w_L(x) = \omega B_L(x) = w_x B_L(x).$$

□

Proof of Lemma 3. Fix $x \in \mathcal{X}$. Let $\mathcal{I} = \{\bullet\} \cup \mathcal{Y}$, where $\bullet$ is an auxiliary marker. For $\iota \in \mathcal{I}$, define the unweighted and weighted upper-cut candidates by

$$\tau_\bullet(x) = m(x), \qquad \tau_t(x) = \ell_{<t}(x) + h_{>t}(x) \quad (t \in \mathcal{Y}),$$

and

$$\tau^w_\bullet(x) = m^w(x), \qquad \tau^w_t(x) = \ell^w_{<t}(x) + h^w_{>t}(x) \quad (t \in \mathcal{Y}),$$

where

$$\ell^w_{<t}(x) = \sum_{i<t} \ell^w_i(x), \qquad h^w_{>t}(x) = \sum_{j>t} h^w_j(x).$$

Thus $\mathcal{I}$ is nonempty even when there are no threshold indices.

1. By Definition 16, the upper cut is the minimum over these candidates:

   $$B_U(x) = \min_{\iota \in \mathcal{I}} \tau_\iota(x), \qquad B^w_U(x) = \min_{\iota \in \mathcal{I}} \tau^w_\iota(x).$$

   For any finite nonempty index set, multiplication by a nonnegative scalar commutes with the minimum:

   $$\min_{\iota \in \mathcal{I}} w_x \tau_\iota(x) = w_x \min_{\iota \in \mathcal{I}} \tau_\iota(x).$$

   Indeed, if $w_x = 0$ both sides are zero, while if $w_x > 0$ the order-preserving map $z \mapsto w_x z$ carries the least element of the finite family to the least element of the scaled family. Hence

   $$w_x B_U(x) = \min_{\iota \in \mathcal{I}} w_x \tau_\iota(x).$$

2. It remains to identify the weighted candidates with the scaled unweighted candidates. For the marker candidate, Lemma 1 gives

$$\tau_\bullet^w(x) = m^w(x) = w_x m(x) = w_x \tau_\bullet(x).$$

For any threshold $t \in \mathcal{Y}$,

$$\ell_{<t}^w(x) = \sum_{i<t} w_x \ell_i(x) = w_x \sum_{i<t} \ell_i(x) = w_x \ell_{<t}(x),$$

and similarly

$$h_{>t}^w(x) = \sum_{j>t} w_x h_j(x) = w_x \sum_{j>t} h_j(x) = w_x h_{>t}(x).$$

Therefore

$$\tau_t^w(x) = \ell_{<t}^w(x) + h_{>t}^w(x) = w_x\big(\ell_{<t}(x) + h_{>t}(x)\big) = w_x \tau_t(x).$$

Thus $\tau_\iota^w(x) = w_x \tau_\iota(x)$ for every $\iota \in \mathcal{I}$.

3. Combining the previous displays,

$$B_U^w(x) = \min_{\iota \in \mathcal{I}} \tau_\iota^w(x) = \min_{\iota \in \mathcal{I}} w_x \tau_\iota(x) = w_x B_U(x).$$

This is the desired identity for the fixed cell x, and x was arbitrary.

□

Proof of Theorem 9. 1. Fix $n \geq 1$, $\lambda \in \Lambda$, and ω. Let $\mathcal{Y} = \{0, \ldots, K-1\}$. The finite guard-event collection used here is

$$\mathcal{E} = \{\mathsf{c}_x : x \in \mathcal{X}\} \cup \{\mathsf{a}_{xz} : x \in \mathcal{X},\ z \in \{0,1\}\} \cup \{\mathsf{s}_{xzvi} : x \in \mathcal{X},\ z, v \in \{0,1\},\ i \in \mathcal{Y}\},$$

where

$$\mathsf{c}_x = \{X = x\}, \qquad \mathsf{a}_{xz} = \{X = x, Z = z\}, \qquad \mathsf{s}_{xzvi} = \{X = x, Z = z, D = v, S = 1, \widetilde{Y} = i\}.$$

Define

$$\delta_{n,\lambda}(\omega) = \max_{\mathsf{e} \in \mathcal{E}} \left| \widehat{P}_{n,\lambda}(\mathsf{e}; \omega) - P_{\mathrm{obs},\lambda}(\mathsf{e}) \right|.$$

The estimator in Algorithm 2 uses the totalized conditional ratio

$$\mathrm{cond}(u, v) = \begin{cases} u/v, & v > 0, \\ 0, & v \leq 0. \end{cases}$$

Set

$$\rho_{\mathrm{cap}} = \frac{64K|\mathcal{X}|}{\varepsilon_Z^2}, \qquad \mathsf{A}_n(\delta) = \rho_{\mathrm{cap}}\delta + |\mathcal{X}|\eta_n.$$

By Definition 23, the law at λ satisfies the structural restrictions, the i.i.d. sampling condition, and Assumption 9. Thus Theorem 1 gives valid observable capacities and $p_x \geq 0$. The overlap condition gives the arm-mass inequality

$$\varepsilon_Z p_x \leq P_{\mathrm{obs},\lambda}(X = x, Z = z), \qquad x \in \mathcal{X}, \quad z \in \{0,1\}.$$

Indeed, when $p_x = 0$ the inequality is immediate from nonnegativity. When $p_x > 0$, write $\pi(x) = P_{\mathrm{obs},\lambda}(Z = 1 \mid X = x)$. For $z = 1$, Assumption 5 gives

$$P_{\mathrm{obs},\lambda}(X = x, Z = 1) = p_x \pi(x) \geq \varepsilon_Z p_x.$$

For $z = 0$, the binary partition and the upper-overlap bound in Assumption 5 give

$$P_{\mathrm{obs},\lambda}(X = x, Z = 0) = p_x\{1 - \pi(x)\} \geq \varepsilon_Z p_x.$$

The uniform aggregate bound gives $m_\star \leq M$.

Suppose $0 \leq \delta$ and $\delta_{n,\lambda}(\omega) \leq \delta$. Fix $x \in \mathcal{X}$, $z, \upsilon \in \{0, 1\}$, and $i \in \mathcal{Y}$. Write

$$\mathfrak{a}_{xz} = P_{\mathrm{obs},\lambda}(X = x, Z = z), \qquad \widehat{\mathfrak{a}}_{xz} = \widehat{P}_{n,\lambda}(X = x, Z = z; \omega),$$

and

$$\mathfrak{u}_{xz\upsilon i} = P_{\mathrm{obs},\lambda}(X = x, Z = z, D = \upsilon, S = 1, \widetilde{Y} = i), \qquad \widehat{\mathfrak{u}}_{xz\upsilon i} = \widehat{P}_{n,\lambda}(X = x, Z = z, D = \upsilon, S = 1, \widetilde{Y} = i; \omega).$$

The definition of $\mathcal{E}$ and the bound $\delta_{n,\lambda}(\omega) \leq \delta$ give

$$|\widehat{p}_x - p_x| \leq \delta, \qquad |\widehat{\mathfrak{a}}_{xz} - \mathfrak{a}_{xz}| \leq \delta, \qquad |\widehat{\mathfrak{u}}_{xz\upsilon i} - \mathfrak{u}_{xz\upsilon i}| \leq \delta.$$

Moreover

$$0 \leq \mathfrak{u}_{xz\upsilon i} \leq \mathfrak{a}_{xz}, \qquad 0 \leq \widehat{\mathfrak{u}}_{xz\upsilon i} \leq \widehat{\mathfrak{a}}_{xz},$$

because each selected-outcome event is contained in its arm event. Therefore

$$|\widehat{p}_x \ \mathrm{cond}(\widehat{\mathfrak{u}}_{xz\upsilon i}, \widehat{\mathfrak{a}}_{xz}) - p_x \ \mathrm{cond}(\mathfrak{u}_{xz\upsilon i}, \mathfrak{a}_{xz})| \leq \frac{8\delta}{\varepsilon_Z^2}.$$

When $\delta = 0$, the three displayed deviation inequalities force $\widehat{p}_x = p_x$, $\widehat{\mathfrak{a}}_{xz} = \mathfrak{a}_{xz}$, and $\widehat{\mathfrak{u}}_{xz\upsilon i} = \mathfrak{u}_{xz\upsilon i}$, so the two totalized conditional expressions are identical, including the case of zero arm mass. Assume now that $\delta > 0$. If $p_x < 4\delta/\varepsilon_Z$, both totalized conditionals lie in $[0, 1]$, $\widehat{p}_x \leq p_x + \delta$, and hence

$$|\widehat{p}_x \, \mathrm{cond}(\widehat{\mathfrak{u}}_{xz\upsilon i}, \widehat{\mathfrak{a}}_{xz}) - p_x \, \mathrm{cond}(\mathfrak{u}_{xz\upsilon i}, \mathfrak{a}_{xz})| \leq \widehat{p}_x + p_x \leq \frac{8\delta}{\varepsilon_Z^2},$$

where the last inequality uses $0 < \varepsilon_Z < 1/2$. If $p_x \geq 4\delta/\varepsilon_Z$, then the arm-overlap inequality gives

$$4\delta \leq \varepsilon_Z p_x \leq \mathfrak{a}_{xz}, \qquad \widehat{\mathfrak{a}}_{xz} \geq \mathfrak{a}_{xz} - \delta \geq \frac{3}{4}\mathfrak{a}_{xz} > 0.$$

Thus the ratio branch is active for both denominators, and

$$\left| \frac{\widehat{\mathfrak{u}}_{xz\upsilon i}}{\widehat{\mathfrak{a}}_{xz}} - \frac{\mathfrak{u}_{xz\upsilon i}}{\mathfrak{a}_{xz}} \right| = \frac{|\widehat{\mathfrak{u}}_{xz\upsilon i}\mathfrak{a}_{xz} - \mathfrak{u}_{xz\upsilon i}\widehat{\mathfrak{a}}_{xz}|}{\widehat{\mathfrak{a}}_{xz}\mathfrak{a}_{xz}} \leq \frac{8\delta}{3\mathfrak{a}_{xz}}.$$

The numerator bound follows from

$$|\widehat{\mathfrak{u}}_{xz\upsilon i}\mathfrak{a}_{xz} - \mathfrak{u}_{xz\upsilon i}\widehat{\mathfrak{a}}_{xz}| \leq |\widehat{\mathfrak{u}}_{xz\upsilon i} - \mathfrak{u}_{xz\upsilon i}|\mathfrak{a}_{xz} + \mathfrak{u}_{xz\upsilon i}|\mathfrak{a}_{xz} - \widehat{\mathfrak{a}}_{xz}| \leq 2\delta\mathfrak{a}_{xz},$$

and the denominator lower bound is $\widehat{\mathfrak{a}}_{xz}\mathfrak{a}_{xz} \geq 3\mathfrak{a}_{xz}^2/4$. Since $p_x/\mathfrak{a}_{xz} \leq 1/\varepsilon_Z$,

$$\begin{aligned}
&\left|\widehat{p}_x \frac{\widehat{\mathfrak{u}}_{xzvi}}{\widehat{\mathfrak{a}}_{xz}} - p_x \frac{\mathfrak{u}_{xzvi}}{\mathfrak{a}_{xz}}\right| \\
&\qquad \leq |\widehat{p}_x - p_x| \frac{\widehat{\mathfrak{u}}_{xzvi}}{\widehat{\mathfrak{a}}_{xz}} + p_x \left|\frac{\widehat{\mathfrak{u}}_{xzvi}}{\widehat{\mathfrak{a}}_{xz}} - \frac{\mathfrak{u}_{xzvi}}{\mathfrak{a}_{xz}}\right| \\
&\qquad \leq \delta + \frac{8\delta}{3\varepsilon_Z} \leq \frac{8\delta}{\varepsilon_Z^2},
\end{aligned}$$

again using $0 < \varepsilon_Z < 1/2$.

Each raw lower or upper capacity coordinate is a difference of two such conditionals, and the projection in Algorithm 2 is coordinatewise positive part, a one-Lipschitz map against the nonnegative population capacities. Hence

$$\max\left\{\left|\widehat{p}_x\widehat{\ell}_i(x) - p_x\ell_i(x)\right|, \left|\widehat{p}_x\widehat{h}_i(x) - p_x h_i(x)\right|\right\} \leq \frac{16\delta}{\varepsilon_Z^2}.$$

Put

$$e_{\text{cap}} = \frac{16\delta}{\varepsilon_Z^2}.$$

For each cell define the unconditional population and empirical projected capacity coordinates by

$$\mathfrak{l}_i(x) = p_x\ell_i(x), \qquad \mathfrak{h}_i(x) = p_x h_i(x), \qquad \widehat{\mathfrak{l}}_{i,n}(x) = \widehat{p}_x\widehat{\ell}_i(x), \qquad \widehat{\mathfrak{h}}_{i,n}(x) = \widehat{p}_x\widehat{h}_i(x).$$

Let $\mathfrak{m}(x)$, $\mathfrak{B}_L(x)$, and $\mathfrak{B}_U(x)$ denote, respectively, the survivor-complier mass and the lower and upper threshold cuts computed from the unconditional capacity array $(\mathfrak{l}_i(x), \mathfrak{h}_i(x))_{i\in\mathcal{Y}}$. Let $\widehat{\mathfrak{m}}_n(x)$, $\widehat{\mathfrak{B}}_{L,n}(x)$, and $\widehat{\mathfrak{B}}_{U,n}(x)$ denote the corresponding quantities computed from $(\widehat{\mathfrak{l}}_{i,n}(x), \widehat{\mathfrak{h}}_{i,n}(x))_{i\in\mathcal{Y}}$. Since $p_x \geq 0$ and $\widehat{p}_x \geq 0$, Lemmas 1 to 3 give the homogeneity identities

$$\mathfrak{m}(x) = p_x m(x), \qquad \mathfrak{B}_L(x) = p_x B_L(x), \qquad \mathfrak{B}_U(x) = p_x B_U(x),$$

and

$$\widehat{\mathfrak{m}}_n(x) = \widehat{p}_x\widehat{m}(x), \qquad \widehat{\mathfrak{B}}_{L,n}(x) = \widehat{p}_x\widehat{B}_L(x), \qquad \widehat{\mathfrak{B}}_{U,n}(x) = \widehat{p}_x\widehat{B}_U(x).$$

The preceding coordinate bound says that every lower and upper coordinate in these unconditional arrays changes by at most e_{cap}. Hence each total $\sum_i \mathfrak{l}_i(x)$ or $\sum_i \mathfrak{h}_i(x)$ changes by at most Ke_{cap}, and the difference of these totals changes by at most $2Ke_{\text{cap}}$. Since $u \mapsto \min\{u, 0\}$, finite maxima, and finite minima are one-Lipschitz in the sup norm, the lower threshold candidate

$$\sum_{i\leq t}\mathfrak{l}_i(x) - \sum_{j\leq t}\mathfrak{h}_j(x) + \min\left\{\sum_j \mathfrak{h}_j(x) - \sum_i \mathfrak{l}_i(x), 0\right\}$$

changes by at most

$$Ke_{\text{cap}} + Ke_{\text{cap}} + 2Ke_{\text{cap}} = 4Ke_{\text{cap}}.$$

The zero candidate in the outer maximum is unchanged, so $p_x B_L(x) = \mathfrak{B}_L(x)$ changes by at most $4Ke_{\text{cap}}$. For the upper cut, the mass candidate changes by at most Ke_{cap}, and each strict-prefix-plus-tail candidate changes by at most $2Ke_{\text{cap}}$; therefore $p_x B_U(x) = \mathfrak{B}_U(x)$ changes by at most $4Ke_{\text{cap}}$. The same argument gives the bound $4Ke_{\text{cap}}$ for the unconditional survivor mass $p_x m(x) = \mathfrak{m}(x)$. Summing over the finite covariate support yields

$$\left|\sum_x \widehat{p}_x \widehat{m}(x) - \sum_x p_x m(x)\right| \le 4K|\mathcal{X}|e_{\text{cap}} = \rho_{\text{cap}}\delta,$$

and

$$\left|\sum_x \widehat{p}_x \widehat{B}_L(x) - \sum_x p_x B_L(x)\right| \le \rho_{\text{cap}}\delta, \qquad \left|\sum_x \widehat{p}_x \widehat{B}_U(x) - \sum_x p_x B_U(x)\right| \le \rho_{\text{cap}}\delta.$$

Before applying the screening rule, record the cut bounds used for both the population capacities and the projected empirical capacities. Fix a covariate cell x and either one of these two nonnegative capacity arrays; write its associated totals, gap, mass, and cuts with a superscript $\circ$:

$$q_0^\circ(x) = \sum_{i\in\mathcal{Y}} \ell_i^\circ(x), \qquad q_1^\circ(x) = \sum_{j\in\mathcal{Y}} h_j^\circ(x),$$

$$\Delta q^\circ(x) = q_1^\circ(x) - q_0^\circ(x), \qquad m^\circ(x) = \min\{q_0^\circ(x), q_1^\circ(x)\}.$$

All coordinates are nonnegative, so $m^\circ(x) \ge 0$. By the formula in Definition 16, $B_L^\circ(x)$ is the maximum of the zero candidate and the threshold candidates, and therefore

$$0 \le B_L^\circ(x).$$

The upper cut is the minimum of $m^\circ(x)$ and nonnegative strict-prefix-plus-tail sums, so

$$0 \le B_U^\circ(x) \qquad \text{and} \qquad B_U^\circ(x) \le m^\circ(x).$$

It remains to bound the lower cut above by the mass. For a threshold $t \in \mathcal{Y}$, define its lower-cut candidate by

$$C_t^\circ(x) = \ell_{\le t}^\circ(x) - h_{\le t}^\circ(x) + \min\{\Delta q^\circ(x), 0\}.$$

If $q_0^\circ(x) \le q_1^\circ(x)$, then $m^\circ(x) = q_0^\circ(x)$, $h_{\le t}^\circ(x) \ge 0$, and $\min\{\Delta q^\circ(x), 0\} \le 0$; hence

$$C_t^\circ(x) \le \ell_{\le t}^\circ(x) \le q_0^\circ(x) = m^\circ(x).$$

If $q_1^\circ(x) \le q_0^\circ(x)$, then $m^\circ(x) = q_1^\circ(x)$ and $\min\{\Delta q^\circ(x), 0\} = q_1^\circ(x) - q_0^\circ(x)$, so

$$C_t^\circ(x) = \ell_{\le t}^\circ(x) - h_{\le t}^\circ(x) + q_1^\circ(x) - q_0^\circ(x) \le q_1^\circ(x) = m^\circ(x),$$

because $\ell_{\le t}^\circ(x) \le q_0^\circ(x)$ and $h_{\le t}^\circ(x) \ge 0$. The zero candidate is also at most $m^\circ(x)$. Thus every candidate in the finite maximum defining the lower cut is at most $m^\circ(x)$, and

$$B_L^\circ(x) \le m^\circ(x).$$

Applying these inequalities first to the valid population capacities and then to the projected empirical capacities gives, for every x,

$$0 \le B_L(x), B_U(x) \le m(x), \qquad 0 \le \widehat{B}_L(x), \widehat{B}_U(x) \le \widehat{m}(x).$$

Screening removes, in each cell, at most η_n from the empirical mass and at most η_n from each empirical numerator. Indeed, a screened-out cell satisfies $\widehat{p}_x\widehat{m}(x) \le \eta_n$, while the just-proved empirical cut bounds give $0 \le \widehat{B}_L(x) \le \widehat{m}(x)$ and $0 \le \widehat{B}_U(x) \le \widehat{m}(x)$. Thus, with

$$N_L = \sum_x p_x B_L(x), \qquad N_U = \sum_x p_x B_U(x),$$

$$\left|\widehat{M}_n - M\right| \le \mathsf{A}_n(\delta), \qquad \left|\widehat{N}_{L,n} - N_L\right| \le \mathsf{A}_n(\delta), \qquad \left|\widehat{N}_{U,n} - N_U\right| \le \mathsf{A}_n(\delta).$$

Since $p_x \ge 0$, $\widehat{p}_x \ge 0$, and the retained-cell indicators are nonnegative, the same cut bounds also imply

$$0 \le N_L, N_U \le M, \qquad 0 \le \widehat{N}_{L,n}, \widehat{N}_{U,n} \le \widehat{M}_n.$$

If $4\mathsf{A}_n(\delta) < m_\star$, then

$$\widehat{M}_n \ge M - \mathsf{A}_n(\delta) \ge m_\star - \mathsf{A}_n(\delta) > \frac{3m_\star}{4},$$

so the ratio branch of $\widehat{\Psi}_n$ is active. For either endpoint numerator $N \in \{N_L, N_U\}$ and its empirical screened counterpart $\widehat{N}$,

$$\left|\frac{\widehat{N}}{\widehat{M}_n} - \frac{N}{M}\right| \le \frac{|\widehat{N} - N|}{\widehat{M}_n} + \frac{N|M - \widehat{M}_n|}{M\widehat{M}_n} \le \frac{2\mathsf{A}_n(\delta)}{\widehat{M}_n} \le \frac{3\mathsf{A}_n(\delta)}{m_\star}.$$

If $4\mathsf{A}_n(\delta) \ge m_\star$, then $\min\{1, 4\mathsf{A}_n(\delta)/m_\star\} = 1$, and both the target endpoints and the plug-in endpoints lie in $[0,1]$. Combining the two cases gives

$$\max\left\{\left|\widehat{\theta}_{L,n,\lambda}(\omega) - \theta_{L,\lambda}\right|, \left|\widehat{\theta}_{U,n,\lambda}(\omega) - \theta_{U,\lambda}\right|\right\} \le \min\left\{1, \frac{4\mathsf{A}_n(\delta)}{m_\star}\right\},$$

and the sharper bound

$$4\mathsf{A}_n(\delta) < m_\star \quad \Longrightarrow \quad \max\left\{\left|\widehat{\theta}_{L,n,\lambda}(\omega) - \theta_{L,\lambda}\right|, \left|\widehat{\theta}_{U,n,\lambda}(\omega) - \theta_{U,\lambda}\right|\right\} \le \frac{3\mathsf{A}_n(\delta)}{m_\star}.$$

2. For $\tau > 0$, the maximal guard deviation has the finite second-moment tail bound

$$\mu_\lambda\{\omega : \tau < \delta_{n,\lambda}(\omega)\} \le \frac{|\mathcal{E}|}{4n\tau^2}.$$

To see this, for $\mathsf{e} \in \mathcal{E}$ put

$$W_{r,\mathsf{e}} = \mathbf{1}\{\mathsf{e}(O_{\lambda,r})\} - P_{\mathrm{obs},\lambda}(\mathsf{e}), \qquad r = 1, \ldots, n.$$

The i.i.d. sampling condition gives independent mean-zero variables with

$$\mathrm{Var}_{\mu_\lambda}(W_{r,\mathsf{e}}) = P_{\mathrm{obs},\lambda}(\mathsf{e})\{1 - P_{\mathrm{obs},\lambda}(\mathsf{e})\} \le \frac{1}{4}.$$

Hence

$$\widehat{P}_{n,\lambda}(\mathsf{e}; \omega) - P_{\mathrm{obs},\lambda}(\mathsf{e}) = \frac{1}{n}\sum_{r=1}^{n} W_{r,\mathsf{e}}(\omega), \qquad \mathrm{Var}_{\mu_\lambda}\left(\frac{1}{n}\sum_{r=1}^{n} W_{r,\mathsf{e}}\right) \le \frac{1}{4n}.$$

Chebyshev's inequality and a finite union bound over $\mathcal{E}$ give the displayed tail bound. Since $\mathcal{X}$ is nonempty, $\mathcal{E}$ contains a cell event, so

$$b_{n,\alpha} = \sqrt{\frac{|\mathcal{E}|}{4n\alpha}} > 0$$

and therefore

$$\mu_\lambda\{\omega : b_{n,\alpha} < \delta_{n,\lambda}(\omega)\} \leq \alpha.$$

3. On the event $\{\omega : \delta_{n,\lambda}(\omega) \leq b_{n,\alpha}\}$, the first step with $\delta = b_{n,\alpha}$ and Algorithm 3 give

$$\max\left\{\left|\widehat{\theta}_{L,n,\lambda}(\omega) - \theta_{L,\lambda}\right|, \left|\widehat{\theta}_{U,n,\lambda}(\omega) - \theta_{U,\lambda}\right|\right\} \leq g_{n,\alpha}.$$

Let $y \in \Theta_I(P_{\text{obs},\lambda})$. By Definition 17,

$$\theta_{L,\lambda} \leq y \leq \theta_{U,\lambda},$$

and the endpoint bounds established from $0 \leq B_L(x), B_U(x) \leq m(x)$ give $0 \leq \theta_{L,\lambda} \leq 1$ and $0 \leq \theta_{U,\lambda} \leq 1$. Thus

$$\max\{0, \widehat{\theta}_{L,n,\lambda}(\omega) - g_{n,\alpha}\} \leq y \leq \min\{1, \widehat{\theta}_{U,n,\lambda}(\omega) + g_{n,\alpha}\},$$

which is exactly $y \in \mathcal{C}_{1-\alpha,n,\lambda}(\omega)$. The tail bound in the previous step therefore implies

$$\mu_\lambda\{\omega : \Theta_I(P_{\text{obs},\lambda}) \subseteq \mathcal{C}_{1-\alpha,n,\lambda}(\omega)\} \geq 1 - \alpha.$$

Taking the infimum over $\lambda \in \Lambda$ proves the finite-sample coverage claim for this n, and $n \geq 1$ was arbitrary.

4. Define

$$\tau_n = \frac{|\mathcal{X}|}{4\rho_{\text{cap}}}\eta_n.$$

Since $K \geq 3$, $|\mathcal{X}| > 0$, and $\varepsilon_Z > 0$, we have $\rho_{\text{cap}} > 0$. Together with the screening-threshold assumption $\eta_n > 0$, this gives $\tau_n > 0$. Moreover,

$$4\{\rho_{\text{cap}}\tau_n + |\mathcal{X}|\eta_n\} = 5|\mathcal{X}|\eta_n \longrightarrow 0,$$

so eventually

$$4\{\rho_{\text{cap}}\tau_n + |\mathcal{X}|\eta_n\} < m_\star.$$

For such n, if $\delta_{n,\lambda}(\omega) \leq \tau_n$, the sharper deterministic bound from the first step gives

$$\max\left\{\left|\widehat{\theta}_{L,n,\lambda}(\omega) - \theta_{L,\lambda}\right|, \left|\widehat{\theta}_{U,n,\lambda}(\omega) - \theta_{U,\lambda}\right|\right\} \leq \frac{3\{\rho_{\text{cap}}\tau_n + |\mathcal{X}|\eta_n\}}{m_\star}.$$

Because $\rho_{\text{cap}}\tau_n = |\mathcal{X}|\eta_n/4$ and $b_{n,\alpha} \geq 0$,

$$\frac{3\{\rho_{\text{cap}}\tau_n + |\mathcal{X}|\eta_n\}}{m_\star} \leq \frac{4\{\rho_{\text{cap}}b_{n,\alpha} + |\mathcal{X}|\eta_n\}}{m_\star}.$$

Together with the bound by 1 from the first deterministic inequality, this yields

$$\max\left\{\left|\widehat{\theta}_{L,n,\lambda}(\omega) - \theta_{L,\lambda}\right|, \left|\widehat{\theta}_{U,n,\lambda}(\omega) - \theta_{U,\lambda}\right|\right\} \leq g_{n,\alpha}.$$

Consequently, eventually in n,

$$\left\{\omega : g_{n,\alpha} < \max\left(\left|\widehat{\theta}_{L,n,\lambda}(\omega) - \theta_{L,\lambda}\right|, \left|\widehat{\theta}_{U,n,\lambda}(\omega) - \theta_{U,\lambda}\right|\right)\right\} \subseteq \{\omega : \tau_n < \delta_{n,\lambda}(\omega)\}.$$

The tail bound gives, uniformly in λ,

$$\mu_\lambda\{\omega : \tau_n < \delta_{n,\lambda}(\omega)\} \leq \frac{|\mathcal{E}|}{4n\tau_n^2} = \frac{4\rho_{\text{cap}}^2|\mathcal{E}|}{n|\mathcal{X}|^2\eta_n^2}.$$

The right-hand side tends to 0 because $\sqrt{n}\,\eta_n \to \infty$. This proves

$$\sup_{\lambda\in\Lambda}\mu_\lambda\left\{\omega : g_{n,\alpha} < \max\left\{\left|\widehat{\theta}_{L,n,\lambda}(\omega) - \theta_{L,\lambda}\right|, \left|\widehat{\theta}_{U,n,\lambda}(\omega) - \theta_{U,\lambda}\right|\right\}\right\} \longrightarrow 0.$$

5. Let

$$\beta_\alpha = \sqrt{\frac{|\mathcal{E}|}{4\alpha}}.$$

For $n \geq 1$,

$$b_{n,\alpha} = \beta_\alpha n^{-1/2}.$$

Therefore Algorithm 3 gives

$$0 \leq g_{n,\alpha} \leq \frac{4}{m_\star}\left\{\rho_{\text{cap}}\beta_\alpha n^{-1/2} + |\mathcal{X}|\eta_n\right\} \leq \frac{4(\rho_{\text{cap}}\beta_\alpha + |\mathcal{X}|)}{m_\star}\left(n^{-1/2} + \eta_n\right),$$

which proves

$$g_{n,\alpha} = O(n^{-1/2} + \eta_n).$$

Now let $L_n > 0$, $L_n \to \infty$, and $L_n = o(\sqrt{n})$, and set

$$\eta_n = n^{-1/2}L_n.$$

For every $n \geq 1$, $\eta_n > 0$. Also

$$\eta_n = \frac{L_n}{\sqrt{n}} \longrightarrow 0, \qquad \sqrt{n}\,\eta_n = L_n \longrightarrow \infty.$$

Since $L_n \to \infty$, eventually $L_n \geq 1$, and hence $n^{-1/2} \leq n^{-1/2}L_n = \eta_n$. The preceding guard bound then gives, eventually,

$$g_{n,\alpha} \leq \frac{8(\rho_{\text{cap}}\beta_\alpha + |\mathcal{X}|)}{m_\star}\,n^{-1/2}L_n,$$

so

$$g_{n,\alpha} = O(n^{-1/2}L_n).$$

□

References

Joshua D. Angrist, Guido W. Imbens, and Donald B. Rubin. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455, 1996. doi: 10.1080/01621459.1996.10476902.

Alexander Balke and Judea Pearl. Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92(439):1171–1176, 1997. doi: 10.1080/01621459.1997.10474074.

Otávio Bartalotti, Désiré Kédagni, and Vitor Possebom. Identifying marginal treatment effects in the presence of sample selection. *Journal of Econometrics*, 234(2):565–584, 2023. doi: 10.1016/j.jeconom.2021.11.011.

Eli Ben-Michael. Partial identification via conditional linear programs: Estimation and policy learning, 2025.

German Blanco, Carlos A. Flores, and Alfonso Flores-Lagunes. Bounds on average and quantile treatment effects of Job Corps training on wages. *Journal of Human Resources*, 48(3):659–701, 2013. doi: 10.3368/jhr.48.3.659.

Xiaohang Chen and Fan Li. Principal stratification with U-statistics under principal ignorability. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, page qkag044, 2026. doi: 10.1093/jrsssb/qkag044.

Xiaohong Chen and Carlos A. Flores. Bounds on treatment effects in the presence of sample selection and noncompliance: The wage effects of Job Corps. *Journal of Business & Economic Statistics*, 33(4):523–540, 2015. doi: 10.1080/07350015.2014.975229.

Xiaohong Chen, Carlos A. Flores, and Alfonso Flores-Lagunes. Going beyond LATE: Bounding average treatment effects of Job Corps training. *Journal of Human Resources*, 53(4):1050–1099, 2018. URL https://jhr.uwpress.org/content/53/4/1050.

Victor Chernozhukov, Han Hong, and Elie Tamer. Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, 75(5):1243–1284, 2007. doi: 10.1111/j.1468-0262.2007.00794.x.

Victor Chernozhukov, Sokbae Lee, and Adam M. Rosen. Intersection bounds: Estimation and inference. *Econometrica*, 81(2):667–737, 2013. doi: 10.3982/ECTA8718.

Johan de Aguas, Sebastian Krumscheid, Johan Pensar, and Guido Biele. The probability of tiered benefit: Partial identification with robust and stable inference. In *Proceedings of the First Conference on Causal Learning and Reasoning*, volume 275 of *Proceedings of Machine Learning Research*, pages 90–113. PMLR, 2025. URL https://proceedings.mlr.press/v275/aguas25a.html.

Yingying Dong and Phillip Heiler. Sharp bounds and inference in sample selection models with treatment endogeneity, 2026.

Guilherme Duarte, Noam Finkelstein, Dean Knox, Jonathan Mummolo, and Ilya Shpitser. An automated approach to causal inference in discrete settings. *Journal of the American Statistical Association*, pages 1–16, 2023. doi: 10.1080/01621459.2023.2216909.

Yanqin Fan and Sang Soo Park. Sharp bounds on the distribution of treatment effects and their statistical inference. *Econometric Theory*, 26(3):931–951, 2010. doi: 10.1017/S0266466609990168.

Zheng Fang and Andres Santos. Inference on directionally differentiable functions. *Review of Economic Studies*, 86(1):377–412, 2019. doi: 10.1093/restud/rdy049.

Constantine E. Frangakis and Donald B. Rubin. Principal stratification in causal inference. *Biometrics*, 58(1):21–29, 2002. doi: 10.1111/j.0006-341X.2002.00021.x.

Erin E. Gabriel, Michael C. Sachs, and Annette K. Jensen. Sharp symbolic nonparametric bounds for measures of benefit in observational and imperfect randomized studies with ordinal outcomes. *Biometrika*, 111(4):1429–1436, 2024. doi: 10.1093/biomet/asae020.

James J. Heckman. Sample selection bias as a specification error. *Econometrica*, 47(1):153–161, 1979. doi: 10.2307/1912352.

Han Hong and Jessie Li. The numerical delta method. *Journal of Econometrics*, 206(2):379–394, 2018. doi: 10.1016/j.jeconom.2018.06.007.

Guido W. Imbens and Joshua D. Angrist. Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475, 1994. doi: 10.2307/2951620.

Guido W. Imbens and Charles F. Manski. Confidence intervals for partially identified parameters. *Econometrica*, 72(6):1845–1857, 2004. doi: 10.1111/j.1468-0262.2004.00555.x.

Edward H. Kennedy, Scott Harris, and Luke J. Keele. Survivor-complier effects in the presence of selection on treatment, with application to a study of prompt ICU admission. *Journal of the American Statistical Association*, 114(525):93–104, 2019. doi: 10.1080/01621459.2018.1469990.

Gevorg Khandamiryan and Vira Semenova. Adaptive estimation of aggregated values of conditional linear programs, 2026.

David S. Lee. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies*, 76(3):1071–1102, 2009. doi: 10.1111/j.1467-937X.2009.00536.x.

Ying-Ying Lee and Chu-An Liu. Lee bounds with a continuous treatment in sample selection, 2024.

Alexander W. Levis, Matteo Bonvini, Edward H. Kennedy, and Luke J. Keele. Covariate-assisted bounds on causal effects with instrumental variables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 87(5):1508–1527, 2025. doi: 10.1093/jrsssb/qkaf028.

Jiannan Lu, Peng Ding, and Tirthankar Dasgupta. Treatment effects on ordinal outcomes: Causal estimands and sharp bounds. *Journal of Educational and Behavioral Statistics*, 43(5):540–567, 2018. doi: 10.3102/1076998618776435.

Jiannan Lu, Yukun Zhang, and Peng Ding. Sharp bounds on the relative treatment effect for ordinal outcomes. *Biometrics*, 76(2):664–669, 2020. doi: 10.1111/biom.13148.

G. D. Makarov. Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed. *Theory of Probability and Its Applications*, 26(4):803–806, 1982. doi: 10.1137/1126086.

Charles F. Manski. Nonparametric bounds on treatment effects. *American Economic Review*, 80(2):319–323, 1990. URL https://www.jstor.org/stable/2006592.

Charles F. Manski. Monotone treatment response. *Econometrica*, 65(6):1311–1334, 1997. doi: 10.2307/2171738.

Charles F. Manski. *Partial Identification of Probability Distributions*. Springer Series in Statistics. Springer, 2003. doi: 10.1007/b97478.

Charles F. Manski and John V. Pepper. Monotone instrumental variables: With an application to the returns to schooling. *Econometrica*, 68(4):997–1010, 2000. doi: 10.1111/1468-0262.00144.

Francesca Molinari. Microeconometrics with partial identification. In Steven N. Durlauf, Lars Peter Hansen, James J. Heckman, and Rosa L. Matzkin, editors, *Handbook of Econometrics*, volume 7A, pages 355–486. Elsevier, 2020.

Claudia Noack. Sensitivity of LATE estimates to violations of the monotonicity assumption, 2026. Working paper.

Vitor Possebom and Flavio Riva. Probability of causation with sample selection: A reanalysis of the impacts of Jóvenes en Acción on formality. *Journal of Business & Economic Statistics*, 43(2):391–400, 2025. doi: 10.1080/07350015.2024.2388639.

Joseph P. Romano and Azeem M. Shaikh. Inference for the identified set in partially identified econometric models. *Econometrica*, 78(1):169–211, 2010. doi: 10.3982/ECTA6706.

Thomas M. Russell. Sharp bounds on functionals of the joint distribution in the analysis of treatment effects. *Journal of Business & Economic Statistics*, 39(2):532–546, 2021. doi: 10.1080/07350015.2019.1684300.

Michael C. Sachs, Erin E. Gabriel, and Arvid Sjölander. A general method for deriving tight symbolic bounds on causal effects. *Journal of Computational and Graphical Statistics*, 32(2):567–576, 2023. doi: 10.1080/10618600.2022.2071905.

Vira Semenova. Generalized Lee bounds. *Journal of Econometrics*, 251:106055, 2025. doi: 10.1016/j.jeconom.2025.106055.

Elie Tamer. Partial identification in econometrics. *Annual Review of Economics*, 2:167–195, 2010. doi: 10.1146/annurev.economics.050708.143401.

Jin Tian and Judea Pearl. Probabilities of causation: Bounds and identification. *Annals of Mathematics and Artificial Intelligence*, 28(1–4):287–313, 2000. doi: 10.1023/A:1018912507879.

Aad W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998. doi: 10.1017/CBO9780511802256.

Cédric Villani. *Optimal Transport: Old and New*. Springer, 2009. doi: 10.1007/978-3-540-71050-9.

Robert C. Williamson and Tom Downs. Probabilistic arithmetic. I. numerical methods for calculating convolutions and dependency bounds. *International Journal of Approximate Reasoning*, 4(2):89–158, 1990. doi: 10.1016/0888-613X(90)90022-T.