Quotient-law inference with latent treatment-effect collisions

CausalSmith*

September 12, 2026

Abstract

This paper studies proxy-based causal inference for latent-class mean treatment effects in a fixed finite-class model. Under the Virk–Mazaheri–Wu proxy moment structure with fixed latent cardinality k (the number of latent classes), fixed proxy dimensions, bounded observable contributions, supplied latent-arm positivity and proxy-rank margins, the target is the quotient latent-effect law ν_P, the population-weighted distribution of latent-class average treatment effects with masses aggregated at coincident class means. The main population result establishes a gap-free Lipschitz modulus from the observable five-block summary $S(P)$, the vector of proxy moment summaries, to ν_P in W_1, the one-Wasserstein distance. The modulus covers homogeneous, partially colliding, and separated effect configurations within the same uniformly conditioned class.

The statistical results turn this modulus into root-n law estimation and confidence reporting. A nearest-summary repair estimator and a structured-lattice estimator, calibrated by the supplied constants $k, d_x, d_z, L, \pi_0, \sigma_0$, are Borel sample maps and achieve uniform root-n W_1 tail bounds over the model class. In the fixed-dimensional unit-cost exact-real model, the lattice estimator has polynomial candidate growth in n and centers an honest finitely represented exact-real Wasserstein outer confidence set. The associated cluster report merges empirically unresolved support points and reports compatible aggregate mass intervals. On separated gap-local strata, effect-ordered latent masses have clipped inverse-gap risk $\min\{1, (\sqrt{n}\, g)^{-1}\}$, where g is the local effect-gap scale. Local and same-class lower bounds from explicit two-class proxy experiments establish root-n sharpness for quotient-law estimation and the inverse-gap rate for ordered weights on the displayed two-class specialization.

1 Introduction

Proxy causal models use measurements related to latent confounding or latent classes to recover causal objects from observational data. In the potential-outcomes tradition [Splawa-Neyman et al., 1990, Rubin, 1974, Holland, 1986, Imbens and Rubin, 2015], latent heterogeneity is typically organized through assumptions linking treatment assignment, outcomes, and observed covariates. Negative-control and proximal approaches enrich that structure by introducing proxy measurements with conditional-independence and rank properties [Miao et al., 2018, Shi et al., 2020, Tchetgen Tchetgen et al., 2024, Cui et al., 2024, Miao et al., 2023, Qi et al., 2024, Liu et al., 2024, Li et al., 2024, Ai and Shan, 2025, Saha et al., 2026]. The finite-class proxy model of

*Machine-generated by the CausalSmith research pipeline. Each displayed formal statement is either mapped to a Lean declaration or marked as a presentation-level definition, and each displayed proof renders a Lean proof, as recorded in the verification appendix (scope, toolchain, commit, and any statement conditional on a published input). Cited work otherwise supplies framing and positioning.

Virk et al. [2026] supplies a compressed observable operator whose spectral structure identifies latent-class mean treatment effects under separated recovery conditions. This paper develops the law-level inference theory for the same proxy moment geometry when equal or nearly equal class-average effects are represented through the probability law of effect values.

The central estimand is the quotient latent-effect law

$$\nu_P = \sum_{u=1}^{k} p_u \delta_{\tau_u},$$

the distribution that places latent-class mass p_u at the class treatment effect $\tau_u = E[Y(1) \mid U = u] - E[Y(0) \mid U = u]$ and aggregates the mass of classes with equal class means. The support points are latent-class average treatment effects; unit-level contrasts $Y(1) - Y(0)$ enter through these conditional means. The law records how much population mass has each class-average effect value, with collisions interpreted as shared support points. The observable input is the five-block summary

$$S(P) = \{M_0(P), M_1(P), N_0(P), N_1(P), m_X(P)\},$$

formed from armwise proxy moments, outcome-weighted proxy moments, and the target-proxy mean.

The model class $\mathcal{M}$, the uniformly conditioned proxy causal class, fixes k, d_x, d_z, imposes the VMW proxy conditional-independence and consistency restrictions, and uses supplied quantitative bounds: the observable envelope L, a lower bound π_0 on every joint latent-arm probability, and a singular-value margin σ_0 for the reference- and target-proxy feature matrices. These conditions yield the observed proxy factorizations and rank margins in Proposition 1, compactness of the feasible summary closure in Proposition 2, and a common compact support interval for latent-class mean effect laws. The fixed-margin formulation is the population domain for all upper bounds and confidence statements in the paper.

The first main result is the gap-free modulus in Theorem 1. It establishes a constant C_{mod}, depending only on the fixed dimensions and conditioning constants, such that

$$W_1(\nu_P, \nu_Q) \le C_{\mathrm{mod}} d_S(S(P), S(Q))$$

for all $P, Q \in \mathcal{M}$. It also extends the summary-to-law map continuously to the compact summary closure. The statement is formulated in the one-Wasserstein metric W_1, so perturbations move aggregate probability mass between effect values and treat simultaneous merges or splits of atoms through the same law-level metric.

This population result leads to two estimators. The nearest-summary repair estimator in Algorithm 1 projects an empirical summary to the compact feasible summary closure and evaluates the continuous extension. The structured-lattice estimator in Algorithm 2 searches over finite grids of signal bases, conditioning matrices, latent masses, and effect locations, using a criterion that matches the empirical compressed operator, target-proxy mean, and anchor equation. Theorem 2 establishes that both estimators are Borel sample maps and achieve uniform root-n W_1 tail bounds over $\mathcal{M}$. The lattice construction in Proposition 4 has polynomial candidate growth in n for fixed dimensions and constants.

The confidence results report law-level uncertainty in forms aligned with the quotient target. Algorithm 3 defines an exact repaired-image summary-inversion set through the repaired map and a lattice-centered Wasserstein outer set with a finite constrained exact-real representation.

Theorem 3 gives simultaneous coverage and root-n diameter bounds for these sets. Algorithm 4 and Theorem 4 then translate the Wasserstein confidence set into support-component intervals and aggregate mass intervals. The data-facing report uses the empirical lattice law, supplied constants, and the algorithmic confidence set; the true-law objects $K_C(P)$ and $\Delta_C(P)$ name the validity quantities used to state coverage and width. Empirical atoms are connected at threshold $4\rho_{n,\alpha}$, while support association and interval expansion use the radius $\rho_{n,\alpha}$; the resulting mass intervals sharpen according to the external separation of the associated true component.

The paper also characterizes effect-ordered latent masses on separated strata. The ordered target $p^{\uparrow}(P)$, the vector of latent masses ordered by increasing effect, is studied on $\mathcal{M}(g)$, the gap-local stratum with positive effect gaps at scale g. Theorem 5 gives the upper risk bound

$$E\|\widehat{p}_n^{\uparrow} - p^{\uparrow}(P)\|_1 \leq C \min\{1, (\sqrt{n}\, g)^{-1}\}.$$

This rate expresses the conversion from law-level transport error to ordered mass recovery when effect locations are separated at scale g.

The lower bounds use explicit two-class proxy experiments. The witness law in Definition 55 verifies that a collision can occur inside the full-rank, strictly positive proxy model. Theorem 6 gives local two-point lower bounds for quotient laws and ordered weights, and Propositions 6 and 7 convert them into same-class minimax rates on the displayed two-class specialization. The quotient-law minimax rate is exactly root-n, while the ordered-weight minimax rate is the clipped inverse-gap rate. Theorem 7 records the corresponding transfer to comparator classes stated in the published VMW vocabulary when those classes contain the relevant witness pair.

The paper proceeds as follows. The related-work section positions the argument within proxy causal identification, spectral finite-mixture recovery, and Wasserstein inference. The setup section defines the proxy model, observable summaries, quotient law, and compact summary closure. The main-results section proves the law-level modulus and develops the repair and lattice estimators. The confidence section constructs Wasserstein confidence sets, cluster reports, and ordered-weight inference on gap-local strata. The lower-bound section gives the local and same-class minimax converses and the VMW transfer statement. The discussion section collects extensions and open questions, and the appendices provide auxiliary probability, selection, spectral, finite-net, proof, and verification material.

2 Related work

This paper connects three literatures: proxy causal identification, finite-mixture and spectral recovery, and Wasserstein inference for singular discrete laws. In the Neyman–Rubin potential-outcome tradition [Splawa-Neyman et al., 1990, Rubin, 1974, Holland, 1986, Imbens and Rubin, 2015], latent treatment-effect heterogeneity is usually interpreted through assumptions that relate treatment assignment, outcomes, and observed covariates. Proxy causal models enrich that structure by using measurements linked to latent confounding or latent classes, as in negative-control and proximal identification arguments [Miao et al., 2018, Tchetgen Tchetgen et al., 2024, Shi et al., 2020, Cui et al., 2024, Miao et al., 2023, Qi et al., 2024, Liu et al., 2024, Li et al., 2024, Ai and Shan, 2025, Saha et al., 2026]. The present analysis studies the finite-class proxy setting in which the observed proxy moments identify the quotient latent-effect law ν_P of Definition 40: the probability law that aggregates latent classes sharing the same class-average treatment effect. This target is tailored to collisions in effect values, because the inferential object is the distribution of class-average effect values and their aggregate masses.

The closest point of contact is Virk et al. [2026]. Their framework supplies the proxy moment structure, the compressed observable operator formalized here in Definition 39, and separated spectral recovery guarantees that make latent-class mean effects estimable when the relevant spectral quantities are well resolved. The present paper uses that operator as the population bridge from observable summaries to the quotient law and develops a W_1 modulus, with W_1 defined in Definition 19, for the law target. This perspective keeps the metric aligned with the identifiable probability measure: when several latent classes have the same class-average effect, their masses are combined at a common atom, and the stability statement is made for the quotient law. Label-sensitive lists enter through the separated ordered-weight results.

The identification argument also draws on classical and modern work on finite mixtures and tensor or spectral decompositions. Moment-based identification begins with Pearson [1894] and continues through finite-mixture identifiability results such as Teicher [1963], Lindsay [1995], and Allman et al. [2009]. Econometric proxy and measurement-error models use related rank and completeness ideas [Hu, 2008, Hu and Schennach, 2008, Kasahara and Shimotsu, 2009, Bonhomme et al., 2016], while algorithmic spectral methods exploit low-rank decompositions and perturbation theory [Hsu et al., 2009, Anandkumar et al., 2014, Sidiropoulos and Bro, 2000, Bauer and Fike, 1960, Kato, 1995, Davis and Kahan, 1970, Stewart and Sun, 1990, Mazaheri et al., 2025]. Within this lineage, the contribution here is a stability and inference theory for the atomic effect law induced by the proxy causal model, with fixed known dimensions and uniform conditioning constants.

A second comparison is with Wasserstein rates for finite mixtures near singularities. Collision phenomena are central in the mixture literature: when component locations approach one another, labeled parameters and weak distributional metrics exhibit different local geometries [Nguyen, 2013, Ho and Nguyen, 2016b, Heinrich and Kahn, 2018, Ho and Nguyen, 2016a, Manole and Ho, 2022, Nguyen et al., 2026]. The rates in Heinrich and Kahn [2018] clarify how overfitting and merging components affect Wasserstein estimation. The present setting is causal and proxy-based, but the same mathematical distinction matters: law-level estimation remains regular for the quotient distribution of latent-class mean effects, while effect-ordered weights are governed by local separation. The paper makes that distinction explicit by pairing collision-uniform quotient-law inference with inverse-gap behavior for ordered weights on the stated gap-local strata.

A quantitative comparison makes the difference in local geometry explicit. The closest results differ from the present paper in three respects at once: the experiment that is observed, the target the loss is measured on, and the conditions under which the stated rate holds.

result	experiment observed	target and loss	conditions and rate
Heinrich and Kahn [2018]	draws from an ordinary finite mixture	mixing measure, Wasserstein	strong identifiability, m components fitted around an m_0-component law: local minimax rate $n^{-1/(4(m-m_0)+2)}$
Doss et al. [2023]	draws from a high-dimensional Gaussian location mixture	mixing measure, Wasserstein	k components in dimension d: rate $\Theta\big((d/n)^{1/4} + n^{-1/(4k-2)}\big)$
Wu and Yang [2020]	draws from a Gaussian mixture	mixing measure, Wasserstein	moment projection onto the feasible moment space, attaining adaptive optimal rates
Virk et al. [2026]	proxy moments in the same causal model	labelled effects, features, and simplex-projected weights	rank, boundedness, positivity, and spectral separation: high-probability finite-sample first-order recovery for separated effects
this paper	proxy moments in the same causal model	quotient effect law ν_P, W_1	fixed k, d_x, d_z and uniform envelope, positivity, and rank margins: root-n, uniformly over collision configurations
this paper	proxy moments in the same causal model	effect-ordered masses $p^{\uparrow}(P)$	additionally a positive effect gap at scale g: clipped inverse-gap risk $\min\{1, (\sqrt{n}\, g)^{-1}\}$

The table locates the source of the difference. The ordinary-mixture results observe draws from a mixture distribution, so the map from the sampling law back to the mixing measure degenerates exactly where components merge, and the exponents above deteriorate with the number of colliding components. The present experiment observes a uniformly conditioned proxy-moment operator instead, and the compressed observable operator of Virk et al. [2026] carries the latent effects in its spectrum. Aggregating the mass of coinciding effects into the quotient law ν_P removes the labelling that the collision destroys, and the surviving map from observable summaries to ν_P stays Lipschitz across merges and splits by Theorem 1. Root-n quotient-law recovery and slower Wasserstein rates for ordinary mixing measures are therefore compatible: they measure different targets in different experiments. The labelled quantity that does retain collision sensitivity here is the effect-ordered mass vector, whose inverse-gap behaviour on gap-local strata is the direct analogue of the mixture exponents, and whose matching converse appears in Theorem 6.

Finally, the confidence-set construction is related to projection and inversion methods for weakly identified or singular models. Generalized method-of-moments traditions emphasize confidence regions obtained by inverting sample restrictions [Hansen, 1982]. Recent Wasserstein approaches for finite mixtures, including Wu and Yang [2020], Doss et al. [2023], Deo and Randrianarisoa [2023], Bing et al. [2026], and Niles-Weed and Rigollet [2022], develop distributional metrics, singular-rate analyses, and confidence constructions for nonregular finite-mixture settings. Here the observable proxy summary is the primitive sample object, and the confidence reports are expressed directly in Wasserstein distance for the quotient law and as cluster-adaptive support and mass summaries. This yields an econometric reporting target that matches the causal estimand while retaining the finite-mixture insight that separated labels and quotient laws have different local statistical behavior.

3 Setup and assumptions

Throughout the paper, k, d_x, and d_z are fixed finite dimensions, and treatment-arm and latent-class indices range over $t \in \{0,1\}$ and $u \in \{1,\ldots,k\}$. Vector norms are Euclidean, matrix norms are operator norms, and singular values are ordered decreasingly. Generic constants may depend on the fixed dimensions and conditioning constants, while asymptotic statements vary only with the sample size n. Expectations and conditional laws are taken under the full-data probability law under discussion unless a subscript specifies otherwise.

The section is organized around one dependency chain, which the reader may find it useful to keep in view: the observed and full-data records fix the sampling object; the causal and proxy assumptions constrain it; the latent masses, means, and effects are the parameters those assumptions govern; the observable population moments and the quotient law are the two derived targets; the model class and the gap-local stratum collect the uniform conditions under which the theorems hold; and the empirical summary is the sample counterpart of the observable moments. The spectral, lattice, and confidence objects that appear alongside these are the machinery that later sections consume, and each is stated where the theorem citing it can be checked against it.

For reading, the logical dependencies are as follows. The observed record and full-data law supply the variables; the causal and proxy assumptions define the uniformly conditioned class $\mathcal{M}$; the population moments $M_t(P)$, $N_t(P)$, and $m_X(P)$ form the observable summary $S(P)$; the quotient law ν_P is the population-weighted law of latent-class average treatment effects; and the empirical summary, spectral thresholding, lattice, confidence radii, and cluster objects are derived from those primitives. Some auxiliary objects are displayed where they are needed by later statements; the surrounding prose identifies their role in this dependency chain.

The statistical procedures are calibrated by supplied values of $k, d_x, d_z, L, \pi_0, \sigma_0$. A full-data law receives the stated guarantees when it satisfies the restrictions of $\mathcal{M}(k, d_x, d_z, L, \pi_0, \sigma_0)$ with those supplied constants. A larger valid envelope L, or smaller valid lower margins π_0 and σ_0, preserves membership after the constants and radii are recomputed, with the numerical bounds adjusted accordingly. The joint latent-arm margin encodes overlap within each latent class and arm, while the proxy-rank margin encodes stable variation in the reference- and target-proxy feature matrices.

The observed data consist of the treatment, two proxy vectors, and the realized outcome. The displayed objects below include both primitives and derived objects used later; the prose after each display identifies how it fits with the record spaces, observed margin, transport metric, and inferential target.

Definition 1. [synth_8] (Thresholded armwise Moore–Penrose inverse) *Let $s = (M_0(s), M_1(s), N_0(s), N_1(s), m_X$ be an observable summary and put $s_0 := \pi_0\sigma_0^2$. For $t \in \{0,1\}$, write a real singular-value expansion $M_t(s) = \sum_j \sigma_j u_j v_j^\top$, with $\sigma_j \geq 0$. The thresholded Moore–Penrose inverse of the armwise proxy moment is*

$$(M_t(s))^\dagger_{\geq s_0/2} := \sum_{\sigma_j \geq s_0/2} \sigma_j^{-1} v_j u_j^\top .$$

The observed record collects the binary treatment T, target proxy X, reference proxy Z, and observed outcome Y. The fixed target- and reference-proxy dimensions are d_x and d_z.

Definition 2. [synth_10] (Finite transport-plan relation) *Let $\nu = \sum_{a=1}^{r} w_a \delta_{x_a}$ and $\widehat{\lambda}_n = \sum_{b=1}^{\widehat{r}} \widehat{w}_b \delta_{\widehat{x}_b}$ be finite probability laws on $[-L_\tau, L_\tau]$. We write $\gamma : \nu \rightsquigarrow \widehat{\lambda}_n$ when*

$\gamma = (\gamma_{ab})_{1\le a\le r,\,1\le b\le \widehat{r}}$ is a finite transport plan with $\gamma_{ab} \ge 0$, $\sum_{b=1}^{\widehat{r}} \gamma_{ab} = w_a$ for every a, and $\sum_{a=1}^{r} \gamma_{ab} = \widehat{w}_b$ for every b. Its absolute-distance transport cost is $\sum_{a=1}^{r}\sum_{b=1}^{\widehat{r}} \gamma_{ab}|x_a - \widehat{x}_b|$.

The distance W_1 compares atomic laws by the minimum absolute-distance transport cost, so it measures the law of effect values. Label-specific comparisons enter later through the ordered-weight target on separated strata.

Definition 3. [synth_18] (VMW Assumption 4) *For a one-unit full-data law P, VMW Assumption 4 holds when $0 < k$, $k \le d_x$, $k \le d_z$, and there exist positive constants $L_X, L_{ZX}, L_{YZX}, \pi, \sigma$ such that $\|X\|_2 \le L_X$, $\|ZX^\top\|_{\mathrm{op}} \le L_{ZX}$, and $\|YZX^\top\|_{\mathrm{op}} \le L_{YZX}$ almost surely under P, both treatment-arm probabilities satisfy $P(T=t) \ge \pi$ for $t \in \{0,1\}$, and there is a population top-k right singular basis V of the stacked proxy moment $[M_0(P); M_1(P)]$ for which*

$$\sigma_k([M_0(P); M_1(P)]) \ge \sigma, \qquad \sigma_k(M_t(P)V) \ge \sigma \quad \text{for } t \in \{0,1\}.$$

The generic full-data law P carries the latent class U, treatment, proxies, potential outcomes $Y(t)$, and observed outcome. The observed map O induces the observed-data marginal P_O, which is the sampling law for the observable record.

Latent treatment-effect heterogeneity is described through class masses, class-specific potential-outcome means, and their differences.

Definition 4. [synth_19] (Summary radius $r_{n,\alpha}$) *For sample size n, miscoverage level α, concentration constant C_0, and envelope constant L, the simultaneous summary radius is*

$$r_{n,\alpha} := C_0 L \sqrt{\frac{\log(C_0/\alpha)}{n}}.$$

This is the radius used in the event $E_P = \{d_S(\widehat{S}_n, S(P)) \le r_{n,\alpha}\}$.

The mass p_u records how much of the population belongs to latent class u.

Definition 5. [synth_21] (Endpoint feasibility for cluster-mass extrema) *For confidence-set data with atom floor $m_\star$, outer radius $R_{n,\alpha}$, computed center $\widehat{\lambda}_n$, association radius $\rho_{n,\alpha}$, empirical component C, and candidate endpoint value m, say that* EndpointFeasible(C, m) *holds when there exists a law $\xi \in \mathcal{P}_{\le k, m_\star}([-L_\tau, L_\tau])$ and a transport plan γ from ξ to $\widehat{\lambda}_n$ such that*

$$\sum_{a,b} \gamma_{ab}|x_a - \widehat{x}_b| \le R_{n,\alpha}$$

and

$$m = \xi(K_C(\xi)), \qquad K_C(\xi) := \{x \in \mathrm{supp}(\xi) : \mathrm{dist}(x, C) \le \rho_{n,\alpha}\}.$$

Thus the lower and upper endpoints of I_C are the infimum and supremum of feasible values m over the same ordered-support, atom-floor, simplex, and transport constraints.

The mean μ_{tu} is the class-specific potential-outcome level for arm t.

Definition 6. [synth_28] (Atomic measure ν_ξ) *For a represented atomic record $\xi = ((w_i)_{i<k}, (a_i)_{i<k})$, define*

$$\nu_\xi = \sum_{i<k} \max\{w_i, 0\}\, \delta_{a_i}.$$

Repeated labelled locations contribute additively through the nonnegative masses $\max\{w_i, 0\}$.

The latent effect τ_u is the class-level treatment contrast.

The proxy structure is encoded by reference- and target-feature matrices. These matrices are the finite-dimensional objects through which the observable conditional moments factor.

Definition 7. [synth_29] (Candidate count N_A) *For a prescribed structured-lattice estimator A, let*

$$N_A \in \mathbb{N}$$

be the cardinality of its finite exhaustive candidate list. The listed candidates are indexed as

$$\vartheta_i = (V_i, R_i, p_i, \tau_i), \qquad i = 1, \ldots, N_A,$$

and this enumeration is the duplicate-free lexicographically ordered list of all well-formed structured lattice points at the stated sample size, conditioning constants, and effect-support radius.

The matrix A_t summarizes the reference proxy within latent classes and treatment arms.

Definition 8. [synth_30] (Matrix action operator) *For $r, c \in \mathbb{N}$ and any real $r \times c$ matrix A, the associated continuous linear operator from $\mathbb{R}^c$ to $\mathbb{R}^r$ is the ordinary finite-dimensional matrix action:*

$$x \longmapsto Ax.$$

The matrix B summarizes the target proxy within latent classes.

Definition 9. [synth_33] (Thresholded empirical compressed operator) *For a rectangular matrix $G = \sum_j \sigma_j u_j v_j^\top$, define the thresholded pseudoinverse*

$$G^\dagger_{\geq s_0/2} := \sum_{\sigma_j \geq s_0/2} \sigma_j^{-1} v_j u_j^\top.$$

The empirical compressed operator is

$$\widehat{D}_n := (\widehat{M}_{1,n})^\dagger_{\geq s_0/2} \widehat{N}_{1,n} - (\widehat{M}_{0,n})^\dagger_{\geq s_0/2} \widehat{N}_{0,n}.$$

The radius L_τ is the common compact interval radius used for latent-class mean effect laws.

Definition 10. [synth_24] (Structured comparison operator) *For a structured lattice point $\vartheta = (V, R, p, \tau)$, define the structured comparison operator by*

$$\mathrm{Op}(\vartheta) := V R^{-1} \operatorname{diag}(\tau) R V^\top.$$

For an empirical summary $s = (M_0, M_1, N_0, N_1, m_X)$ and threshold $s_0/2 = \pi_0 \sigma_0^2/2$, define the empirical comparison operator by

$$\widehat{\mathrm{Op}}(s) := M^\dagger_{1,\geq s_0/2} N_1 - M^\dagger_{0,\geq s_0/2} N_0,$$

where $M^\dagger_{t,\geq s_0/2}$ is the singular-value-thresholded pseudoinverse retaining singular directions with singular value at least $s_0/2$. In the lattice criterion, these operators are compared through

$$\| \mathrm{Op}(\vartheta) - \widehat{\mathrm{Op}}(\widehat{S}_n) \|_{\mathrm{op}}.$$

The gap $\delta(P)$ measures separation among distinct positive-mass effect values and is used only for label-sensitive ordered-weight statements.

Definition 11. [synth_35] (Matrix operator norm $\|A\|_{\mathrm{op}}$) *For natural numbers r and c and a real $r \times c$ matrix A, A acts as the ordinary finite-dimensional linear map*

$$x \mapsto Ax, \qquad x \in \mathbb{R}^c.$$

Its Euclidean operator norm is

$$\|A\|_{\mathrm{op}} := \sup_{\|x\|_2 \le 1} \|Ax\|_2.$$

Thus comparisons of matrix-valued operators use the ordinary matrix operator norm, for example

$$\|\operatorname{Op}(\vartheta) - \widehat{\operatorname{Op}}(\widehat{S}_n)\|_{\mathrm{op}}.$$

The moments $M_t(P)$, $N_t(P)$, and $m_X(P)$ are the population coordinates observed through the proxy experiment.

The analysis works with represented atomic laws while interpreting coincident atom locations as a single aggregate effect value.

Definition 12. [synth_36] (Quotient atomic law structure) *For every $k \in \mathbb{N}$ and $r \in \mathbb{R}$, consider the quotient of valid labelled at-most-k-atomic probability laws supported in the radius-r interval by equality of their induced measures. For a quotient law ν in this space,* representative(ν) *is the noncomputably chosen valid labelled representative of the equivalence class ν.*

The topology on this quotient space is the topology coinduced by the canonical quotient map from valid labelled laws to their induced-measure equivalence classes. Its measurable structure is the mapped measurable structure along the same quotient map. With this coinduced topology, the quotient space is compact.

Definition 13. [synth_32] (Structured lattice candidate spaces) *Put $m_\star := \pi_0$, $s_0 := \pi_0 \sigma_0^2$,*

$$H_n := \lceil \sqrt{n} \rceil + \lceil \pi_0^{-1} \rceil + 2k + \lceil 4\sqrt{d_x k} \rceil + \lceil 2k/\sigma_0 \rceil, \qquad q_n := H_n^{-1},$$

and $\ell_n := \lceil m_\star H_n \rceil$. Let $\mathcal{V}_n$ consist of the polar factors $G(G^\top G)^{-1/2}$ of all $G \in (q_n\mathbb{Z})^{d_x \times k}$ satisfying $|G_{ij}| \le 1$ and $\sigma_k(G) \ge 1/2$. Let $\mathcal{R}_n$ consist of all $R \in (q_n\mathbb{Z})^{k \times k}$ satisfying $\sigma_k(R) \ge \sigma_0/2$ and $\|R\|_{\mathrm{op}} \le 2\sqrt{k}L$. Define

$$\mathcal{P}_n := \left\{ H_n^{-1}(a_1, \ldots, a_k)^\top : a_u \in \{\ell_n, \ldots, H_n\},\ \sum_{u=1}^k a_u = H_n \right\},$$

and let $\mathcal{T}_n$ be the k-fold product of the clipped q_n-lattice in $[-L_\tau, L_\tau]$.

Definition 14. [synth_34] (Lattice-implied operator and quotient law) *For $\vartheta = (V, R, p, \tau) \in \mathcal{V}_n \times \mathcal{R}_n \times \mathcal{P}_n \times \mathcal{T}_n$, define*

$$D_\vartheta := VR^{-1} \operatorname{diag}(\tau) RV^\top, \qquad m_\vartheta := VR^\top p, \qquad b_\vartheta := RV^\top e_1,$$

and define the associated quotient law by

$$\lambda_\vartheta := \sum_{u=1}^k p_u \delta_{\tau_u}.$$

Definition 15. [synth_7] (All-ones vector $\mathbf{1}_k$) *The vector $\mathbf{1}_k \in \mathbb{R}^k$ is the all-ones vector, with coordinate $(\mathbf{1}_k)_u = 1$ for every latent-coordinate index $u = 1, \ldots, k$.*

The ordered target $p^{\uparrow}(P)$ is used when the effects are separated enough to support label-sensitive mass recovery.

Definition 16. [synth_37] (Structured-lattice objective $J_n(\vartheta; s)$) *Put $s_0 := \pi_0\sigma_0^2$. For a summary $s = (M_0, M_1, N_0, N_1, m_X) \in (\mathbb{R}^{d_z \times d_x})^4 \times \mathbb{R}^{d_x}$, define the thresholded pseudoinverse by writing a rectangular matrix $G = \sum_j \sigma_j u_j v_j^\top$ and setting $G^\dagger_{\geq s_0/2} := \sum_{\sigma_j \geq s_0/2} \sigma_j^{-1} v_j u_j^\top$. Set $\widehat{D}(s) := M^\dagger_{1 \geq s_0/2} N_1 - M^\dagger_{0 \geq s_0/2} N_0$. For $\vartheta = (V, R, p, \tau) \in \mathcal{V}_n \times \mathcal{R}_n \times \mathcal{P}_n \times \mathcal{T}_n$, with $V \in \mathbb{R}^{d_x \times k}$, $R \in \mathbb{R}^{k \times k}$ invertible on $\mathcal{R}_n$, $p \in \Delta^{k-1}$, and $\tau \in [-L_\tau, L_\tau]^k$, define $D_\vartheta := VR^{-1}\operatorname{diag}(\tau)RV^\top$, $m_\vartheta := VR^\top p$, $b_\vartheta := RV^\top e_1$, and $\lambda_\vartheta := \sum_{u=1}^k p_u \delta_{\tau_u}$. The structured-lattice criterion at summary s is*

$$J_n(\vartheta; s) := \|D_\vartheta - \widehat{D}(s)\|_{\mathrm{op}} + \|m_\vartheta - m_X\|_2 + \|b_\vartheta - \mathbf{1}_k\|_2.$$

For the empirical summary $\widehat{S}_n$, write $J_n(\vartheta) := J_n(\vartheta; \widehat{S}_n)$.

The model assumptions combine the proxy conditional-independence structure of Virk et al. [2026] with fixed quantitative margins. We first state the structural restrictions linking proxies, potential outcomes, treatment assignment, and the anchor coordinate.

Definition 17. [synth_9] (Operation count for the structured-lattice estimator) *For the structured-lattice estimator $\widehat{\lambda}_n$, let $\mathcal{A}_n := \mathcal{V}_n \times \mathcal{R}_n \times \mathcal{P}_n \times \mathcal{T}_n$ denote its lexicographically ordered finite candidate list. In the fixed-dimensional unit-cost exact-real model, define*

$$\operatorname{ops}(\widehat{\lambda}_n) := n + \#\mathcal{A}_n.$$

The term n counts formation of the empirical summary $\widehat{S}_n$, and $\#\mathcal{A}_n$ counts exhaustive evaluation of the fixed-dimensional objective $J_n(\vartheta)$ once for each $\vartheta \in \mathcal{A}_n$, including the fixed-size arithmetic, comparison, singular-value, polar-decomposition, and root-isolation operations used by that candidate evaluation.

Reference-proxy separation is the VMW reference-proxy conditional independence condition [Virk et al., 2026]. It makes the reference proxy vary with the latent class and treatment arm while separating it from the target proxy and outcome once those variables are fixed.

Definition 18. [def:observed-record-space] (Observed-record space O_{d_x,d_z}) *Write O_{d_x,d_z} for the observed-record space of quadruples*

$$O = (T, X, Z, Y),$$

where $T \in \{0, 1\}$, $X \in \mathbb{R}^{d_x}$, $Z \in \mathbb{R}^{d_z}$, and $Y \in \mathbb{R}$.

Target-proxy separation is the VMW target-proxy conditional independence condition [Virk et al., 2026]. It makes X a latent-class measurement whose conditional mean is stable across treatment arms.

Definition 19. [def:wasserstein-distance] (Wasserstein distance $W_1(\nu, \xi)$) *For represented atomic laws*

$$\nu = \sum_i p_i \delta_{x_i}, \qquad \xi = \sum_j q_j \delta_{y_j},$$

define $W_1(\nu,\xi)$ *by*

$$W_1(\nu,\xi) = \min_{\gamma_{ij}\geq 0} \sum_{i,j} \gamma_{ij}|x_i - y_j|,$$

where the minimum is over all nonnegative couplings $(\gamma_{ij})_{i,j}$ *whose marginals are the represented masses:*

$$\sum_j \gamma_{ij} = p_i \quad \text{for every } i, \qquad \sum_i \gamma_{ij} = q_j \quad \text{for every } j.$$

Observed-outcome consistency is the VMW causal consistency condition [Virk et al., 2026]. It connects the realized outcome to the treatment-indexed potential outcome.

Definition 20. [`def:observed-margin`] (Observed-data marginal P_O) *For a full-data probability law* P*, let*

$$O = (T, X, Z, Y)$$

be the observed record. The observed-data marginal of P *is*

$$P_O := P \circ O^{-1}.$$

Latent ignorability is the VMW armwise latent ignorability condition [Virk et al., 2026]. It places treatment assignment variation within latent classes in the role needed for the armwise proxy moments.

Definition 21. [`def:latent-mass`] (Latent-class mass p_u) *For each latent class* u*, its latent-class mass is*

$$p_u := P(U = u).$$

The target-proxy anchor is the VMW anchor-coordinate normalization [Virk et al., 2026]. It fixes the scale needed by the spectral anchors below.

The next conditions impose bounded contributions and uniform conditioning. The boundedness assumptions are standard VMW envelope conditions, while the positivity and rank margins are the fixed-margin structure specific to this analysis.

Definition 22. [`def:latent-mean`] (Latent conditional mean μ_{tu}) *For treatment arm* t *and latent class* u*, the latent conditional potential-outcome mean is*

$$\mu_{tu} := E_P[Y(t) \mid U = u]\,.$$

Bounded target-proxy contributions follow the VMW bounded target-proxy condition [Virk et al., 2026]. The radius L supplies the common envelope used for summary concentration and compactness.

Definition 23. [`def:latent-effect`] (Latent-class effect τ_u) *For each latent class* u*, its latent-class treatment effect is*

$$\tau_u := \mu_{1u} - \mu_{0u}.$$

The proxy-product bound is the VMW bounded proxy cross-moment contribution condition [Virk et al., 2026]. It controls the armwise $ZX^\top$ moment coordinates.

Definition 24. [`def:reference-feature`] (Reference-proxy feature matrix A_t) *For treatment arm* t*, the reference-proxy feature matrix* $A_t \in \mathbb{R}^{d_z \times k}$ *is defined entrywise by*

$$A_t(i,u) := E_P[Z_i \mid U = u, T = t]\,.$$

The outcome-weighted proxy-product bound is the VMW bounded outcome-weighted proxy contribution condition [Virk et al., 2026]. It controls the armwise $YZX^\top$ moment coordinates.

Definition 25. [`def:target-feature`] (Target-proxy feature matrix B) *The target-proxy feature matrix $B \in \mathbb{R}^{d_x \times k}$ is defined entrywise by*

$$B(i,u) := E_P[X_i \mid U = u] \,.$$

Latent-arm positivity is specific to this analysis. The supplied margin π_0 gives every latent class and treatment arm a fixed amount of population mass, which yields uniform arm probabilities and atom-mass control.

Definition 26. [`def:effect-radius`] (Effect-support radius L_τ) *The derived effect-support radius is*

$$L_\tau := \frac{4L\sqrt{d_z}}{\sigma_0}.$$

The proxy-rank margin is specific to this analysis. The supplied singular-value margin σ_0 gives the proxy feature matrices uniform conditioning, which is used to form stable compressed operators and fixed-radius effect supports.

For ordered weights, the paper restricts attention to local strata where all positive-mass latent effects are separated at a prescribed scale.

Definition 27. [`def:effect-gap`] (Effect gap $\delta(P)$) *The effect gap of P is*

$$\delta(P) := \min\bigl\{|\tau_u - \tau_v| : \ p_u > 0,\ p_v > 0,\ \tau_u \neq \tau_v\bigr\},$$

with the convention $\delta(P) = +\infty$ when the quotient law has a single distinct support point.

The gap window is specific to this analysis. The local scale g records the separation regime for ordered-weight rates.

Definition 28. [`def:observable-population-moments`] (Observable population moments $M_t(P)$, $N_t(P)$, and $m_X(P)$) *For each $t \in \{0,1\}$, the observable armwise population moments and target-proxy population mean are*

$$M_t(P) := E_P[ZX^\top \mid T = t], \qquad N_t(P) := E_P[YZX^\top \mid T = t], \qquad m_X(P) := E_P[X].$$

Distinct latent effects are the VMW spectral separation condition [Virk et al., 2026]. In this paper the condition is used for the gap-local ordered-weight target, while the quotient law below aggregates coincident effects.

Three representation levels are used for atomic laws. A raw labelled record consists of k weight coordinates and k location coordinates. A valid represented probability law is a raw record whose masses are nonnegative and sum to one, whose locations lie in the support interval, and whose repeated locations are interpreted by aggregating their masses. A quotient law is the equality class of valid represented laws that induce the same probability measure. The coordinate container in Definition 29 records the raw labelled representation; the W_1, atom-floor, estimator, and confidence-set statements use the valid represented-law and quotient-law levels indicated in their hypotheses and conclusions.

We can now collect the full-data restrictions into the uniformly conditioned proxy model class.

Definition 29. [`def:atomic-law-class`] (Atomic law coordinates) *For $k \in \mathbb{N}$ and radius parameter $L_\tau \in \mathbb{R}$, the represented atomic-law class*

$$\mathcal{P}_{\leq k}([-L_\tau, L_\tau])$$

is the labelled coordinate space whose element ξ consists of two real coordinate vectors,

$$\xi = \big((w_i)_{i<k}, (a_i)_{i<k}\big), \qquad w_i \in \mathbb{R}, \quad a_i \in \mathbb{R}.$$

The abbreviation $\mathcal{P}_k([-L_\tau, L_\tau])$ denotes the same represented space. The associated measure on $\mathbb{R}$ is

$$\nu_\xi = \sum_{i<k} \max\{w_i, 0\}\, \delta_{a_i},$$

so coincident labelled locations contribute additively at that location. The topology on this represented space is induced by the coordinate map

$$\xi \mapsto \big((w_i)_{i<k}, (a_i)_{i<k}\big),$$

and its measurable structure is the pullback of the measurable structure on the two real coordinate vectors under the same map.

Three distinct levels of representation are in play, and it is worth separating them once here because later definitions move between them. At the first level sits the labelled coordinate container just displayed: an element records k real weights and k real locations with no constraint tying them to a probability law. At the second level sit the valid labelled laws, those coordinate records whose weights are nonnegative, sum to one, and whose locations lie in the stated interval; the atom floor of Definition 31, the estimators, and the confidence sets are stated for laws of this kind. At the third level sits the quotient of Definition 12, formed from the valid labelled laws by identifying records that induce the same measure. The bridge between the levels is the induced measure ν_ξ displayed above, which sends coincident labelled locations to a single atom carrying their aggregate mass; W_1 compares laws through this measure, so it is insensitive to the labelling that a collision destroys. The estimand ν_P of Definition 40 is an object of the third level, which is what makes it the regular target under collisions, while the effect-ordered mass vector of Definition 30 retains second-level labelling information and is correspondingly governed by the effect gap.

The class $\mathcal{M}$ fixes the latent cardinality k, the boundedness radius, and the two conditioning margins. The qualitative proxy restrictions align with Virk et al. [2026]; the fixed quantitative margins make the subsequent stability and concentration statements uniform over the class.

The observable summary is the finite-dimensional population object to which the statistical results are applied.

Definition 30. [`def:ordered-mass-target`] (Ordered mass target $p^{\uparrow}(P)$) *When the k latent effects are distinct, $p^{\uparrow}(P)$ is the vector of latent-class masses ordered by increasing effect. Otherwise,*

$$p^{\uparrow}(P) := (1/k, \ldots, 1/k).$$

The summary $S(P)$ and metric d_S put the four matrix moments and target-proxy mean in a single product space.

Definition 31. [`def:atom-floor`] (Atom floor $\mathrm{AtomFloor}(m, \xi)$) *For a probability law ξ, $\mathrm{AtomFloor}(m, \xi)$ is the condition that every distinct support atom of ξ has aggregate mass at least m.*

The arm count $N_{t,n}$, empirical moments $\widehat{M}_{t,n}$ and $\widehat{N}_{t,n}$, empirical mean $\widehat{m}_{X,n}$, and empirical summary $\widehat{S}_n$ are total functions of the observed sample.

Assumption 1. [`ass:reference-proxy-separation`] (Reference-proxy separation) *Under the full-data law P,*

$$Z \perp (X, Y) \mid (T, U).$$

Assumption 2. [`ass:target-proxy-separation`] (Target-proxy separation) *Under the full-data law P,*

$$X \perp (Y, T) \mid U.$$

The vector e_1 implements the anchor coordinate from Assumption 5.

The compressed operator is the population spectral object inherited from the VMW moment structure.

Assumption 3. [`ass:consistency`] (Observed-outcome consistency) *Under the full-data law P,*

$$Y = Y(T)$$

almost surely.

Assumption 4. [`ass:latent-ignorability`] (Latent ignorability) *Under the full-data law P, for each $t \in \{0, 1\}$,*

$$Y(t) \perp T \mid U.$$

The signal basis $V(P)$, compressed operator $\Delta Q(P)$, and anchors $a(P)$ and $c(P)$ provide the population bridge from observable summaries to latent-class mean effect moments.

Assumption 5. [`ass:anchor`] (Target-proxy anchor) *Under the full-data law P,*

$$X_1 = 1$$

almost surely.

Assumption 6. [`ass:bounded-x`] (Bounded target proxy) *For the boundedness radius L, under the full-data law P,*

$$\|X\|_2 \le L$$

almost surely.

Assumption 7. [`ass:bounded-proxy-product`] (Bounded proxy product) *Under P,*

$$\|ZX^\top\|_{\mathrm{op}} \le L$$

almost surely.

The quotient law is the primary inferential target: it records the distribution of latent-class mean effect values with the aggregate mass at each distinct value.

Assumption 8. [`ass:bounded-outcome-proxy-product`] (Bounded outcome-proxy product) *Under P,*

$$\|YZX^\top\|_{\mathrm{op}} \le L$$

almost surely.

The stratum $\mathcal{M}(g)$ is the domain for ordered-weight results, where the gap scale and distinctness conditions provide an effect ordering.

Assumption 9. [`ass:latent-arm-positivity`] (Latent-arm positivity) *The latent class and treatment arms satisfy*

$$\min_{1 \le u \le k,\ t \in \{0,1\}} P(U = u, T = t) \ge \pi_0.$$

The experiment $\mathcal{E}_n$ contains the sample laws $Q_P^{(n)}$ generated by all model laws in $\mathcal{M}$.

The next definition records the closure of the admissible summary image and the quotient-law functional on it.

Assumption 10. [`ass:proxy-rank-margin`] (Proxy rank margin) *The reference- and target-proxy feature matrices satisfy*

$$\min\{\sigma_k(A_0), \sigma_k(A_1), \sigma_k(B)\} \ge \sigma_0.$$

The admissible image $\mathscr{S}$, closure $\mathcal{K}$, and functional F define the population domain for the stability theorem in the next section. The moment identity is the cited spectral bridge from Virk et al. [2026]; the homogeneous-effect clause records the corresponding single-atom specialization with common effect $\tau_\star$.

Assumption 11. [`ass:gap-window`] (Effect gap window) *The smallest positive effect gap satisfies*

$$g/2 \le \delta(P) \le 2g.$$

Assumption 12. [`ass:distinct-effects`] (Distinct latent effects) *The positive-mass latent effects τ_u take exactly k distinct values.*

Proposition 1 establishes the observable consequences of the full-data restrictions. It derives armwise positivity, the proxy moment factorization, compressed and stacked singular-value margins, observable envelopes, and the effect support radius. These conclusions place every $P \in \mathcal{M}$ inside the VMW moment structure under the fixed margins used here, while preserving the quantitative bounds needed for the quotient-law target.

Definition 32. [`def:model-class`] (Uniform proxy model class) *The uniformly conditioned proxy causal model class $\mathcal{M}$ is the class of full-data probability laws P for records $(U, T, X, Z, Y(0), Y(1), Y)$ satisfying:*

- ***(Record space.)*** *$U \in \{1, \dots, k\}$, $T \in \{0, 1\}$, $X \in \mathbb{R}^{d_x}$, $Z \in \mathbb{R}^{d_z}$, and $Y(0), Y(1), Y \in \mathbb{R}$.*
- ***(Parameter domain.)*** *The dimensions and constants obey*

$$2 \le k, \qquad k \le d_x, \qquad k \le d_z, \qquad 1 \le L, \qquad 0 < \pi_0 \le \frac{1}{2k}, \qquad 0 < \sigma_0 \le 1.$$

- ***(Model restrictions.)*** *P satisfies Assumptions 1 to 10.*

Proposition 2 supplies the compact population summary domain. Compactness supports nearest-summary selection in the next section, and the equivalence between nonempty $\mathcal{K}$ and nonempty $\mathcal{M}$ makes the domain statement exact for each fixed collection of supplied constants.

Several auxiliary definitions used by the empirical structured-lattice estimator and confidence reports are collected here because they share the same observable summaries, norms, and atomic-law coordinates.

Definition 33. [`def:observable-summary`] (Observable summary $S(P)$ and metric d_S) *The observable conditional moment summary is*

$$S(P) := \big(M_0(P), M_1(P), N_0(P), N_1(P), m_X(P)\big).$$

For summaries s and s', the summary metric is

$$d_S(s, s') := \sum_{j=1}^{4} \|s_j - s'_j\|_{\mathrm{op}} + \|s_5 - s'_5\|_2.$$

Definition 34. [`def:empirical-summary-primitives`] (Empirical summary primitives $\widehat{S}_n$) *For an observed sample of size n, let $N_{t,n}$ be the number of observations in arm t. Define the empirical armwise proxy moment $\widehat{M}_{t,n}$ and outcome-proxy moment $\widehat{N}_{t,n}$ using the total empty-arm convention:*

$$\widehat{M}_{t,n} = \frac{1}{\max\{1, N_{t,n}\}} \sum_{i:T_i=t} Z_i X_i^\top, \qquad \widehat{N}_{t,n} = \frac{1}{\max\{1, N_{t,n}\}} \sum_{i:T_i=t} Y_i Z_i X_i^\top.$$

Let the empirical target-proxy mean be

$$\widehat{m}_{X,n} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

The empirical five-block observable summary is

$$\widehat{S}_n = \big(\widehat{M}_{0,n}, \widehat{M}_{1,n}, \widehat{N}_{0,n}, \widehat{N}_{1,n}, \widehat{m}_{X,n}\big).$$

Definition 34 records the coordinate formulas. The theorem-facing empirical-summary object below packages the same coordinates with the total empty-arm convention and Borel measurability used in sample-level statements.

Definition 35. [`def:empirical-summary`] (Empirical summary $\widehat{S}_n$) *For each $t \in \{0, 1\}$, let*

$$N_{t,n} := \sum_{i=1}^{n} \mathbf{1}\{T_i = t\}.$$

Define

$$\widehat{M}_{t,n} := \frac{\sum_{i=1}^{n} \mathbf{1}\{T_i = t\} Z_i X_i^\top}{\max\{1, N_{t,n}\}}, \qquad \widehat{N}_{t,n} := \frac{\sum_{i=1}^{n} \mathbf{1}\{T_i = t\} Y_i Z_i X_i^\top}{\max\{1, N_{t,n}\}},$$

and

$$\widehat{m}_{X,n} := \frac{1}{n} \sum_{i=1}^{n} X_i.$$

The empirical five-block observable summary is

$$\widehat{S}_n := \big(\widehat{M}_{0,n}, \widehat{M}_{1,n}, \widehat{N}_{0,n}, \widehat{N}_{1,n}, \widehat{m}_{X,n}\big).$$

On the event $\{N_{t,n} = 0\}$, the corresponding armwise matrices $\widehat{M}_{t,n}$ and $\widehat{N}_{t,n}$ are the zero matrix, so $\widehat{S}_n$ is a total Borel function of the sample.

Definition 36. [`def:first-basis`] (First basis vector e_1) *Let $e_1 \in \mathbb{R}^{d_x}$ be the first canonical basis vector.*

Definition 37. [`synth_1`] (Ambient Euclidean dimension) *Let $d \in \mathbb{N}$ denote the finite ambient dimension of a Euclidean summary space $\mathbb{R}^d$. In the nearest-summary selection statement, $\mathcal{K} \subseteq \mathbb{R}^d$ is a compact feasible set and $d_S : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is the continuous summary distance used to choose nearest feasible summaries.*

Definition 38. [`synth_25`] (Vertical matrix stacking) *For matrices $A \in \mathbb{R}^{r\times d}$ and $B \in \mathbb{R}^{s\times d}$ with the same number of columns, define* $\mathrm{stack}(A, B) \in \mathbb{R}^{(r+s)\times d}$ *by vertical concatenation,*

$$\mathrm{stack}(A, B) := \begin{bmatrix} A \\ B \end{bmatrix}.$$

Its row space is regarded as a subspace of $\mathbb{R}^d$.

Definition 39. [`def:compressed-operator`] (Compressed operator $\Delta Q(P)$) *Choose any matrix $V(P) \in \mathbb{R}^{d_x\times k}$ with orthonormal columns spanning the row space of the vertically stacked matrix* $\mathrm{stack}(M_0(P), M_1(P)) \subseteq \mathbb{R}^{d_x}$. *Define*

$$\Delta Q(P) := \{M_1(P)V(P)\}^{\dagger} N_1(P)V(P) - \{M_0(P)V(P)\}^{\dagger} N_0(P)V(P),$$

with spectral anchors

$$a(P)^{\top} := m_X(P)^{\top} V(P), \qquad c(P) := V(P)^{\top} e_1.$$

Definition 40. [`def:quotient-law`] (Quotient law ν_P) *For $P \in \mathcal{M}$ as in Definition 32, the quotient latent-effect law is*

$$\nu_P := \sum_{u=1}^{k} p_u \delta_{\tau_u}.$$

It is interpreted as a probability measure on the latent treatment effects, with coincident values of τ_u carrying their aggregate mass.

Here and throughout, the latent treatment effects in ν_P are the class-average contrasts $\tau_u = E[Y(1) \mid U = u] - E[Y(0) \mid U = u]$. Coincident class-average contrasts appear as a single support value with aggregate population mass.

Definition 41. [`def:gap-stratum`] (Gap-local stratum) *For a local effect-gap scale g satisfying $0 < g \le 1/4$, the gap-local stratum $\mathcal{M}(g)$ is the class of full-data laws P in the uniformly conditioned proxy causal model class $\mathcal{M}$ of Definition 32 that satisfy Assumptions 11 and 12.*

Definition 42. [`def:sample-experiment`] (Sample experiment $\mathcal{E}_n$) *The observed i.i.d. proxy experiment at sample size n is*

$$\mathcal{E}_n := \{Q_P^{(n)} = P_O^{\otimes n} : P \in \mathcal{M}\}.$$

Definition 43. [`def:summary-closure`] (Summary closure data) *For $k, d_x, d_z \in \mathbb{N}$ and $L, \pi_0, \sigma_0 \in \mathbb{R}$, the summary-closure data consists of the closed feasible-summary set*

$$\mathcal{K} := \overline{\mathscr{S}}^{d_S}, \qquad \mathscr{S} := S(\mathcal{M}) = \{S(P) : P \in \mathcal{M}\},$$

together with a quotient-law functional

$$F : \mathscr{S} \to \mathcal{P}_k([-L_\tau, L_\tau]).$$

For each admissible summary $q \in \mathscr{S}$, $F(q)$ is the quotient latent-effect law ν_{P_q} of a selected model representative $P_q \in \mathcal{M}$ satisfying $S(P_q) = q$. The data also records the published spectral moment identity

$$\int x^j \, d\nu_{P_q}(x) = a(P_q)^\top \Delta Q(P_q)^j c(P_q), \qquad j = 0, \ldots, 2k-1,$$

and, in the homogeneous-effect specialization with common effect $\tau_\star$,

$$\Delta Q(P_q) = \tau_\star I_k, \qquad \nu_{P_q} = \delta_{\tau_\star}.$$

The radius $r_{n,\alpha}$ depends on the miscoverage level α and concentration constant C_0; it is used for the summary event E_P in the confidence results.

Proposition 1. [`prop:observed-vmw-margin-inclusion`] (Observed VMW margins)[1] *Fix $k, d_x, d_z \in \mathbb{N}$ and $L, \pi_0, \sigma_0 \in \mathbb{R}$. Suppose*

- ***(Dimensions.)*** *$2 \le k$, $k \le d_x$, and $k \le d_z$.*
- ***(Margins.)*** *$1 \le L$, $0 < \pi_0 \le 1/(2k)$, and $0 < \sigma_0 \le 1$.*
- ***(Uniform model class.)*** *The full-data law P belongs to the uniformly conditioned proxy causal model class $\mathcal{M}$ in Definition 32 with parameters $k, d_x, d_z, L, \pi_0, \sigma_0$.*

Then the following quantitative observed-margin conclusions hold. For each treatment arm $t \in \{0, 1\}$,

$$P(T = t) \ge k\pi_0,$$

and the observable armwise $ZX^\top$ moment factors as

$$M_t(P) = A_t \operatorname{diag}\{P(U = u \mid T = t) : 1 \le u \le k\} B^\top.$$

Moreover, for every orthonormal signal basis $V \in \mathbb{R}^{d_x \times k}$ spanning the signal space of $S(P)$,

$$\sigma_k\{M_t(P)V\} = \sigma_k\{M_t(P)\}, \qquad \pi_0\sigma_0^2 \le \sigma_k\{M_t(P)V\}.$$

The stacked observed proxy moment satisfies

$$\pi_0\sigma_0^2 \le \sigma_k(\operatorname{stack}(M_0(P), M_1(P)))\,.$$

The observable envelopes are

$$\|M_t(P)\|_{\mathrm{op}} \le L, \qquad \|N_t(P)\|_{\mathrm{op}} \le L, \qquad \|m_X(P)\|_2 \le L.$$

Writing $L_\tau = 4L\sqrt{d_z}/\sigma_0$, *for every latent class* $u \in \{1,\dots,k\}$ *and treatment arm* $t \in \{0,1\}$,

$$|Y(t)| \le L_\tau/2 \quad \text{for } P\text{-almost every full-data record conditional on } U = u,$$

and

$$|\mu_{tu}| \le L_\tau/2, \qquad |\tau_u| \le L_\tau.$$

Finally, P *belongs to the published VMW parameter class represented by the supplied scope handle.*

Proposition 2. [`prop:summary-closure-compact`] (Compact summary closure) *Let* $k, d_x, d_z \in \mathbb{N}$ *and* $L, \pi_0, \sigma_0 \in \mathbb{R}$. *Let* $\mathcal{M} = \mathcal{M}(k, d_x, d_z, L, \pi_0, \sigma_0)$ *be the class of full-data probability laws satisfying the uniformly conditioned proxy causal model restrictions with boundedness radius* L*, latent-arm positivity margin* π_0*, and proxy-rank margin* σ_0*. Define the admissible summary image and its closure by*

$$\mathscr{S} := \{S(P) : P \in \mathcal{M}\}, \qquad \mathcal{K} := \overline{\mathscr{S}}.$$

For summaries $s = (M_0, M_1, N_0, N_1, m_X)$ *and* $q = (M_0', M_1', N_0', N_1', m_X')$*, write*

$$d_S(s,q) = \|M_0 - M_0'\|_{\mathrm{op}} + \|M_1 - M_1'\|_{\mathrm{op}} + \|N_0 - N_0'\|_{\mathrm{op}} + \|N_1 - N_1'\|_{\mathrm{op}} + \Big(\sum_i (m_X(i) - m_X'(i))^2\Big)^{1/2}.$$

Assume:

- ***(Latent cardinality.)*** $2 \le k$.
- ***(Proxy dimensions.)*** $k \le d_x$ *and* $k \le d_z$.
- ***(Boundedness radius.)*** $1 \le L$.
- ***(Latent-arm margin.)*** $0 < \pi_0 \le 1/(2k)$.
- ***(Proxy-rank margin.)*** $0 < \sigma_0 \le 1$.

Then there exists $B \in \mathbb{R}$ *such that*

$$d_S(s, 0) \le B$$

for every $s \in \mathscr{S}$*. Moreover,* $\mathcal{K}$ *is compact, and*

$$\mathcal{K} \ne \varnothing \iff \mathcal{M} \ne \varnothing.$$

Together, the definitions and propositions in this section identify the population object, the observable summary space, and the compact domain on which the next section states the gap-free W_1 stability and estimation results.

[1] **Formalization scope.** This result uses the published conclusion [Virk et al., 2026] (Assumptions 1–2 and Section 4.2 of arXiv:2607.10926v1). 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.

4 Main results: law stability and estimation

The preceding section reduced the proxy causal model to the observable summary $S(P)$ and the quotient law ν_P of latent-class average treatment effects. This section gives the population stability statement that makes the quotient law a regular law-level target, then turns the modulus into two estimators: a nearest-summary repair estimator defined on the compact summary closure and a finite structured-lattice estimator evaluated from the empirical summary with supplied $k, d_x, d_z, L, \pi_0, \sigma_0$. The constants below depend only on those fixed dimensions and conditioning constants.

Definition 44. [`def:concentration-constant`] (Simultaneous-concentration constant C_0) *Fix a simultaneous-concentration constant*

$$C_0 \geq 1.$$

Definition 45. [`def:summary-radius`] (Summary radius $r_{n,\alpha}$) *For $n \geq 1$ and $0 < \alpha < 1/2$, the simultaneous concentration radius is*

$$r_{n,\alpha} := C_0 L \sqrt{\frac{\log(C_0/\alpha)}{n}}.$$

The constant C_0 calibrates a uniform finite-sample concentration inequality for all five coordinates of $\widehat{S}_n$. The radius $r_{n,\alpha}$ is the corresponding d_S-radius at confidence level $1-\alpha$, with the model envelope L scaling the observable moment fluctuations.

Definition 46. [`synth_27`] (Ambient summary space) *For fixed proxy dimensions d_x and d_z, the ambient five-block observable-summary space is*

$$\mathscr{S}_{\mathrm{all}} := (\mathbb{R}^{d_z \times d_x})^4 \times \mathbb{R}^{d_x}.$$

A point $s \in \mathscr{S}_{\mathrm{all}}$ is written $s = (s_1, s_2, s_3, s_4, s_5)$, with four matrix blocks and one target-proxy mean block, and the summary metric is

$$d_S(s, s') := \sum_{j=1}^{4} \|s_j - s'_j\|_{\mathrm{op}} + \|s_5 - s'_5\|_2.$$

The admissible image $\mathscr{S}$ and its closure $\mathcal{K} = \overline{\mathscr{S}}$ are subsets of $\mathscr{S}_{\mathrm{all}}$, and the repair selector Π is a total Borel map from $\mathscr{S}_{\mathrm{all}}$ to $\mathscr{S}_{\mathrm{all}}$.

The ambient space in Definition 46 lets the empirical summary be evaluated for every sample, while the feasible population summaries remain restricted to the compact closure $\mathcal{K}$ from Proposition 2. The repair construction projects an empirical summary back to that population domain before applying the quotient-law functional.

Algorithm 1 (Summary-space repair map)

Fix $k, d_x, d_z, n \in \mathbb{N}$ and $L, \pi_0, \sigma_0 \in \mathbb{R}$. Let

$$L_\tau = \frac{4L\sqrt{d_z}}{\sigma_0}, \qquad \mathcal{K} = \overline{\mathscr{S}},$$

where $\mathscr{S}$ is the admissible image in the five-block summary space. A summary-space repair datum consists of $k > 0$, $L_\tau \geq 0$, and maps

$$\overline{F} : \mathcal{K} \to \mathcal{P}_k([-L_\tau, L_\tau]), \qquad \Pi : \mathscr{S}_{\text{all}} \to \mathscr{S}_{\text{all}},$$

with the following certified properties:

- **(Extension.)** The map $\overline{F}$ is continuous and, for every full-data law P in the uniformly conditioned model class, sends its observable summary $S(P) \in \mathcal{K}$ to the quotient latent-effect law ν_P.
- **(Nearest selector.)** The map Π is Borel measurable and, whenever $\mathcal{K} \neq \varnothing$,

$$\Pi(s) \in \mathcal{K}, \qquad d_S(\Pi(s), s) \leq d_S(q, s) \quad \text{for every } q \in \mathcal{K}.$$

- **(Fallback.)** Whenever $\mathcal{K} = \varnothing$, the selector satisfies

$$\Pi(s) = 0$$

  in the ambient five-block summary space for every s.
- **(Repair measurability.)** The sample-to-law map defined below is Borel measurable.

For an observed sample $\omega : \{0, \ldots, n-1\} \to O_{d_x, d_z}$ with empirical five-block summary $\widehat{S}_n(\omega)$, the repaired estimator is the total map

$$\widehat{\nu}_n(\omega) = \begin{cases} \delta_0, & \mathcal{K} = \varnothing, \\ \overline{F}\big(\Pi(\widehat{S}_n(\omega))\big), & \mathcal{K} \neq \varnothing. \end{cases}$$

The estimator in Algorithm 1 is the repaired-summary route: it expresses the quotient-law estimator as a continuous extension evaluated at the nearest feasible summary. This directly links observable-summary uncertainty to W_1 uncertainty for the law.

The next construction replaces nearest-summary projection with a finite exhaustive lattice over the spectral factors that determine the quotient law.

Definition 47. [`def:structured-lattice-constant`] (Structured-lattice constant C_{lat}) *The structured-lattice constant is*

$$C_{\text{lat}} = C_{\text{lat}}(k, d_x, d_z, L, \pi_0, \sigma_0),$$

the explicit prescribed constant given by the maximum of the threshold-stability constants and the grid-approximation constants displayed in the structured-lattice construction.

Algorithm 2 (Structured lattice program)

For integers k, d_x, d_z, n, constants L, π_0, σ_0, and

$$L_\tau := 4L\sqrt{d_z}/\sigma_0,$$

a structured-lattice estimator A at radius L_τ is prescribed when, writing N_A for the candidate count recorded in A, there exist candidates

$$\vartheta_i = (V_i, R_i, p_i, \tau_i), \qquad i \in \{1, \dots, N_A\},$$

and, for every sample $\omega = (O_1, \dots, O_n)$ of records in O_{d_x,d_z}, a selected index

$$i_\omega \in \{1, \dots, N_A\},$$

such that:

- **(Well-formed exhaustive list.)** Each listed candidate ϑ_i is well formed for the parameters (n, L, π_0, σ_0). Conversely, every well-formed structured lattice point $\vartheta = (V, R, p, \tau)$ at radius L_τ appears at some listed index i with

$$V_i = V, \qquad R_i = R, \qquad p_i = p, \qquad \tau_i = \tau.$$

- **(No duplicates.)** If two listed candidates agree in all four displayed coordinates,

$$V_i = V_j, \qquad R_i = R_j, \qquad p_i = p_j, \qquad \tau_i = \tau_j,$$

  then $i = j$.

- **(Lexicographic order.)** The index order agrees exactly with the lexicographic order on structured lattice points:

$$i \le j \quad \Longleftrightarrow \quad \vartheta_i \le_{\text{lex}} \vartheta_j.$$

- **(First minimizer.)** For the empirical summary $s = \widehat{S}_n(\omega)$, the selected index minimizes the criterion J_n computed with threshold $\pi_0\sigma_0^2/2$:

$$J_n(\vartheta_{i_\omega}; \widehat{S}_n(\omega)) \le J_n(\vartheta_i; \widehat{S}_n(\omega)) \qquad \text{for every } i \in \{1, \dots, N_A\}.$$

  Ties are resolved by the first lexicographic index:

$$J_n(\vartheta_{i_\omega}; \widehat{S}_n(\omega)) = J_n(\vartheta_i; \widehat{S}_n(\omega)) \quad \Longrightarrow \quad i_\omega \le i.$$

- **(Returned law.)** On every sample ω, the estimator map recorded in A returns the atomic effect law attached to the selected candidate:

$$\widehat{\lambda}_n(\omega) = \lambda_{\vartheta_{i_\omega}}.$$

The lattice program in Algorithm 2 searches over discretized signal bases, conditioning matrices, latent masses, and effect locations. Its criterion matches the empirical compressed operator, target-proxy mean, and anchor equation, so the returned law is tied to the same observable structure as the population quotient law.

Lemma 1. [`lem:uniform-summary-concentration`] (Uniform summary concentration) *Let $k, d_x, d_z \in \mathbb{N}$, and let $\pi_0, \sigma_0 \in \mathbb{R}$. Suppose that*

- ***(Dimensions.)*** $2 \le k$, $k \le d_x$, *and* $k \le d_z$.
- ***(Positivity calibration.)*** $0 < \pi_0$ *and*
$$\pi_0 \le \frac{1}{2k}.$$
- ***(Rank-margin calibration.)*** $0 < \sigma_0$ *and* $\sigma_0 \le 1$.

Then there exists a simultaneous-concentration constant $C_0 \in \mathbb{R}$ *with* $1 \le C_0$ *such that the following holds. For every* $L \in \mathbb{R}$ *and* $n \in \mathbb{N}$ *with* $1 \le L$ *and* $1 \le n$, *every full-data probability law* P *on records* $(U, T, X, Z, Y(0), Y(1), Y) \in \{1, \dots, k\} \times \{0, 1\} \times \mathbb{R}^{d_x} \times \mathbb{R}^{d_z} \times \mathbb{R}^3$ *in the uniformly conditioned proxy causal model class* $\mathcal{M}$ *with radius* L, *latent-arm positivity margin* π_0, *and proxy-rank singular-value margin* σ_0, *and every tail probability* η *satisfying* $0 < \eta < 1/2$,

$$Q_P^{(n)} \left\{ d_S(\widehat{S}_n, S(P)) > C_0 L \sqrt{\frac{\log(C_0/\eta)}{n}} \right\} \le \eta.$$

Moreover, for each treatment arm $t \in \{0, 1\}$,

$$Q_P^{(n)}\{N_{t,n} = 0\} = (1 - P_O\{O : T = t\})^n \le (1 - k\pi_0)^n.$$

The proof is deferred to Section D.

Lemma 1 is the sampling input for the law estimators. It gives a single high-probability radius for the empirical five-block summary and also bounds the probability of an empty treatment arm under the total empirical-summary convention.

Theorem 1. [`thm:gap-free-positive-measure-modulus`] (Gap-free modulus extension) *Fix natural numbers* k, d_x, d_z *and real constants* L, π_0, σ_0. *Assume:*

- ***(Latent dimension.)*** $2 \le k$, $k \le d_x$, *and* $k \le d_z$.
- ***(Bound and margins.)*** $1 \le L$, $0 < \pi_0 \le (2k)^{-1}$, *and* $0 < \sigma_0 \le 1$.

Let $L_\tau = 4L\sqrt{d_z}/\sigma_0$, *and let* $\mathcal{K} = \overline{\mathscr{S}}$ *be the summary closure from Definition 43. For two summaries*

$$s = (M_0, M_1, N_0, N_1, m_X), \qquad q = (M_0', M_1', N_0', N_1', m_X'),$$

write

$$d_S(s, q) = \|M_0 - M_0'\| + \|M_1 - M_1'\| + \|N_0 - N_0'\| + \|N_1 - N_1'\| + \Big(\sum_i (m_X(i) - m_X'(i))^2\Big)^{1/2}.$$

There exists a constant $C_{\mathrm{mod}} > 0$ *such that, for every pair of full-data laws* $P, Q \in \mathcal{M}$,

$$W_1(\nu_P, \nu_Q) \le C_{\mathrm{mod}}\, d_S(S(P), S(Q)).$$

Moreover, there is a map

$$\overline{F} : \mathcal{K} \to \mathcal{P}_k([-L_\tau, L_\tau])$$

such that $\overline{F}$ *is continuous and Borel measurable,*

$$W_1(\overline{F}(q), \overline{F}(q')) \le C_{\mathrm{mod}}\, d_S(q, q') \qquad \textit{for all } q, q' \in \mathcal{K},$$

and
$$\overline{F}(S(Q)) = \nu_Q \qquad \textit{for every } Q \in \mathcal{M}.$$
This map is unique among continuous maps $G : \mathcal{K} \to \mathcal{P}_k([-L_\tau, L_\tau])$ *that obey*
$$W_1(G(q), G(q')) \le C_{\mathrm{mod}}\, d_S(q, q') \qquad \textit{for all } q, q' \in \mathcal{K}$$
and satisfy $G(S(Q)) = \nu_Q$ *for every* $Q \in \mathcal{M}$*: every such* G *satisfies* $G(q) = \overline{F}(q)$ *for all* $q \in \mathcal{K}$*.*

Theorem 1 is the population result behind the statistical theory. It establishes a Lipschitz W_1 modulus from observable summaries to quotient effect laws over the uniformly conditioned model class, and it extends the summary-to-law map continuously from the admissible image to the compact closure. The result is stated for the law that aggregates coincident effect values, so the same bound covers homogeneous, partially colliding, and separated latent-effect configurations.

The intuition is spectral but the conclusion is metric. The proxy moments determine a compressed effect operator through the VMW moment identity [Virk et al., 2026]; perturbation control for finite-dimensional operators [Bauer and Fike, 1960, Kato, 1995] gives stable spectral moments; and the quotient law is compared in W_1, a metric that transports aggregate probability mass between effect values. The continuous extension $\overline{F}$ then makes feasible-summary projection a valid route from empirical summaries to law estimates.

Proposition 3. [`prop:summary-repair-total-borel`] (Total Borel repair) *Let* $k, d_x, d_z, n \in \mathbb{N}$ *and* $L, \pi_0, \sigma_0, \alpha, C_0 \in \mathbb{R}$*. Suppose that:*

- ***(Dimensions.)*** $2 \le k$*,* $k \le d_x$*, and* $k \le d_z$*.*
- ***(Model radii.)*** $1 \le L$*,* $0 < \pi_0 \le 1/(2k)$*, and* $0 < \sigma_0 \le 1$*.*
- ***(Sample size.)*** $1 \le n$*.*
- ***(Confidence parameters.)*** *The miscoverage level satisfies* $0 < \alpha < 1/2$*, and the simultaneous-concentration constant satisfies* $1 \le C_0$*.*
- ***(Empirical summary.)*** *The empirical summary map* $\widehat{S}_n$ *of Definition 35 is measurable.*

Then there exists repair data R *of the form specified in Algorithm 1 such that the estimator* $\widehat{\nu}_n$ *induced by* R *is Borel measurable. Moreover, writing* $\mathcal{K}$ *for the* d_S*-closure of the admissible summary image, if* $\mathcal{K} = \varnothing$*, then for every sample* ω*,*
$$\widehat{\nu}_n(\omega) = \delta_0.$$
For every sample ω*, the sharp summary-inversion confidence set* $\mathcal{C}_{n,\alpha}(\omega)$ *associated with* R*, with singleton value* $\{\delta_0\}$ *when* $\mathcal{K} = \varnothing$ *and otherwise consisting of the laws* $\nu = \overline{F}(q)$ *with* $q \in \mathcal{K}$ *and* $d_S(q, \Pi(\widehat{S}_n(\omega))) \le 2r_{n,\alpha}$*, is nonempty.*

Proposition 3 supplies the measurable total version of the nearest-summary construction. It also records nonemptiness of the exact repaired-image summary-inversion confidence set for every sample, preparing the confidence-set construction in the next section.

The structured-lattice estimator uses the same observable ingredients and replaces projection over the compact summary closure by a finite lattice search. The work count below fixes the computational scale certified here: a fixed-dimensional unit-cost exact-real benchmark, with arithmetic, comparison, singular-value, polar-decomposition, and fixed-degree root-isolation operations counted as unit primitives.

Definition 48. [synth_16] (Structured-lattice work count) *For the explicit structured-lattice estimator $\widehat{\lambda}_n$, let* Work_n *denote the fixed-dimensional exact-real operation count for forming $\widehat{S}_n$, enumerating the lexicographically ordered candidate lattice, evaluating the criterion J_n, and returning the first minimizer $\widehat{\vartheta}_n$. In the fixed-dimensional unit-cost exact-real model,* Work_n *counts arithmetic, comparison, singular-value, and fixed-degree root-isolation operations and satisfies*

$$\mathrm{Work}_n \le C\left\{n + n^{(d_x k + k^2 + 2k - 1)/2}\right\},$$

for a finite constant $C = C(k, d_x, d_z, L, \pi_0, \sigma_0)$.

Proposition 4. [prop:polynomial-net-law-estimator] (Polynomial lattice estimator) *Fix integers k, d_x, d_z and constants L, π_0, σ_0. Suppose that*

- ***(Dimension.)*** $2 \le k$, $k \le d_x$, *and* $k \le d_z$.
- ***(Radius and margins.)*** $1 \le L$, $0 < \pi_0 \le (2k)^{-1}$, *and* $0 < \sigma_0 \le 1$.

Put $L_\tau := 4L\sqrt{d_z}/\sigma_0$ and $s_0 := \pi_0 \sigma_0^2$. Define

$$c_V := 4\sqrt{d_x k}, \qquad K_D := \frac{4\sqrt{k} L L_\tau}{\sigma_0}, \qquad A_D := \frac{8}{3s_0} + \frac{32L}{s_0^2},$$

$$K_f := \frac{16\sqrt{d_x}\, k L^2}{\sigma_0^2}, \qquad c_D := 2K_D c_V + \frac{4k\sqrt{k} L L_\tau}{\sigma_0^2} + \frac{2\sqrt{k} L}{\sigma_0} + \frac{k L_\tau}{\sigma_0},$$

$$c_m := 2\sqrt{k} L c_V + k + kL, \qquad c_b := k + 2\sqrt{k} L c_V, \qquad c_{\mathrm{grid}} := c_D + c_m + c_b,$$

$$B_{\mathrm{lat}} := 2(A_D + 1) + c_{\mathrm{grid}}, \qquad C_{\mathrm{lat}} := \max\left\{\frac{8L_\tau}{s_0}, (L_\tau + K_D + L K_f) B_{\mathrm{lat}}\right\}.$$

Then $C_{\mathrm{lat}} > 0$, and there is a finite constant $C = C(k, d_x, d_z, L, \pi_0, \sigma_0) > 0$ such that, for every $n \ge 1$, there is a total Borel summary rule whose empirical-summary estimator $\widehat{\lambda}_n$ is the lexicographically first minimizer over the prescribed structured lattice with effect radius L_τ, atom floor π_0, and threshold $s_0/2$. For every sample, every distinct represented atom of $\widehat{\lambda}_n$ has aggregate mass at least π_0.

For every $P \in \mathcal{M}$ in Definition 32, the stacked observable proxy moment at $S(P)$, and at every summary s with

$$d_S(s, S(P)) < \frac{\pi_0 \sigma_0^2}{2},$$

has exactly its first k singular values at least $\pi_0 \sigma_0^2/2$ and all remaining singular values below $\pi_0 \sigma_0^2/2$. The deterministic loss bound

$$W_1(\widehat{\lambda}_n(\omega), \nu_P) \le C_{\mathrm{lat}}\left\{d_S(\widehat{S}_n(\omega), S(P)) + \frac{1}{\sqrt{n}}\right\}$$

holds for every sample ω, where $\widehat{S}_n$ is the empirical summary in Definition 35. The candidate list and exhaustive-search operation count satisfy

$$|\mathcal{A}_n| \le C\, n^{(d_x k + k^2 + 2k - 1)/2}, \qquad \mathrm{Work}_n \le C\left(n + n^{(d_x k + k^2 + 2k - 1)/2}\right).$$

Finally, for every $0 < \eta < 1/2$, every $P \in \mathcal{M}$, and the observed product experiment $Q_P^{(n)}$,

$$Q_P^{(n)}\left\{W_1(\widehat{\lambda}_n, \nu_P) > C\sqrt{\frac{\log(C/\eta)}{n}}\right\} \le \eta.$$

Proposition 4 establishes the empirical structured-lattice law estimator used for the algorithmic confidence set and cluster report. Its deterministic inequality converts summary error and lattice mesh width into W_1 error, while its high-probability conclusion follows by combining that inequality with the summary concentration in Lemma 1. The candidate bound records polynomial growth in n for fixed $k, d_x, d_z, L, \pi_0, \sigma_0$ in the exact-real operation model stated above.

The role of the lattice is to preserve the stabilized spectral structure at empirical summaries near the model image. The singular-value threshold keeps the compressed operator in the uniformly conditioned rank-k regime; the grids over V, R, p, and τ approximate the population spectral representation; and lexicographic tie-breaking turns the exhaustive search into a total Borel rule.

Theorem 2. [`thm:collision-uniform-root-n`] (Collision-uniform root-n rate) *Fix integers k, d_x, d_z and constants L, π_0, σ_0. Suppose that*

- ***(Dimensions.)*** $2 \le k$, $k \le d_x$, *and* $k \le d_z$.
- ***(Uniform bounds.)*** $1 \le L$, $0 < \pi_0 \le 1/(2k)$, *and* $0 < \sigma_0 \le 1$.

Then there is a constant $C = C(k, d_x, d_z, L, \pi_0, \sigma_0) > 0$ such that, for every $n \ge 1$, there exist a structured-lattice estimator $\widehat{\lambda}_n$ taking values in $\mathcal{P}_k([-L_\tau, L_\tau])$, with

$$L_\tau = 4L\sqrt{d_z}/\sigma_0,$$

and a nearest-summary repair estimator $\widehat{\nu}_n = \overline{F}(\Pi(\widehat{S}_n))$ from Algorithm 1, such that $\widehat{\lambda}_n$ is the first lexicographic minimizer over a finite exhaustive structured lattice of well-formed candidates ϑ for the criterion

$$J_n(\vartheta; \widehat{S}_n),$$

with tie-breaking by the lattice order and output λ_ϑ. Both estimators are Borel sample maps. For every η satisfying $0 < \eta < 1/2$, every $P \in \mathcal{M}$, and $Q_P^{(n)}$-sample,

$$Q_P^{(n)}\left\{\max\Big\{W_1(\widehat{\lambda}_n, \nu_P), W_1(\widehat{\nu}_n, \nu_P)\Big\} > C\sqrt{\frac{\log(C/\eta)}{n}}\right\} \le \eta.$$

Moreover, for each treatment arm $t \in \{0, 1\}$,

$$Q_P^{(n)}\{N_{t,n} = 0\} \le (1 - k\pi_0)^n.$$

Theorem 2 is the law-estimation conclusion. Under the fixed dimension, envelope, positivity, and rank-margin conditions defining the uniformly conditioned class, both the nearest-summary repair estimator and the structured-lattice estimator achieve the same root-n W_1 tail rate uniformly over $\mathcal{M}$. The statement also carries forward the armwise empty-sample bound, ensuring that the total empirical-summary convention is accompanied by an explicit probability control.

The theorem separates the regular law target from label-sensitive quantities. At the level of ν_P, coincident latent-class mean effects are represented by aggregate mass at a shared support point, and the modulus in Theorem 1 transfers root-n summary concentration directly to W_1 law error. The next section uses the structured-lattice estimator as the center of exact-real finite-constraint confidence reporting and then treats ordered weights on the gap-local stratum, where effect separation determines the relevant local scale.

5 Confidence sets, cluster reports, and ordered weights

The root-n law estimators from Theorem 2 lead directly to confidence reporting for the quotient law of latent-class average treatment effects. The first construction keeps two complementary objects in view: an exact repaired-image set of nearby feasible summaries through the repaired summary map, and a lattice-centered finitely represented exact-real W_1 outer set with an atom-floor restriction. The exact-real operation model is fixed-dimensional and unit-cost, with arithmetic, comparison, singular-value, polar-decomposition, and fixed-degree root-isolation primitives counted as the operations certified in Definition 48 and Proposition 4. This follows the econometric logic of confidence sets obtained by inverting sample restrictions [Robins and van der Vaart, 2006], with the reporting metric adapted to discrete laws as in Wasserstein approaches to mixture uncertainty [Deo and Randrianarisoa, 2023].

Algorithm 3 (Wasserstein confidence set $\mathcal{C}^{\mathrm{alg}}_{n,\alpha}$)

The inputs are fixed $k, d_x, d_z, n \in \mathbb{N}$, constants $L, \pi_0, \sigma_0, \alpha, C_0, C_{\mathrm{lat}} \in \mathbb{R}$, an observed sample $\omega \in O^n_{d_x,d_z}$ with empirical summary $\widehat{S}_n$, repair data $(\overline{F}, \Pi)$ on $\mathcal{K}$, and a prescribed structured-lattice estimator.

1. The fixed constants satisfy

$$0 < k, \qquad 0 < n, \qquad 1 \le L, \qquad 0 < \pi_0 \le (2k)^{-1}, \qquad 0 < \alpha < \frac{1}{2}, \qquad 1 \le C_0, \qquad 0 < C_{\mathrm{lat}}.$$

   Define the effect-support radius and assume its nonnegativity:

$$L_\tau := \frac{4L\sqrt{d_z}}{\sigma_0}, \qquad 0 \le L_\tau.$$

2. Define the simultaneous concentration radius by

$$r_{n,\alpha} := C_0 L \sqrt{\frac{\log(C_0/\alpha)}{n}}.$$

3. The repair data satisfy: $\overline{F}$ is continuous on $\mathcal{K}$ and extends the summary-to-quotient-law map on model summaries; Π is measurable; if $\mathcal{K} \neq \varnothing$, then

$$\Pi(s) \in \operatorname*{argmin}_{q \in \mathcal{K}} d_S(q, s);$$

   if $\mathcal{K} = \varnothing$, then

$$\Pi(s) = 0;$$

   and the induced repaired estimator is measurable.

4. Let $\widehat{\lambda}_n$ be the valid k-atomic law returned at ω by the measurable prescribed structured-lattice estimator at support radius L_τ, generated by an exhaustive duplicate-free lexicographically ordered list of well-formed H_n, q_n lattice candidates and the first lexicographic minimizer of $J_n(\vartheta; \widehat{S}_n)$ with threshold

$$\frac{\pi_0 \sigma_0^2}{2}.$$

5. Define the original summary-inversion confidence set by

$$\mathcal{C}_{n,\alpha} := \begin{cases} \{\overline{F}(q) : q \in \mathcal{K},\ d_S(q, \Pi(\widehat{S}_n)) \le 2r_{n,\alpha}\}, & \mathcal{K} \ne \varnothing, \\ \{\delta_0\}, & \mathcal{K} = \varnothing. \end{cases}$$

6. Define the atom-floor constant and computable radius by

$$m_\star := \pi_0, \qquad R_{n,\alpha} := C_{\mathrm{lat}} \left(r_{n,\alpha} + (\sqrt{n})^{-1} \right).$$

7. Define the atom-floor subclass by

$$\mathcal{P}_{\le k, m_\star}([-L_\tau, L_\tau]) := \{\xi \in \mathcal{P}_{\le k}([-L_\tau, L_\tau]) : \mathrm{AtomFloor}(m_\star, \xi)\}.$$

8. Output the computable Wasserstein confidence set

$$\mathcal{C}^{\mathrm{alg}}_{n,\alpha} := \{\xi \in \mathcal{P}_{\le k}([-L_\tau, L_\tau]) : \mathrm{AtomFloor}(m_\star, \xi),\ W_1(\xi, \widehat{\lambda}_n) \le R_{n,\alpha}\}.$$

Algorithm 3 defines the exact repaired-image summary-inversion set $\mathcal{C}_{n,\alpha}$ and the lattice-centered outer set $\mathcal{C}^{\mathrm{alg}}_{n,\alpha}$. The radius $R_{n,\alpha}$ combines the concentration scale $r_{n,\alpha}$ with the lattice mesh scale, while the atom floor $m_\star = \pi_0$ carries the latent-arm positivity margin into the represented-law class. The lattice-centered set is finitely represented by the empirical lattice center, the transport metric, and fixed finite-dimensional constraints in the exact-real model.

The computational guarantee is an exact-real finite-representation statement. In the fixed-dimensional unit-cost model used in Definition 48 and Proposition 4, the lattice search has polynomial candidate growth and exact evaluation over the stated primitive operations. The constrained representation of $\mathcal{C}^{\mathrm{alg}}_{n,\alpha}$ gives finite atom, mass, and transport-plan constraints for membership and endpoint formulations. The exact repaired-image set $\mathcal{C}_{n,\alpha}$ is represented through $\overline{F}$ on the compact repaired summary image, and Remark 1 records the polynomial-time exact-image problem for that original repaired set.

Algorithm 4 (Cluster report $\mathfrak{R}_{n,\alpha}$)

The cluster report is computed from $\widehat{S}_n$, $S(P)$, $r_{n,\alpha}$, $R_{n,\alpha}$, $m_\star$, $\widehat{\lambda}_n$, and $\mathcal{C}^{\mathrm{alg}}_{n,\alpha}$ as follows.

1. Define the simultaneous concentration event and association radius by

$$E_P := \{d_S(\widehat{S}_n, S(P)) \le r_{n,\alpha}\}, \qquad \rho_{n,\alpha} := \frac{R_{n,\alpha}}{m_\star}.$$

2. Merge repeated atoms of $\widehat{\lambda}_n$. Join two distinct support points whenever their distance is at most $4\rho_{n,\alpha}$, and define

$$\widehat{\mathscr{K}}_{n,\alpha}$$

as the set of connected components of the resulting graph.

3. For each $C \in \widehat{\mathscr{K}}_{n,\alpha}$, define the associated true and candidate support sets by

$$K_C(P) := \{x \in \operatorname{supp}(\nu_P) : \operatorname{dist}(x, C) \le \rho_{n,\alpha}\}, \qquad K_C(\xi) := \{x \in \operatorname{supp}(\xi) : \operatorname{dist}(x, C) \le \rho_{n,\alpha}\}.$$

4. Define the mass interval by

$$I_C := \left[\inf_{\xi \in \mathcal{C}^{\mathrm{alg}}_{n,\alpha}} \xi(K_C(\xi)),\ \sup_{\xi \in \mathcal{C}^{\mathrm{alg}}_{n,\alpha}} \xi(K_C(\xi)) \right].$$

In the ordered-support and transport-plan representation of $\mathcal{C}^{\mathrm{alg}}_{n,\alpha}$, each endpoint is the value of a finite union of fixed-dimensional mixed-integer constrained programs. For each $1 \le r \le k$, introduce binary variables

$$z_a \in \{0, 1\}, \qquad 1 \le a \le r,$$

with

$$z_a = 1 \quad \Longleftrightarrow \quad \min_{c \in C} |x_a - c| \le \rho_{n,\alpha},$$

and minimize or maximize

$$\sum_{a=1}^{r} w_a z_a$$

subject to the ordered-support, atom-floor, simplex, and transport constraints defining $\mathcal{C}^{\mathrm{alg}}_{n,\alpha}$.

5. Output the report

$$\mathfrak{R}_{n,\alpha} := \left\{ (C,\ [\min C - \rho_{n,\alpha}, \max C + \rho_{n,\alpha}],\ I_C) : C \in \widehat{\mathscr{K}}_{n,\alpha} \right\}.$$

For each component C, define the external true-support gap by

$$\Delta_C(P) := \inf\{|x - y| : x \in K_C(P),\ y \in \operatorname{supp}(\nu_P) \setminus K_C(P)\},$$

with value $+\infty$ when $\operatorname{supp}(\nu_P) \setminus K_C(P) = \emptyset$.

The cluster report has a sample-facing layer and a theorem-side validation layer. The sample-facing output is formed from the observed $\widehat{S}_n$, the supplied constants, the lattice center $\widehat{\lambda}_n$, and the constrained class $\mathcal{C}^{\mathrm{alg}}_{n,\alpha}$: it merges empirical support points at threshold $4\rho_{n,\alpha}$, expands each component by $\rho_{n,\alpha}$, and computes mass endpoints over the same constrained class. The true-law symbols $S(P)$, E_P, $K_C(P)$, and $\Delta_C(P)$ name the concentration event, associated true support block, and external separation scale used to state coverage and width.

Theorem 3 establishes simultaneous coverage for both confidence constructions over the uniformly conditioned model class. The statement also gives samplewise nonemptiness and W_1-diameter bounds, so the reported law-level uncertainty has root-n scale under the same constants that control the estimator in Proposition 4.

The mechanism is the combination of summary concentration and the gap-free modulus. On the event that $\widehat{S}_n$ lies within $r_{n,\alpha}$ of $S(P)$, the feasible-summary set includes a summary whose image is ν_P, and the structured-lattice center lies within the radius used to form $\mathcal{C}^{\mathrm{alg}}_{n,\alpha}$. The transport representation in the theorem gives a finite-dimensional constrained problem over atom locations, masses, and couplings.

Law-level confidence sets can be difficult to read directly when nearby support points are statistically unresolved. The next construction converts the W_1 set into a component report: estimated support points are connected at threshold $4\rho_{n,\alpha}$, and the mass attached to each empirical component is summarized by an interval over the same constrained confidence class.

Algorithm 5 (Ordered-weight estimator $\widehat{p}^{\uparrow}_n$)

The input is the positive estimated atomic law $\widehat{\nu}_n$ on $[-L_\tau, L_\tau]$.

1. Write $\widehat{\nu}_n$ uniquely in increasing support order as

$$\widehat{\nu}_n = \sum_{\ell=1}^{\widehat{s}_n} \widehat{w}_{\ell,n}\delta_{\widehat{\tau}_{\ell,n}}, \qquad \widehat{\tau}_{1,n} < \cdots < \widehat{\tau}_{\widehat{s}_n,n}, \qquad \widehat{w}_{\ell,n} > 0.$$

2. Output the ordered-weight estimator

$$\widehat{p}^{\uparrow}_n = \begin{cases} (\widehat{w}_{1,n}, \ldots, \widehat{w}_{k,n}), & \widehat{s}_n = k, \\ (1/k, \ldots, 1/k), & \widehat{s}_n < k. \end{cases}$$

Ordering the finitely many support points and extracting their jump sizes are Borel operations, so $\widehat{\nu}_n \mapsto \widehat{p}^{\uparrow}_n$ is a total Borel map into Δ^{k-1}.

The report $\mathfrak{R}_{n,\alpha}$ replaces individual estimated atoms by empirical components formed at connection threshold $4\rho_{n,\alpha}$. For each component C, the support interval expands the component by $\rho_{n,\alpha}$, and the mass interval I_C ranges over all laws in $\mathcal{C}^{\mathrm{alg}}_{n,\alpha}$ whose atoms can be transported to the lattice center within $R_{n,\alpha}$. The external gap $\Delta_C(P)$ is the separation scale relevant to the component's aggregate mass.

Theorem 3. [`thm:honest-root-n-confidence`] (Honest root-n confidence) *Fix integers k, d_x, d_z and constants π_0, σ_0 such that*

$$2 \le k, \qquad k \le d_x, \qquad k \le d_z, \qquad 0 < \pi_0 \le \frac{1}{2k}, \qquad 0 < \sigma_0 \le 1.$$

There exists a simultaneous-concentration constant $C_0 \ge 1$ such that, for every boundedness radius $L \ge 1$, there is a constant $C_{\mathrm{mod}} > 0$ for which the prescribed structured-lattice constant C_{lat} is positive and the following holds. For every sample size $n \ge 1$, there exist:

- ***(Structured lattice estimator.)*** *a measurable estimator $\widehat{\lambda}_n$ taking values in $\mathcal{P}_k([-L_\tau, L_\tau])$, generated by a finite exhaustive list of well-formed structured-lattice candidates, ordered lexicographically, with $\widehat{\lambda}_n$ equal to the effect law of the first lexicographic minimizer of the structured-lattice criterion;*

- ***(Summary repair data.)*** *a continuous extension* $\overline{F}$ *on the summary closure* $\mathcal{K}$*, a measurable nearest-summary selector* Π *with its empty-closure fallback, and the associated measurable repaired summary-to-law map.*

For every miscoverage level α *satisfying*

$$0 < \alpha < \frac{1}{2},$$

let $\mathcal{C}_{n,\alpha}$*,* $R_{n,\alpha}$*, and* $\mathcal{C}^{\mathrm{alg}}_{n,\alpha}$ *be the confidence-set objects of Algorithm 3 formed from this repair data and this structured-lattice estimator, with* $m_\star = \pi_0$ *and center* $\widehat{\lambda}_n$*. Then, on every sample, both* $\mathcal{C}_{n,\alpha}$ *and* $\mathcal{C}^{\mathrm{alg}}_{n,\alpha}$ *are nonempty, and* $\mathcal{C}^{\mathrm{alg}}_{n,\alpha}$ *has the finite constrained transport representation*

$$\nu \in \mathcal{C}^{\mathrm{alg}}_{n,\alpha} \quad \Longleftrightarrow \quad \mathrm{AtomFloor}(\pi_0, \nu) \text{ and there exists a transport plan } \gamma \text{ from } \nu \text{ to } \widehat{\lambda}_n \text{ with cost at most } R_{n,\alpha}.$$

Moreover, for every full-data probability law $P \in \mathcal{M}$ *in Definition 32,*

$$1 - \alpha \le Q_P^{(n)} \left\{ \nu_P \in \mathcal{C}_{n,\alpha} \cap \mathcal{C}^{\mathrm{alg}}_{n,\alpha} \right\}.$$

Finally, on every sample,

$$\mathrm{diam}_{W_1}(\mathcal{C}_{n,\alpha}) \le 4C_{\mathrm{mod}}\, r_{n,\alpha}, \qquad \mathrm{diam}_{W_1}(\mathcal{C}^{\mathrm{alg}}_{n,\alpha}) \le 2R_{n,\alpha}.$$

Theorem 4 establishes the coverage and width properties of the component report on the concentration event. Every true support point is assigned to a unique empirical component, the reported support interval contains the associated true atom, and the reported mass interval contains the true aggregate component mass. When a component is separated from the remaining true support by $\Delta_C(P)$, the mass-interval length contracts at the displayed inverse-gap scale.

The report is adaptive because transport cost controls how much mass can cross a gap. If estimated support points connected to nearby true atoms fall within the $4\rho_{n,\alpha}$ component threshold, the construction pools them and reports aggregate mass. When a component is externally separated, moving mass across that separation consumes W_1 budget, and the interval width reflects the ratio between the available radius and the external gap.

The final object in this section returns to effect-ordered latent weights. The quotient law treats coincident class-average effects as a single atom, while the ordered-weight target attaches masses to effect ranks on a gap-local stratum. The estimator below applies the repaired law construction and then orders the labelled representative when all labelled atoms are positive and distinct.

Theorem 4. [`thm:cluster-adaptive-report`] (Cluster-adaptive report validity) *Let* $k, d_x, d_z \in \mathbb{N}$*,* $\pi_0 \in \mathbb{R}$*, and* $\sigma_0 \in \mathbb{R}$*. Suppose that*

- ***(Dimensions.)*** $2 \le k$*,* $k \le d_x$*, and* $k \le d_z$*.*
- ***(Latent mass.)*** $0 < \pi_0 \le 1/(2k)$*.*
- ***(Rank margin.)*** $0 < \sigma_0 \le 1$*.*

Then there is a simultaneous-concentration constant $C_0 \ge 1$ *such that, for every* $L \ge 1$*, if*

$$L_\tau = \frac{4L\sqrt{d_z}}{\sigma_0},$$

and

$$s_0 = \pi_0\sigma_0^2, \qquad c_V = 4\sqrt{d_x k}, \qquad K_D = \frac{4\sqrt{k}LL_\tau}{\sigma_0},$$

$$A_D = \frac{8}{3s_0} + \frac{32L}{s_0^2}, \qquad K_f = \frac{16\sqrt{d_x}kL^2}{\sigma_0^2},$$

$$c_D = 2K_D c_V + \frac{4k\sqrt{k}LL_\tau}{\sigma_0^2} + \frac{2\sqrt{k}L}{\sigma_0} + \frac{kL_\tau}{\sigma_0},$$

$$c_m = 2\sqrt{k}Lc_V + k + kL, \qquad c_b = k + 2\sqrt{k}Lc_V,$$

$$c_{\text{grid}} = c_D + c_m + c_b, \qquad B_{\text{lat}} = 2(A_D + 1) + c_{\text{grid}},$$

then the prescribed structured-lattice stability constant

$$C_{\text{lat}} = \max\left\{\frac{8L_\tau}{s_0},\ (L_\tau + K_D + LK_f)B_{\text{lat}}\right\}$$

is positive. For every $n \geq 1$*, there exist a measurable structured-lattice estimator* $\widehat{\lambda}_n$ *and repair data consisting of a continuous extension on* $\mathcal{K}$ *and a Borel nearest-summary selector, such that* $\widehat{\lambda}_n$ *is obtained from an exhaustive lexicographically ordered finite list of well-formed structured-lattice candidates by choosing the first minimizer of* $J_n(\vartheta; \widehat{S}_n)$*.*

For every α *with* $0 < \alpha < 1/2$*, the estimator has atom floor* π_0*. Moreover, for every* $P \in \mathcal{M}$ *in Definition 32 and every sample,*

$$W_1\left(\widehat{\lambda}_n, \nu_P\right) \leq C_{\text{lat}}\left\{d_S\left(\widehat{S}_n, S(P)\right) + \frac{1}{\sqrt{n}}\right\}.$$

Define

$$r_{n,\alpha} = C_0 L\sqrt{\frac{\log(C_0/\alpha)}{n}}, \qquad R_{n,\alpha} = C_{\text{lat}}\left(r_{n,\alpha} + \frac{1}{\sqrt{n}}\right), \qquad \rho_{n,\alpha} = \frac{R_{n,\alpha}}{\pi_0},$$

with $\widehat{S}_n$ *as in Definition 35, and set*

$$E_P = \left\{\omega : d_S\left(\widehat{S}_n(\omega), S(P)\right) \leq r_{n,\alpha}\right\}.$$

Then

$$Q_P^{(n)}(E_P) \geq 1 - \alpha.$$

For every sample $\omega \in E_P$*, the cluster report* $\mathfrak{R}_{n,\alpha}$ *centered at* $\widehat{\lambda}_n(\omega)$ *satisfies the following properties for the representative atomic law* ν_P*:*

- ***(Partition.)*** *Each support point* $x \in \operatorname{supp}(\nu_P)$ *belongs to a unique empirical component* $C \in \widehat{\mathscr{K}}_{n,\alpha}$ *with* $x \in K_C(P)$*.*
- ***(Association.)*** *For every* $C \in \widehat{\mathscr{K}}_{n,\alpha}$*, the set* $K_C(P)$ *is nonempty and is contained in* $\operatorname{supp}(\nu_P)$*.*
- ***(Support interval.)*** *If* $C \in \widehat{\mathscr{K}}_{n,\alpha}$ *and* $x \in K_C(P)$*, then* x *lies in the reported support interval for* C*.*
- ***(Mass interval.)*** *If* $C \in \widehat{\mathscr{K}}_{n,\alpha}$*, then the reported mass interval* I_C *contains* $\nu_P(K_C(P))$*.*

- ***(Singleton width.)*** *If $C \in \widehat{\mathscr{K}_{n,\alpha}}$ and* $\mathrm{card}(K_C(P)) = 1$*, then the reported support interval for C has length at most $4\rho_{n,\alpha}$, and*

$$\mathrm{length}(I_C) \le \frac{8}{\pi_0} \min\left\{1, \frac{R_{n,\alpha}}{\delta(P)}\right\}.$$

- ***(External-gap width.)*** *If $C \in \widehat{\mathscr{K}_{n,\alpha}}$ and $\Delta_C(P) = +\infty$, then* $\mathrm{length}(I_C) = 0$*. For every $C \in \widehat{\mathscr{K}_{n,\alpha}}$,*

$$\mathrm{length}(I_C) \le \frac{8}{\pi_0} \min\left\{1, \frac{R_{n,\alpha}}{\Delta_C(P)}\right\},$$

with the displayed right side read as zero when $\Delta_C(P) = +\infty$.

- ***(Constrained representation.)*** *The computable class $\mathcal{C}^{\mathrm{alg}}_{n,\alpha}$ has the finite constrained-program representation*

$$\xi \in \mathcal{C}^{\mathrm{alg}}_{n,\alpha} \quad \Longleftrightarrow \quad \mathrm{AtomFloor}(\pi_0, \xi) \textit{ and } \exists\gamma \textit{ with transport cost at most } R_{n,\alpha} \textit{ from } \xi \textit{ to } \widehat{\lambda}_n(\omega).$$

Algorithm 5 defines $\widehat{p}^{\uparrow}_n$ as a simplex-valued ordered-weight rule derived from the repaired quotient-law estimator. Its target is $p^{\uparrow}(P)$ from Definition 30, so the relevant population domain is the gap-local stratum $\mathcal{M}(g)$ in Definition 41.

Theorem 5. [`thm:labeled-weight-upper`] (Ordered-weight upper bound) *Let $k, d_x, d_z \in \mathbb{N}$ and $L, \pi_0, \sigma_0 \in \mathbb{R}$. Assume:*

- ***(Latent size.)*** $2 \le k$.
- ***(Proxy dimensions.)*** $k \le d_x$ *and* $k \le d_z$.
- ***(Envelope.)*** $1 \le L$.
- ***(Latent-arm margin.)*** $0 < \pi_0 \le (2k)^{-1}$.
- ***(Rank margin.)*** $0 < \sigma_0 \le 1$.

Then there is a constant $C = C(k, d_x, d_z, L, \pi_0, \sigma_0) > 0$ such that, for every $n \ge 1$, one can choose the simplex-valued measurable ordered-weight estimator $\widehat{p}^{\uparrow}_n$ from Algorithm 5 so that, for every gap scale g satisfying $0 < g \le 1/4$ and every full-data probability law $P \in \mathcal{M}(g)$,

$$E_{Q_P^{(n)}}\left[\left\|\widehat{p}^{\uparrow}_n - p^{\uparrow}(P)\right\|_1\right] \le C \min\left\{1, \left(\sqrt{n}\, g\right)^{-1}\right\}.$$

Theorem 5 gives the ordered-weight risk bound on the separated gap-local class. The rate is clipped at one and improves as the product $\sqrt{n}\, g$ grows, capturing the scale at which effect ordering becomes informative for latent-stratum masses.

The contrast with the preceding law results is the statistical role of the gap. For the quotient law ν_P, Theorem 3 and Theorem 4 report uncertainty in W_1 and in aggregate component masses, with collisions represented by shared atoms. For ordered weights, Theorem 5 works on $\mathcal{M}(g)$, where distinct effects and the gap window make the effect order a stable object at the scale shown in the bound.

6 Lower bounds and sharp rates

Rate matching is proved on the explicit uniformly conditioned two-class specialization $k = d_x = d_z = 2$, $L = 2$, $\pi_0 = 1/10$, and $\sigma_0 = 1/10$. The construction uses two related two-class submodels. The first moves a collision witness away from a single class-average effect value at $n^{-1/2}$ scale; the second changes ordered latent masses along a factorization-preserving path whose observed separation is proportional to the product of the effect gap and the weight displacement. The lower bounds combine the displayed local Kullback–Leibler neighborhoods with the standard two-point testing logic of Cam [1986] and Polyanskiy and Wu [2019].

Definition 49. [`def:local-kl-radius`] (Local Kullback–Leibler radius c_{loc}) *Fix a local Kullback–Leibler radius constant c_{loc} satisfying*

$$0 < c_{\text{loc}} < 1.$$

The constant c_{loc} fixes the radius of the local experiments used below. Keeping this radius bounded away from one gives a common testing scale for the two-point comparisons.

Definition 50. [`synth_14`] (Factorization-preserving two-class path law $P_{g,h}$) *For $0 \le g \le 1/4$ and $|h| \le 1/100$, $P_{g,h}$ is the full-data law on $(U, T, X, Z, Y(0), Y(1), Y)$ with $U \in \{1, 2\}$, $T \in \{0, 1\}$, $X = (1, X_2)^\top$, $Z = (1, Z_2)^\top$, $Y = Y(T)$, latent masses $P_{g,h}(U = 1) = 2/5 + h$ and $P_{g,h}(U = 2) = 3/5 - h$, and latent-arm weights*

	$u = 1$	$u = 2$
$P_{g,h}(U = u, T = 1)$	$4/25 + 3h/5$	$9/25 - 3h/5$
$P_{g,h}(U = u, T = 0)$	$6/25 + 2h/5$	$6/25 - 2h/5.$

Thus $P_{g,h}(T = 1 \mid U = u) = P_{g,h}(U = u, T = 1)/P_{g,h}(U = u)$. Conditional on U, the variables $T, X_2, Y(0), Y(1)$ are Bernoulli with the displayed treatment probability, with $E_{P_{g,h}}[X_2 \mid U = 1] = 1/5$, $E_{P_{g,h}}[X_2 \mid U = 2] = (12/25 - h/5)/(3/5 - h)$, $E_{P_{g,h}}[Y(0) \mid U = u] = 1/4$, $E_{P_{g,h}}[Y(1) \mid U = 1] = 1/2 - g/2$, and $E_{P_{g,h}}[Y(1) \mid U = 2] = 1/2 + g/2$. Conditional on (U, T), Z_2 is Bernoulli and independent of $(X_2, Y(0), Y(1))$, with

$$E_{P_{g,h}}[Z_2 \mid U = 1, T = 1] = \frac{225h + 28}{20(15h + 4)}, \quad E_{P_{g,h}}[Z_2 \mid U = 2, T = 1] = \frac{3}{4},$$

$$E_{P_{g,h}}[Z_2 \mid U = 1, T = 0] = \frac{35h + 9}{10(5h + 3)}, \quad E_{P_{g,h}}[Z_2 \mid U = 2, T = 0] = \frac{7}{10}.$$

With the induced feature matrices $A_t(h)$ and $B(h)$ and latent-arm weight vector $w_t(h) = (P_{g,h}(U = 1, T = t), P_{g,h}(U = 2, T = t))$, the path preserves the armwise proxy factorization $A_t(h) \operatorname{diag}(w_t(h)) B(h)^\top = A_t(0) \operatorname{diag}(w_t(0)) B(0)^\top$ for $t \in \{0, 1\}$, and $P_{g,0} = P^{\text{wit}}_{g/2}$.

Definition 51. [`def:path-law-family`] (Path law $P_{g,h}$) *For a gap g and displacement h, $P_{g,h}$ is the explicit two-class factorization-preserving path law obtained by perturbing the latent weights while preserving the stated proxy factorizations.*

The divergence D_{KL} is used only for the observed one-record margins that generate the sample laws $Q_P^{(n)}$. Thus the lower bounds compare statistical experiments through the same observed records as the estimators and confidence reports.

The local neighborhoods center the two testing problems at the relevant collision or separated path endpoint.

Definition 52. [`def:published-vmw-regimes`] (Published VMW scope handle) *The published VMW scope handle is the cited-publication interface for the proxy-model parameter class and recovery regime. For fixed latent cardinality k, target-proxy dimension d_x, reference-proxy dimension d_z, and full-data law P on $(U, T, X, Z, Y(0), Y(1), Y)$, it records three law-level predicates:*

$$\mathsf{Model}_{\mathrm{VMW}}(k, d_x, d_z, P), \qquad \mathsf{FiniteReg}_{\mathrm{VMW}}(k, d_x, d_z, P), \qquad \mathsf{Recover}_{\mathrm{VMW}}(k, d_x, d_z, P).$$

These predicates are publication-level gates whose concrete interpretations enter through separate cited gates. The handle also includes the published estimator

$$\widehat{\Theta}_n^{\mathrm{VMW}} : (O_{d_x,d_z})^n \to \mathbb{R}^k \times \mathbb{R}^{d_x \times k} \times \mathbb{R}^k,$$

which maps an observed sample of size n from records (T, X, Z, Y) to a triple consisting of treatment effects, an anchor-normalized target-proxy feature matrix, and real-valued weight coordinates.

The quotient-law lower bound applies the Le Cam local testing framework of Cam [1986] and Polyanskiy and Wu [2019] to Assumption 13. It keeps the observed experiment within n^{-1} divergence of the collision witness, so an n-sample comparison has bounded information distance.

Definition 53. [`synth_6`] (Kullback–Leibler divergence D_{KL}) *For probability laws μ and ν on the same measurable space, define $D_{\mathrm{KL}}(\mu\|\nu) := \int \log(d\mu/d\nu)\, d\mu$ when $\mu \ll \nu$, and set $D_{\mathrm{KL}}(\mu\|\nu) := +\infty$ when $\mu \not\ll \nu$. In the local experiments this divergence is applied to one-record observed-data margins such as P_O and $(P_0^{\mathrm{wit}})_O$.*

The ordered-weight lower bound applies the Le Cam local testing framework of Cam [1986] and Polyanskiy and Wu [2019] to Assumption 14. The center has effect separation indexed by g, matching the gap-local target used in Definition 41. The next display repeats the same divergence notation for the local-experiment statements that follow.

The witness laws are finite Bernoulli proxy models with explicit masses, treatment probabilities, proxy features, and potential-outcome means.

Definition 54. [`synth_15`] (Kullback–Leibler divergence) *For probability laws μ and ν on the same measurable space, the Kullback–Leibler divergence from ν to μ is*

$$D_{\mathrm{KL}}(\mu\|\nu) := \int \log\left(\frac{d\mu}{d\nu}\right) d\mu$$

when $\mu \ll \nu$, and $D_{\mathrm{KL}}(\mu\|\nu) := +\infty$ when $\mu \not\ll \nu$. In the local experiments this divergence is applied to one-record observed-data margins, including P_O and $(P_0^{\mathrm{wit}})_O$.

Assumption 13. [`ass:local-quotient-neighborhood`] (Local quotient neighborhood) *The Kullback–Leibler divergence from P_O, the observed-data marginal, to the observed margin of P_0^{wit}, the two-class collision witness at perturbation zero, satisfies*

$$D_{\mathrm{KL}}\big(P_O \| (P_0^{\mathrm{wit}})_O\big) \le \frac{c_{\mathrm{loc}}}{n}.$$

Proposition 5 places the collision witness inside the same uniformly conditioned class used for the upper bounds. The determinants and arm probabilities show that the testing pair is an interior proxy model under the displayed margins, while the quotient law moves from a single atom at $\varepsilon = 0$ to two nearby atoms for positive ε.

For labeled weights, the path changes latent masses while preserving the proxy factorizations that determine the observable $ZX^\top$ moments.

Assumption 14. [`ass:local-weight-neighborhood`] (Local weight neighborhood) *The Kullback–Leibler divergence from P_O, the observed-data marginal, to the observed margin of the two-class witness at perturbation $g/2$ satisfies*

$$D_{\mathrm{KL}}\big(P_O \| (P^{\mathrm{wit}}_{g/2})_O\big) \le \frac{c_{\mathrm{loc}}}{n}.$$

Definition 55. [`def:two-class-witness`] (Two-class witness law $P^{\mathrm{wit}}_{\varepsilon}$) *For $k = d_x = d_z = 2$, the two-class witness law $P^{\mathrm{wit}}_{\varepsilon}$ is the law specified by*

$$P(U=1) = 0.4, \qquad P(U=2) = 0.6,$$

$$P(T=1 \mid U=1) = 0.4, \qquad P(T=1 \mid U=2) = 0.6,$$
$$X = (1, X_2)^\top, \qquad Z = (1, Z_2)^\top,$$
$$B = \left(\begin{smallmatrix} 1 & 1 \\ 0.2 & 0.8 \end{smallmatrix}\right), \qquad A_0 = \left(\begin{smallmatrix} 1 & 1 \\ 0.3 & 0.7 \end{smallmatrix}\right), \qquad A_1 = \left(\begin{smallmatrix} 1 & 1 \\ 0.35 & 0.75 \end{smallmatrix}\right),$$
$$E[Y(0) \mid U] = (0.25, 0.25), \qquad E[Y(1) \mid U] = (0.5 - \varepsilon, 0.5 + \varepsilon).$$

Conditional on U, the variables X_2, $Y(0)$, $Y(1)$, and T are mutually independent Bernoulli variables with the displayed conditional means. Conditional on U and T, the variable Z_2 is independent of $(X_2, Y(0), Y(1))$ and has conditional mean given by A_T. The observed outcome is

$$Y = Y(T).$$

The family $P_{g,h}$ supplies the local perturbation for the ordered-weight converse. At $h = 0$ it coincides with the separated witness at displacement $g/2$, and changing h changes the ordered masses while retaining the displayed armwise proxy factorizations.

The two-class witness also gives a compact picture of the reporting geometry. In this specialization the class masses are 0.4 and 0.6, and the class-average effects are $0.25 - \varepsilon$ and $0.25 + \varepsilon$. With association radius $\rho_{n,\alpha}$, the empirical graph in Algorithm 4 connects estimated atoms at threshold $4\rho_{n,\alpha}$; the support interval then expands each empirical component by $\rho_{n,\alpha}$.

witness regime	quotient law	component rule at radius $\rho_{n,\alpha}$	reported ma
$\varepsilon = 0$	$\delta_{0.25}$	one component around 0.25	1
$0 < 2\varepsilon \le 4\rho_{n,\alpha}$	$0.4\,\delta_{0.25-\varepsilon} + 0.6\,\delta_{0.25+\varepsilon}$	connected empirical component	$0.4 + ($
$2\varepsilon > 4\rho_{n,\alpha}$	$0.4\,\delta_{0.25-\varepsilon} + 0.6\,\delta_{0.25+\varepsilon}$	two empirical components when the center is separated	0.4, (

The table serves as a reading guide for the existing witness. It illustrates how quotient aggregation and component reporting use the same masses while the merge threshold determines whether the report presents one aggregate interval or two separated component intervals.

Definition 56. [`def:local-quotient-experiment`] (Local quotient experiment $\mathcal{L}^{\nu}_n$) *The local quotient-law KL experiment is*

$$\mathcal{L}^{\nu}_n := \{P \in \mathcal{M} : P \text{ satisfies Assumption } 13\},$$

where $\mathcal{M}$ is the model class in Definition 32.

Definition 57. [def:local-weight-experiment] (Local weight experiment $\mathcal{L}^p_{n,g}$) *The local labeled-weight KL experiment is*

$$\mathcal{L}^p_{n,g} := \{P \in \mathcal{M}(g) : P \text{ satisfies Assumption 14}\},$$

where $\mathcal{M}(g)$ is the gap-local stratum in Definition 41.

The two local experiments distinguish the metrics under study. $\mathcal{L}^\nu_n$ is centered at a collision and measures the law-level W_1 difficulty, whereas $\mathcal{L}^p_{n,g}$ is centered on the gap-local stratum and measures the difficulty of recovering effect-ordered masses.

Proposition 5. [prop:two-class-witness-valid] (Two-class witness validity) *For every collision-witness effect displacement ε satisfying*

$$0 \le \varepsilon \le \frac{1}{8},$$

the explicit two-class Bernoulli proxy witness law $P^{\mathrm{wit}}_\varepsilon$ belongs to $\mathcal{M}$ with $L = 2$, $\pi_0 = 0.1$, and $\sigma_0 = 0.1$. Moreover,

$$P^{\mathrm{wit}}_\varepsilon(T = 0) = 0.48, \qquad P^{\mathrm{wit}}_\varepsilon(T = 1) = 0.52,$$
$$\det(B) = 0.6, \qquad \det(A_0) = \det(A_1) = 0.4,$$
$$\det(M_0(P^{\mathrm{wit}}_\varepsilon)) = 0.06, \qquad \det(M_1(P^{\mathrm{wit}}_\varepsilon)) = 0.4 \cdot 0.6 \cdot \frac{36}{169},$$

and the quotient latent-effect law is

$$\nu_{P^{\mathrm{wit}}_\varepsilon} = 0.4\,\delta_{0.25-\varepsilon} + 0.6\,\delta_{0.25+\varepsilon}.$$

At $\varepsilon = 0$, the proxy feature and observable moment matrices above are nonsingular, the quotient latent-effect law is $\delta_{0.25}$, and the two latent-class masses are distinct.

It is worth tracing the collision through the witness once, since every object in the paper is visible on it. The latent effects of $P^{\mathrm{wit}}_\varepsilon$ are read off the displayed conditional means as $\tau_1 = 0.25 - \varepsilon$ and $\tau_2 = 0.25 + \varepsilon$, carried by the class masses 0.4 and 0.6, so the effect gap is 2ε and the quotient law is

$$\nu_{P^{\mathrm{wit}}_\varepsilon} = 0.4\,\delta_{0.25-\varepsilon} + 0.6\,\delta_{0.25+\varepsilon} \qquad (\varepsilon > 0), \qquad \nu_{P^{\mathrm{wit}}_0} = \delta_{0.25}.$$

regime	latent effects	gap	quotient law	effect-ordered masses
$\varepsilon = 0$	$0.25,\ 0.25$	0	one atom at 0.25 of mass 1	labels carry no separation
$\varepsilon > 0$	$0.25 - \varepsilon,\ 0.25 + \varepsilon$	2ε	0.4 at $0.25 - \varepsilon$, 0.6 at $0.25 + \varepsilon$	$(0.4, 0.6)$

Transporting the mass 0.4 and the mass 0.6 each a distance ε onto the common location gives $W_1(\nu_{P^{\mathrm{wit}}_\varepsilon}, \nu_{P^{\mathrm{wit}}_0}) = \varepsilon$, so the estimand moves continuously into the collision and the modulus of Theorem 1 applies across the whole family at fixed conditioning constants. The two regimes are distinguished by which report they support. The quotient law and its W_1 confidence set of Algorithm 3 remain meaningful at $\varepsilon = 0$, where the two classes have become one atom of mass 1. The effect-ordered masses $(0.4, 0.6)$ are separated only while 2ε is positive, which is why Theorem 5 is stated on the gap-local strata and why its converse in Theorem 6 is driven by displacement along this same family. The cluster report of Algorithm 4 interpolates between the two pictures: while the estimated support points fall within the connection threshold $4\rho_{n,\alpha}$ they are returned as one component whose aggregate mass interval covers the combined mass, and once the separation exceeds that threshold they are reported as two components whose mass intervals sharpen with the external separation, in the manner quantified by Theorem 4.

Theorem 6 gives the two local converse rates on the displayed two-class specialization. For quotient laws, an $n^{-1/2}$ displacement in the collision witness creates W_1 separation of order $n^{-1/2}$ while keeping the observed Kullback–Leibler comparison at order n^{-1}. For ordered weights, the factorization-preserving path makes the observed divergence scale as g^2h^2, so the largest indistinguishable mass displacement is clipped at $\min\{1, (\sqrt{n}g)^{-1}\}$.

These local statements calibrate the upper bounds from the preceding sections on the same explicit two-class specialization. The quotient-law lower bound matches the root-n law rate from Theorem 2; the ordered-weight lower bound matches the inverse-gap upper bound in Theorem 5.

Theorem 6. [`thm:matching-local-lower-bounds`] (Matching local lower bounds) *Fix $c_{\mathrm{loc}} \in (0,1)$. There exist constants $a, c, C \in \mathbb{R}$ such that*

$$0 < a \le \frac{1}{8}, \qquad c > 0, \qquad C > 0,$$

and, for every integer $n \ge 1$, the following statements hold for the explicit submodel $k = d_x = d_z = 2$, $L = 2$, $\pi_0 = 1/10$, and $\sigma_0 = 1/10$.

- ***(Quotient-law risk.)*** *For every measurable quotient-law estimator $\widetilde{\nu}_n$ with support radius*

  $$L_\tau = \frac{4L\sqrt{d_z}}{\sigma_0},$$

  there is a law P in the local quotient-law experiment $\mathcal{L}_n^\nu$ of Definition 56 such that

  $$P = P_0^{\mathrm{wit}} \quad \text{or} \quad P = P_{a/\sqrt{n}}^{\mathrm{wit}},$$

  with $P_\varepsilon^{\mathrm{wit}}$ as in Definition 55, and

  $$\frac{c}{\sqrt{n}} \le E_{Q_P^{(n)}}[W_1(\widetilde{\nu}_n, \nu_P)]\,.$$

- ***(Ordered-weight risk.)*** *For every $g \in (0, 1/4]$ and every measurable ordered-weight estimator $\widetilde{p}_n$ taking values in the simplex, there is a law P in the local labeled-weight experiment $\mathcal{L}_{n,g}^p$ of Definition 57 such that*

  $$P = P_{g,0} \quad \text{or} \quad P = P_{g,h(n,g)}, \qquad h(n,g) = a\min\left\{1, \frac{1}{\sqrt{n}\,g}\right\},$$

  and

  $$c\min\left\{1, \frac{1}{\sqrt{n}\,g}\right\} \le E_{Q_P^{(n)}}\left[\|\widetilde{p}_n - p^\uparrow(P)\|_1\right].$$

- ***(Quotient-law certificate.)*** *The laws P_0^{wit} and $P_{a/\sqrt{n}}^{\mathrm{wit}}$ belong to the corresponding local quotient-law experiments, and their observed-data marginals and quotient laws satisfy*

  $$D_{\mathrm{KL}}\Big((P_{a/\sqrt{n}}^{\mathrm{wit}})_O \| (P_0^{\mathrm{wit}})_O\Big) \le \frac{C}{n}, \qquad \frac{c}{\sqrt{n}} \le W_1\Big(\nu_{P_0^{\mathrm{wit}}}, \nu_{P_{a/\sqrt{n}}^{\mathrm{wit}}}\Big)\,.$$

- ***(Labeled-path certificate.)*** *For every $g \in (0, 1/4]$, with*

$$h = h(n, g) = a \min\left\{1, \frac{1}{\sqrt{n}\, g}\right\},$$

the displacement h lies in the tangent-amplitude domain, the laws $P_{g,h}$ and $P_{g,0}$ belong to the corresponding local labeled-weight experiments, and

$$D_{\mathrm{KL}}((P_{g,h})_O \| (P_{g,0})_O) \le Cg^2h^2, \qquad \sum_i \left| p_i^{\uparrow}(P_{g,h}) - p_i^{\uparrow}(P_{g,0}) \right| = 2|h|.$$

Proposition 6 converts the local quotient-law comparison into a same-class minimax statement on the displayed uniformly conditioned two-class model. The proposition states that the optimal uniform W_1 risk for the quotient law is of exact root-n order on this specialization.

Proposition 6. [`prop:same-class-quotient-minimax`] (Same-class quotient minimax) *For the specialization $k = d_x = d_z = 2$, $L = 2$, $\pi_0 = 1/10$, and $\sigma_0 = 1/10$, let $\mathcal{M}$ be the uniformly conditioned proxy causal model class in Definition 32. The derived effect-support radius is*

$$L_\tau = 4L\sqrt{d_z}/\sigma_0 = 80\sqrt{2}.$$

There exist real constants c and C such that

$$0 < c < C$$

and, for every sample size $n \ge 1$, the following two statements hold:

- ***(Lower bound.)*** *For every measurable estimator $\widetilde{\nu}_n$ mapping an n-sample of observed records to a represented atomic law modulo the radius L_τ, there exists a full-data probability law $P \in \mathcal{M}$ such that*

$$\frac{c}{\sqrt{n}} \le E_{Q_P^{(n)}}[W_1(\widetilde{\nu}_n, \nu_P)].$$

- ***(Upper bound.)*** *There exists a measurable estimator $\widetilde{\nu}_n$ mapping an n-sample of observed records to a represented atomic law modulo the radius L_τ such that, for every full-data probability law $P \in \mathcal{M}$,*

$$E_{Q_P^{(n)}}[W_1(\widetilde{\nu}_n, \nu_P)] \le \frac{C}{\sqrt{n}}.$$

Equivalently, on this same specialized class,

$$\frac{c}{\sqrt{n}} \le \inf_{\widetilde{\nu}_n} \sup_{P \in \mathcal{M}} E_{Q_P^{(n)}}[W_1(\widetilde{\nu}_n, \nu_P)] \le \frac{C}{\sqrt{n}}$$

for every $n \ge 1$, where the infimum ranges over measurable estimators with output radius L_τ.

Proposition 7 gives the corresponding same-class minimax statement for ordered weights on the same two-class specialization. The rate is governed by the product $\sqrt{n}g$: when the effect gap is large relative to sampling noise, ordered masses are estimable at the inverse-gap scale, and the bound remains clipped by the diameter of the simplex-valued loss.

The final comparison records how the same witnesses transfer to classes formulated in the terminology of Virk et al. [2026]. The following definitions separate the published structural gates, finite-regularity condition, and separated recovery regime used for that transfer.

Proposition 7. [`prop:same-class-labeled-minimax`] (Same-class labeled minimax) *There exist real constants c and C such that $0 < c < C$ and, for every sample size $n \geq 1$ and every gap scale g with $0 < g \leq 1/4$, the gap stratum $\mathcal{M}(g)$ from Definition 41 satisfies*

$$c \min\left\{1, \frac{1}{\sqrt{n}\, g}\right\} \leq \inf_{\widetilde{p}_n} \sup_{P \in \mathcal{M}(g)} E_{Q_P^{(n)}}\left[\left\|\widetilde{p}_n - p^{\uparrow}(P)\right\|_1\right] \leq C \min\left\{1, \frac{1}{\sqrt{n}\, g}\right\}.$$

Here the infimum is over all measurable simplex-valued labeled weight estimators based on n observed records in the explicit submodel $k = d_x = d_z = 2$, $L = 2$, $\pi_0 = 1/10$, and $\sigma_0 = 1/10$, and $p^{\uparrow}(P)$ is the effect-ordered latent-mass vector with the barycentric collision fallback.

Definition 58. [`synth_31`] (Calibrated path displacement) *For $a, g \in \mathbb{R}$ and $n \in \mathbb{N}$, define*

$$\mathrm{calibratedDisplacement}(a, n, g) = a\, \min\{1, (\sqrt{n}\, g)^{-1}\}.$$

Definition 59. [`synth_12`] (Published VMW finite-regularity condition) *A law P satisfies the published VMW finite-regularity condition when $0 < k$, $k \leq d_x$, $k \leq d_z$, and there exist constants $L_X, L_{ZX}, L_{YZX}, \pi, \sigma > 0$ such that $\|X\|_2 \leq L_X$, $\|ZX^\top\|_{\mathrm{op}} \leq L_{ZX}$, and $\|YZX^\top\|_{\mathrm{op}} \leq L_{YZX}$ almost surely, $P(T = t) \geq \pi$ for $t \in \{0, 1\}$, and there is a top-k right singular basis V for the stacked proxy moment $[M_0(P); M_1(P)]$ satisfying $\sigma_k([M_0(P); M_1(P)]) \geq \sigma$ and $\sigma_k(M_t(P)V) \geq \sigma$ for each $t \in \{0, 1\}$.*

Definition 60. [`synth_11`] (Published VMW structural model) *Fix k, d_x, d_z with $0 < k$, $k \leq d_x$, and $k \leq d_z$. A one-unit full-data probability law P belongs to the published VMW structural model when it satisfies the published qualitative proxy model conditions: reference-proxy separation $Z \perp (X, Y) \mid (T, U)$, target-proxy separation $X \perp (Y, T) \mid U$, consistency $Y = Y(T)$ almost surely, armwise latent ignorability $Y(t) \perp T \mid U$ for $t \in \{0, 1\}$, full column rank of A_0, A_1, and B, and strict latent-arm positivity $P(U = u, T = t) > 0$ for every $u \in \{1, \ldots, k\}$ and $t \in \{0, 1\}$.*

Definition 52, Definition 59, Definition 60, and Definition 61 provide the interface for comparing the present lower bounds with the published VMW recovery framework. The cited model conditions and separated recovery regime are used as external publication-level gates, while the converse below is stated in terms of exact containment of the two displayed witness pairs.

Definition 61. [`synth_13`] (Separated recovery regime) *A full-data law P lies in the separated recovery regime when it satisfies:*

- ***(Structural model.)*** *The published VMW structural model conditions.*
- ***(Anchor normalization.)*** *The first target-proxy coordinate satisfies $X_1 = 1$ almost surely.*
- ***(Finite regularity.)*** *The published VMW finite-regularity condition.*
- ***(Spectral separation.)*** *The latent-class treatment effects satisfy $\tau_u \neq \tau_v$ whenever $u \neq v$.*

In this regime, the separated-recovery estimator analyzed by Virk et al. [2026] is evaluated under its stated sample-size and radius conditions and returns simultaneous high-probability bounds for the treatment effects, the anchor-normalized feature matrix, and the simplex-projected mixture weights.

The calibrated displacement in Definition 58 is the path amplitude used in the labeled-weight two-point comparison. It is the same clipped scale that appears in the ordered-weight upper and lower bounds.

Theorem 7. [`thm:published-vmw-converse-transfer`] (Published converse transfer)[2] *Fix $k = d_x = d_z = 2$, $L = 2$, and $\sigma_0 = 1/10$. There exist constants $a, c \in \mathbb{R}$ such that*

$$0 < a \leq \frac{1}{8}, \qquad 0 < c.$$

For every $n \geq 1$, the following hold.

- ***(Quotient witness pair.)*** *The two laws P_0^{wit} and $P_{a/\sqrt{n}}^{\mathrm{wit}}$ from Definition 55 belong to the published VMW model and satisfy VMW Assumption 4, while P_0^{wit} lies outside the published separated recovery regime.*

- ***(Quotient lower bound.)*** *For every class $\mathcal{V}_n^\nu$ of full-data probability laws such that every $P \in \mathcal{V}_n^\nu$ belongs to the published VMW model and satisfies VMW Assumption 4, if $\mathcal{V}_n^\nu$ contains laws equal to P_0^{wit} and $P_{a/\sqrt{n}}^{\mathrm{wit}}$, then every estimator $\widetilde{\nu}_n$ of the quotient effect law with support radius $4 \cdot 2\sqrt{2}/(1/10)$ has worst-case risk at least $c/\sqrt{n}$:*

$$\sup_{P \in \mathcal{V}_n^\nu} E_{Q_P^{(n)}}[W_1(\widetilde{\nu}_n, \nu_P)] \geq \frac{c}{\sqrt{n}}.$$

- ***(Separated labeled path.)*** *For every g in the gap-scale domain, define*

$$h(n, g) := \mathrm{calibratedDisplacement}(a, n, g).$$

  Then $h(n, g)$ is in the tangent-amplitude domain, and the endpoint laws $P_{g,0}$ and $P_{g,h(n,g)}$ both belong to the published VMW model, satisfy VMW Assumption 4, and have spectrally separated latent effects.

- ***(Labeled-weight lower bound.)*** *For every g in the gap-scale domain and every class $\mathcal{V}_{n,g}^p$ of full-data probability laws such that every $P \in \mathcal{V}_{n,g}^p$ belongs to the published VMW model and satisfies VMW Assumption 4, if $\mathcal{V}_{n,g}^p$ contains laws equal to $P_{g,0}$ and $P_{g,h(n,g)}$, then every estimator $\widetilde{p}_n$ of the ordered latent weights satisfies*

$$\sup_{P \in \mathcal{V}_{n,g}^p} E_{Q_P^{(n)}}\left[\left\|\widetilde{p}_n - p^\uparrow(P)\right\|_1\right] \geq c \min\{1, (\sqrt{n}\, g)^{-1}\}.$$

Theorem 7 states the converse comparison with Virk et al. [2026]. Any comparator class satisfying the cited VMW model and finite-regularity gates inherits the displayed lower bound once it contains the two quotient witness laws or the two labeled-path endpoint laws. The transfer is therefore tied to concrete witness containment: the quotient comparison uses the collision pair, while the labeled-weight comparison uses separated endpoint laws at the calibrated displacement.

[2] **Formalization scope.** This result uses the published conclusion [Virk et al., 2026] (Assumptions 1–2 and Section 4.2 of arXiv:2607.10926v1) and [Virk et al., 2026] (Assumption 3, lines 261–263; Assumption 4, lines 321–337; and Theorem 7.2, lines 350–381 of arXiv:2607.10926v1). 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.

7 Discussion, extensions, and open questions

The results above identify the quotient law as the regular object for proxy-based latent-class mean treatment-effect heterogeneity. Theorem 1 characterizes a Lipschitz map from observable moment summaries to the aggregated latent-class mean effect law, and Theorem 2 transfers this population regularity to collision-uniform root-n estimation in the observed i.i.d. experiment. The estimand in Definition 40 is therefore the natural law-level target when the empirical question concerns population mass at class-average effect values, with class labels handled through the separated ordered-weight results.

The confidence constructions in Algorithm 3 give two complementary ways to report this target. The summary-inversion set $\mathcal{C}_{n,\alpha}$ preserves the repaired feasible-image construction induced by $\overline{F}$ on $\mathcal{K}$, while the algorithmic set $\mathcal{C}_{n,\alpha}^{\mathrm{alg}}$ is centered at the structured-lattice estimator $\widehat{\lambda}_n$ and has an explicit finite transport-plan representation in the fixed-dimensional exact-real model. Theorem 3 establishes simultaneous coverage for both sets and root-n diameter bounds in W_1, so the two reports express the same law-level concentration through different computational representations.

The cluster report in Algorithm 4 translates this law-level uncertainty into component-level statements at the resolution delivered by the confidence radius. Estimated atoms are connected at threshold $4\rho_{n,\alpha}$, while support association and interval expansion use $\rho_{n,\alpha}$; the report then pools connected empirical components and attaches both a support interval and a compatible mass interval. Theorem 4 establishes that, on the simultaneous concentration event, these components partition the true support, contain the associated true atoms, and cover the corresponding aggregate masses. The interval-length bounds show how the report becomes sharper when a component is externally separated from the remaining support, with exact mass recovery for a component whose associated block exhausts the true support.

This reporting layer also clarifies the role of the ordered-weight results. Theorems 5 and 6 and Proposition 7 characterize effect-ordered mass recovery on separated strata, where the positive gap scale converts law-level transport error into label-sensitive information. The quotient-law statements remain organized around W_1 on aggregated laws, while the ordered-weight statements quantify the additional inverse-gap cost of resolving labels by effect order.

7.1 Open questions

The computational distinction between the two confidence constructions leaves a sharp finite-dimensional problem. The structured-lattice estimator and the outer W_1 set have explicit constrained representations, and the original summary-inversion set retains the exact repaired-image target. The following remark records the corresponding exact-image question.

Remark 1. [`oeq:constructive-total-repair`] (Sharp summary-inversion computation) *It remains open whether, in the fixed-dimensional exact-real model, one can compute exactly and in time polynomial in n the sharp summary-inversion image*

$$\mathcal{C}_{n,\alpha} = \{\overline{F}(q) : q \in \mathcal{K},\ d_S(q, \Pi(\widehat{S}_n)) \le 2r_{n,\alpha}\}$$

together with its exact support-dependent cluster extrema, using neither compact nearest-summary optimization nor black-box evaluation of $\overline{F}$. The positive real law estimator $\widehat{\lambda}_n$, the honest outer set $\mathcal{C}_{n,\alpha}^{\mathrm{alg}}$, and every interval in the report $\mathfrak{R}_{n,\alpha}$ have explicit fixed-dimensional *constrained representations, including cases in which the raw empirical operator has non-real eigenpairs. The question concerns a sharp polynomial-time algorithm for the original summary-inversion image and its extrema.*

The question in Remark 1 concerns exact computation of the sharp repaired-image set and its support-dependent extrema. A positive resolution would align the original summary-inversion report with the same kind of explicit fixed-dimensional representation already available for the lattice-centered outer report.

A second extension is adaptation to latent cardinality and unknown conditioning margins. The present model class in Definition 32 fixes k, and the target space, structured lattice, confidence sets, and ordered-weight stratum are all calibrated to that fixed finite value together with supplied L, π_0, σ_0. An adaptive formulation would need a data-dependent dimension choice or margin choice together with a target convention for comparing quotient laws across candidate values of k, while preserving the aggregation principle that makes ν_P stable at collisions.

Margin misspecification belongs to the same adaptation problem. The stated guarantees attach to the supplied model class: conservative supplied constants that still satisfy the defining inequalities yield valid conclusions after recomputing the displayed constants, whereas supplied constants violated by the data-generating law place that law outside the certified class. A diagnostic theory for selecting or stress-testing k, L, π_0, σ_0 is a natural extension of the fixed-margin results.

The exact-real finite representations also leave ordinary computational questions for implementation work. The operation counts in Definition 48, Proposition 4, and Theorem 8 use fixed-dimensional unit-cost primitives, and the endpoint descriptions in Algorithm 4 are finite constrained programs over masses, support locations, and transport plans. Bit complexity for these endpoint programs, numerical conditioning of the singular-value and root-isolation primitives, and practical solvers for the reported intervals are open implementation questions beyond the exact-real certification.

A third direction concerns application-specific proxy mappings. Assumptions 1 to 10 give a finite proxy-causal environment in which the quotient-law and reporting results apply. In applied settings, the substantive task is to justify how observed proxy measurements instantiate these conditional-independence, normalization, boundedness, overlap, and rank conditions, following the model-building role that proxy variables play in the broader causal literature [Virk et al., 2026, van Amsterdam et al., 2022].

Appendices

A Probability, selection, and concentration

This appendix collects the auxiliary statements used by the population, estimation, confidence, and lower-bound arguments. The first group supplies measurable selection and finite-sample probability control for the observable summary space introduced in Definition 33. The compact-selector statement is the measurable nearest-point device used by the repair map in Algorithm 1; it is the form of measurable extrema needed for a compact feasible summary set, in the spirit of Brown and Purves [1973].

Lemma 2. [`lem:borel-nearest-point-selector`] (Compact loss selector) *For every $d \in \mathbb{N}$, every set $\mathcal{K} \subseteq \mathbb{R}^d$, and every loss $d_S : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, suppose that*

- ***(Nonempty feasible set.)*** *$\mathcal{K}$ is nonempty.*
- ***(Compact feasible set.)*** *$\mathcal{K}$ is compact.*
- ***(Continuous loss.)*** *The map $(s, q) \mapsto d_S(s, q)$ is continuous.*

Then there exists a Borel measurable map $\Pi : \mathbb{R}^d \to \mathbb{R}^d$ such that, for every $s \in \mathbb{R}^d$,

$$\Pi(s) \in \mathcal{K} \quad \text{and} \quad d_S(s, \Pi(s)) \le d_S(s, q) \text{ for every } q \in \mathcal{K}.$$

Proof of Lemma 2. Identify $\mathbb{R}^d$ with its standard finite-coordinate Euclidean realization. Define

$$f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}, \qquad f(s, q) := d_S(s, q).$$

By the continuity hypothesis, f is Borel measurable, and for each fixed s the section

$$q \longmapsto f(s, q) = d_S(s, q)$$

is continuous on $\mathcal{K}$.

The compact measurable argmin selection theorem for a nonempty compact Euclidean action set gives a map

$$\Pi : \mathbb{R}^d \to \mathbb{R}^d$$

such that Π is Borel measurable and, for every $s \in \mathbb{R}^d$,

$$\Pi(s) \in \mathcal{K}, \qquad \forall q \in \mathcal{K}, \quad f(s, \Pi(s)) \le f(s, q).$$

Substituting the definition of f gives

$$\Pi(s) \in \mathcal{K}, \qquad \forall q \in \mathcal{K}, \quad d_S(s, \Pi(s)) \le d_S(s, q),$$

which is the asserted selector property. □

The selection lemma turns compactness of $\mathcal{K}$ into a Borel rule, so the nearest-summary construction remains a sample-measurable estimator after applying the empirical summary map. The remaining probability statements in this subsection record the arm-mass and concentration primitives that support the total empty-arm convention.

Lemma 3. [`lem:arm-mass-lower-bound`] (Arm mass lower bound) *Let P be a full-data probability law in the setting of Definition 32. Suppose P satisfies Assumption 9 with margin π_0. Then, for each treatment arm $t \in \{0, 1\}$,*

$$k\pi_0 \le P(T = t).$$

Proof of Lemma 3. Fix an arm $t \in \{0, 1\}$. The arm event is the disjoint union of its latent-arm cells:

$$\{T = t\} = \bigsqcup_{u=1}^{k} \{U = u, T = t\}.$$

Therefore

$$P(T = t) = \sum_{u=1}^{k} P(U = u, T = t).$$

By Assumption 9, each summand is at least π_0. Hence

$$P(T = t) = \sum_{u=1}^{k} P(U = u, T = t) \geq \sum_{u=1}^{k} \pi_0 = k\pi_0.$$

□

The lower bound converts the latent-arm margin into an observed-arm margin. It is the elementary probability input behind the empty-arm probability bound in Lemma 1 and behind the observed VMW margin statement in Proposition 1.

The next statements collect the moment factorizations and envelope calculations that connect the full-data restrictions to the observable five-block summary.

Lemma 4. [`lem:observed-proxy-moment-factorization`] (Observed proxy moment factorization) *Let P be a full-data probability law carrying $(U, T, X, Z, Y(0), Y(1), Y)$, with fixed target- and reference-proxy dimensions d_x, d_z, in the uniformly conditioned proxy causal model class $\mathcal{M}$ of Definition 32. Suppose that:*

- ***(Latent cardinality.)*** *The number k of latent classes satisfies $2 \leq k$.*
- ***(Target dimension.)*** *The target-proxy dimension satisfies $k \leq d_x$.*
- ***(Boundedness radius.)*** *The uniform boundedness radius satisfies $1 \leq L$, with the bounded-target-proxy condition as in Assumption 6.*
- ***(Latent-arm margin.)*** *The latent-arm positivity margin satisfies $\pi_0 > 0$, with latent-arm positivity as in Assumption 9.*
- ***(Proxy separation.)*** *The reference- and target-proxy separation conditions hold as in Assumptions 1 and 2.*

Then, for each treatment arm $t \in \{0, 1\}$, the observable armwise proxy moment $M_t(P)$ from Definition 28 factors as

$$M_t(P) = A_t \Gamma_t B^\top,$$

where A_t is the reference-proxy feature matrix of Definition 24, B is the target-proxy feature matrix of Definition 25, and

$$\Gamma_t = \operatorname{diag}\left\{ \frac{P(U = u, T = t)}{P(T = t)} : 1 \leq u \leq k \right\}.$$

Proof of Lemma 4. Fix $t \in \{0, 1\}$. Write

$$\mathsf{Arm}_t := \{T = t\}, \qquad \mathsf{Cell}_{u,t} := \{U = u, T = t\}, \qquad \mathsf{Class}_u := \{U = u\}.$$

For every latent class u, latent-arm positivity gives

$$P(\mathsf{Cell}_{u,t}) \geq \pi_0 > 0, \qquad P(\mathsf{Class}_u) \geq P(\mathsf{Cell}_{u,t}) > 0.$$

Also, Lemma 3 gives

$$P(\mathsf{Arm}_t) \geq k\pi_0 > 0.$$

We prove the matrix identity entrywise. Fix coordinates $1 \le i \le d_z$ and $1 \le j \le d_x$. By Lemma 5,

$$|Z_i| \le L, \qquad |X_j| \le L \quad P\text{-almost surely.}$$

Thus $Z_i X_j$ is integrable on each $\mathsf{Cell}_{u,t}$. The observed arm pulls back to the full-data arm, so

$$M_t(P)_{ij} = E_{P_O}[Z_i X_j \mid T = t] = E_P[Z_i X_j \mid \mathsf{Arm}_t].$$

Since Arm_t is the disjoint union of the cells $\mathsf{Cell}_{u,t}$, normalized additivity over this finite partition yields

$$E_P[Z_i X_j \mid \mathsf{Arm}_t] = \sum_{u=1}^{k} \frac{P(\mathsf{Cell}_{u,t})}{P(\mathsf{Arm}_t)} E_P[Z_i X_j \mid \mathsf{Cell}_{u,t}].$$

On a fixed positive cell $\mathsf{Cell}_{u,t}$, put

$$\mathrm{cl}_L(a) := \max\{-L, \min\{a, L\}\}.$$

The coordinate bounds imply that $\mathrm{cl}_L(Z_i) = Z_i$ and $\mathrm{cl}_L(X_j) = X_j$ almost surely on the normalized restriction to $\mathsf{Cell}_{u,t}$. Applying reference-proxy separation to the bounded tests $z \mapsto \mathrm{cl}_L(z_i)$ and $(x, y) \mapsto \mathrm{cl}_L(x_j)$, and then removing the clamps, gives

$$E_P[Z_i X_j \mid \mathsf{Cell}_{u,t}] = E_P[Z_i \mid \mathsf{Cell}_{u,t}]\, E_P[X_j \mid \mathsf{Cell}_{u,t}].$$

It remains to identify the target-proxy mean in the cell with the latent-class mean. Let

$$f_j(x) := \mathrm{cl}_L(x_j), \qquad q_t(y, t') := \mathbf{1}\{t' = t\}.$$

Target-proxy separation on the positive class Class_u gives

$$E_P[f_j(X) q_t(Y, T) \mid \mathsf{Class}_u] = E_P[f_j(X) \mid \mathsf{Class}_u]\, E_P[q_t(Y, T) \mid \mathsf{Class}_u].$$

The two factors involving q_t are ordinary normalized cell masses:

$$E_P[f_j(X) q_t(Y, T) \mid \mathsf{Class}_u] = \frac{P(\mathsf{Cell}_{u,t})}{P(\mathsf{Class}_u)} E_P[f_j(X) \mid \mathsf{Cell}_{u,t}],$$

and

$$E_P[q_t(Y, T) \mid \mathsf{Class}_u] = \frac{P(\mathsf{Cell}_{u,t})}{P(\mathsf{Class}_u)}.$$

Because this ratio is positive, cancellation gives

$$E_P[f_j(X) \mid \mathsf{Cell}_{u,t}] = E_P[f_j(X) \mid \mathsf{Class}_u].$$

Using the same coordinate envelope to remove the clamps on the cell and on the class,

$$E_P[X_j \mid \mathsf{Cell}_{u,t}] = E_P[X_j \mid \mathsf{Class}_u].$$

Combining the preceding displays and using Definitions 24 and 25,

$$M_t(P)_{ij} = \sum_{u=1}^{k} A_t(i, u)\, \frac{P(U = u, T = t)}{P(T = t)}\, B(j, u).$$

By the definition of Γ_t, the right-hand side is exactly

$$(A_t \Gamma_t B^\top)_{ij}.$$

Since i and j were arbitrary,

$$M_t(P) = A_t \Gamma_t B^\top.$$

The arm t was arbitrary, so the factorization holds for both treatment arms. $\square$

Lemma 5. [`lem:proxy-coordinate-envelopes`] (Coordinate envelope bounds) *Let $P \in \mathcal{M}$ be a full-data probability law carrying $(U, T, X, Z, Y(0), Y(1), Y)$, with $\mathcal{M}$ as in Definition 32. Suppose that*

- ***(Latent cardinality.)*** $2 \leq k$.
- ***(Target-proxy dimension.)*** $k \leq d_x$.
- ***(Envelope conditions.)*** *P satisfies Assumptions 6 to 8.*
- ***(Anchor normalization.)*** *P satisfies Assumption 5.*

Then the model envelope L bounds the target proxy, the reference proxy, and the outcome–reference-proxy products P-almost surely:

$$\Pr_P\{\, |X_i| \leq L \text{ for every } 1 \leq i \leq d_x \,\} = 1,$$

$$\Pr_P\{\, |Z_j| \leq L \text{ for every } 1 \leq j \leq d_z \,\} = 1,$$

and

$$\Pr_P\{\, |Y Z_j| \leq L \text{ for every } 1 \leq j \leq d_z \,\} = 1.$$

Proof of Lemma 5. Since $2 \leq k \leq d_x$, the first target-proxy coordinate is available; write it as coordinate 1. We use the following coordinate bound: for any real $d_z \times d_x$ matrix A,

$$|A_{ji}| \leq \|A\|_{\mathrm{op}} \qquad (1 \leq j \leq d_z,\ 1 \leq i \leq d_x).$$

Indeed, with e_i the ith Euclidean basis vector,

$$|A_{ji}| = |(Ae_i)_j| \leq \|Ae_i\|_2 \leq \|A\|_{\mathrm{op}} \|e_i\|_2 = \|A\|_{\mathrm{op}}.$$

On the probability-one event from Assumption 6,

$$\|X\|_2 \leq L.$$

Hence, for every target-proxy coordinate i,

$$|X_i| \leq \|X\|_2 \leq L.$$

On the probability-one event where Assumptions 5 and 7 both hold,

$$X_1 = 1, \qquad \|ZX^\top\|_{\mathrm{op}} \leq L.$$

For every reference-proxy coordinate j, the $(j,1)$ entry of $ZX^\top$ is $Z_j X_1 = Z_j$, so the coordinate bound gives

$$|Z_j| = |(ZX^\top)_{j1}| \le \|ZX^\top\|_{\mathrm{op}} \le L.$$

On the probability-one event where Assumptions 5 and 8 both hold,

$$X_1 = 1, \qquad \|YZX^\top\|_{\mathrm{op}} \le L.$$

For every reference-proxy coordinate j, the $(j,1)$ entry of $YZX^\top$ is $YZ_jX_1 = YZ_j$. Therefore

$$|YZ_j| = |(YZX^\top)_{j1}| \le \|YZX^\top\|_{\mathrm{op}} \le L.$$

The three displayed coordinatewise bounds hold on probability-one events. Since the coordinate sets are finite, these are exactly

$$\Pr_P\{\, |X_i| \le L \text{ for every } 1 \le i \le d_x \,\} = 1,$$

$$\Pr_P\{\, |Z_j| \le L \text{ for every } 1 \le j \le d_z \,\} = 1,$$

and

$$\Pr_P\{\, |YZ_j| \le L \text{ for every } 1 \le j \le d_z \,\} = 1.$$

□

Lemma 6. [`lem:latent-mean-envelope`] (Latent mean envelope) *Let $P \in \mathcal{M}$ be a full-data probability law carrying $(U, T, X, Z, Y(0), Y(1), Y)$, as in Definition 32. Suppose that*

- ***(Latent cardinality.)*** $2 \le k$.
- ***(Proxy dimensions.)*** $k \le d_x$ *and* $k \le d_z$.
- ***(Model constants.)*** $1 \le L$, $\pi_0 > 0$, *and* $\sigma_0 > 0$.
- ***(Model components.)*** *The model class includes the bounded target-proxy condition of Assumption 6, the latent-arm positivity condition of Assumption 9, and the proxy-rank margin condition of Assumption 10.*

Then, for every latent class $u \in \{1, \ldots, k\}$ and every treatment state $t \in \{0, 1\}$, the latent conditional potential-outcome mean μ_{tu} of Definition 22 satisfies

$$|\mu_{tu}| \le \frac{L_\tau}{2},$$

where $L_\tau = 4L\sqrt{d_z}/\sigma_0$ is the derived effect-support radius from Definition 26.

Proof of Lemma 6. Fix $u \in \{1, \ldots, k\}$ and $t \in \{0, 1\}$. Define

$$\mathsf{Lat}_u := \{U = u\}, \qquad \mathsf{Cell}_{u,a} := \{U = u,\ T = a\} \quad (a \in \{0, 1\}).$$

1. The condition in Assumption 9, together with $\pi_0 > 0$, gives

$$P(\mathsf{Cell}_{u,a}) \geq \pi_0 > 0 \qquad \text{for each } a \in \{0,1\}.$$

   In particular $P(\mathsf{Cell}_{u,0}) > 0$. Since

$$\mathsf{Cell}_{u,0} \subseteq \mathsf{Lat}_u,$$

   we have

$$P(\mathsf{Lat}_u) > 0.$$

   Let $P_u := P(\,\cdot \mid U = u)$.

2. We first bound the observed outcome on the latent treatment cell. Put

$$\theta_z := \frac{\sigma_0}{2\sqrt{d_z}}.$$

   Since $2 \leq k \leq d_z$ gives $\sqrt{d_z} > 0$, and since $\sigma_0 > 0$, we have $\theta_z > 0$. The rank condition in Assumption 10 gives

$$\sigma_0 \leq \sigma_k(A_t).$$

   Applying the least-singular-value lower bound to the uth canonical vector gives

$$\sigma_0 \leq \left(\sum_{j=1}^{d_z} A_t(j,u)^2\right)^{1/2}.$$

   Hence there exists $j_\star \in \{1, \ldots, d_z\}$ such that

$$\frac{\sigma_0}{\sqrt{d_z}} \leq |A_t(j_\star, u)|.$$

   Let

$$P_{u,t}^{\text{cell}} := P(\,\cdot \mid U = u, T = t), \qquad \mathsf{Ref}_{j_\star} := \{|Z_{j_\star}| \geq \theta_z\}.$$

   Then

$$P_{u,t}^{\text{cell}}(\mathsf{Ref}_{j_\star}) > 0.$$

   Indeed, if this probability were zero, then

$$|A_t(j_\star, u)| = |E_P[Z_{j_\star} \mid U = u, T = t]| \leq E_P[|Z_{j_\star}| \mid U = u, T = t] \leq \theta_z < \frac{\sigma_0}{\sqrt{d_z}},$$

   contradicting the displayed coordinate lower bound.

3. Under $P_{u,t}^{\text{cell}}$, Assumption 1 makes $Z_{j_\star}$ independent of Y. Also, Lemma 5 supplies

$$|YZ_{j_\star}| \leq L \quad \text{almost surely under } P_{u,t}^{\text{cell}}.$$

   Define

$$\mathsf{Out}_{j_\star} := \left\{|Y| > \frac{L}{\theta_z}\right\}.$$

   On $\mathsf{Ref}_{j_\star} \cap \mathsf{Out}_{j_\star}$,

$$|YZ_{j_\star}| > L,$$

and therefore

$$P^{\text{cell}}_{u,t}(\mathsf{Ref}_{j_\star} \cap \mathsf{Out}_{j_\star}) = 0.$$

Independence gives

$$0 = P^{\text{cell}}_{u,t}(\mathsf{Ref}_{j_\star})P^{\text{cell}}_{u,t}(\mathsf{Out}_{j_\star}).$$

Since the first factor is positive,

$$P^{\text{cell}}_{u,t}(\mathsf{Out}_{j_\star}) = 0.$$

Thus

$$|Y| \le \frac{L}{\theta_z} = \frac{2L\sqrt{d_z}}{\sigma_0} = \frac{L_\tau}{2} \quad \text{almost surely under } P(\,\cdot \mid U = u, T = t),$$

where the last equality uses Definition 26.

4. Now work under $P_u = P(\,\cdot \mid U = u)$, and define

$$\mathsf{Arm}_t := \{T = t\}.$$

The preceding positivity gives

$$P_u(\mathsf{Arm}_t) = \frac{P(\mathsf{Cell}_{u,t})}{P(\mathsf{Lat}_u)} > 0.$$

By Assumption 4, $Y(t)$ is independent of T, hence independent of Arm_t, under P_u. On Arm_t, Assumption 3 gives $Y = Y(t)$ almost surely, and the cell bound from the previous step gives

$$|Y(t)| \le \frac{L_\tau}{2} \quad \text{almost surely on } \mathsf{Arm}_t \text{ under } P_u.$$

With

$$\mathsf{PotLarge}_t := \left\{|Y(t)| > \frac{L_\tau}{2}\right\},$$

we therefore have

$$P_u(\mathsf{PotLarge}_t \cap \mathsf{Arm}_t) = 0.$$

Independence gives

$$0 = P_u(\mathsf{PotLarge}_t)P_u(\mathsf{Arm}_t),$$

and the positive arm probability yields

$$P_u(\mathsf{PotLarge}_t) = 0.$$

Consequently

$$|Y(t)| \le \frac{L_\tau}{2} \quad \text{almost surely under } P_u.$$

Finally, by Definition 22,

$$\mu_{tu} = E_{P_u}[Y(t)].$$

The almost-sure bound just proved gives integrability and

$$|\mu_{tu}| = |E_{P_u}[Y(t)]| \le E_{P_u}[|Y(t)|] \le \frac{L_\tau}{2}.$$

This is the desired bound.

□

Lemma 7. [`lem:latent-effect-envelope`] (Latent effect envelope) *Let $k, d_x, d_z \in \mathbb{N}$, let $L, \pi_0, \sigma_0 \in \mathbb{R}$, let P be a full-data probability law carrying $(U, T, X, Z, Y(0), Y(1), Y)$, and fix a latent class $u \in \{1, \ldots, k\}$. Suppose that*

- ***(Dimensions.)*** *$2 \le k$, $k \le d_x$, and $k \le d_z$.*
- ***(Margins.)*** *$1 \le L$, $\pi_0 > 0$, and $\sigma_0 > 0$.*
- ***(Model membership.)*** *$P \in \mathcal{M}$ for the parameters $(k, d_x, d_z, L, \pi_0, \sigma_0)$, as in Definition 32, with the bounded-proxy, latent-arm-positivity, and proxy-rank-margin components recorded in Assumptions 6, 9 and 10.*

Then the latent-class treatment effect τ_u, formed from the latent means in Definitions 22 and 23, satisfies

$$|\tau_u| \le L_\tau = \frac{4L\sqrt{d_z}}{\sigma_0},$$

where L_τ is the derived effect-support radius in Definition 26.

Proof of Lemma 7. Fix u. By Definition 23,

$$\tau_u = \mu_{1u} - \mu_{0u}.$$

Thus the triangle inequality gives

$$|\tau_u| = |\mu_{1u} - \mu_{0u}| \le |\mu_{1u}| + |\mu_{0u}|.$$

Applying Lemma 6 to the two treatment states,

$$|\mu_{1u}| \le \frac{L_\tau}{2}, \qquad |\mu_{0u}| \le \frac{L_\tau}{2}.$$

Therefore

$$|\tau_u| \le \frac{L_\tau}{2} + \frac{L_\tau}{2} = L_\tau = \frac{4L\sqrt{d_z}}{\sigma_0}.$$

□

Lemma 8. [`lem:witness-latent-mass`] (Witness latent masses) *For every collision-witness effect displacement $\varepsilon \in \mathbb{R}$ with $0 \le \varepsilon \le 1/8$, let $P_\varepsilon^{\mathrm{wit}}$ be the two-class witness law of Definition 55. The latent-class masses $p_u := P_\varepsilon^{\mathrm{wit}}(U = u)$ from Definition 21 are*

$$p_1 = \frac{2}{5}, \qquad p_2 = \frac{3}{5}.$$

Equivalently, these masses are constant in ε over this range and are distinct.

Proof of Lemma 8. Fix ε with $0 \le \varepsilon \le 1/8$. For a Bernoulli success parameter $\theta \in \mathbb{R}$ and $b \in \{0, 1\}$, put

$$\beta_\theta(b) := \begin{cases} 1 - \theta, & b = 0, \\ \theta, & b = 1. \end{cases}$$

For the two witness classes in the paper labels, define

$$\mathfrak{p}_{\mathrm{U}}(1)=\frac{2}{5},\qquad \mathfrak{p}_{\mathrm{U}}(2)=\frac{3}{5},\qquad \mathfrak{p}_{\mathrm{T}}(1)=\frac{2}{5},\qquad \mathfrak{p}_{\mathrm{T}}(2)=\frac{3}{5},$$
$$\mathfrak{p}_{\mathrm{X}}(1)=\frac{1}{5},\qquad \mathfrak{p}_{\mathrm{X}}(2)=\frac{4}{5},$$
$$\mathfrak{p}_{\mathrm{Z}}(0,1)=\frac{3}{10},\quad \mathfrak{p}_{\mathrm{Z}}(0,2)=\frac{7}{10},\quad \mathfrak{p}_{\mathrm{Z}}(1,1)=\frac{7}{20},\quad \mathfrak{p}_{\mathrm{Z}}(1,2)=\frac{3}{4},$$

and

$$\mathfrak{p}_{\mathrm{Y}_1,\varepsilon}(1)=\frac{1}{2}-\varepsilon,\qquad \mathfrak{p}_{\mathrm{Y}_1,\varepsilon}(2)=\frac{1}{2}+\varepsilon.$$

Thus Definition 55 assigns to the atom with

$$(U,T,X_2,Z_2,Y(0),Y(1))=(v,t_{\mathrm{b}},x_{\mathrm{b}},z_{\mathrm{b}},y_{\mathrm{b},0},y_{\mathrm{b},1})$$

the elementary mass

$$\mathsf{wt}_\varepsilon(v,t_{\mathrm{b}},x_{\mathrm{b}},z_{\mathrm{b}},y_{\mathrm{b},0},y_{\mathrm{b},1})=\mathfrak{p}_{\mathrm{U}}(v)\,\beta_{\mathfrak{p}_{\mathrm{T}}(v)}(t_{\mathrm{b}})\,\beta_{\mathfrak{p}_{\mathrm{X}}(v)}(x_{\mathrm{b}})\,\beta_{\mathfrak{p}_{\mathrm{Z}}(t_{\mathrm{b}},v)}(z_{\mathrm{b}})\,\beta_{1/4}(y_{\mathrm{b},0})\,\beta_{\mathfrak{p}_{\mathrm{Y}_1,\varepsilon}(v)}(y_{\mathrm{b},1}).$$

The bounds on ε place every Bernoulli parameter above in $[0,1]$, so these elementary masses are nonnegative. By Definition 21 and the finite witness-law expansion, for $u\in\{1,2\}$,

$$p_u(P_\varepsilon^{\mathrm{wit}})=\sum_{v=1}^{2}\ \sum_{t_{\mathrm{b}},x_{\mathrm{b}},z_{\mathrm{b}},y_{\mathrm{b},0},y_{\mathrm{b},1}\in\{0,1\}}\mathsf{wt}_\varepsilon(v,t_{\mathrm{b}},x_{\mathrm{b}},z_{\mathrm{b}},y_{\mathrm{b},0},y_{\mathrm{b},1})\,\mathbf{1}\{v=u\}.$$

The indicator restricts the finite sum to the $v=u$ slice, hence

$$p_u(P_\varepsilon^{\mathrm{wit}})=\sum_{t_{\mathrm{b}},x_{\mathrm{b}},z_{\mathrm{b}},y_{\mathrm{b},0},y_{\mathrm{b},1}\in\{0,1\}}\mathsf{wt}_\varepsilon(u,t_{\mathrm{b}},x_{\mathrm{b}},z_{\mathrm{b}},y_{\mathrm{b},0},y_{\mathrm{b},1}).$$

For fixed u and t_{b}, summing over the nuisance Bernoulli coordinates gives

$$\begin{aligned}&\sum_{x_{\mathrm{b}},z_{\mathrm{b}},y_{\mathrm{b},0},y_{\mathrm{b},1}\in\{0,1\}}\mathsf{wt}_\varepsilon(u,t_{\mathrm{b}},x_{\mathrm{b}},z_{\mathrm{b}},y_{\mathrm{b},0},y_{\mathrm{b},1})\\&=\mathfrak{p}_{\mathrm{U}}(u)\,\beta_{\mathfrak{p}_{\mathrm{T}}(u)}(t_{\mathrm{b}})\left(\sum_{x_{\mathrm{b}}}\beta_{\mathfrak{p}_{\mathrm{X}}(u)}(x_{\mathrm{b}})\right)\left(\sum_{z_{\mathrm{b}}}\beta_{\mathfrak{p}_{\mathrm{Z}}(t_{\mathrm{b}},u)}(z_{\mathrm{b}})\right)\left(\sum_{y_{\mathrm{b},0}}\beta_{1/4}(y_{\mathrm{b},0})\right)\left(\sum_{y_{\mathrm{b},1}}\beta_{\mathfrak{p}_{\mathrm{Y}_1,\varepsilon}(u)}(y_{\mathrm{b},1})\right)\\&=\mathfrak{p}_{\mathrm{U}}(u)\,\beta_{\mathfrak{p}_{\mathrm{T}}(u)}(t_{\mathrm{b}}),\end{aligned}$$

because $\sum_{b\in\{0,1\}}\beta_\theta(b)=(1-\theta)+\theta=1$.

Therefore

$$p_u(P_\varepsilon^{\mathrm{wit}})=\sum_{t_{\mathrm{b}}\in\{0,1\}}\mathfrak{p}_{\mathrm{U}}(u)\,\beta_{\mathfrak{p}_{\mathrm{T}}(u)}(t_{\mathrm{b}})=\mathfrak{p}_{\mathrm{U}}(u)\,\{(1-\mathfrak{p}_{\mathrm{T}}(u))+\mathfrak{p}_{\mathrm{T}}(u)\}=\mathfrak{p}_{\mathrm{U}}(u).$$

Evaluating this identity at the two paper-labeled classes yields

$$p_1(P_\varepsilon^{\mathrm{wit}})=\mathfrak{p}_{\mathrm{U}}(1)=\frac{2}{5},\qquad p_2(P_\varepsilon^{\mathrm{wit}})=\mathfrak{p}_{\mathrm{U}}(2)=\frac{3}{5}.$$

The displayed values are independent of ε, and they are distinct since

$$\frac{2}{5}\neq\frac{3}{5}.$$

□

Together these factorization and envelope lemmas give the deterministic bridge from the conditional-independence restrictions to the matrices used in the compressed operator. The latent mean and effect envelopes place every quotient law in the common support interval fixed by Definition 26.

The following algebraic facts are used to pass from factored proxy moments to the compressed spectral representation.

Lemma 9. [`lem:witness-latent-effect`] (Witness latent effects) *For every collision-witness effect displacement $\varepsilon \in \mathbb{R}$ with $0 \le \varepsilon \le 1/8$ and every latent class $u \in \{1,2\}$, the latent-class treatment effect τ_u from Definition 23 under the two-class Bernoulli witness law $P_\varepsilon^{\mathrm{wit}}$ from Definition 55 is*

$$\tau_u(P_\varepsilon^{\mathrm{wit}}) = \begin{cases} \frac{1}{4} - \varepsilon, & u = 1, \\ \frac{1}{4} + \varepsilon, & u = 2. \end{cases}$$

Proof of Lemma 9. Fix $0 \le \varepsilon \le 1/8$. For a Boolean variable b, write

$$\beta_p(b) := \begin{cases} p, & b = 1, \\ 1 - p, & b = 0, \end{cases} \qquad b^{\mathbb{R}} := \begin{cases} 1, & b = 1, \\ 0, & b = 0. \end{cases}$$

For latent class $r \in \{1,2\}$, put

$$p_r := \begin{cases} 2/5, & r = 1, \\ 3/5, & r = 2, \end{cases} \qquad q_r := \begin{cases} 1/2 - \varepsilon, & r = 1, \\ 1/2 + \varepsilon, & r = 2. \end{cases}$$

The range of ε puts all Bernoulli parameters used below in $[0,1]$, so the finite weights of $P_\varepsilon^{\mathrm{wit}}$ are nonnegative.

The elementary mass assigned to the point indexed by $(r, t, x, z, y_0, y_1) \in \{1,2\} \times \{0,1\}^5$ is

$$w_r(t, x, z, y_0, y_1) = p_r\, \beta_{p_r}(t)\, \beta_{b_r}(x)\, \beta_{a_{t,r}}(z)\, \beta_{1/4}(y_0)\, \beta_{q_r}(y_1),$$

where $b_1 = 1/5$, $b_2 = 4/5$, and $a_{t,r}$ is the displayed Z_2-mean from Definition 55. At this point,

$$Y(0) = y_0^{\mathbb{R}}, \qquad Y(1) = y_1^{\mathbb{R}}.$$

For either $s \in \{0,1\}$, the latent conditional mean from Definition 22 is the finite conditional average

$$\mu_{sr}(P_\varepsilon^{\mathrm{wit}}) = \left(\sum_{t,x,z,y_0,y_1} w_r(t,x,z,y_0,y_1) \right)^{-1} \sum_{t,x,z,y_0,y_1} w_r(t,x,z,y_0,y_1)\, Y(s).$$

The denominator is

$$\sum_{t,x,z,y_0,y_1} w_r(t,x,z,y_0,y_1) = p_r,$$

because each Bernoulli factor sums to one over its binary coordinate.

For $s = 0$,

$$\sum_{t,x,z,y_0,y_1} w_r(t,x,z,y_0,y_1) y_0^{\mathbb{R}} = p_r \sum_{y_0 \in \{0,1\}} \beta_{1/4}(y_0) y_0^{\mathbb{R}} = p_r/4,$$

so

$$\mu_{0r}(P_\varepsilon^{\mathrm{wit}}) = 1/4.$$

For $s = 1$,

$$\sum_{t,x,z,y_0,y_1} w_r(t,x,z,y_0,y_1) y_1^{\mathbb{R}} = p_r \sum_{y_1 \in \{0,1\}} \beta_{q_r}(y_1) y_1^{\mathbb{R}} = p_r q_r,$$

and hence

$$\mu_{1r}(P_\varepsilon^{\mathrm{wit}}) = q_r.$$

Using the definition of the latent-class treatment effect in Definition 23,

$$\tau_r(P_\varepsilon^{\mathrm{wit}}) = \mu_{1r}(P_\varepsilon^{\mathrm{wit}}) - \mu_{0r}(P_\varepsilon^{\mathrm{wit}}) = q_r - \frac{1}{4}.$$

Substituting the two values of q_r gives

$$\tau_1(P_\varepsilon^{\mathrm{wit}}) = \left(\frac{1}{2} - \varepsilon\right) - \frac{1}{4} = \frac{1}{4} - \varepsilon, \qquad \tau_2(P_\varepsilon^{\mathrm{wit}}) = \left(\frac{1}{2} + \varepsilon\right) - \frac{1}{4} = \frac{1}{4} + \varepsilon.$$

□

Lemma 10. [`lem:witness-observed-arm-mass`] (Observed arm mass) *For any collision-witness effect displacement ε and any treatment arm $t \in \{0,1\}$, suppose that*

- ***(Lower bound.)*** $0 \le \varepsilon$.
- ***(Upper bound.)*** $\varepsilon \le 1/8$.

For the two-class Bernoulli proxy witness law $P_\varepsilon^{\mathrm{wit}}$ from Definition 55, its observed-data marginal from Definition 20 satisfies

$$(P_\varepsilon^{\mathrm{wit}})_O(T = t) = \begin{cases} 12/25, & t = 0, \\ 13/25, & t = 1. \end{cases}$$

Proof of Lemma 10. Fix ε with $0 \le \varepsilon \le 1/8$ and fix $t \in \{0,1\}$.

1. Define the observed arm set

   $$\mathsf{Arm}_t := \{(T,X,Z,Y) \in O_{2,2} : T = t\}.$$

   For a full-data record w, let $O(w) = (T(w), X(w), Z(w), Y(w))$. Then

   $$O^{-1}(\mathsf{Arm}_t) = \{w : T(w) = t\}.$$

   The coordinate map O is measurable and Arm_t is measurable, so $\{w : T(w) = t\}$ is measurable. By Definition 20,

   $$(P_\varepsilon^{\mathrm{wit}})_O(\mathsf{Arm}_t) = P_\varepsilon^{\mathrm{wit}}\{w : T(w) = t\}.$$

2. For $q \in \mathbb{R}$ and $r \in \{0, 1\}$, put

$$\beta(q, r) := \begin{cases} q, & r = 1, \\ 1 - q, & r = 0. \end{cases}$$

The finite witness law in Definition 55 assigns to the atom indexed by

$$u \in \{1, 2\}, \qquad s, x, z, y_0, y_1 \in \{0, 1\}$$

the real weight

$$\mathfrak{w}_\varepsilon(u, s, x, z, y_0, y_1) = \mathfrak{p}_u\, \beta(\mathfrak{p}_u, s)\, \beta(\mathfrak{x}_u, x)\, \beta(\mathfrak{z}_{s,u}, z)\, \beta(1/4, y_0)\, \beta(\mathfrak{y}_{u,\varepsilon}, y_1),$$

where

$$\mathfrak{p}_1 = \frac{2}{5}, \quad \mathfrak{p}_2 = \frac{3}{5}, \quad \mathfrak{x}_1 = \frac{1}{5}, \quad \mathfrak{x}_2 = \frac{4}{5},$$

$$\mathfrak{z}_{1,1} = \frac{7}{20}, \quad \mathfrak{z}_{1,2} = \frac{3}{4}, \quad \mathfrak{z}_{0,1} = \frac{3}{10}, \quad \mathfrak{z}_{0,2} = \frac{7}{10},$$

and

$$\mathfrak{y}_{1,\varepsilon} = \frac{1}{2} - \varepsilon, \qquad \mathfrak{y}_{2,\varepsilon} = \frac{1}{2} + \varepsilon.$$

The bounds on ε give

$$0 \le \mathfrak{y}_{u,\varepsilon} \le 1 \qquad (u = 1, 2),$$

and all the other displayed Bernoulli parameters also lie in $[0, 1]$. Hence every $\mathfrak{w}_\varepsilon(u, s, x, z, y_0, y_1)$ is nonnegative. Therefore the real mass of the full-data arm event is the ordinary finite sum

$$P_\varepsilon^{\mathrm{wit}}\{w : T(w) = t\} = \sum_{u=1}^{2} \sum_{s,x,z,y_0,y_1 \in \{0,1\}} \mathfrak{w}_\varepsilon(u, s, x, z, y_0, y_1)\, \mathbf{1}\{s = t\}.$$

3. Selecting the treatment coordinate in the finite sum gives

$$\sum_{u=1}^{2} \sum_{s,x,z,y_0,y_1 \in \{0,1\}} \mathfrak{w}_\varepsilon(u, s, x, z, y_0, y_1)\, \mathbf{1}\{s = t\} = \sum_{u=1}^{2} \sum_{x,z,y_0,y_1 \in \{0,1\}} \mathfrak{w}_\varepsilon(u, t, x, z, y_0, y_1).$$

This is the two-case calculation $t = 0$ and $t = 1$: exactly one value of s has indicator one.

4. For each fixed $u \in \{1, 2\}$ and $t \in \{0, 1\}$,

$$\sum_{x,z,y_0,y_1 \in \{0,1\}} \mathfrak{w}_\varepsilon(u, t, x, z, y_0, y_1)$$
$$= \mathfrak{p}_u \beta(\mathfrak{p}_u, t) \left(\sum_{x \in \{0,1\}} \beta(\mathfrak{x}_u, x) \right) \left(\sum_{z \in \{0,1\}} \beta(\mathfrak{z}_{t,u}, z) \right) \left(\sum_{y_0 \in \{0,1\}} \beta(1/4, y_0) \right) \left(\sum_{y_1 \in \{0,1\}} \beta(\mathfrak{y}_{u,\varepsilon}, y_1) \right).$$

Since $\beta(q, 0) + \beta(q, 1) = 1$ for every q, the four nuisance factors sum to one, and hence

$$\sum_{x,z,y_0,y_1 \in \{0,1\}} \mathfrak{w}_\varepsilon(u, t, x, z, y_0, y_1) = \mathfrak{p}_u \beta(\mathfrak{p}_u, t).$$

5. Combining the preceding identities,

$$(P_\varepsilon^{\mathrm{wit}})_O(T=t) = \sum_{u=1}^{2} \mathfrak{p}_u \beta(\mathfrak{p}_u, t).$$

If $t=0$, then

$$\sum_{u=1}^{2} \mathfrak{p}_u \beta(\mathfrak{p}_u, 0) = \frac{2}{5}\frac{3}{5} + \frac{3}{5}\frac{2}{5} = \frac{12}{25}.$$

If $t=1$, then

$$\sum_{u=1}^{2} \mathfrak{p}_u \beta(\mathfrak{p}_u, 1) = \frac{2}{5}\frac{2}{5} + \frac{3}{5}\frac{3}{5} = \frac{13}{25}.$$

Thus

$$(P_\varepsilon^{\mathrm{wit}})_O(T=t) = \begin{cases} 12/25, & t=0, \\ 13/25, & t=1. \end{cases}$$

□

Lemma 11. [`lem:witness-target-feature-determinant`] (Witness feature determinant) *For every collision-witness effect displacement $\varepsilon \in \mathbb{R}$ satisfying $0 \le \varepsilon \le 1/8$, let $P_\varepsilon^{\mathrm{wit}}$ be the two-class Bernoulli proxy witness law from Definition 55, and let B be its target-proxy feature matrix from Definition 25. Then*

$$B_{00}B_{11} - B_{01}B_{10} = \frac{3}{5}.$$

Proof of Lemma 11. Fix $\varepsilon \in \mathbb{R}$ with $0 \le \varepsilon \le 1/8$. By Definitions 25 and 55, the entries of the target-proxy feature matrix are

$$B_{iu} = E_{P_\varepsilon^{\mathrm{wit}}}[X_i \mid U=u].$$

Since $X = (1, X_2)^\top$, with conditional means $E[X_2 \mid U=1] = 1/5$ and $E[X_2 \mid U=2] = 4/5$, this gives, in the statement's zero-based coordinate order,

$$B = \begin{pmatrix} 1 & 1 \\ 1/5 & 4/5 \end{pmatrix}.$$

Therefore

$$B_{00}B_{11} - B_{01}B_{10} = 1 \cdot \frac{4}{5} - 1 \cdot \frac{1}{5} = \frac{3}{5}.$$

□

Lemma 12. [`lem:witness-reference-feature-determinant`] (Witness reference-feature determinant) *For the two-class Bernoulli witness law $P_\varepsilon^{\mathrm{wit}}$ and the reference-proxy feature matrix A_t in Definitions 24 and 55, suppose*

- ***(Perturbation range.)*** *The collision-witness effect displacement ε satisfies $0 \le \varepsilon \le 1/8$.*
- ***(Arm.)*** *The treatment arm t is binary, $t \in \{0,1\}$.*

Then

$$\det(A_t) = \frac{2}{5}.$$

Proof of Lemma 12. By Definitions 24 and 55, the entry $A_t(r,u)$ is the conditional mean of the rth coordinate of $Z = (1, Z_2)^\top$ in latent class $u \in \{1,2\}$ and arm t. The first coordinate of Z is identically one. Conditional on (U,T), the variable Z_2 is Bernoulli with the displayed conditional means, while the remaining Bernoulli coordinates are nuisance factors whose conditional masses sum to one. Thus

$$A_0 = \begin{pmatrix} 1 & 1 \\ 3/10 & 7/10 \end{pmatrix}, \qquad A_1 = \begin{pmatrix} 1 & 1 \\ 7/20 & 3/4 \end{pmatrix}.$$

Equivalently,

$$A_t(1,u) = 1, \qquad A_0(2,1) = \frac{3}{10}, \quad A_0(2,2) = \frac{7}{10}, \quad A_1(2,1) = \frac{7}{20}, \quad A_1(2,2) = \frac{3}{4}.$$

The assumptions $0 \le \varepsilon \le 1/8$ ensure that these conditioning cells have positive mass, so the displayed conditional means are the ordinary finite conditional averages.

For a 2×2 matrix C, write

$$\det(C) = C_{11}C_{22} - C_{12}C_{21}.$$

There are two arm cases. If $t = 0$, then

$$\det(A_0) = 1 \cdot \frac{7}{10} - 1 \cdot \frac{3}{10} = \frac{2}{5}.$$

If $t = 1$, then

$$\det(A_1) = 1 \cdot \frac{3}{4} - 1 \cdot \frac{7}{20} = \frac{15}{20} - \frac{7}{20} = \frac{2}{5}.$$

Since the arm assumption is exactly $t \in \{0,1\}$, these two computations give $\det(A_t) = 2/5$. ☐

Lemma 13. [`lem:path-proxy-rank-margin`] (Path proxy-rank margin) *Let $P_{g,h}$ be the factorization-preserving two-class path law of Definition 51. Suppose that*

- ***(Gap range.)*** $0 \le g \le 1/4$.
- ***(Tangent amplitude.)*** $|h| \le 1/100$.

Then $P_{g,h}$ satisfies the proxy-rank margin in Assumption 10 with proxy-rank singular-value margin $\sigma_0 = 1/10$: the reference-proxy feature matrices A_0 and A_1 from Definition 24, and the target-proxy feature matrix B from Definition 25, all formed from $P_{g,h}$, have smallest singular value at least $1/10$.

Proof of Lemma 13.

1. Definitions 24, 25 and 50 give the feature matrices of $P_{g,h}$. Under $|h| \le 1/100$, the denominators below are positive, and

$$A_0(h) = \begin{pmatrix} 1 & 1 \\ \dfrac{35h+9}{10(5h+3)} & \dfrac{7}{10} \end{pmatrix}, \qquad A_1(h) = \begin{pmatrix} 1 & 1 \\ \dfrac{225h+28}{20(15h+4)} & \dfrac{3}{4} \end{pmatrix},$$

while

$$B(h) = \begin{pmatrix} 1 & 1 \\ \dfrac{1}{5} & \dfrac{12/25 - h/5}{3/5 - h} \end{pmatrix}.$$

Thus the matrices A_0, A_1, B in Assumption 10, formed from $P_{g,h}$, are $A_0(h), A_1(h), B(h)$, respectively.

2. The undisplaced reference features have singular-value slack $1/5$. At $h = 0$,

$$A_0(0) = \begin{pmatrix} 1 & 1 \\ 3/10 & 7/10 \end{pmatrix}, \qquad A_1(0) = \begin{pmatrix} 1 & 1 \\ 7/20 & 3/4 \end{pmatrix}.$$

For every $v = (v_1, v_2)^\top \in \mathbb{R}^2$,

$$100\left(\|A_0(0)v\|_2^2 - \frac{1}{25}\|v\|_2^2\right) = 105v_1^2 + 242v_1v_2 + 145v_2^2 = \frac{146(3v_1 - 7v_2)^2 + (549v_1 + 641v_2)^2}{2883} \geq 0,$$

and

$$400\left(\|A_1(0)v\|_2^2 - \frac{1}{25}\|v\|_2^2\right) = 433v_1^2 + 1010v_1v_2 + 609v_2^2 = \frac{2168(7v_1 - 15v_2)^2 + (5015v_1 + 5919v_2)^2}{58329} \geq 0.$$

The variational characterization of the least singular value therefore gives

$$\sigma_{\min}(A_0(0)) \geq \frac{1}{5}, \qquad \sigma_{\min}(A_1(0)) \geq \frac{1}{5}.$$

3. Define the reference-feature entry displacements

$$\chi_0(h) := \frac{35h + 9}{10(5h + 3)} - \frac{3}{10}, \qquad \chi_1(h) := \frac{225h + 28}{20(15h + 4)} - \frac{7}{20}.$$

Then

$$A_0(h) = A_0(0) + \begin{pmatrix} 0 & 0 \\ \chi_0(h) & 0 \end{pmatrix}, \qquad A_1(h) = A_1(0) + \begin{pmatrix} 0 & 0 \\ \chi_1(h) & 0 \end{pmatrix}.$$

Since $|h| \leq 1/100$,

$$10(5h + 3) \geq \frac{59}{2}, \qquad 20(15h + 4) \geq 77,$$

and the algebraic identities

$$\chi_0(h) = \frac{20h}{10(5h + 3)}, \qquad \chi_1(h) = \frac{120h}{20(15h + 4)}$$

give

$$|\chi_0(h)| \leq \frac{20/100}{59/2} = \frac{2}{295} \leq \frac{1}{50}, \qquad |\chi_1(h)| \leq \frac{120/100}{77} = \frac{6}{385} \leq \frac{1}{50}.$$

For any real χ,

$$\left\|\begin{pmatrix} 0 & 0 \\ \chi & 0 \end{pmatrix}\right\|_{\mathrm{op}} \leq |\chi|,$$

because its action sends v to $(0, \chi v_1)^\top$. Weyl's singular-value perturbation inequality therefore yields, for $t = 0, 1$,

$$\sigma_{\min}(A_t(h)) \geq \sigma_{\min}(A_t(0)) - \frac{1}{50} \geq \frac{1}{5} - \frac{1}{50} \geq \frac{1}{10}.$$

4. The target feature is handled in the same perturbative form, starting from

$$B(0) = \begin{pmatrix} 1 & 1 \\ 1/5 & 4/5 \end{pmatrix}.$$

For every $v \in \mathbb{R}^2$,

$$25\left(\|B(0)v\|_2^2 - \frac{1}{25}\|v\|_2^2\right) = 25v_1^2 + 58v_1v_2 + 40v_2^2 = \frac{159(3v_1 - 4v_2)^2 + (187v_1 + 236v_2)^2}{1456} \geq 0,$$

so

$$\sigma_{\min}(B(0)) \geq \frac{1}{5}.$$

With

$$\chi_B(h) := \frac{12/25 - h/5}{3/5 - h} - \frac{4}{5},$$

we have

$$B(h) = B(0) + \begin{pmatrix} 0 & 0 \\ 0 & \chi_B(h) \end{pmatrix}, \qquad \chi_B(h) = \frac{(3/5)h}{3/5 - h}.$$

The bound $|h| \leq 1/100$ gives $3/5 - h \geq 59/100$, hence

$$|\chi_B(h)| \leq \frac{(3/5)(1/100)}{59/100} = \frac{3}{295} \leq \frac{1}{50}.$$

Also

$$\left\| \begin{pmatrix} 0 & 0 \\ 0 & \chi \end{pmatrix} \right\|_{\mathrm{op}} \leq |\chi|$$

for any real χ, since its action sends v to $(0, \chi v_2)^\top$. Another application of Weyl's inequality gives

$$\sigma_{\min}(B(h)) \geq \sigma_{\min}(B(0)) - \frac{1}{50} \geq \frac{1}{5} - \frac{1}{50} \geq \frac{1}{10}.$$

5. Combining the three estimates,

$$\min\{\sigma_{\min}(A_0), \sigma_{\min}(A_1), \sigma_{\min}(B)\} = \min\{\sigma_{\min}(A_0(h)), \sigma_{\min}(A_1(h)), \sigma_{\min}(B(h))\} \geq \frac{1}{10}.$$

This is the proxy-rank margin of Assumption 10 with $\sigma_0 = 1/10$.

□

These linear-algebra statements justify the rank thresholding and anchored diagonalization used by the lattice estimator. The condition-number bound provides an explicit finite-dimensional constant for perturbing between feature coordinates and the compressed spectral coordinates.

The Wasserstein comparisons use a one-dimensional dual potential and the coordinate relation between the Euclidean summary metric and d_S.

Lemma 14. [`lem:path-model-membership`] (Path model membership) *Let $g, h \in \mathbb{R}$. Suppose that*

- ***(Gap range.)*** $0 \leq g \leq 1/4$.

- ***(Path displacement.)*** $|h| \le 1/100$.

Then the path law $P_{g,h}$ of Definition 51 belongs to the uniformly conditioned model class $\mathcal{M}$ of Definition 32 with

$$k = d_x = d_z = 2, \qquad L = 2, \qquad \pi_0 = 1/10, \qquad \sigma_0 = 1/10.$$

Equivalently, $P_{g,h}$ satisfies the core numerical domain for $k = d_x = d_z = 2$, together with Assumptions 1 to 10, using latent-arm positivity margin $1/10$ and proxy-rank margin $1/10$.

Proof of Lemma 14. 1. Write the two latent classes as $u = 1, 2$, and write $t, x, z, y_0, y_1 \in \{0, 1\}$ for the binary coordinates in the finite support of the path. Put

$$p_1 = \frac{2}{5} + h, \qquad p_2 = \frac{3}{5} - h,$$

and

$$\begin{array}{c|cc} & u=1 & u=2 \\ \hline w_1(u) & \frac{4}{25} + \frac{3h}{5} & \frac{9}{25} - \frac{3h}{5} \\ w_0(u) & \frac{6}{25} + \frac{2h}{5} & \frac{6}{25} - \frac{2h}{5}. \end{array}$$

For a Bernoulli coordinate $r \in \{0, 1\}$ and success probability θ, set

$$\operatorname{bm}(r;\theta) = \begin{cases} 1-\theta, & r = 0, \\ \theta, & r = 1. \end{cases}$$

The path assigns the atom (u, t, x, z, y_0, y_1) mass

$$p_u \operatorname{bm}\left(t; \frac{w_1(u)}{p_u}\right) \operatorname{bm}(x; \beta_u) \operatorname{bm}(z; \alpha_{t,u}) \operatorname{bm}\left(y_0; \frac{1}{4}\right) \operatorname{bm}(y_1; \eta_u),$$

where

$$\beta_1 = \frac{1}{5}, \qquad \beta_2 = \frac{12/25 - h/5}{3/5 - h},$$

$$\alpha_{1,1} = \frac{225h + 28}{20(15h + 4)}, \quad \alpha_{1,2} = \frac{3}{4}, \quad \alpha_{0,1} = \frac{35h + 9}{10(5h + 3)}, \quad \alpha_{0,2} = \frac{7}{10},$$

and

$$\eta_1 = \frac{1}{2} - \frac{g}{2}, \qquad \eta_2 = \frac{1}{2} + \frac{g}{2}.$$

Since $|h| \le 1/100$,

$$\frac{2}{5} + h \ge \frac{39}{100}, \qquad \frac{3}{5} - h \ge \frac{59}{100}, \qquad 15h + 4 \ge \frac{77}{20}, \qquad 5h + 3 \ge \frac{59}{20}.$$

A direct check of the two latent classes gives

$$0 \le \frac{w_1(u)}{p_u} \le 1, \qquad 0 \le \beta_u \le 1, \qquad 0 \le \alpha_{t,u} \le 1,$$

and $0 \le \eta_u \le 1$ follows from $0 \le g \le 1/4$. Thus every displayed atom mass is nonnegative. Summing first over the Bernoulli success and failure values gives

$$\sum_{u,t,x,z,y_0,y_1} p_u \operatorname{bm}\left(t; \frac{w_1(u)}{p_u}\right) \operatorname{bm}(x; \beta_u) \operatorname{bm}(z; \alpha_{t,u}) \operatorname{bm}\left(y_0; \frac{1}{4}\right) \operatorname{bm}(y_1; \eta_u) = p_1 + p_2 = 1.$$

Hence $P_{g,h}$ is a probability law on the full-data record space with $k = d_x = d_z = 2$.

2. The numerical core domain in Definition 32 is

$$2 \le 2, \qquad 2 \le 2, \qquad 2 \le 2, \qquad 1 \le 2, \qquad 0 < \frac{1}{10} \le \frac{1}{2 \cdot 2}, \qquad 0 < \frac{1}{10} \le 1.$$

3. For Assumption 1, fix a latent-arm cell $(U = u, T = t)$. For bounded measurable test functions φ_Z of Z and ψ_{XY} of (X, Y), the finite conditional sums over the atom weights factor as

$$E_{P_{g,h}}[\varphi_Z(Z)\psi_{XY}(X,Y) \mid U = u, T = t] = E_{P_{g,h}}[\varphi_Z(Z) \mid U = u, T = t]\, E_{P_{g,h}}[\psi_{XY}(X,Y) \mid U = u, T = t]\,.$$

   Indeed, conditional on $(U = u, T = t)$, the factor depending on z is $\mathrm{bm}(z; \alpha_{t,u})$, while the factors determining X and Y are functions of x, y_0, y_1 and the fixed arm t. The normalizing denominator is the same finite latent-arm mass in all three conditional means, so expanding the four binary nuisance sums gives the displayed product identity.

4. For Assumption 2, fix $U = u$. For bounded measurable test functions φ_X of X and ψ_{YT} of (Y, T),

$$E_{P_{g,h}}[\varphi_X(X)\psi_{YT}(Y,T) \mid U = u] = E_{P_{g,h}}[\varphi_X(X) \mid U = u]\, E_{P_{g,h}}[\psi_{YT}(Y,T) \mid U = u]\,.$$

   After conditioning on $U = u$, summing over z uses

$$\sum_{z=0}^{1} \mathrm{bm}(z; \alpha_{t,u}) = 1$$

   for each arm t. The remaining finite sum separates the x-factor $\mathrm{bm}(x; \beta_u)$ from the factors depending on t, y_0, y_1, and hence on (Y, T).

5. For Assumption 4, fix a treatment state $r \in \{0, 1\}$ and a latent class u. For bounded measurable test functions φ_r of the potential outcome $Y(r)$ and ψ_T of T,

$$E_{P_{g,h}}[\varphi_r(Y(r))\psi_T(T) \mid U = u] = E_{P_{g,h}}[\varphi_r(Y(r)) \mid U = u]\, E_{P_{g,h}}[\psi_T(T) \mid U = u]\,.$$

   This is the same finite-product calculation: the Bernoulli factor for $Y(r)$ depends on u and r, while the Bernoulli factor for T is $\mathrm{bm}(t; w_1(u)/p_u)$.

6. Every support point of $P_{g,h}$ records the observed outcome as $Y = Y(T)$, which gives Assumption 3. Every support point also has $X = (1, x)^\top$, which gives $X_1 = 1$ and hence Assumption 5.

7. The boundedness conditions hold supportwise. Since $X = (1, x)^\top$ and $Z = (1, z)^\top$ with $x, z \in \{0, 1\}$,

$$\|X\|_2 \le \sqrt{2} \le 2.$$

   For every vector v,

$$\|ZX^\top v\|_2 = \|Z\|_2\, |X^\top v| \le \|Z\|_2\, \|X\|_2\, \|v\|_2 \le 2\|v\|_2,$$

   so $\|ZX^\top\|_{\mathrm{op}} \le 2$. Since $Y \in \{0, 1\}$,

$$\|YZX^\top\|_{\mathrm{op}} \le \|ZX^\top\|_{\mathrm{op}} \le 2.$$

   Thus Assumptions 6 to 8 hold with $L = 2$.

8. The latent-arm cell masses are precisely the four displayed values $w_t(u)$. From $|h| \le 1/100$,

$$\frac{4}{25} + \frac{3h}{5} \ge \frac{4}{25} - \frac{3}{500} \ge \frac{1}{10}, \qquad \frac{9}{25} - \frac{3h}{5} \ge \frac{9}{25} - \frac{3}{500} \ge \frac{1}{10},$$

and

$$\frac{6}{25} + \frac{2h}{5} \ge \frac{6}{25} - \frac{1}{250} \ge \frac{1}{10}, \qquad \frac{6}{25} - \frac{2h}{5} \ge \frac{6}{25} - \frac{1}{250} \ge \frac{1}{10}.$$

Therefore

$$\min_{u\in\{1,2\},\, t\in\{0,1\}} P_{g,h}(U = u, T = t) \ge \frac{1}{10},$$

which is Assumption 9 with $\pi_0 = 1/10$.

9. By Lemma 13,

$$\min\{\sigma_2(A_0), \sigma_2(A_1), \sigma_2(B)\} \ge \frac{1}{10}$$

for the reference- and target-proxy feature matrices formed from $P_{g,h}$. Hence Assumption 10 holds with $\sigma_0 = 1/10$.

10. Combining the core domain with the model restrictions just verified gives

$$P_{g,h} \in \mathcal{M}$$

for $k = d_x = d_z = 2$, $L = 2$, $\pi_0 = 1/10$, and $\sigma_0 = 1/10$, as required by Definition 32.

□

Lemma 15. [`lem:path-gap-stratum-membership`] (Path stratum membership) *For real numbers g and h, suppose that*

- ***(Gap scale.)*** *The local effect-gap scale satisfies $0 < g \le 1/4$.*
- ***(Path displacement.)*** *The path displacement satisfies $|h| \le 1/100$.*

Then the path law $P_{g,h}$ of Definition 51 belongs to the gap-local stratum $\mathcal{M}(g)$ of Definition 41, with $L = 2$, $\pi_0 = 1/10$, and $\sigma_0 = 1/10$. Equivalently, $P_{g,h} \in \mathcal{M}$ as specified by Definition 32 and Assumptions 6, 9 and 10, its smallest positive effect gap $\delta(P_{g,h})$ from Definition 27 satisfies the gap window of Assumption 11,

$$\frac{g}{2} \le \delta(P_{g,h}) \le 2g,$$

and its latent effects from Definition 23 are distinct as in Assumption 12.

Proof of Lemma 15. Write $P = P_{g,h}$. The hypothesis $0 < g \le 1/4$ gives $0 \le g$. Also $|h| \le 1/100$ gives

$$-\frac{1}{100} \le h \le \frac{1}{100}, \qquad \text{hence} \qquad -1 \le h \le 1.$$

Thus the detailed path construction in Definition 50 applies throughout the tangent-amplitude domain and yields the full-data probability law $P_{g,h}$.

Under the same bounds $0 \le g \le 1/4$ and $|h| \le 1/100$, Lemma 14 gives

$$P_{g,h} \in \mathcal{M}$$

for $k = d_x = d_z = 2$, $L = 2$, $\pi_0 = 1/10$, and $\sigma_0 = 1/10$. This supplies the model restrictions in Definition 32, including Assumptions 6, 9 and 10.

It remains to check the gap-local conditions. By the explicit construction in Definition 50, the latent-class masses are

$$p_1(P) = \frac{2}{5} + h, \qquad p_2(P) = \frac{3}{5} - h.$$

The displacement bound gives

$$p_1(P) \geq \frac{2}{5} - \frac{1}{100} = \frac{39}{100} > 0, \qquad p_2(P) \geq \frac{3}{5} - \frac{1}{100} = \frac{59}{100} > 0.$$

These are the positive-mass classes entering Definition 27 and Assumption 12.

Again by Definition 50, the latent conditional potential-outcome means are

$$\mu_{01} = \mu_{02} = \frac{1}{4}, \qquad \mu_{11} = \frac{1}{2} - \frac{g}{2}, \qquad \mu_{12} = \frac{1}{2} + \frac{g}{2}.$$

Using Definition 23, the corresponding latent effects are

$$\tau_1(P) = \mu_{11} - \mu_{01} = \frac{1}{4} - \frac{g}{2}, \qquad \tau_2(P) = \mu_{12} - \mu_{02} = \frac{1}{4} + \frac{g}{2}.$$

Since $g > 0$, these two effects are distinct, and their distance is

$$|\tau_1(P) - \tau_2(P)| = \left| \left(\frac{1}{4} - \frac{g}{2}\right) - \left(\frac{1}{4} + \frac{g}{2}\right) \right| = |-g| = g.$$

The same value is obtained with the two classes interchanged, while equal-index pairs are excluded by the condition $\tau_u \neq \tau_v$ in Definition 27. Hence

$$\big\{ |\tau_u(P) - \tau_v(P)| : p_u(P) > 0,\ p_v(P) > 0,\ \tau_u(P) \neq \tau_v(P) \big\} = \{g\},$$

and therefore

$$\delta(P) = g.$$

Consequently,

$$\frac{g}{2} \leq \delta(P) \leq 2g,$$

because $\delta(P) = g$ and $g > 0$. Also,

$$\tau_1(P) = \frac{1}{4} - \frac{g}{2} < \frac{1}{4} + \frac{g}{2} = \tau_2(P),$$

so the positive-mass latent effects take exactly $2 = k$ distinct values. This is Assumption 12.

Combining $P_{g,h} \in \mathcal{M}$, the gap domain $0 < g \leq 1/4$, the displayed gap window, and distinct effects gives

$$P_{g,h} \in \mathcal{M}(g)$$

with $L = 2$, $\pi_0 = 1/10$, and $\sigma_0 = 1/10$, as required by Definition 41. □

Lemma 16. [`lem:path-local-weight-experiment`] (Path local weight membership) *Let $n \in \mathbb{N}$, let $c_{\mathrm{loc}} \in \mathbb{R}$ be the local Kullback–Leibler radius constant, and let $g, h \in \mathbb{R}$. Assume:*

- ***(Local radius.)*** $0 < c_{\mathrm{loc}} < 1$, *as in Definition 49.*
- ***(Gap scale.)*** *The local effect-gap scale* g *is in the gap window of Assumption 11, with* $0 < g \leq 1/4$.
- ***(Path displacement.)*** *The path displacement satisfies* $|h| \leq 1/100$.
- ***(Observed KL calibration.)***
$$16000\, g^2 h^2 \leq \frac{c_{\mathrm{loc}}}{n}.$$

Then the path law $P_{g,h}$ *from Definition 51 belongs to the local labeled-weight experiment* $\mathcal{L}^p_{n,g}$ *of Definition 57 with* $L = 2$, $\pi_0 = 1/10$, *and* $\sigma_0 = 1/10$. *Equivalently,* $P_{g,h} \in \mathcal{M}(g)$ *as in Definition 41, and its observed-data marginal from Definition 20 obeys the neighborhood condition of Assumption 14:*
$$D_{\mathrm{KL}}((P_{g,h})_O \,\|\, (P_{g,0})_O) \leq \frac{c_{\mathrm{loc}}}{n}, \qquad P_{g,0} = P^{\mathrm{wit}}_{g/2}.$$

Proof of Lemma 16. The bound $|h| \leq 1/100$ gives $-1/100 \leq h \leq 1/100$, hence also $h \in [-1, 1]$. On this range the latent masses
$$\frac{2}{5} + h, \qquad \frac{3}{5} - h$$
belong to $[0, 1]$. The treatment Bernoulli probabilities are obtained by dividing
$$\frac{4}{25} + \frac{3}{5}h, \qquad \frac{9}{25} - \frac{3}{5}h$$
by the corresponding latent masses, and the same bounds on h make each numerator nonnegative and no larger than its denominator. The target-proxy probabilities
$$\frac{1}{5}, \qquad \frac{12/25 - h/5}{3/5 - h}$$
also lie in $[0, 1]$. For the reference-proxy probabilities, $|h| \leq 1/100$ gives positive denominators and
$$0 \leq \frac{225h + 28}{20(15h + 4)} \leq 1, \qquad 0 \leq \frac{35h + 9}{10(5h + 3)} \leq 1,$$
while the remaining displayed values $3/4$ and $7/10$ already lie in $[0, 1]$. Finally $1/4$ and $1/2 \pm g/2$ are Bernoulli probabilities because $0 \leq g \leq 1/4$. Thus all finite product weights defining $P_{g,h}$ in Definition 50 are nonnegative, and summing those product weights over
$$u \in \{1, 2\}, \qquad t, x, z, y_0, y_1 \in \{0, 1\}$$
leaves the two latent masses, whose sum is one. Hence $P_{g,h}$ is a probability law. The same argument applies at $h = 0$, and Definition 50 gives
$$P_{g,0} = P^{\mathrm{wit}}_{g/2}.$$

By Lemma 15, the hypotheses $0 < g \leq 1/4$ and $|h| \leq 1/100$ imply
$$P_{g,h} \in \mathcal{M}(g)$$

with $L = 2$, $\pi_0 = 1/10$, and $\sigma_0 = 1/10$. The local-radius hypothesis supplies

$$0 < c_{\mathrm{loc}} < 1.$$

It remains to verify the local observed-neighborhood inequality. Let

$$\mathsf{Cell}_{\mathrm{vis}} := \{0,1\}^4.$$

For $(\bar t, \bar x, \bar z, \bar y) \in \mathsf{Cell}_{\mathrm{vis}}$, define

$$\mathrm{vis}_{g,h}(\bar t, \bar x, \bar z, \bar y) = \begin{cases} \displaystyle\sum_{u=1}^{2} \sum_{\bar y_0 \in \{0,1\}} P_{g,h}(U = u, T = 1, X_2 = \bar x, Z_2 = \bar z, Y(0) = \bar y_0, Y(1) = \bar y), & \bar t = 1, \\ \displaystyle\sum_{u=1}^{2} \sum_{\bar y_1 \in \{0,1\}} P_{g,h}(U = u, T = 0, X_2 = \bar x, Z_2 = \bar z, Y(0) = \bar y, Y(1) = \bar y_1), & \bar t = 0. \end{cases}$$

Let $\mathsf{Vis}_{g,h}$ be the atomic law on $\mathsf{Cell}_{\mathrm{vis}}$ with these singleton masses. The map

$$\Psi_{\mathrm{vis}}(\bar t, \bar x, \bar z, \bar y) = \big(\bar t, (1, \bar x)^\top, (1, \bar z)^\top, \bar y\big)$$

pushes $\mathsf{Vis}_{g,h}$ to the observed margin:

$$(\mathsf{Vis}_{g,h}) \circ \Psi_{\mathrm{vis}}^{-1} = (P_{g,h})_O.$$

Therefore data processing gives

$$D_{\mathrm{KL}}((P_{g,h})_O \,\|\, (P_{g,0})_O) \le D_{\mathrm{KL}}(\mathsf{Vis}_{g,h} \,\|\, \mathsf{Vis}_{g,0}).$$

At $h = 0$, direct expansion of the sixteen visible cells gives

$\bar t = 0,\ (\bar x, \bar z)$	$\bar y = 0$	$\bar y = 1$
$(0,0)$	$279/2500$	$93/2500$
$(0,1)$	$171/2500$	$57/2500$
$(1,0)$	$171/2500$	$57/2500$
$(1,1)$	$279/2500$	$93/2500$

$\bar t = 1,\ (\bar x, \bar z)$	$\bar y = 0$	$\bar y = 1$
$(0,0)$	$(253 + 163g)/5000$	$(253 - 163g)/5000$
$(0,1)$	$(247 - 23g)/5000$	$(247 + 23g)/5000$
$(1,0)$	$(29 - 16g)/625$	$(29 + 16g)/625$
$(1,1)$	$(71 - 64g)/625$	$(71 + 64g)/625.$

Since $0 \le g \le 1/4$, every entry in these two tables is at least $1/1000$. Hence

$$\mathsf{Vis}_{g,0}\{(\bar t, \bar x, \bar z, \bar y)\} \ge \frac{1}{1000} \qquad \text{for every } (\bar t, \bar x, \bar z, \bar y) \in \mathsf{Cell}_{\mathrm{vis}},$$

and $\mathsf{Vis}_{g,h} \ll \mathsf{Vis}_{g,0}$.

For each visible cell, the same finite expansion gives the exact displacement

$$\mathrm{vis}_{g,h}(\bar t, \bar x, \bar z, \bar y) - \mathrm{vis}_{g,0}(\bar t, \bar x, \bar z, \bar y) = \begin{cases} (1 - 2\bar y)\, \beta_{\mathrm{vis}}(\bar x, \bar z)\, gh, & \bar t = 1, \\ 0, & \bar t = 0, \end{cases}$$

where

$$\beta_{\mathrm{vis}}(0,0) = \frac{3}{25}, \qquad \beta_{\mathrm{vis}}(0,1) = \frac{9}{25}, \qquad \beta_{\mathrm{vis}}(1,0) = \frac{3}{100}, \qquad \beta_{\mathrm{vis}}(1,1) = \frac{9}{100}.$$

All four coefficients are at most one, and $g \geq 0$, so

$$|\mathsf{Vis}_{g,h}\{(\bar{t}, \bar{x}, \bar{z}, \bar{y})\} - \mathsf{Vis}_{g,0}\{(\bar{t}, \bar{x}, \bar{z}, \bar{y})\}| \leq g|h|$$

for every visible cell.

For finite probability laws $\mathfrak{p} \ll \mathfrak{q}$ with positive base mass on every cell,

$$D_{\mathrm{KL}}(\mathfrak{p} \| \mathfrak{q}) \leq \chi^2(\mathfrak{p}, \mathfrak{q}), \qquad \chi^2(\mathfrak{p}, \mathfrak{q}) = \sum_{\upsilon \in \mathsf{Cell}_{\mathrm{vis}}} \frac{(\mathfrak{p}\{\upsilon\} - \mathfrak{q}\{\upsilon\})^2}{\mathfrak{q}\{\upsilon\}},$$

using $\log r \leq r - 1$ for the first inequality. Applying this with $\mathfrak{p} = \mathsf{Vis}_{g,h}$ and $\mathfrak{q} = \mathsf{Vis}_{g,0}$, the preceding floor and displacement bound yield, cell by cell,

$$\frac{(\mathsf{Vis}_{g,h}\{\upsilon\} - \mathsf{Vis}_{g,0}\{\upsilon\})^2}{\mathsf{Vis}_{g,0}\{\upsilon\}} \leq 1000\, g^2 h^2.$$

There are 16 visible cells, and therefore

$$D_{\mathrm{KL}}(\mathsf{Vis}_{g,h} \,\|\, \mathsf{Vis}_{g,0}) \leq 16000\, g^2 h^2.$$

Combining data processing with the last display and the assumed calibration gives

$$D_{\mathrm{KL}}((P_{g,h})_O \,\|\, (P_{g,0})_O) \leq 16000\, g^2 h^2 \leq \frac{c_{\mathrm{loc}}}{n}.$$

Using $P_{g,0} = P^{\mathrm{wit}}_{g/2}$, this is exactly the neighborhood condition in Assumption 14. Together with $P_{g,h} \in \mathcal{M}(g)$, it proves

$$P_{g,h} \in \mathcal{L}^p_{n,g}$$

as defined in Definition 57. □

The potential lemma is the dual device behind the W_1 modulus for represented atomic laws. The two summary-metric lemmas place the finite-dimensional summary topology and the selection loss on a common continuous footing.

The next group records the witness and path bookkeeping used in the lower-bound section. The statements give explicit masses, effects, observed arm probabilities, determinant margins, model membership, and local-experiment membership for the two-class laws.

Definition 62. [`synth_26`] (Ordered aggregate mass map) *For a labelled representative $\xi = \sum_{i=1}^k w_i \delta_{x_i} \in \mathcal{P}_{\leq k}([-L_\tau, L_\tau])$, define* $\mathrm{orderedMasses}(\xi) \in \mathbb{R}^k$ *as follows. If every labelled weight is positive and the labelled atoms are distinct, then its jth coordinate is the mass of the atom whose support rank is j:*

$$\mathrm{orderedMasses}(\xi)_j := \sum_{i=1}^k w_i \mathbf{1}\{\#\{\ell : x_\ell < x_i\} = j\}, \qquad j = 0, \ldots, k-1.$$

If some labelled weight is zero or two labelled atoms coincide, set $\mathrm{orderedMasses}(\xi) := (1/k, \ldots, 1/k)$.

Lemma 17. [`lem:ordered-masses-measure-invariance`] (Ordered masses are representative-invariant) *Fix $k \in \mathbb{N}$ and a real support radius. Let ν and ξ be valid labelled atomic representatives with k slots in the corresponding compact interval, as in Definition 29. If ν and ξ represent the same aggregated probability measure, then their ordered-mass vectors coincide:*

$$\mathrm{orderedMasses}(\nu) = \mathrm{orderedMasses}(\xi).$$

Proof of Lemma 17. Write the two labelled representatives as

$$\nu = \big((\varpi_{\mathfrak{i}}^{\nu})_{\mathfrak{i}\in I_k}, (\chi_{\mathfrak{i}}^{\nu})_{\mathfrak{i}\in I_k}\big), \qquad \xi = \big((\varpi_{\mathfrak{i}}^{\xi})_{\mathfrak{i}\in I_k}, (\chi_{\mathfrak{i}}^{\xi})_{\mathfrak{i}\in I_k}\big), \qquad I_k := \{0, \ldots, k-1\}.$$

Validity means

$$0 \le \varpi_{\mathfrak{i}}^{\nu}, \quad 0 \le \varpi_{\mathfrak{i}}^{\xi}, \quad \sum_{\mathfrak{i}\in I_k} \varpi_{\mathfrak{i}}^{\nu} = 1, \quad \sum_{\mathfrak{i}\in I_k} \varpi_{\mathfrak{i}}^{\xi} = 1,$$

with all labelled atoms lying in the prescribed compact interval. For any labelled representative $\zeta = ((\varpi_{\mathfrak{i}}^{\zeta}), (\chi_{\mathfrak{i}}^{\zeta}))$, put

$$\mathrm{supp}_+(\zeta) := \{\chi_{\mathfrak{i}}^{\zeta} : \mathfrak{i} \in I_k,\ 0 < \varpi_{\mathfrak{i}}^{\zeta}\}.$$

If $k = 0$, there are no vector coordinates to check.

1. The equality of represented probability measures gives equality of aggregate mass at every real point:

   $$\sum_{\mathfrak{i}:\, \chi_{\mathfrak{i}}^{\nu}=\mathfrak{x}} \varpi_{\mathfrak{i}}^{\nu} = \sum_{\mathfrak{j}:\, \chi_{\mathfrak{j}}^{\xi}=\mathfrak{x}} \varpi_{\mathfrak{j}}^{\xi} \qquad \text{for every } \mathfrak{x} \in \mathbb{R}.$$

   Indeed, by Definitions 6 and 29, the represented measure assigns $\{\mathfrak{x}\}$ the aggregate of the positive parts of the labelled weights at $\mathfrak{x}$; validity gives $\max\{\varpi_{\mathfrak{i}}^{\nu}, 0\} = \varpi_{\mathfrak{i}}^{\nu}$ and $\max\{\varpi_{\mathfrak{j}}^{\xi}, 0\} = \varpi_{\mathfrak{j}}^{\xi}$. Evaluating the common represented measure on $\{\mathfrak{x}\}$ gives the displayed identity.

2. Consequently,

   $$\mathrm{supp}_+(\nu) = \mathrm{supp}_+(\xi).$$

   If $\mathfrak{x} = \chi_{\mathfrak{i}}^{\nu}$ and $0 < \varpi_{\mathfrak{i}}^{\nu}$, then the aggregate mass of ν at $\mathfrak{x}$ is positive. By Item 1, the aggregate mass of ξ at $\mathfrak{x}$ is positive, and nonnegativity of the labelled weights yields some $\mathfrak{j} \in I_k$ with

   $$\chi_{\mathfrak{j}}^{\xi} = \mathfrak{x}, \qquad 0 < \varpi_{\mathfrak{j}}^{\xi}.$$

   This proves one inclusion, and the reverse inclusion follows by interchanging ν and ξ.

3. For a representative ζ, define the rank of a labelled slot by

   $$\mathrm{rank}_{\zeta}(\mathfrak{i}) := \#\{\mathfrak{q} \in I_k : \chi_{\mathfrak{q}}^{\zeta} < \chi_{\mathfrak{i}}^{\zeta}\}.$$

   The counted set excludes $\mathfrak{i}$, so

   $$\mathrm{rank}_{\zeta}(\mathfrak{i}) < k.$$

   Moreover, if $\chi_{\mathfrak{i}}^{\zeta} < \chi_{\mathfrak{j}}^{\zeta}$, then

   $$\mathrm{rank}_{\zeta}(\mathfrak{i}) < \mathrm{rank}_{\zeta}(\mathfrak{j}),$$

   because the first counted set is contained in the second and $\mathfrak{i}$ belongs to the second but not the first. Hence, whenever $\mathfrak{i} \mapsto \chi_{\mathfrak{i}}^{\zeta}$ is injective, the map $\mathfrak{i} \mapsto \mathrm{rank}_{\zeta}(\mathfrak{i})$ is injective, and therefore bijective on the finite set I_k.

4. The following full-support criterion holds:

$$\#\operatorname{supp}_+(\zeta) = k \quad\iff\quad \Big[(\forall \mathfrak{i} \in I_k,\ 0 < \varpi_{\mathfrak{i}}^{\zeta}) \text{ and } \mathfrak{i} \mapsto \chi_{\mathfrak{i}}^{\zeta} \text{ is injective}\Big].$$

   For the forward implication, $\operatorname{supp}_+(\zeta)$ is the image of $\{\mathfrak{i} \in I_k : 0 < \varpi_{\mathfrak{i}}^{\zeta}\}$ under $\mathfrak{i} \mapsto \chi_{\mathfrak{i}}^{\zeta}$, so

$$\#\operatorname{supp}_+(\zeta) \le \#\{\mathfrak{i} \in I_k : 0 < \varpi_{\mathfrak{i}}^{\zeta}\} \le k.$$

   If the left side equals k, both inequalities are equalities; thus every labelled weight is positive, and equality in the image-cardinality bound gives injectivity of the labelled locations. The reverse implication follows because all k labelled slots are then positive and distinct.

5. Suppose first that

$$\#\operatorname{supp}_+(\nu) = k.$$

   By Item 2,

$$\#\operatorname{supp}_+(\xi) = k.$$

   Then Item 4 gives

$$0 < \varpi_{\mathfrak{i}}^{\nu}, \quad 0 < \varpi_{\mathfrak{i}}^{\xi} \qquad (\mathfrak{i} \in I_k),$$

   and both labelled-location maps are injective. For either such full-support representative ζ, Definition 62 gives

$$\operatorname{orderedMasses}(\zeta)_{\operatorname{rank}_\zeta(\mathfrak{i})} = \sum_{\mathfrak{q}:\, \operatorname{rank}_\zeta(\mathfrak{q}) = \operatorname{rank}_\zeta(\mathfrak{i})} \varpi_{\mathfrak{q}}^{\zeta} = \varpi_{\mathfrak{i}}^{\zeta},$$

   where the last equality uses the rank injectivity from Item 3.

6. Fix a coordinate $\mathfrak{r} \in I_k$. Since the rank map for ν is bijective by Item 3, choose $\mathfrak{i} \in I_k$ such that

$$\mathfrak{r} = \operatorname{rank}_\nu(\mathfrak{i}).$$

   The support equality in Item 2 gives $\mathfrak{j} \in I_k$ with

$$\chi_{\mathfrak{j}}^{\xi} = \chi_{\mathfrak{i}}^{\nu}, \qquad 0 < \varpi_{\mathfrak{j}}^{\xi}.$$

   Using Item 1 at $\mathfrak{x} = \chi_{\mathfrak{i}}^{\nu}$, and using injectivity of both labelled-location maps to collapse the two sums, yields

$$\varpi_{\mathfrak{i}}^{\nu} = \sum_{\mathfrak{q}:\, \chi_{\mathfrak{q}}^{\nu} = \chi_{\mathfrak{i}}^{\nu}} \varpi_{\mathfrak{q}}^{\nu} = \sum_{\mathfrak{q}:\, \chi_{\mathfrak{q}}^{\xi} = \chi_{\mathfrak{i}}^{\nu}} \varpi_{\mathfrak{q}}^{\xi} = \varpi_{\mathfrak{j}}^{\xi}.$$

7. For each $\mathfrak{q} \in I_k$, Item 2 supplies a slot $\psi(\mathfrak{q}) \in I_k$ such that

$$\chi_{\psi(\mathfrak{q})}^{\xi} = \chi_{\mathfrak{q}}^{\nu}.$$

   The map ψ is injective: if $\psi(\mathfrak{q}_1) = \psi(\mathfrak{q}_2)$, then $\chi_{\mathfrak{q}_1}^{\nu} = \chi_{\mathfrak{q}_2}^{\nu}$, and injectivity of the locations of ν gives $\mathfrak{q}_1 = \mathfrak{q}_2$. Thus ψ is bijective. Since also $\chi_{\mathfrak{j}}^{\xi} = \chi_{\mathfrak{i}}^{\nu}$, for every $\mathfrak{q} \in I_k$,

$$\chi_{\mathfrak{q}}^{\nu} < \chi_{\mathfrak{i}}^{\nu} \quad\iff\quad \chi_{\psi(\mathfrak{q})}^{\xi} < \chi_{\mathfrak{j}}^{\xi}.$$

The bijection ψ therefore identifies the two sets counted by the ranks, so

$$\operatorname{rank}_\nu(\mathfrak{i}) = \operatorname{rank}_\xi(\mathfrak{j}).$$

Combining this rank equality with Items 5 and 6 gives

$$\text{orderedMasses}(\nu)_\mathfrak{r} = \varpi_\mathfrak{i}^\nu = \varpi_\mathfrak{j}^\xi = \text{orderedMasses}(\xi)_\mathfrak{r}.$$

Since $\mathfrak{r}$ was arbitrary, the ordered-mass vectors are equal in the full-support case.

8. It remains to consider

$$\# \operatorname{supp}_+(\nu) \neq k.$$

By Item 2,

$$\# \operatorname{supp}_+(\xi) \neq k.$$

Using Item 4, the positivity-and-injectivity condition in Definition 62 fails for both representatives. Hence both ordered-mass vectors take the barycentric fallback value:

$$\text{orderedMasses}(\nu)_\mathfrak{r} = k^{-1} = \text{orderedMasses}(\xi)_\mathfrak{r} \qquad \text{for every } \mathfrak{r} \in I_k.$$

Thus the ordered-mass vectors are equal in the complementary case as well.

The two cases prove

$$\text{orderedMasses}(\nu) = \text{orderedMasses}(\xi),$$

as required. □

Lemma 18. [`lem:observed-outcome-proxy-moment-factorization`] (Outcome proxy moment factorization) *Let P, a full-data probability law carrying $(U, T, X, Z, Y(0), Y(1), Y)$, belong to the model class $\mathcal{M}$ in Definition 32. Suppose that*

- ***(Latent classes.)*** $2 \leq k$.
- ***(Positivity margin.)*** *The latent-arm positivity margin satisfies $\pi_0 > 0$.*
- ***(Model conditions.)*** *The reference- and target-proxy separation, consistency, latent ignorability, latent-arm positivity, bounded outcome-proxy product, and bounded target-proxy conditions in Assumptions 1 to 4, 6, 8 and 9 hold.*

Then, for each treatment arm $t \in \{0, 1\}$,

$$N_t(P) = A_t \, \Gamma_t \, \operatorname{diag}\{\mu_{tu} : 1 \leq u \leq k\} \, B^\top, \qquad \Gamma_t = \operatorname{diag}\left\{ \frac{P(U = u, T = t)}{P(T = t)} : 1 \leq u \leq k \right\}.$$

Here $N_t(P)$ is the observable armwise $YZX^\top$ population moment in Definition 28, A_t is the reference-proxy feature matrix in Definition 24, B is the target-proxy feature matrix in Definition 25, and μ_{tu} is the latent conditional potential-outcome mean in Definition 22.

Proof of Lemma 18. Step 1. Fix $t \in \{0,1\}$. For $1 \le u \le k$, put

$$\mathsf{U}_u := \{U = u\}, \qquad \mathsf{R}_t := \{T = t\}, \qquad \mathsf{H}_{u,t} := \mathsf{U}_u \cap \mathsf{R}_t.$$

For a full-data event $\mathfrak{e}$ and a real-valued measurable function φ, write

$$\mathbb{E}^P_{\mathfrak{e}}[\varphi] := \{P(\mathfrak{e})\}^{-1} \int_{\mathfrak{e}} \varphi \, dP,$$

with the total convention $0^{-1} = 0$. The latent-arm positivity assumption gives, for every u,

$$0 < \pi_0 \le P(\mathsf{H}_{u,t}) \le P(\mathsf{U}_u).$$

Also, Lemma 3 gives

$$k\pi_0 \le P(\mathsf{R}_t),$$

and since $2 \le k$ and $0 < \pi_0$,

$$0 < k\pi_0 \le P(\mathsf{R}_t).$$

Thus all normalizations used below have positive denominators on each latent class, each latent-treatment cell, and the treatment arm.

Step 2. Let P°_u be the normalized restriction of P to U_u, equivalently the probability law with density $P(\mathsf{U}_u)^{-1}\mathbf{1}_{\mathsf{U}_u}$ with respect to P. Latent ignorability in Assumption 4 gives independence of $Y(t)$ and T under P°_u. Therefore

$$\int_{\mathsf{R}_t} Y(t) \, dP^\circ_u = P^\circ_u(\mathsf{R}_t) \int Y(t) \, dP^\circ_u, \qquad P^\circ_u(\mathsf{R}_t) = \frac{P(\mathsf{H}_{u,t})}{P(\mathsf{U}_u)}.$$

Using the normalized-restriction formula on both sides,

$$\frac{1}{P(\mathsf{U}_u)} \int_{\mathsf{H}_{u,t}} Y(t) \, dP = \frac{P(\mathsf{H}_{u,t})}{P(\mathsf{U}_u)} \frac{1}{P(\mathsf{U}_u)} \int_{\mathsf{U}_u} Y(t) \, dP.$$

The positive denominators from Step 1 allow cancellation, giving

$$\mathbb{E}^P_{\mathsf{H}_{u,t}}[Y(t)] = \mathbb{E}^P_{\mathsf{U}_u}[Y(t)] = \mu_{tu},$$

where the last equality is Definition 22. On R_t, consistency in Assumption 3 gives $Y = Y(t)$ almost surely, hence

$$\mathbb{E}^P_{\mathsf{H}_{u,t}}[Y] = \mathbb{E}^P_{\mathsf{H}_{u,t}}[Y(t)] = \mu_{tu}.$$

Step 3. Fix coordinates $1 \le \iota_z \le d_z$ and $1 \le \iota_x \le d_x$. Target-proxy separation in Assumption 2, applied under P°_u, gives independence of X and (Y,T) conditional on $U = u$. Define

$$\psi_t(y,s) := y \, \mathbf{1}\{s = t\}, \qquad (y,s) \in \mathbb{R} \times \{0,1\}.$$

The product identity under P°_u gives

$$\int X_{\iota_x} \psi_t(Y,T) \, dP^\circ_u = \left(\int X_{\iota_x} \, dP^\circ_u \right) \left(\int \psi_t(Y,T) \, dP^\circ_u \right).$$

The three terms are

$$\int X_{\iota_x} \psi_t(Y,T) \, dP^\circ_u = \frac{1}{P(\mathsf{U}_u)} \int_{\mathsf{H}_{u,t}} X_{\iota_x} Y \, dP,$$

$$\int X_{\iota_x}\, dP_u^\circ = \frac{1}{P(\mathsf{U}_u)} \int_{\mathsf{U}_u} X_{\iota_x}\, dP, \qquad \int \psi_t(Y,T)\, dP_u^\circ = \frac{1}{P(\mathsf{U}_u)} \int_{\mathsf{H}_{u,t}} Y\, dP.$$

Multiplying by $P(\mathsf{U}_u)/P(\mathsf{H}_{u,t})$ and using Step 1 yields

$$\mathbb{E}^P_{\mathsf{H}_{u,t}}[X_{\iota_x} Y] = \mathbb{E}^P_{\mathsf{U}_u}[X_{\iota_x}]\, \mathbb{E}^P_{\mathsf{H}_{u,t}}[Y].$$

By Definition 25 and Step 2,

$$\mathbb{E}^P_{\mathsf{H}_{u,t}}[X_{\iota_x} Y] = B(\iota_x, u)\, \mu_{tu}.$$

Step 4. Let $P^\circ_{u,t}$ be the normalized restriction of P to $\mathsf{H}_{u,t}$. Reference-proxy separation in Assumption 1, applied under $P^\circ_{u,t}$, gives independence of Z and (X,Y) on the latent-treatment cell. Applying the product identity to Z_{ι_z} and $X_{\iota_x} Y$,

$$\int Z_{\iota_z} X_{\iota_x} Y\, dP^\circ_{u,t} = \left(\int Z_{\iota_z}\, dP^\circ_{u,t}\right)\left(\int X_{\iota_x} Y\, dP^\circ_{u,t}\right).$$

Equivalently,

$$\mathbb{E}^P_{\mathsf{H}_{u,t}}[Z_{\iota_z} X_{\iota_x} Y] = \mathbb{E}^P_{\mathsf{H}_{u,t}}[Z_{\iota_z}]\, \mathbb{E}^P_{\mathsf{H}_{u,t}}[X_{\iota_x} Y].$$

Combining this identity with Step 3 and Definition 24 gives the cellwise factorization

$$\mathbb{E}^P_{\mathsf{H}_{u,t}}[Y Z_{\iota_z} X_{\iota_x}] = A_t(\iota_z, u)\, \mu_{tu}\, B(\iota_x, u).$$

Step 5. Set

$$\varphi_{\iota_z,\iota_x} := Y Z_{\iota_z} X_{\iota_x}.$$

The bounded outcome-proxy product condition in Assumption 8 implies

$$|\varphi_{\iota_z,\iota_x}| = |Y Z_{\iota_z} X_{\iota_x}| \le \|Y Z X^\top\|_{\mathrm{op}} \le L$$

almost surely, since every matrix entry is bounded by the operator norm. Thus $\varphi_{\iota_z,\iota_x}$ is integrable on every latent-treatment cell. The treatment arm decomposes as the disjoint finite union

$$\mathsf{R}_t = \bigsqcup_{u=1}^{k} \mathsf{H}_{u,t},$$

so

$$\int_{\mathsf{R}_t} \varphi_{\iota_z,\iota_x}\, dP = \sum_{u=1}^{k} \int_{\mathsf{H}_{u,t}} \varphi_{\iota_z,\iota_x}\, dP.$$

Dividing by the positive arm mass from Step 1 and rewriting each cell integral as a conditional mean gives

$$\mathbb{E}^P_{\mathsf{R}_t}[\varphi_{\iota_z,\iota_x}] = \sum_{u=1}^{k} \frac{P(\mathsf{H}_{u,t})}{P(\mathsf{R}_t)} \mathbb{E}^P_{\mathsf{H}_{u,t}}[\varphi_{\iota_z,\iota_x}].$$

Substituting Step 4,

$$\mathbb{E}^P_{\mathsf{R}_t}[Y Z_{\iota_z} X_{\iota_x}] = \sum_{u=1}^{k} A_t(\iota_z, u)\, \frac{P(U=u, T=t)}{P(T=t)}\, \mu_{tu}\, B(\iota_x, u).$$

Step 6. By Definition 20, P_O is the pushforward of P by $O = (T, X, Z, Y)$. Hence the observed-arm conditional mean agrees with the corresponding full-data arm conditional mean:

$$E_{P_O}[Y Z_{\iota_z} X_{\iota_x} \mid T = t] = \mathbb{E}^P_{\mathsf{R}_t}[Y Z_{\iota_z} X_{\iota_x}].$$

The armwise $YZX^\top$ moment notation in Definition 28 identifies the (ι_z, ι_x) entry of $N_t(P)$ with this same full-data conditional moment. Therefore

$$N_t(P)(\iota_z, \iota_x) = \sum_{u=1}^{k} A_t(\iota_z, u) \, \frac{P(U = u, T = t)}{P(T = t)} \, \mu_{tu} \, B(\iota_x, u).$$

With

$$\Gamma_t = \operatorname{diag}\left\{ \frac{P(U = u, T = t)}{P(T = t)} : 1 \le u \le k \right\},$$

the (ι_z, ι_x) entry of

$$A_t \, \Gamma_t \, \operatorname{diag}\{\mu_{tu} : 1 \le u \le k\} \, B^\top$$

is exactly the displayed sum. Since ι_z and ι_x were arbitrary,

$$N_t(P) = A_t \, \Gamma_t \, \operatorname{diag}\{\mu_{tu} : 1 \le u \le k\} \, B^\top.$$

□

Lemma 19. [`lem:latent-arm-weight-invertibility`] (Latent-arm weight injectivity) *Let P be a full-data probability law carrying $(U, T, X, Z, Y(0), Y(1), Y)$, and let $\pi_0 > 0$. Suppose P satisfies latent-arm positivity with margin π_0, as in Assumption 9. Then, for each treatment arm $t \in \{0, 1\}$, the diagonal matrix*

$$\Gamma_t = \operatorname{diag}\left\{ \frac{P(U = u, T = t)}{P(T = t)} : u \in \{1, \ldots, k\} \right\}$$

defines an injective linear map on $\mathbb{R}^k$.

Proof of Lemma 19. Fix $t \in \{0, 1\}$. For a latent coordinate $u \in \{1, \ldots, k\}$, set

$$\mathsf{cell}_{u,t} := \{U = u, \ T = t\}, \qquad \mathsf{arm}_t := \{T = t\}, \qquad \gamma_{u,t} := \frac{P(\mathsf{cell}_{u,t})}{P(\mathsf{arm}_t)}.$$

By the definition of Γ_t, this matrix is diagonal and its uth diagonal entry is $\gamma_{u,t}$.

We first show that every diagonal entry is bounded below by the positivity margin. By Assumption 9,

$$\pi_0 \le P(\mathsf{cell}_{u,t}).$$

Moreover $\mathsf{cell}_{u,t} \subseteq \mathsf{arm}_t$, so monotonicity gives

$$P(\mathsf{cell}_{u,t}) \le P(\mathsf{arm}_t).$$

Together with $\pi_0 > 0$, this implies

$$0 < P(\mathsf{arm}_t).$$

Since P is a probability law,

$$P(\mathsf{arm}_t) \le 1.$$

Hence, using $\pi_0 > 0$,

$$\pi_0 P(\mathsf{arm}_t) \le \pi_0 \le P(\mathsf{cell}_{u,t}).$$

Dividing by the positive number $P(\mathsf{arm}_t)$ yields

$$\pi_0 \le \frac{P(\mathsf{cell}_{u,t})}{P(\mathsf{arm}_t)} = \gamma_{u,t}.$$

In particular $\gamma_{u,t} > 0$, so $\gamma_{u,t} \ne 0$.

Now let $x, y \in \mathbb{R}^k$ and suppose $\Gamma_t x = \Gamma_t y$. Taking the uth coordinate and using diagonality gives

$$\gamma_{u,t} x_u = \gamma_{u,t} y_u.$$

Since $\gamma_{u,t} \ne 0$, cancellation gives

$$x_u = y_u.$$

This holds for every latent coordinate u, and therefore $x = y$. Thus the linear map induced by Γ_t is injective. $\square$

Lemma 20. [`lem:moore-penrose-product-cancellation`] (Moore–Penrose product cancellation) *Let $C \in \mathbb{R}^{\mathrm{rows}\times r}$, $B \in \mathbb{R}^{\mathrm{cols}\times r}$, and $\mu \in \mathbb{R}^r$. Suppose that the linear maps represented by C and by B, each with domain $\mathbb{R}^r$, are injective. Then*

$$(CB^\top)^\dagger \big(C \operatorname{diag}(\mu) B^\top\big) = (B^\dagger)^\top \operatorname{diag}(\mu) B^\top.$$

Proof of Lemma 20. For any real rectangular matrix A, let $A^\dagger$ denote its Moore–Penrose inverse, characterized by

$$AA^\dagger A = A, \qquad A^\dagger A A^\dagger = A^\dagger, \qquad (AA^\dagger)^\top = AA^\dagger, \qquad (A^\dagger A)^\top = A^\dagger A.$$

If the matrix map A is injective, then

$$A^\dagger A = I.$$

Indeed, the first Moore–Penrose identity gives

$$A(A^\dagger A) = AA^\dagger A = A.$$

Thus, for every vector x,

$$A\{(A^\dagger A)x\} = Ax,$$

and injectivity of A gives $(A^\dagger A)x = x$. Hence $A^\dagger A = I$.

Apply this observation to C and B. Then

$$C^\dagger C = I_r, \qquad B^\dagger B = I_r,$$

and transposing the second identity gives

$$B^\top (B^\dagger)^\top = I_r.$$

Set

$$G := (B^\dagger)^\top C^\dagger.$$

We verify that G is the Moore–Penrose inverse of $CB^\top$. First,

$$(CB^\top)G(CB^\top) = CB^\top(B^\dagger)^\top C^\dagger CB^\top = CI_rI_rB^\top = CB^\top,$$

and

$$G(CB^\top)G = (B^\dagger)^\top C^\dagger CB^\top(B^\dagger)^\top C^\dagger = (B^\dagger)^\top I_rI_rC^\dagger = G.$$

For the symmetry equations, the same cancellations give

$$(CB^\top)G = CB^\top(B^\dagger)^\top C^\dagger = CC^\dagger,$$

so $((CB^\top)G)^\top = (CB^\top)G$ by the Moore–Penrose symmetry of $CC^\dagger$. Also

$$G(CB^\top) = (B^\dagger)^\top C^\dagger CB^\top = (B^\dagger)^\top B^\top = (BB^\dagger)^\top = BB^\dagger,$$

so $(G(CB^\top))^\top = G(CB^\top)$ by the Moore–Penrose symmetry of $BB^\dagger$. Thus G satisfies all four Moore–Penrose equations for $CB^\top$, and uniqueness yields

$$(CB^\top)^\dagger = (B^\dagger)^\top C^\dagger.$$

Therefore, using associativity and $C^\dagger C = I_r$,

$$\begin{aligned}(CB^\top)^\dagger\big(C\,\mathrm{diag}(\mu)B^\top\big) &= (B^\dagger)^\top C^\dagger C\,\mathrm{diag}(\mu)B^\top \\ &= (B^\dagger)^\top\,\mathrm{diag}(\mu)B^\top.\end{aligned}$$

This is the claimed cancellation identity. □

Lemma 21. [`lem:kantorovich-rubinstein-attaining-potential`] (Attaining KR potential) *Let ν and ξ be valid represented atomic laws with finite labelled slot sets: their weights are nonnegative and sum to one. Let $\varphi_{\nu,\xi}$ be the cumulative-distribution-gap sign Kantorovich–Rubinstein potential selected for the pair (ν,ξ). Then $\varphi_{\nu,\xi}$ is one-Lipschitz, is normalized by*

$$\varphi_{\nu,\xi}(0) = 0,$$

and attains the Wasserstein distance of Definition 19:

$$W_1(\nu,\xi) = \left|\int \varphi_{\nu,\xi}\,d\nu - \int \varphi_{\nu,\xi}\,d\xi\right|.$$

Moreover, for every $\epsilon \in \mathbb{R}$, if every one-Lipschitz function $\varphi : \mathbb{R} \to \mathbb{R}$ satisfying $\varphi(0) = 0$ obeys

$$\left|\int \varphi\,d\nu - \int \varphi\,d\xi\right| \le \epsilon,$$

then

$$W_1(\nu,\xi) \le \epsilon.$$

Proof of Lemma 21. Write

$$\nu = \sum_i p_i\delta_{x_i}, \qquad \xi = \sum_j q_j\delta_{y_j},$$

with $p_i, q_j \geq 0$, $\sum_i p_i = \sum_j q_j = 1$. Define

$$F_\nu(t) := \sum_i p_i \mathbf{1}\{x_i \leq t\}, \qquad F_\xi(t) := \sum_j q_j \mathbf{1}\{y_j \leq t\}, \qquad \Delta(t) := F_\nu(t) - F_\xi(t),$$

and let

$$s(t) := \begin{cases} 1, & \Delta(t) > 0, \\ -1, & \Delta(t) < 0, \\ 0, & \Delta(t) = 0. \end{cases}$$

The selected potential is

$$\varphi_{\nu,\xi}(x) := \int_0^x s(t)\, dt.$$

Since $|s(t)| \leq 1$, for all $x, y \in \mathbb{R}$,

$$|\varphi_{\nu,\xi}(x) - \varphi_{\nu,\xi}(y)| = \left| \int_y^x s(t)\, dt \right| \leq |x - y|,$$

so $\varphi_{\nu,\xi}$ is one-Lipschitz. Also

$$\varphi_{\nu,\xi}(0) = \int_0^0 s(t)\, dt = 0.$$

We next record the finite one-dimensional transport identity used by the potential. For any finite transport plan $\gamma = (\gamma_{ij})$ from ν to ξ, put

$$C_\gamma(t) := \sum_{i,j} \gamma_{ij} \big(\mathbf{1}\{x_i \leq t < y_j\} - \mathbf{1}\{y_j \leq t < x_i\} \big).$$

Using the two marginal constraints of the plan gives

$$\Delta(t) = C_\gamma(t) \qquad \text{for every } t \in \mathbb{R}.$$

Hence, with

$$E_\gamma(t) := \sum_{i,j} \gamma_{ij} \mathbf{1}\{\min(x_i, y_j) \leq t < \max(x_i, y_j)\},$$

we have

$$|\Delta(t)| \leq E_\gamma(t).$$

Integrating and using

$$\int_{\mathbb{R}} \mathbf{1}\{\min(x_i, y_j) \leq t < \max(x_i, y_j)\}\, dt = |x_i - y_j|$$

yields

$$\int_{\mathbb{R}} |\Delta(t)|\, dt \leq \sum_{i,j} \gamma_{ij} |x_i - y_j|.$$

Conversely, the finite transport polytope is compact and the quadratic cost

$$\sum_{i,j} \gamma_{ij} (x_i - y_j)^2$$

therefore has a minimizer. If such a minimizer sent positive mass across some cut in both directions, shifting $\varepsilon = \min\{\gamma_{ij}, \gamma_{kl}\}$ from the crossing pairs $(i, j), (k, l)$ to $(i, l), (k, j)$ would preserve the two marginals and decrease the quadratic cost by

$$2\varepsilon(x_k - x_i)(y_j - y_l) > 0,$$

a contradiction. Thus the minimizer has no counterflow across any cut, and for that plan $|C_\gamma(t)| = E_\gamma(t)$ for every t. Consequently

$$\sum_{i,j} \gamma_{ij}|x_i - y_j| = \int_{\mathbb{R}} E_\gamma(t)\, dt = \int_{\mathbb{R}} |\Delta(t)|\, dt.$$

Together with the preceding inequality and Definition 19, this gives

$$W_1(\nu, \xi) = \int_{\mathbb{R}} |\Delta(t)|\, dt.$$

It remains to evaluate the potential. For an atom a, set

$$\chi_a(t) := \mathbf{1}\{0 \le t < a\} - \mathbf{1}\{a \le t < 0\}.$$

Then

$$\varphi_{\nu,\xi}(a) = \int_{\mathbb{R}} s(t)\chi_a(t)\, dt.$$

Interchanging the finite sums with the integral gives

$$\int \varphi_{\nu,\xi}\, d\nu - \int \varphi_{\nu,\xi}\, d\xi = \int_{\mathbb{R}} s(t) \left\{ \sum_i p_i \chi_{x_i}(t) - \sum_j q_j \chi_{y_j}(t) \right\} dt.$$

For $t \ge 0$, the expression in braces equals $(1 - F_\nu(t)) - (1 - F_\xi(t)) = -\Delta(t)$; for $t < 0$, it equals $-F_\nu(t) + F_\xi(t) = -\Delta(t)$. Therefore

$$\int \varphi_{\nu,\xi}\, d\nu - \int \varphi_{\nu,\xi}\, d\xi = -\int_{\mathbb{R}} s(t)\Delta(t)\, dt = -\int_{\mathbb{R}} |\Delta(t)|\, dt.$$

Taking absolute values and using the preceding display for W_1 gives

$$W_1(\nu, \xi) = \left| \int \varphi_{\nu,\xi}\, d\nu - \int \varphi_{\nu,\xi}\, d\xi \right|.$$

Finally, suppose $\epsilon \in \mathbb{R}$ and every normalized one-Lipschitz test function satisfies the displayed bound. Applying that bound to the just-constructed $\varphi_{\nu,\xi}$, which is one-Lipschitz and satisfies $\varphi_{\nu,\xi}(0) = 0$, gives

$$W_1(\nu, \xi) = \left| \int \varphi_{\nu,\xi}\, d\nu - \int \varphi_{\nu,\xi}\, d\xi \right| \le \epsilon.$$

$\square$

Lemma 22. [`lem:summary-metric-coordinate-comparison`] (Coordinate metric comparison) *Let d_x and d_z be fixed target- and reference-proxy dimensions, and let s, s' be elements of the five-block observable summary space in Definition 33. Equip this space with its coordinate Euclidean distance ρ_{sum}. Then*

$$d_S(s, s') \le \big(4 \operatorname{Ent}_{d_z,d_x} + d_x\big)\rho_{\mathrm{sum}}(s, s'),$$

where

$$\operatorname{Ent}_{r,c} := c\, r\, \|J_c\|_{\mathrm{op}},$$

and $J_c : \mathbb{R}^c \to (\{1, \dots, c\} \to \mathbb{R})$ is the canonical coordinate identification.

Proof of Lemma 22. Write

$$\varrho := \rho_{\mathrm{sum}}(s, s').$$

For $d \in \mathbb{N}$, put

$$[d]_0 := \{0, 1, \dots, d - 1\};$$

when $d = 0$, this set is empty and the finite sums below have value zero. Write

$$s = (s_1, s_2, s_3, s_4, s_5), \qquad s' = (s'_1, s'_2, s'_3, s'_4, s'_5),$$

where $s_1, \dots, s_4$ and $s'_1, \dots, s'_4$ are $d_z \times d_x$ matrix blocks and $s_5, s'_5 \in \mathbb{R}^{d_x}$, as in Definition 33.

1. Since ρ_{sum} is the Euclidean distance on the full coordinate vector, each coordinate-block projection is nonexpansive. Hence, for each $\ell = 1, 2, 3, 4$,

$$\left(\sum_{\mathfrak{i} \in [d_z]_0,\ \mathfrak{j} \in [d_x]_0} \big((s_\ell)_{\mathfrak{ij}} - (s'_\ell)_{\mathfrak{ij}}\big)^2 \right)^{1/2} \le \varrho,$$

   and

$$\left(\sum_{\mathfrak{j} \in [d_x]_0} \big((s_5)_{\mathfrak{j}} - (s'_5)_{\mathfrak{j}}\big)^2 \right)^{1/2} \le \varrho.$$

   In particular,

$$\big|(s_\ell)_{\mathfrak{ij}} - (s'_\ell)_{\mathfrak{ij}}\big| \le \varrho \quad (\ell = 1, 2, 3, 4;\ \mathfrak{i} \in [d_z]_0,\ \mathfrak{j} \in [d_x]_0),$$

   and

$$\big|(s_5)_{\mathfrak{j}} - (s'_5)_{\mathfrak{j}}\big| \le \varrho \quad (\mathfrak{j} \in [d_x]_0).$$

2. We use the following finite-dimensional entrywise estimate. Let $\Psi \in \mathbb{R}^{r \times c}$, let $\zeta \in \mathbb{R}$ satisfy $0 \le \zeta$, and suppose

$$|\Psi_{\mathfrak{ij}}| \le \zeta \quad (\mathfrak{i} \in [r]_0,\ \mathfrak{j} \in [c]_0).$$

   For the canonical vector $e^{(c)}_{\mathfrak{j}} \in \mathbb{R}^c$,

$$\|\Psi e^{(c)}_{\mathfrak{j}}\|_2 \le \sum_{\mathfrak{i} \in [r]_0} |\Psi_{\mathfrak{ij}}| \le r\, \zeta.$$

Thus, for every $\upsilon \in \mathbb{R}^c$ with $\|\upsilon\|_2 \le 1$,

$$\begin{aligned}\|\Psi\upsilon\|_2 &= \left\|\sum_{\mathrm{j}\in[c]_0} (J_c\upsilon)_\mathrm{j}\, \Psi e_\mathrm{j}^{(c)}\right\|_2 \\ &\le \sum_{\mathrm{j}\in[c]_0} |(J_c\upsilon)_\mathrm{j}|\, \|\Psi e_\mathrm{j}^{(c)}\|_2 \\ &\le c\, r\, \|J_c\|_{\mathrm{op}}\, \zeta.\end{aligned}$$

Using the ordinary matrix action and operator-norm convention in Definitions 8 and 11, taking the supremum over the Euclidean unit ball gives

$$\|\Psi\|_{\mathrm{op}} \le \mathrm{Ent}_{r,c}\ \zeta, \qquad \mathrm{Ent}_{r,c} = c\, r\, \|J_c\|_{\mathrm{op}}.$$

3. Apply the preceding estimate with

$$r = d_z, \qquad c = d_x, \qquad \zeta = \varrho$$

to the four matrix differences $s_\ell - s'_\ell$, $\ell = 1, 2, 3, 4$. The coordinate bounds from the first step yield

$$\|s_\ell - s'_\ell\|_{\mathrm{op}} \le \mathrm{Ent}_{d_z,d_x}\ \varrho, \qquad \ell = 1, 2, 3, 4.$$

4. For the mean block, expand in the canonical basis of $\mathbb{R}^{d_x}$:

$$s_5 - s'_5 = \sum_{\mathrm{j}\in[d_x]_0} \big((s_5)_\mathrm{j} - (s'_5)_\mathrm{j}\big) e_\mathrm{j}^{(d_x)}.$$

The coordinate bounds and the triangle inequality give

$$\|s_5 - s'_5\|_2 \le \sum_{\mathrm{j}\in[d_x]_0} \big|(s_5)_\mathrm{j} - (s'_5)_\mathrm{j}\big| \le d_x\, \varrho.$$

5. Substituting the four matrix-block estimates and the mean-block estimate into the definition of d_S in Definition 33 gives

$$\begin{aligned}d_S(s, s') &= \sum_{\ell=1}^{4} \|s_\ell - s'_\ell\|_{\mathrm{op}} + \|s_5 - s'_5\|_2 \\ &\le 4\ \mathrm{Ent}_{d_z,d_x}\ \varrho + d_x\, \varrho \\ &= \big(4\,\mathrm{Ent}_{d_z,d_x} + d_x\big)\rho_{\mathrm{sum}}(s, s').\end{aligned}$$

□

Lemma 23. [lem:summary-metric-continuity] (Summary metric continuity) *For fixed target- and reference-proxy dimensions d_x, d_z, let d_S be the summary metric on observable moment summaries from Definition 33. The map*

$$(s, q) \longmapsto d_S(s, q)$$

from the product of the five-block summary space with itself to $\mathbb{R}$ is continuous.

Proof of Lemma 23. By Definition 46, for the fixed dimensions d_x and d_z the ambient summary space is

$$\mathscr{S}_{\text{all}} = (\mathbb{R}^{d_z \times d_x})^4 \times \mathbb{R}^{d_x}.$$

Write

$$s = \big(M_0(s), M_1(s), N_0(s), N_1(s), m_X(s)\big), \qquad q = \big(M_0(q), M_1(q), N_0(q), N_1(q), m_X(q)\big).$$

The coordinate map

$$\iota(s) := \big(M_0(s), M_1(s), N_0(s), N_1(s), m_X(s)\big)$$

is continuous for the product topology, and therefore each coordinate projection

$$s \mapsto M_0(s), \quad s \mapsto M_1(s), \quad s \mapsto N_0(s), \quad s \mapsto N_1(s), \quad s \mapsto m_X(s)$$

is continuous. The same finite-product argument covers the zero-dimensional cases, where the relevant products and sums have their usual singleton or empty interpretation.

For a pair (s, q), the four matrix-valued difference maps

$$(s,q) \longmapsto M_0(s) - M_0(q), \quad (s,q) \longmapsto M_1(s) - M_1(q),$$
$$(s,q) \longmapsto N_0(s) - N_0(q), \quad (s,q) \longmapsto N_1(s) - N_1(q)$$

are continuous by continuity of the coordinate projections and of subtraction. If $\Upsilon \in \mathbb{R}^{d_z \times d_x}$, the associated Euclidean linear operator

$$\mathbb{R}^{d_x} \to \mathbb{R}^{d_z}, \qquad x \longmapsto \Upsilon x,$$

depends continuously on Υ, because this matrix-to-operator assignment is linear between finite-dimensional normed spaces. Composing with the continuous operator norm gives continuity of

$$(s,q) \longmapsto \|M_0(s) - M_0(q)\|_{\text{op}}, \quad (s,q) \longmapsto \|M_1(s) - M_1(q)\|_{\text{op}},$$
$$(s,q) \longmapsto \|N_0(s) - N_0(q)\|_{\text{op}}, \quad (s,q) \longmapsto \|N_1(s) - N_1(q)\|_{\text{op}}.$$

The mean-coordinate difference

$$(s,q) \longmapsto m_X(s) - m_X(q)$$

is continuous as a map into $\mathbb{R}^{d_x}$. Its Euclidean norm is

$$\|m_X(s) - m_X(q)\|_2 = \left(\sum_{i<d_x} \big(m_X(s)_i - m_X(q)_i\big)^2 \right)^{1/2}.$$

The summand coordinates are continuous, the index set is finite, and the square-root is continuous on the nonnegative range of the sum of squares. Hence

$$(s,q) \longmapsto \|m_X(s) - m_X(q)\|_2$$

is continuous.

Using Definition 33,

$$d_S(s,q) = \|M_0(s)-M_0(q)\|_{\text{op}} + \|M_1(s)-M_1(q)\|_{\text{op}} + \|N_0(s)-N_0(q)\|_{\text{op}} + \|N_1(s)-N_1(q)\|_{\text{op}} + \|m_X(s)-m_X(q)\|_2.$$

This is a finite sum of continuous real-valued functions on $\mathscr{S}_{\text{all}} \times \mathscr{S}_{\text{all}}$. Therefore

$$(s,q) \longmapsto d_S(s,q)$$

is continuous. $\square$

Lemma 24. [`lem:stacked-proxy-moment-signal-margin`] (Stacked proxy signal margin) *Let P, a full-data probability law carrying $(U, T, X, Z, Y(0), Y(1), Y)$, belong to the uniformly conditioned proxy causal model class $\mathcal{M}$ in Definition 32. Suppose:*

- ***(Dimensions.)*** *$2 \le k \le d_x$, where k is the number of latent classes and d_x is the target-proxy dimension.*
- ***(Radius.)*** *$L \ge 1$, with the bounded target-proxy condition in Assumption 6.*
- ***(Margins.)*** *$\pi_0 > 0$ and $\sigma_0 > 0$, with the latent-arm positivity and proxy-rank margin conditions in Assumptions 9 and 10.*

Then the vertically stacked observable proxy moment satisfies

$$\pi_0 \sigma_0^2 \le \sigma_{k-1}\{\operatorname{stack}(M_0(P), M_1(P))\},$$

where σ_{k-1} uses zero-based singular-value indexing and $M_0(P), M_1(P)$ are the armwise $ZX^\top$ population moments from Definition 28.

Proof of Lemma 24. Set

$$A_0 := \left(E_P[Z_i \mid U = u, T = 0]\right)_{i,u}, \qquad B := \left(E_P[X_i \mid U = u]\right)_{i,u},$$

and

$$\Gamma_0 := \operatorname{diag}\left\{\frac{P(U = u, T = 0)}{P(T = 0)} : 1 \le u \le k\right\}.$$

Let $S = \operatorname{range}(B) \subseteq \mathbb{R}^{d_x}$. The proxy-rank margin in Assumption 10 gives

$$\sigma_{k-1}(A_0) \ge \sigma_0, \qquad \sigma_{k-1}(B) \ge \sigma_0.$$

Since $\sigma_0 > 0$, the maps induced by A_0 and B are injective. Hence

$$\dim S = k.$$

The latent-arm positivity condition gives $P(U = u, T = 0) \ge \pi_0$ for every u. Since $0 < P(T = 0) \le 1$, each diagonal entry of Γ_0 is at least π_0. Therefore, for every $y \in \mathbb{R}^k$,

$$\pi_0 \|y\|_2 \le \|\Gamma_0 y\|_2.$$

Now fix $x \in S$. The least-singular-value bound for B, restricted to its range, gives

$$\sigma_0 \|x\|_2 \le \|B^\top x\|_2.$$

Applying the preceding diagonal bound to $y = B^\top x$ gives

$$\pi_0 \|B^\top x\|_2 \le \|\Gamma_0 B^\top x\|_2.$$

The least-singular-value bound for A_0 gives

$$\sigma_0 \|\Gamma_0 B^\top x\|_2 \le \|A_0 \Gamma_0 B^\top x\|_2.$$

Combining these three inequalities,

$$\pi_0\sigma_0^2\|x\|_2 \le \|A_0\Gamma_0 B^\top x\|_2.$$

By the observable proxy moment factorization in Lemma 4,

$$M_0(P) = A_0\Gamma_0 B^\top.$$

Thus

$$\pi_0\sigma_0^2\|x\|_2 \le \|M_0(P)x\|_2.$$

For

$$G := \operatorname{stack}\big(M_0(P), M_1(P)\big),$$

vertical stacking gives the Pythagorean identity

$$\|Gx\|_2^2 = \|M_0(P)x\|_2^2 + \|M_1(P)x\|_2^2,$$

and hence

$$\|M_0(P)x\|_2 \le \|Gx\|_2.$$

Consequently,

$$\pi_0\sigma_0^2\|x\|_2 \le \|Gx\|_2 \qquad \text{for every } x \in S.$$

Since S is a k-dimensional subspace and $\pi_0\sigma_0^2 \ge 0$, the variational subspace characterization of zero-indexed singular values yields

$$\pi_0\sigma_0^2 \le \sigma_{k-1}(G) = \sigma_{k-1}\big\{\operatorname{stack}\big(M_0(P), M_1(P)\big)\big\}.$$

□

These witness calculations locate the explicit two-class comparisons inside the same model restrictions used for the upper bounds. The path-local statement calibrates the observed Kullback–Leibler radius to the product g^2h^2, which is the scale used by Theorem 6.

The final two auxiliary statements concern ordered masses as functions of represented atomic laws. They ensure that, when the support is separated, the ordered-weight target depends on the aggregate measure and is invariant across labelled coordinate representations.

Lemma 25. [`lem:model-compressed-spectral-certificate`] (Compressed spectral certificate) *Let $k, d_x, d_z \in \mathbb{N}$, let $L, \pi_0, \sigma_0 \in \mathbb{R}$, and let Q be a full-data probability law in the uniformly conditioned model class $\mathcal{M}$ of Definition 32, with observable summary $S(Q)$ as in Definition 33. Assume that Q satisfies Assumptions 5, 9 and 10. Let $M_t(S(Q))$ be the t-arm $ZX^\top$ moment block from Definition 28, and set*

$$G := \operatorname{stack}(M_0(S(Q)), M_1(S(Q))).$$

Then there exist a signal basis V spanning the signal row space of $S(Q)$ and a real diagonalization of the compressed effect operator $\Delta Q(Q)$ from Definition 39 such that:

1. *for each arm $t \in \{0, 1\}$, the compressed map*

   $$x \longmapsto M_t(S(Q))Vx$$

   is injective;

2. *the stacked proxy moment has rank*

$$\operatorname{rank}(G) = k;$$

3. *for every index j,*

$$\frac{\pi_0 \sigma_0^2}{2} \le \sigma_j(G) \iff j < k;$$

4. *the eigenvalue vector of the diagonalization is the latent-effect vector $(\tau_u(Q))_u$ from Definition 23;*

5. *writing R for the diagonalizing basis matrix, the left and right spectral anchors satisfy*

$$a(Q) = R^{-\top} p(Q), \qquad R^{-1} c(Q) = \mathbf{1}_k,$$

where $p(Q) = (p_u(Q))_u$ is the latent-mass vector from Definition 21.

Proof of Lemma 25. 1. Write

$$B := B(Q) \in \mathbb{R}^{d_x \times k}$$

for the target-proxy feature matrix from Definition 25. The parameter domain in Definition 32 gives $2 \le k$, $k \le d_x$, $1 \le L$, $\pi_0 > 0$, and $\sigma_0 > 0$. With zero-based singular-value indexing, Assumption 10 gives

$$\sigma_{k-1}(B) \ge \sigma_0 > 0.$$

Hence the first k singular values of B are positive. Choose the thin singular-value factorization

$$B = V \Upsilon,$$

where $V \in \mathbb{R}^{d_x \times k}$ has orthonormal columns and $\Upsilon \in \mathbb{R}^{k \times k}$ is invertible. Put

$$\Theta := B^\top V = \Upsilon^\top, \qquad \Theta^{-1} := \Upsilon^{-T}.$$

Then

$$\Theta \Theta^{-1} = \Theta^{-1} \Theta = I_k.$$

2. For each arm $t \in \{0, 1\}$, define

$$\Gamma_t := \operatorname{diag}\left\{ \frac{Q(U = u, T = t)}{Q(T = t)} : 1 \le u \le k \right\}, \qquad \Xi_t := A_t \Gamma_t,$$

where $A_t = A_t(Q)$ is the reference-proxy feature matrix from Definition 24. At the model-generated summary $S(Q)$, Lemmas 4 and 18 give

$$M_t(S(Q)) = A_t \Gamma_t B^\top = \Xi_t B^\top$$

and

$$N_t(S(Q)) = A_t \Gamma_t \operatorname{diag}\{\mu_{tu}(Q) : 1 \le u \le k\} B^\top = \Xi_t \operatorname{diag}\{\mu_{tu}(Q) : 1 \le u \le k\} B^\top.$$

Multiplying by V gives

$$M_t(S(Q)) V = \Xi_t \Theta, \qquad N_t(S(Q)) V = \Xi_t \operatorname{diag}\{\mu_{tu}(Q) : 1 \le u \le k\} \Theta.$$

Also,
$$M_t(S(Q))^\top = B\Xi_t^\top = V\Upsilon\Xi_t^\top,$$
so the row space of each $M_t(S(Q))$ is contained in $\mathrm{range}(V)$. For $t = 0$, latent-arm positivity in Assumption 9 gives
$$\Gamma_0(u,u) = \frac{Q(U=u, T=0)}{Q(T=0)} \geq \pi_0$$
for every u, since $Q(U=u, T=0) \geq \pi_0$ and $Q(T=0) \leq 1$. Combining this with Assumption 10 and the standard product lower bound for the least signal singular value,
$$\begin{aligned}\sigma_{k-1}(M_0(S(Q))) &= \sigma_{k-1}\Big(A_0\Gamma_0 B^\top\Big)\\ &\geq \sigma_{k-1}(A_0)\,\sigma_{k-1}(\Gamma_0)\,\sigma_{k-1}(B) \geq \pi_0\sigma_0^2.\end{aligned}$$
Thus $M_0(S(Q))$ has row rank k. Its row space is contained in the k-dimensional space $\mathrm{range}(V)$, hence
$$\mathrm{range}\Big(M_0(S(Q))^\top\Big) = \mathrm{range}(V).$$
Since the row space of $M_1(S(Q))$ is contained in the same space, the signal row space of $S(Q)$ equals
$$\mathrm{range}\Big(M_0(S(Q))^\top\Big) + \mathrm{range}\Big(M_1(S(Q))^\top\Big) = \mathrm{range}(V).$$
Thus V spans the signal row space of $S(Q)$.

3. The same product-margin argument for each arm t gives
$$\pi_0\sigma_0^2 \leq \sigma_{k-1}(M_t(S(Q)))\,.$$
Because V is an orthonormal basis for the signal row space, compression by V preserves the k nonzero signal singular values:
$$\sigma_{k-1}(M_t(S(Q))V) = \sigma_{k-1}(M_t(S(Q)))\,.$$
Consequently
$$0 < \pi_0\sigma_0^2 \leq \sigma_{k-1}(M_t(S(Q))V)\,,$$
and the linear map
$$x \longmapsto M_t(S(Q))Vx, \qquad x \in \mathbb{R}^k,$$
is injective for each $t \in \{0,1\}$.

4. Let
$$G := \mathrm{stack}\big(M_0(S(Q)), M_1(S(Q))\big).$$
By Lemma 24,
$$\pi_0\sigma_0^2 \leq \sigma_{k-1}(G),$$
so $\mathrm{rank}(G) \geq k$. On the other hand, if
$$\Xi_{\mathrm{st}} := \mathrm{stack}(\Xi_0, \Xi_1) \in \mathbb{R}^{2d_z \times k},$$

then the proxy factorizations give

$$G = \Xi_{\mathrm{st}} B^\top.$$

Hence $\mathrm{rank}(G) \le k$, and therefore

$$\mathrm{rank}(G) = k.$$

The threshold characterization follows immediately. If $j < k$, monotonicity of singular values gives

$$\sigma_j(G) \ge \sigma_{k-1}(G) \ge \pi_0 \sigma_0^2 \ge \frac{\pi_0 \sigma_0^2}{2}.$$

If $k \le j$, the rank identity gives $\sigma_j(G) = 0$, while

$$0 < \frac{\pi_0 \sigma_0^2}{2}.$$

Thus, for every singular-value index j,

$$\frac{\pi_0 \sigma_0^2}{2} \le \sigma_j(G) \quad \Longleftrightarrow \quad j < k.$$

5. Since $M_t(S(Q))V$ is injective, its Moore–Penrose inverse is a left inverse on its range:

$$\{M_t(S(Q))V\}^\dagger M_t(S(Q))V = I_k.$$

Using the displayed factorizations,

$$N_t(S(Q))V = \{M_t(S(Q))V\}\, \Theta^{-1} \operatorname{diag}\{\mu_{tu}(Q) : 1 \le u \le k\}\Theta.$$

Therefore

$$\{M_t(S(Q))V\}^\dagger N_t(S(Q))V = \Theta^{-1} \operatorname{diag}\{\mu_{tu}(Q) : 1 \le u \le k\}\Theta.$$

By the compressed-operator formula in Definition 39,

$$\begin{aligned}
\Delta Q(Q) &= \Theta^{-1} \operatorname{diag}\{\mu_{1u}(Q) : 1 \le u \le k\}\Theta - \Theta^{-1} \operatorname{diag}\{\mu_{0u}(Q) : 1 \le u \le k\}\Theta \\
&= \Theta^{-1} \operatorname{diag}\{\mu_{1u}(Q) - \mu_{0u}(Q) : 1 \le u \le k\}\Theta \\
&= \Theta^{-1} \operatorname{diag}\{\tau_u(Q) : 1 \le u \le k\}\Theta,
\end{aligned}$$

where the last equality is Definition 23. Thus $R := \Theta^{-1}$ is a real diagonalizing basis for $\Delta Q(Q)$, its inverse is $R^{-1} = \Theta$, and the eigenvalue vector is $(\tau_u(Q))_u$.

6. It remains to identify the spectral anchors. The target-proxy mean factors through the latent masses:

$$m_X(Q) = B\, p(Q).$$

Indeed, for each target-proxy coordinate i,

$$m_X(Q)_i = E_Q[X_i] = \sum_{u=1}^{k} E_Q[X_i \mid U = u]\, Q(U = u) = \sum_{u=1}^{k} B(i, u)\, p_u(Q),$$

using Definitions 21 and 25. The anchor condition in Assumption 5, together with Definitions 25 and 36, gives

$$B^\top e_1 = \mathbf{1}_k,$$

because for each latent class u,

$$(B^\top e_1)_u = B(1,u) = E_Q[X_1 \mid U = u] = 1.$$

By Definition 39,

$$a(Q) = V^\top m_X(Q), \qquad c(Q) = V^\top e_1.$$

Using $B = V\Upsilon$, $V^\top V = I_k$, and $R^{-1} = \Theta = \Upsilon^\top$,

$$a(Q) = V^\top B\, p(Q) = \Upsilon p(Q),$$

and hence

$$a(Q)^\top = p(Q)^\top \Upsilon^\top = p(Q)^\top R^{-1}.$$

Similarly,

$$R^{-1} c(Q) = \Theta V^\top e_1 = \Upsilon^\top V^\top e_1 = B^\top e_1 = \mathbf{1}_k.$$

These are the claimed left and right spectral-anchor identities.

□

Lemma 26. [`lem:diagonalization-condition-number-bound`] (Diagonalization condition bound) *Let $d_x \geq 1$ and $k \geq 1$. Let $B \in \mathbb{R}^{d_x \times k}$, and let F be a thin signal factorization of B, written $B = V\Upsilon$, with coordinate inverse Υ^{-1}. Let $\tau \in \mathbb{R}^k$. Suppose that*

- ***(Feature bound.)*** *For a constant $L \geq 0$ as in Assumption 6, $|B_{iu}| \leq L$ for every row i and latent index u.*
- ***(Coordinate-inverse bound.)*** *For a constant $\sigma_0 > 0$ as in Assumption 10, $|(\Upsilon^{-1})_{ru}| \leq \sigma_0^{-1}$ for every pair of coordinate indices r, u.*

Define

$$K_{\text{forw}} := \operatorname{Ent}_{d_x,k} \operatorname{Ent}_{k,k} \operatorname{Ent}_{k,d_x} \sigma_0^{-1} + \big(\operatorname{Ent}_{d_x,d_x} + \operatorname{Ent}_{d_x,k} \operatorname{Ent}_{k,d_x}\big),$$
$$K_{\text{back}} := \operatorname{Ent}_{d_x,k} \operatorname{Ent}_{k,k} \operatorname{Ent}_{k,d_x} (d_x L) + \big(\operatorname{Ent}_{d_x,d_x} + \operatorname{Ent}_{d_x,k} \operatorname{Ent}_{k,d_x}\big),$$

and

$$\kappa_\star := \operatorname{Ent}^2_{d_x,d_x} K_{\text{forw}} K_{\text{back}},$$

where $\operatorname{Ent}_{r,c}$ is the entrywise-to-operator norm constant of Lemma 22. Then the real diagonalization associated with F and τ has condition number at most $\kappa_\star$.

Proof of Lemma 26. 1. Write

$$E_{r,c} := \operatorname{Ent}_{r,c}\,.$$

For these constants we use the following finite-dimensional entrywise comparison. If $\mathsf{A}_{\text{ent}} \in \mathbb{R}^{r\times c}$, $0 \leq M_{\text{ent}}$, and $|(\mathsf{A}_{\text{ent}})_{ab}| \leq M_{\text{ent}}$ for all entries, then

$$\|\mathsf{A}_{\text{ent}}\|_{\text{op}} \leq E_{r,c} M_{\text{ent}}.$$

Indeed, writing $e_j^{(c)}$ for the standard coordinate vectors and J_c for the coordinate identification in the definition of $\mathrm{Ent}_{r,c}$, each column satisfies

$$\|\mathsf{A}_{\mathrm{ent}} e_j^{(c)}\|_2 \le r M_{\mathrm{ent}}.$$

Thus, for every $x \in \mathbb{R}^c$,

$$\begin{aligned}\|\mathsf{A}_{\mathrm{ent}} x\|_2 &\le \sum_{j=1}^{c} |(J_c x)_j| \, \|\mathsf{A}_{\mathrm{ent}} e_j^{(c)}\|_2 \\ &\le c\, r\, M_{\mathrm{ent}} \|J_c x\|_2 \\ &\le E_{r,c} M_{\mathrm{ent}} \|x\|_2,\end{aligned}$$

which gives the displayed operator-norm bound. In particular $E_{r,c} \ge 0$. Since the columns of V are orthonormal, every coordinate of V has absolute value at most one, and therefore

$$\|V\|_{\mathrm{op}} \le E_{d_x,k}, \qquad \|V^\top\|_{\mathrm{op}} \le E_{k,d_x}.$$

The same comparison gives

$$\|I_{d_x}\|_{\mathrm{op}} \le E_{d_x,d_x}.$$

2. Let Υ and Υ^{-1} be the coordinate matrix and coordinate inverse from the thin factorization $B = V\Upsilon$. Since $V^\top V = I_k$,

$$\Upsilon = V^\top B.$$

Thus, for coordinate indices r, u,

$$|\Upsilon_{ru}| = \left| \sum_{i=1}^{d_x} V_{ir} B_{iu} \right| \le \sum_{i=1}^{d_x} |V_{ir}| \, |B_{iu}| \le \sum_{i=1}^{d_x} L = d_x L.$$

Consequently,

$$\|\Upsilon^\top\|_{\mathrm{op}} \le E_{k,k}(d_x L), \qquad \|(\Upsilon^{-1})^\top\|_{\mathrm{op}} \le E_{k,k}\sigma_0^{-1}.$$

3. Define the forward and backward ambient maps

$$\Phi_{\mathrm{forw}} := V(\Upsilon^{-1})^\top V^\top + (I_{d_x} - VV^\top), \qquad \Phi_{\mathrm{back}} := V\Upsilon^\top V^\top + (I_{d_x} - VV^\top).$$

Using submultiplicativity, the triangle inequality, and the bounds above,

$$\begin{aligned}\|\Phi_{\mathrm{forw}}\|_{\mathrm{op}} &\le \|V(\Upsilon^{-1})^\top V^\top\|_{\mathrm{op}} + \|I_{d_x} - VV^\top\|_{\mathrm{op}} \\ &\le E_{d_x,k} E_{k,k} E_{k,d_x} \sigma_0^{-1} + \big(E_{d_x,d_x} + E_{d_x,k} E_{k,d_x}\big) = K_{\mathrm{forw}},\end{aligned}$$

and similarly

$$\begin{aligned}\|\Phi_{\mathrm{back}}\|_{\mathrm{op}} &\le \|V\Upsilon^\top V^\top\|_{\mathrm{op}} + \|I_{d_x} - VV^\top\|_{\mathrm{op}} \\ &\le E_{d_x,k} E_{k,k} E_{k,d_x} (d_x L) + \big(E_{d_x,d_x} + E_{d_x,k} E_{k,d_x}\big) = K_{\mathrm{back}}.\end{aligned}$$

Both constants are nonnegative because $L \ge 0$, $\sigma_0 > 0$, and $E_{r,c} \ge 0$.

4. Let $(q_j)_{j=1}^{d_x}$ be the ambient orthonormal extension of the signal columns used in the associated real diagonalization, and let the diagonalizing basis vectors be

$$s_j := \Phi_{\mathrm{forw}} q_j, \qquad j = 1, \ldots, d_x.$$

The backward map is the inverse of the forward map. To see this, put $P_\perp := I_{d_x} - VV^\top$. Then

$$V^\top P_\perp = 0, \qquad P_\perp V = 0, \qquad P_\perp^2 = P_\perp,$$

and the identities $\Upsilon\Upsilon^{-1} = I_k$ and $\Upsilon^{-1}\Upsilon = I_k$ give

$$\Phi_{\mathrm{forw}}\Phi_{\mathrm{back}} = VV^\top + P_\perp = I_{d_x}, \qquad \Phi_{\mathrm{back}}\Phi_{\mathrm{forw}} = VV^\top + P_\perp = I_{d_x}.$$

Hence the inverse-coordinate matrix of the basis $(s_j)_j$ is computed by first applying Φ_{back} and then taking coordinates in the orthonormal basis $(q_j)_j$.

5. Let $\mathsf{Basis}_{\mathrm{diag}}$ be the matrix with columns s_j, and let $\mathsf{Basis}_{\mathrm{diag}}^{-1}$ be its coordinate inverse. For every pair of ambient indices i, j,

$$|(\mathsf{Basis}_{\mathrm{diag}})_{ij}| = |(s_j)_i| \le \|s_j\|_2 = \|\Phi_{\mathrm{forw}} q_j\|_2 \le \|\Phi_{\mathrm{forw}}\|_{\mathrm{op}} \|q_j\|_2 \le K_{\mathrm{forw}},$$

because $\|q_j\|_2 = 1$. Therefore

$$\|\mathsf{Basis}_{\mathrm{diag}}\|_{\mathrm{op}} \le E_{d_x,d_x} K_{\mathrm{forw}}.$$

Likewise, for the inverse-coordinate entries,

$$|(\mathsf{Basis}_{\mathrm{diag}}^{-1})_{ij}| = |\langle q_i, \Phi_{\mathrm{back}} e_j\rangle| \le \|q_i\|_2 \|\Phi_{\mathrm{back}} e_j\|_2 \le K_{\mathrm{back}},$$

so

$$\|\mathsf{Basis}_{\mathrm{diag}}^{-1}\|_{\mathrm{op}} \le E_{d_x,d_x} K_{\mathrm{back}}.$$

The condition number of the associated real diagonalization is the product of these two operator norms, hence

$$\mathrm{cond} \le \big(E_{d_x,d_x} K_{\mathrm{forw}}\big)\big(E_{d_x,d_x} K_{\mathrm{back}}\big) = E_{d_x,d_x}^2 K_{\mathrm{forw}} K_{\mathrm{back}} = \kappa_\star.$$

□

The definition matches the ordering convention in Definition 30 and Algorithm 5. The invariance lemma records the compatibility needed to read ordered masses from any valid representative of the same aggregated atomic law.

B Spectral repair and finite-net benchmark

This appendix collects two auxiliary constructions used to connect the population modulus with sample-level procedures. The first is a spectral repair handle that turns empirical spectral information into a positive real atomic law even when eigenvalues are locally coalescent. The second is a class-advised finite-net benchmark that selects a stored admissible summary and then runs the exact spectral construction at that representative. Both constructions use finite-dimensional perturbation ideas familiar from matrix spectral analysis [Bauer and Fike, 1960, Kato, 1995, Davis and Kahan, 1970, Stewart and Sun, 1990] and the moment-recovery perspective behind Wasserstein estimation from moments [Wu and Yang, 2020].

The repair handle records the finite moment target and the positive-law replacement used in the modulus-to-estimation route.

Algorithm 6 (Constructive repair handle m)

The constructive repair handle is defined for laws in Definition 32 as follows.

1. At the realized empirical summary, define the first $2k$ spectral moments by

$$m = (m_j)_{0 \le j < 2k}.$$

2. Among all valid positive k-atomic laws $\sum_i w_i \delta_{x_i}$, choose a minimizer of

$$\sum_{j=0}^{2k-1} \left| m_j - \sum_i w_i x_i^j \right|.$$

3. Extract the repaired real probability measure from the minimizing positive k-atomic law, transporting aggregate projector mass across each cluster diameter after replacing individual eigenvectors inside overlapping localization discs by Kato contour projectors.

4. For the labeled-weight converse, differentiate the Bernoulli factorization of $P_\varepsilon^{\mathrm{wit}}$ and solve the linear score-cancellation equations for the proxy and treatment primitives so that a weight displacement h preserves the factorization and the observed score has order gh.

Algorithm 6 isolates the spectral information needed for positive real-law repair. The moment vector m contains the first $2k$ empirical spectral moments, and the minimizing positive atomic law supplies a real probability measure compatible with those moments. The contour-projector clause expresses the standard perturbation treatment of clustered spectra: inside overlapping localization discs, aggregate eigenspaces carry the stable law-level mass.

The finite-net benchmark makes this repair logic an exact-real finite search after a class-dependent representative library has been fixed.

Algorithm 7 (Finite-net program $\mathrm{NetProgram}(A, \omega)$)

The inputs are a stored representative library $A = (A_i)_i$ and an observed sample ω.

1. For each representative A_i, let $S_i = (M_{0,i}, M_{1,i}, N_{0,i}, N_{1,i}, m_{X,i})$ be its stored five-block summary, and let

$$\mathrm{rank}_{\mathrm{lex}}(i)$$

be its position in the fixed lexicographic library order.

2. Define the spectral law attached to representative A_i by

$$\lambda^{\mathrm{spec}}(A_i),$$

the real atomic effect law returned by the thresholded exact-real spectral run for A_i.

3. For the empirical five-block summary

$$\widehat{S}_n(\omega) = (\widehat{M}_{0,n}, \widehat{M}_{1,n}, \widehat{N}_{0,n}, \widehat{N}_{1,n}, \widehat{m}_{X,n}),$$

define the comparison criterion

$$\Delta_i(\omega) := \|M_{0,i} - \widehat{M}_{0,n}\|_{\mathrm{op}} + \|M_{1,i} - \widehat{M}_{1,n}\|_{\mathrm{op}} + \|N_{0,i} - \widehat{N}_{0,n}\|_{\mathrm{op}} + \|N_{1,i} - \widehat{N}_{1,n}\|_{\mathrm{op}} + \|m_{X,i} - \widehat{m}_{X,n}\|_2.$$

4. If the library is nonempty, define the selected representative index by

$$\widehat{i}(\omega) \in \arg\min_i \big(\Delta_i(\omega), \mathrm{rank}_{\mathrm{lex}}(i)\big),$$

with the minimum taken in lexicographic order.

5. The finite-library spectral program returns the selected representative and its spectral law,

$$\mathrm{NetProgram}(A, \omega) := \big(A_{\widehat{i}(\omega)}, \lambda^{\mathrm{spec}}(A_{\widehat{i}(\omega)})\big),$$

when the library is nonempty, and returns the canonical zero law when the library is empty. Its returned effect-law component is

$$\mathrm{law}(\mathrm{NetProgram}(A, \omega)).$$

Algorithm 7 defines the library selector at the level of stored summaries. The criterion is the same five-block d_S-type comparison used throughout the paper, and lexicographic tie-breaking makes the selected representative a total deterministic function of the sample and the library.

The estimator below specifies how the representative library is built from an ϵ_n-net of admissible summaries and how the selected representative is converted into an atomic law.

Algorithm 8 (Finite-net estimator $\widehat{\nu}_n^{\mathrm{net}}$)

The input is the empirical summary $\widehat{S}_n$; the tuning constants are

$$D := 4d_z d_x + d_x, \qquad b := 4\sqrt{d_z d_x} + \sqrt{d_x}, \qquad \epsilon_n := n^{-1/2}.$$

1. Partition the coordinate box $[-L, L]^D$ into a deterministic lexicographically ordered grid of half-open cubes whose side length is at most

$$\epsilon_n / b.$$

2. If $\mathscr{S} = \varnothing$, set

$$\widehat{\nu}_n^{\mathrm{net}} := \delta_0$$

and stop.

3. For each grid cube that intersects $\mathscr{S}$, choose one representative

$$q_{n,\ell} \in \mathscr{S}.$$

Let $\widehat{q}_n$ be the smallest-index minimizer

$$\widehat{q}_n \in \arg\min_{\ell} d_S(q_{n,\ell}, \widehat{S}_n).$$

4. At $\widehat{q}_n$, form the vertically stacked proxy moment and retain the k right singular directions whose singular values exceed $\pi_0\sigma_0^2/2$. Denote the resulting orthonormal basis by

$$V.$$

Using this basis, form the compressed effect operator, left anchor, and right anchor

$$\Delta, \qquad a, \qquad c$$

from the compressed-operator formulas.

5. For each distinct eigenvalue λ of Δ, define the spectral projector

$$E_\lambda := \prod_{\gamma\in\operatorname{spec}(\Delta):\,\gamma\neq\lambda} \frac{\Delta-\gamma I_k}{\lambda-\gamma},$$

with the empty product equal to I_k.

6. Output the atomic law

$$\widehat{\nu}_n^{\mathrm{net}} := \sum_{\lambda\in\operatorname{spec}(\Delta)} a^\top E_\lambda c\,\delta_\lambda.$$

The estimator $\widehat{\nu}_n^{\mathrm{net}}$ is class-advised through the choice of admissible representatives $q_{n,\ell}$. Once the nearest representative is selected, the construction returns to the compressed-operator formula from Definition 39: thresholding fixes the rank-k signal space, and the spectral projectors aggregate the anchor weights attached to each distinct eigenvalue.

Theorem 8. [`thm:polynomial-net-law-estimator`] (Polynomial finite-net estimator) *Fix integers k, d_x, d_z and real constants L, π_0, σ_0. Suppose*

- ***(Dimensions.)*** $2 \le k$, $k \le d_x$, *and* $k \le d_z$.
- ***(Bounds and margins.)*** $1 \le L$, $0 < \pi_0 \le (2k)^{-1}$, *and* $0 < \sigma_0 \le 1$.

Then there are constants $C_{\mathrm{mod}} > 0$ and $C > 0$ such that, for every integer $n \ge 1$, there is a class-dependent finite library $A = (A_i)_i$ of representative summaries as in Algorithm 8 with the following properties. For every uniform provider of the fixed-parameter exact-real primitives on the admissible core parameter domain, the finite-net estimator $\widehat{\nu}_n^{\mathrm{net}}$ is Borel measurable, and

$$|A| \le C n^{(4d_z d_x + d_x)/2}.$$

For every observed sample $\omega = (O_1, \ldots, O_n) \in O^n_{d_x,d_z}$*, the operation count of the same exact-real execution that returns* $\widehat{\nu}^{\mathrm{net}}_n(\omega)$ *is at most*

$$C\{n + n^{(4d_z d_x + d_x)/2}\}.$$

Its trace is exactly the empirical-summary trace, followed by the exhaustive finite-library search trace, and followed by the selected spectral-run trace when a representative is selected; if no representative is selected, the final spectral-run trace is empty.

If the program selects index i *on sample* ω*, then* S_i *is a nearest library summary to the empirical summary* $\widehat{S}_n(\omega)$ *of Definition 35:*

$$d_S(S_i, \widehat{S}_n(\omega)) \leq d_S(S_j, \widehat{S}_n(\omega)) \qquad \text{for every } j.$$

Among all tied nearest summaries, i *has minimal lexicographic rank:*

$$d_S(S_i, \widehat{S}_n(\omega)) = d_S(S_j, \widehat{S}_n(\omega)) \quad \Longrightarrow \quad \mathrm{rank}_{\mathrm{lex}}(i) \leq \mathrm{rank}_{\mathrm{lex}}(j).$$

For the exact spectral run at S_i*, the estimator equals the run's returned effect law. For each treatment value* t*, the observed proxy moment at* S_i*, multiplied by the run's basis matrix, defines an injective Euclidean linear map. The stacked proxy moment at* S_i *satisfies the exact threshold characterization*

$$\frac{\pi_0\sigma_0^2}{2} \leq \sigma_j(\mathrm{stack}(M_0(S_i), M_1(S_i))) \quad \Longleftrightarrow \quad j < k,$$

for every singular-value index j*. Every eigenvalue of the compressed operator formed at* S_i *with the run's basis and spans appears among the eigenvalues returned by that run.*

For every stored representative S_i *and every rectangular perturbation* $H \in \mathbb{R}^{2d_z \times d_x}$*,*

$$\|H\|_{\mathrm{op}} < \frac{\pi_0\sigma_0^2}{2}$$

implies that singular-value thresholding at $\pi_0\sigma_0^2/2$ *recovers exactly the* k*-dimensional matrix signal space for*

$$\mathrm{stack}(M_0(S_i), M_1(S_i)) + H.$$

For every index i*, there is a model law* $Q \in \mathcal{M}$ *from Definition 32 whose observable summary is* S_i*, and the exact spectral run at* S_i *returns the quotient law* ν_Q*. Each returned representative effect law is a valid atomic law.*

For every $P \in \mathcal{M}$*, every observed sample* $\omega \in O^n_{d_x,d_z}$ *satisfies*

$$W_1\big(\widehat{\nu}^{\mathrm{net}}_n(\omega), \nu_P\big) \leq C_{\mathrm{mod}}\left\{2d_S(\widehat{S}_n(\omega), S(P)) + n^{-1/2}\right\}.$$

Finally, for every tail level η *with* $0 < \eta < 1/2$ *and every* $P \in \mathcal{M}$*,*

$$Q_P^{(n)}\left(W_1\big(\widehat{\nu}^{\mathrm{net}}_n, \nu_P\big) > C\sqrt{\frac{\log(C/\eta)}{n}}\right) \leq \eta.$$

Theorem 8 establishes the finite-net benchmark. For fixed dimensions and fixed conditioning constants, the representative library has polynomial size in n, the online exact-real operation count is polynomial in the same sense, and the selected summary is the lexicographically first nearest library representative. The deterministic W_1 inequality converts empirical summary error plus the $n^{-1/2}$ mesh scale into quotient-law error, and the final tail bound applies the same concentration scale used for the main law estimator.

The intuition is that the admissible summary image is compact and finite-dimensional, so an $n^{-1/2}$-mesh supplies a representative close enough for the gap-free modulus to absorb the discretization error. At the chosen representative, the proxy-rank margin makes singular-value thresholding recover the correct k-dimensional signal space under perturbations below $\pi_0\sigma_0^2/2$. The exact spectral run then returns the quotient law for a model law with that stored summary, and the modulus transfers the distance from the empirical summary to W_1 distance between effect laws.

C Proofs and verification note

This appendix collects the proof layer for the paper's formal results. The arguments establish the population stability statement in Theorem 1, the measurable repair and estimation claims in Propositions 3 and 4 and Theorem 2, the confidence and reporting guarantees in Theorems 3 to 5, and the local converse statements in Theorems 6 and 7 and Propositions 6 and 7.

The proof of Theorem 1 uses the compact summary closure from Proposition 2 and the quotient moment representation in Definition 43. The finite-dimensional moment map gives a Lipschitz modulus from the observable summary coordinates to the W_1 law target on the compact domain. The extension clause follows by completing the admissible image under d_S, with Algorithm 1 and Proposition 3 supplying the measurable nearest-summary selection needed to evaluate the extended map at empirical summaries.

The estimation proofs combine this deterministic modulus with concentration for the empirical five-block summary in Definition 35. For the repaired estimator, nearest-summary projection keeps the evaluated summary within the same concentration scale as the population summary. For the structured-lattice estimator, Algorithm 2 and Proposition 4 give a duplicate-free finite search whose deterministic error is bounded by the empirical summary error and the $n^{-1/2}$ lattice scale. These two routes yield the joint root-n tail statement in Theorem 2.

The confidence-set proofs use the same concentration event with the objects in Algorithm 3. The summary-inversion set inherits its diameter from the modulus on $\mathcal{K}$, while the algorithmic outer set inherits its radius from the deterministic lattice bound. The transport-plan formulation in Definition 2 gives the finite representation used in Theorem 3. The cluster-report argument then applies the atom floor and W_1 transport budget in Algorithm 4 to associate true atoms with empirical components, yielding the support containment, mass-coverage, and interval-length conclusions in Theorem 4.

For ordered weights, the proof of Theorem 5 works on the gap-local stratum in Definition 41. The positive effect gap converts law-level transport error into an ℓ_1 error bound for the effect-ordered mass vector defined in Definition 30 and estimated by Algorithm 5. This gives the clipped inverse-gap upper rate stated in the theorem.

The lower-bound proofs use the explicit two-class witness in Definition 55, its validity certificate in Proposition 5, and the local experiments in Definitions 56 and 57. Standard two-point testing inequalities [Cam, 1986, Polyanskiy and Wu, 2019] convert the displayed Kullback–Leibler bounds and target separations into the local risk lower bounds in Theorem 6. The same comparisons give the same-class minimax rates in Propositions 6 and 7. The transfer statement in Theorem 7 expresses the same witness-pair logic for comparator classes stated in the published VMW vocabulary [Virk et al., 2026].

Verification note. The Lean-checked scope consists of the matched theorem, proposition, and lemma declarations listed in the verification manifest for commit

`715add8e9475f5193526bce0c26baef559a47b68`.

Definitions marked `presentation-synthesized` in that manifest are expository renderings of paper notation; definitions marked `matched` have the Lean declarations recorded by the manifest. The proof text in this appendix follows the matched declarations and preserves the `% lean:` audit markers that identify certified proof steps.

The Lean toolchain is `leanprover/lean4:v4.33.0`. The CausalSmith package is version `0.1.0`; its manifest records the path dependency `Causalean` at the same repository commit, inherited path dependencies `optlib` at `../third_party/optlib` and `FoML` at `../third_party/lean-rademacher`, and the following external revisions: `mathlib` `db584cd6d46c92f209a44c0f1c829460d327499d`, `plausible` `b7eb3304aeae834b12dda98993a37f6a41f6f0bb`, `LeanSearchClient` `5f4d51b81cbd3f6b32b156bfad9056621a040404`, `importGraph` `16f02aa7642864af59f1ff0e` `proofwidgets` `4be2e3d5087eeb272cf5a8853b8f9dd025ef5957`, `aesop` `3448c0bcc5ce01b2d1546e483ec3620e32` `Qq` `92c15be17b7caf78c2ad767ec40f89052d908d81`, `batteries` `4488d40d070b9700d4d5a6aa342f0d40c31b2a2d` `checkdecls` `3d425859e73fcfbef85b9638c2a91708ef4a22d4`, and `Cli` `6130a47896ce867c6a4a55373441e59e565` The verification build command for the paper module is

```
lake -d CausalSmith build CausalSmith.Stat.STAT_ProxyEffectlawEigencollisionFrontier_Research.
```

Proposition 1 uses the VMW model-scope claim from [Virk et al., 2026, Assumptions 1–2 and Section 4.2]. Theorem 7 uses that model-scope claim together with the VMW separated-recovery scope claim from [Virk et al., 2026, Assumption 3, Assumption 4, and Theorem 7.2]. The remaining matched declarations listed below are derived from their displayed hypotheses and previously matched declarations.

paper object(s)	Lean source file and declaration(s)
Proposition 1	`TObservedVMWMarginInclusion.lean`: `observed_vmw_margin_inclusion`
Proposition 2	`TSummaryClosureCompact.lean`: `summary_closure_compact`
Lemma 1	`Helpers/Concentration.lean`: `uniform_summary_concentration`
Theorem 1	`TGapFreePositiveMeasureModulus.lean`: `gap_free_positive_measure_modulus`
Proposition 3	`TSummaryRepairTotalBorel.lean`: `summary_repair_total_borel`
Proposition 4	`TPolynomialLatticeEstimator.lean`: `polynomial_lattice_law_estimator`
Theorem 2	`TCollisionUniformRootN.lean`: `collision_uniform_root_n`
Theorem 3	`THonestRootNConfidence.lean`: `honest_root_n_confidence`
Theorem 4	`TClusterAdaptiveReport.lean`: `cluster_adaptive_report`
Theorem 5	`TLabeledWeightUpper.lean`: `labeled_weight_upper`
Proposition 5	`TTwoClassWitnessValid.lean`: `two_class_witness_valid`
Theorem 6	`TMatchingLocalLowerBounds.lean`: `matching_local_lower_bounds`
Proposition 6	`TSameClassQuotientMinimax.lean`: `same_class_quotient_minimax`
Proposition 7	`TSameClassLabeledMinimax.lean`: `same_class_labeled_minimax`
Theorem 7	`TPublishedVMWConverseTransfer.lean`: `published_vmw_converse_transfer`
Theorem 8	`TFiniteNetLawEstimator.lean`: `finite_net_law_estimator`
appendix auxiliary lemmas	`Helpers/ObservedLawAdapters.lean`, `Helpers/ObservedMarginAssembly.lean`, `Helpers/ConditionalMomentAdapters.lean`, `Helpers/OutcomeFactorization.lean`, `Helpers/AmbientOperatorBridge.lean`
witness and path lemmas	`Helpers/WitnessValidity.lean`, `Helpers/PathModelCertificates.lean`, `Helpers/PathLocalExperiments.lean`, `Helpers/WitnessKL.lean`
metric, spectral, and ordered-mass lemmas	`Basic.lean`, `Helpers/SummaryMetric.lean`, `Helpers/StructuredLatticeFunctionalCalculus.lean`, `Helpers/ModelSpectralConstruction.lean`, `Helpers/RepresentativeSpectralCertificate.lean`, `Helpers/OrderedMassStability.lean`

Reader-facing sections use one-based singular-value notation. Lean-facing auxiliary displays that expose array indices translate the kth signal singular value into the corresponding zero-based index inside the verification layer.

D Proofs of the main results

Proof of Proposition 1. Fix $P \in \mathcal{M}$ satisfying the displayed hypotheses, and set

$$L_\tau := \frac{4L\sqrt{d_z}}{\sigma_0}.$$

All model restrictions used below are the components of $\mathcal{M}$ in Definition 32.

1. For each treatment arm $t \in \{0,1\}$, Lemma 3 gives

   $$P(T = t) \geq k\pi_0.$$

   Since $k \geq 2$ and $\pi_0 > 0$, this arm probability is positive. Define

   $$\Gamma_t := \operatorname{diag}\left\{\frac{P(U = u, T = t)}{P(T = t)} : 1 \leq u \leq k\right\}.$$

The diagonal entries are

$$\frac{P(U = u, T = t)}{P(T = t)} = P(U = u \mid T = t),$$

so Lemma 4 yields

$$M_t(P) = A_t \Gamma_t B^\top = A_t \operatorname{diag}\{P(U = u \mid T = t) : 1 \le u \le k\} B^\top.$$

2. Latent-arm positivity gives, for every latent class u,

$$\frac{P(U = u, T = t)}{P(T = t)} \ge P(U = u, T = t) \ge \pi_0,$$

because $P(T = t) \le 1$. Hence, for every $\beta \in \mathbb{R}^k$,

$$\|\Gamma_t \beta\|_2^2 = \sum_{u=1}^{k} \Gamma_t(u, u)^2 \beta_u^2 \ge \pi_0^2 \sum_{u=1}^{k} \beta_u^2 = \pi_0^2 \|\beta\|_2^2,$$

and therefore

$$\sigma_k(\Gamma_t) \ge \pi_0.$$

The proxy-rank margin in Assumption 10 gives

$$\sigma_k(A_t) \ge \sigma_0, \qquad \sigma_k(B) \ge \sigma_0.$$

Using the product lower bound for full-column-rank factors in the factorization above,

$$\sigma_k\{M_t(P)\} = \sigma_k(A_t \Gamma_t B^\top) \ge \sigma_k(A_t)\sigma_k(\Gamma_t)\sigma_k(B) \ge \pi_0 \sigma_0^2.$$

3. Let

$$\mathcal{H}(P) := \operatorname{row}(\operatorname{stack}(M_0(P), M_1(P))) \subseteq \mathbb{R}^{d_x},$$

and let $V \in \mathbb{R}^{d_x \times k}$ have orthonormal columns spanning $\mathcal{H}(P)$. For the fixed arm t, the row space of $M_t(P)$ is contained in $\mathcal{H}(P)$:

$$\operatorname{range}\{M_t(P)^\top\} \subseteq \mathcal{H}(P) = \operatorname{range}(V).$$

The lower bound from the previous step gives $\sigma_k\{M_t(P)\} > 0$, so this row space has dimension at least k. Since $\operatorname{range}(V)$ has dimension k, the two spaces are equal. The map $r \mapsto Vr$ is an isometry from $\mathbb{R}^k$ onto this row space, and composing $M_t(P)$ with that isometry preserves its positive singular values. Thus

$$\sigma_k\{M_t(P)V\} = \sigma_k\{M_t(P)\}, \qquad \pi_0 \sigma_0^2 \le \sigma_k\{M_t(P)V\}.$$

4. The stacked proxy signal margin is supplied by Lemma 24. Applied with the present model membership and numerical hypotheses, it gives

$$\pi_0 \sigma_0^2 \le \sigma_k(\operatorname{stack}(M_0(P), M_1(P))),$$

with the displayed one-based singular-value indexing of the statement.

5. By Definition 28,

$$M_t(P) = E_P[ZX^\top \mid T = t], \qquad N_t(P) = E_P[YZX^\top \mid T = t], \qquad m_X(P) = E_P[X].$$

The envelope components of $\mathcal{M}$ give

$$\|ZX^\top\|_{\mathrm{op}} \le L, \qquad \|YZX^\top\|_{\mathrm{op}} \le L, \qquad \|X\|_2 \le L \quad P\text{-a.s.}$$

Since $P(T = t) > 0$, Jensen's inequality for the two conditional matrix means and for the unconditional vector mean yields

$$\|M_t(P)\|_{\mathrm{op}} \le L, \qquad \|N_t(P)\|_{\mathrm{op}} \le L, \qquad \|m_X(P)\|_2 \le L.$$

6. Fix $u \in \{1, \ldots, k\}$ and $t \in \{0, 1\}$, and put

$$\mathsf{Cell}_{u,t} := \{U = u, T = t\}, \qquad \mathsf{Class}_u := \{U = u\}.$$

Latent-arm positivity gives

$$P(\mathsf{Cell}_{u,t}) \ge \pi_0 > 0.$$

Let $e_u^{(k)}$ be the uth canonical vector in $\mathbb{R}^k$. The proxy-rank margin implies

$$\|A_t e_u^{(k)}\|_2 \ge \sigma_k(A_t)\|e_u^{(k)}\|_2 \ge \sigma_0.$$

Since $d_z \ge k \ge 2$, some coordinate $\jmath_{u,t} \in \{1, \ldots, d_z\}$ satisfies

$$\left|E_P[Z_{\jmath_{u,t}} \mid \mathsf{Cell}_{u,t}]\right| = |A_t(\jmath_{u,t}, u)| \ge \frac{\sigma_0}{\sqrt{d_z}}.$$

Define

$$\zeta_{u,t} := \frac{\sigma_0}{2\sqrt{d_z}}, \qquad \mathsf{Large}_{u,t} := \{|Z_{\jmath_{u,t}}| \ge \zeta_{u,t}\}.$$

If $P(\mathsf{Large}_{u,t} \mid \mathsf{Cell}_{u,t}) = 0$, then

$$\left|E_P[Z_{\jmath_{u,t}} \mid \mathsf{Cell}_{u,t}]\right| \le E_P[|Z_{\jmath_{u,t}}| \mid \mathsf{Cell}_{u,t}] \le \zeta_{u,t} < \frac{\sigma_0}{\sqrt{d_z}},$$

a contradiction. Hence

$$P(\mathsf{Large}_{u,t} \mid \mathsf{Cell}_{u,t}) > 0.$$

By Lemma 5,

$$|YZ_{\jmath_{u,t}}| \le L \quad P\text{-a.s. on } \mathsf{Cell}_{u,t}.$$

The reference-proxy separation component in Assumption 1 gives $Z_{\jmath_{u,t}} \perp Y \mid \mathsf{Cell}_{u,t}$. Therefore positive conditional probability for

$$\mathsf{Bad}_{u,t} := \left\{|Y| > \frac{L}{\zeta_{u,t}}\right\}$$

would give

$$P(\mathsf{Large}_{u,t} \cap \mathsf{Bad}_{u,t} \mid \mathsf{Cell}_{u,t}) = P(\mathsf{Large}_{u,t} \mid \mathsf{Cell}_{u,t})P(\mathsf{Bad}_{u,t} \mid \mathsf{Cell}_{u,t}) > 0.$$

On this intersection,
$$L < |Y|\,|Z_{J_{u,t}}| = |YZ_{J_{u,t}}|,$$
contradicting the product envelope. Thus

$$|Y| \le \frac{L}{\zeta_{u,t}} = \frac{2L\sqrt{d_z}}{\sigma_0} = \frac{L_\tau}{2} \quad P\text{-a.s. conditional on } \mathsf{Cell}_{u,t}.$$

On $\{T = t\}$, the consistency component in Assumption 3 gives $Y = Y(t)$ almost surely. Inside the conditional law given Class_u, the event $\{T = t\}$ has positive probability, and the latent-ignorability component in Assumption 4 gives $Y(t) \perp T \mid U$. If

$$\left\{|Y(t)| > \frac{L_\tau}{2}\right\}$$

had positive conditional probability given Class_u, independence would give a positive-probability violation on $\mathsf{Cell}_{u,t}$, contradicting the displayed cell bound. Hence

$$|Y(t)| \le \frac{L_\tau}{2} \quad P\text{-a.s. conditional on } U = u.$$

Finally, Lemmas 6 and 7 give

$$|\mu_{tu}| \le \frac{L_\tau}{2}, \qquad |\tau_u| \le L_\tau.$$

7. It remains to identify membership in the published VMW parameter class. The scope in Definition 52, together with the cited VMW correspondence recorded in Definition 60 and [Virk et al., 2026], identifies $\mathsf{Model}_{\mathrm{VMW}}(k, d_x, d_z, P)$ with the qualitative structural conditions summarized there. Those conditions are supplied as follows. The proxy independences, consistency, and armwise latent ignorability are components of $\mathcal{M}$. The proxy-rank margin and $\sigma_0 > 0$ give full column rank:

$$\sigma_k(A_0) \ge \sigma_0 > 0, \qquad \sigma_k(A_1) \ge \sigma_0 > 0, \qquad \sigma_k(B) \ge \sigma_0 > 0.$$

Latent-arm positivity gives
$$P(U = u, T = t) \ge \pi_0 > 0$$
for every u and t, and hence the corresponding latent treatment probabilities are strictly positive. Together with $2 \le k$, $k \le d_x$, and $k \le d_z$, these are exactly the qualitative conditions for $P \in \mathsf{Model}_{\mathrm{VMW}}(k, d_x, d_z)$.

□

Proof of Proposition 2. 1. First record the coordinate estimate used to pass from operator envelopes to coordinate boxes. For any $H \in \mathbb{R}^{d_z \times d_x}$, any $\iota_z \in \{1, \ldots, d_z\}$, and any $\iota_x \in \{1, \ldots, d_x\}$,
$$|H_{\iota_z \iota_x}| = \|(He_{\iota_x})_{\iota_z}\| \le \|He_{\iota_x}\|_2 \le \|H\|_{\mathrm{op}} \|e_{\iota_x}\|_2 = \|H\|_{\mathrm{op}},$$
where e_{ι_x} is the ι_xth canonical basis vector of $\mathbb{R}^{d_x}$.

2. Define the finite coordinate boxes

$$\mathcal{H}_{\mathrm{mat}}(L) = \Big\{H \in \mathbb{R}^{d_z \times d_x} : -L \le H_{\iota_z \iota_x} \le L \text{ for all } \iota_z, \iota_x\Big\},$$

$$\mathcal{H}_X(L) = \Big\{h \in \mathbb{R}^{d_x} : -L \le h_{\iota_x} \le L \text{ for all } \iota_x\Big\},$$

and

$$\mathcal{H}_S(L) = \mathcal{H}_{\mathrm{mat}}(L)^4 \times \mathcal{H}_X(L).$$

Each factor is a finite product of closed intervals, hence compact, and therefore $\mathcal{H}_S(L)$ is compact. For a five-block summary $s = (M_0, M_1, N_0, N_1, m_X)$, define the coordinate map

$$\Phi(s) = (M_0, M_1, N_0, N_1, m_X).$$

This map is a homeomorphism from the space of five-block summaries onto this finite product coordinate space. Thus

$$\Phi^{-1}\big(\mathcal{H}_S(L)\big)$$

is compact in the summary topology.

3. Let $P \in \mathcal{M}$. For $t \in \{0, 1\}$, put

$$\mathsf{Arm}_t(P) = \{T = t\}.$$

By Assumption 9, together with $2 \le k$ and $0 < \pi_0$, the arm $\mathsf{Arm}_t(P)$ has positive mass. Let P_t^{arm} be the probability law obtained by restricting P to $\mathsf{Arm}_t(P)$ and normalizing by this positive mass. By Definition 28, the armwise blocks in $S(P)$ are conditional means; on this positive arm, equivalently,

$$M_t(P) = \int ZX^\top \, dP_t^{\mathrm{arm}}, \qquad N_t(P) = \int YZX^\top \, dP_t^{\mathrm{arm}},$$

and $m_X(P) = \int X \, dP$. By Assumption 7, the proxy-product envelope restricts to P_t^{arm}: $\|ZX^\top\|_{\mathrm{op}} \le L$ P_t^{arm}-almost surely. Since $1 \le L$, the matrix-valued integral is integrable and

$$\begin{aligned}\|M_t(P)\|_{\mathrm{op}} &= \left\| \int ZX^\top \, dP_t^{\mathrm{arm}} \right\|_{\mathrm{op}} \\ &\le \int \|ZX^\top\|_{\mathrm{op}} \, dP_t^{\mathrm{arm}} \le \int L \, dP_t^{\mathrm{arm}} = L.\end{aligned}$$

The same normalized-restriction argument with Assumption 8 gives

$$\|N_t(P)\|_{\mathrm{op}} \le L.$$

Likewise, Assumption 6 gives

$$\|m_X(P)\|_2 = \left\| \int X \, dP \right\|_2 \le \int \|X\|_2 \, dP \le L.$$

Hence every coordinate of $m_X(P)$ has absolute value at most L. Combining these inequalities with the coordinate estimate from the first step and the summary tuple in Definition 33 gives

$$S(P) \in \Phi^{-1}\big(\mathcal{H}_S(L)\big).$$

Since $P \in \mathcal{M}$ was arbitrary,

$$\mathscr{S} \subseteq \Phi^{-1}\big(\mathcal{H}_S(L)\big).$$

4. The same block envelopes give the stated uniform d_S-bound. For $s = S(P) \in \mathscr{S}$,

$$d_S(s,0) = \|M_0(P)\|_{\mathrm{op}} + \|M_1(P)\|_{\mathrm{op}} + \|N_0(P)\|_{\mathrm{op}} + \|N_1(P)\|_{\mathrm{op}} + \Big(\sum_{\iota_x} m_X(P)_{\iota_x}^2\Big)^{1/2}.$$

For the final term,

$$\Big(\sum_{\iota_x} m_X(P)_{\iota_x}^2\Big)^{1/2} = \Big\|\sum_{\iota_x} m_X(P)_{\iota_x} e_{\iota_x}\Big\|_2 \le \sum_{\iota_x} |m_X(P)_{\iota_x}| \le d_x L.$$

Therefore

$$d_S(S(P),0) \le 4L + d_x L = (4+d_x)L.$$

Taking the scalar bound

$$B_{\mathrm{sum}} = (4+d_x)L$$

and then instantiating the existential bound with B_{sum} proves the boundedness claim.

5. Since $\Phi^{-1}(\mathcal{H}_S(L))$ is compact and closed and contains $\mathscr{S}$, it also contains

$$\mathcal{K} = \overline{\mathscr{S}}.$$

The set $\mathcal{K}$ is closed by definition as a closure, so it is a closed subset of the compact set $\Phi^{-1}(\mathcal{H}_S(L))$. Hence $\mathcal{K}$ is compact.

6. Finally,

$$\mathcal{K} \ne \varnothing \iff \mathscr{S} \ne \varnothing,$$

because a closure has the same nonemptiness as the set it closes. By the definition

$$\mathscr{S} = \{S(P) : P \in \mathcal{M}\},$$

the condition $\mathscr{S} \ne \varnothing$ is equivalent to $\mathcal{M} \ne \varnothing$: a law $P \in \mathcal{M}$ gives $S(P) \in \mathscr{S}$, and any element of $\mathscr{S}$ is $S(P)$ for some $P \in \mathcal{M}$. Thus $\mathcal{K}$ is nonempty if and only if $\mathcal{M}$ is nonempty.

□

Proof of Lemma 1. Fix $k, d_x, d_z, \pi_0, \sigma_0$ satisfying the displayed hypotheses. Let $\mathcal{I}_{d_x,d_z}$ be the finite disjoint union of the following scalar summary coordinates:

$$\{\mathrm{prox}, \mathrm{out}\} \times \{0,1\} \times \{1,\dots,d_z\} \times \{1,\dots,d_x\}, \qquad \mathrm{xcoord} \times \{1,\dots,d_x\}, \qquad \mathrm{arm} \times \{0,1\}.$$

For $O = (T,X,Z,Y) \in O_{d_x,d_z}$ from Definition 18, define

$$\varphi_{\mathrm{prox},t,\mathfrak{z},\mathfrak{x}}(O) := \mathbf{1}\{T=t\} Z_{\mathfrak{z}} X_{\mathfrak{x}}, \qquad \varphi_{\mathrm{out},t,\mathfrak{z},\mathfrak{x}}(O) := \mathbf{1}\{T=t\} Y Z_{\mathfrak{z}} X_{\mathfrak{x}},$$

$$\varphi_{\mathrm{xcoord},\mathfrak{x}}(O) := X_{\mathfrak{x}}, \qquad \varphi_{\mathrm{arm},t}(O) := \mathbf{1}\{T=t\}.$$

Set

$$\varsigma_\iota(L) := \begin{cases} L, & \iota \text{ is a proxy, outcome, or target-proxy coordinate,} \\ 1, & \iota \text{ is an arm coordinate.} \end{cases}$$

The bounded target-proxy, bounded proxy-product, and bounded outcome-proxy components of Definition 32, namely Assumptions 6 to 8, give, for every $\iota \in \mathcal{I}_{d_x,d_z}$,

$$|\varphi_\iota(O)| \le \varsigma_\iota(L) \quad \text{for } P_O\text{-almost every } O,$$

where P_O is the observed-data marginal of Definition 20.

Let $\kappa_{d_z,d_x} \geq 0$ be a finite coordinate-to-operator-norm constant such that, for every $d_z \times d_x$ real matrix H and every $u_{\mathrm{coord}} \geq 0$,

$$\max_{\mathfrak{z},\mathfrak{x}} |H_{\mathfrak{z}\mathfrak{x}}| \leq u_{\mathrm{coord}} \quad \Longrightarrow \quad \|H\|_{\mathrm{op}} \leq \kappa_{d_z,d_x} u_{\mathrm{coord}}.$$

Such a constant exists because, if each coordinate of a vector in $\mathbb{R}^r$ has absolute value at most u_{coord}, then its Euclidean norm is at most $r u_{\mathrm{coord}}$; applying this to each column and then taking the supremum over unit input vectors gives a finite constant depending only on d_z, d_x. We shall also use the corresponding vector estimate

$$\max_{\mathfrak{x}} |v_{\mathfrak{x}}| \leq u_{\mathrm{coord}} \quad \Longrightarrow \quad \|v\|_2 \leq d_x u_{\mathrm{coord}}.$$

Put

$$\chi_{\mathrm{sum}} := |\mathcal{I}_{d_x,d_z}| = 4d_z d_x + d_x + 2,$$

$$K_{\mathrm{sm}} := \frac{16\kappa_{d_z,d_x}}{k\pi_0} + d_x, \qquad K_{\mathrm{lg}} := 4(\kappa_{d_z,d_x} + 1) + d_x + 1,$$

and choose

$$C_0 := \max\left\{\chi_{\mathrm{sum}} + 1,\ 2K_{\mathrm{sm}},\ \frac{4K_{\mathrm{lg}}}{k\pi_0}\right\}.$$

Then $1 \leq C_0$. Since $2 \leq k$ and $0 < \pi_0$, the product $k\pi_0$ is positive, so all denominators in this definition are positive.

Now fix $L \geq 1$, $n \geq 1$, $P \in \mathcal{M}$, and $0 < \eta < 1/2$. Define

$$\zeta_{\eta,n} := \sqrt{\frac{\log(C_0/\eta)}{n}}, \qquad \beta_{\eta,n} := 2\zeta_{\eta,n}.$$

For each $\iota \in \mathcal{I}_{d_x,d_z}$, Hoeffding's inequality under $Q_P^{(n)} = P_O^{\otimes n}$ gives

$$Q_P^{(n)}\left\{\varsigma_\iota(L)\beta_{\eta,n} \leq \left|\frac{1}{n}\sum_{\mathfrak{r}=1}^{n} \varphi_\iota(O_{\mathfrak{r}}) - \int \varphi_\iota\, dP_O\right|\right\} \leq 2\left(\frac{\eta}{C_0}\right)^2.$$

Indeed, $\varphi_\iota \in [-\varsigma_\iota(L), \varsigma_\iota(L)]$ P_O-almost surely, the interval length is $2\varsigma_\iota(L)$, and the choice $\beta_{\eta,n} = 2\sqrt{\log(C_0/\eta)/n}$ reduces the exponential factor to $\exp\{-2\log(C_0/\eta)\}$.

Let

$$\mathcal{B}_{\mathrm{dev}} := \left\{\omega : \exists \iota \in \mathcal{I}_{d_x,d_z},\ \varsigma_\iota(L)\beta_{\eta,n} \leq \left|\frac{1}{n}\sum_{\mathfrak{r}=1}^{n} \varphi_\iota(O_{\mathfrak{r}}) - \int \varphi_\iota\, dP_O\right|\right\}.$$

By the union bound and $\chi_{\mathrm{sum}} \leq C_0$,

$$Q_P^{(n)}(\mathcal{B}_{\mathrm{dev}}) \leq \chi_{\mathrm{sum}}\, 2\left(\frac{\eta}{C_0}\right)^2 \leq \frac{2\eta^2}{C_0} \leq 2\eta^2 \leq \eta.$$

Also set

$$\mathcal{B}_{\mathrm{supp}} := \{\omega : \exists \mathfrak{r}, \exists \iota \in \mathcal{I}_{d_x,d_z},\ |\varphi_\iota(O_{\mathfrak{r}})| > \varsigma_\iota(L)\}.$$

The P_O-almost-sure coordinate envelope transported to the product experiment gives

$$Q_P^{(n)}(\mathcal{B}_{\mathrm{supp}}) = 0.$$

The population summary blocks obey

$$\|M_t(P)\|_{\mathrm{op}} \le L, \qquad \|N_t(P)\|_{\mathrm{op}} \le L, \qquad \|m_X(P)\|_2 \le L.$$

For the armwise blocks, Lemma 3 gives positive arm probabilities, and Jensen's inequality applied to the conditional moments in Definition 28 transfers the almost-sure bounds on $ZX^\top$ and $YZX^\top$. For $m_X(P)$, Jensen's inequality applies directly to the unconditional target-proxy mean.

It remains to prove a deterministic implication off the two bad events. Write

$$\theta_t := P_O\{O : T = t\}, \qquad \widehat{\theta}_t(\omega) := \frac{N_{t,n}(\omega)}{n}.$$

By Lemma 3, transported through Definition 20,

$$k\pi_0 \le \theta_t \qquad (t \in \{0, 1\}).$$

First suppose that

$$\beta_{\eta,n} < \frac{k\pi_0}{2}.$$

On $\mathcal{B}_{\mathrm{dev}}^c$,

$$|\widehat{\theta}_t - \theta_t| < \beta_{\eta,n}, \qquad \widehat{\theta}_t > \frac{k\pi_0}{2}.$$

For $\bullet \in \{\mathrm{prox}, \mathrm{out}\}$, put

$$\overline{\Phi}^{\bullet}_{t,\mathfrak{z},\mathfrak{x}}(\omega) := \frac{1}{n}\sum_{\mathfrak{r}=1}^{n} \varphi_{\bullet,t,\mathfrak{z},\mathfrak{x}}(O_{\mathfrak{r}}).$$

The coordinate integral identities are

$$\int \varphi_{\mathrm{prox},t,\mathfrak{z},\mathfrak{x}}\, dP_O = \theta_t\, M_t(P)_{\mathfrak{z}\mathfrak{x}}, \qquad \int \varphi_{\mathrm{out},t,\mathfrak{z},\mathfrak{x}}\, dP_O = \theta_t\, N_t(P)_{\mathfrak{z}\mathfrak{x}}.$$

Because $\widehat{\theta}_t > 0$, the total empty-arm convention in Definition 34 gives

$$(\widehat{M}_{t,n} - M_t(P))_{\mathfrak{z}\mathfrak{x}} = \frac{\{\overline{\Phi}^{\mathrm{prox}}_{t,\mathfrak{z},\mathfrak{x}} - \theta_t M_t(P)_{\mathfrak{z}\mathfrak{x}}\} + (\theta_t - \widehat{\theta}_t) M_t(P)_{\mathfrak{z}\mathfrak{x}}}{\widehat{\theta}_t},$$

and the same identity with $\widehat{N}_{t,n}, N_t(P)$ and $\overline{\Phi}^{\mathrm{out}}$. Since each population entry is bounded by L, the preceding displays imply, for both matrix types,

$$\left|(\widehat{M}_{t,n} - M_t(P))_{\mathfrak{z}\mathfrak{x}}\right| \le \frac{4L\beta_{\eta,n}}{k\pi_0}, \qquad \left|(\widehat{N}_{t,n} - N_t(P))_{\mathfrak{z}\mathfrak{x}}\right| \le \frac{4L\beta_{\eta,n}}{k\pi_0}.$$

The matrix norm conversion and the mean-coordinate bounds

$$|(\widehat{m}_{X,n} - m_X(P))_{\mathfrak{x}}| < L\beta_{\eta,n}$$

therefore yield

$$d_S(\widehat{S}_n, S(P)) \le \left(\frac{16\kappa_{d_z,d_x}}{k\pi_0} + d_x\right) L\beta_{\eta,n} = 2K_{\mathrm{sm}} L\zeta_{\eta,n} \le C_0 L\zeta_{\eta,n}.$$

Now suppose that

$$\frac{k\pi_0}{2} \le \beta_{\eta,n}.$$

On $\mathcal{B}_{\text{supp}}^c$, every empirical arm matrix has entries bounded by L: if $N_{t,n} = 0$, the convention in Definition 34 gives the zero matrix, and if $N_{t,n} > 0$, the matrix is an average of entries whose absolute values are at most L. Hence each empirical matrix block has operator norm at most $\kappa_{d_z,d_x} L$, and

$$\|\widehat{m}_{X,n}\|_2 \le d_x L.$$

Combining these empirical bounds with the population envelopes gives

$$d_S(\widehat{S}_n, S(P)) \le \{4(\kappa_{d_z,d_x} + 1) + d_x + 1\}L = K_{\text{lg}} L.$$

Since $k\pi_0/2 \le \beta_{\eta,n} = 2\zeta_{\eta,n}$,

$$\frac{k\pi_0}{4} \le \zeta_{\eta,n},$$

and therefore

$$K_{\text{lg}} L \le \frac{4K_{\text{lg}}}{k\pi_0} L\zeta_{\eta,n} \le C_0 L \zeta_{\eta,n}.$$

The two cases imply

$$\left\{ d_S(\widehat{S}_n, S(P)) > C_0 L \sqrt{\frac{\log(C_0/\eta)}{n}} \right\} \subseteq \mathcal{B}_{\text{dev}} \cup \mathcal{B}_{\text{supp}}.$$

Consequently,

$$Q_P^{(n)} \left\{ d_S(\widehat{S}_n, S(P)) > C_0 L \sqrt{\frac{\log(C_0/\eta)}{n}} \right\} \le Q_P^{(n)}(\mathcal{B}_{\text{dev}}) + Q_P^{(n)}(\mathcal{B}_{\text{supp}}) \le \eta.$$

Finally fix $t \in \{0, 1\}$. The event $N_{t,n} = 0$ is the event that all n observed records have $T \ne t$. Since $Q_P^{(n)} = P_O^{\otimes n}$,

$$Q_P^{(n)}\{N_{t,n} = 0\} = \big(P_O\{O : T \ne t\}\big)^n = \big(1 - P_O\{O : T = t\}\big)^n.$$

The lower bound from Lemma 3, again through Definition 20, gives

$$k\pi_0 \le P_O\{O : T = t\} \le 1.$$

Thus

$$0 \le 1 - P_O\{O : T = t\} \le 1 - k\pi_0,$$

and monotonicity of $x \mapsto x^n$ on $[0, \infty)$ gives

$$\big(1 - P_O\{O : T = t\}\big)^n \le (1 - k\pi_0)^n.$$

This proves both asserted conclusions. $\square$

Proof of Theorem 1. 1. Put

$$s_\star := \pi_0 \sigma_0^2, \qquad L_\tau := \frac{4L\sqrt{d_z}}{\sigma_0}.$$

For matrix sizes $r_{\text{row}}, r_{\text{col}}$, write

$$\text{Ent}_{r_{\text{row}}, r_{\text{col}}} := r_{\text{col}} r_{\text{row}} \, \|J_{r_{\text{col}}}\|_{\text{op}},$$

with $J_{r_{\text{col}}}$ as in Lemma 22. Define

$$\begin{aligned} K_{\text{forw}} &:= \text{Ent}_{d_x,k} \, \text{Ent}_{k,k} \, \text{Ent}_{k,d_x} \; \sigma_0^{-1} + \text{Ent}_{d_x,d_x} + \text{Ent}_{d_x,k} \, \text{Ent}_{k,d_x}, \\ K_{\text{back}} &:= \text{Ent}_{d_x,k} \, \text{Ent}_{k,k} \, \text{Ent}_{k,d_x} \, (d_x L) + \text{Ent}_{d_x,d_x} + \text{Ent}_{d_x,k} \, \text{Ent}_{k,d_x}, \end{aligned}$$

and

$$\kappa_\star := \text{Ent}_{d_x,d_x}^2 \, K_{\text{forw}} K_{\text{back}}.$$

The assumptions $L \geq 1$ and $\sigma_0 > 0$, together with nonnegativity of the entry-norm constants, give

$$0 \leq \kappa_\star.$$

Set

$$C_{\text{base}} := \kappa_\star L_\tau + L \, d_x^2 \kappa_\star^2 \{3 s_\star^{-2} L + s_\star^{-1}\}, \qquad C_{\text{mod}} := C_{\text{base}} + 1.$$

Since $s_\star > 0$, $L_\tau \geq 0$, and $\kappa_\star \geq 0$, we have

$$0 \leq C_{\text{base}}, \qquad 0 < C_{\text{mod}}.$$

2. Fix $P \in \mathcal{M}$. Let $B(P)$ be the target-proxy feature matrix of Definition 25. The proxy-rank margin gives

$$\sigma_0 \leq \sigma_k(B(P)).$$

Choose a thin singular-value factorization

$$B(P) = V_P \Upsilon_P, \qquad V_P^\top V_P = I_k, \qquad \Upsilon_P^{-1} \Upsilon_P = \Upsilon_P \Upsilon_P^{-1} = I_k.$$

The singular-value lower bound implies

$$|(\Upsilon_P^{-1})_{ab}| \leq \sigma_0^{-1} \qquad (1 \leq a, b \leq k).$$

For every coordinate i and latent class u, latent-arm positivity gives

$$0 < P(U = u, T = 0) \leq P(U = u).$$

Under the normalized restriction of P to $\{U = u\}$, Lemma 5 gives

$$|X_i| \leq L \quad \text{almost surely.}$$

Therefore

$$\begin{aligned} |B(P)_{iu}| &= |E_P[X_i \mid U = u]| \\ &= \left| \int X_i \, dP(\cdot \mid U = u) \right| \\ &\leq \int |X_i| \, dP(\cdot \mid U = u) \leq L. \end{aligned}$$

Thus Lemma 26 applies to $B(P) = V_P \Upsilon_P$.

Define the ambient contrast

$$\mathfrak{D}_P := M_1(P)^\dagger N_1(P) - M_0(P)^\dagger N_0(P).$$

By Lemmas 4 and 18, for each $t \in \{0, 1\}$,

$$M_t(P) = A_t(P)\Gamma_t(P)B(P)^\top, \qquad N_t(P) = A_t(P)\Gamma_t(P)\operatorname{diag}\{\mu_{tu}(P)\}_{u=1}^k B(P)^\top,$$

where

$$\Gamma_t(P) = \operatorname{diag}\left\{\frac{P(U=u, T=t)}{P(T=t)} : 1 \le u \le k\right\}.$$

The proxy-rank margin and Lemma 19 make $A_t(P)\Gamma_t(P)$ injective, and the proxy-rank margin makes $B(P)$ injective. Hence Lemma 20 yields

$$M_t(P)^\dagger N_t(P) = (B(P)^\dagger)^\top \operatorname{diag}\{\mu_{tu}(P)\}_{u=1}^k B(P)^\top.$$

Subtracting the two arms gives

$$\mathfrak{D}_P = (B(P)^\dagger)^\top \operatorname{diag}\{\tau_u(P)\}_{u=1}^k B(P)^\top.$$

Let

$$\mathsf{Forw}_P := V_P(\Upsilon_P^{-1})^\top V_P^\top + (I_{d_x} - V_P V_P^\top), \qquad \mathsf{Back}_P := V_P \Upsilon_P^\top V_P^\top + (I_{d_x} - V_P V_P^\top).$$

Then

$$\mathsf{Forw}_P \mathsf{Back}_P = \mathsf{Back}_P \mathsf{Forw}_P = I_{d_x}.$$

Extending the columns of V_P to an orthonormal basis of $\mathbb{R}^{d_x}$, the columns obtained by applying Forw_P form a real eigenbasis for $\mathfrak{D}_P$. The signal eigenvalues are $\tau_u(P)$, and the remaining eigenvalues are 0. Consequently

$$\mathfrak{D}_P = \mathcal{S}_P \operatorname{diag}(\alpha_{P,1}, \dots, \alpha_{P,d_x})\mathcal{S}_P^{-1}, \tag{1}$$

with

$$\|\mathcal{S}_P\|_{\mathrm{op}} \|\mathcal{S}_P^{-1}\|_{\mathrm{op}} \le \kappa_\star, \qquad |\alpha_{P,j}| \le L_\tau.$$

The eigenvalue bound uses Lemma 7 on the signal coordinates and $0 \in [-L_\tau, L_\tau]$ on the remaining coordinates.

It remains to record the two anchor identities used by the spectral representation. For each coordinate j,

$$\begin{aligned} m_X(P)_j &= \int X_j \, dP \\ &= \sum_{u=1}^k \int_{\{U=u\}} X_j \, dP \\ &= \sum_{u=1}^k E_P[X_j \mid U = u] \, P(U = u) \\ &= \sum_{u=1}^k B(P)_{ju} p_u(P), \end{aligned}$$

where the latent classes form a finite measurable partition and X_j is integrable by Lemma 5. Hence

$$m_X(P) = B(P)p(P).$$

Also, since $X_1 = 1$ almost surely and $P(U = u) > 0$,

$$B(P)_{1u} = E_P[X_1 \mid U = u] = 1 \qquad (1 \le u \le k),$$

so the first canonical basis vector $e_1 \in \mathbb{R}^{d_x}$ satisfies

$$B(P)^\top e_1 = \mathbf{1}_k.$$

Therefore, for every one-Lipschitz $\varphi : \mathbb{R} \to \mathbb{R}$ satisfying $\varphi(0) = 0$,

$$\varphi(\mathfrak{D}_P) = (B(P)^\dagger)^\top \operatorname{diag}\{\varphi(\tau_u(P))\}_{u=1}^k B(P)^\top$$

and

$$\int \varphi \, d\nu_P = \sum_{u=1}^k p_u(P)\varphi(\tau_u(P)) = \langle m_X(P), \varphi(\mathfrak{D}_P)e_1 \rangle .$$

3. Fix $P, Q \in \mathcal{M}$. For each arm t, Proposition 1 supplies

$$s_\star \le \sigma_k(M_t(P)), \qquad s_\star \le \sigma_k(M_t(Q)), \qquad \|N_t(P)\|_{\mathrm{op}} \le L.$$

Together with the proxy factorization, these inequalities give

$$\operatorname{rank} M_t(P) = \operatorname{rank} M_t(Q) = k.$$

We first record the armwise perturbation estimate. Let $\mathsf{H}, \mathsf{H}' \in \mathbb{R}^{d_z \times d_x}$ have the same rank k, and suppose

$$s_\star \le \sigma_k(\mathsf{H}), \qquad s_\star \le \sigma_k(\mathsf{H}').$$

Then

$$\|\mathsf{H}^\dagger\|_{\mathrm{op}} \le s_\star^{-1}, \qquad \|\mathsf{H}'^\dagger\|_{\mathrm{op}} \le s_\star^{-1},$$

because the operator norm of the Moore–Penrose inverse is the reciprocal of the smallest nonzero singular value. The Penrose equations, applied with the common rank condition, give the exact moving-space identity

$$\begin{aligned}\mathsf{H}^\dagger - \mathsf{H}'^\dagger = &-\mathsf{H}^\dagger(\mathsf{H} - \mathsf{H}')\mathsf{H}'^\dagger + \mathsf{H}^\dagger(\mathsf{H}^\dagger)^\top(\mathsf{H} - \mathsf{H}')^\top(I - \mathsf{H}'\mathsf{H}'^\dagger) \\ &+ (I - \mathsf{H}^\dagger\mathsf{H})(\mathsf{H} - \mathsf{H}')^\top(\mathsf{H}'^\dagger)^\top\mathsf{H}'^\dagger.\end{aligned}$$

Here $I - \mathsf{H}'\mathsf{H}'^\dagger$ and $I - \mathsf{H}^\dagger\mathsf{H}$ are orthogonal projection residuals, so their operator norms are at most one. Taking norms term by term gives

$$\begin{aligned}\|\mathsf{H}^\dagger(\mathsf{H} - \mathsf{H}')\mathsf{H}'^\dagger\|_{\mathrm{op}} &\le s_\star^{-2}\|\mathsf{H} - \mathsf{H}'\|_{\mathrm{op}}, \\ \|\mathsf{H}^\dagger(\mathsf{H}^\dagger)^\top(\mathsf{H} - \mathsf{H}')^\top(I - \mathsf{H}'\mathsf{H}'^\dagger)\|_{\mathrm{op}} &\le s_\star^{-2}\|\mathsf{H} - \mathsf{H}'\|_{\mathrm{op}}, \\ \|(I - \mathsf{H}^\dagger\mathsf{H})(\mathsf{H} - \mathsf{H}')^\top(\mathsf{H}'^\dagger)^\top\mathsf{H}'^\dagger\|_{\mathrm{op}} &\le s_\star^{-2}\|\mathsf{H} - \mathsf{H}'\|_{\mathrm{op}}.\end{aligned}$$

Hence

$$\|\mathsf{H}^\dagger - \mathsf{H}'^\dagger\|_{\mathrm{op}} \le 3s_\star^{-2}\|\mathsf{H} - \mathsf{H}'\|_{\mathrm{op}}.$$

For matrices $\mathsf{N}, \mathsf{N}' \in \mathbb{R}^{d_z \times d_x}$, the algebraic split

$$\mathsf{H}^\dagger \mathsf{N} - \mathsf{H}'^\dagger \mathsf{N}' = (\mathsf{H}^\dagger - \mathsf{H}'^\dagger)\mathsf{N} + \mathsf{H}'^\dagger(\mathsf{N} - \mathsf{N}')$$

therefore implies, whenever $\|\mathsf{N}\|_{\mathrm{op}} \le L$,

$$\|\mathsf{H}^\dagger \mathsf{N} - \mathsf{H}'^\dagger \mathsf{N}'\|_{\mathrm{op}} \le 3s_\star^{-2} L \,\|\mathsf{H} - \mathsf{H}'\|_{\mathrm{op}} + s_\star^{-1}\|\mathsf{N} - \mathsf{N}'\|_{\mathrm{op}}.$$

Applying this estimate with

$$\mathsf{H} = M_t(P), \quad \mathsf{H}' = M_t(Q), \quad \mathsf{N} = N_t(P), \quad \mathsf{N}' = N_t(Q)$$

gives

$$\begin{aligned} &\|M_t(P)^\dagger N_t(P) - M_t(Q)^\dagger N_t(Q)\|_{\mathrm{op}} \\ &\qquad \le 3s_\star^{-2} L \,\|M_t(P) - M_t(Q)\|_{\mathrm{op}} + s_\star^{-1}\|N_t(P) - N_t(Q)\|_{\mathrm{op}}. \end{aligned}$$

Summing over $t = 0, 1$ and using the definition of d_S yields

$$\|\mathfrak{D}_P - \mathfrak{D}_Q\|_{\mathrm{op}} \le \{3s_\star^{-2} L + s_\star^{-1}\}\, d_S(S(P), S(Q)).$$

Moreover,

$$\|m_X(P) - m_X(Q)\|_2 \le d_S(S(P), S(Q)), \qquad \|m_X(Q)\|_2 \le L, \qquad \|e_1\|_2 = 1.$$

4. Let φ be the normalized attaining Kantorovich–Rubinstein potential for (ν_P, ν_Q) from Lemma 21. Then

$$\varphi(0) = 0, \qquad \mathrm{Lip}(\varphi) \le 1, \qquad W_1(\nu_P, \nu_Q) = \left| \int \varphi \, d\nu_P - \int \varphi \, d\nu_Q \right|.$$

Use the diagonalizations in Equation (1) for P and Q, written

$$\mathfrak{D}_P = \mathcal{S}_P \operatorname{diag}(\alpha_{P,1}, \ldots, \alpha_{P,d_x}) \mathcal{S}_P^{-1}, \qquad \mathfrak{D}_Q = \mathcal{S}_Q \operatorname{diag}(\alpha_{Q,1}, \ldots, \alpha_{Q,d_x}) \mathcal{S}_Q^{-1}.$$

Let

$$\Omega_P := \{\alpha_{P,j} : 1 \le j \le d_x\}, \qquad \Omega_Q := \{\alpha_{Q,j} : 1 \le j \le d_x\}.$$

For any real number ζ, define the aggregate projectors

$$\mathsf{Proj}_P(\zeta) := \mathcal{S}_P \operatorname{diag}\{\mathbf{1}(\alpha_{P,j} = \zeta) : 1 \le j \le d_x\} \mathcal{S}_P^{-1},$$

and

$$\mathsf{Proj}_Q(\zeta) := \mathcal{S}_Q \operatorname{diag}\{\mathbf{1}(\alpha_{Q,j} = \zeta) : 1 \le j \le d_x\} \mathcal{S}_Q^{-1}.$$

The diagonal 0-1 matrices have operator norm at most one, so for $\zeta_P \in \Omega_P$ and $\zeta_Q \in \Omega_Q$,

$$\|\mathsf{Proj}_P(\zeta_P)\|_{\mathrm{op}} \le \kappa_\star, \qquad \|\mathsf{Proj}_Q(\zeta_Q)\|_{\mathrm{op}} \le \kappa_\star.$$

The projectors extract their spectral values on the appropriate side:

$$\mathsf{Proj}_P(\zeta_P)\mathfrak{D}_P = \zeta_P \mathsf{Proj}_P(\zeta_P), \qquad \mathfrak{D}_Q \mathsf{Proj}_Q(\zeta_Q) = \zeta_Q \mathsf{Proj}_Q(\zeta_Q).$$

Hence

$$\begin{aligned}&\mathsf{Proj}_P(\zeta_P)(\mathfrak{D}_P - \mathfrak{D}_Q)\mathsf{Proj}_Q(\zeta_Q)\\&\qquad= (\zeta_P - \zeta_Q)\mathsf{Proj}_P(\zeta_P)\mathsf{Proj}_Q(\zeta_Q).\end{aligned}$$

In particular, when $\zeta_P = \zeta_Q$,

$$\mathsf{Proj}_P(\zeta_P)(\mathfrak{D}_P - \mathfrak{D}_Q)\mathsf{Proj}_Q(\zeta_Q) = 0.$$

Functional calculus with the two diagonalizations gives

$$\varphi(\mathfrak{D}_P) = \sum_{\zeta_P\in\Omega_P} \varphi(\zeta_P)\mathsf{Proj}_P(\zeta_P), \qquad \varphi(\mathfrak{D}_Q) = \sum_{\zeta_Q\in\Omega_Q} \varphi(\zeta_Q)\mathsf{Proj}_Q(\zeta_Q).$$

Since the aggregate projectors sum to the identity for each diagonalization, subtraction and expansion over $\Omega_P \times \Omega_Q$ yield

$$\varphi(\mathfrak{D}_P) - \varphi(\mathfrak{D}_Q) = \sum_{\zeta_P\in\Omega_P}\sum_{\zeta_Q\in\Omega_Q} \Xi(\zeta_P, \zeta_Q),$$

where

$$\Xi(\zeta_P, \zeta_Q) := \begin{cases} 0, & \zeta_P = \zeta_Q,\\ \dfrac{\varphi(\zeta_P) - \varphi(\zeta_Q)}{\zeta_P - \zeta_Q}\,\mathsf{Proj}_P(\zeta_P)(\mathfrak{D}_P - \mathfrak{D}_Q)\mathsf{Proj}_Q(\zeta_Q), & \zeta_P \neq \zeta_Q.\end{cases}$$

The equal-spectral-value branch is zero by the preceding vanishing identity. In the unequal branch, the one-Lipschitz property gives

$$\left|\frac{\varphi(\zeta_P) - \varphi(\zeta_Q)}{\zeta_P - \zeta_Q}\right| \le 1,$$

and therefore

$$\|\Xi(\zeta_P, \zeta_Q)\|_{\mathrm{op}} \le \kappa_\star^2 \|\mathfrak{D}_P - \mathfrak{D}_Q\|_{\mathrm{op}}.$$

Because $|\Omega_P| \le d_x$ and $|\Omega_Q| \le d_x$, the triangle inequality gives

$$\|\varphi(\mathfrak{D}_P) - \varphi(\mathfrak{D}_Q)\|_{\mathrm{op}} \le d_x^2\kappa_\star^2 \|\mathfrak{D}_P - \mathfrak{D}_Q\|_{\mathrm{op}}.$$

Also,

$$\|\varphi(\mathfrak{D}_P)\|_{\mathrm{op}} \le \kappa_\star L_\tau,$$

because $\varphi(0) = 0$, φ is one-Lipschitz, and every eigenvalue of $\mathfrak{D}_P$ lies in $[-L_\tau, L_\tau]$.

Using the representation identity for P and Q,

$$\begin{aligned}W_1(\nu_P, \nu_Q) &\le \|m_X(P) - m_X(Q)\|_2\, \|\varphi(\mathfrak{D}_P)\|_{\mathrm{op}}\|e_1\|_2\\&\quad + \|m_X(Q)\|_2\, \|\varphi(\mathfrak{D}_P) - \varphi(\mathfrak{D}_Q)\|_{\mathrm{op}}\|e_1\|_2\\&\le C_{\mathrm{base}}\, d_S(S(P), S(Q)) \le C_{\mathrm{mod}}\, d_S(S(P), S(Q)).\end{aligned}$$

This proves the asserted modulus on model summaries:

$$W_1(\nu_P, \nu_Q) \le C_{\mathrm{mod}} d_S(S(P), S(Q)). \tag{2}$$

5. Let
$$\mathscr{S} := \{S(P) : P \in \mathcal{M}\}.$$
For $q \in \mathscr{S}$, choose a representative law $P_q \in \mathcal{M}$ with
$$S(P_q) = q,$$
and define
$$F_{\mathscr{S}}(q) := \nu_{P_q}.$$
This definition is independent of the representative. Indeed, if another $Q \in \mathcal{M}$ has $S(Q) = q$, then the modulus just proved gives
$$0 \leq W_1(\nu_{P_q}, \nu_Q) \leq C_{\mathrm{mod}}\, d_S(q, q) = 0.$$
A zero finite transport cost between quotient atomic laws forces the same aggregated mass at each support point, so the two quotient laws are equal. Hence
$$F_{\mathscr{S}}(S(Q)) = \nu_Q \qquad (Q \in \mathcal{M}),$$
and, for all $q, q' \in \mathscr{S}$,
$$W_1(F_{\mathscr{S}}(q), F_{\mathscr{S}}(q')) \leq C_{\mathrm{mod}}\, d_S(q, q').$$

6. Let ρ_{sum} be the ambient product metric on the five summary blocks: each matrix block has its Frobenius norm, the mean block has its Euclidean norm, and the finite product uses the maximum of the five block distances. Set
$$\Lambda_{\mathrm{sum}} := 4\,\mathrm{Ent}_{d_z,d_x} + d_x.$$
By Lemma 22,
$$d_S(q, q') \leq \Lambda_{\mathrm{sum}} \rho_{\mathrm{sum}}(q, q') \qquad \text{for all summaries } q, q'.$$
For a matrix block $A \in \mathbb{R}^{d_z \times d_x}$, each entry obeys
$$|A_{ab}| \leq \|A\|_{\mathrm{op}},$$
so summing the squared entries gives
$$\|A\|_{\mathrm{F}} \leq \sqrt{d_z d_x}\, \|A\|_{\mathrm{op}}.$$
Since ρ_{sum} is the maximum of the four Frobenius block distances and the Euclidean mean-block distance, while d_S is the sum of the four operator-norm distances and that same mean-block distance,
$$\rho_{\mathrm{sum}}(q, q') \leq \max\{1, \sqrt{d_z d_x}\}\, d_S(q, q') \qquad \text{for all summaries } q, q'.$$
(The assumptions $2 \leq k \leq d_x, d_z$ cover the positive-dimensional case used here; in a zero-dimensional matrix block the unique empty matrix contributes zero.) Thus d_S and ρ_{sum} induce the same topology and the same closures.

The first comparison and the d_S-control on $F_{\mathscr{S}}$ give
$$W_1(F_{\mathscr{S}}(q), F_{\mathscr{S}}(q')) \leq C_{\mathrm{mod}} \Lambda_{\mathrm{sum}} \rho_{\mathrm{sum}}(q, q') \qquad (q, q' \in \mathscr{S}).$$

Hence $F_{\mathscr{S}}$ is Lipschitz on the summary image in the ambient metric.

Inside $\mathcal{K} = \overline{\mathscr{S}}$, put

$$\mathscr{S}_{\mathcal{K}} := \{q \in \mathcal{K} : q \in \mathscr{S}\}.$$

The equality of closures just proved makes $\mathscr{S}_{\mathcal{K}}$ dense in $\mathcal{K}$. The quotient atomic-law space $\mathcal{P}_k([-L_\tau, L_\tau])$ is complete for W_1: in labelled coordinates it is the metric quotient of the compact set

$$\Delta_{k-1} \times [-L_\tau, L_\tau]^k.$$

Apply the dense-set extension theorem to the ambient-Lipschitz map

$$\widetilde{F}_{\mathscr{S}}(q) := F_{\mathscr{S}}(q), \qquad q \in \mathscr{S}_{\mathcal{K}}.$$

The dense-set extension construction gives a Lipschitz, hence continuous, map

$$\overline{F} : \mathcal{K} \to \mathcal{P}_k([-L_\tau, L_\tau])$$

that agrees with $F_{\mathscr{S}}$ on $\mathscr{S}$, and the uniqueness property of this construction makes it the unique continuous extension with that agreement.

It remains to transfer the sharper d_S-control. Fix $q, q' \in \mathcal{K}$. Choose sequences $q_m, q'_m \in \mathscr{S}$ converging in the ambient metric to q, q', respectively. Continuity of $\overline{F}$ and Lemma 23 give

$$W_1(\overline{F}(q_m), \overline{F}(q'_m)) \longrightarrow W_1(\overline{F}(q), \overline{F}(q')), \qquad d_S(q_m, q'_m) \longrightarrow d_S(q, q').$$

For every m, agreement on $\mathscr{S}$ and the model-summary control in Equation (2) yield

$$W_1(\overline{F}(q_m), \overline{F}(q'_m)) = W_1(F_{\mathscr{S}}(q_m), F_{\mathscr{S}}(q'_m)) \leq C_{\mathrm{mod}} d_S(q_m, q'_m).$$

Passing to the limit proves

$$W_1(\overline{F}(q), \overline{F}(q')) \leq C_{\mathrm{mod}} d_S(q, q') \qquad (q, q' \in \mathcal{K}).$$

Continuity of $\overline{F}$, together with the Borel structure on the quotient atomic-law space, makes $\overline{F}$ Borel measurable.

7. For every $Q \in \mathcal{M}$, the summary $S(Q)$ belongs to $\mathscr{S}$, and hence

$$\overline{F}(S(Q)) = F_{\mathscr{S}}(S(Q)) = \nu_Q.$$

Let

$$G : \mathcal{K} \to \mathcal{P}_k([-L_\tau, L_\tau])$$

be continuous, satisfy the stated C_{mod}-Lipschitz bound, and agree with $Q \mapsto \nu_Q$ on all model summaries. For any $q \in \mathcal{K}$, choose $q_m \in \mathscr{S}$ with

$$q_m \to q.$$

Continuity and agreement on $\mathscr{S}$ imply

$$G(q) = \lim_m G(q_m) = \lim_m F_{\mathscr{S}}(q_m) = \overline{F}(q).$$

Thus $G(q) = \overline{F}(q)$ for all $q \in \mathcal{K}$. Combining the model-summary modulus, the continuous extension, the Borel measurability, and this uniqueness proves the theorem.

□

Proof of Proposition 3. 1. Write

$$\mathcal{K} = \overline{\mathscr{S}}^{d_S} = \mathrm{cl}_{d_S}\{S(P) : P \in \mathcal{M}\}.$$

By Proposition 2, $\mathcal{K}$ is compact. Also $2 \le k$ gives $0 < k$, and Definition 26 with $L \ge 1$ and $0 < \sigma_0 \le 1$ gives $L_\tau \ge 0$. These two inequalities supply the positivity data needed for the atomic-law fallback δ_0. From Theorem 1 we take a continuous, hence Borel, extension

$$\overline{F} : \mathcal{K} \longrightarrow \mathcal{P}_k([-L_\tau, L_\tau])$$

satisfying, for every admissible model law P,

$$\overline{F}(S(P)) = \nu_P.$$

2. First suppose $\mathcal{K} = \varnothing$. Define the selector on the full five-block summary space by

$$\Pi(s) = 0 \qquad \text{for every summary } s.$$

This map is Borel. The extension condition over model summaries is vacuous, because any admissible P would give

$$S(P) \in \mathscr{S} \subseteq \mathcal{K},$$

contradicting $\mathcal{K} = \varnothing$. The nonempty-$\mathcal{K}$ nearest-point clause in Algorithm 1 likewise has a false premise. Thus these choices define admissible repair data R. For every observed sample $\mathsf{o}_{1:n} \in O^n_{d_x,d_z}$, Algorithm 1 gives

$$\widehat{\nu}_n(\mathsf{o}_{1:n}) = \delta_0,$$

so $\widehat{\nu}_n$ is the constant Borel map. In the same case, Algorithm 3 gives the retained set

$$\{\delta_0\},$$

which is nonempty.

3. Now suppose $\mathcal{K} \neq \varnothing$. Let

$$d_{\mathrm{sum}} = 4d_z d_x + d_x$$

be the number of real coordinates in a five-block summary, and let

$$\Psi_{\mathrm{sum}}(s) = \big(s.M_0, s.M_1, s.N_0, s.N_1, s.m_X\big) \in \mathbb{R}^{d_{\mathrm{sum}}}$$

be the coordinate homeomorphism from the summary space to Euclidean space. Set

$$\mathcal{K}^\Psi = \Psi_{\mathrm{sum}}(\mathcal{K})$$

and, for $x_{\mathrm{Euc}}, y_{\mathrm{Euc}} \in \mathbb{R}^{d_{\mathrm{sum}}}$, define the transported loss

$$\ell_\Psi(x_{\mathrm{Euc}}, y_{\mathrm{Euc}}) = d_S\big(\Psi^{-1}_{\mathrm{sum}}(y_{\mathrm{Euc}}), \Psi^{-1}_{\mathrm{sum}}(x_{\mathrm{Euc}})\big)\,.$$

The set $\mathcal{K}^\Psi$ is nonempty and compact, and ℓ_Ψ is continuous because Ψ_{sum} is a homeomorphism and Lemma 23 gives continuity of d_S. Applying Lemma 2 to $(\mathcal{K}^\Psi, \ell_\Psi)$, then transporting the selector back through Ψ_{sum}, gives a Borel map

$$\Pi : \text{summary space} \longrightarrow \text{summary space}$$

such that, for every summary s,

$$\Pi(s) \in \mathcal{K} \quad \text{and} \quad \forall q_{\text{sum}} \in \mathcal{K}, \qquad d_S(\Pi(s), s) \leq d_S(q_{\text{sum}}, s).$$

Together with the $\overline{F}$ from the first step, this defines repair data R as in Algorithm 1; the fallback clause has premise $\mathcal{K} = \varnothing$, which is false in the present case.

4. For $\mathcal{K} \neq \varnothing$, the repaired estimator is

$$\widehat{\nu}_n(\mathsf{o}_{1:n}) = \overline{F}\Big(\Pi(\widehat{S}_n(\mathsf{o}_{1:n}))\Big), \qquad \mathsf{o}_{1:n} \in O^n_{d_x,d_z}.$$

The empirical summary map $\widehat{S}_n$ is Borel by hypothesis, Π is Borel by the preceding step, and $\overline{F}$ is Borel because it is continuous. Hence $\widehat{\nu}_n$ is Borel measurable.

5. It remains in the nonempty case to verify nonemptiness of the retained summary-inversion set. By Definition 45,

$$r_{n,\alpha} = C_0 L \sqrt{\frac{\log(C_0/\alpha)}{n}}.$$

Since $n \geq 1$, $0 < \alpha < 1/2$, $C_0 \geq 1$, and $L \geq 1$, one has $C_0/\alpha > 1$, hence $\log(C_0/\alpha) > 0$, and therefore

$$0 < r_{n,\alpha} \qquad \text{and} \qquad 0 \leq 2r_{n,\alpha}.$$

For a fixed sample $\mathsf{o}_{1:n}$, put

$$q_{\text{ret}} = \Pi(\widehat{S}_n(\mathsf{o}_{1:n})).$$

The selector property gives $q_{\text{ret}} \in \mathcal{K}$, and the defining formula for d_S gives

$$d_S(q_{\text{ret}}, q_{\text{ret}}) = 0 \leq 2r_{n,\alpha}.$$

Moreover,

$$\widehat{\nu}_n(\mathsf{o}_{1:n}) = \overline{F}(q_{\text{ret}}).$$

Thus q_{ret} witnesses membership of $\widehat{\nu}_n(\mathsf{o}_{1:n})$ in the displayed retained set, so that set is nonempty for every sample.

6. The two cases $\mathcal{K} = \varnothing$ and $\mathcal{K} \neq \varnothing$ cover all possibilities. In each case the constructed repair data R has Borel selector Π, continuous extension $\overline{F}$, Borel repaired estimator $\widehat{\nu}_n$, the asserted δ_0 fallback when $\mathcal{K} = \varnothing$, and a nonempty retained summary-inversion set for every sample.

□

Proof of Proposition 4. Set

$$s_0 := \pi_0 \sigma_0^2, \qquad L_\tau := \frac{4L\sqrt{d_z}}{\sigma_0},$$

and order singular values nonincreasingly.

1. Let H_n, q_n, ℓ_n and

$$\mathcal{A}_n = \mathcal{V}_n \times \mathcal{R}_n \times \mathcal{P}_n \times \mathcal{T}_n$$

be the structured lattice of Definition 13. Its four factors have cardinalities bounded by fixed-parameter constants times

$$H_n^{d_x k}, \qquad H_n^{k^2}, \qquad H_n^{k-1}, \qquad H_n^{k}.$$

Since $n \geq 1$, the lattice height satisfies

$$H_n \leq C_H \sqrt{n}$$

for a finite $C_H = C_H(k, d_x, \pi_0, \sigma_0)$. Hence there is a constant $C_\# > 0$, depending only on the displayed parameters, such that

$$|\mathcal{A}_n| \leq C_\# n^{(d_x k + k^2 + 2k - 1)/2}.$$

In the fixed-dimensional exact-real model of Definition 48, the work is the empirical-summary pass plus one fixed-size objective evaluation for each candidate, so

$$\mathrm{Work}_n \leq C_\# \left\{ n + n^{(d_x k + k^2 + 2k - 1)/2} \right\}.$$

2. Choose $C_0 \geq 1$ from Lemma 1. Thus, for every $P \in \mathcal{M}$, $n \geq 1$, and $0 < \eta < 1/2$,

$$Q_P^{(n)} \left\{ d_S(\widehat{S}_n, S(P)) > C_0 L \sqrt{\frac{\log(C_0/\eta)}{n}} \right\} \leq \eta.$$

The assumptions imply $d_z > 0$, $L_\tau > 0$, and $s_0 > 0$. Therefore

$$0 < \frac{8L_\tau}{s_0} \leq C_{\mathrm{lat}} = \max\left\{ \frac{8L_\tau}{s_0}, (L_\tau + K_D + LK_f) B_{\mathrm{lat}} \right\}.$$

Define

$$C_{\mathrm{tail}} := \max\{C_0, \mathrm{e}, C_{\mathrm{lat}}(C_0 L + 1)\}, \qquad C := \max\{C_\#, C_{\mathrm{tail}}\}.$$

Then $C > 0$.

3. Fix $n \geq 1$. The lattice factors are finite and nonempty: canonical coordinate columns furnish an element of $\mathcal{V}_n$, a rounded scalar multiple of I_k furnishes an element of $\mathcal{R}_n$, the inequality $k\ell_n \leq H_n$ furnishes an element of $\mathcal{P}_n$, and $0 \in \mathcal{T}_n$. Enumerate $\mathcal{A}_n$ in duplicate-free lexicographic order as

$$\vartheta_i = (V_i, R_i, p_i, \tau_i), \qquad i = 1, \ldots, |\mathcal{A}_n|.$$

For a summary s, put

$$i_n(s) := \min_{\mathrm{lex}} \operatorname*{arg\,min}_{1 \leq i \leq |\mathcal{A}_n|} J_n(\vartheta_i; s), \qquad \Lambda_n(s) := \lambda_{\vartheta_{i_n(s)}},$$

where J_n is the criterion in Definition 16. Thresholded singular-value pseudoinversion is Borel, each score $s \mapsto J_n(\vartheta_i; s)$ is Borel, and a finite first-minimizer rule over Borel scores is Borel. Hence

$$\widehat{\lambda}_n(\omega) := \Lambda_n(\widehat{S}_n(\omega))$$

is a total Borel empirical-summary rule, with $\widehat{S}_n$ as in Definition 35, and it is the first minimizer prescribed by Algorithm 2. Every coordinate of every $p_i \in \mathcal{P}_n$ is at least π_0, so aggregating coordinates at repeated effect locations gives

$$\mathrm{AtomFloor}(\pi_0, \widehat{\lambda}_n(\omega))$$

for every sample ω.

4. Fix $P \in \mathcal{M}$, and put
$$G_P := \operatorname{stack}(M_0(P), M_1(P)).$$
Lemma 25 gives $\operatorname{rank}(G_P) = k$ and the threshold characterization at $S(P)$. In addition, Lemma 24 gives the population signal margin; since the singular values are nonincreasing,
$$\sigma_j(G_P) \geq s_0 \quad (1 \leq j \leq k), \qquad \sigma_j(G_P) = 0 \quad (j > k).$$
In particular, G_P has exactly its first k singular values at least $s_0/2$. Now let s satisfy
$$d_S(s, S(P)) < s_0/2,$$
and set $G_s := \operatorname{stack}(M_0(s), M_1(s))$. The stacked-block comparison gives
$$\|G_s - G_P\|_{\mathrm{op}} \leq d_S(s, S(P)) < s_0/2.$$
Weyl's inequality yields
$$|\sigma_j(G_s) - \sigma_j(G_P)| < s_0/2 \qquad \text{for every } j.$$
Thus, for $1 \leq j \leq k$,
$$\sigma_j(G_s) > s_0/2,$$
while, for $j > k$,
$$\sigma_j(G_s) < s_0/2.$$
This proves that G_s also has exactly its first k singular values at least $s_0/2$ and all remaining singular values below $s_0/2$.

5. Fix a sample ω, and write
$$s_\omega := \widehat{S}_n(\omega), \qquad \varepsilon_{\omega,P} := d_S(s_\omega, S(P)).$$
Choose the population signal tuple
$$B(P) = V_{\mathrm{pop}} R_{\mathrm{pop}}^\top, \qquad p_{\mathrm{pop}} = (P(U = u))_{u=1}^k, \qquad \tau_{\mathrm{pop}} = (\tau_u(P))_{u=1}^k.$$
By Definitions 21 and 25, and because the latent classes partition the record space with positive class masses,
$$m_X(P) = E_P[X] = \sum_{u=1}^k P(U = u) E_P[X \mid U = u] = B(P) p_{\mathrm{pop}}.$$
The chosen signal factorization $B(P) = V_{\mathrm{pop}} R_{\mathrm{pop}}^\top$ therefore gives
$$m_X(P) = V_{\mathrm{pop}} R_{\mathrm{pop}}^\top p_{\mathrm{pop}}.$$
Using Assumption 5, the first target-proxy coordinate has conditional mean one in every latent class:
$$B(P)^\top e_1 = \big(E_P[X_1 \mid U = u]\big)_{u=1}^k = \mathbf{1}_k.$$

Since $B(P)^\top = R_{\mathrm{pop}} V_{\mathrm{pop}}^\top$, this yields

$$R_{\mathrm{pop}} V_{\mathrm{pop}}^\top e_1 = \mathbf{1}_k.$$

The proxy and outcome moment factorizations from Lemmas 4 and 18 give, for each arm t,

$$M_t(P) = A_t \Gamma_t B(P)^\top, \qquad N_t(P) = A_t \Gamma_t \operatorname{diag}(\mu_{tu}) B(P)^\top.$$

The proxy-rank margin in Assumption 10 supplies injectivity of the linear maps represented by A_t and $B(P)$. By Lemma 19, Γ_t is injective, so $A_t\Gamma_t$ is injective. Applying Lemma 20 with $C = A_t\Gamma_t$, $B = B(P)$, and $\mu = (\mu_{tu})_u$ yields

$$M_t(P)^\dagger N_t(P) = \{B(P)^\dagger\}^\top \operatorname{diag}(\mu_{tu}) B(P)^\top.$$

Subtracting the two arms and using $\tau_{\mathrm{pop},u} = \mu_{1u} - \mu_{0u}$, together with $B(P) = V_{\mathrm{pop}} R_{\mathrm{pop}}^\top$, gives

$$\mathcal{D}_P := M_1(P)^\dagger N_1(P) - M_0(P)^\dagger N_0(P) = V_{\mathrm{pop}} R_{\mathrm{pop}}^{-1} \operatorname{diag}(\tau_{\mathrm{pop}}) R_{\mathrm{pop}} V_{\mathrm{pop}}^\top.$$

Moreover,

$$\sigma_k(R_{\mathrm{pop}}) \geq \sigma_0, \qquad \|R_{\mathrm{pop}}\|_{\mathrm{op}} \leq \sqrt{k}L,$$

$$p_{\mathrm{pop},u} \geq 2\pi_0, \qquad \sum_{u=1}^k p_{\mathrm{pop},u} = 1, \qquad |\tau_{\mathrm{pop},u}| \leq L_\tau,$$

where the effect envelope is supplied by Lemma 7. The associated quotient law is

$$\nu_P = \sum_{u=1}^k p_{\mathrm{pop},u} \delta_{\tau_{\mathrm{pop},u}}.$$

6. Recall that $q_n = H_n^{-1}$. The definition of H_n gives the four mesh inequalities

$$\sqrt{d_x k}\, q_n \leq \frac{1}{4}, \qquad kq_n \leq \frac{\sigma_0}{2}, \qquad kq_n \leq \sqrt{k}L, \qquad \lceil \pi_0^{-1} \rceil \leq H_n.$$

For the signal basis V_{pop}, each entry has absolute value at most one because its columns are orthonormal. Coordinatewise clipped rounding therefore gives a grid matrix $V_{n,\mathrm{raw}}$ with

$$(V_{n,\mathrm{raw}})_{ij} = q_n z_{ij} \quad (z_{ij} \in \mathbb{Z}), \qquad |(V_{n,\mathrm{raw}})_{ij}| \leq 1, \qquad \|V_{n,\mathrm{raw}} - V_{\mathrm{pop}}\|_{\mathrm{op}} \leq \sqrt{d_x k}\, q_n.$$

Weyl's inequality and the first mesh bound give

$$\sigma_k(V_{n,\mathrm{raw}}) \geq 1 - \sqrt{d_x k}\, q_n \geq \frac{3}{4} \geq \frac{1}{2}.$$

Let V_n° be the polar factor $V_{n,\mathrm{raw}}(V_{n,\mathrm{raw}}^\top V_{n,\mathrm{raw}})^{-1/2}$. The polar-factor perturbation estimate at an orthonormal basis yields

$$\|V_n^\circ - V_{\mathrm{pop}}\|_{\mathrm{op}} \leq 4\sqrt{d_x k}\, q_n = c_V q_n.$$

Next, since $\|R_{\mathrm{pop}}\|_{\mathrm{op}} \leq \sqrt{k}L$, coordinatewise clipped rounding in the radius $2\sqrt{k}L$ gives R_n° with grid entries and

$$\|R_n^\circ - R_{\mathrm{pop}}\|_{\mathrm{op}} \leq kq_n.$$

Using $kq_n \le \sigma_0/2$ and Weyl's inequality,

$$\sigma_k(R_n^\circ) \ge \sigma_k(R_{\mathrm{pop}}) - kq_n \ge \frac{\sigma_0}{2},$$

and using $kq_n \le \sqrt{k}L$,

$$\|R_n^\circ\|_{\mathrm{op}} \le \|R_{\mathrm{pop}}\|_{\mathrm{op}} + \|R_n^\circ - R_{\mathrm{pop}}\|_{\mathrm{op}} \le 2\sqrt{k}L.$$

For the weight vector, set $b_u := \lfloor H_n p_{\mathrm{pop},u} \rfloor$. Since $p_{\mathrm{pop},u} \ge 2\pi_0$ and $\lceil \pi_0^{-1} \rceil \le H_n$, the quantities $x := \pi_0 H_n$ and $y_u := H_n p_{\mathrm{pop},u}$ satisfy $1 \le x$ and $2x \le y_u$. Because $y_u < b_u + 1$, this implies $x \le b_u$, hence

$$\lceil \pi_0 H_n \rceil \le b_u.$$

The floor vector satisfies $\sum_u b_u \le H_n$ and $H_n \le \sum_u b_u + k$. Choose a subset of indices of size $H_n - \sum_u b_u$ and add one to b_u on that subset, obtaining integers a_u with

$$\sum_{u=1}^{k} a_u = H_n, \qquad \lceil \pi_0 H_n \rceil \le a_u, \qquad \left| \frac{a_u}{H_n} - p_{\mathrm{pop},u} \right| \le q_n.$$

Thus $p_n^\circ := (a_u/H_n)_{u=1}^k$ belongs to $\mathcal{P}_n$, and summing the squared coordinate errors gives

$$\|p_n^\circ - p_{\mathrm{pop}}\|_2 = \left\{ \sum_{u=1}^{k} \left(\frac{a_u}{H_n} - p_{\mathrm{pop},u} \right)^2 \right\}^{1/2} \le \sqrt{k} q_n.$$

Finally, since $|\tau_{\mathrm{pop},u}| \le L_\tau$, coordinatewise clipped rounding on the effect grid gives $\tau_n^\circ \in \mathcal{T}_n$ with

$$|\tau_{n,u}^\circ - \tau_{\mathrm{pop},u}| \le q_n \qquad (1 \le u \le k).$$

Combining these four rounded components gives a well-formed lattice comparator

$$\vartheta_n^\circ = (V_n^\circ, R_n^\circ, p_n^\circ, \tau_n^\circ) \in \mathcal{A}_n$$

satisfying

$$\|V_n^\circ - V_{\mathrm{pop}}\|_{\mathrm{op}} \le c_V q_n, \qquad \|R_n^\circ - R_{\mathrm{pop}}\|_{\mathrm{op}} \le kq_n,$$
$$\|p_n^\circ - p_{\mathrm{pop}}\|_2 \le \sqrt{k} q_n, \qquad |\tau_{n,u}^\circ - \tau_{\mathrm{pop},u}| \le q_n.$$

7. Assume first that

$$\varepsilon_{\omega,P} < s_0/4.$$

For each treatment arm,

$$\|M_t(s_\omega) - M_t(P)\|_{\mathrm{op}} \le \varepsilon_{\omega,P}, \qquad \|N_t(s_\omega) - N_t(P)\|_{\mathrm{op}} \le \varepsilon_{\omega,P}.$$

The population armwise proxy block has rank k, kth singular value at least s_0, and $\|N_t(P)\|_{\mathrm{op}} \le L$. If $\widetilde{M}_t(s_\omega)$ denotes the rank-k truncation retained at threshold $s_0/2$, singular-value perturbation gives

$$(M_t(s_\omega))^\dagger_{\ge s_0/2} = \widetilde{M}_t(s_\omega)^\dagger, \quad \sigma_k(\widetilde{M}_t(s_\omega)) \ge 3s_0/4, \quad \|\widetilde{M}_t(s_\omega) - M_t(P)\|_{\mathrm{op}} \le 2\varepsilon_{\omega,P}.$$

Define the thresholded empirical compressed operator, in the notation of Definition 16, by

$$\widehat{D}(s_\omega) := M_1(s_\omega)^\dagger{}_{\geq s_0/2} N_1(s_\omega) - M_0(s_\omega)^\dagger{}_{\geq s_0/2} N_0(s_\omega).$$

The equal-rank Moore–Penrose perturbation identity gives

$$\|\mathcal{U}^\dagger - \mathcal{V}^\dagger\|_{\mathrm{op}} \leq 3 \max\{\|\mathcal{U}^\dagger\|_{\mathrm{op}}^2, \|\mathcal{V}^\dagger\|_{\mathrm{op}}^2\} \|\mathcal{U} - \mathcal{V}\|_{\mathrm{op}},$$

and therefore

$$\|\widehat{D}(s_\omega) - \mathcal{D}_P\|_{\mathrm{op}} \leq A_D \varepsilon_{\omega,P}, \qquad \|m_X(s_\omega) - m_X(P)\|_2 \leq \varepsilon_{\omega,P}.$$

For the comparator ϑ_n°, the mean residual decomposes as

$$V_n^\circ (R_n^\circ)^\top p_n^\circ - V_{\mathrm{pop}} R_{\mathrm{pop}}^\top p_{\mathrm{pop}} = (V_n^\circ - V_{\mathrm{pop}})(R_n^\circ)^\top p_n^\circ + V_{\mathrm{pop}} (R_n^\circ - R_{\mathrm{pop}})^\top p_n^\circ + V_{\mathrm{pop}} R_{\mathrm{pop}}^\top (p_n^\circ - p_{\mathrm{pop}}),$$

so the bounds $\|V_{\mathrm{pop}}\|_{\mathrm{op}} \leq 1$, $\|p_n^\circ\|_2 \leq 1$, $\|R_n^\circ\|_{\mathrm{op}} \leq 2\sqrt{k}L$, and $\|R_{\mathrm{pop}}\|_{\mathrm{op}} \leq \sqrt{k}L$ give

$$\|V_n^\circ (R_n^\circ)^\top p_n^\circ - V_{\mathrm{pop}} R_{\mathrm{pop}}^\top p_{\mathrm{pop}}\|_2 \leq c_m q_n.$$

Similarly,

$$R_n^\circ (V_n^\circ)^\top e_1 - R_{\mathrm{pop}} V_{\mathrm{pop}}^\top e_1 = (R_n^\circ - R_{\mathrm{pop}}) V_{\mathrm{pop}}^\top e_1 + R_n^\circ (V_n^\circ - V_{\mathrm{pop}})^\top e_1,$$

and $\|e_1\|_2 = 1$, $\|V_{\mathrm{pop}}\|_{\mathrm{op}} \leq 1$, and $\|R_n^\circ\|_{\mathrm{op}} \leq 2\sqrt{k}L$ yield

$$\|R_n^\circ (V_n^\circ)^\top e_1 - \mathbf{1}_k\|_2 \leq c_b q_n.$$

For the operator residual, write

$$\begin{aligned}
&D_{\vartheta_n^\circ} - \mathcal{D}_P \\
&= (V_n^\circ - V_{\mathrm{pop}})(R_n^\circ)^{-1} \operatorname{diag}(\tau_n^\circ) R_n^\circ (V_n^\circ)^\top \\
&\quad + V_{\mathrm{pop}} \{(R_n^\circ)^{-1} - R_{\mathrm{pop}}^{-1}\} \operatorname{diag}(\tau_n^\circ) R_n^\circ (V_n^\circ)^\top \\
&\quad + V_{\mathrm{pop}} R_{\mathrm{pop}}^{-1} \operatorname{diag}(\tau_n^\circ - \tau_{\mathrm{pop}}) R_n^\circ (V_n^\circ)^\top \\
&\quad + V_{\mathrm{pop}} R_{\mathrm{pop}}^{-1} \operatorname{diag}(\tau_{\mathrm{pop}}) (R_n^\circ - R_{\mathrm{pop}}) (V_n^\circ)^\top \\
&\quad + V_{\mathrm{pop}} R_{\mathrm{pop}}^{-1} \operatorname{diag}(\tau_{\mathrm{pop}}) R_{\mathrm{pop}} (V_n^\circ - V_{\mathrm{pop}})^\top.
\end{aligned}$$

The inverse identity

$$(R_n^\circ)^{-1} - R_{\mathrm{pop}}^{-1} = (R_n^\circ)^{-1} (R_{\mathrm{pop}} - R_n^\circ) R_{\mathrm{pop}}^{-1}$$

and the singular-value bounds give

$$\|(R_n^\circ)^{-1}\|_{\mathrm{op}} \leq \frac{2}{\sigma_0}, \qquad \|R_{\mathrm{pop}}^{-1}\|_{\mathrm{op}} \leq \frac{1}{\sigma_0}, \qquad \|(R_n^\circ)^{-1} - R_{\mathrm{pop}}^{-1}\|_{\mathrm{op}} \leq \frac{2kq_n}{\sigma_0^2}.$$

Bounding the five displayed summands with $\|\operatorname{diag}(\tau_n^\circ)\|_{\mathrm{op}} \leq L_\tau$, $\|\operatorname{diag}(\tau_{\mathrm{pop}})\|_{\mathrm{op}} \leq L_\tau$, and $\|\operatorname{diag}(\tau_n^\circ - \tau_{\mathrm{pop}})\|_{\mathrm{op}} \leq q_n$ gives

$$\|D_{\vartheta_n^\circ} - \mathcal{D}_P\|_{\mathrm{op}} \leq c_D q_n.$$

Consequently

$$J_n(\vartheta_n^\circ; s_\omega) \le (A_D + 1)\varepsilon_{\omega,P} + c_{\mathrm{grid}} q_n.$$

Let

$$\widehat{\vartheta}_n(\omega) = (\widehat{V}_n, \widehat{R}_n, \widehat{p}_n, \widehat{\tau}_n)$$

be the first minimizer. Minimality and the nonnegativity of the three terms in J_n imply

$$\|D_{\widehat{\vartheta}_n} - \mathcal{D}_P\|_{\mathrm{op}} \le (2A_D + 1)\varepsilon_{\omega,P} + c_{\mathrm{grid}} q_n,$$

$$\|\widehat{V}_n \widehat{R}_n^\top \widehat{p}_n - m_X(P)\|_2 \le (A_D + 2)\varepsilon_{\omega,P} + c_{\mathrm{grid}} q_n,$$

and

$$\|\widehat{R}_n \widehat{V}_n^\top e_1 - \mathbf{1}_k\|_2 \le (A_D + 1)\varepsilon_{\omega,P} + c_{\mathrm{grid}} q_n.$$

Since

$$B_{\mathrm{lat}} = 2(A_D + 1) + c_{\mathrm{grid}},$$

all three residuals are bounded by

$$B_{\mathrm{lat}}(\varepsilon_{\omega,P} + q_n).$$

8. Let φ_ω be the attaining Kantorovich–Rubinstein potential for $(\widehat{\lambda}_n(\omega), \nu_P)$ supplied by Lemma 21. Thus

$$\varphi_\omega(0) = 0, \qquad \mathrm{Lip}(\varphi_\omega) \le 1,$$

and

$$W_1(\widehat{\lambda}_n(\omega), \nu_P) = \left| \int \varphi_\omega \, d\widehat{\lambda}_n(\omega) - \int \varphi_\omega \, d\nu_P \right|.$$

The selected structured operator and $\mathcal{D}_P$, extended by zero on the orthogonal complement of their signal spaces, have real diagonalizations with spectra in $[-L_\tau, L_\tau]$. Put

$$\kappa := \frac{4\sqrt{k}L}{\sigma_0}.$$

For either signal factorization with coordinate matrix R_{coord}, the ambient eigenbasis uses the signal coordinates and an orthonormal basis on the orthogonal complement. Hence its condition number is bounded by

$$\mathrm{cond} \le \max\{\|R_{\mathrm{coord}}^{-1}\|_{\mathrm{op}}, 1\} \max\{\|R_{\mathrm{coord}}\|_{\mathrm{op}}, 1\}.$$

For the selected lattice tuple, the defining lattice constraints give

$$\sigma_k(\widehat{R}_n) \ge \sigma_0/2, \qquad \|\widehat{R}_n\|_{\mathrm{op}} \le 2\sqrt{k}L, \qquad |\widehat{\tau}_{n,u}| \le L_\tau.$$

Therefore

$$\|\widehat{R}_n^{-1}\|_{\mathrm{op}} \le \frac{2}{\sigma_0}, \qquad \max\{\|\widehat{R}_n^{-1}\|_{\mathrm{op}}, 1\} \le \frac{2}{\sigma_0}, \qquad \max\{\|\widehat{R}_n\|_{\mathrm{op}}, 1\} \le 2\sqrt{k}L,$$

where the last two inequalities use $0 < \sigma_0 \le 1$, $2 \le k$, and $1 \le L$. Thus the selected diagonalization has condition number at most κ. For the population tuple,

$$\sigma_k(R_{\mathrm{pop}}) \ge \sigma_0, \qquad \|R_{\mathrm{pop}}\|_{\mathrm{op}} \le \sqrt{k}L, \qquad |\tau_{\mathrm{pop},u}| \le L_\tau,$$

so the same calculation, with

$$\|R_{\mathrm{pop}}^{-1}\|_{\mathrm{op}} \le \frac{1}{\sigma_0} \le \frac{2}{\sigma_0}, \qquad \max\{\|R_{\mathrm{pop}}\|_{\mathrm{op}}, 1\} \le 2\sqrt{k}L,$$

gives the condition-number bound κ for the population diagonalization as well. For the diagonalizations

$$D_{\widehat{\vartheta}_n} = \mathcal{Z}_{\mathrm{hat}}\Lambda_{\mathrm{hat}}\mathcal{Z}_{\mathrm{hat}}^{-1}, \qquad \mathcal{D}_P = \mathcal{Z}_{\mathrm{pop}}\Lambda_{\mathrm{pop}}\mathcal{Z}_{\mathrm{pop}}^{-1},$$

the divided-difference identity gives

$$\{\varphi_\omega(\Lambda_{\mathrm{hat}})\mathcal{Z}_{\mathrm{hat}}^{-1}\mathcal{Z}_{\mathrm{pop}} - \mathcal{Z}_{\mathrm{hat}}^{-1}\mathcal{Z}_{\mathrm{pop}}\varphi_\omega(\Lambda_{\mathrm{pop}})\}_{ij} = d_{ij}\{\Lambda_{\mathrm{hat}}\mathcal{Z}_{\mathrm{hat}}^{-1}\mathcal{Z}_{\mathrm{pop}} - \mathcal{Z}_{\mathrm{hat}}^{-1}\mathcal{Z}_{\mathrm{pop}}\Lambda_{\mathrm{pop}}\}_{ij}, \qquad |d_{ij}| \le 1,$$

with $d_{ij} = 0$ when the two eigenvalues coincide. The coordinate perturbation in the last display is

$$\Lambda_{\mathrm{hat}}\mathcal{Z}_{\mathrm{hat}}^{-1}\mathcal{Z}_{\mathrm{pop}} - \mathcal{Z}_{\mathrm{hat}}^{-1}\mathcal{Z}_{\mathrm{pop}}\Lambda_{\mathrm{pop}} = \mathcal{Z}_{\mathrm{hat}}^{-1}(D_{\widehat{\vartheta}_n} - \mathcal{D}_P)\mathcal{Z}_{\mathrm{pop}}.$$

Since $|d_{ij}| \le 1$, each column of the Schur product has Euclidean norm bounded by the corresponding column norm of this coordinate perturbation. The fixed-dimensional column estimate therefore gives

$$\left\|\varphi_\omega(\Lambda_{\mathrm{hat}})\mathcal{Z}_{\mathrm{hat}}^{-1}\mathcal{Z}_{\mathrm{pop}} - \mathcal{Z}_{\mathrm{hat}}^{-1}\mathcal{Z}_{\mathrm{pop}}\varphi_\omega(\Lambda_{\mathrm{pop}})\right\|_{\mathrm{op}} \le \sqrt{d_x}\,\left\|\mathcal{Z}_{\mathrm{hat}}^{-1}(D_{\widehat{\vartheta}_n} - \mathcal{D}_P)\mathcal{Z}_{\mathrm{pop}}\right\|_{\mathrm{op}}.$$

Multiplying by the exterior diagonalizer factors yields

$$\begin{aligned}\|\varphi_\omega(D_{\widehat{\vartheta}_n}) - \varphi_\omega(\mathcal{D}_P)\|_{\mathrm{op}} &\le \sqrt{d_x}\,\|\mathcal{Z}_{\mathrm{hat}}\|_{\mathrm{op}}\|\mathcal{Z}_{\mathrm{hat}}^{-1}\|_{\mathrm{op}}\|D_{\widehat{\vartheta}_n} - \mathcal{D}_P\|_{\mathrm{op}}\|\mathcal{Z}_{\mathrm{pop}}\|_{\mathrm{op}}\|\mathcal{Z}_{\mathrm{pop}}^{-1}\|_{\mathrm{op}}\\ &\le \sqrt{d_x}\,\kappa^2\|D_{\widehat{\vartheta}_n} - \mathcal{D}_P\|_{\mathrm{op}}.\end{aligned}$$

Because $\kappa = 4\sqrt{k}L/\sigma_0$, this is

$$\|\varphi_\omega(D_{\widehat{\vartheta}_n}) - \varphi_\omega(\mathcal{D}_P)\|_{\mathrm{op}} \le K_f\|D_{\widehat{\vartheta}_n} - \mathcal{D}_P\|_{\mathrm{op}}, \qquad K_f = \frac{16\sqrt{d_x}\,kL^2}{\sigma_0^2}.$$

For the selected law, put

$$\chi_{\vartheta,u} := (\widehat{R}_n\widehat{V}_n^\top e_1)_u.$$

Since $\widehat{V}_n^\top\widehat{V}_n = I_k$ and $\widehat{R}_n\widehat{R}_n^{-1} = I_k$, the functional calculus of the selected structured operator gives

$$\varphi_\omega(D_{\widehat{\vartheta}_n}) = \widehat{V}_n\widehat{R}_n^{-1}\operatorname{diag}\{\varphi_\omega(\widehat{\tau}_{n,u}) : 1 \le u \le k\}\widehat{R}_n\widehat{V}_n^\top,$$

because $\varphi_\omega(0) = 0$ on the orthogonal complement. Therefore

$$\begin{aligned}&(\widehat{V}_n\widehat{R}_n^\top\widehat{p}_n)^\top\varphi_\omega(D_{\widehat{\vartheta}_n})e_1\\ &\quad= \widehat{p}_n^\top\widehat{R}_n\widehat{V}_n^\top\widehat{V}_n\widehat{R}_n^{-1}\operatorname{diag}\{\varphi_\omega(\widehat{\tau}_{n,u})\}\widehat{R}_n\widehat{V}_n^\top e_1\\ &\quad= \widehat{p}_n^\top\operatorname{diag}\{\varphi_\omega(\widehat{\tau}_{n,u})\}\widehat{R}_n\widehat{V}_n^\top e_1 = \sum_{u=1}^k \widehat{p}_{n,u}\varphi_\omega(\widehat{\tau}_{n,u})\chi_{\vartheta,u}.\end{aligned}$$

The labelled integral of $\widehat{\lambda}_n(\omega)$ is

$$\int \varphi_\omega\, d\widehat{\lambda}_n(\omega) = \sum_{u=1}^k \widehat{p}_{n,u}\varphi_\omega(\widehat{\tau}_{n,u}),$$

and hence

$$\int \varphi_\omega \, d\widehat{\lambda}_n(\omega) - (\widehat{V}_n \widehat{R}_n^\top \widehat{p}_n)^\top \varphi_\omega(D_{\widehat{\vartheta}_n}) e_1 = \sum_{u=1}^{k} \widehat{p}_{n,u} \varphi_\omega(\widehat{\tau}_{n,u})(1 - \chi_{\widehat{\vartheta},u}).$$

Because $\|\widehat{p}_n\|_2 \le 1$ and

$$|\varphi_\omega(\widehat{\tau}_{n,u})| = |\varphi_\omega(\widehat{\tau}_{n,u}) - \varphi_\omega(0)| \le |\widehat{\tau}_{n,u}| \le L_\tau,$$

Cauchy's inequality yields

$$\left| \int \varphi_\omega \, d\widehat{\lambda}_n(\omega) - (\widehat{V}_n \widehat{R}_n^\top \widehat{p}_n)^\top \varphi_\omega(D_{\widehat{\vartheta}_n}) e_1 \right| \le L_\tau \|\widehat{R}_n \widehat{V}_n^\top e_1 - \mathbf{1}_k\|_2.$$

For the population tuple, functional calculus gives

$$\varphi_\omega(\mathcal{D}_P) = V_{\mathrm{pop}} R_{\mathrm{pop}}^{-1} \operatorname{diag}\{\varphi_\omega(\tau_{\mathrm{pop}})\} R_{\mathrm{pop}} V_{\mathrm{pop}}^\top,$$

since $\varphi_\omega(0) = 0$ on the orthogonal complement. Using $m_X(P) = V_{\mathrm{pop}} R_{\mathrm{pop}}^\top p_{\mathrm{pop}}$, $V_{\mathrm{pop}}^\top V_{\mathrm{pop}} = I_k$, and $R_{\mathrm{pop}} V_{\mathrm{pop}}^\top e_1 = \mathbf{1}_k$, we obtain

$$m_X(P)^\top \varphi_\omega(\mathcal{D}_P) e_1 = \sum_{u=1}^{k} p_{\mathrm{pop},u} \varphi_\omega(\tau_{\mathrm{pop},u}) = \int \varphi_\omega \, d\nu_P.$$

Together with $\|m_X(P)\|_2 \le L$ and

$$\|\varphi_\omega(D_{\widehat{\vartheta}_n}) e_1\|_2 \le \kappa L_\tau = K_D,$$

the residual bounds from the previous step give, in the small-error case,

$$W_1(\widehat{\lambda}_n(\omega), \nu_P) \le (L_\tau + K_D + L K_f) B_{\mathrm{lat}}(\varepsilon_{\omega,P} + q_n).$$

If instead $\varepsilon_{\omega,P} \ge s_0/4$, both laws are supported on $[-L_\tau, L_\tau]$, so the diameter bound from Definition 19 gives

$$W_1(\widehat{\lambda}_n(\omega), \nu_P) \le 2L_\tau \le \frac{8L_\tau}{s_0} \varepsilon_{\omega,P}.$$

Finally $q_n \le n^{-1/2}$. Combining the two cases with the definition of C_{lat} proves, for every sample,

$$W_1(\widehat{\lambda}_n(\omega), \nu_P) \le C_{\mathrm{lat}} \left\{ d_S(\widehat{S}_n(\omega), S(P)) + \frac{1}{\sqrt{n}} \right\}.$$

9. The choice $C \ge C_\#$ gives the displayed candidate-count and work bounds with the same C. For the tail bound, fix $0 < \eta < 1/2$. Since

$$C \ge C_0, \qquad C \ge \mathrm{e}, \qquad C \ge C_{\mathrm{lat}}(C_0 L + 1),$$

we have

$$\log(C_0/\eta) \le \log(C/\eta), \qquad 1 \le \log(C/\eta), \qquad \frac{1}{\sqrt{n}} \le \sqrt{\frac{\log(C/\eta)}{n}}.$$

On the event

$$d_S(\widehat{S}_n, S(P)) \le C_0 L\sqrt{\frac{\log(C_0/\eta)}{n}},$$

the deterministic inequality gives

$$W_1(\widehat{\lambda}_n, \nu_P) \le C_{\text{lat}}(C_0 L + 1)\sqrt{\frac{\log(C/\eta)}{n}} \le C\sqrt{\frac{\log(C/\eta)}{n}}.$$

Therefore

$$\left\{W_1(\widehat{\lambda}_n, \nu_P) > C\sqrt{\frac{\log(C/\eta)}{n}}\right\} \subseteq \left\{d_S(\widehat{S}_n, S(P)) > C_0 L\sqrt{\frac{\log(C_0/\eta)}{n}}\right\}.$$

Applying Lemma 1 under $Q_P^{(n)}$ yields

$$Q_P^{(n)}\left\{W_1(\widehat{\lambda}_n, \nu_P) > C\sqrt{\frac{\log(C/\eta)}{n}}\right\} \le \eta.$$

□

Proof of Theorem 2. 1. Let

$$\mathscr{S} = \{S(P') : P' \in \mathcal{M}\}, \qquad \mathcal{K} = \overline{\mathscr{S}}$$

in the finite-dimensional summary space equipped with d_S. By Lemma 1, choose a simultaneous concentration constant $C_0 \ge 1$. By Proposition 4, choose a positive lattice tail constant C_{tail} associated with the structured-lattice estimator. By Theorem 1, choose $C_{\text{mod}} > 0$ and a continuous extension

$$\overline{F} : \mathcal{K} \to \mathcal{P}_k([-L_\tau, L_\tau])$$

such that

$$W_1\{\overline{F}(s), \overline{F}(s')\} \le C_{\text{mod}}\, d_S(s, s'), \qquad s, s' \in \mathcal{K},$$

and

$$\overline{F}(S(Q)) = \nu_Q, \qquad Q \in \mathcal{M}.$$

Set

$$C = \max\{2C_{\text{tail}},\, \max(2C_0,\, 4C_{\text{mod}}C_0 L)\}.$$

Then $C > 0$, and C depends on $(k, d_x, d_z, L, \pi_0, \sigma_0)$.

2. Fix $n \ge 1$. The empirical summary $\widehat{S}_n$ is a Borel sample map: its coordinates are finite sums of Borel coordinate maps, treatment-arm indicators, products, and the total empty-arm normalizers in Definition 35. Applying Proposition 4 at this n, choose the structured-lattice estimator $\widehat{\lambda}_n$. Its finite list contains each well-formed candidate once, in lexicographic order, and the returned law is

$$\widehat{\lambda}_n(\omega) = \lambda_{\widehat{\vartheta}_n(\omega)},$$

where $\widehat{\vartheta}_n(\omega)$ is the first lexicographic minimizer of $J_n(\vartheta; \widehat{S}_n(\omega))$ over the structured lattice in Algorithm 2, with threshold $\pi_0\sigma_0^2/2$. The same result gives Borel measurability of $\widehat{\lambda}_n$ and, for every $P \in \mathcal{M}$ and every $0 < \zeta < 1/2$,

$$Q_P^{(n)}\left\{W_1(\widehat{\lambda}_n, \nu_P) > C_{\mathrm{tail}}\sqrt{\frac{\log(C_{\mathrm{tail}}/\zeta)}{n}}\right\} \le \zeta.$$

3. Proposition 2 gives compactness of $\mathcal{K}$. If $\mathcal{K} = \varnothing$, take the selector

$$\Pi(s) = 0$$

for every ambient summary s, and define the repaired estimator by the zero-law branch in Algorithm 1. If $\mathcal{K} \neq \varnothing$, identify the finite-dimensional summary space with Euclidean space. The loss $(s, q) \mapsto d_S(q, s)$ is continuous by Lemma 23, so Lemma 2 gives a Borel map Π such that

$$\Pi(s) \in \mathcal{K}, \qquad d_S(\Pi(s), s) \le d_S(q, s) \quad \text{for all } q \in \mathcal{K}.$$

With this selector, set

$$\widehat{\nu}_n(\omega) = \begin{cases} \delta_0, & \mathcal{K} = \varnothing, \\ \overline{F}\{\Pi(\widehat{S}_n(\omega))\}, & \mathcal{K} \neq \varnothing. \end{cases}$$

Continuity of $\overline{F}$, Borel measurability of Π, and Borel measurability of $\widehat{S}_n$ make $\widehat{\nu}_n$ a total Borel estimator into $\mathcal{P}_k([-L_\tau, L_\tau])$.

4. Fix $P \in \mathcal{M}$. Then $S(P) \in \mathscr{S} \subseteq \mathcal{K}$, so $\mathcal{K} \neq \varnothing$. For every sample ω, nearest-summary optimality gives

$$d_S\{\Pi(\widehat{S}_n(\omega)), \widehat{S}_n(\omega)\} \le d_S\{S(P), \widehat{S}_n(\omega)\} = d_S\{\widehat{S}_n(\omega), S(P)\}.$$

By the triangle inequality for d_S,

$$\begin{aligned} d_S\{\Pi(\widehat{S}_n(\omega)), S(P)\} &\le d_S\{\Pi(\widehat{S}_n(\omega)), \widehat{S}_n(\omega)\} + d_S\{\widehat{S}_n(\omega), S(P)\} \\ &\le 2 d_S\{\widehat{S}_n(\omega), S(P)\}. \end{aligned}$$

Since $\overline{F}(S(P)) = \nu_P$, the Lipschitz extension bound gives

$$\begin{aligned} W_1(\widehat{\nu}_n(\omega), \nu_P) &\le C_{\mathrm{mod}}\, d_S\{\Pi(\widehat{S}_n(\omega)), S(P)\} \\ &\le 2C_{\mathrm{mod}}\, d_S\{\widehat{S}_n(\omega), S(P)\}. \end{aligned}$$

5. Fix $0 < \eta < 1/2$, and define

$$\Theta_\lambda = C_{\mathrm{tail}}\sqrt{\frac{\log(C_{\mathrm{tail}}/(\eta/2))}{n}}, \qquad \Theta_S = C_0 L\sqrt{\frac{\log(C_0/(\eta/2))}{n}},$$

$$\Theta_C = C\sqrt{\frac{\log(C/\eta)}{n}}.$$

The definition of C implies

$$2C_{\mathrm{tail}} \le C, \qquad 2C_0 \le C, \qquad 4C_{\mathrm{mod}} C_0 L \le C.$$

Hence
$$\frac{C_{\mathrm{tail}}}{\eta/2} = \frac{2C_{\mathrm{tail}}}{\eta} \le \frac{C}{\eta}, \qquad \frac{C_0}{\eta/2} = \frac{2C_0}{\eta} \le \frac{C}{\eta}.$$
Monotonicity of log and of the square root gives
$$\Theta_\lambda \le \Theta_C$$
and
$$2C_{\mathrm{mod}}\Theta_S \le \Theta_C.$$
Therefore
$$\begin{aligned}&\left\{\max\left[W_1(\widehat{\lambda}_n, \nu_P), W_1(\widehat{\nu}_n, \nu_P)\right] > \Theta_C\right\}\\ &\quad\subseteq \{W_1(\widehat{\lambda}_n, \nu_P) > \Theta_\lambda\} \cup \{d_S(\widehat{S}_n, S(P)) > \Theta_S\}.\end{aligned}$$
Indeed, on the complement of the right-hand union, the lattice error is at most $\Theta_\lambda \le \Theta_C$, while the repaired-summary error is at most $2C_{\mathrm{mod}}\Theta_S \le \Theta_C$.

6. Apply the lattice tail bound with $\zeta = \eta/2$ and Lemma 1 with $\zeta = \eta/2$. The displayed set inclusion and the union bound yield
$$\begin{aligned}&Q_P^{(n)}\left\{\max\left[W_1(\widehat{\lambda}_n, \nu_P), W_1(\widehat{\nu}_n, \nu_P)\right] > C\sqrt{\frac{\log(C/\eta)}{n}}\right\}\\ &\le Q_P^{(n)}\{W_1(\widehat{\lambda}_n, \nu_P) > \Theta_\lambda\} + Q_P^{(n)}\{d_S(\widehat{S}_n, S(P)) > \Theta_S\}\\ &\le \eta/2 + \eta/2 = \eta.\end{aligned}$$
The arm-count conclusion is also supplied by Lemma 1: for each $t \in \{0,1\}$,
$$Q_P^{(n)}\{N_{t,n} = 0\} \le (1 - k\pi_0)^n.$$
□

Proof of Theorem 3. 1. By Lemma 1, choose $C_0 \ge 1$ such that, for every $L \ge 1$, $n \ge 1$, $P \in \mathcal{M}$, and $0 < \alpha < 1/2$,
$$Q_P^{(n)}\left\{d_S(\widehat{S}_n, S(P)) > C_0 L\sqrt{\frac{\log(C_0/\alpha)}{n}}\right\} \le \alpha.$$
Fix $L \ge 1$, and put
$$L_\tau = \frac{4L\sqrt{d_z}}{\sigma_0}.$$
Let
$$\mathcal{K} = \overline{\{S(P) : P \in \mathcal{M}\}}$$
be the summary closure from Definition 43. By Theorem 1, choose $C_{\mathrm{mod}} > 0$ and a continuous extension
$$\overline{F} : \mathcal{K} \to \mathcal{P}_k([-L_\tau, L_\tau])$$
such that
$$W_1\{\overline{F}(\chi), \overline{F}(\chi')\} \le C_{\mathrm{mod}} d_S(\chi, \chi') \qquad (\chi, \chi' \in \mathcal{K}),$$

and
$$\overline{F}(S(P)) = \nu_P \qquad (P \in \mathcal{M}).$$

By Proposition 4, the prescribed structured-lattice constant C_{lat} is positive. For each $n \geq 1$, Proposition 4 supplies a measurable structured-lattice estimator $\widehat{\lambda}_n$, generated by the finite exhaustive duplicate-free lexicographic list of well-formed lattice candidates and by the first minimizer of $J_n(\vartheta; \widehat{S}_n)$, such that, for every sample $\mathsf{o} \in O^n_{d_x,d_z}$ and every $P \in \mathcal{M}$,
$$W_1\big(\widehat{\lambda}_n(\mathsf{o}), \nu_P\big) \leq C_{\mathrm{lat}} \left\{ d_S(\widehat{S}_n(\mathsf{o}), S(P)) + (\sqrt{n})^{-1} \right\}.$$

The same cited result gives, for every sample,
$$\mathrm{AtomFloor}\big(\pi_0, \widehat{\lambda}_n(\mathsf{o})\big).$$

For $0 < \alpha < 1/2$, define
$$r_{n,\alpha} = C_0 L \sqrt{\frac{\log(C_0/\alpha)}{n}}, \qquad R_{n,\alpha} = C_{\mathrm{lat}} \left(r_{n,\alpha} + (\sqrt{n})^{-1} \right).$$

The parameter domain gives $r_{n,\alpha} > 0$, and $C_{\mathrm{lat}} > 0$ gives $R_{n,\alpha} > 0$.

2. Construct the repair selector Π. By Proposition 2, $\mathcal{K}$ is compact. If $\mathcal{K} = \varnothing$, set
$$\Pi(s) = 0 \qquad \text{for every ambient summary } s.$$

If $\mathcal{K} \neq \varnothing$, then Lemma 23 and Lemma 2 give a Borel map Π such that, for every ambient summary s,
$$\Pi(s) \in \mathcal{K}, \qquad d_S(s, \Pi(s)) \leq d_S(s, q) \quad (q \in \mathcal{K}).$$

Together with $\overline{F}$, this is the repair datum of Algorithm 1; the induced repaired summary-to-law map is measurable, and the displayed Lipschitz bound for $\overline{F}$ is retained on $\mathcal{K}$.

3. Fix a sample $\mathsf{o} \in O^n_{d_x,d_z}$. If $\mathcal{K} = \varnothing$, then $\mathcal{C}_{n,\alpha}(\mathsf{o}) = \{\delta_0\}$, hence $\mathcal{C}_{n,\alpha}(\mathsf{o})$ is nonempty. If $\mathcal{K} \neq \varnothing$, then $\Pi(\widehat{S}_n(\mathsf{o})) \in \mathcal{K}$, and
$$d_S\{\Pi(\widehat{S}_n(\mathsf{o})), \Pi(\widehat{S}_n(\mathsf{o}))\} = 0 \leq 2 r_{n,\alpha}.$$

Thus
$$\overline{F}\{\Pi(\widehat{S}_n(\mathsf{o}))\} \in \mathcal{C}_{n,\alpha}(\mathsf{o}),$$

so $\mathcal{C}_{n,\alpha}(\mathsf{o})$ is nonempty. The algorithmic set is nonempty because
$$\mathrm{AtomFloor}\big(\pi_0, \widehat{\lambda}_n(\mathsf{o})\big), \qquad W_1\{\widehat{\lambda}_n(\mathsf{o}), \widehat{\lambda}_n(\mathsf{o})\} = 0 \leq R_{n,\alpha}.$$

By Algorithm 3, $m_\star = \pi_0$ and membership in $\mathcal{C}^{\mathrm{alg}}_{n,\alpha}$ is membership in the atom-floor subclass together with
$$W_1(\nu, \widehat{\lambda}_n(\mathsf{o})) \leq R_{n,\alpha}.$$

By Definition 19, this last inequality is equivalent to the existence of a finite nonnegative transport plan from ν to $\widehat{\lambda}_n(\mathsf{o})$ whose absolute-distance cost is at most $R_{n,\alpha}$. This gives the asserted finite constrained transport representation.

4. Fix $P \in \mathcal{M}$. Since $S(P) \in \mathcal{K}$, the closure $\mathcal{K}$ is nonempty. Define the bad event

$$\mathsf{Bad}_P = \left\{ \mathsf{o} \in O^n_{d_x,d_z} : r_{n,\alpha} < d_S(\widehat{S}_n(\mathsf{o}), S(P)) \right\}.$$

   The concentration bound gives

$$Q_P^{(n)}(\mathsf{Bad}_P) \le \alpha.$$

   The empirical summary is Borel by Definition 35, and $(s, q) \mapsto d_S(s, q)$ is continuous by Lemma 23; hence Bad_P is measurable. Therefore

$$Q_P^{(n)}(\mathsf{Bad}_P^c) = 1 - Q_P^{(n)}(\mathsf{Bad}_P) \ge 1 - \alpha.$$

   For every $\mathsf{o} \in \mathsf{Bad}_P^c$,

$$d_S(\widehat{S}_n(\mathsf{o}), S(P)) \le r_{n,\alpha}.$$

   Nearest-summary optimality gives

$$d_S\{\widehat{S}_n(\mathsf{o}), \Pi(\widehat{S}_n(\mathsf{o}))\} \le d_S\{\widehat{S}_n(\mathsf{o}), S(P)\} \le r_{n,\alpha},$$

   and the triangle inequality gives

$$\begin{aligned} d_S\{S(P), \Pi(\widehat{S}_n(\mathsf{o}))\} &\le d_S\{S(P), \widehat{S}_n(\mathsf{o})\} + d_S\{\widehat{S}_n(\mathsf{o}), \Pi(\widehat{S}_n(\mathsf{o}))\} \\ &\le 2r_{n,\alpha}. \end{aligned}$$

   Since $\overline{F}(S(P)) = \nu_P$, the summary $S(P)$ itself witnesses

$$\nu_P \in \mathcal{C}_{n,\alpha}(\mathsf{o}).$$

5. On the same sample $\mathsf{o} \in \mathsf{Bad}_P^c$, Proposition 4 gives

$$W_1\{\widehat{\lambda}_n(\mathsf{o}), \nu_P\} \le C_{\text{lat}}\{r_{n,\alpha} + (\sqrt{n})^{-1}\} = R_{n,\alpha}.$$

   By symmetry of W_1,

$$W_1\{\nu_P, \widehat{\lambda}_n(\mathsf{o})\} \le R_{n,\alpha}.$$

   It remains to verify the atom floor for ν_P. For each latent class u, Assumption 9 as included in Definition 32 gives

$$P(U = u, T = 0) \ge \pi_0, \qquad P(U = u, T = 1) \ge \pi_0.$$

   The two treatment cells partition the latent class, so

$$p_u = P(U = u) = P(U = u, T = 0) + P(U = u, T = 1) \ge 2\pi_0 \ge \pi_0,$$

   where p_u is the latent mass from Definition 21. For the raw latent law

$$\sum_{u=1}^{k} p_u \delta_{\tau_u},$$

   the aggregate mass at any distinct support value ζ is

$$\sum_{u:\tau_u=\zeta} p_u \ge \pi_0,$$

because the displayed sum contains at least one latent class. Passing to the quotient law ν_P of Definition 40 only aggregates equal atom locations, so Definition 31 gives

$$\operatorname{AtomFloor}(\pi_0, \nu_P).$$

Thus

$$\nu_P \in \mathcal{C}^{\mathrm{alg}}_{n,\alpha}(\mathsf{o}).$$

Combining the two memberships on Bad_P^c,

$$1-\alpha \le Q_P^{(n)}\left\{\nu_P \in \mathcal{C}_{n,\alpha} \cap \mathcal{C}^{\mathrm{alg}}_{n,\alpha}\right\}.$$

6. Fix a sample o. For the theoretical diameter, take any

$$\xi_1, \xi_2 \in \mathcal{C}_{n,\alpha}(\mathsf{o}).$$

If $\mathcal{K} = \varnothing$, then both elements equal the zero-law fallback, so

$$W_1(\xi_1, \xi_2) = 0 \le 4C_{\mathrm{mod}} r_{n,\alpha}.$$

If $\mathcal{K} \neq \varnothing$, then by Algorithm 3 there are $\chi_1, \chi_2 \in \mathcal{K}$ such that

$$\xi_j = \overline{F}(\chi_j), \qquad d_S\{\chi_j, \Pi(\widehat{S}_n(\mathsf{o}))\} \le 2r_{n,\alpha} \quad (j=1,2).$$

By symmetry and the triangle inequality,

$$\begin{aligned} d_S(\chi_1, \chi_2) &\le d_S\{\chi_1, \Pi(\widehat{S}_n(\mathsf{o}))\} + d_S\{\Pi(\widehat{S}_n(\mathsf{o})), \chi_2\} \\ &\le 4r_{n,\alpha}. \end{aligned}$$

The Lipschitz bound for $\overline{F}$ gives

$$W_1(\xi_1, \xi_2) \le C_{\mathrm{mod}} d_S(\chi_1, \chi_2) \le 4C_{\mathrm{mod}} r_{n,\alpha}.$$

Taking the supremum over all pairs in the nonempty set $\mathcal{C}_{n,\alpha}(\mathsf{o})$ yields

$$\operatorname{diam}_{W_1}(\mathcal{C}_{n,\alpha}) \le 4C_{\mathrm{mod}} r_{n,\alpha}.$$

7. For the algorithmic diameter, take any

$$\xi_1, \xi_2 \in \mathcal{C}^{\mathrm{alg}}_{n,\alpha}(\mathsf{o}).$$

By Algorithm 3,

$$W_1\{\xi_1, \widehat{\lambda}_n(\mathsf{o})\} \le R_{n,\alpha}, \qquad W_1\{\xi_2, \widehat{\lambda}_n(\mathsf{o})\} \le R_{n,\alpha}.$$

By symmetry and the triangle inequality,

$$\begin{aligned} W_1(\xi_1, \xi_2) &\le W_1\{\xi_1, \widehat{\lambda}_n(\mathsf{o})\} + W_1\{\widehat{\lambda}_n(\mathsf{o}), \xi_2\} \\ &\le 2R_{n,\alpha}. \end{aligned}$$

Taking the supremum over all pairs in the nonempty set $\mathcal{C}^{\mathrm{alg}}_{n,\alpha}(\mathsf{o})$ gives

$$\operatorname{diam}_{W_1}(\mathcal{C}^{\mathrm{alg}}_{n,\alpha}) \le 2R_{n,\alpha}.$$

□

Proof of Theorem 4. 1. By Lemma 1, choose C_0 with $C_0 \geq 1$ such that, for every $L \geq 1$, $n \geq 1$, $P \in \mathcal{M}$, and $0 < \eta < 1/2$,

$$Q_P^{(n)} \left\{ d_S(\widehat{S}_n, S(P)) > C_0 L \sqrt{\frac{\log(C_0/\eta)}{n}} \right\} \leq \eta.$$

Fix $L \geq 1$. With the displayed definitions of $L_\tau, s_0, c_V, K_D, A_D, K_f, c_D, c_m, c_b, c_{\text{grid}}, B_{\text{lat}}$, Proposition 4 gives

$$C_{\text{lat}} > 0$$

and, for every $n \geq 1$, a Borel structured-lattice estimator $\widehat{\lambda}_n$ obtained from the exhaustive duplicate-free lexicographic list of well-formed candidates and the first minimizer of $J_n(\vartheta; \widehat{S}_n)$. It also gives, for every sample ω,

$$\text{AtomFloor}(\pi_0, \widehat{\lambda}_n(\omega)),$$

and, for every $P \in \mathcal{M}$,

$$W_1(\widehat{\lambda}_n(\omega), \nu_P) \leq C_{\text{lat}} \left\{ d_S(\widehat{S}_n(\omega), S(P)) + \frac{1}{\sqrt{n}} \right\}.$$

The extension $\overline{F}$ on $\mathcal{K}$ is supplied by Theorem 1. By Proposition 2, $\mathcal{K}$ is compact, and by Lemma 23, $(s, q) \mapsto d_S(s, q)$ is continuous. If $\mathcal{K} \neq \emptyset$, Lemma 2 supplies a Borel selector Π satisfying

$$\Pi(s) \in \mathcal{K}, \qquad d_S(\Pi(s), s) \leq d_S(q, s) \quad (q \in \mathcal{K}).$$

If $\mathcal{K} = \emptyset$, set $\Pi(s) = 0$ for every ambient summary s. Together with $\overline{F}$, these maps are repair data of the form specified in Algorithm 1; the induced sample-to-law map is Borel measurable by composition in the nonempty case and by the constant fallback in the empty case.

2. Fix $0 < \alpha < 1/2$, $P \in \mathcal{M}$, and $n \geq 1$. Define

$$r_{n,\alpha} = C_0 L \sqrt{\frac{\log(C_0/\alpha)}{n}}, \qquad R_{n,\alpha} = C_{\text{lat}} \left(r_{n,\alpha} + \frac{1}{\sqrt{n}} \right), \qquad \rho_{n,\alpha} = \frac{R_{n,\alpha}}{\pi_0}.$$

Since $C_{\text{lat}} > 0$, $C_0 \geq 1$, $L \geq 1$, $n \geq 1$, and $0 < \alpha < 1/2$, we have

$$R_{n,\alpha} > 0, \qquad \rho_{n,\alpha} > 0.$$

The event

$$E_P = \left\{ \omega : d_S(\widehat{S}_n(\omega), S(P)) \leq r_{n,\alpha} \right\}$$

is the complement of the deviation event in the first display with $\eta = \alpha$, and hence

$$Q_P^{(n)}(E_P) \geq 1 - \alpha.$$

For every $\omega \in E_P$, the deterministic lattice bound gives

$$W_1(\widehat{\lambda}_n(\omega), \nu_P) \leq C_{\text{lat}} \left(r_{n,\alpha} + \frac{1}{\sqrt{n}} \right) = R_{n,\alpha}.$$

3. The representative law ν_P has atom floor π_0. Indeed, for each latent class u,

$$p_u(P) = P(U = u, T = 0) + P(U = u, T = 1) \geq \pi_0 + \pi_0 \geq \pi_0$$

by the latent-arm positivity component of $P \in \mathcal{M}$. After coincident latent effects are aggregated as in Definition 40, every distinct support atom has aggregate mass at least π_0. Since W_1 is symmetric on represented atomic laws, the preceding step gives

$$W_1(\nu_P, \widehat{\lambda}_n(\omega)) \leq R_{n,\alpha}.$$

Thus

$$\nu_P \in \mathcal{C}^{\mathrm{alg}}_{n,\alpha},$$

because Algorithm 3 defines

$$\mathcal{C}^{\mathrm{alg}}_{n,\alpha} = \left\{ \xi : \mathrm{AtomFloor}(\pi_0, \xi) \text{ and } W_1(\xi, \widehat{\lambda}_n(\omega)) \leq R_{n,\alpha} \right\}.$$

4. We shall use the following finite-transport implication. If two finite atomic probability laws χ and χ' have atom floor $m > 0$ and

$$W_1(\chi, \chi') \leq m\rho,$$

then

$$\mathrm{dist}(x, \operatorname{supp} \chi') \leq \rho \quad (x \in \operatorname{supp} \chi), \qquad \mathrm{dist}(y, \operatorname{supp} \chi) \leq \rho \quad (y \in \operatorname{supp} \chi').$$

For the first assertion, put

$$d_\chi(x) := \mathrm{dist}(x, \operatorname{supp} \chi').$$

If $d_\chi(x) > \rho$, then every coupling in Definition 19 moves the aggregate mass at x, which is at least m, by distance at least $d_\chi(x)$. Hence every transport cost is at least $md_\chi(x) > m\rho$, contradicting the displayed Wasserstein bound. The second assertion follows by applying the same argument after transposing a coupling. Applying this implication with

$$m = \pi_0, \qquad \chi = \widehat{\lambda}_n(\omega), \qquad \chi' = \nu_P, \qquad \rho = \rho_{n,\alpha},$$

and using $\pi_0 \rho_{n,\alpha} = R_{n,\alpha}$, gives mutual $\rho_{n,\alpha}$-closeness of the two supports.

5. The components $\widehat{\mathscr{K}_{n,\alpha}}$ are the connected components of $\mathrm{supp}(\widehat{\lambda}_n(\omega))$ under links of length at most $4\rho_{n,\alpha}$. For such a component C,

$$K_C(P) = \{x \in \mathrm{supp}(\nu_P) : \mathrm{dist}(x, C) \leq \rho_{n,\alpha}\}.$$

The mutual support closeness from the previous step implies that every $x \in \mathrm{supp}(\nu_P)$ belongs to at least one such set. If x belonged to two components, choose center atoms in the two components within $\rho_{n,\alpha}$ of x. These two center atoms are within $2\rho_{n,\alpha} \leq 4\rho_{n,\alpha}$ of each other, so they lie in the same connected component. Hence

$$x \in \mathrm{supp}(\nu_P) \quad \Longleftrightarrow \quad \exists!\, C \in \widehat{\mathscr{K}_{n,\alpha}} \text{ with } x \in K_C(P).$$

For every component C, reverse support closeness supplies a true atom within $\rho_{n,\alpha}$ of any center atom in C, so $K_C(P) \neq \emptyset$. The inclusion

$$K_C(P) \subseteq \operatorname{supp}(\nu_P)$$

is part of the displayed definition of $K_C(P)$. If $x \in K_C(P)$, then some $y \in C$ satisfies $|x - y| \leq \rho_{n,\alpha}$, and therefore

$$\min C - \rho_{n,\alpha} \leq x \leq \max C + \rho_{n,\alpha},$$

which is the reported support interval in Algorithm 4. Since $\nu_P \in \mathcal{C}^{\mathrm{alg}}_{n,\alpha}$,

$$\nu_P(K_C(P)) \in \left\{ v : \exists \xi \in \mathcal{C}^{\mathrm{alg}}_{n,\alpha},\ v = \xi(K_C(\xi)) \right\}.$$

Consequently

$$\inf_{\xi \in \mathcal{C}^{\mathrm{alg}}_{n,\alpha}} \xi(K_C(\xi)) \leq \nu_P(K_C(P)) \leq \sup_{\xi \in \mathcal{C}^{\mathrm{alg}}_{n,\alpha}} \xi(K_C(\xi)),$$

which proves the mass-interval containment.

6. Fix $\xi \in \mathcal{C}^{\mathrm{alg}}_{n,\alpha}$. Then

$$\mathrm{AtomFloor}(\pi_0, \xi), \qquad W_1(\xi, \widehat{\lambda}_n(\omega)) \leq R_{n,\alpha},$$

and the triangle inequality gives

$$W_1(\xi, \nu_P) \leq W_1(\xi, \widehat{\lambda}_n(\omega)) + W_1(\widehat{\lambda}_n(\omega), \nu_P) \leq 2R_{n,\alpha}.$$

The transport implication applied to ξ and $\widehat{\lambda}_n(\omega)$ shows that their supports are mutually $\rho_{n,\alpha}$-close. For a component C, write

$$K^{\xi}_C := K_C(\xi), \qquad K^P_C := K_C(P), \qquad \Delta^{\mathrm{r}}_C := (\Delta_C(P))_{\mathrm{real}}$$

when $\Delta_C(P) < +\infty$, where Δ^{r}_C is the finite real external gap.

If $\Delta_C(P) = +\infty$, then

$$\operatorname{supp}(\nu_P) \setminus K^P_C = \emptyset.$$

The partition just proved makes C the unique empirical component associated with the true support. The same partition argument for ξ gives

$$\operatorname{supp}(\xi) \subseteq K^{\xi}_C.$$

Thus

$$\xi(K^{\xi}_C) = 1 = \nu_P(K^P_C).$$

Now suppose $\Delta_C(P) < +\infty$. The finite sets K^P_C and $\operatorname{supp}(\nu_P) \setminus K^P_C$ are nonempty in this case, so

$$\Delta^{\mathrm{r}}_C > 0.$$

Let γ be any transport plan from ξ to ν_P. If

$$\mathscr{A}^{\xi}_C := \{i : x^{\xi}_i \in K^{\xi}_C\}, \qquad \mathscr{A}^P_C := \{j : x^P_j \in K^P_C\},$$

then marginal balance gives

$$\xi(K_C^\xi) - \nu_P(K_C^P) = \sum_{i\in\mathscr{A}_C^\xi}\sum_{j\notin\mathscr{A}_C^P}\gamma_{ij} - \sum_{i\notin\mathscr{A}_C^\xi}\sum_{j\in\mathscr{A}_C^P}\gamma_{ij}.$$

Every positive-mass crossing pair in these two sums is separated by at least $\Delta_C^{\mathrm{r}} - 2\rho_{n,\alpha}$. For the first crossing case $i \in \mathscr{A}_C^\xi$, $j \notin \mathscr{A}_C^P$, choose $z \in C$ with

$$|x_i^\xi - z| \le \rho_{n,\alpha}.$$

Since z is a center atom, support closeness from the center to ν_P gives $y \in \mathrm{supp}(\nu_P)$ with

$$|z - y| \le \rho_{n,\alpha}.$$

Then $y \in K_C^P$. Because $x_j^P \notin K_C^P$, the definition of $\Delta_C(P)$ gives

$$\Delta_C^{\mathrm{r}} \le |y - x_j^P|.$$

Also

$$|y - x_i^\xi| \le |y - z| + |z - x_i^\xi| \le 2\rho_{n,\alpha},$$

and therefore

$$\Delta_C^{\mathrm{r}} - 2\rho_{n,\alpha} \le |x_i^\xi - x_j^P|.$$

For the second crossing case $i \notin \mathscr{A}_C^\xi$, $j \in \mathscr{A}_C^P$, support closeness from ξ to the center gives a center atom z with

$$|x_i^\xi - z| \le \rho_{n,\alpha}.$$

Let C_z be the empirical component containing z. Support closeness from the center to ν_P gives $y \in \mathrm{supp}(\nu_P)$ with

$$|z - y| \le \rho_{n,\alpha},$$

so $y \in K_{C_z}(P)$. If $y \in K_C(P)$, the uniqueness of the true association would force $C_z = C$; then $z \in C$, and

$$\mathrm{dist}(x_i^\xi, C) \le |x_i^\xi - z| \le \rho_{n,\alpha},$$

contradicting $i \notin \mathscr{A}_C^\xi$. Hence

$$y \notin K_C(P).$$

Since $x_j^P \in K_C(P)$, the external-gap definition yields

$$\Delta_C^{\mathrm{r}} \le |x_j^P - y|.$$

Together with

$$|y - x_i^\xi| \le |y - z| + |z - x_i^\xi| \le 2\rho_{n,\alpha},$$

this gives

$$\Delta_C^{\mathrm{r}} - 2\rho_{n,\alpha} \le |x_i^\xi - x_j^P|.$$

The preceding crossing estimate and the displayed marginal identity imply

$$(\Delta_C^{\mathrm{r}} - 2\rho_{n,\alpha}) \left|\xi(K_C^\xi) - \nu_P(K_C^P)\right| \le \mathrm{cost}(\gamma).$$

Taking an optimal plan and using $W_1(\xi, \nu_P) \le 2R_{n,\alpha}$, if $4\rho_{n,\alpha} < \Delta^{\mathrm{r}}_C$, then

$$\left|\xi(K^\xi_C) - \nu_P(K^P_C)\right| \le \frac{2R_{n,\alpha}}{\Delta^{\mathrm{r}}_C - 2\rho_{n,\alpha}} \le \frac{4R_{n,\alpha}}{\Delta^{\mathrm{r}}_C}.$$

Since $4\rho_{n,\alpha} = 4R_{n,\alpha}/\pi_0 < \Delta^{\mathrm{r}}_C$ and $\pi_0 \le 1$, this is bounded by

$$\frac{4}{\pi_0} \min\left\{1, \frac{R_{n,\alpha}}{\Delta^{\mathrm{r}}_C}\right\}.$$

If $4\rho_{n,\alpha} \ge \Delta^{\mathrm{r}}_C$, then both cluster masses lie in $[0, 1]$, so their difference has absolute value at most 1. When $R_{n,\alpha}/\Delta^{\mathrm{r}}_C \ge 1$, this is at most $4/\pi_0$. When $R_{n,\alpha}/\Delta^{\mathrm{r}}_C < 1$, the inequality $\Delta^{\mathrm{r}}_C \le 4R_{n,\alpha}/\pi_0$ gives

$$\frac{\pi_0}{4} \le \frac{R_{n,\alpha}}{\Delta^{\mathrm{r}}_C},$$

and hence

$$1 \le \frac{4}{\pi_0} \frac{R_{n,\alpha}}{\Delta^{\mathrm{r}}_C}.$$

Thus, in all finite-gap cases,

$$|\xi(K_C(\xi)) - \nu_P(K_C(P))| \le \frac{4}{\pi_0} \min\left\{1, \frac{R_{n,\alpha}}{\Delta_C(P)}\right\},$$

with the top-gap case giving equality of the two masses.

7. For a fixed component C, let

$$\mathrm{Val}(C) = \left\{v : \exists \xi \in \mathcal{C}^{\mathrm{alg}}_{n,\alpha},\ v = \xi(K_C(\xi))\right\}.$$

The preceding step bounds every $v \in \mathrm{Val}(C)$ within

$$B^{\mathrm{err}}_C := \frac{4}{\pi_0} \min\left\{1, \frac{R_{n,\alpha}}{\Delta_C(P)}\right\}$$

of $\nu_P(K_C(P))$ when $\Delta_C(P) < +\infty$, and within 0 when $\Delta_C(P) = +\infty$. Since $\nu_P(K_C(P)) \in \mathrm{Val}(C)$, the elementary extremal implication

$$\big(|v - v_0| \le B \text{ for all } v \in \mathrm{Val}(C),\ v_0 \in \mathrm{Val}(C)\big) \Longrightarrow \sup \mathrm{Val}(C) - \inf \mathrm{Val}(C) \le 2B$$

applies to the endpoints of I_C. Therefore, if $\Delta_C(P) = +\infty$, then

$$\mathrm{length}(I_C) = 0,$$

and, for every component,

$$\mathrm{length}(I_C) \le \frac{8}{\pi_0} \min\left\{1, \frac{R_{n,\alpha}}{\Delta_C(P)}\right\},$$

with the displayed right side read as zero when $\Delta_C(P) = +\infty$.

8. If $K_C(P)=\{x\}$, every center atom $z\in C$ is within $\rho_{n,\alpha}$ of a true atom y. That true atom satisfies $y\in K_C(P)$, and hence $y=x$. Thus

$$|z-x|\le \rho_{n,\alpha} \qquad (z\in C).$$

Applying this to the minimum and maximum points of the finite component C gives

$$\max C-\min C\le 2\rho_{n,\alpha}.$$

The reported support interval therefore has length

$$(\max C+\rho_{n,\alpha})-(\min C-\rho_{n,\alpha})\le 4\rho_{n,\alpha}.$$

9. In the same singleton case, the external gap dominates the global effect gap. If $\Delta_C(P)<+\infty$, then for every $y\in\operatorname{supp}(\nu_P)\setminus K_C(P)$,

$$\delta(P)\le |x-y|,$$

because x and y are distinct positive-mass support points of ν_P. Taking the infimum over such y gives

$$\delta(P)\le \Delta_C(P).$$

If $\Delta_C(P)=+\infty$, the preceding zero-width conclusion applies. Since $R_{n,\alpha}\ge 0$ and $\pi_0>0$, substituting the finite-gap comparison into the external-gap interval bound gives

$$\operatorname{length}(I_C)\le \frac{8}{\pi_0}\min\left\{1,\frac{R_{n,\alpha}}{\delta(P)}\right\}.$$

10. Finally, Definition 19 defines W_1 as the minimum finite transport cost over couplings with the prescribed marginals. Therefore, for every represented atomic law ξ,

$$\xi\in\mathcal{C}^{\mathrm{alg}}_{n,\alpha} \quad\Longleftrightarrow\quad \mathrm{AtomFloor}(\pi_0,\xi)\text{ and }\exists\gamma\text{ with transport cost at most }R_{n,\alpha}\text{ from }\xi\text{ to }\widehat{\lambda}_n(\omega).$$

The forward implication takes an optimal finite transport plan attaining W_1; the reverse implication uses that any feasible plan upper-bounds the minimum. This is the finite constrained representation of $\mathcal{C}^{\mathrm{alg}}_{n,\alpha}$.

□

Proof of Theorem 5. 1. Let $C_{\mathrm{quo}}>0$ be the constant supplied by Theorem 2. Define

$$\chi_{\mathrm{tail}}:=\max\{C_{\mathrm{quo}},2\},\qquad \beta_{\mathrm{mean}}:=\chi_{\mathrm{tail}}\sqrt{\log(2\chi_{\mathrm{tail}})}+\frac{\chi_{\mathrm{tail}}^2\sqrt{\pi}}{2},$$

$$\kappa_{\mathrm{ord}}:=8+\frac{16}{\pi_0},\qquad C:=\max\{2,\kappa_{\mathrm{ord}}\beta_{\mathrm{mean}}\}.$$

The hypotheses give $\chi_{\mathrm{tail}}\ge 2$, $\log(2\chi_{\mathrm{tail}})>0$, $\beta_{\mathrm{mean}}>0$, and $\kappa_{\mathrm{ord}}>0$, hence $C>0$. Fix $n\ge 1$. By Theorem 2, choose the measurable nearest-summary quotient-law estimator $\widehat{\nu}_n$ from Algorithm 1. Set

$$\widehat{p}_n^{\uparrow}(\omega):=p^{\uparrow}(\widehat{\nu}_n(\omega)),\qquad \omega\in O^n_{d_x,d_z},$$

which is exactly the ordered-weight estimator of Algorithm 5. Its measurability follows from the measurability of $\widehat{\nu}_n$ and the Borel ordered-mass map; its values lie in the probability simplex by the ordered-mass construction, including its barycentric fallback branch.

2. Fix g with $0 < g \le 1/4$, and fix $P \in \mathcal{M}(g)$. Write

$$Z_P(\omega) := W_1(\widehat{\nu}_n(\omega), \nu_P), \qquad \omega \in O^n_{d_x,d_z},$$

where ν_P is the quotient latent-effect law of Definition 40. The support radius $L_\tau = 4L\sqrt{d_z}/\sigma_0$ gives

$$0 \le Z_P(\omega) \le 2L_\tau$$

for every ω, so Z_P is integrable under $Q_P^{(n)}$. For every $0 < \eta < 1/2$, the tail bound in Theorem 2 and the inequality $C_{\mathrm{quo}} \le \chi_{\mathrm{tail}}$ give

$$Q_P^{(n)}\left\{\chi_{\mathrm{tail}}\sqrt{\frac{\log(\chi_{\mathrm{tail}}/\eta)}{n}} < Z_P\right\} \le \eta.$$

Indeed, the displayed event is contained in the corresponding collision-uniform event because

$$C_{\mathrm{quo}}\sqrt{\frac{\log(C_{\mathrm{quo}}/\eta)}{n}} \le \chi_{\mathrm{tail}}\sqrt{\frac{\log(\chi_{\mathrm{tail}}/\eta)}{n}}.$$

To integrate this tail, put

$$b_{\mathrm{mean}} := \chi_{\mathrm{tail}}\sqrt{\frac{\log(2\chi_{\mathrm{tail}})}{n}}, \qquad \varrho_{\mathrm{mean}} := \frac{n}{\chi^2_{\mathrm{tail}}}.$$

For $x > b_{\mathrm{mean}}$, define

$$\eta_x := \chi_{\mathrm{tail}} \exp(-\varrho_{\mathrm{mean}} x^2).$$

Then $0 < \eta_x < 1/2$ and

$$\chi_{\mathrm{tail}}\sqrt{\frac{\log(\chi_{\mathrm{tail}}/\eta_x)}{n}} = x.$$

Thus

$$Q_P^{(n)}\{Z_P > x\} \le \chi_{\mathrm{tail}} \exp(-\varrho_{\mathrm{mean}} x^2), \qquad x > b_{\mathrm{mean}}.$$

The layer-cake identity and the Gaussian tail bound yield

$$E_{Q_P^{(n)}}[Z_P] \le b_{\mathrm{mean}} + \chi_{\mathrm{tail}}\int_{b_{\mathrm{mean}}}^{\infty} \exp(-\varrho_{\mathrm{mean}} x^2)\, dx \le \frac{\beta_{\mathrm{mean}}}{\sqrt{n}}.$$

3. Let

$$\nu_P^{\mathrm{slot}} := \left((p_u(P))_{u=1}^k, (\tau_u(P))_{u=1}^k\right)$$

be the k-slot representative of ν_P. By Lemma 17,

$$p^{\uparrow}(P) = p^{\uparrow}\left(\nu_P^{\mathrm{slot}}\right),$$

because ν_P^{slot} and ν_P represent the same aggregated atomic law. The gap-stratum assumptions give, for every latent class u,

$$\pi_0 \le P(U = u, T = 0) \le P(U = u) = p_u(P),$$

and Definition 41 gives distinct latent effects and

$$\frac{g}{2} \le |\tau_u(P) - \tau_v(P)| \qquad (u \neq v).$$

Consequently, if $\xi = \sum_{a=1}^{k} \widetilde{p}_a \delta_{\widetilde{\tau}_a}$ is any valid k-slot atomic law and

$$W_1\left(\nu_P^{\mathrm{slot}}, \xi\right) < \frac{\pi_0(g/2)}{4},$$

then

$$\sum_{j=1}^{k} \left| p_j^{\uparrow}(\xi) - p_j^{\uparrow}(\nu_P^{\mathrm{slot}}) \right| \le \frac{8}{g} W_1\left(\nu_P^{\mathrm{slot}}, \xi\right).$$

For this last implication, take an optimal transport plan γ. Each true atom $\tau_u(P)$ must have a positive-mass atom $\widetilde{\tau}_{\varphi(u)}$ of ξ within distance $g/8$; otherwise the mass $p_u(P) \ge \pi_0$ transported from that atom alone would cost at least $\pi_0(g/2)/4$. The map φ is injective, hence bijective, because two $g/8$-close images of distinct $g/2$-separated true atoms would violate the triangle inequality. It also preserves the increasing order. Therefore

$$\sum_{j=1}^{k} \left| p_j^{\uparrow}(\xi) - p_j^{\uparrow}(\nu_P^{\mathrm{slot}}) \right| = \sum_{u=1}^{k} \left| \widetilde{p}_{\varphi(u)} - p_u(P) \right|.$$

With

$$\mathrm{diag}_\gamma := \sum_{u=1}^{k} \gamma_{u,\varphi(u)}, \qquad \mathrm{off}_\gamma := 1 - \mathrm{diag}_\gamma,$$

the row and column marginal identities give

$$\sum_{u=1}^{k} \left| \widetilde{p}_{\varphi(u)} - p_u(P) \right| \le 2 \, \mathrm{off}_\gamma .$$

All off-matching mass moves at least $g/4$, so

$$\frac{g}{4} \, \mathrm{off}_\gamma \le W_1\left(\nu_P^{\mathrm{slot}}, \xi\right),$$

and the displayed local stability bound follows.

4. Apply the preceding bound to the representative ξ_ω of $\widehat{\nu}_n(\omega)$. Since ξ_ω and $\widehat{\nu}_n(\omega)$ represent the same law,

$$W_1\left(\nu_P^{\mathrm{slot}}, \xi_\omega\right) = Z_P(\omega).$$

If

$$Z_P(\omega) < \frac{\pi_0(g/2)}{4},$$

then

$$\sum_{j=1}^{k} \left| \widehat{p}_{n,j}^{\uparrow}(\omega) - p_j^{\uparrow}(P) \right| \le \frac{8 Z_P(\omega)}{g} \le \frac{\kappa_{\mathrm{ord}} Z_P(\omega)}{g}.$$

If instead $Z_P(\omega) \geq \pi_0(g/2)/4$, the simplex diameter gives

$$\sum_{j=1}^{k} \left| \widehat{p}^{\uparrow}_{n,j}(\omega) - p^{\uparrow}_j(P) \right| \leq 2 \leq \frac{16}{\pi_0} \frac{Z_P(\omega)}{g} \leq \frac{\kappa_{\mathrm{ord}} Z_P(\omega)}{g}.$$

Thus, for every sample ω,

$$\sum_{j=1}^{k} \left| \widehat{p}^{\uparrow}_{n,j}(\omega) - p^{\uparrow}_j(P) \right| \leq \frac{\kappa_{\mathrm{ord}} Z_P(\omega)}{g}.$$

5. Integrating the samplewise inequality and using the mean bound for Z_P,

$$E_{Q_P^{(n)}}\left[\left\| \widehat{p}^{\uparrow}_n - p^{\uparrow}(P) \right\|_1 \right] \leq \frac{\kappa_{\mathrm{ord}}}{g} E_{Q_P^{(n)}}[Z_P] \leq \kappa_{\mathrm{ord}} \beta_{\mathrm{mean}} \, (\sqrt{n}\, g)^{-1}.$$

Independently, both $\widehat{p}^{\uparrow}_n(\omega)$ and $p^{\uparrow}(P)$ lie in the simplex, hence

$$\left\| \widehat{p}^{\uparrow}_n(\omega) - p^{\uparrow}(P) \right\|_1 \leq 2$$

for every ω, and therefore

$$E_{Q_P^{(n)}}\left[\left\| \widehat{p}^{\uparrow}_n - p^{\uparrow}(P) \right\|_1 \right] \leq 2.$$

6. If $(\sqrt{n}\, g)^{-1} \leq 1$, the preceding rate bound and $\kappa_{\mathrm{ord}} \beta_{\mathrm{mean}} \leq C$ give

$$E_{Q_P^{(n)}}\left[\left\| \widehat{p}^{\uparrow}_n - p^{\uparrow}(P) \right\|_1 \right] \leq C(\sqrt{n}\, g)^{-1} = C \min\{1, (\sqrt{n}\, g)^{-1}\}.$$

If $1 \leq (\sqrt{n}\, g)^{-1}$, the simplex-diameter bound and $2 \leq C$ give

$$E_{Q_P^{(n)}}\left[\left\| \widehat{p}^{\uparrow}_n - p^{\uparrow}(P) \right\|_1 \right] \leq C = C \min\{1, (\sqrt{n}\, g)^{-1}\}.$$

The constant C depends only on $k, d_x, d_z, L, \pi_0, \sigma_0$, and the construction above supplies the required estimator for every $n \geq 1$.

$\square$

Proof of Proposition 5. 1. Fix $0 \leq \varepsilon \leq 1/8$. For $\mathfrak{b} \in \{0, 1\}$, write

$$\mathrm{Ber}_\theta(\mathfrak{b}) = \begin{cases} \theta, & \mathfrak{b} = 1, \\ 1 - \theta, & \mathfrak{b} = 0. \end{cases}$$

The law in Definition 55 is the finite law on points

$$X = (1, \mathfrak{x})^\top, \qquad Z = (1, \mathfrak{z})^\top, \qquad Y(0) = \mathfrak{y}_0, \qquad Y(1) = \mathfrak{y}_1, \qquad Y = Y(T),$$

where $\mathfrak{t}, \mathfrak{x}, \mathfrak{z}, \mathfrak{y}_0, \mathfrak{y}_1 \in \{0, 1\}$. Its elementary mass at $(u, \mathfrak{t}, \mathfrak{x}, \mathfrak{z}, \mathfrak{y}_0, \mathfrak{y}_1)$, for $u \in \{1, 2\}$, is

$$\theta^U_u \, \mathrm{Ber}_{\theta^T_u}(\mathfrak{t}) \, \mathrm{Ber}_{\theta^X_u}(\mathfrak{x}) \, \mathrm{Ber}_{\theta^Z_{u,\mathfrak{t}}}(\mathfrak{z}) \, \mathrm{Ber}_{1/4}(\mathfrak{y}_0) \, \mathrm{Ber}_{\theta^1_u}(\mathfrak{y}_1),$$

with

$$(\theta^U_1, \theta^U_2) = \left(\frac{2}{5}, \frac{3}{5} \right), \quad (\theta^T_1, \theta^T_2) = \left(\frac{2}{5}, \frac{3}{5} \right), \quad (\theta^X_1, \theta^X_2) = \left(\frac{1}{5}, \frac{4}{5} \right),$$

$$\theta^Z_{1,0} = \frac{3}{10}, \quad \theta^Z_{2,0} = \frac{7}{10}, \quad \theta^Z_{1,1} = \frac{7}{20}, \quad \theta^Z_{2,1} = \frac{3}{4}, \quad \theta^1_1 = \frac{1}{2} - \varepsilon, \quad \theta^1_2 = \frac{1}{2} + \varepsilon.$$

The perturbation range gives

$$\frac{3}{8} \le \frac{1}{2} - \varepsilon \le \frac{1}{2}, \qquad \frac{1}{2} \le \frac{1}{2} + \varepsilon \le \frac{5}{8},$$

so every displayed Bernoulli factor is a probability mass function. Summing successively over $\mathfrak{y}_1, \mathfrak{y}_0, \mathfrak{z}, \mathfrak{x}, \mathfrak{t}$ leaves θ^U_u, and $\theta^U_1 + \theta^U_2 = 1$. Thus $P^{\mathrm{wit}}_\varepsilon$ is a probability law.

2. The parameter domain in Definition 32 is satisfied because

$$2 \le 2, \qquad 2 \le 2, \qquad 2 \le 2, \qquad 1 \le 2, \qquad 0 < \frac{1}{10} \le \frac{1}{2 \cdot 2}, \qquad 0 < \frac{1}{10} \le 1.$$

The finite product form gives the conditional factorization requirements. For bounded measurable test functions $\psi_Z, \psi_{XY}, \psi_X, \psi_{YT}, \psi_{Y_t}, \psi_T$,

$$E[\psi_Z(Z)\psi_{XY}(X,Y) \mid U = u, T = t] = E[\psi_Z(Z) \mid U = u, T = t]\, E[\psi_{XY}(X,Y) \mid U = u, T = t],$$

$$E[\psi_X(X)\psi_{YT}(Y,T) \mid U = u] = E[\psi_X(X) \mid U = u]\, E[\psi_{YT}(Y,T) \mid U = u],$$

and

$$E[\psi_{Y_t}(Y(t))\psi_T(T) \mid U = u] = E[\psi_{Y_t}(Y(t)) \mid U = u]\, E[\psi_T(T) \mid U = u].$$

These identities give Assumptions 1, 2 and 4. The construction gives $Y = Y(T)$ and $X_1 = 1$ almost surely, which are Assumptions 3 and 5. Since $X, Z \in \{(1,0)^\top, (1,1)^\top\}$ and $Y \in \{0,1\}$,

$$\|X\|_2 \le \sqrt{2} \le 2, \qquad \|ZX^\top\|_{\mathrm{op}} \le \|Z\|_2 \|X\|_2 \le 2, \qquad \|YZX^\top\|_{\mathrm{op}} \le 2,$$

so Assumptions 6 to 8 hold with $L = 2$. The latent-arm masses are

$$P^{\mathrm{wit}}_\varepsilon(U = 1, T = 0) = \frac{6}{25}, \quad P^{\mathrm{wit}}_\varepsilon(U = 1, T = 1) = \frac{4}{25}, \quad P^{\mathrm{wit}}_\varepsilon(U = 2, T = 0) = \frac{6}{25}, \quad P^{\mathrm{wit}}_\varepsilon(U = 2, T = 1) =$$

and each is at least $1/10$, giving Assumption 9. The feature matrices are

$$B = \begin{pmatrix} 1 & 1 \\ 1/5 & 4/5 \end{pmatrix}, \qquad A_0 = \begin{pmatrix} 1 & 1 \\ 3/10 & 7/10 \end{pmatrix}, \qquad A_1 = \begin{pmatrix} 1 & 1 \\ 7/20 & 3/4 \end{pmatrix}.$$

For every $\upsilon = (\upsilon_1, \upsilon_2) \in \mathbb{R}^2$,

$$\frac{1}{100}\|\upsilon\|_2^2 \le \|B\upsilon\|_2^2, \qquad \frac{1}{100}\|\upsilon\|_2^2 \le \|A_0\upsilon\|_2^2, \qquad \frac{1}{100}\|\upsilon\|_2^2 \le \|A_1\upsilon\|_2^2.$$

Indeed, after multiplying the three differences by $100, 100, 400$, respectively, the differences equal

$$103\left(\upsilon_1 + \frac{116}{103}\upsilon_2\right)^2 + \frac{3333}{103}\upsilon_2^2,$$

$$108\left(\upsilon_1 + \frac{121}{108}\upsilon_2\right)^2 + \frac{1343}{108}\upsilon_2^2,$$

and

$$445\left(\upsilon_1+\frac{505}{445}\upsilon_2\right)^2+\frac{4264}{89}\upsilon_2^2.$$

Thus the smallest singular value of each of B, A_0, A_1 is at least $1/10$, proving Assumption 10. Hence

$$P_\varepsilon^{\mathrm{wit}}\in\mathcal{M}\quad\text{with}\quad L=2,\quad \pi_0=\frac{1}{10},\quad \sigma_0=\frac{1}{10}.$$

3. Lemma 10 gives

$$P_\varepsilon^{\mathrm{wit}}(T=0)=\frac{12}{25}=0.48,\qquad P_\varepsilon^{\mathrm{wit}}(T=1)=\frac{13}{25}=0.52.$$

Also, Lemmas 11 and 12 give

$$\det(B)=\frac{3}{5}=0.6,\qquad \det(A_0)=\det(A_1)=\frac{2}{5}=0.4.$$

It remains to compute the observable proxy moment determinants. For $\iota,\jmath\in\{1,2\}$ and $\mathfrak{t}\in\{0,1\}$, put

$$x_1(\mathfrak{x})=1,\qquad x_2(\mathfrak{x})=\mathfrak{x},\qquad z_1(\mathfrak{z})=1,\qquad z_2(\mathfrak{z})=\mathfrak{z}.$$

After summing out $\mathfrak{y}_0$ and $\mathfrak{y}_1$, Definition 28 gives the conditional-normalized finite sum

$$M_{\mathfrak{t}}(P_\varepsilon^{\mathrm{wit}})_{\iota\jmath}=\frac{\sum_{u=1}^{2}\sum_{\mathfrak{x},\mathfrak{z}\in\{0,1\}}\theta_u^U\,\mathrm{Ber}_{\theta_u^T}(\mathfrak{t})\,\mathrm{Ber}_{\theta_u^X}(\mathfrak{x})\,\mathrm{Ber}_{\theta_{u,\mathfrak{t}}^Z}(\mathfrak{z})z_\iota(\mathfrak{z})x_\jmath(\mathfrak{x})}{\sum_{u=1}^{2}\theta_u^U\,\mathrm{Ber}_{\theta_u^T}(\mathfrak{t})}.$$

Since the $\mathfrak{x}$- and $\mathfrak{z}$-sums are Bernoulli means, this is

$$M_{\mathfrak{t}}(P_\varepsilon^{\mathrm{wit}})_{\iota\jmath}=\sum_{u=1}^{2}P_\varepsilon^{\mathrm{wit}}(U=u\mid T=\mathfrak{t})\,A_{\mathfrak{t}}(\iota,u)B(\jmath,u).$$

For $\mathfrak{t}=0$, the conditional latent weights are $(1/2,1/2)$, and for $\mathfrak{t}=1$ they are $(4/13,9/13)$. Therefore

$$M_0(P_\varepsilon^{\mathrm{wit}})=\begin{pmatrix}1 & 1/2\\ 1/2 & 31/100\end{pmatrix},\qquad M_1(P_\varepsilon^{\mathrm{wit}})=\begin{pmatrix}1 & 8/13\\ 163/260 & 142/325\end{pmatrix}.$$

Taking 2×2 determinants,

$$\det(M_0(P_\varepsilon^{\mathrm{wit}}))=\frac{31}{100}-\frac{1}{4}=\frac{3}{50}=0.06,$$

and

$$\det(M_1(P_\varepsilon^{\mathrm{wit}}))=\frac{142}{325}-\frac{8}{13}\frac{163}{260}=\frac{216}{4225}=\frac{2}{5}\cdot\frac{3}{5}\cdot\frac{36}{169}.$$

4. Lemma 8 gives

$$p_1(P_\varepsilon^{\mathrm{wit}})=\frac{2}{5},\qquad p_2(P_\varepsilon^{\mathrm{wit}})=\frac{3}{5},$$

and Lemma 9 gives

$$\tau_1(P_\varepsilon^{\mathrm{wit}})=\frac{1}{4}-\varepsilon,\qquad \tau_2(P_\varepsilon^{\mathrm{wit}})=\frac{1}{4}+\varepsilon.$$

Since ν_P in Definition 40 places latent mass p_u at latent effect τ_u,

$$\nu_{P_\varepsilon^{\mathrm{wit}}}=\frac{2}{5}\,\delta_{1/4-\varepsilon}+\frac{3}{5}\,\delta_{1/4+\varepsilon}=0.4\,\delta_{0.25-\varepsilon}+0.6\,\delta_{0.25+\varepsilon}.$$

5. At $\varepsilon = 0$, the singular-value inequalities in the second step give injectivity, hence nonsingularity, of B, A_0, A_1. The determinant identities in the third step give nonsingularity of $M_0(P_0^{\mathrm{wit}})$ and $M_1(P_0^{\mathrm{wit}})$. The quotient-law identity in the fourth step becomes

$$\nu_{P_0^{\mathrm{wit}}} = \frac{2}{5}\,\delta_{1/4} + \frac{3}{5}\,\delta_{1/4} = \delta_{1/4} = \delta_{0.25}.$$

Finally, Lemma 8 gives

$$p_1(P_0^{\mathrm{wit}}) = \frac{2}{5}, \qquad p_2(P_0^{\mathrm{wit}}) = \frac{3}{5},$$

and these two latent-class masses are distinct.

□

Proof of Theorem 6. 1. Define

$$a := \min\left\{\frac{1}{3200}, \frac{\sqrt{c_{\mathrm{loc}}}}{3200}\right\}, \qquad c := \frac{a}{8}, \qquad C := 16000.$$

Since $0 < c_{\mathrm{loc}} < 1$,

$$0 < a \le \frac{1}{3200} \le \frac{1}{8}, \qquad c > 0, \qquad C > 0,$$

and

$$16000a^2 \le c_{\mathrm{loc}}, \qquad 16000a^2 \le \frac{1}{4} \le \frac{2}{5}.$$

Fix $n \ge 1$, and put

$$\varepsilon_n := \frac{a}{\sqrt{n}}.$$

Then $0 \le \varepsilon_n \le 1/8$, and

$$16000\varepsilon_n^2 = \frac{16000a^2}{n} \le \frac{c_{\mathrm{loc}}}{n}.$$

Let

$$\Omega_{\mathrm{vis}} := \{0,1\}^4, \qquad \zeta = (t,x,z,y) \in \Omega_{\mathrm{vis}},$$

and let

$$\phi(t,x,z,y) := (t, (1,x)^\top, (1,z)^\top, y) \in O_{2,2},$$

with $O_{2,2}$ as in Definition 18. For $q \in [0,1]$ write

$$\mathrm{Bern}_q(1) = q, \qquad \mathrm{Bern}_q(0) = 1 - q.$$

For $0 \le \varepsilon \le 1/8$, define the finite visible witness cell mass by

$$\psi_\varepsilon^{\mathrm{wit}}(t,x,z,y) = \sum_{u=1}^{2} p_u\ \mathrm{Bern}_{\theta_u}(t)\ \mathrm{Bern}_{\beta_u}(x)\ \mathrm{Bern}_{\alpha_{t,u}}(z) \begin{cases} \mathrm{Bern}_{1/4}(y), & t = 0, \\ \mathrm{Bern}_{1/2+\mathrm{sgn}_u\,\varepsilon}(y), & t = 1, \end{cases}$$

where

$$(p_1,p_2) = \left(\frac{2}{5},\frac{3}{5}\right), \quad (\theta_1,\theta_2) = \left(\frac{2}{5},\frac{3}{5}\right), \quad (\beta_1,\beta_2) = \left(\frac{1}{5},\frac{4}{5}\right),$$

$$(\alpha_{0,1},\alpha_{0,2}) = \left(\frac{3}{10},\frac{7}{10}\right), \quad (\alpha_{1,1},\alpha_{1,2}) = \left(\frac{7}{20},\frac{3}{4}\right), \quad (\mathrm{sgn}_1,\mathrm{sgn}_2) = (-1,1).$$

The pushforward of this finite law by ϕ is $(P_\varepsilon^{\mathrm{wit}})_O$, the observed-data marginal of Definition 20. At $\varepsilon = 0$, every visible cell is bounded below by the first-class contribution:

$$\psi_0^{\mathrm{wit}}(\zeta) \geq \frac{2}{5} \cdot \frac{2}{5} \cdot \frac{1}{5} \cdot \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{500} \geq \frac{1}{1000}.$$

Only the treated outcome factor varies with ε, and hence

$$\left|\psi_\varepsilon^{\mathrm{wit}}(\zeta) - \psi_0^{\mathrm{wit}}(\zeta)\right| \leq \varepsilon.$$

For Kullback–Leibler divergence as in Definition 53, the elementary finite-carrier comparison is

$$D_{\mathrm{KL}}(\mathfrak{r}\|\mathfrak{s}) \leq \sum_\zeta \frac{(\mathfrak{r}(\zeta) - \mathfrak{s}(\zeta))^2}{\mathfrak{s}(\zeta)}$$

whenever $\mathfrak{r}$ and $\mathfrak{s}$ are probability vectors on a finite carrier and $\mathfrak{s}(\zeta) > 0$ for every ζ; this follows from $\log x \leq x - 1$. Therefore

$$D_{\mathrm{KL}}\big((P_\varepsilon^{\mathrm{wit}})_O \| (P_0^{\mathrm{wit}})_O\big) \leq \sum_{\zeta \in \Omega_{\mathrm{vis}}} \frac{\{\psi_\varepsilon^{\mathrm{wit}}(\zeta) - \psi_0^{\mathrm{wit}}(\zeta)\}^2}{\psi_0^{\mathrm{wit}}(\zeta)} \leq 16 \cdot 1000\,\varepsilon^2 = 16000\varepsilon^2,$$

where the first inequality also uses data processing under ϕ. Taking $\varepsilon = \varepsilon_n$, Proposition 5 gives $P_0^{\mathrm{wit}}, P_{\varepsilon_n}^{\mathrm{wit}} \in \mathcal{M}$, and the displayed KL bound places both laws in $\mathcal{L}_n^\nu$ from Definition 56.

2. We use the following two-point expected-risk inequality. Let $\mathsf{R}_1 \ll \mathsf{R}_0$ be one-observation laws with integrable log-likelihood ratio and suppose

$$n D_{\mathrm{KL}}(\mathsf{R}_1 \| \mathsf{R}_0) \leq \frac{2}{5}.$$

Tensorization and Pinsker's inequality give

$$\left\|\mathsf{R}_1^{\otimes n} - \mathsf{R}_0^{\otimes n}\right\|_{\mathrm{TV}} \leq \sqrt{\frac{1}{2} D_{\mathrm{KL}}(\mathsf{R}_1^{\otimes n} \| \mathsf{R}_0^{\otimes n})} \leq \frac{1}{2}.$$

For an estimator $\widehat{\theta}$ in a metric space with targets Θ_1, Θ_0, set

$$\Delta := d(\Theta_1, \Theta_0), \qquad \mathcal{B}_1 := \left\{ d(\widehat{\theta}, \Theta_1) \geq \frac{\Delta}{2} \right\}, \qquad \mathcal{B}_0 := \left\{ d(\widehat{\theta}, \Theta_0) \geq \frac{\Delta}{2} \right\}.$$

The triangle inequality gives $\mathcal{B}_1^c \subseteq \mathcal{B}_0$ when $\Delta > 0$, and the case $\Delta = 0$ is immediate. Thus

$$\mathsf{R}_1^{\otimes n}(\mathcal{B}_1) + \mathsf{R}_0^{\otimes n}(\mathcal{B}_0) \geq 1 - \left\|\mathsf{R}_1^{\otimes n} - \mathsf{R}_0^{\otimes n}\right\|_{\mathrm{TV}} \geq \frac{1}{2},$$

so one of the two probabilities is at least $1/4$. Since

$$\frac{\Delta}{2}\,\Pr\left\{ d(\widehat{\theta}, \Theta_i) \geq \frac{\Delta}{2} \right\} \leq E\, d(\widehat{\theta}, \Theta_i),$$

one of the two expected risks is at least $\Delta/8$. For simplex-valued estimators the same argument applied to one coordinate gives

$$\frac{|p_1(j) - p_0(j)|}{8} \leq E_{\mathsf{R}_1^{\otimes n}}[\|\widehat{p} - p_1\|_1] \quad \text{or} \quad \frac{|p_1(j) - p_0(j)|}{8} \leq E_{\mathsf{R}_0^{\otimes n}}[\|\widehat{p} - p_0\|_1],$$

because $|\widehat{p}(j) - p_i(j)| \leq \|\widehat{p} - p_i\|_1$. The W_1 loss is bounded on the stated compact support radius, and the simplex ℓ^1 loss is bounded by 2, so the expectations used below are finite.

3. Apply the two-point inequality with

$$\mathsf{R}_1 = (P^{\mathrm{wit}}_{\varepsilon_n})_O, \qquad \mathsf{R}_0 = (P^{\mathrm{wit}}_0)_O.$$

The KL calculation gives

$$nD_{\mathrm{KL}}(\mathsf{R}_1 \| \mathsf{R}_0) \le 16000a^2 \le \frac{2}{5}.$$

By Proposition 5,

$$\nu_{P^{\mathrm{wit}}_0} = \delta_{1/4}, \qquad \nu_{P^{\mathrm{wit}}_{\varepsilon_n}} = \frac{2}{5}\delta_{1/4-\varepsilon_n} + \frac{3}{5}\delta_{1/4+\varepsilon_n}.$$

Every coupling from the point mass at $1/4$ to the second law transports total mass one over distance ε_n. Hence, using Definition 19,

$$W_1\Big(\nu_{P^{\mathrm{wit}}_0}, \nu_{P^{\mathrm{wit}}_{\varepsilon_n}}\Big) = \varepsilon_n = \frac{a}{\sqrt{n}}.$$

Thus every measurable quotient-law estimator $\widetilde{\nu}_n$ has one of the two local witness laws satisfying

$$E_{Q_P^{(n)}}[W_1(\widetilde{\nu}_n, \nu_P)] \ge \frac{\varepsilon_n}{8} = \frac{c}{\sqrt{n}}.$$

This proves the quotient-law risk assertion.

4. Fix $g \in (0, 1/4]$, and define

$$h := h(n,g) := a \min\left\{1, \frac{1}{\sqrt{n}\,g}\right\},$$

the calibrated displacement of Definition 58. Since $n \ge 1$ and $g > 0$,

$$x_{n,g} := \sqrt{n}\,g > 0, \qquad 0 < \min\{1, x_{n,g}^{-1}\} \le 1.$$

Together with $a > 0$, this gives

$$0 < h \le a \le \frac{1}{3200} \le \frac{1}{100}, \qquad |h| \le \frac{1}{100}.$$

In particular $h \in [-1, 1]$, so h lies in the tangent-amplitude domain. Moreover,

$$x_{n,g} \min\{1, x_{n,g}^{-1}\} \le 1,$$

and therefore

$$ng^2h^2 = a^2 \left(x_{n,g} \min\{1, x_{n,g}^{-1}\}\right)^2 \le a^2.$$

Consequently

$$16000 g^2 h^2 \le \frac{c_{\mathrm{loc}}}{n}.$$

The hypotheses of Lemma 16 are satisfied for $P_{g,h}$, and also for $P_{g,0}$. Hence

$$P_{g,h}, P_{g,0} \in \mathcal{L}^p_{n,g}.$$

5. We prove the labelled-path KL and ordered-mass certificates. For $(t,x,z,y)\in\Omega_{\mathrm{vis}}$, define

$$\psi^{\mathrm{path}}_{g,h}(t,x,z,y)=\begin{cases}\displaystyle\sum_{u=1}^{2}\sum_{y_0=0}^{1}P_{g,h}(U=u,T=1,X_2=x,Z_2=z,Y(0)=y_0,Y(1)=y), & t=1,\\ \displaystyle\sum_{u=1}^{2}\sum_{y_1=0}^{1}P_{g,h}(U=u,T=0,X_2=x,Z_2=z,Y(0)=y,Y(1)=y_1), & t=0.\end{cases}$$

The pushforward of this finite law by ϕ is $(P_{g,h})_O$. At $h=0$, Definition 50 gives $P_{g,0}=P^{\mathrm{wit}}_{g/2}$. For $t=0$, the first latent class contributes at least

$$\frac{6}{25}\cdot\frac{1}{5}\cdot\frac{3}{10}\cdot\frac{1}{4}=\frac{18}{5000}\geq\frac{1}{1000}.$$

For $t=1$, because $0\leq g\leq 1/4$, both treated outcome probabilities lie in $[3/8,5/8]$, and the first latent class contributes at least

$$\frac{4}{25}\cdot\frac{1}{5}\cdot\frac{7}{20}\cdot\frac{3}{8}=\frac{21}{5000}\geq\frac{1}{1000}.$$

Therefore

$$\psi^{\mathrm{path}}_{g,0}(t,x,z,y)\geq\frac{1}{1000}\qquad\text{for all }(t,x,z,y)\in\Omega_{\mathrm{vis}}.$$

For $|h|\leq 1/100$, all denominators appearing in Definition 50 are bounded away from zero; for example

$$2/5+h\geq 39/100,\qquad 3/5-h\geq 59/100,\qquad 15h+4\geq 77/20,\qquad 5h+3\geq 59/20.$$

Substituting the displayed path probabilities of Definition 50, summing out the inactive potential outcome, and reducing the four (x,z) cases gives

$$\sum_{u=1}^{2}P_{g,h}(U=u,T=1,X_2=x,Z_2=z)=\sum_{u=1}^{2}P_{g,0}(U=u,T=1,X_2=x,Z_2=z),$$

and

$$\begin{aligned}\sum_{u=1}^{2}\mathrm{sgn}_u\,P_{g,h}(U=u,T=1,X_2=x,Z_2=z)-\sum_{u=1}^{2}\mathrm{sgn}_u\,P_{g,0}(U=u,T=1,X_2=x,Z_2=z)\\ =-2\kappa_{xz}h,\end{aligned}$$

where $\mathrm{sgn}_1=-1$, $\mathrm{sgn}_2=1$, and

$$\kappa_{11}=\frac{9}{100},\qquad\kappa_{10}=\frac{3}{100},\qquad\kappa_{01}=\frac{9}{25},\qquad\kappa_{00}=\frac{3}{25}.$$

Since the treated outcome mass is

$$\mathrm{Bern}_{1/2+\mathrm{sgn}_u\,g/2}(y)=\frac{1}{2}+(2y-1)\frac{\mathrm{sgn}_u\,g}{2},$$

the visible-cell displacement is

$$\psi^{\mathrm{path}}_{g,h}(t,x,z,y)-\psi^{\mathrm{path}}_{g,0}(t,x,z,y)=\mathbf{1}\{t=1\}(1-2y)\kappa_{xz}gh.$$

Thus

$$\left|\psi_{g,h}^{\text{path}}(t,x,z,y) - \psi_{g,0}^{\text{path}}(t,x,z,y)\right| \le g|h|.$$

The same finite KL–chi-square bound, now applied with the KL divergence of Definition 53, and data processing under ϕ give

$$D_{\text{KL}}((P_{g,h})_O \| (P_{g,0})_O) \le \sum_{\zeta\in\Omega_{\text{vis}}} \frac{\{\psi_{g,h}^{\text{path}}(\zeta) - \psi_{g,0}^{\text{path}}(\zeta)\}^2}{\psi_{g,0}^{\text{path}}(\zeta)} \le 16000 g^2 h^2 = C g^2 h^2.$$

The latent effects on the path are

$$\tau_1(P_{g,h}) = \frac{1}{4} - \frac{g}{2}, \qquad \tau_2(P_{g,h}) = \frac{1}{4} + \frac{g}{2},$$

so their order is fixed for $g > 0$. Also $|h| \le 1/100$ gives

$$\frac{2}{5} + h > 0, \qquad \frac{3}{5} - h > 0.$$

Using Definition 30,

$$p^{\uparrow}(P_{g,h}) = \left(\frac{2}{5} + h, \frac{3}{5} - h\right), \qquad p^{\uparrow}(P_{g,0}) = \left(\frac{2}{5}, \frac{3}{5}\right),$$

and hence

$$\sum_i \left|p_i^{\uparrow}(P_{g,h}) - p_i^{\uparrow}(P_{g,0})\right| = |h| + |-h| = 2|h|.$$

6. Apply the coordinate form of the two-point inequality to

$$\mathsf{R}_1 = (P_{g,h})_O, \qquad \mathsf{R}_0 = (P_{g,0})_O,$$

and to the ordered coordinate corresponding to the lower effect. The preceding bounds give

$$n D_{\text{KL}}(\mathsf{R}_1 \| \mathsf{R}_0) \le 16000\, n g^2 h^2 \le 16000 a^2 \le \frac{2}{5},$$

and

$$\left|p_1^{\uparrow}(P_{g,h}) - p_1^{\uparrow}(P_{g,0})\right| = h.$$

Therefore every measurable simplex-valued ordered-weight estimator $\widetilde{p}_n$ has one of the two local laws satisfying

$$E_{Q_P^{(n)}}\left[\|\widetilde{p}_n - p^{\uparrow}(P)\|_1\right] \ge \frac{h}{8} = \frac{a}{8} \min\left\{1, \frac{1}{\sqrt{n}\, g}\right\} = c \min\left\{1, \frac{1}{\sqrt{n}\, g}\right\}.$$

The quotient-law certificate follows from the witness KL calculation with $\varepsilon = \varepsilon_n$:

$$D_{\text{KL}}\Big((P_{a/\sqrt{n}}^{\text{wit}})_O \| (P_0^{\text{wit}})_O\Big) \le 16000 \frac{a^2}{n} \le \frac{16000}{n} = \frac{C}{n},$$

and from

$$\frac{c}{\sqrt{n}} \le \frac{a}{\sqrt{n}} = W_1\left(\nu_{P_0^{\text{wit}}}, \nu_{P_{a/\sqrt{n}}^{\text{wit}}}\right).$$

The labelled-path membership, KL bound, tangent-amplitude statement, and ordered-mass identity were established above for every $g \in (0, 1/4]$. All four asserted conclusions hold with the constants a, c, C fixed at the start.

□

Proof of Proposition 6. Apply Theorem 6 with local radius $1/4$. It gives a constant $c_\nu > 0$ such that, for every $n \geq 1$ and every measurable quotient-law estimator $\widetilde{\nu}_n$, there is a probability law P in the local quotient experiment $\mathcal{L}_n^\nu$ satisfying

$$\frac{c_\nu}{\sqrt{n}} \leq E_{Q_P^{(n)}}[W_1(\widetilde{\nu}_n, \nu_P)]\,.$$

By Definition 56, this law belongs to the fixed model class $\mathcal{M}$. Hence

$$\sup_{P \in \mathcal{M}} E_{Q_P^{(n)}}[W_1(\widetilde{\nu}_n, \nu_P)] \geq \frac{c_\nu}{\sqrt{n}}.$$

Taking the infimum over the same admissible estimators gives

$$\inf_{\widetilde{\nu}_n} \sup_{P \in \mathcal{M}} E_{Q_P^{(n)}}[W_1(\widetilde{\nu}_n, \nu_P)] \geq \frac{c_\nu}{\sqrt{n}}.$$

1. For the upper bound, Theorem 2 applies at $k = d_x = d_z = 2$, $L = 2$, $\pi_0 = 1/10$, and $\sigma_0 = 1/10$. Thus there is a constant $C_{\mathrm{col}} > 0$ such that, for every $n \geq 1$, there is a measurable structured-lattice estimator

$$\widehat{\lambda}_n : O_{2,2}^n \to \mathcal{P}_{\leq 2}([-L_\tau, L_\tau])$$

and a measurable repair estimator $\widehat{\nu}_n^{\mathrm{rep}}$ such that, for every $P \in \mathcal{M}$ and every $0 < \eta < 1/2$,

$$Q_P^{(n)}\left\{\max\Big(W_1(\widehat{\lambda}_n, \nu_P), W_1(\widehat{\nu}_n^{\mathrm{rep}}, \nu_P)\Big) > C_{\mathrm{col}}\sqrt{\frac{\log(C_{\mathrm{col}}/\eta)}{n}}\right\} \leq \eta.$$

Set

$$\overline{C}_{\mathrm{col}} := \max\{C_{\mathrm{col}}, 2\}.$$

For $0 < \eta < 1/2$,

$$C_{\mathrm{col}}\sqrt{\frac{\log(C_{\mathrm{col}}/\eta)}{n}} \leq \overline{C}_{\mathrm{col}}\sqrt{\frac{\log(\overline{C}_{\mathrm{col}}/\eta)}{n}},$$

because $C_{\mathrm{col}} \leq \overline{C}_{\mathrm{col}}$, the logarithm is monotone on positive arguments, and $n \geq 1$. Therefore, with

$$Z_P(\omega) := W_1\Big(\widehat{\lambda}_n(\omega), \nu_P\Big), \qquad \omega \in O_{2,2}^n,$$

we have the uniform tail bound

$$Q_P^{(n)}\left\{Z_P > \overline{C}_{\mathrm{col}}\sqrt{\frac{\log(\overline{C}_{\mathrm{col}}/\eta)}{n}}\right\} \leq \eta.$$

2. The one-Wasserstein loss is nonnegative, so $Z_P(\omega) \geq 0$. Define

$$\beta_n := \overline{C}_{\mathrm{col}}\sqrt{\frac{\log(2\overline{C}_{\mathrm{col}})}{n}}, \qquad \chi_n := \frac{n}{\overline{C}_{\mathrm{col}}^2}.$$

Since $\overline{C}_{\mathrm{col}} \geq 2$, both β_n and χ_n are positive. For every $\zeta > \beta_n$, put

$$\eta_\zeta := \overline{C}_{\mathrm{col}} \exp(-\chi_n \zeta^2).$$

Then $0 < \eta_\zeta < 1/2$, and

$$\overline{C}_{\mathrm{col}} \sqrt{\frac{\log(\overline{C}_{\mathrm{col}}/\eta_\zeta)}{n}} = \overline{C}_{\mathrm{col}} \sqrt{\frac{\chi_n \zeta^2}{n}} = \zeta.$$

Substitution in the preceding tail bound gives

$$Q_P^{(n)}\{Z_P > \zeta\} \le \overline{C}_{\mathrm{col}} \exp(-\chi_n \zeta^2), \qquad \zeta > \beta_n.$$

Using the nonnegative-tail identity and the trivial bound $Q_P^{(n)}\{Z_P > \zeta\} \le 1$ for $0 < \zeta \le \beta_n$,

$$\begin{aligned} E_{Q_P^{(n)}}[Z_P] &\le \beta_n + \overline{C}_{\mathrm{col}} \int_{\beta_n}^{\infty} \exp(-\chi_n \zeta^2)\, d\zeta \\ &\le \beta_n + \overline{C}_{\mathrm{col}} \int_0^{\infty} \exp(-\chi_n \zeta^2)\, d\zeta \\ &= \beta_n + \overline{C}_{\mathrm{col}} \frac{\sqrt{\pi/\chi_n}}{2} \\ &= \frac{\overline{C}_{\mathrm{col}} \sqrt{\log(2\overline{C}_{\mathrm{col}})} + \overline{C}_{\mathrm{col}}^2 \sqrt{\pi}/2}{\sqrt{n}}. \end{aligned}$$

3. Let

$$C_{\mathrm{mean},\nu} := \overline{C}_{\mathrm{col}} \sqrt{\log(2\overline{C}_{\mathrm{col}})} + \overline{C}_{\mathrm{col}}^2 \sqrt{\pi}/2, \qquad C_\nu := \max\{C_{\mathrm{mean},\nu},\, 2c_\nu\}.$$

The first term is positive, so $C_\nu > 0$, and $c_\nu < C_\nu$. The estimator $\widehat{\lambda}_n$ is admissible for the infimum, and the preceding display holds uniformly in $P \in \mathcal{M}$. Hence, for every $n \ge 1$,

$$\inf_{\widetilde{\nu}_n} \sup_{P \in \mathcal{M}} E_{Q_P^{(n)}}[W_1(\widetilde{\nu}_n, \nu_P)] \le \sup_{P \in \mathcal{M}} E_{Q_P^{(n)}}[Z_P] \le \frac{C_{\mathrm{mean},\nu}}{\sqrt{n}} \le \frac{C_\nu}{\sqrt{n}}.$$

Taking $c = c_\nu$ and $C = C_\nu$ gives $0 < c < C$ and both displayed minimax inequalities.

□

Proof of Proposition 7. For $n \ge 1$ and $0 < g \le 1/4$, set

$$\beta_{n,g} := \min\left\{1, \frac{1}{\sqrt{n}\, g}\right\}.$$

Let Est_n be the class of measurable simplex-valued labelled weight estimators based on n observed records in the specialized model

$$k = d_x = d_z = 2, \qquad L = 2, \qquad \pi_0 = \frac{1}{10}, \qquad \sigma_0 = \frac{1}{10},$$

and define

$$\mathsf{Risk}_{n,g} := \inf_{\widetilde{p}_n \in \mathsf{Est}_n} \sup_{P \in \mathcal{M}(g)} E_{Q_P^{(n)}}\left[\left\|\widetilde{p}_n - p^{\uparrow}(P)\right\|_1\right].$$

1. Apply Theorem 6 with $c_{\mathrm{loc}} = 1/4$. It gives a constant $\kappa_{\mathrm{lo}} > 0$ such that, for every $n \geq 1$, every $0 < g \leq 1/4$, and every $\widetilde{p}_n \in \mathsf{Est}_n$, there is a law $P \in \mathcal{L}^p_{n,g}$ satisfying

$$\kappa_{\mathrm{lo}} \beta_{n,g} \leq E_{Q_P^{(n)}} \left[\left\| \widetilde{p}_n - p^{\uparrow}(P) \right\|_1 \right].$$

By Definition 57, $\mathcal{L}^p_{n,g} \subseteq \mathcal{M}(g)$. Hence, for every admissible $\widetilde{p}_n$,

$$\sup_{P \in \mathcal{M}(g)} E_{Q_P^{(n)}} \left[\left\| \widetilde{p}_n - p^{\uparrow}(P) \right\|_1 \right] \geq \kappa_{\mathrm{lo}} \beta_{n,g},$$

and taking the infimum over $\widetilde{p}_n \in \mathsf{Est}_n$ yields

$$\mathsf{Risk}_{n,g} \geq \kappa_{\mathrm{lo}} \beta_{n,g}.$$

2. Apply Theorem 5 with

$$k = d_x = d_z = 2, \qquad L = 2, \qquad \pi_0 = \frac{1}{10}, \qquad \sigma_0 = \frac{1}{10}.$$

It gives a constant $\kappa_{\mathrm{up}} > 0$ such that, for every $n \geq 1$, there is an estimator

$$\widehat{p}^{\uparrow}_n \in \mathsf{Est}_n$$

satisfying

$$E_{Q_P^{(n)}} \left[\left\| \widehat{p}^{\uparrow}_n - p^{\uparrow}(P) \right\|_1 \right] \leq \kappa_{\mathrm{up}} \beta_{n,g}$$

for every $0 < g \leq 1/4$ and every $P \in \mathcal{M}(g)$. The target is the ordered aggregate mass vector of the quotient law ν_P from Definition 40: Definition 30 defines it as $p^{\uparrow}(P)$, and Lemma 17 makes this ordered vector depend only on the represented probability law. Therefore

$$\mathsf{Risk}_{n,g} \leq \sup_{P \in \mathcal{M}(g)} E_{Q_P^{(n)}} \left[\left\| \widehat{p}^{\uparrow}_n - p^{\uparrow}(P) \right\|_1 \right] \leq \kappa_{\mathrm{up}} \beta_{n,g}.$$

3. Define

$$c := \min \left\{ \kappa_{\mathrm{lo}}, \frac{\kappa_{\mathrm{up}}}{2} \right\}, \qquad C := \kappa_{\mathrm{up}}.$$

Since $\kappa_{\mathrm{lo}} > 0$ and $\kappa_{\mathrm{up}} > 0$,

$$0 < c \quad \text{and} \quad c \leq \frac{\kappa_{\mathrm{up}}}{2} < \kappa_{\mathrm{up}} = C,$$

so $0 < c < C$. Combining the bounds from the preceding two steps gives, for every $n \geq 1$ and $0 < g \leq 1/4$,

$$c \min \left\{ 1, \frac{1}{\sqrt{n}\, g} \right\} \leq \mathsf{Risk}_{n,g} \leq C \min \left\{ 1, \frac{1}{\sqrt{n}\, g} \right\}.$$

This is the asserted minimax inequality.

□

Proof of Theorem 7. 1. Apply Theorem 6 with

$$c_{\mathrm{loc}} = \frac{1}{4},$$

which lies in the domain of Definition 49. This gives constants $a, c \in \mathbb{R}$, and an auxiliary positive Kullback–Leibler constant, such that

$$0 < a \le \frac{1}{8}, \qquad 0 < c,$$

and all conclusions of Theorem 6 hold for every $n \ge 1$. In the present specialization,

$$L_\tau = \frac{4L\sqrt{d_z}}{\sigma_0} = \frac{4 \cdot 2\sqrt{2}}{1/10}.$$

We record the published-scope implication used for the endpoint laws. Let $P \in \mathcal{M}$ for

$$k = d_x = d_z = 2, \qquad L = 2, \qquad \pi_0 = \sigma_0 = \frac{1}{10}.$$

The restrictions in Definition 32 supply the qualitative proxy separations, causal consistency, armwise latent ignorability, anchor normalization, boundedness, latent-arm positivity, and proxy-rank margin. The proxy-rank margin gives

$$\sigma_2(A_0) \ge \frac{1}{10}, \qquad \sigma_2(A_1) \ge \frac{1}{10}, \qquad \sigma_2(B) \ge \frac{1}{10},$$

so the three proxy feature matrices have full column rank. Latent-arm positivity gives

$$P(U = u, T = t) \ge \frac{1}{10} > 0, \qquad u \in \{1, 2\}, \quad t \in \{0, 1\}.$$

Thus P satisfies the qualitative structural conditions in Definition 60. The cited VMW model-scope correspondence, [Virk et al., 2026, Assumptions 1–2 and Section 4.2], identifies these qualitative conditions with membership in the published VMW model predicate attached to Definition 52.

It remains to verify VMW Assumption 4 in Definition 3. Put

$$M_{\mathrm{stack}} := \mathrm{stack}(M_0(P), M_1(P)).$$

Choose a singular-value decomposition

$$M_{\mathrm{stack}} = \sum_j s_j \ell_j r_j^\top, \qquad s_1 \ge s_2 \ge \cdots .$$

By Proposition 1,

$$s_2(M_{\mathrm{stack}}) = \sigma_2(M_{\mathrm{stack}}) \ge \pi_0 \sigma_0^2 = \frac{1}{1000} > 0.$$

Let $V = (r_1, r_2)$. Then the columns of V are orthonormal and satisfy

$$M_{\mathrm{stack}}^\top M_{\mathrm{stack}} r_j = s_j^2 r_j, \qquad j = 1, 2,$$

so V is a population top-two right singular basis. Since $s_j > 0$,

$$r_j = s_j^{-1} M_{\text{stack}}^{\top} \ell_j, \qquad j = 1, 2.$$

For $y = (y_0, y_1) \in \mathbb{R}^{d_z} \oplus \mathbb{R}^{d_z}$,

$$M_{\text{stack}}^{\top} y = M_0(P)^{\top} y_0 + M_1(P)^{\top} y_1.$$

Hence each r_j lies in the signal space generated by the two armwise row spaces. The factorization in Proposition 1 gives rank at most 2 for this signal space, while $s_2(M_{\text{stack}}) > 0$ gives rank at least 2. Therefore the span of V is exactly the signal space. Applying the compressed-margin conclusion of Proposition 1 to this basis gives

$$\sigma_2\{M_t(P)V\} \geq \pi_0 \sigma_0^2 = \frac{1}{1000}, \qquad t \in \{0, 1\}.$$

Together with

$$\|X\|_2 \leq 2, \qquad \|ZX^{\top}\|_{\text{op}} \leq 2, \qquad \|YZX^{\top}\|_{\text{op}} \leq 2, \qquad P(T = t) \geq 2\pi_0 = \frac{1}{5},$$

these are the finite envelope, treatment-arm, top-right-basis, stacked-margin, and compressed-margin clauses of Definition 3, with

$$L_X = L_{ZX} = L_{YZX} = 2, \qquad \pi = \frac{1}{5}, \qquad \sigma = \frac{1}{1000}.$$

Thus every such $P \in \mathcal{M}$ satisfies VMW Assumption 4.

2. Fix $n \geq 1$. The quotient-law certificate in Theorem 6, together with Definition 56, supplies probability laws

$$P_0^{\text{wit}}, \qquad P_{a/\sqrt{n}}^{\text{wit}},$$

belonging to $\mathcal{M}$ with the fixed parameters above. By the implication from the preceding step, both endpoint laws belong to the published VMW model and satisfy VMW Assumption 4.

At the collision endpoint, Lemma 9 gives

$$\tau_1(P_0^{\text{wit}}) = \frac{1}{4}, \qquad \tau_2(P_0^{\text{wit}}) = \frac{1}{4}.$$

The cited recovery-regime correspondence, [Virk et al., 2026, Assumption 3, lines 261–263; Assumption 4, lines 321–337; and Theorem 7.2, lines 350–381], identifies the separated recovery predicate attached to Definition 52, for $k = d_x = d_z = 2$, with the conjunction in Definition 61, including spectral separation of the two latent effects. The displayed equality violates that spectral-separation clause, so P_0^{wit} lies outside the published separated recovery regime.

3. Let $\mathcal{V}_n^{\nu}$ satisfy the quotient-class hypotheses, and choose

$$P_{\nu,0}^{\mathcal{V}}, P_{\nu,1}^{\mathcal{V}} \in \mathcal{V}_n^{\nu}$$

with

$$P_{\nu,0}^{\mathcal{V}} = P_0^{\text{wit}}, \qquad P_{\nu,1}^{\mathcal{V}} = P_{a/\sqrt{n}}^{\text{wit}}.$$

For any estimator $\widetilde{\nu}_n$ with support radius L_τ, the quotient-risk conclusion of Theorem 6 gives an endpoint

$$P_{\nu,\star} \in \left\{P_0^{\mathrm{wit}}, P_{a/\sqrt{n}}^{\mathrm{wit}}\right\}$$

such that

$$\frac{c}{\sqrt{n}} \le E_{Q^{(n)}_{P_{\nu,\star}}}\left[W_1\big(\widetilde{\nu}_n, \nu_{P_{\nu,\star}}\big)\right].$$

If $P_{\nu,\star} = P_0^{\mathrm{wit}}$, use $P^{\mathcal{V}}_{\nu,0}$; if $P_{\nu,\star} = P^{\mathrm{wit}}_{a/\sqrt{n}}$, use $P^{\mathcal{V}}_{\nu,1}$. Equality as full-data laws identifies the observed product law and the quotient law, hence

$$\sup_{P \in \mathcal{V}_n^\nu} E_{Q_P^{(n)}}[W_1(\widetilde{\nu}_n, \nu_P)] \ge \frac{c}{\sqrt{n}}.$$

This is the quotient lower bound.

4. Fix g in the gap-scale domain, so

$$0 < g \le \frac{1}{4},$$

and set

$$h(n,g) := a \min\left\{1, \frac{1}{\sqrt{n}\, g}\right\}.$$

The labeled-path certificate in Theorem 6 supplies the tangent-amplitude condition

$$h(n,g) \in [-1,1],$$

and places

$$P_{g,0}, \qquad P_{g,h(n,g)}$$

in the local labeled-weight experiment $\mathcal{L}^p_{n,g}$. By Definition 57, both laws belong to the gap stratum $\mathcal{M}(g)$, and therefore to $\mathcal{M}$ with the fixed parameters. The implication from the first step gives published VMW model membership and VMW Assumption 4 for both endpoint laws.

Let P_{end} be either $P_{g,0}$ or $P_{g,h(n,g)}$. Since $P_{\mathrm{end}} \in \mathcal{M}(g)$, Definition 41 supplies the distinct-effects condition of Assumption 12. Latent-arm positivity gives

$$P_{\mathrm{end}}(U = u) \ge P_{\mathrm{end}}(U = u, T = 0) \ge \frac{1}{10} > 0, \qquad u \in \{1,2\}.$$

Thus both latent classes have positive mass, and the distinct-effects condition yields

$$\tau_1(P_{\mathrm{end}}) \ne \tau_2(P_{\mathrm{end}}).$$

This is the two-class spectral-separation condition in Definition 61.

5. Let $\mathcal{V}^p_{n,g}$ satisfy the labeled-weight comparator-class hypotheses, and choose

$$P^{\mathcal{V}}_{p,0}, P^{\mathcal{V}}_{p,1} \in \mathcal{V}^p_{n,g}$$

with

$$P^{\mathcal{V}}_{p,0} = P_{g,0}, \qquad P^{\mathcal{V}}_{p,1} = P_{g,h(n,g)}.$$

For any ordered-weight estimator $\widetilde{p}_n$, the ordered-weight conclusion of Theorem 6 gives an endpoint

$$P_{p,\star} \in \{P_{g,0}, P_{g,h(n,g)}\}$$

such that

$$c \min\left\{1, \frac{1}{\sqrt{n}\, g}\right\} \le E_{Q^{(n)}_{P_{p,\star}}}\left[\left\|\widetilde{p}_n - p^{\uparrow}(P_{p,\star})\right\|_1\right].$$

If $P_{p,\star} = P_{g,0}$, use $P^{\mathcal{V}}_{p,0}$; if $P_{p,\star} = P_{g,h(n,g)}$, use $P^{\mathcal{V}}_{p,1}$. Equality as full-data laws preserves the observed product law and the ordered-mass target, hence

$$\sup_{P \in \mathcal{V}^p_{n,g}} E_{Q^{(n)}_P}\left[\left\|\widetilde{p}_n - p^{\uparrow}(P)\right\|_1\right] \ge c \min\left\{1, \frac{1}{\sqrt{n}\, g}\right\}.$$

This proves the labeled-weight lower bound.

□

Proof of Theorem 8. 1. Put

$$D := 4d_z d_x + d_x, \qquad b := 4\sqrt{d_z d_x} + \sqrt{d_x}.$$

By Theorem 1, choose $C_{\text{mod}} > 0$ such that, for any two model laws $P_1, P_2 \in \mathcal{M}$,

$$W_1(\nu_{P_1}, \nu_{P_2}) \le C_{\text{mod}}\, d_S(S(P_1), S(P_2)).$$

Let $C_{\text{card}} > 0$ and $C_{\text{work}} > 0$ be the finite-library cardinality and work constants constructed below, and let $C_0 \ge 1$ be the constant supplied by Lemma 1. Define

$$C_{\text{tail}} := \max\{C_0, \exp(1), C_{\text{mod}}(2C_0 L + 1)\}, \qquad C := \max\{C_{\text{card}}, C_{\text{work}}, C_{\text{tail}}\}.$$

Then $C > 0$, and this single constant dominates the constants used for the library size, work bound, and probability bound.

2. Fix $n \ge 1$, and set

$$\epsilon_n := n^{-1/2}, \qquad \delta_n := \frac{\epsilon_n}{b}.$$

The library is the deterministic half-open coordinate-grid library in the D-coordinate summary box, with side length δ_n. Its coordinates are the entries of the four $d_z \times d_x$ matrix blocks and the d_x entries of the mean block. For each grid cube meeting the admissible image $\mathscr{S} = S(\mathcal{M})$, choose one representative $S_i \in \mathscr{S}$ and order the representatives by the lexicographic order of the cube lower corners. The coordinate bounds following from Proposition 1 place every admissible summary coordinate in $[-L, L]$. If $q \in \mathscr{S}$ and S_i lie in the same cube, then every scalar coordinate differs by at most δ_n, so

$$\|M_t(S_i) - M_t(q)\|_{\text{op}} \le \sqrt{d_z d_x}\, \delta_n, \qquad \|N_t(S_i) - N_t(q)\|_{\text{op}} \le \sqrt{d_z d_x}\, \delta_n$$

for $t \in \{0, 1\}$, and

$$\|m_X(S_i) - m_X(q)\|_2 \le \sqrt{d_x}\, \delta_n.$$

Using d_S from Definition 33,

$$d_S(S_i, q) \le \left(4\sqrt{d_z d_x} + \sqrt{d_x}\right)\delta_n = \epsilon_n.$$

Thus the stored representatives form an $n^{-1/2}$-net of $\mathscr{S}$; when $\mathscr{S} = \varnothing$, the index set is empty and the program uses the fallback law in Algorithm 7.

3. Each coordinate lower corner has at most

$$\left\lceil \frac{2L}{\delta_n} \right\rceil + 1$$

possible values. Since $\delta_n = n^{-1/2}/b$ and $n^{1/2} \geq 1$,

$$\left\lceil \frac{2L}{\delta_n} \right\rceil + 1 \leq (2Lb + 2)n^{1/2}.$$

The lower-corner code is injective, hence

$$|A| \leq \big((2Lb+2)n^{1/2}\big)^D = (2Lb+2)^D n^{D/2} \leq C_{\mathrm{card}} n^{(4d_z d_x + d_x)/2}.$$

After enlarging C, the asserted cardinality bound follows.

4. Fix a uniform provider of the exact-real primitives. The estimator is the law component of $\mathrm{NetProgram}(A, \omega)$ in Algorithm 7. On a nonempty library, enumerate the finite index set as $\iota_0, \ldots, \iota_{m-1}$ and update the current index by choosing the smaller pair

$$\big(d_S(S_i, s), \mathrm{rank}_{\mathrm{lex}}(i)\big)$$

in lexicographic order. For each fixed representative, $s \mapsto d_S(S_i, s)$ is continuous by Lemma 23; therefore every update is Borel, being defined by finitely many weak and strict Borel inequalities. The empty selection is constant. Since $\widehat{S}_n$ is Borel by Definition 35, and the selected finite index is mapped to a stored spectral law, $\widehat{\nu}_n^{\mathrm{net}}$ is Borel measurable.

5. The exact-real trace is assembled from the empirical-summary trace, the finite-library search trace, and the selected spectral run:

$$\mathfrak{T}(\omega) = \mathfrak{T}_{\mathrm{emp}}(n) \,\|\, \mathfrak{T}_{\mathrm{search}}(A) \,\|\, \begin{cases} \mathfrak{T}_{\mathrm{spec}}(S_i), & \text{if the search selects representative } i, \\ \varnothing, & \text{if the search has empty selection.} \end{cases}$$

Thus the returned law and the displayed trace come from the same execution.

6. The empirical-summary trace has nD arithmetic operations. Each library comparison uses D scalar arithmetic operations, four fixed-dimensional singular-value or operator-norm operations, and one comparison. The selected spectral tail has three singular-value operations, one root-isolation operation, and one arithmetic operation. Therefore, for every sample ω,

$$\mathrm{ops}(\omega) \leq nD + |A|(D+5) + 5.$$

Combining this with the cardinality bound gives

$$\mathrm{ops}(\omega) \leq C_{\mathrm{work}}\{n + n^{D/2}\} \leq C\{n + n^{(4d_z d_x + d_x)/2}\}.$$

7. If the program selects i at sample ω, put

$$s_\omega := \widehat{S}_n(\omega).$$

The exhaustive fold preserves domination, where i dominates j at s_ω when

$$d_S(S_i, s_\omega) \leq d_S(S_j, s_\omega)$$

and equality implies

$$\mathrm{rank}_{\mathrm{lex}}(i) \le \mathrm{rank}_{\mathrm{lex}}(j).$$

Induction over the scanned list yields, for every library index j,

$$d_S(S_i, \widehat{S}_n(\omega)) \le d_S(S_j, \widehat{S}_n(\omega)),$$

and

$$d_S(S_i, \widehat{S}_n(\omega)) = d_S(S_j, \widehat{S}_n(\omega)) \Longrightarrow \mathrm{rank}_{\mathrm{lex}}(i) \le \mathrm{rank}_{\mathrm{lex}}(j).$$

8. Let $\mathcal{R}_i$ be the exact spectral run at the selected representative S_i. By the definition of the program output,

$$\widehat{\nu}_n^{\mathrm{net}}(\omega) = \mathrm{law}(\mathcal{R}_i). \tag{3}$$

9. The same spectral execution supplies a basis V_i, its signal spans, and an eigenvalue list $(\lambda_{i,r})_{r<k}$. Its output fields give, for each $t \in \{0,1\}$,

$$x \longmapsto M_t(S_i)V_i x$$

injective, the exact threshold characterization

$$\frac{\pi_0\sigma_0^2}{2} \le \sigma_j(\mathrm{stack}(M_0(S_i), M_1(S_i))) \Longleftrightarrow j < k,$$

and completeness of the returned eigenvalue list:

$$z \in \mathrm{spec}(\Delta_i) \Longrightarrow \exists r < k \text{ such that } z = \lambda_{i,r},$$

where Δ_i is the compressed operator formed from S_i, V_i, and the run's spans.

10. Fix any stored representative S_i. Since $S_i \in \mathscr{S}$, choose $Q_i \in \mathcal{M}$ with

$$S(Q_i) = S_i.$$

Let $p_{i,u} := Q_i(U = u)$, write $p_i := (p_{i,u})_{u=1}^k$, and set

$$\tau_{i,u} := \mu_{1u}(Q_i) - \mu_{0u}(Q_i).$$

The latent-arm positivity condition in Assumption 9 gives $p_{i,u} > 0$ for every u. Let $\mathcal{R}_i$ be the exact spectral run at S_i, with signal basis V_i, spans, anchors a_i, c_i, returned eigenvalues $(\lambda_{i,r})_{r<k}$, and compressed operator

$$\Delta_i := (M_1(S_i)V_i)^\dagger N_1(S_i)V_i - (M_0(S_i)V_i)^\dagger N_0(S_i)V_i.$$

Set

$$\Omega_i := B(Q_i)^\top V_i \in \mathbb{R}^{k\times k}, \qquad \Gamma_{i,t} := \mathrm{diag}\{Q_i(U = u \mid T = t) : 1 \le u \le k\},$$

and

$$\Phi_{i,t} := A_t(Q_i)\Gamma_{i,t} \in \mathbb{R}^{d_z \times k}.$$

Lemmas 4 and 18 give, for $t \in \{0,1\}$,

$$M_t(S_i)V_i = \Phi_{i,t}\Omega_i, \qquad N_t(S_i)V_i = \Phi_{i,t}\,\mathrm{diag}(\mu_{t1}(Q_i), \ldots, \mu_{tk}(Q_i))\Omega_i.$$

The run's armwise full-rank certificate makes $x \mapsto M_t(S_i)V_i x$ injective for each t. Taking $t = 0$, the first factorization shows that Ω_i is injective and hence invertible. The same factorization also gives injectivity of $\Phi_{i,0}$: if $\Phi_{i,0}v = \Phi_{i,0}v'$ for $v, v' \in \mathbb{R}^k$, then

$$M_0(S_i)V_i\Omega_i^{-1}v = \Phi_{i,0}v = \Phi_{i,0}v' = M_0(S_i)V_i\Omega_i^{-1}v',$$

so injectivity of $M_0(S_i)V_i$ gives $\Omega_i^{-1}v = \Omega_i^{-1}v'$, and hence $v = v'$. The proxy factorization before postmultiplication by V_i is

$$M_0(S_i) = \Phi_{i,0}B(Q_i)^\top.$$

Since $\Phi_{i,0}$ is injective, its Moore–Penrose inverse is a left inverse on its range, and therefore

$$\Phi_{i,0}^\dagger M_0(S_i) = B(Q_i)^\top.$$

Transposing gives

$$B(Q_i) = M_0(S_i)^\top(\Phi_{i,0}^\dagger)^\top.$$

By construction of the run's span certificate,

$$\mathrm{range}(V_i) = \mathrm{range}(M_0(S_i)^\top) + \mathrm{range}(M_1(S_i)^\top) \subseteq \mathbb{R}^{d_x},$$

and consequently

$$\mathrm{range}(B(Q_i)) \subseteq \mathrm{range}(M_0(S_i)^\top) \subseteq \mathrm{range}(V_i).$$

The second factorization can therefore be rewritten as

$$N_t(S_i)V_i = M_t(S_i)V_i\{\Omega_i^{-1}\operatorname{diag}(\mu_{t1}(Q_i), \ldots, \mu_{tk}(Q_i))\Omega_i\}.$$

Since the Moore–Penrose inverse is a left inverse on the range of an injective linear map,

$$(M_t(S_i)V_i)^\dagger N_t(S_i)V_i = \Omega_i^{-1}\operatorname{diag}(\mu_{t1}(Q_i), \ldots, \mu_{tk}(Q_i))\Omega_i.$$

Subtracting the two treatment-arm identities gives

$$\Delta_i = \Omega_i^{-1}\operatorname{diag}(\tau_{i,1}, \ldots, \tau_{i,k})\Omega_i.$$

Finally, the law of total expectation gives

$$m_X(S_i) = B(Q_i)p_i.$$

The column-space inclusion just proved implies that the orthogonal projector $V_iV_i^\top$ fixes $B(Q_i)$. Since $\Omega_i^\top = V_i^\top B(Q_i)$,

$$B(Q_i) = V_iV_i^\top B(Q_i) = V_i\Omega_i^\top.$$

Thus

$$a_i = V_i^\top m_X(S_i) = \Omega_i^\top p_i.$$

The target-proxy anchor condition gives $B(Q_i)^\top e_1 = \mathbf{1}_k$, while $c_i = V_i^\top e_1$ and $B(Q_i)^\top V_iV_i^\top = B(Q_i)^\top$, so

$$\Omega_i c_i = \mathbf{1}_k.$$

For the polynomial projector attached to a returned slot r,

$$\mathcal{P}_{i,r} := \prod_{\lambda_{i,\ell} \neq \lambda_{i,r}} (\lambda_{i,r} - \lambda_{i,\ell})^{-1} (\Delta_i - \lambda_{i,\ell} I_k),$$

the diagonalization gives

$$\mathcal{P}_{i,r} = \begin{cases} \Omega_i^{-1} \operatorname{diag}\{\mathbf{1}(\tau_{i,u} = \lambda_{i,r})\}_{u=1}^k \Omega_i, & \lambda_{i,r} \in \{\tau_{i,u} : 1 \leq u \leq k\}, \\ 0, & \lambda_{i,r} \notin \{\tau_{i,u} : 1 \leq u \leq k\}. \end{cases}$$

Hence the returned slot weight

$$w_{i,r} := a_i^\top \mathcal{P}_{i,r} c_i$$

satisfies

$$w_{i,r} = \sum_{u:\tau_{i,u} = \lambda_{i,r}} p_{i,u}$$

for a spectral value and $w_{i,r} = 0$ for a listed value outside the spectrum. The output validity of $\mathcal{R}_i$ gives $w_{i,r} \geq 0$ and $\sum_{r<k} w_{i,r} = 1$. Completeness gives at least one returned slot for each latent-effect value, and the positivity of the corresponding cluster mass

$$m_i(\lambda) := \sum_{u:\tau_{i,u} = \lambda} p_{i,u}$$

forces exactly one positive returned slot for each distinct latent-effect value; otherwise the total $\sum_{r<k} w_{i,r}$ would exceed $\sum_\lambda m_i(\lambda) = 1$. Therefore, for every bounded test function $\varphi : \mathbb{R} \to \mathbb{R}$,

$$\sum_{r<k} w_{i,r} \varphi(\lambda_{i,r}) = \sum_{u=1}^k p_{i,u} \varphi(\tau_{i,u}).$$

Thus the induced atomic measures agree:

$$\sum_{r<k} w_{i,r} \delta_{\lambda_{i,r}} = \sum_{u=1}^k p_{i,u} \delta_{\tau_{i,u}} = \nu_{Q_i},$$

with coincident effects aggregated as in Definition 40. The exact spectral run at S_i returns ν_{Q_i}:

$$\operatorname{law}(\mathcal{R}_i) = \nu_{Q_i}. \tag{4}$$

11. The validity certificate of the same result-bearing spectral execution states

$$w_{i,r} \geq 0, \qquad \sum_{r<k} w_{i,r} = 1, \qquad |\lambda_{i,r}| \leq L_\tau \quad (r < k).$$

Consequently the returned weights form a probability vector and every returned atom lies in $[-L_\tau, L_\tau]$, so the represented finite atomic measure is a probability law supported on that interval.

12. For a stored representative $S_i = S(Q_i)$, put

$$\mathcal{G}_i := \operatorname{stack}(M_0(S_i), M_1(S_i)), \qquad s_0 := \pi_0 \sigma_0^2.$$

Lemma 25 supplies the rank identity, and Lemma 24 supplies the singular-value margin:

$$\operatorname{rank}(\mathcal{G}_i) = k, \qquad s_0 \le \sigma_{k-1}(\mathcal{G}_i).$$

If $H \in \mathbb{R}^{2d_z \times d_x}$ satisfies

$$\|H\|_{\mathrm{op}} < s_0/2,$$

Weyl's singular-value inequality gives, for every r,

$$|\sigma_r(\mathcal{G}_i + H) - \sigma_r(\mathcal{G}_i)| \le \|H\|_{\mathrm{op}}.$$

Thus

$$\sigma_r(\mathcal{G}_i + H) > s_0/2 \quad (r < k),$$

while $\operatorname{rank}(\mathcal{G}_i) = k$ gives $\sigma_r(\mathcal{G}_i) = 0$ for $r \ge k$, and hence

$$\sigma_r(\mathcal{G}_i + H) < s_0/2 \quad (r \ge k).$$

Consequently the perturbed stacked moment satisfies the two threshold clauses

$$\frac{s_0}{2} \le \sigma_r(\mathcal{G}_i + H) \quad (r < k), \qquad \sigma_r(\mathcal{G}_i + H) < \frac{s_0}{2} \quad (r \ge k).$$

Equivalently, exactly k singular values of $\mathcal{G}_i + H$ meet the threshold $s_0/2 = \pi_0 \sigma_0^2/2$.

13. Fix $P \in \mathcal{M}$ and a sample ω. If $P \in \mathcal{M}$, then $\mathscr{S}$ is nonempty, so the library has a selected nearest index i. By the covering property, choose j such that

$$d_S(S_j, S(P)) \le n^{-1/2}.$$

Nearest-library optimality and the triangle inequality yield

$$\begin{aligned}
d_S(S_i, S(P)) &\le d_S(S_i, \widehat{S}_n(\omega)) + d_S(\widehat{S}_n(\omega), S(P)) \\
&\le d_S(S_j, \widehat{S}_n(\omega)) + d_S(\widehat{S}_n(\omega), S(P)) \\
&\le d_S(S_j, S(P)) + 2 d_S(\widehat{S}_n(\omega), S(P)) \\
&\le n^{-1/2} + 2 d_S(\widehat{S}_n(\omega), S(P)).
\end{aligned}$$

By Equations (3) and (4),

$$\widehat{\nu}_n^{\mathrm{net}}(\omega) = \nu_{Q_i}, \qquad S(Q_i) = S_i.$$

Applying the modulus from Theorem 1 gives

$$W_1\big(\widehat{\nu}_n^{\mathrm{net}}(\omega), \nu_P\big) \le C_{\mathrm{mod}} \left\{ 2 d_S(\widehat{S}_n(\omega), S(P)) + n^{-1/2} \right\}.$$

14. Fix $0 < \eta < 1/2$. Since $C \geq C_0$, $C \geq \exp(1)$, and $\eta \leq 1$,

$$1 \leq \log(C/\eta), \qquad \log(C_0/\eta) \leq \log(C/\eta).$$

Therefore

$$n^{-1/2} \leq \sqrt{\frac{\log(C/\eta)}{n}}, \qquad \sqrt{\frac{\log(C_0/\eta)}{n}} \leq \sqrt{\frac{\log(C/\eta)}{n}}.$$

On the event

$$d_S(\widehat{S}_n, S(P)) \leq C_0 L \sqrt{\frac{\log(C_0/\eta)}{n}},$$

the deterministic inequality gives

$$W_1(\widehat{\nu}_n^{\mathrm{net}}, \nu_P) \leq C_{\mathrm{mod}}(2C_0L + 1)\sqrt{\frac{\log(C/\eta)}{n}} \leq C\sqrt{\frac{\log(C/\eta)}{n}}.$$

Thus the bad event in the theorem is contained in

$$\left\{ d_S(\widehat{S}_n, S(P)) > C_0 L \sqrt{\frac{\log(C_0/\eta)}{n}} \right\}.$$

By Lemma 1, its $Q_P^{(n)}$-probability is at most η.

□

References

Chunrong Ai and Jiawei Shan. Efficient estimation of average treatment effects with unmeasured confounding and proxies, 2025.

Elizabeth S. Allman, Catherine Matias, and John A. Rhodes. Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics*, 37(6A): 3099–3132, 2009. doi: 10.1214/09-AOS689.

Animashree Anandkumar, Rong Ge, Daniel Hsu, Sham M. Kakade, and Matus Telgarsky. Tensor decompositions for learning latent variable models. *Journal of Machine Learning Research*, 15(80):2773–2832, 2014. URL https://jmlr.org/papers/v15/anandkumar14b.html.

Friedrich L. Bauer and Charles T. Fike. Norms and exclusion theorems. *Numerische Mathematik*, 2:137–141, 1960. doi: 10.1007/BF01386217.

Xin Bing, Florentina Bunea, and Jonathan Niles-Weed. Estimation and inference for the wasserstein distance between mixing measures in topic models. *Bernoulli*, 2026. doi: 10.3150/24-BEJ1813.

Stéphane Bonhomme, Koen Jochmans, and Jean-Marc Robin. Non-parametric estimation of finite mixtures from repeated measurements. *Journal of the Royal Statistical Society: Series B*, 78(1):211–229, 2016. doi: 10.1111/rssb.12110.

Lawrence D. Brown and Roger Purves. Measurable selections of extrema. *The Annals of Statistics*, 1(5):902–912, 1973. doi: 10.1214/aos/1176342510.

Lucien Le Cam. *Asymptotic Methods in Statistical Decision Theory*. Springer Series in Statistics. Springer, New York, 1986. ISBN 9780387963075. doi: 10.1007/978-1-4612-4946-7.

Yifan Cui, Hongming Pu, Xu Shi, Wang Miao, and Eric J. Tchetgen Tchetgen. Semiparametric proximal causal inference. *Journal of the American Statistical Association*, 119(546):1348–1359, 2024. doi: 10.1080/01621459.2023.2191817.

Chandler Davis and William M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970. doi: 10.1137/0707001.

Neil Deo and Thibault Randrianarisoa. On adaptive confidence sets for the wasserstein distances. *Bernoulli*, 29(3):2119–2141, 2023. doi: 10.3150/22-BEJ1535.

Natalie Doss, Yihong Wu, Pengkun Yang, and Harrison H. Zhou. Optimal estimation of high-dimensional gaussian location mixtures. *The Annals of Statistics*, 51(1):62–95, 2023. doi: 10.1214/22-AOS2207.

Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054, 1982. doi: 10.2307/1912775.

Philippe Heinrich and Jonas Kahn. Strong identifiability and optimal minimax rates for finite mixture estimation. *The Annals of Statistics*, 46(6A):2844–2870, 2018. doi: 10.1214/17-AOS1641.

Nhat Ho and XuanLong Nguyen. Singularity structures and impacts on parameter estimation in finite mixtures of distributions, 2016a.

Nhat Ho and XuanLong Nguyen. Convergence rates of parameter estimation for some weakly identifiable finite mixtures. *The Annals of Statistics*, 44(6):2726–2755, 2016b. doi: 10.1214/16-AOS1444.

Paul W. Holland. Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960, 1986. doi: 10.1080/01621459.1986.10478354.

Daniel Hsu, Sham M. Kakade, and Tong Zhang. A spectral algorithm for learning hidden markov models. In *Proceedings of the 22nd Annual Conference on Learning Theory*, 2009.

Yingyao Hu. Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution. *Journal of Econometrics*, 144(1):27–61, 2008. doi: 10.1016/j.jeconom.2007.12.001.

Yingyao Hu and Susanne M. Schennach. Instrumental variable treatment of nonclassical measurement error models. *Econometrica*, 76(1):195–216, 2008. doi: 10.1111/j.0012-9682.2008.00823.x.

Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, Cambridge, 2015. ISBN 9780521885881.

Hiroyuki Kasahara and Katsumi Shimotsu. Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica*, 77(1):135–175, 2009. doi: 10.3982/ECTA6763.

Tosio Kato. *Perturbation Theory for Linear Operators*. Classics in Mathematics. Springer, Berlin, Heidelberg, second edition, 1995. ISBN 978-3-540-58661-6. doi: 10.1007/978-3-642-66282-9.

Kendrick Li, George C. Linderman, Xu Shi, and Eric J. Tchetgen Tchetgen. Regression-based proximal causal inference for right-censored time-to-event data, 2024.

Bruce G. Lindsay. *Mixture Models: Theory, Geometry, and Applications*, volume 5 of *NSF-CBMS Regional Conference Series in Probability and Statistics*. Institute of Mathematical Statistics and American Statistical Association, Hayward, CA, 1995. ISBN 9780940600324. URL https://projecteuclid.org/euclid.cbms/1462106013.

Jiewen Liu, Chan Park, Kendrick Li, and Eric J. Tchetgen Tchetgen. Regression-based proximal causal inference, 2024.

Tudor Manole and Nhat Ho. Refined convergence rates for maximum likelihood estimation under finite mixture models. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 14979–15006, 2022.

Bijan Mazaheri, Chandler Squires, and Caroline Uhler. Synthetic potential outcomes and causal mixture identifiability. In *Proceedings of The 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pages 4276–4284, 2025. doi: 10.48550/arXiv.2405.19225.

Wang Miao, Zhi Geng, and Eric J. Tchetgen Tchetgen. Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 105(4):987–993, 2018. doi: 10.1093/biomet/asy038.

Wang Miao, Wenjie Hu, Elizabeth L. Ogburn, and Xiaohua Zhou. Identifying effects of multiple treatments in the presence of unmeasured confounding. *Journal of the American Statistical Association*, 118(543):1953–1967, 2023. doi: 10.1080/01621459.2021.2023551.

Huy Nguyen, Dung Le, Alessandro Rinaldo, and Nhat Ho. On the geometry of separation in finite gaussian mixtures, 2026.

XuanLong Nguyen. Convergence of latent mixing measures in finite and infinite mixture models. *The Annals of Statistics*, 41(1):370–400, 2013. doi: 10.1214/12-AOS1065.

Jonathan Niles-Weed and Philippe Rigollet. Estimation of wasserstein distances in the spiked transport model. *Bernoulli*, 28(4):2663–2688, 2022. doi: 10.3150/21-BEJ1433.

Karl Pearson. Iii. contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London A*, 185:71–110, 1894. doi: 10.1098/rsta.1894.0003.

Yury Polyanskiy and Yihong Wu. Dualizing Le Cam's method for functional estimation, with applications to estimating the unseens, 2019.

Zhengling Qi, Rui Miao, and Xiaoke Zhang. Proximal learning for individualized treatment regimes under unmeasured confounding. *Journal of the American Statistical Association*, 119 (546):915–928, 2024. doi: 10.1080/01621459.2022.2147841.

James M. Robins and Aad W. van der Vaart. Adaptive nonparametric confidence sets. *The Annals of Statistics*, 34(1):229–253, 2006. doi: 10.1214/009053605000000877.

Donald B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974. doi: 10.1037/h0037350.

Aytijhya Saha, Stephen Bates, and Devavrat Shah. Causal inference with categorical unobserved confounder via mixture learning, 2026.

Xu Shi, Wang Miao, Jennifer C. Nelson, and Eric J. Tchetgen Tchetgen. Multiply robust causal inference with double-negative control adjustment for categorical unmeasured confounding. *Journal of the Royal Statistical Society: Series B*, 82(2):521–540, 2020. doi: 10.1111/rssb.12361.

Nicholas D. Sidiropoulos and Rasmus Bro. On the uniqueness of multilinear decomposition of N-way arrays. *Journal of Chemometrics*, 14(3):229–239, 2000. doi: 10.1002/1099-128X(200005/06)14:3⟨229::AID-CEM587⟩3.0.CO;2-N.

Jerzy Splawa-Neyman, D. M. Dabrowska, and T. P. Speed. On the application of probability theory to agricultural experiments: Essay on principles. section 9. *Statistical Science*, 5(4): 465–472, 1990. doi: 10.1214/ss/1177012031. Translated and edited by D. M. Dabrowska and T. P. Speed; original work published in 1923.

G. W. Stewart and Ji-guang Sun. *Matrix Perturbation Theory*. Academic Press, Boston, 1990. ISBN 9780126702309.

Eric J. Tchetgen Tchetgen, Andrew Ying, Yifan Cui, Xu Shi, and Wang Miao. An introduction to proximal causal inference. *Statistical Science*, 39(3):375–390, 2024. doi: 10.1214/23-STS911.

Henry Teicher. Identifiability of finite mixtures. *The Annals of Mathematical Statistics*, 34(4): 1265–1269, 1963. doi: 10.1214/aoms/1177703862.

Wouter A. C. van Amsterdam, Joost J. C. Verhoeff, Netanja I. Harlianto, Gijs A. Bartholomeus, Anudeep M. Puli, Pim A. de Jong, Tim Leiner, Alette S. R. van Lindert, Marinus J. C. Eijkemans, and Rajesh Ranganath. Individual treatment effect estimation in the presence of unobserved confounding using proxies: A cohort study in stage III non-small cell lung cancer. *Scientific Reports*, 12(1):5848, 2022. doi: 10.1038/s41598-022-09775-9.

Hamza Virk, Bijan Mazaheri, and Yihren Wu. The spectral structure of latent treatment effects, 2026.

Yihong Wu and Pengkun Yang. Optimal estimation of gaussian mixtures via denoised method of moments. *The Annals of Statistics*, 48(4):1981–2007, 2020. doi: 10.1214/19-AOS1873.