Honest expected length for transported complier effects with weak first stages

CausalSmith*

August 25, 2026

Abstract

This paper studies honest confidence sets for a transported complier effect in a two-sample encouragement design. Outcomes, receipt, encouragement, and covariates are observed in a source population, while covariates are observed in a target population. Under the transported complier-effect model class with binary receipt, bounded outcomes, instrument overlap, monotonicity, and target-to-source covariate domination, the target complier effect θ_T, the target complier-conditional outcome contrast, is identified as a transported reduced-form ratio and lies in the compact causal range $\Theta = [-1, 1]$. The expected-length frontier is governed by the effective strength $t_n = n\mu_n^2/\kappa_n$, where μ_n is the transported first-stage mean and κ_n is the Kish second moment of the transport weights. For every fixed threshold $t_0 > 0$, oracle honest confidence sets have minimax expected length of order $\min\{1, t_0^{-1/2}\}$, both globally and conditional on each admissible deterministic source-covariate, transport-weight, and propensity geometry. The lower bound is matched by an oracle Anderson–Rubin/Fieller score inversion. In uniform finite-cell designs, empirical target cell frequencies yield a score inversion that is sample-only in the uniform class with balanced assignment, while the regular nonuniform extension additionally uses known source cell probabilities and a known cell-varying propensity, the covariate being observed as a finite-cell label and the cell count obeying $k_n = o(\sqrt{n})$, attaining the same order; the same conclusion extends to regular nonuniform source cells with known cell probabilities and known cell-varying propensity.

1 Introduction

Transported complier effects combine two sources of statistical fragility. The first is the ratio structure of the local average treatment effect: the causal contrast for compliers is recovered by dividing an outcome reduced form by a first-stage receipt contrast [Imbens and Angrist, 1994, Angrist et al., 1996]. When the first stage is weak, conventional ratio inference can behave poorly, motivating Anderson–Rubin and Fieller inversions and the weak-IV literature on procedures with reliable coverage under small denominators [Anderson and Rubin, 1949, Fieller, 1954, Staiger and Stock, 1997, Stock et al., 2002, Moreira, 2003, Andrews et al., 2006, Mikusheva, 2010, Andrews et al., 2019]. The second is transport: the source encouragement design is reweighted to represent a target covariate distribution, so the effective information

*Machine-generated by the CausalSmith research pipeline. The displayed formal statements and proofs are Lean-verified under the paper's theorem-local citation interface; where a result from the literature is a formal dependency, it enters as a published input rather than a Lean-checked step, and cited work otherwise supplies framing and positioning.

in the source sample depends on the dispersion of the target-to-source density ratio [Cole and Stuart, 2010, Tipton, 2013, Dahabreh et al., 2020b, Chen and Huang, 2025, Zeng et al., 2024].

This paper characterizes the minimax expected-length rate order, up to constants, of honest confidence sets for the transported complier ratio, at each fixed effective-strength threshold and within the oracle, fixed-geometry, uniform finite-cell, and known regular-cell classes studied here. The source sample contains encouragement, treatment receipt, outcomes, and covariates; the target sample contains covariates. The maintained model class in Definition 1 imposes the standard binary encouragement-design conditions, conditional transport of the outcome reduced form, integrated transport of the first stage, target complier positivity, target-to-source covariate domination, and deterministic controls on transport-weight dispersion. The target parameter is the target complier-conditional outcome contrast. Under these conditions, Theorem 1 identifies the parameter as

$$\theta_T = \frac{\mu_{Y,n}}{\mu_n}$$

and places it in the compact range $\Theta = [-1, 1]$, where $\mu_{Y,n}$ is the transported reduced-form outcome mean and μ_n is the transported first-stage mean.

The main rate index is

$$t_n = \frac{n\mu_n^2}{\kappa_n},$$

where $\kappa_n = \mathbb{E}_S[w(X)^2]$ is the Kish dispersion of the transport weights. This scalar combines the source sample size, the transported first-stage magnitude, and the unevenness of target reweighting. A design with the same first-stage mean has less effective strength when the target population is represented by more dispersed source weights. In the no-shift case $w(X) = 1$, Theorem 4 gives $\kappa_n = 1$ and reduces the scale to $n\mu_n^2$.

The oracle results give matching lower and upper expected-length rates. The oracle class in Definition 3 allows the procedure to use the target-to-source density ratio and the source assignment propensity. This construction serves as an oracle benchmark for the general model; feasible target-weight learning is established in the finite-cell classes of Theorems 6 and 7. For every fixed effective-strength threshold $t_0 > 0$, Theorem 2 gives a lower bound of order $\min\{1, t_0^{-1/2}\}$ for oracle honest procedures. Theorem 3 matches this order with a score inversion for the transported moment of $Y - \vartheta D$. The resulting confidence set keeps a candidate value $\vartheta \in \Theta$ when the transported empirical score is within a radius proportional to $\sqrt{\widehat{\kappa}_n/n}$. This is the transported complier analogue of weak-ratio inversion in the Anderson–Rubin and Fieller tradition.

The same frontier holds after conditioning on deterministic geometry. The admissible geometry class in Definition 8 fixes the source covariate marginal, the transport weight, the target covariate law induced by that weight, and the source propensity. Within a fixed-geometry slice, the causal contrasts and the first-stage strength vary subject to the transported model restrictions. Theorem 5 establishes the same two-sided order $\min\{1, t_0^{-1/2}\}$ uniformly over admissible deterministic geometries. This conditional statement isolates the weak-first-stage and outcome-tilt components of the problem from the deterministic transport geometry.

The feasible part of the paper studies finite-cell target-weight learning. In the uniform finite-cell submodel of Definition 2, the source covariate law assigns equal mass to k_n cells and the source propensity is balanced. The target covariate sample estimates the target cell probabilities, and the source sample estimates cellwise outcome and receipt contrasts. Theorem 6 shows that the resulting score inversion using sample-supplied target weights is honest over the uniform finite-cell class and attains the oracle order $\min\{1, t_0^{-1/2}\}$. Theorem 7 extends the construction to regular nonuniform source cells when the source cell probabilities and cell-varying propensity are known.

The analysis connects transported CACE identification and estimation [Rudolph and Laan, 2016, Chen and Huang, 2025] with weak-identification-robust ratio inference [Gleser and Hwang, 1987, Dufour, 1997, Choi et al., 2018, Choi and Shen, 2019, Ma, 2023, Smucler et al., 2025]. It also complements transport and generalizability theory for average treatment effects [Joseph Hotz et al., 2005, Pearl and Bareinboim, 2011, Bareinboim and Pearl, 2013, 2016, Dahabreh et al., 2020a, Nguyen et al., 2017, Zeng et al., 2024] by treating the transported denominator as part of the identification problem. The finite-cell constructions estimate target-cell weights from the target covariate sample while retaining the compact-range and weak-ratio features of the transported complier estimand.

The formal derivations state their assumptions and conclusions explicitly.

The rest of the paper proceeds as follows. The related-work section places the results in the literatures on complier effects, weak ratio inference, and transportability. The setup section defines the two-sample transported encouragement design, the model classes, the target estimand, and the expected-length criterion. The oracle minimax section proves the global and fixed-geometry frontiers and gives the oracle score inversion. The cell-weight-learning section gives the uniform and regular finite-cell feasible procedures. The discussion section interprets the effective-strength scale and records directions for future work.

2 Related work

2.1 Comparison with the closest results

The nearest results are best placed by naming, for each, the estimand, the sampling design, the first-stage regime, the status of the transport weights, the geometry of the parameter space, and the guarantee delivered.

Chen and Huang [2025] study the transported complier average causal effect in a two-sample generalizability design, with a first stage bounded away from zero, transport weights estimated as nuisances, an unrestricted parameter space, and asymptotic normality with a consistent variance estimator as the guarantee. The present paper adopts that estimand and design and extends the guarantee to honest coverage with a two-sided expected-length rate that remains meaningful as the transported first stage degenerates.

Choi et al. [2018] and Choi and Shen [2019] study structural coefficients in two-sample instrumental variables within a single population, under weak identification, on unbounded parameter spaces, delivering weak-identification-robust tests and the confidence sets obtained by inverting them. The transported problem here adds a covariate shift, and the dispersion of the transport weight enters the rate as a Kish effective-sample-size deflation, which is the specific mechanism this paper isolates.

Ma [2023] and Smucler et al. [2025] construct weak-identification-robust confidence sets for the local average treatment effect in a single population, the former with machine-learning nuisances, and deliver coverage guarantees. This paper contributes the matching expected-length frontier, in the transported setting, alongside coverage.

Zeng et al. [2024] develop minimax theory for transported average treatment effects, where the functional is point identified under strong overlap and the rates concern estimation error. Here the estimand is a ratio whose denominator is permitted to vanish, and the compact parameter space forced by bounded outcomes and binary receipt is what keeps the minimax expected length finite.

This paper belongs to the literature on causal effects for compliers in encouragement designs. The local average treatment effect framework of Imbens and Angrist [1994] and Angrist et al. [1996] formalizes the ratio of an intention-to-treat outcome contrast to a first-stage receipt contrast under random assignment, exclusion, and monotonicity, building on randomized encouragement designs such as Bloom [1984]. Principal stratification gives the same object a broader potential-outcome interpretation, including bounds and strata-specific estimands in settings with post-assignment behavior [Balke and Pearl, 1997, Imbens and Rubin, 1997, Frangakis and Rubin, 2002]. Selection and marginal-treatment-effect approaches provide related ways to organize heterogeneity in treatment take-up and response [Heckman and Vytlacil, 2005]. The analysis here keeps the complier-ratio target and studies its transported version when outcomes, receipt, and encouragement are observed in a source population and covariates are observed in a target population.

A second line of work develops inference for ratios under weak first stages. Anderson–Rubin and Fieller methods provide classical confidence sets for ratio parameters [Anderson and Rubin, 1949, Fieller, 1954]; Gleser and Hwang [1987] and Dufour [1997] characterize fundamental constraints on bounded expected length and uniform confidence coverage for ratio-type problems. Weak-instrument asymptotics and robust IV procedures were developed by Bound et al. [1995], Staiger and Stock [1997], Stock et al. [2002], Moreira [2003], Andrews et al. [2006], Mikusheva [2010], and Andrews et al. [2019], with modern refinements including Bertanha and Moreira [2020]. Recent work on LATE confidence sets under weak identification includes Ma [2023] and Smucler et al. [2025]. The present paper uses the same ratio-inference logic to study honest expected length for a bounded transported complier effect, with the first-stage strength summarized together with transport-weight dispersion.

The transportability and generalizability literature studies how causal conclusions move from an observed population to a target population. Early econometric work on external validity includes Joseph Hotz et al. [2005]; graphical and causal-transport formulations are developed by Pearl and Bareinboim [2011], Bareinboim and Pearl [2013], and Bareinboim and Pearl [2016]. Weighting and sampling-score approaches for trial generalization and transport appear in Cole and Stuart [2010], Stuart et al. [2011], Tipton [2013], Dahabreh et al. [2020b], Dahabreh et al. [2020a], and Nguyen et al. [2017]. Efficiency and minimax theory for transported average treatment effects is studied by Zeng et al. [2024]. This paper focuses on the transported complier ratio, where the denominator is itself a transported first-stage contrast and the target population enters through a covariate-density ratio.

The closest transported-noncompliance benchmarks are Rudolph and Laan [2016] and Chen and Huang [2025]. Rudolph and Laan [2016] analyze transported stochastic direct and indirect effects in encouragement-design settings, while Chen and Huang [2025] study transported CACE identification and regular estimation. Related two-sample weak-IV work by Choi et al. [2018] and Choi and Shen [2019] develops weak-instrument-robust inference when information is split across samples. Recent work by Ross et al. [2026] and Ren [2025] studies transported adherence and extrapolated LATE-type targets, while Rudolph et al. [2025] contributes recent efficiency theory for transported average treatment effects. The contribution here is the fixed-geometry honest expected-length characterization for bounded transported complier-ratio confidence sets, together with oracle and finite-cell target-weight-learning procedures attaining the same rate order.

3 Setup and assumptions

All objects are indexed by the source sample size n, and limits are taken along the triangular array. Generic positive constants may change from line to line unless they are named in a formal statement. Expectations and probabilities are evaluated under the law indicated by the subscript, and equalities involving conditional laws are understood up to the corresponding almost-sure version. For a set $C \subseteq [-1,1]$, $\lambda(C)$ denotes its Lebesgue length restricted to the displayed interval. The symbols introduced below are reused throughout the paper.

3.1 Full-data, sampling, and instrument structure

At each index, the two-population full-data potential-outcome world is carried by the full-data law P_n^F. The population indicator S distinguishes the source and target populations; X denotes covariates; $D(z)$ and $Y(d)$ denote treatment-receipt and outcome potential variables. The first condition fixes the bounded binary structure used by the ratio problem.

Assumption 1. [`ass:full-data-support`] (Full-data support) *The full-data vector is* $(S, X, D(0), D(1), Y(0), Y(1))$, *where* X *is the covariate taking values in* $\mathcal{X}$. *The full-data law* P_n^F *satisfies*

$$P_n^F\{S \in \{0,1\},\ D(0), D(1) \in \{0,1\},\ Y(0), Y(1) \in [0,1]\} = 1.$$

Assumption 1 is specific to this analysis: it encodes binary receipt, bounded outcomes, and the two-population indexing in a single support condition. The bounded outcome scale is the source of the compact causal range established below.

Assumption 2. [`ass:population-presence`] (Population presence) *For each* $s \in \{0,1\}$,

$$P_n^F(S = s) > 0.$$

Assumption 2 is the standard two-population support condition, ensuring that both source and target conditional laws are well defined [Chen and Huang, 2025]. Sampling then proceeds from the observed source law P_S and target covariate law P_T.

Assumption 3. [`ass:two-sample-array`] (Two-sample array) *The source observed law* P_S, *target covariate law* P_T, *target sample size* N_n, *and asymptotic target-to-source sample-size ratio* c *satisfy:*

- ***(Source law.)*** *For every* n, P_S *at array index* n *is a probability measure on the source observation* O_i^S.
- ***(Target law.)*** *For every* n, P_T *at array index* n *is a probability measure on the target covariate* X_j^T.
- ***(Sample-size ratio.)*** $0 < c$ *and* $N_n/n \to c$.

The observed two-sample triangular-array sampling world comprises source observations O_i^S, target covariate draws X_j^T, target sample size N_n, and sample-size limit c. Assumption 3 is the standard independent two-sample triangular-array sampling condition for transported CACE settings [Chen and Huang, 2025].

The next group imposes the source encouragement-design restrictions. Let $e(X)$ be the source assignment propensity and ε the overlap constant.

Assumption 4. [`ass:instrument-overlap`] (Instrument overlap) *For the source assignment propensity $e(X)$ and instrument-overlap constant ε, $0 < \varepsilon < 1/2$, and*

$$\varepsilon \leq e(X) \leq 1 - \varepsilon$$

P_S^X-almost surely, where P_S^X denotes the covariate marginal of the source law P_S.

Assumption 4 is the standard instrument positivity condition [Chen and Huang, 2025]. It keeps the inverse-propensity source contrasts controlled over the source covariate marginal P_S^X.

Assumption 5. [`ass:source-observation`] (Source observation) *Conditional on $S = 1$, the observed source law satisfies:*

- ***(Assignment propensity.)*** *$P(Z = 1 \mid X, S = 1) = e(X)$.*
- ***(Treatment receipt.)*** *$D = D(Z)$.*
- ***(Observed outcome.)*** *$Y = Y(D)$.*

With source encouragement Z, observed receipt D, and observed outcome Y, Assumption 5 is the standard source assignment and consistency condition [Chen and Huang, 2025]. It connects the observed source variables to the potential-outcome objects in Assumption 1.

Assumption 6. [`ass:iv-randomization`] (Instrument randomization) *Conditional on $(X, S = 1)$,*

$$Z \perp (Y(0), Y(1), D(0), D(1)).$$

Assumption 6 is the standard conditional instrument exchangeability condition [Chen and Huang, 2025]. It makes the source encouragement contrast interpretable within covariate strata.

Assumption 7. [`ass:iv-exclusion`] (Instrument exclusion) *For each $z, s \in \{0, 1\}$, under the full-data law P_n^F,*

$$Y^Z(z) = Y(D(z))$$

almost surely conditional on $S = s$.

The assignment-generated outcome $Y^Z(z)$ satisfies the standard instrument exclusion restriction in Assumption 7 [Chen and Huang, 2025]. The restriction ties assignment effects on outcomes to assignment effects on treatment receipt.

Assumption 8. [`ass:iv-monotonicity`] (Instrument monotonicity) *For each $s \in \{0, 1\}$, under the full-data law P_n^F,*

$$D(1) \geq D(0)$$

almost surely conditional on $S = s$.

Assumption 8 is the standard monotonicity condition for the encouragement design [Chen and Huang, 2025]. Together, Assumptions 5 to 8 supply the IV structure used to read source reduced-form and receipt contrasts as complier-weighted quantities.

3.2 Transport and dispersion

Transport is organized through the covariate space $\mathcal{X}$ and the density ratio $w(X)$ from the target covariate law to the source covariate marginal. The transport and overlap geometry world collects the source marginal, density ratio, and propensity array that govern extrapolation from the source sample to the target population.

Assumption 9. [`ass:outcome-transport`] (Outcome transport) *The source conditional outcome contrast is*

$$\Delta_Y(x) = \mathbb{E}_{P_n^F}\left[Y^Z(1) - Y^Z(0) \mid X = x, S = 1\right].$$

For P_T-almost every x, where P_T is the target covariate law,

$$\mathbb{E}_{P_n^F}\left[Y^Z(1) - Y^Z(0) \mid X = x, S = 0\right] = \Delta_Y(x).$$

The transported reduced-form outcome mean is $\mu_{Y,n} := \mathbb{E}_{X|S=0}[\Delta_Y(X)]$.

The source conditional outcome contrast $\Delta_Y(x)$ is transported conditionally in Assumption 9. This is the standard conditional transport of the assignment-outcome contrast condition [Chen and Huang, 2025].

Assumption 10. [`ass:receipt-transport`] (Receipt transport) *The source conditional receipt contrast is*

$$\Delta_D(x) = \mathbb{E}_{P_n^F}[D(1) - D(0) \mid X = x, S = 1].$$

Under the full-data law P_n^F,

$$\mathbb{E}_{X|S=0}\Big[\mathbb{E}_{P_n^F}[D(1) - D(0) \mid X, S = 0]\Big] = \mathbb{E}_{X|S=0}[\Delta_D(X)].$$

This common value is the transported first-stage mean $\mu_n := \mathbb{E}_{X|S=0}[\Delta_D(X)]$.

The source conditional receipt contrast $\Delta_D(x)$ enters through the integrated first-stage transport condition in Assumption 10. This is the standard integrated transport of the first stage [Chen and Huang, 2025], matching the target average receipt contrast to the transported source contrast.

Assumption 11. [`ass:target-complier-positivity`] (Target complier positivity) *Under the full-data law P_n^F,*

$$P_n^F(D(1) > D(0) \mid S = 0) > 0.$$

Assumption 11 is the standard positive target complier share condition [Chen and Huang, 2025]. It gives a positive denominator for the transported complier ratio.

Assumption 12. [`ass:transport-domination`] (Transport domination) *The target covariate law P_T and the source covariate marginal P_S^X are laws on the covariate space $\mathcal{X}$, and P_T is absolutely continuous with respect to P_S^X:*

$$P_T \ll P_S^X.$$

We write

$$w := \frac{dP_T}{dP_S^X}$$

for the resulting target-to-source covariate density ratio.

Assumption 12 is the standard transport covariate-support inclusion condition [Chen and Huang, 2025]. It provides the density-ratio representation used by the transported moments.

The remaining conditions are specific to this analysis. They encode the weight-dispersion and weak-first-stage array along which the expected-length results are evaluated.

Assumption 13. [`ass:weight-envelope`] (Weight envelope) *The target-to-source covariate density ratio satisfies*

$$0 \leq w(X) \leq 2k_n$$

P_S-almost surely, where k_n is the transport-weight envelope scale.

Assumption 13 is specific to this analysis. It bounds the pointwise transport weight by the envelope and cell-count scale k_n, allowing target reweighting while keeping a finite deterministic envelope.

Assumption 14. [`ass:weight-second-moment`] (Weight second moment) *The second moment of the transport weight,*

$$\kappa_n := \mathbb{E}_S\left[w(X)^2\right],$$

satisfies

$$\kappa_n \leq k_n.$$

The second moment κ_n is the Kish dispersion scale of the transport weights. Assumption 14 is specific to this analysis and links dispersion to the same scale that controls the weight envelope.

Assumption 15. [`ass:degrading-array`] (Degrading array rates) *The array satisfies the following rate conditions:*

- ***(Envelope scale.)*** *$k_n \to \infty$.*
- ***(Sampling scale.)*** *$k_n = o(n^{1/2})$.*
- ***(First-stage strength.)*** *$\mu_n \to 0$.*

The transported first-stage mean μ_n is allowed to drift toward zero under Assumption 15. The condition is specific to this analysis and defines the weak-first-stage asymptotic regime while keeping the weight scale sub-root in the source sample size.

Assumption 16. [`ass:finite-cell-source`] (Uniform finite-cell source) *The source covariate marginal P_S^X is supported on exactly k_n measurable cells, assigns mass $1/k_n$ to each cell, and satisfies*

$$e(X) = \frac{1}{2}$$

P_S-almost surely.

Assumption 16 is specific to this analysis. It defines the uniform finite-cell design used for the feasible target-weight-learning construction, with equal source cell masses and a known balanced assignment propensity.

3.3 Model classes, estimand, and risk

The main model class collects the support, sampling, IV, transport, and dispersion restrictions above.

Definition 1. [`def:model-class`] (Transported complier-effect model class $\mathcal{P}_n$) *For each n, the transported complier-effect model class is*

$$\mathcal{P}_n = \left\{ P = (P_n^F, e) : P \text{ satisfies the conditions below at index } n \right\}.$$

The conditions are:

- ***(Full-data and population structure.)*** *P satisfies Assumptions 1 and 2.*
- ***(Sampling structure.)*** *P satisfies Assumption 3.*
- ***(Instrument and observation structure.)*** *P satisfies Assumptions 4 to 8.*
- ***(Transport and target-complier structure.)*** *P satisfies Assumptions 9 to 12.*
- ***(Envelope and degradation structure.)*** *P satisfies Assumptions 13 to 15.*

The transported complier-effect model class $\mathcal{P}_n$ is the domain for the oracle confidence-set analysis. Its elements bundle a full-data law and source propensity satisfying the restrictions at a fixed array index.

Definition 2. [`def:finite-cell-class`] (Finite-cell submodel $\mathcal{N}_n$) *For each n, the finite-cell submodel is*

$$\mathcal{N}_n = \left\{ P = (P_n^F, e) : P \text{ satisfies the conditions below at index } n \right\}.$$

The conditions are:

- ***(Full-data and population structure.)*** *P satisfies Assumptions 1 and 2.*
- ***(Sampling structure.)*** *P satisfies Assumption 3.*
- ***(Instrument and observation structure.)*** *P satisfies Assumptions 4 to 8.*
- ***(Transport and target-complier structure.)*** *P satisfies Assumptions 9 to 12.*
- ***(Envelope and degradation structure.)*** *P satisfies Assumptions 13 to 15.*
- ***(Finite-cell source structure.)*** *P satisfies Assumption 16.*

The finite-cell submodel $\mathcal{N}_n$ adds the uniform source-cell structure in Assumption 16. It is the benchmark for procedures that learn target weights from the covariate-only target sample.

Before defining the expected-length criterion, we record the identification and range statement supplied by the assumptions. The identification and effective-strength scalar world uses the transported reduced-form outcome mean $\mu_{Y,n}$, first-stage mean μ_n, and target complier contrast θ_T.

Theorem 1. [`prop:compact-causal-range`] (Compact causal range) *Fix n, P, N_n, k_n, c, and the instrument-overlap constant ε. Suppose that:*

- ***(Model class.)*** *P belongs to the transported complier-effect model class $\mathcal{P}_n$ of Definition 1.*
- ***(Source contrasts.)*** *The source conditional outcome contrast $\Delta_Y(x)$ and the source conditional receipt contrast $\Delta_D(x)$ are measurable and integrable with respect to the source covariate marginal P_S^X.*

Then Δ_Y and Δ_D represent the observed source contrasts: for every measurable covariate set A,

$$\int_{\{O^S:X\in A\}} H(O^S)Y\,dP_S = \int_A \Delta_Y(x)\,dP_S^X(x), \qquad \int_{\{O^S:X\in A\}} H(O^S)D\,dP_S = \int_A \Delta_D(x)\,dP_S^X(x).$$

Moreover, with

$$\mu_{Y,n} = \int \big(Y(1)-Y(0)\big)\mathbf{1}\{D(1)=1,\ D(0)=0\}\,dP_n^F(\,\cdot\mid S=0)$$

and

$$\mu_n = P_n^F\{D(1)=1,\ D(0)=0 \mid S=0\},$$

the transported reduced-form and first-stage means satisfy

$$\int w(x)\Delta_Y(x)\,dP_S^X(x) = \mu_{Y,n}, \qquad \int w(x)\Delta_D(x)\,dP_S^X(x) = \mu_n.$$

Consequently,

$$\frac{\mu_{Y,n}}{\mu_n} = \theta_T = \mathbb{E}_{P_n^F}[Y(1)-Y(0)\mid D(1)=1,\ D(0)=0,\ S=0] \in \Theta = [-1,1].$$

Theorem 1 establishes that the transported source reduced-form and first-stage contrasts identify the target complier-conditional contrast, and that the contrast lies in the compact parameter space $[-1,1]$. The intuition is direct: bounded potential outcomes place every complier outcome contrast in the same bounded interval, while monotone binary receipt turns the transported first stage into the target complier share.

The confidence-set definitions use the compact parameter space Θ. The procedure and minimax-risk world treats a confidence-set sequence as a random subset of this space, with the oracle version allowed to use the transport weights and assignment propensity.

Definition 3. [`def:oracle-honesty`] (Oracle honesty class $\mathfrak{C}^{\mathrm{or}}$, estimand θ_T, and space Θ) *Let the target complier estimand and compact parameter space be*

$$\theta_T = \frac{\mu_{Y,n}}{\mu_n}, \qquad \Theta = [-1,1].$$

For a noncoverage level α, the oracle honest confidence-set class is

$$\mathfrak{C}^{\mathrm{or}} = \left\{(C_n)_{n\ge 1} : C_n = C_n\Big((O_i^S)_{i=1}^n, (X_j^T)_{j=1}^{N_n}, w, e\Big) \subseteq \Theta,\ \liminf_{n\to\infty}\inf_{P\in\mathcal{P}_n} P\{\theta_T\in C_n\} \ge 1-\alpha\right\}.$$

The oracle class $\mathfrak{C}^{\mathrm{or}}$ formalizes honest coverage uniformly over $\mathcal{P}_n$ for random sets C_n. The noncoverage level is α.

Definition 4. [`def:frontier-risk`] (Frontier risk $R((C_n), t_0)$) *For a confidence set $C_n \subseteq \Theta$, let $\lambda(C_n)$ denote its Lebesgue length restricted to Θ. The effective identification strength is*

$$t_n = \frac{n\mu_n^2}{\kappa_n}.$$

Fix a threshold $t_0 > 0$ and let $(C_n)_{n\geq 1}$ be an oracle procedure: each

$$C_n = C_n\Big((O_i^S)_{i=1}^n, (X_j^T)_{j=1}^{N_n}, w, e\Big) \subseteq \Theta$$

is formed from the observed samples together with the transport weight w and the propensity e. For each fixed measurable pair (w, e), the sample-by-parameter graph $\{(sample, \vartheta) : \vartheta \in C_n\}$ is measurable; and under each law in $\mathcal{P}_n$, replacing w by another version of that law's transport weight, evaluated at the true propensity, leaves C_n almost surely unchanged. The frontier risk of $(C_n)_{n\geq 1}$ at t_0 is

$$R((C_n), t_0) = \limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n : t_n \geq t_0} \mathbb{E}_P[\lambda(C_n)].$$

Here the supremum over an empty set of laws is zero, so a sample size at which no law meets the strength restriction contributes nothing to the frontier risk.

The effective identification strength t_n combines source sample size, transported first-stage magnitude, and weight dispersion. For a fixed threshold t_0, the risk $R((C_n), t_0)$ measures limiting worst-case expected length over laws whose effective strength exceeds that threshold.

Definition 5. [`def:oracle-value`] (Oracle value $V^\star(t_0)$)

$$V^\star(t_0) = \inf_{(C_n)\in\mathfrak{C}^{\mathrm{or}}} R((C_n), t_0).$$

The oracle value $V^\star(t_0)$ is the minimax expected length over oracle-honest procedures. It is the population-weight-known benchmark for the rate results in the next section.

Definition 6. [`def:finite-cell-oracle-value`] (Finite-cell oracle value $V_{\mathcal{N}}^\star(t_0)$)

$$V_{\mathcal{N}}^\star(t_0) = \inf_{(C_n)\in\mathfrak{C}^{\mathrm{or}}} \limsup_{n\to\infty} \sup_{P\in\mathcal{N}_n : t_n\geq t_0} \mathbb{E}_P[\lambda(C_n)].$$

The finite-cell oracle value $V_{\mathcal{N}}^\star(t_0)$ evaluates the same oracle procedures on the finite-cell submodel. It separates the statistical effect of the finite-cell geometry from the additional task of learning target weights.

Definition 7. [`def:finite-cell-feasible-honesty`] (Finite-cell feasible class $\mathfrak{C}^{\mathrm{cell}}$)

$$\mathfrak{C}^{\mathrm{cell}} = \left\{(C_n)_{n\geq 1} : C_n = C_n\Big((O_i^S)_{i=1}^n, (X_j^T)_{j=1}^{N_n}\Big) \subseteq \Theta,\ \liminf_{n\to\infty} \inf_{P\in\mathcal{N}_n} P\{\theta_T \in C_n\} \geq 1-\alpha\right\}.$$

The feasible finite-cell class $\mathfrak{C}^{\mathrm{cell}}$ restricts procedures to the observed source sample and target covariate sample. This class is the setting for the finite-cell target-weight-learning construction in Theorem 6.

3.4 Reading the design restrictions

The feasible classes carry restrictions worth separating into design features, approximations, and technical regimes. Uniform source cells of mass $1/k_n$ and balanced assignment describe a stratified experiment in which the sampling frame fixes equal-sized strata and randomizes within them; this is a design the experimenter can implement, not an assumption about nature. In the regular-cell class the cell probabilities $q_{x,n}$ and the cell-varying propensity are supplied by that same frame — they are known because the design records them — so the result should be read as covering registry- or survey-based settings where stratum sizes and assignment rates are administrative facts rather than quantities to be estimated. The growth condition $k_n = o(\sqrt{n})$ is the technical regime: it keeps the number of cells small enough that per-cell frequencies concentrate fast enough for the collision estimator to learn the target weights, and it is the condition under which learning the weights costs nothing in rate. The theorems accordingly cover designs with known strata, supplied assignment probabilities, and cell counts obeying $k_n = o(\sqrt{n})$, the regime in which the cell-frequency and collision terms calibrate at the oracle rate.

4 Oracle minimax rates

The expected-length criterion in Definition 4 reduces the transported weak-first-stage problem to the scalar

$$t_n = \frac{n\mu_n^2}{\kappa_n}.$$

The numerator is the squared transported first stage scaled by the source sample size, and the denominator is the Kish dispersion of the target-to-source weights. Thus a design with the same transported first stage can be statistically weaker when the target population is represented by a more uneven reweighting of the source covariate distribution.

The oracle benchmark treats the target-to-source density ratio and source assignment propensity as known. This leads to an Anderson–Rubin and Fieller style inversion of a transported reduced-form score [Anderson and Rubin, 1949, Fieller, 1954]. The candidate value is retained when the weighted source score is small relative to the oracle weight-dispersion radius.

Algorithm 1 (Score-inversion set C_n)

With oracle-known w and e, define C_n by the following source-sample inversion procedure.

1. For each $\vartheta \in \Theta$, define the inverse-propensity score

$$\psi_i(\vartheta) = w(X_i)\left\{\frac{Z_i}{e(X_i)} - \frac{1-Z_i}{1-e(X_i)}\right\}(Y_i - \vartheta D_i).$$

2. Define the empirical score mean

$$\widehat{m}_n(\vartheta) = \frac{1}{n}\sum_{i=1}^{n}\psi_i(\vartheta).$$

3. Define the oracle critical radius

$$r_n = L_\alpha \left\{ \frac{1}{n} \cdot \frac{1}{n} \sum_{i=1}^{n} w(X_i)^2 \right\}^{1/2},$$

where L_α is the critical constant supplied to the procedure at noncoverage level α.

4. Output the score-inversion confidence set

$$C_n = \{\vartheta \in \Theta : |\widehat{m}_n(\vartheta)| \le r_n\}.$$

Algorithm 1 is the transported analogue of weak-IV score inversion. For a fixed candidate ϑ, the score tests whether the transported reduced-form contrast is compatible with ϑ times the transported first-stage contrast. The critical radius has order $\sqrt{\widehat{\kappa}_n/n}$, so it expands with the empirical second moment of the oracle weights.

The lower bound applies to the oracle value over the full transported model and to the finite-cell oracle value. It shows that the same effective-strength index governing the score radius also governs the minimax expected length.

Theorem 2. [`thm:oracle-converse`] (Oracle lower frontier) *Fix $c > 0$, an instrument-overlap constant $\varepsilon \in (0, 1/2)$, and a noncoverage level $\alpha \in (0, 1)$. There exist constants $c_0 > 0$ and $t_c > 0$, depending only on (α, ε, c), such that the following holds for every measurable covariate space $\mathcal{X}$.*

Let N_n and k_n be integer sequences satisfying:

- ***(Target-source ratio.)*** *$N_n/n \to c$.*
- ***(Positive cells.)*** *$k_n > 0$ for every n.*
- ***(Diverging cells.)*** *$k_n \to \infty$.*
- ***(Sub-root growth.)*** *$k_n/\sqrt{n} \to 0$.*
- ***(Measurable carrier.)*** *For each n, $\mathcal{X}$ contains a measurable k_n-cell carrier.*
- ***(Transported-IV model.)*** *Every array in the transported complier-effect class $\mathcal{P}_n$ of Definition 1 satisfies Assumptions 1 to 15.*
- ***(Finite-cell source.)*** *Every array in the finite-cell class $\mathcal{N}_n$ of Definition 2 satisfies Assumption 16.*

Then, for every effective-strength threshold $t_0 > 0$,

$$c_0 \min\{1, t_0^{-1/2}\} \le V^\star(t_0)$$

and

$$c_0 \min\{1, t_0^{-1/2}\} \le V^\star_{\mathcal{N}}(t_0),$$

where $V^\star(t_0)$ and $V^\star_{\mathcal{N}}(t_0)$ are the oracle values in Definitions 5 and 6. Moreover, whenever $0 < t_0 \le t_c$,

$$c_0 \le V^\star(t_0).$$

The converse in Theorem 2 gives the information-theoretic side of the rate. Above a fixed threshold t_0, honest expected length is bounded below by the smaller of a constant and $t_0^{-1/2}$. At very small effective strength, the final display gives the order-one elbow: honest oracle procedures retain nonvanishing worst-case expected length.

The matching upper bound is attained by the oracle-known-weight score-inversion rule, which takes the true transport-weight version and the source propensity as oracle inputs. Writing

$$H_i = \frac{Z_i}{e(X_i)} - \frac{1 - Z_i}{1 - e(X_i)},$$

the procedure can be expressed as an inversion of the weighted reduced-form mean against the weighted first-stage mean.

Theorem 3. [`thm:oracle-score-inversion-attainment`] (Oracle score inversion) *Let N_n and k_n be integer sequences, and let c, ε, and α be constants. Suppose that:*

- ***(Sampling ratio.)*** *$c > 0$ and $N_n/n \to c$.*
- ***(Overlap and coverage constants.)*** *$0 < \varepsilon < 1/2$ and $0 < \alpha < 1$.*
- ***(Envelope scale.)*** *$k_n > 0$ for every n, $k_n \to \infty$, and $k_n/\sqrt{n} \to 0$.*
- ***(Admissible geometry.)*** *The measurable covariate space admits a deterministic geometry $\mathfrak{g}$ satisfying Definition 8 with envelope scale k_n and overlap constant ε.*
- ***(Model-class implications.)*** *For every n and every transported array P, membership of P in the transported complier-effect model class $\mathcal{P}_n$ of Definition 1 implies Assumptions 3, 4 and 13 to 15.*

Define

$$L_\alpha = \left(\frac{8}{\alpha\varepsilon^2}\right)^{1/2}, \qquad H_i = \frac{Z_i}{e(X_i)} - \frac{1 - Z_i}{1 - e(X_i)},$$

and

$$\widehat{A}_n = \frac{1}{n}\sum_{i=1}^{n} w(X_i) H_i Y_i, \qquad \widehat{B}_n = \frac{1}{n}\sum_{i=1}^{n} w(X_i) H_i D_i, \qquad \widehat{\kappa}_n = \frac{1}{n}\sum_{i=1}^{n} w(X_i)^2.$$

Let the score-inversion confidence set be

$$C_n = \left\{\vartheta \in \Theta : \left|\widehat{A}_n - \vartheta \widehat{B}_n\right| \le L_\alpha \sqrt{\widehat{\kappa}_n / n}\right\}.$$

Then the sequence (C_n) is oracle honest, $(C_n) \in \mathfrak{C}^{\mathrm{or}}$ in the sense of Definition 3, and for every fixed effective-strength threshold $t_0 > 0$,

$$R((C_n), t_0) \le C_0 \min\{1, t_0^{-1/2}\}, \qquad C_0 = \max\left\{2, 4L_\alpha + \frac{8}{\varepsilon^2}\right\}.$$

Consequently, for the oracle minimax value $V^\star(t_0)$ of Definition 5,

$$\frac{3(1-\alpha)^2}{16} \min\{1, t_0^{-1/2}\} \le V^\star(t_0) \le C_0 \min\{1, t_0^{-1/2}\},$$

and for every admissible geometry $\mathfrak{g}' \in \mathcal{G}$ the fixed-geometry value of Definition 11 obeys the same two-sided bounds,

$$\frac{3(1-\alpha)^2}{16} \min\{1, t_0^{-1/2}\} \le V^\star_{\mathfrak{g}'}(t_0) \le C_0 \min\{1, t_0^{-1/2}\}.$$

Theorem 3 supplies the constructive half of the oracle frontier. The set is always restricted to $\Theta = [-1, 1]$, and its random width is controlled by the same dispersion-normalized standard-error scale that appears in t_n. Combining Theorems 2 and 3 yields the matched oracle rate $\min\{1, t_0^{-1/2}\}$ for $V^\star(t_0)$, up to constants depending on the stated coverage, overlap, and sampling-ratio inputs.

It is useful to separate the causal array from the deterministic geometry that carries the source covariate law, transport weights, and source propensity. The next definitions fix this geometry and evaluate honest expected length conditionally on it.

Definition 8. [`def:admissible-geometry-class`] (Admissible geometry class $\mathcal{G}$) *$\mathcal{G}$ is the class of deterministic arrays*

$$\mathfrak{g} = (P_S^X, w, e)_{n\geq 1}$$

such that, for each n,

- ***(Source marginal.)*** *P_S^X is a probability law on $\mathcal{X}$.*
- ***(Instrument overlap.)*** *The source assignment propensity satisfies*

$$e : \mathcal{X} \to [\varepsilon, 1-\varepsilon].$$

- ***(Transport weights.)*** *The target-to-source covariate density ratio satisfies*

$$w : \mathcal{X} \to [0, 2k_n], \qquad \mathbb{E}_S[w(X)] = 1.$$

- ***(Dispersion.)*** *The second moment of the transport weight is*

$$\kappa_n = \mathbb{E}_S[w(X)^2] \leq k_n.$$

- ***(Target law.)*** *For every measurable $A \subseteq \mathcal{X}$,*

$$P_T(A) = \mathbb{E}_S[w(X)\mathbf{1}\{X \in A\}].$$

Definition 8 records the deterministic ingredients shared by the lower and upper bounds: an overlapped assignment propensity, a normalized target-to-source density ratio, a second-moment bound, and the induced target covariate law. Holding these objects fixed isolates the weak-first-stage and outcome-tilt variation inside the transported model class.

Definition 9. [`def:fixed-geometry-slice`] (Fixed-geometry slice $\mathcal{P}_n(\mathfrak{g})$) *For $\mathfrak{g} \in \mathcal{G}$,*

$$\mathcal{P}_n(\mathfrak{g}) = \{P \in \mathcal{P}_n : (P_S^X, w, e) \text{ equals the index-}n\text{ component of } \mathfrak{g}\}.$$

Definition 9 restricts attention to laws whose source covariate marginal, transport weight, and propensity match the chosen geometry. The causal contrasts and first-stage strength vary within that slice subject to the model restrictions.

Definition 10. [`def:fixed-geometry-honesty`] (Fixed-geometry oracle class $\mathfrak{C}^{\mathrm{or}}(\mathfrak{g})$) *For $\mathfrak{g} \in \mathcal{G}$ as in Definition 8,*

$$\mathfrak{C}^{\mathrm{or}}(\mathfrak{g}) = \left\{(C_n)_{n\geq 1} : C_n = C_n\Big((O_i^S)_{i=1}^n, (X_j^T)_{j=1}^{N_n}, w, e\Big) \subseteq \Theta,\ \liminf_{n\to\infty} \inf_{P\in\mathcal{P}_n(\mathfrak{g})} P(\theta_T \in C_n) \geq 1-\alpha\right\}.$$

The fixed-geometry honesty class in Definition 10 requires uniform asymptotic coverage only over the slice in Definition 9. The procedure remains oracle in the same sense as Definition 3: it may use the fixed weight and propensity inputs.

Definition 11. [`def:fixed-geometry-value`] (Fixed-geometry value $V^\star_{\mathfrak{g}}(t_0)$) *For $\mathfrak{g} \in \mathcal{G}$ as in Definition 8,*

$$V^\star_{\mathfrak{g}}(t_0) = \inf_{(C_n)\in\mathfrak{C}^{\mathrm{or}}(\mathfrak{g})} \limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n(\mathfrak{g}):\, t_n\geq t_0} \mathbb{E}_P[\lambda(C_n)].$$

The value in Definition 11 is the conditional minimax expected length at effective-strength threshold t_0. It keeps the weighting and overlap geometry fixed while retaining the same length and strength criteria used in Definition 4.

A transparent special case is the absence of covariate shift. When the density ratio is identically one, the target and source covariate laws coincide and the effective strength becomes the familiar sample-size-scaled squared first stage.

Theorem 4. [`prop:no-shift-reduction`] (No-shift reduction) *Fix a measurable covariate space and deterministic arrays N_n, k_n, and $\mathfrak{g}$. Suppose:*

- ***(Sample-size ratio.)*** *$c > 0$ and $N_n/n \to c$.*
- ***(Coverage and overlap constants.)*** *$0 < \alpha < 1$ and $0 < \varepsilon < 1/2$.*
- ***(Cell-scale regime.)*** *For every n, $k_n > 0$, with $k_n \to \infty$ and $k_n/\sqrt{n} \to 0$.*
- ***(Admissible geometry.)*** *The geometry $\mathfrak{g}$ belongs to the admissible class $\mathcal{G}$ in Definition 8.*
- ***(No covariate shift.)*** *For every n, $w_n(X) = 1$ $P^X_{S,n}$-almost surely.*

Then, for every n, the geometry has

$$\kappa_n = \int w_n(x)^2\, dP^X_{S,n}(x) = 1$$

and $P_{T,n} = P^X_{S,n}$. Moreover, for every transported array P^F_n in the fixed-geometry slice $\mathcal{P}_n(\mathfrak{g})$ of Definition 9,

$$t_n = \frac{n\mu_n^2}{\kappa_n} = n\mu_n^2.$$

Finally, for every $t_0 > 0$, define

$$L_\alpha = \sqrt{\frac{8}{\alpha\varepsilon^2}}, \qquad c_\alpha = \frac{3(1-\alpha)^2}{16}, \qquad C_\alpha = \max\left\{2,\, 4L_\alpha + \frac{8}{\varepsilon^2}\right\}.$$

The fixed-geometry minimax value $V^\star_{\mathfrak{g}}(t_0)$ in Definition 11 satisfies

$$c_\alpha \min\{1, t_0^{-1/2}\} \leq V^\star_{\mathfrak{g}}(t_0) \leq C_\alpha \min\{1, t_0^{-1/2}\}.$$

Theorem 4 anchors the transported rate to the single-population weak-ratio scale. With $w = 1$, there is no Kish deflation and the denominator of t_n is one. The displayed bounds then express the same order-one elbow and $t_0^{-1/2}$ decay in terms of $n\mu_n^2$.

The fixed-geometry theorem gives the conditional version of the full oracle characterization. It shows that the lower and upper bounds hold uniformly over admissible deterministic geometry arrays.

Theorem 5. [`thm:fixed-geometry-frontier`] (Fixed-geometry length frontier) *Let $c > 0$, let N_n be target sample sizes with $N_n/n \to c$, let k_n be positive integers with $k_n \to \infty$ and $k_n/\sqrt{n} \to 0$, let $\varepsilon \in (0, 1/2)$, and let $\alpha \in (0, 1)$. Let $\mathfrak{g}$ be a deterministic geometry array satisfying Definition 8 at scale k_n and overlap parameter ε. Assume that, for every n, every law in the fixed-geometry slice $\mathcal{P}_n(\mathfrak{g})$ of Definition 9 satisfies*

- ***(Two samples.)*** *Assumption 3 with target-to-source ratio c.*
- ***(Overlap.)*** *Assumption 4 with overlap parameter ε.*
- ***(Envelope.)*** *Assumption 13 at scale k_n.*
- ***(Second moment.)*** *Assumption 14 at scale k_n.*
- ***(Degradation.)*** *Assumption 15 at scale k_n.*

Define

$$L_\alpha = \left(\frac{8}{\alpha\varepsilon^2}\right)^{1/2}, \qquad c_\alpha = \frac{3(1-\alpha)^2}{16}, \qquad C_\alpha = \max\{2, 4L_\alpha + 8\varepsilon^{-2}\}.$$

Then, for every fixed effective-strength threshold $t_0 > 0$, the fixed-geometry minimax expected-length value $V_{\mathfrak{g}}^\star(t_0)$ of Definition 11 satisfies

$$c_\alpha \min\{1, t_0^{-1/2}\} \le V_{\mathfrak{g}}^\star(t_0) \le C_\alpha \min\{1, t_0^{-1/2}\}.$$

The constants c_α and C_α depend only on α and ε, uniformly over admissible deterministic geometry arrays $\mathfrak{g}$.

Theorem 5 establishes the same rate after conditioning on the source-covariate, transport-weight, and propensity array. The lower bound treats Kish dispersion as part of the information scale, and the upper bound is attained by the same score radius $L_\alpha\sqrt{\widehat{\kappa}_n/n}$ used in Theorem 3. Together, the oracle, no-shift, and fixed-geometry results pin the minimax expected-length rate order, up to constants, as a function of the effective-strength threshold, within the procedure classes stated.

5 Cell weight learning

The oracle procedure in Theorem 3 uses the target-to-source weights as inputs. This section turns that benchmark into a feasible finite-cell construction. In the uniform finite-cell experiment of Definition 2, the target covariate sample estimates the cell probabilities directly, and the source sample estimates the cellwise reduced-form and first-stage contrasts. The resulting score inversion has the same expected-length order as the oracle benchmark, with the dispersion scale estimated from target-cell coincidences. This connects transported noncompliance inference to the weighting and two-sample transport ideas in Rudolph and Laan [2016], Chen and Huang [2025], and Zeng et al. [2024], while retaining the weak-ratio inversion logic used in robust IV procedures [Choi et al., 2018, Choi and Shen, 2019].

Theorem 6. [`thm:finite-cell-unknown-weight-attainment`] (Finite-cell sample-only attainment) *Let N_n and k_n be positive integer sequences, let $c, \varepsilon, \alpha \in \mathbb{R}$, and suppose the following conditions hold.*

- ***(Sampling ratio.)*** *$c > 0$ and $N_n/n \to c$.*
- ***(Overlap and coverage levels.)*** *$0 < \varepsilon < 1/2$ and $0 < \alpha < 1$.*
- ***(Cell growth.)*** *$k_n > 0$ for every n, $k_n \to \infty$, and $k_n/\sqrt{n} \to 0$.*
- ***(Ambient carrier.)*** *For every n, the measurable covariate space contains an injected copy of $\{1, \ldots, k_n\}$ whose singleton images are measurable.*
- ***(Finite-cell rows.)*** *For every row n and every transported array $P \in \mathcal{N}_n$, with $\mathcal{N}_n$ as in Definition 2, the row satisfies the two-sample condition in Assumption 3, the uniform finite-cell source condition in Assumption 16, and the degrading-array condition in Assumption 15.*

Put

$$B_c = 32(1 + c^{-1}), \qquad L_{\alpha,c} = \left(\frac{2B_c}{\alpha}\right)^{1/2}, \qquad C_0 = \max\{2, 4\sqrt{2}\, L_{\alpha,c} + 8B_c\}.$$

Then $C_0 > 0$, and there exists a sample-only finite-cell confidence-set sequence $C = (C_n)$ in the feasible class $\mathfrak{C}^{\mathrm{cell}}$ of Definition 7 such that, for every n and every $P \in \mathcal{N}_n$, C_n agrees P-almost surely under the two-sample law with the finite-cell score-inversion rule using critical constant $L_{\alpha,c}$. Equivalently, on the finite-cell experiment, this rule is given by

$$H_i = 2(2Z_i - 1), \qquad \widehat{p}_{x,n} = \frac{1}{N_n} \sum_{j=1}^{N_n} \mathbf{1}\{X_j^T = x\},$$

and, for $G \in \{Y, D\}$,

$$\widehat{m}_{G,n}(x) = \frac{k_n}{n} \sum_{i=1}^{n} \mathbf{1}\{X_i = x\} H_i G_i, \qquad \widehat{M}_{G,n} = \sum_{x=1}^{k_n} \widehat{p}_{x,n} \widehat{m}_{G,n}(x).$$

For $N_n \geq 2$, define

$$\widehat{\kappa}_{U,n} = \frac{k_n}{N_n(N_n - 1)} \sum_{j \neq \ell} \mathbf{1}\{X_j^T = X_\ell^T\}, \qquad \widehat{K}_n = 1 + \widehat{\kappa}_{U,n},$$

and set

$$C_n = \left\{ \vartheta \in \Theta : |\widehat{M}_{Y,n} - \vartheta \widehat{M}_{D,n}| \leq L_{\alpha,c} \sqrt{\widehat{K}_n / n} \right\},$$

with $C_n = \Theta$ at indices for which $N_n < 2$. For every $t_0 > 0$,

$$\limsup_{n \to \infty} \sup_{P \in \mathcal{N}_n : t_n \geq t_0} \mathbb{E}_P[\lambda(C_n)] \leq C_0 \min\{1, t_0^{-1/2}\}.$$

Moreover, for every $t_0 > 0$, the feasible finite-cell honest minimax risk satisfies

$$\frac{3(1-\alpha)^2}{16} \min\{1, t_0^{-1/2}\} \leq \inf_{\widetilde{C} \in \mathfrak{C}^{\mathrm{cell}}} \limsup_{n \to \infty} \sup_{P \in \mathcal{N}_n : t_n \geq t_0} \mathbb{E}_P[\lambda(\widetilde{C}_n)].$$

Theorem 6 establishes that empirical target cell frequencies can be substituted for oracle transport weights in the uniform finite-cell design while preserving the order $\min\{1, t_0^{-1/2}\}$. The statistic $\widehat{p}_{x,n}$ learns the target mass of each cell; $\widehat{m}_{G,n}(x)$ estimates the source encouragement contrast in that cell; and $\widehat{M}_{G,n}$ transports those source contrasts to the target population. The U-statistic $\widehat{\kappa}_{U,n}$ estimates the collision component of the weight dispersion, so the feasible radius uses the same strength scale that governs the oracle expected length.

The lower bound in Theorem 6 matches the oracle order from Theorem 2 on the same finite-cell submodel. Thus the target-weight-learning step contributes a fully sample-based construction whose honest expected length remains governed by the effective-strength threshold. The condition $k_n/\sqrt{n} \to 0$ is the cell-growth regime under which the theorem calibrates the source-cell contrast error and the target-frequency error at the displayed rate.

The regular extension allows the source cells to have known nonuniform probabilities. This formulation keeps the source cell masses and the assignment propensity available to the procedure, matching settings in which the sampling frame or design supplies these quantities, as in standard inverse-probability and semiparametric weighting constructions [Hirano et al., 2003, Robins et al., 1994, van der Laan and Rubin, 2006, Chernozhukov et al., 2018].

Definition 12. [`def:regular-finite-cell-class`] (Regular finite-cell class $\mathcal{N}_n^{\mathrm{reg}}$)

$$\mathcal{N}_n^{\mathrm{reg}} = \left\{ P \in \mathcal{P}_n : \begin{array}{l} P_S^X \textit{ is supported on exactly } k_n \textit{ measurable cells,} \\ q_{x,n} = P_S^X(X \textit{ lies in cell } x), \\ c_-/k_n \le q_{x,n} \le c_+/k_n \textit{ for every } x \end{array} \right\},$$

where $0 < c_- \le 1 \le c_+ < \infty$ *are fixed.*

Definition 12 replaces equal source-cell masses with regular cell masses bounded above and below by constants times $1/k_n$. The known probabilities $q_{x,n}$ let the source contrast estimator normalize each cell by its actual source mass, while the constants c_- and c_+ keep the finite-cell geometry comparable across cells.

Definition 13. [`def:regular-cell-feasible-honesty`] (Regular-cell feasible honesty class $\mathfrak{C}^{\mathrm{reg}}$)

$$\mathfrak{C}^{\mathrm{reg}} = \left\{ (C_n)_{n \ge 1} : C_n = C_n\left((O_i^S)_{i=1}^n, (X_j^T)_{j=1}^{N_n}, (q_{x,n})_{x=1}^{k_n}, e \right) \subseteq \Theta,\ \liminf_{n\to\infty} \inf_{P \in \mathcal{N}_n^{\mathrm{reg}}} P(\theta_T \in C_n) \ge 1 - \alpha \right\}.$$

Definition 13 records the information available to the regular-cell procedure: the two samples, the known source-cell probabilities, and the propensity. Honesty is uniform over the regular finite-cell class, using the same compact parameter space and target complier contrast introduced in Definition 3 and Theorem 1.

Theorem 7. [`thm:regular-cell-unknown-weight-attainment`] (Regular-cell score attainment) *Fix $c, \varepsilon, \alpha, c_-, c_+ \in \mathbb{R}$, target sample sizes N_n, and cell counts k_n. Suppose that*

- ***(Sample-size ratio.)*** $0 < c$ *and* $N_n/n \to c$.
- ***(Overlap and coverage levels.)*** $0 < \varepsilon < 1/2$ *and* $0 < \alpha < 1$.
- ***(Regular cell constants.)*** $0 < c_- \le 1$ *and* $1 \le c_+$.

- ***(Cell growth.)*** *$k_n > 0$ for every n, $k_n \to \infty$, and $k_n/\sqrt{n} \to 0$.*
- ***(Measurable carrier.)*** *For every n, the covariate carrier contains k_n injected measurable singleton cells.*
- ***(Model restrictions.)*** *Every array $P \in \mathcal{N}_n^{\mathrm{reg}}$ from Definition 12 with constants c, ε, c_-, c_+ satisfies Assumption 3, Assumption 4, and Assumption 15.*

On $\mathcal{N}_n^{\mathrm{reg}}$, let $q_{x,n} = P_S^X(X = x)$ and define

$$H_i = \frac{Z_i}{e(X_i)} - \frac{1 - Z_i}{1 - e(X_i)}.$$

For $G \in \{Y, D\}$, set

$$\widehat{p}_{x,n} = \frac{1}{N_n} \sum_{j=1}^{N_n} \mathbf{1}\{X_j^T = x\}, \qquad \widehat{m}_{G,n}(x) = \frac{1}{n q_{x,n}} \sum_{i=1}^{n} \mathbf{1}\{X_i = x\} H_i G_i,$$

and

$$\widehat{M}_{G,n} = \sum_{x=1}^{k_n} \widehat{p}_{x,n} \widehat{m}_{G,n}(x).$$

For $N_n \geq 2$, set

$$\widehat{\kappa}_{U,n} = \frac{1}{N_n(N_n - 1)} \sum_{j \neq \ell} \frac{\mathbf{1}\{X_j^T = X_\ell^T\}}{q_{X_j^T,n}}, \qquad \widehat{K}_n = 1 + \widehat{\kappa}_{U,n},$$

and put $C_n^{\mathrm{reg}} = \Theta$ when $N_n < 2$. Let

$$B_{\varepsilon,c} = 8(\varepsilon^{-2} + c^{-1}).$$

For every $L \geq \{2B_{\varepsilon,c}/\alpha\}^{1/2}$, define

$$C_0 = \max\{2, 4\sqrt{2}L + 8B_{\varepsilon,c}\}.$$

Then $0 < C_0$, and there exists a regular-cell procedure $C = (C_n)_{n\geq 1}$ such that, for every n and every $P \in \mathcal{N}_n^{\mathrm{reg}}$, C_n agrees almost surely under the two-sample law with

$$C_n^{\mathrm{reg}} = \left\{ \vartheta \in \Theta : |\widehat{M}_{Y,n} - \vartheta \widehat{M}_{D,n}| \leq L\sqrt{\widehat{K}_n/n} \right\}$$

for $N_n \geq 2$, together with the preceding convention for $N_n < 2$. This procedure belongs to $\mathfrak{C}^{\mathrm{reg}}$ from Definition 13. Moreover, for every fixed $t_0 > 0$,

$$\limsup_{n\to\infty} \sup_{P \in \mathcal{N}_n^{\mathrm{reg}} : t_n \geq t_0} \mathbb{E}_P[\lambda(C_n)] \leq C_0 \min\{1, t_0^{-1/2}\}.$$

For every fixed $t_0 > 0$, the minimax risk over $\mathfrak{C}^{\mathrm{reg}}$ satisfies

$$\frac{3(1-\alpha)^2}{16} \min\{1, t_0^{-1/2}\} \leq \inf_{C \in \mathfrak{C}^{\mathrm{reg}}} \limsup_{n\to\infty} \sup_{P \in \mathcal{N}_n^{\mathrm{reg}} : t_n \geq t_0} \mathbb{E}_P[\lambda(C_n)].$$

Thus, with known regular source-cell probabilities and a known cell-varying propensity satisfying the stated overlap, the feasible regular-cell construction attains the oracle frontier order on $\mathcal{N}_n^{\mathrm{reg}}$.

Theorem 7 extends the sample-only construction to regular nonuniform source cells. The inverse-propensity score H_i now uses the known cell-varying assignment propensity, and the source contrast estimator rescales observations by $q_{x,n}$. The dispersion estimator similarly weights target-cell coincidences by the known source mass of the matched cell. Under the displayed regularity and growth conditions, the resulting score-inversion confidence set is honest over $\mathcal{N}_n^{\mathrm{reg}}$ and has worst-case expected length bounded by a constant times $\min\{1, t_0^{-1/2}\}$.

In the uniform finite-cell class and in the known regular nonuniform-cell class with known source cell probabilities and known cell-varying propensity, Theorems 6 and 7 show that empirical target-cell frequencies and collision dispersion estimates attain the oracle expected-length order. The finite-cell procedures use the target sample only through empirical cell frequencies and dispersion estimates, while the source sample supplies the outcome and receipt contrasts. In both finite-cell classes, the effective-strength scalar t_n remains the quantity that determines the transition between the compact-range regime and the square-root contraction regime.

6 Discussion and limitations

The results identify the effective-strength scalar $t_n = n\mu_n^2/\kappa_n$ as the organizing quantity for honest expected length in the transported complier-ratio problem. The lower and upper bounds in Theorems 2 and 3 give the order

$$\min\{1, t_0^{-1/2}\}$$

for oracle honest procedures above a fixed threshold t_0. Thus the expected-length curve has an order-one elbow: near the weak-identification region, the compact parameter range governs the achievable length, while stronger effective first stages support square-root contraction in the threshold scale.

The compact range is a substantive part of the conclusion rather than a normalization artifact. Theorem 1 identifies the target complier-conditional contrast as the ratio of the transported reduced-form outcome mean to the transported first-stage mean and places that contrast in $\Theta = [-1, 1]$. This bounded target space gives honest procedures a finite fallback length when the transported first stage is weak. It also makes the expected-length criterion in Definition 4 interpretable on the causal scale of outcome contrasts, with $\lambda(C_n)$ measuring the part of the confidence set that lies in the feasible causal range.

The role of transport enters through the Kish dispersion κ_n. Fixed instrument overlap, encoded in Assumption 4, controls the source inverse-propensity scores, while Assumptions 13 and 14 control the size and second moment of the target-to-source weights. The resulting scalar $t_n = n\mu_n^2/\kappa_n$ captures the way transported covariate reweighting reduces effective source information. Theorem 4 makes this interpretation explicit: when $w(X) = 1$, the target covariate law equals the source covariate marginal, $\kappa_n = 1$, and the effective-strength scale reduces to $n\mu_n^2$.

The oracle and feasible constructions also have the same statistical shape. The oracle score inversion in Algorithm 1 and Theorem 3 tests each candidate ratio by centering the transported score for $Y - \vartheta D$, following the Anderson–Rubin and Fieller logic for weak ratios. The finite-cell procedures in Theorems 6 and 7 replace oracle weights with empirical target cell frequencies and use target-sample collision statistics to estimate the dispersion scale. Their matched lower and upper bounds show that, in the stated finite-cell designs, in the uniform finite-cell class with balanced assignment, and in the regular nonuniform-cell class with known cell probabilities and known cell-varying propensity, learning the target weights attains the same honest minimax expected-length order as the oracle benchmark.

6.1 Limitations and future work

The feasible weight-learning results are stated for the uniform finite-cell class in Definition 2 and the known regular nonuniform-cell class in Definition 12. Extending the same honest expected-length characterization to richer feasible weighting schemes, including continuous-covariate approximations and data-adaptive partitions, remains an open direction.

The finite-cell procedures use the assignment propensity information specified by their model classes. Future work can study joint learning of the source assignment propensity and target transport weights while preserving weak-ratio honesty. Such an extension would connect the present score-inversion approach with semiparametric nuisance-estimation tools, while retaining the finite-sample caution emphasized in weak-identification analysis [Armstrong and Kolesár, 2021].

The analysis fixes the effective-strength threshold t_0 when evaluating limiting expected length. Moving-threshold regimes, sharp constants, and refinements of the order-one elbow are natural next targets. These questions would sharpen the comparison between compact-range behavior and square-root contraction without changing the core scale $n\mu_n^2/\kappa_n$.

Appendices

A Proofs, auxiliary arguments, and verification note

This appendix collects the auxiliary statements used by the identification, oracle, fixed-geometry, and finite-cell arguments in the order in which the main text relies on them. The first group supplies the fixed-geometry least-favorable family and the inhabitation facts that make the lower-bound comparisons well posed on the stated rows.

The lower-bound argument works with a fixed deterministic geometry and varies the complier probability and outcome tilt inside that geometry. The construction below pins down the law used for that comparison.

Definition 14. [`def:least-favorable-witness`] (Least-favorable law $Q^{\mathfrak{g}}_{n,h}$) *For* $\mathfrak{g} = (P_S^X, w, e)_{n\geq 1} \in \mathcal{G}$*,* $t_0 > 0$*, and* $|h| \leq H(t_0)$*,* $Q^{\mathfrak{g}}_{n,h}$ *is the joint observed-data law generated by the following construction. Set*

$$\mu_n = \left(\frac{t_0\kappa_n}{n}\right)^{1/2}, \qquad p_n(x) = \frac{\mu_n w(x)}{\kappa_n}, \qquad \rho = \frac{1-\alpha}{8}, \qquad H(t_0) = \min\{1/4, \rho t_0^{-1/2}\},$$

where κ_n *is the second moment of the transport weight and* α *is the noncoverage level. For a source observation* $O^S = (X, Z, D, Y)$*, define the inverse-propensity instrument score*

$$H(O^S) = \frac{Z}{e(X)} - \frac{1-Z}{1-e(X)}.$$

Conditional on $X = x$*, assign compliance types by*

$$\Pr\{(D(0), D(1)) = (0,1) \mid X = x\} = p_n(x),$$

$$\Pr\{(D(0), D(1)) = (1,1) \mid X = x\} = \Pr\{(D(0), D(1)) = (0,0) \mid X = x\} = \frac{1-p_n(x)}{2}.$$

Set $Y(0) = 0$. For compliers,

$$Y(1) \sim \mathrm{Bernoulli}(1/2 + h),$$

and for always-takers and never-takers,

$$Y(1) \sim \mathrm{Bernoulli}(1/2).$$

The same conditional full-data law is used in both populations, source encouragement is drawn with propensity $e(X)$, and the resulting joint observed-data law together with the fixed oracle inputs is denoted by $Q^{\mathfrak{g}}_{n,h}$.

Definition 14 fixes the local experiment behind the oracle converse. The complier probability is proportional to the transport weight, so the transported first stage is calibrated to the effective-strength threshold, while the scalar tilt h moves the target complier contrast within the compact causal range.

Definition 15. [`def:geometry-handle`] (Geometry handle at (n, t_0, α)) *Fix n, a threshold $t_0 > 0$, a noncoverage level $0 < \alpha < 1$, and an admissible geometry array $\mathfrak{g} = (P_S^X, w, e) \in \mathcal{G}$, and set*

$$\mu_n = \left(\frac{t_0 \kappa_n}{n}\right)^{1/2}, \qquad p_n(x) = \frac{\mu_n w(x)}{\kappa_n}, \qquad \rho = \frac{1-\alpha}{8}.$$

Say that $\mathfrak{g}$ admits a geometry handle at (n, t_0, α) when all of the following hold.

- ***(Valid compliance probabilities.)*** *$0 \le p_n(x) \le 1$ for every x.*
- ***(Transported first stage.)*** *$\int w(x) p_n(x)\, dP_S^X(x) = \mu_n$.*
- ***(Exact strength.)*** *$n\mu_n^2/\kappa_n = t_0$.*
- ***(Least-favourable family.)*** *There is a family $h \mapsto Q^{\mathfrak{g}}_{n,h}$ such that for every $h \in \mathbb{R}$ with*

  $$|h| \le \min\{1/4, \rho t_0^{-1/2}\},$$

  the law $Q^{\mathfrak{g}}_{n,h}$ is a least-favourable fixed-geometry witness in the sense of Definition 14 and satisfies, relative to the member $Q^{\mathfrak{g}}_{n,0}$ of that family at $h = 0$,

  $$1 + \chi^2(Q^{\mathfrak{g}}_{n,h} \| Q^{\mathfrak{g}}_{n,0}) \le \exp(8 t_0 h^2), \qquad \|Q^{\mathfrak{g}}_{n,h} - Q^{\mathfrak{g}}_{n,0}\|_{\mathrm{TV}} \le 2\rho.$$

The handle in Definition 15 packages the validity and distance calculations needed for the Le Cam comparison. Its chi-square and total-variation bounds are the probability-distance inputs used to translate honest coverage into expected-length lower bounds, following the weak-ratio logic associated with Anderson–Rubin and Fieller comparisons [Anderson and Rubin, 1949, Fieller, 1954, Gleser and Hwang, 1987, Dufour, 1997].

Lemma 1. [`lem:geometry-handle-eventually-inhabited`] (Eventual geometry-handle existence) *Assume the following conditions.*

- ***(Sample-size ratio.)*** *The asymptotic target-to-source sample-size ratio c satisfies $0 < c$, and $N_n/n \to c$ as in Assumption 3.*

- ***(Overlap level.)*** *The instrument-overlap constant satisfies* $0 < \varepsilon < 1/2$*, as in Assumption 4.*
- ***(Coverage level and calibration constant.)*** *The coverage level* α *of Definition 3 satisfies* $0 < \alpha < 1$*, and the Le Cam calibration constant is* $\rho = (1-\alpha)/8$*.*
- ***(Strength threshold.)*** *The fixed effective-strength threshold satisfies* $t_0 > 0$*, for the effective identification strength* t_n *of Definition 4.*
- ***(Envelope growth.)*** *For every* n*,* $k_n > 0$*, and* $k_n \to \infty$ *with* $k_n/\sqrt{n} \to 0$*, as in Assumptions 13 and 15.*
- ***(Admissible geometry.)*** *The deterministic geometry array* $\mathfrak{g} = (P_S^X, w, e)_{n\geq 1}$ *belongs to the admissible class* $\mathcal{G}$ *of Definition 8 at envelope scale* k_n *and overlap constant* ε*.*

Then, for all sufficiently large n*, the geometry handle of Definition 15 exists at* $\mathfrak{g}$*. In particular, writing*

$$\mu_n = \left(\frac{t_0\kappa_n}{n}\right)^{1/2}, \qquad p_n(x) = \frac{\mu_n w_n(x)}{\kappa_n},$$

one has

$$0 \leq p_n(x) \leq 1, \qquad \int w_n(x)p_n(x)\, dP_{S,n}^X(x) = \mu_n, \qquad \frac{n\mu_n^2}{\kappa_n} = t_0,$$

and the least-favorable family $h \mapsto Q_{n,h}^{\mathfrak{g}}$ *of Definition 14, defined for every* $|h| \leq H(t_0) = \min\{1/4, \rho\, t_0^{-1/2}\}$*, takes values in the fixed-geometry slice* $\mathcal{P}_n(\mathfrak{g})$ *of Definition 9 and satisfies, writing* $Q_{n,h}^{\mathfrak{g},N}$ *for the joint law of the source sample of size* n *and the target covariate sample of size* N_n *under* $Q_{n,h}^{\mathfrak{g}}$*,*

$$1 + \chi^2\left(Q_{n,h}^{\mathfrak{g},N} \,\middle\|\, Q_{n,0}^{\mathfrak{g},N}\right) \leq \exp(8t_0h^2), \qquad \left\|Q_{n,h}^{\mathfrak{g},N} - Q_{n,0}^{\mathfrak{g},N}\right\|_{\mathrm{TV}} \leq 2\rho.$$

The conclusion is eventual in n*: admissibility,* $N_n/n \to c$*, and the growth conditions on* k_n *make the compliance function valid and the calibration hold from some index onwards, not at every* n*.*

Proof of Lemma 1. Write

$$\kappa_n = \int w_n(x)^2\, dP_{S,n}^X(x), \qquad \mu_n = \left(\frac{t_0\kappa_n}{n}\right)^{1/2}, \qquad p_n(x) = \frac{\mu_n w_n(x)}{\kappa_n}.$$

Admissibility gives $0 \leq w_n(x) \leq 2k_n$, $\int w_n\, dP_{S,n}^X = 1$, and $\kappa_n \leq k_n$. Since w_n is bounded and $P_{S,n}^X$ is a probability law, expanding the nonnegative square gives

$$0 \leq \int (w_n(x)-1)^2\, dP_{S,n}^X(x) = \int w_n(x)^2\, dP_{S,n}^X(x) - 2\int w_n(x)\, dP_{S,n}^X(x) + 1 = \kappa_n - 1.$$

Thus $1 \leq \kappa_n$. Since $k_n/\sqrt{n} \to 0$ and $t_0 > 0$, for all sufficiently large n with $n \geq 1$,

$$\frac{k_n}{\sqrt{n}} < \frac{1}{2\sqrt{t_0}}.$$

For such n, using $\kappa_n \geq 1$, $w_n(x) \leq 2k_n$, and

$$p_n(x) = \frac{\sqrt{t_0}\, w_n(x)}{\sqrt{n}\,\sqrt{\kappa_n}},$$

we obtain

$$0 \leq p_n(x) \leq \frac{\sqrt{t_0}\, 2k_n}{\sqrt{n}} \leq 1 \qquad \text{for every } x.$$

The transported first stage is exact:

$$\int w_n(x)p_n(x)\, dP^X_{S,n}(x) = \frac{\mu_n}{\kappa_n}\int w_n(x)^2\, dP^X_{S,n}(x) = \mu_n.$$

Also, since $n \geq 1$, $t_0 > 0$, and $\kappa_n > 0$,

$$\frac{n\mu_n^2}{\kappa_n} = \frac{n}{\kappa_n}\cdot\frac{t_0\kappa_n}{n} = t_0.$$

The construction used for the least-favorable family is first defined, at every row m, with the clamped compliance function

$$p^*_m(x) = \min\{1, p_m(x)\}.$$

On the same eventual set of indices, the preceding bound gives $p^*_n = p_n$ pointwise. The transported array whose row n is denoted $Q^{\mathfrak{g}}_{n,h}$ is generated from the source and target covariate laws of $\mathfrak{g}$, the propensity e_m, and the clamped compliance probabilities p^*_m, so that at the fixed row n the compliance probabilities are p_n. For every

$$|h| \leq H(t_0) = \min\{1/4, \rho t_0^{-1/2}\},$$

the bound $|h| \leq 1/4$ makes the Bernoulli means $1/2 + h$ valid, while the positivity of

$$\int p_n\, dP_{T,n} = \int w_n(x)p_n(x)\, dP^X_{S,n}(x) = \mu_n > 0$$

supplies the target-complier positivity clause.

It remains in the class-membership step to account for the sampling, envelope, and degrading-array clauses in Definitions 1 and 9. The hypotheses give $0 < \varepsilon < 1/2$, $c > 0$, $N_m/m \to c$, $k_m \to \infty$, and $k_m/\sqrt{m} \to 0$. Admissibility of $\mathfrak{g}$ supplies the source and target covariate probability laws, the propensity bounds $e_m \in [\varepsilon, 1-\varepsilon]$, transport domination by w_m, the envelope bound $0 \leq w_m \leq 2k_m$, and the second-moment bound $\int w_m^2\, dP^X_{S,m} \leq k_m$. The witness construction supplies the full-data support, population presence, source-observation, randomization, exclusion, monotonicity, and transport clauses. For the first-stage part of the degrading array, the clamp is eventually inactive and

$$\int w_m(x)p^*_m(x)\, dP^X_{S,m}(x) = \int w_m(x)p_m(x)\, dP^X_{S,m}(x) = \mu_m$$

on that tail, while $\mu_m \to 0$ by the following chain. Since $\kappa_m \leq k_m$, the definition of μ_m gives

$$\mu_m = \left(\frac{t_0\kappa_m}{m}\right)^{1/2} \leq \left(\frac{t_0 k_m}{m}\right)^{1/2}.$$

The envelope growth condition supplies $k_m/\sqrt{m} \to 0$, and therefore

$$\frac{k_m}{m} = \frac{k_m}{\sqrt{m}} \cdot \frac{1}{\sqrt{m}} \longrightarrow 0,$$

being a product of a null sequence and a bounded one. Hence $t_0 k_m/m \to 0$ and the displayed bound forces $\mu_m \to 0$. Thus the transported first-stage sequence of the constructed array tends to zero. Consequently $Q^{\mathfrak{g}}_{n,h}$ satisfies the transported-IV class clauses, has

$$P^X_{S,Q^{\mathfrak{g}}_{n,h}} = P^X_{S,n}, \qquad P_{T,Q^{\mathfrak{g}}_{n,h}} = P_{T,n}, \qquad w_{Q^{\mathfrak{g}}_{n,h}} = w_n \quad P^X_{S,n}\text{-a.e.}, \qquad e_{Q^{\mathfrak{g}}_{n,h}} = e_n,$$

and therefore belongs to $\mathcal{P}_n(\mathfrak{g})$. Its transported first stage is μ_n, its weight second moment is κ_n, and its effective strength is t_0.

It remains to record the calibration. Let Q^S_h denote the one-source-observation marginal under $Q^{\mathfrak{g}}_{n,h}$ and Q^S_0 the corresponding marginal at $h = 0$. For $|h| \le 1/4$, the single-observation chi-square divergence is

$$\chi^2(Q^S_h \| Q^S_0) = \int \frac{8e_n(x)p_n(x)^2h^2}{1 + p_n(x)}\, dP^X_{S,n}(x).$$

Because $0 \le e_n(x) \le 1$ and $p_n(x) \ge 0$,

$$\frac{8e_n(x)p_n(x)^2h^2}{1 + p_n(x)} \le 8p_n(x)^2h^2.$$

Hence

$$\chi^2(Q^S_h \| Q^S_0) \le 8h^2 \int p_n(x)^2\, dP^X_{S,n}(x) = \frac{8\mu_n^2 h^2}{\kappa_n}.$$

The target covariate sample has the same law under h and under 0, so it is ancillary for this divergence, and independence of the n source observations gives

$$1 + \chi^2\Big(Q^{\mathfrak{g},N}_{n,h} \Big\| Q^{\mathfrak{g},N}_{n,0}\Big) = \big(1 + \chi^2(Q^S_h \| Q^S_0)\big)^n.$$

Using $1 + u \le e^u$, the preceding one-observation bound, and $n\mu_n^2/\kappa_n = t_0$,

$$1 + \chi^2\Big(Q^{\mathfrak{g},N}_{n,h} \Big\| Q^{\mathfrak{g},N}_{n,0}\Big) \le \exp\big(n\,\chi^2(Q^S_h \| Q^S_0)\big) \le \exp(8t_0h^2).$$

For the total-variation bound, the radius condition also gives

$$t_0 h^2 \le \rho^2.$$

Set $u = 8\rho^2$. Since $\rho = (1-\alpha)/8$ and $0 < \alpha < 1$, one has $0 < \rho \le 1/8$, hence $0 \le u \le 1$. The elementary exponential remainder bound on $[0,1]$ yields

$$\exp(u) - 1 \le u + u^2 \le 9\rho^2.$$

Combining this with the chi-square bound gives

$$\chi^2\Big(Q^{\mathfrak{g},N}_{n,h} \Big\| Q^{\mathfrak{g},N}_{n,0}\Big) \le 9\rho^2.$$

Finally, the standard comparison between total variation and chi-square distance gives

$$\Big\| Q^{\mathfrak{g},N}_{n,h} - Q^{\mathfrak{g},N}_{n,0} \Big\|_{\mathrm{TV}} \le \frac{1}{2}\sqrt{\chi^2\Big(Q^{\mathfrak{g},N}_{n,h} \Big\| Q^{\mathfrak{g},N}_{n,0}\Big)} \le \frac{3}{2}\rho \le 2\rho.$$

All displayed properties therefore hold for every sufficiently large n, which is exactly the asserted eventual geometry handle. $\square$

Lemma 1 connects the abstract handle to the admissible geometries used in Theorem 5. The sub-root envelope condition makes the compliance probabilities valid eventually, and the calibrated family then belongs to the fixed-geometry slice with exactly the displayed effective strength.

The next three inhabitation lemmas record that the finite-cell and fixed-geometry classes used by the minimax criteria contain rows satisfying the stated structural restrictions.

Lemma 2. [`lem:finite-cell-class-inhabited`] (Finite-cell class nonemptiness) *Suppose the following conditions hold.*

- ***(Measurable finite cells.)*** *For every n, the covariate space $\mathcal{X}$ of Assumption 12 contains k_n distinct injected points whose singletons are measurable.*
- ***(Sample-size ratio.)*** *The constant c satisfies $0 < c$, and $N_n/n \to c$ as in Assumption 3.*
- ***(Instrument overlap.)*** *The constant ε satisfies $0 < \varepsilon < 1/2$, as in Assumption 4.*
- ***(Cell-count sequence.)*** *For every n, $k_n > 0$, with $k_n \to \infty$ as in Assumption 15 and*

$$\frac{k_n}{\sqrt{n}} \to 0$$

as in Assumption 13.

Then, for every n, there exists a transported array law P, consisting of full-data and observed source laws together with the associated propensity, assignment-outcome, assignment-contrast, and receipt-contrast functions, such that P belongs to the finite-cell submodel $\mathcal{N}_n$ of Definition 2. In particular, P satisfies the finite source-cell condition in Assumption 16 and the clauses of the transported model class $\mathcal{P}_n$ in Definition 1.

Proof of Lemma 2. Choose the injected measurable cells simultaneously for all rows: for each row m, let

$$i \longmapsto x_{i,m}, \qquad i \in \{1, \dots, k_m\},$$

be the injection supplied by the carrier assumption. Define whole covariate-law, weight, propensity, and complier-probability arrays by

$$P^X_{S,m} = P_{T,m} = \frac{1}{k_m} \sum_{i=1}^{k_m} \delta_{x_{i,m}}, \qquad w_m(x) = 1, \qquad e_m(x) = \frac{1}{2}, \qquad p_m(x) = \frac{1}{m+1}.$$

The hypotheses $k_m > 0$ and singleton measurability make these displayed finite laws probability measures on the covariate space. For every measurable A,

$$P_{T,m}(A) = P^X_{S,m}(A) = \int_A 1 \, dP^X_{S,m},$$

so the target-to-source covariate density ratio is represented by $w_m \equiv 1$. The overlap and weight bounds hold rowwise:

$$\varepsilon \le \frac{1}{2} \le 1 - \varepsilon, \qquad 0 \le w_m(x) = 1 \le 2k_m, \qquad \int w_m(x)^2 \, dP^X_{S,m}(x) = 1 \le k_m.$$

Here the first inequality uses $0 < \varepsilon < 1/2$, and the last two use $k_m > 0$.

Construct the full-data law in each row as the half-source, half-target mixture with

$$P_m^F(S=1) = P_m^F(S=0) = \frac{1}{2}, \qquad X \mid S=1 \sim P_{S,m}^X, \qquad X \mid S=0 \sim P_{T,m}.$$

Conditional on $(S,X) = (s,x)$, draw a complier indicator with probability $p_m(x)$, draw an independent common noncomplier receipt indicator with probability $1/2$, and draw an independent treated-outcome indicator with probability $1/2$. If the complier indicator is one, set $(D(0), D(1)) = (0,1)$; otherwise set $D(0) = D(1)$ equal to the common receipt indicator. Set $Y(0) = 0$, set $Y(1)$ equal to the treated-outcome indicator read as a number in $\{0,1\}$, and take the assignment potential outcome to be $Y^Z(z) = Y(D(z))$. Since $0 \le p_m(x) \le 1$, this specifies probability kernels and hence the full-data laws. The construction has full-data support, positive mass for both populations, exclusion, and monotonicity by inspection, and it gives, for every measurable A and $s \in \{0,1\}$,

$$\int_{\{X\in A\}} \big(Y^Z(1) - Y^Z(0)\big)\, dP_m^F(\cdot \mid S=s) = \int_A \frac{p_m(x)}{2}\, dP_{s,m}^X(x),$$

$$\int_{\{X\in A\}} \big(D(1) - D(0)\big)\, dP_m^F(\cdot \mid S=s) = \int_A p_m(x)\, dP_{s,m}^X(x),$$

where $P_{1,m}^X = P_{S,m}^X$ and $P_{0,m}^X = P_{T,m}$. Thus the population outcome and receipt contrasts are $p_m/2$ and p_m, respectively, in both populations.

The observed source law is induced from the $S=1$ component by drawing Z conditionally on the full data with probability $e_m(X) = 1/2$, then observing $D = D(Z)$ and $Y = Y(D(Z))$. This assigned-source law has the $S=1$ full-data marginal, and for every measurable A,

$$P_{S,m}\{X \in A, Z=1\} = \int_A e_m(x)\, dP_{S,m}^X(x).$$

The same construction pins the observed source score contrasts: for every measurable A, with the Boolean variables read as 0-1 variables,

$$\int_{\{X\in A\}} \left(\frac{Z}{e_m(X)} - \frac{1-Z}{1-e_m(X)}\right) Y\, dP_{S,m} = \int_A \frac{p_m(x)}{2}\, dP_{S,m}^X(x),$$

$$\int_{\{X\in A\}} \left(\frac{Z}{e_m(X)} - \frac{1-Z}{1-e_m(X)}\right) D\, dP_{S,m} = \int_A p_m(x)\, dP_{S,m}^X(x).$$

Therefore the source observation, randomization, outcome-transport, and receipt-transport clauses hold for the transported array.

Now fix the row n. The target complier share under the target population is

$$P_n^F\{D(1)=1,\ D(0)=0 \mid S=0\} = \int p_n(x)\, dP_{T,n}(x) = \frac{1}{n+1} > 0,$$

and the conditioning event has mass $1/2$ by the mixture construction. The transport domination, envelope, and second-moment clauses are the row-n statements already verified above. For the degrading-array clause, the whole array construction gives

$$\mu_m = \int w_m(x) p_m(x)\, dP_{S,m}^X(x) = \frac{1}{m+1} \longrightarrow 0,$$

while the hypotheses give $k_m \to \infty$, $k_m/\sqrt{m} \to 0$, and $N_m/m \to c$. Hence the constructed array satisfies all clauses of $\mathcal{P}_n$ in Definition 1.

Finally, the source covariate law in row n is exactly $P^X_{S,n}$. It is supported on the injected cells, and each injected singleton has mass

$$P^X_{S,n}(\{x_{i,n}\}) = \frac{1}{k_n}.$$

Moreover $e_n(X) = 1/2$ almost surely under the source covariate law. Thus the finite source-cell condition in Assumption 16 holds, and by Definition 2 the constructed transported array belongs to $\mathcal{N}_n$. Since n was arbitrary, the finite-cell class is inhabited for every row. □

Lemma 2 supplies the uniform finite-cell rows used in the finite-cell oracle and feasible comparisons. The measurable injected carrier gives a concrete cell support, while the overlap and growth conditions align those rows with the model class.

Lemma 3. [`lem:fixed-geometry-slice-inhabited`] (Fixed slice inhabitation) *Let $\mathfrak{g}$, the deterministic geometry array, be fixed, and let N_n and k_n be integer-valued sequences. Suppose that:*

- ***(Admissible geometry.)*** *The geometry $\mathfrak{g}$ is admissible for the envelope sequence k_n and the instrument-overlap constant ε in the sense of Definition 8.*
- ***(Sample-size ratio.)*** *The constant c, the asymptotic target-to-source sample-size ratio, satisfies $c > 0$, and $N_n/n \to c$ as in Assumption 3.*
- ***(Overlap level.)*** *The instrument-overlap constant satisfies $0 < \varepsilon < 1/2$, as in Assumption 4.*
- ***(Strength threshold.)*** *The fixed effective-strength threshold satisfies $t_0 > 0$.*
- ***(Envelope growth.)*** *The envelope sequence satisfies $k_n > 0$ for every n, $k_n \to \infty$, and $k_n/\sqrt{n} \to 0$, as in Assumptions 13 and 15.*

Then, for all sufficiently large n, there exists a transported-array law P such that $P \in \mathcal{P}_n(\mathfrak{g})$, the fixed-geometry slice of Definition 9, and its effective identification strength $t_n(P) = n\mu_n(P)^2/\kappa_n(P)$ from Definition 4 satisfies

$$t_n(P) \geq t_0.$$

Equivalently, for all sufficiently large n, the strength-restricted fixed-geometry slice

$$\{P \in \mathcal{P}_n(\mathfrak{g}) : t_n(P) \geq t_0\}$$

is nonempty.

Proof of Lemma 3. Write the fixed geometry at row m as

$$(P^X_{S,m}, P_{T,m}, w_m, e_m),$$

and put

$$\kappa_m = \int w_m(x)^2 \, dP^X_{S,m}(x), \qquad \mu_m = \left(\frac{t_0 \kappa_m}{m}\right)^{1/2}, \qquad p_m(x) = \min\left\{1, \frac{\mu_m w_m(x)}{\kappa_m}\right\}.$$

The geometry supplies the measurability of w_m and e_m. Admissibility in Definition 8 supplies probability laws $P^X_{S,m}, P_{T,m}$, the overlap bounds

$$\varepsilon \le e_m(x) \le 1 - \varepsilon,$$

the weight bounds

$$0 \le w_m(x) \le 2k_m, \qquad \kappa_m \le k_m,$$

the normalization

$$\int w_m(x)\, dP^X_{S,m}(x) = 1,$$

and the transport identity

$$P_{T,m}(A) = \int_A w_m(x)\, dP^X_{S,m}(x)$$

for every measurable A.

The construction uses the untruncated value

$$\bar{p}_m(x) = \frac{\mu_m w_m(x)}{\kappa_m}.$$

The envelope bound and measurability make w_m and w_m^2 integrable. Also $\kappa_m > 0$: if $\kappa_m = 0$, then the nonnegative integral of w_m^2 is zero, hence $w_m = 0$ $P^X_{S,m}$-almost surely, contradicting $\int w_m\, dP^X_{S,m} = 1$. Moreover,

$$0 \le \int (w_m(x) - 1)^2\, dP^X_{S,m}(x) = \int w_m(x)^2\, dP^X_{S,m}(x) - 2\int w_m(x)\, dP^X_{S,m}(x) + 1 = \kappa_m - 1,$$

so $1 \le \kappa_m$. For all sufficiently large m, with $m \ge 1$, the assumption $k_m/\sqrt{m} \to 0$ gives

$$\frac{k_m}{\sqrt{m}} < \frac{1}{2\sqrt{t_0}}.$$

For such m and every x,

$$0 \le \bar{p}_m(x) = \frac{\sqrt{t_0}\, w_m(x)}{\sqrt{m}\,\sqrt{\kappa_m}} \le \frac{2\sqrt{t_0}\, k_m}{\sqrt{m}} < 1.$$

Thus, eventually,

$$p_m(x) = \bar{p}_m(x) = \frac{\mu_m w_m(x)}{\kappa_m} \qquad \text{for every } x.$$

For every row, p_m is measurable and takes values in $[0, 1]$.

Define a transported array row by row as follows. In the source population draw $X \sim P^X_{S,m}$, and in the target population draw $X \sim P_{T,m}$. Conditional on $X = x$, draw a complier indicator with probability $p_m(x)$, a common-receipt indicator with probability $1/2$, and a treated-outcome indicator with probability $1/2$. Set $(D(0), D(1)) = (0, 1)$ for compliers and $(D(0), D(1)) = (U, U)$ otherwise, where U is the common-receipt draw; set $Y(0) = 0$ and $Y(1) = V$, where V is the treated-outcome draw. The full-data law mixes the source and target population laws with weights $1/2$ and $1/2$. The assigned source law then draws Z with propensity $e_m(X)$ and observes $D = D(Z)$ and $Y = Y(D(Z))$. The associated assignment and receipt contrasts are $p_m/2$ and p_m, respectively.

Fix a sufficiently large n in the eventual set above. In the target population component of this witness, the complier event $\{D(1) = 1, D(0) = 0\}$ is exactly the success event of the complier draw. Hence the target complier share is

$$\int p_n(x)\, dP_{T,n}(x).$$

Using the transport identity to integrate against $P_{T,n}$ and using the eventual identity $p_n = \bar{p}_n$,

$$\int p_n(x)\, dP_{T,n}(x) = \int w_n(x) p_n(x)\, dP^X_{S,n}(x) = \int w_n(x) \frac{\mu_n w_n(x)}{\kappa_n}\, dP^X_{S,n}(x) = \mu_n.$$

This value is strictly positive because $t_0 > 0$, $n > 0$, and $\kappa_n > 0$.

The displayed construction supplies the clauses of the transported model class in Definition 1. The full-data support and population-presence clauses follow from the probability laws and the binary support of the three Bernoulli draws. The two-sample clause uses $c > 0$, $N_n/n \to c$, and the source and target probability laws. The instrument-overlap clause is the inherited bound on e_n. The source-observation, randomization, exclusion, and monotonicity clauses follow from drawing Z conditionally on X, observing $D(Z)$ and $Y(D(Z))$, and setting $D(1) \geq D(0)$ by construction. The outcome- and receipt-transport clauses hold because the conditional contrasts are the same functions, $p_m/2$ and p_m, in the source and target population components. The target-complier-positivity clause is the strict positivity just displayed. The transport identity identifies $P_{T,n}$ with the weighted measure $w_n\, P^X_{S,n}$, so the Radon–Nikodym transport weight of the witness equals w_n $P^X_{S,n}$-almost surely; this gives transport domination and the weight-envelope clause, while the second-moment clause is $\int w_n^2\, dP^X_{S,n} \leq k_n$.

For the degrading-array clause, the witness first stage satisfies, eventually in m,

$$\mu_m(P) = \int w_m(x) p_m(x)\, dP^X_{S,m}(x) = \mu_m.$$

Furthermore,

$$0 \leq \mu_m = \left(\frac{t_0 \kappa_m}{m}\right)^{1/2} \leq \left(\frac{t_0 k_m}{m}\right)^{1/2} \longrightarrow 0,$$

since $k_m/m = (k_m/\sqrt{m})(1/\sqrt{m}) \to 0$. Together with $k_m \to \infty$ and $k_m/\sqrt{m} \to 0$, this verifies the degrading-array requirement. Therefore the witness belongs to $\mathcal{P}_n$.

The source covariate law, target covariate law, propensity, and transport weight of the witness agree with the row-n geometry:

$$P^{X,P}_{S,n} = P^X_{S,n}, \qquad P^P_{T,n} = P_{T,n}, \qquad e^P_n = e_n,$$

and the transport weight is w_n $P^X_{S,n}$-almost surely by the Radon–Nikodym identification above. Hence the witness belongs to the fixed-geometry slice $\mathcal{P}_n(\mathfrak{g})$ of Definition 9.

For this witness,

$$\mu_n(P) = \int w_n(x) p_n(x)\, dP^X_{S,n}(x) = \mu_n, \qquad \kappa_n(P) = \int w_n(x)^2\, dP^X_{S,n}(x) = \kappa_n.$$

Using the effective-strength definition in Definition 4,

$$t_n(P) = \frac{n \mu_n(P)^2}{\kappa_n(P)} = \frac{n \mu_n^2}{\kappa_n} = t_0,$$

where the last equality uses $\mu_n^2 = t_0\kappa_n/n$, $n > 0$, and $\kappa_n > 0$. Thus $t_n(P) \geq t_0$ for every sufficiently large n, proving eventual nonemptiness of the strength-restricted fixed-geometry slice. $\square$

The fixed-geometry value in Definition 11 takes a supremum over strength-restricted slices. Lemma 3 ensures that those slices are populated eventually under the same admissible-geometry and growth conditions used for the fixed-geometry characterization.

Lemma 4. [`lem:regular-finite-cell-class-inhabited`] (Regular class nonemptiness) *Assume the following conditions.*

- ***(Finite measurable cells.)*** *For every n, the covariate space $\mathcal{X}$ of Assumption 12 contains k_n distinct points, indexed by $\{1, \ldots, k_n\}$, whose singletons are measurable.*
- ***(Sample-size ratio.)*** *The asymptotic target-to-source sample-size ratio c satisfies $0 < c$, and $N_n/n \to c$ as in Assumption 3.*
- ***(Instrument overlap.)*** *The instrument-overlap constant ε satisfies $0 < \varepsilon < 1/2$, as in Assumption 4.*
- ***(Regular-cell bounds.)*** *The fixed regular-cell mass constants c_- and c_+ satisfy $0 < c_- \leq 1 \leq c_+$.*
- ***(Cell-count growth.)*** *For every n, $k_n > 0$, and $k_n \to \infty$ with $k_n/\sqrt{n} \to 0$, as in Assumptions 13 and 15.*

Then, for every n, there exists a transported array P, consisting of full-data laws, assigned-source laws, propensities, assignment-outcome maps, assignment contrasts, and receipt contrasts with the usual measurability structure, such that $P \in \mathcal{N}_n^{\mathrm{reg}}$, the regular finite-cell model class of Definition 12. In particular, the source cell probabilities $q_{x,n}$ on the witnessing measurable cells satisfy

$$c_-/k_n \leq q_{x,n} \leq c_+/k_n.$$

Proof of Lemma 4. Use the same uniform finite-cell witness constructed in the proof of Lemma 2. Thus, for each fixed n, choose injected measurable cells $x_{i,n}$, set

$$P_{S,n}^X = P_{T,n} = \frac{1}{k_n}\sum_{i=1}^{k_n}\delta_{x_{i,n}}, \qquad w_n \equiv 1, \qquad e_n \equiv \frac{1}{2}, \qquad p_n \equiv \frac{1}{n+1},$$

and generate the corresponding transported array. The preceding verification gives membership in $\mathcal{P}_n$ and the finite source-cell structure.

For the witnessing cells, the source cell probabilities are

$$q_{i,n} = P_{S,n}^X(\{x_{i,n}\}) = \frac{1}{k_n}.$$

Since $k_n > 0$, multiplication by $1/k_n$ preserves the inequalities $c_- \leq 1 \leq c_+$. Hence, for every cell i,

$$\frac{c_-}{k_n} \leq \frac{1}{k_n} = q_{i,n} \leq \frac{c_+}{k_n}.$$

The constructed array therefore satisfies the regular finite-cell bounds in Definition 12, and so belongs to $\mathcal{N}_n^{\mathrm{reg}}$. Since n was arbitrary, the regular finite-cell class is inhabited for every n. $\square$

Lemma 4 provides the regular nonuniform-cell analogue. The bounded source-cell probabilities make the regular class compatible with the finite-cell sampling structure used by Theorem 7.

We next record the measure and source-observation facts that underwrite the identification and score calculations. These statements translate setwise contrast identities into almost-sure bounds and integrability conclusions.

Lemma 5. [`lem:setwise-domination-unit-bound`] (Setwise domination gives an almost-sure unit bound) *Let μ be a finite measure on a measurable space Ω, and let $f : \Omega \to \mathbb{R}$ be a measurable function. Suppose that*

$$\left| \int_A f \, d\mu \right| \le \mu(A)$$

for every measurable set $A \subseteq \Omega$. Then

$$|f(x)| \le 1 \qquad \text{for } \mu\text{-almost every } x.$$

No integrability hypothesis on f is imposed. The truncated measurable slices

$$\{q < f \le m\} \quad \text{and} \quad \{-m \le f < -q\}, \qquad q > 1,\ m \in \mathbb{N},$$

carry a bounded restriction of f and therefore a genuine integral, so the displayed domination may be applied on them; exhausting the regions $\{f > q\}$ and $\{f < -q\}$ as $m \to \infty$ and then letting $q \downarrow 1$ gives the stated almost-everywhere bound. The intended application takes $\mu = P_S^X$, the source covariate marginal of Assumption 12.

Proof of Lemma 5. Fix $q > 1$. For $m \in \mathbb{N}$, put

$$A_m(q) = \{x : q < f(x) \le m\}.$$

This set is measurable, and $f\mathbf{1}_{A_m(q)}$ is integrable because $|f| \le m$ on $A_m(q)$ and μ is finite. On this slice,

$$q\,\mu(A_m(q)) = \int_{A_m(q)} q \, d\mu \le \int_{A_m(q)} f \, d\mu \le \left| \int_{A_m(q)} f \, d\mu \right| \le \mu(A_m(q)).$$

Since $q > 1$, this gives $\mu(A_m(q)) = 0$ for every m. If $f(x) > q$, choose $m \in \mathbb{N}$ with $f(x) \le m$; then $x \in A_m(q)$. Hence

$$\mu\{x : f(x) > q\} = 0, \qquad \text{so} \qquad f \le q \quad \mu\text{-a.e.}$$

The lower tail is identical after reversing signs. For $m \in \mathbb{N}$, set

$$B_m(q) = \{x : -m \le f(x) < -q\}.$$

Again $f\mathbf{1}_{B_m(q)}$ is integrable, and

$$\int_{B_m(q)} f \, d\mu \le \int_{B_m(q)} (-q) \, d\mu = -q\,\mu(B_m(q)).$$

The assumed domination also gives

$$-\mu(B_m(q)) \le \int_{B_m(q)} f \, d\mu.$$

Therefore $(q-1)\mu(B_m(q)) \le 0$, and $\mu(B_m(q)) = 0$. Taking m with $-f(x) \le m$ shows

$$\mu\{x : f(x) < -q\} = 0, \qquad \text{so} \qquad -q \le f \quad \mu\text{-a.e.}$$

Apply the preceding two conclusions to

$$q_j = 1 + \frac{1}{j+1}, \qquad j = 0, 1, 2, \ldots.$$

Outside a null set, $-q_j \le f(x) \le q_j$ for every j. Letting $j \to \infty$ gives

$$-1 \le f(x) \le 1$$

outside that null set, equivalently $|f(x)| \le 1$ μ-almost surely. □

Lemma 5 is the elementary setwise domination step used to obtain almost-sure unit bounds for conditional contrasts. Those bounds are what make the compact range in Theorem 1 available directly from the model restrictions.

Lemma 6. [`lem:fixed-geometry-value-le-frontier-risk`] (Fixed-geometry value below global frontier risk) *Assume the following conditions.*

- ***(Sample-size ratio.)*** *The asymptotic target-to-source sample-size ratio c satisfies $0 < c$, and $N_n/n \to c$ as in Assumption 3.*
- ***(Overlap level.)*** *The instrument-overlap constant satisfies $0 < \varepsilon < 1/2$, as in Assumption 4.*
- ***(Envelope growth.)*** *For every n, $k_n > 0$, and $k_n \to \infty$ with $k_n/\sqrt{n} \to 0$, as in Assumptions 13* and *15.*
- ***(Admissible geometry.)*** *The deterministic geometry array $\mathfrak{g}$ belongs to the admissible class $\mathcal{G}$ of Definition 8 at envelope scale k_n and overlap constant ε.*
- ***(Strength threshold.)*** *The fixed effective-strength threshold satisfies $t_0 > 0$.*

Then, for every oracle-honest sequence $(C_n) \in \mathfrak{C}^{\mathrm{or}}$ of Definition 3, the fixed-geometry minimax value $V_{\mathfrak{g}}^{\star}(t_0)$ of Definition 11 and the frontier risk $R((C_n), t_0)$ of Definition 4 satisfy

$$V_{\mathfrak{g}}^{\star}(t_0) \le R((C_n), t_0).$$

The globally honest sequence is first restricted to a fixed-geometry honest sequence in $\mathfrak{C}^{\mathrm{or}}(\mathfrak{g})$ of Definition 10, whose strength-restricted rows over $\mathcal{P}_n(\mathfrak{g})$ of Definition 9 are eventually nonempty; no pointwise identification of almost-everywhere equal transport-weight versions is needed. Taking the infimum over $(C_n) \in \mathfrak{C}^{\mathrm{or}}$ gives

$$V_{\mathfrak{g}}^{\star}(t_0) \le V^{\star}(t_0)$$

for the oracle minimax value of Definition 5.

Proof of Lemma 6. Fix an oracle-honest sequence $C = (C_n) \in \mathfrak{C}^{\mathrm{or}}$. Let $\widetilde{C}$ be the same oracle procedure evaluated with the fixed geometry $\mathfrak{g}$:

$$\widetilde{C}_n\big((O_i^S)_{i=1}^n, (X_j^T)_{j=1}^{N_n}\big) = C_n\big((O_i^S)_{i=1}^n, (X_j^T)_{j=1}^{N_n}, w, e\big),$$

where w, e are the nth weight and propensity components of $\mathfrak{g}$. If $P \in \mathcal{P}_n(\mathfrak{g})$, then $P \in \mathcal{P}_n$, the propensity input equals the nth geometry propensity, and the nth geometry weight is source-almost surely a valid transport-weight version for P, by Definition 9. The oracle procedure is evaluated on admissible transport-weight versions modulo source-null changes, so under the two-sample law generated by such a P,

$$\mathbf{1}\{\theta_T \in \widetilde{C}_n\} = \mathbf{1}\{\theta_T \in C_n\}, \qquad \lambda(\widetilde{C}_n) = \lambda(C_n) \quad P\text{-a.s.}$$

Consequently, for every $P \in \mathcal{P}_n(\mathfrak{g})$,

$$P\{\theta_T \in \widetilde{C}_n\} = P\{\theta_T \in C_n\}, \qquad \mathbb{E}_P[\lambda(\widetilde{C}_n)] = \mathbb{E}_P[\lambda(C_n)].$$

By Lemma 3, for all sufficiently large n there exists $P_0 \in \mathcal{P}_n(\mathfrak{g})$ with $t_n(P_0) \geq t_0$. Hence the fixed-geometry slice and the global class are inhabited on those rows. On such rows, the preceding coverage identity and the inclusion $\mathcal{P}_n(\mathfrak{g}) \subseteq \mathcal{P}_n$ give

$$\inf_{P \in \mathcal{P}_n} P\{\theta_T \in C_n\} \leq \inf_{P \in \mathcal{P}_n(\mathfrak{g})} P\{\theta_T \in \widetilde{C}_n\}.$$

The same nonempty-row reduction gives the bounds used to pass this comparison to liminf:

$$0 \leq \inf_{P \in \mathcal{P}_n} P\{\theta_T \in C_n\}, \qquad \inf_{P \in \mathcal{P}_n(\mathfrak{g})} P\{\theta_T \in \widetilde{C}_n\} \leq 1$$

for all sufficiently large n, since coverage probabilities lie in $[0, 1]$. Therefore the eventual row comparison yields

$$\liminf_{n\to\infty} \inf_{P \in \mathcal{P}_n(\mathfrak{g})} P\{\theta_T \in \widetilde{C}_n\} \geq \liminf_{n\to\infty} \inf_{P \in \mathcal{P}_n} P\{\theta_T \in C_n\} \geq 1 - \alpha,$$

and the $0 < \alpha < 1$ clauses are inherited from $C \in \mathfrak{C}^{\mathrm{or}}$. Thus $\widetilde{C} \in \mathfrak{C}^{\mathrm{or}}(\mathfrak{g})$ by Definition 10.

The eventual strength-restricted fixed-geometry rows are also nonempty by Lemma 3. On those rows, the expected-length identity and the inclusion of the fixed slice into the global class give

$$\sup_{P \in \mathcal{P}_n(\mathfrak{g}) : t_n(P) \geq t_0} \mathbb{E}_P[\lambda(\widetilde{C}_n)] \leq \sup_{P \in \mathcal{P}_n : t_n(P) \geq t_0} \mathbb{E}_P[\lambda(C_n)].$$

For every set $A \subseteq \Theta = [-1, 1]$, the restricted length satisfies $0 \leq \lambda(A) \leq 2$. For the displayed real suprema over strength-restricted rows, an empty indexing set gives the real supremum of the empty set, which is 0; on nonempty rows, the same length bound makes each fixed-geometry row nonnegative and each global row bounded above by 2. These bounds are the hypotheses needed to pass the eventual row comparison through limsup, so

$$\limsup_{n\to\infty} \sup_{P \in \mathcal{P}_n(\mathfrak{g}) : t_n(P) \geq t_0} \mathbb{E}_P[\lambda(\widetilde{C}_n)] \leq R((C_n), t_0).$$

It remains to take the infimum over fixed-geometry honest procedures. For any oracle procedure E, the same real-supremum convention and the bound $0 \leq \lambda(A) \leq 2$ imply that each strength-restricted fixed-geometry risk row for E lies in $[0, 2]$, and hence

$$0 \leq \limsup_{n\to\infty} \sup_{P \in \mathcal{P}_n(\mathfrak{g}) : t_n(P) \geq t_0} \mathbb{E}_P[\lambda(E_n)].$$

Thus the collection of fixed-geometry risks over $\mathfrak{C}^{\mathrm{or}}(\mathfrak{g})$ is bounded below by 0. Since the constructed $\widetilde{C}$ belongs to $\mathfrak{C}^{\mathrm{or}}(\mathfrak{g})$, Definition 11 gives

$$V_{\mathfrak{g}}^{\star}(t_0) \le \limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n(\mathfrak{g}):\, t_n(P)\ge t_0} \mathbb{E}_P[\lambda(\widetilde{C}_n)].$$

Combining this inequality with the preceding risk comparison gives

$$V_{\mathfrak{g}}^{\star}(t_0) \le R((C_n), t_0).$$

□

The comparison in Lemma 6 places each fixed-geometry lower bound underneath the global oracle criterion. This is the bridge from the fixed-geometry least-favorable experiment to the oracle lower bound stated in Theorem 2.

Lemma 7. [`lem:source-observation-facts`] (Source observation facts) *Let P belong to the transported complier-effect model class $\mathcal{P}_n$ of Definition 1, for the sample-size array N_n, envelope scale k_n, target-to-source ratio c, and instrument-overlap constant ε. In particular, the source-observation quantities below are governed by Assumptions 1, 3 to 5, 7 to 10 and 12. Let $P_{\mathrm{assign},S}$ denote the assigned-source overlay on (F,Z), where $F = (S, X, D(0), D(1), Y(0), Y(1))$, and let P_S be its observed-source image under $(F,Z) \mapsto (X, Z, D(Z), Y(D(Z)))$. Write P_S^X for the source covariate marginal and*

$$H(O^S) = \frac{\mathbf{1}\{Z=1\}}{e(X)} - \frac{\mathbf{1}\{Z=0\}}{1-e(X)}.$$

Then $P_{\mathrm{assign},S}$ and P_S are probability measures, the full-data marginal of $P_{\mathrm{assign},S}$ is $P_n^F(\cdot \mid S=1)$, and the observed outcome satisfies $Y \in [0,1]$ P_S-almost surely. Moreover, for every measurable $A \subseteq \mathcal{X}$,

$$P_S\{X\in A,\ Z=1\} = \int_A e(x)\, dP_S^X(x),$$

$$\int_{\{X\in A\}} H(O^S) Y\, dP_S = \int_A \Delta_Y(x)\, dP_S^X(x),$$

and

$$\int_{\{X\in A\}} H(O^S) D\, dP_S = \int_A \Delta_D(x)\, dP_S^X(x).$$

Finally, $\Delta_Y(X)\in[-1,1]$ and $\Delta_D(X)\in[0,1]$ hold P_S^X-almost surely.

Proof of Lemma 7. Unpacking the assigned-source part of Assumption 5, the overlay $P_{\mathrm{assign},S}$ is a probability law on (F,Z) whose full-data marginal is

$$(P_{\mathrm{assign},S}) \circ F^{-1} = P_n^F(\cdot \mid S=1).$$

The observed source law is the image of this overlay under $(F,Z) \mapsto (X, Z, D(Z), Y(D(Z)))$, and Assumption 3 gives that this image P_S is a probability measure. Therefore the source covariate marginal is the image of the source-population full-data law under $F \mapsto X(F)$, so

$$P_S^X = P_n^F(X \in \cdot \mid S=1).$$

The support condition in Assumption 1, restricted to the source population, gives $Y(0), Y(1) \in [0,1]$ under $P_n^F(\cdot \mid S = 1)$. Pulling this statement back to $P_{\text{assign},S}$ and then applying the observation map gives

$$P_S\{0 \le Y \le 1\} = 1.$$

Let $A \subseteq \mathcal{X}$ be measurable. The assignment-propensity component of Assumption 5 gives the observed assignment slice

$$P_S\{X \in A,\ Z = 1\} = \int_A e(x)\, dP_S^X(x).$$

For the inverse-propensity identities, the assignment-propensity component of Assumption 5 together with the conditional randomization in Assumption 6 gives the two assigned-source slices: for every measurable full-data set B,

$$P_{\text{assign},S}\{F \in B,\ Z = 1\} = \int_B e(X(F))\, dP_n^F(F \mid S = 1),$$

and

$$P_{\text{assign},S}\{F \in B,\ Z = 0\} = \int_B \{1 - e(X(F))\}\, dP_n^F(F \mid S = 1).$$

The overlap bounds in Assumption 4 make these weights nonnegative and make $e(X)$ and $1 - e(X)$ positive on the almost-sure sets where the cancellations below are made.

Define $F_z(F) = Y(D(z))$ for $z \in \{0,1\}$. On the assigned-source space,

$$H(O^S)Y = \begin{cases} e(X)^{-1} F_1(F), & Z = 1, \\ -\{1 - e(X)\}^{-1} F_0(F), & Z = 0. \end{cases}$$

Splitting the integral over $Z = 1$ and $Z = 0$, applying the two slice identities above, and cancelling the propensity factors yields

$$\int_{\{X \in A\}} H(O^S)Y\, dP_S = \int_{\{X(F) \in A\}} \{Y(D(1)) - Y(D(0))\}\, dP_n^F(F \mid S = 1).$$

By Assumption 7, the integrand agrees with $Y^Z(1) - Y^Z(0)$ under the source population law, and the source part of Assumption 9 gives

$$\int_{\{X(F) \in A\}} \{Y^Z(1) - Y^Z(0)\}\, dP_n^F(F \mid S = 1) = \int_A \Delta_Y(x)\, dP_S^X(x).$$

Thus

$$\int_{\{X \in A\}} H(O^S)Y\, dP_S = \int_A \Delta_Y(x)\, dP_S^X(x).$$

The receipt identity uses the same two-slice calculation with $G_z(F) = D(z)$, reading binary receipts as their 0-1 values. On the assigned-source space,

$$H(O^S)D = \begin{cases} e(X)^{-1} G_1(F), & Z = 1, \\ -\{1 - e(X)\}^{-1} G_0(F), & Z = 0, \end{cases}$$

and therefore

$$\int_{\{X\in A\}} H(O^S)D\,dP_S = \int_{\{X(F)\in A\}} \{D(1)-D(0)\}\,dP_n^F(F\mid S=1).$$

The source-population identity in Assumption 10 then gives

$$\int_{\{X\in A\}} H(O^S)D\,dP_S = \int_A \Delta_D(x)\,dP_S^X(x).$$

It remains to record the pointwise ranges of the two covariate contrasts. From Assumptions 1 and 7,

$$\left|Y^Z(1)-Y^Z(0)\right| \le 1$$

under $P_n^F(\cdot\mid S=1)$. Combining this bound with the source part of Assumption 9 gives, for every measurable A,

$$\left|\int_A \Delta_Y(x)\,dP_S^X(x)\right| \le P_S^X(A).$$

Since Δ_Y is measurable and P_S^X is finite, Lemma 5 gives

$$\Delta_Y(X)\in[-1,1] \qquad P_S^X\text{-a.s.}$$

Similarly, binary receipt gives $|D(1)-D(0)|\le 1$, and the source part of Assumption 10 gives

$$\left|\int_A \Delta_D(x)\,dP_S^X(x)\right| \le P_S^X(A)$$

for every measurable A. Since Δ_D is measurable, Lemma 5 gives $|\Delta_D(X)|\le 1$ P_S^X-almost surely. Monotonicity in Assumption 8 gives $D(1)-D(0)\ge 0$ under the source population law, so

$$\int_A \Delta_D(x)\,dP_S^X(x) \ge 0$$

for every measurable A. The almost-sure unit bound makes Δ_D integrable with respect to P_S^X, and the set-integral characterization of nonnegativity yields $\Delta_D(X)\ge 0$ P_S^X-almost surely. Combining the two bounds gives

$$\Delta_D(X)\in[0,1] \qquad P_S^X\text{-a.s.}$$

□

Lemma 7 gathers the source-law identities needed by both identification and score inversion. It supplies the inverse-propensity representations of the source conditional outcome and receipt contrasts, together with the boundedness inherited from bounded outcomes and monotone binary receipt.

Lemma 8. [`lem:transported-contrast-integrable`] (Transported contrast integrability) *Fix an index n, a transported array P, target sample-size and envelope schedules N_n and k_n, a target-to-source ratio c, and the instrument-overlap constant ε. Suppose that $P\in\mathcal{P}_n$ as in Definition 1 for these schedules and constants. In particular, Assumptions 1, 4, 7 to 10 and 12 hold at index n. Then the source conditional outcome contrast Δ_Y and the source conditional receipt contrast Δ_D are integrable with respect to the source covariate marginal P_S^X:*

$$\Delta_Y\in L^1(P_S^X) \quad \textit{and} \quad \Delta_D\in L^1(P_S^X).$$

Proof of Lemma 8. By Lemma 7, the source covariate marginal agrees with the source-population covariate marginal:

$$P_S^X = P_n^F(\,\cdot \mid S = 1) \circ X^{-1}.$$

The same result gives the assigned-source probability structure needed to identify the source-population conditional law.

From Assumption 1, under $P_n^F(\cdot \mid S = 1)$,

$$Y(0), Y(1) \in [0, 1].$$

Using Assumption 7, the assignment-outcome contrast equals

$$Y(D(1)) - Y(D(0))$$

almost surely, and hence

$$|Y(D(1)) - Y(D(0))| \le 1 \qquad P_n^F(\cdot \mid S = 1)\text{-almost surely.}$$

Therefore, by the conditional contrast representation in Assumption 9, for every measurable A,

$$\left| \int_A \Delta_Y(x)\, dP_S^X(x) \right| \le P_S^X(A).$$

A standard level-set argument turns this setwise domination into an almost-sure pointwise bound. If a measurable function f satisfies

$$\left| \int_A f\, d\mu \right| \le \mu(A)$$

for every measurable A, then $|f| \le 1$ μ-almost surely: apply the inequality to the sets $\{q < f \le m\}$ and $\{-m \le f < -q\}$, with $q > 1$ and $m \in \mathbb{N}$, to show those sets have measure zero, then let $q \downarrow 1$. Applying this to $f = \Delta_Y$ and $\mu = P_S^X$ yields

$$|\Delta_Y(x)| \le 1 \qquad P_S^X\text{-almost surely.}$$

Since Δ_Y is measurable and P_S^X is a probability measure, this bound implies

$$\Delta_Y \in L^1(P_S^X).$$

For the receipt contrast, the binary nature of $D(0)$ and $D(1)$ gives

$$|D(1) - D(0)| \le 1 \qquad P_n^F(\cdot \mid S = 1)\text{-almost surely.}$$

By Assumption 10, for every measurable A,

$$\left| \int_A \Delta_D(x)\, dP_S^X(x) \right| \le P_S^X(A).$$

The same level-set domination argument gives

$$|\Delta_D(x)| \le 1 \qquad P_S^X\text{-almost surely.}$$

Because Δ_D is measurable and P_S^X is a probability measure, it follows that

$$\Delta_D \in L^1(P_S^X).$$

$\square$

The integrability statement in Lemma 8 discharges the side conditions attached to the identification result. Once a law belongs to the transported model class, the conditional contrasts can be integrated against the source covariate marginal.

Lemma 9. [`lem:class-compact-causal-range`] (Compact causal range on a model class) *Fix an index n, target sample-size and envelope schedules N_n and k_n, a target-to-source ratio c, and the instrument-overlap constant ε. Let $\mathcal{C}$ be a family of transported arrays with the property that every $P \in \mathcal{C}$ satisfies each clause of the transported complier-effect model class $\mathcal{P}_n$ of Definition 1 at index n, that is, Assumptions 1 to 15 hold at index n. The class $\mathcal{P}_n$ itself and, for $\mathfrak{g} \in \mathcal{G}$ as in Definition 8, the fixed-geometry slice $\mathcal{P}_n(\mathfrak{g})$ of Definition 9 are such families. Then, for every $P \in \mathcal{C}$, the conclusions of Theorem 1 hold with no further hypothesis. In particular Δ_Y and Δ_D agree P_S^X-almost everywhere with the source conditional outcome and receipt contrasts,*

$$\mu_{Y,n} = \int \big(Y(1)-Y(0)\big)\mathbf{1}\{D(1)=1,\ D(0)=0\}\, dP_n^F(\cdot \mid S=0), \qquad \mu_n = P_n^F\{D(1)=1,\ D(0)=0 \mid S=0\},$$

and

$$\theta_T = \frac{\mu_{Y,n}}{\mu_n} \in \Theta = [-1,1].$$

The measurability and integrability side conditions on Δ_Y and Δ_D that Theorem 1 carries as hypotheses are supplied from class membership alone by Lemma 8, so a consumer working at $P \in \mathcal{P}_n$ or at $P \in \mathcal{P}_n(\mathfrak{g})$ need not discharge them separately.

Proof of Lemma 9. Fix $P \in \mathcal{C}$. By the defining property of $\mathcal{C}$, P satisfies all clauses of the transported model class $\mathcal{P}_n$ at the index n. In particular, it is an admissible input for the source-observation facts of Lemma 7.

First supply the side conditions required by Theorem 1. The source-observation facts give, for every measurable $A \subseteq \mathcal{X}$,

$$\int_{\{X\in A\}} H(O^S)Y\, dP_S = \int_A \Delta_Y(x)\, dP_S^X(x), \qquad \int_{\{X\in A\}} H(O^S)D\, dP_S = \int_A \Delta_D(x)\, dP_S^X(x),$$

and also

$$\Delta_Y(X) \in [-1,1], \qquad \Delta_D(X) \in [0,1], \qquad P_S^X\text{-a.s.}$$

Consequently

$$\Delta_Y \in L^1(P_S^X), \qquad \Delta_D \in L^1(P_S^X).$$

Now apply Theorem 1 to this P with those two integrability facts. It gives the observed source contrast representations

$$\int_{\{X\in A\}} H(O^S)Y\, dP_S = \int_A \Delta_Y(x)\, dP_S^X(x), \qquad \int_{\{X\in A\}} H(O^S)D\, dP_S = \int_A \Delta_D(x)\, dP_S^X(x)$$

for every measurable A, and the transported target identities

$$\int w(x)\Delta_Y(x)\, dP_S^X(x) = \mu_{Y,n}, \qquad \int w(x)\Delta_D(x)\, dP_S^X(x) = \mu_n.$$

The same application identifies the Wald ratio with the target complier contrast:

$$\frac{\mu_{Y,n}}{\mu_n} = \theta_T = \mathbb{E}_{P_n^F}\left[Y(1)-Y(0) \mid D(1)=1,\ D(0)=0,\ S=0\right],$$

where

$$\mu_{Y,n} = \int \big(Y(1) - Y(0)\big) \mathbf{1}\{D(1) = 1,\ D(0) = 0\}\, dP_n^F(\,\cdot \mid S = 0), \qquad \mu_n = P_n^F\{D(1) = 1,\ D(0) = 0 \mid S = 0\}.$$

Since $Y(0), Y(1) \in [0, 1]$ on the full-data support and the target complier event has positive probability by target-complier positivity, this conditional contrast lies in

$$\Theta = [-1, 1].$$

Thus all conclusions of Theorem 1 hold for the fixed $P \in \mathcal{C}$. Since P was arbitrary, they hold for every member of $\mathcal{C}$. □

Lemma 9 makes Theorem 1 available uniformly on the model classes used by the minimax statements. In particular, it records that the fixed-geometry slice inherits the same identified target contrast and compact parameter space.

The final two auxiliary results support the expected-length bounds for score inversion. The first is a generic moment inequality for the first-stage denominator, and the second is the conditional mean and variance calculation used by the regular-cell feasible rule.

Lemma 10. [`lem:score-receipt-bad-slope-probability`] (Bad-slope probability for the first-stage mean) *Assume the following conditions.*

- ***(Sample size and overlap.)*** *$n \geq 1$ and the instrument-overlap constant satisfies $\varepsilon > 0$, as in Assumption 4.*
- ***(Strength inputs.)*** *$P \in \mathcal{P}_n$ of Definition 1 has transported first-stage mean $\mu_n(P) > 0$ as in Assumption 15 and transport-weight second moment $\kappa_n(P) > 0$ as in Assumption 14, and $t_n(P) = n\mu_n(P)^2/\kappa_n(P)$ is the effective identification strength of Definition 4.*
- ***(Moment facts.)*** *$\widehat{B}$ is a square-integrable statistic of the source sample with*

$$\mathbb{E}_P[\widehat{B}] = \mu_n(P), \qquad \mathrm{Var}_P(\widehat{B}) \leq \frac{\kappa_n(P)}{\varepsilon^2 n}.$$

Then

$$P\left(\frac{\mu_n(P)}{2} < \left|\widehat{B} - \mu_n(P)\right|\right) \leq \frac{4}{\varepsilon^2 t_n(P)}.$$

The bound is law-agnostic: only the two displayed moment facts are used. It therefore applies verbatim to the weighted first-stage sample mean

$$\widehat{B}_n = \frac{1}{n} \sum_{i=1}^{n} w(X_i) H(O_i^S) D_i$$

of Lemma 7 and Assumption 13, once that estimator has separately been shown to satisfy $\mathbb{E}_P[\widehat{B}_n] = \mu_n(P)$ and $\mathrm{Var}_P(\widehat{B}_n) \leq \kappa_n(P)/(\varepsilon^2 n)$, and to the fixed-geometry version of the same construction.

Proof of Lemma 10. Let $\mu = \mu_n(P)$, $\kappa = \kappa_n(P)$, and $t = t_n(P)$. By assumption, $\mu > 0$, $\kappa > 0$, and

$$t = \frac{n\mu^2}{\kappa}.$$

Apply Chebyshev's inequality to the square-integrable statistic $\widehat{B}$, with center μ, variance bound $\kappa/(\varepsilon^2 n)$, and threshold $a = \mu/2 > 0$. This gives

$$P\Big(\frac{\mu}{2} < |\widehat{B} - \mu|\Big) \le \frac{\kappa/(\varepsilon^2 n)}{(\mu/2)^2}.$$

The right-hand side is

$$\frac{\kappa/(\varepsilon^2 n)}{(\mu/2)^2} = \frac{4\kappa}{\varepsilon^2 n\mu^2} = \frac{4}{\varepsilon^2\, t},$$

and substituting back $\mu = \mu_n(P)$ and $t = t_n(P)$ proves the claim. □

Lemma 10 controls the event on which the estimated first-stage slope is too small relative to its mean. The bound depends on the effective strength t_n, matching the scale that appears in the oracle and finite-cell expected-length rates.

Lemma 11. [`lem:regular-cell-conditional-moment-bound`] (Regular-cell conditional mean and variance) *Assume the following conditions.*

- ***(Regular-cell law.)*** *$n \ge 1$ and $P \in \mathcal{N}_n^{\mathrm{reg}}$, the regular finite-cell class of Definition 12, with known source cell probabilities $q_{x,n}$ on the witnessing measurable cells and propensity $e(\cdot)$; in particular the instrument overlap $\varepsilon \le e(X) \le 1 - \varepsilon$ of Assumption 4 holds, and the source observation (X, Z, D, Y) of Assumption 5 has $Y \in [0, 1]$ and $D \in \{0, 1\}$.*

- ***(Candidate value.)*** *$\vartheta \in \Theta = [-1, 1]$, as in Algorithm 1.*

- ***(Target sample.)*** *A realized target covariate sample $(X_j^T)_{j=1}^{N_n}$ of Assumption 3, with empirical cell frequencies $\widehat{p}_{x,n}$, is held fixed, and the source observations are i.i.d. from the source law.*

Let $\widehat{M}_{Y,n}$ and $\widehat{M}_{D,n}$ be the transported plug-in contrast estimators built from the inverse-propensity source scores $H(O^S)$ of Lemma 7 with cell weights $\widehat{p}_{x,n}/q_{x,n}$. Then, conditionally on the target sample,

$$\mathbb{E}_P\Big[\widehat{M}_{Y,n} - \vartheta\widehat{M}_{D,n} \mid (X_j^T)_{j=1}^{N_n}\Big] = \sum_x \widehat{p}_{x,n}\,(\Delta_Y(x) - \vartheta\Delta_D(x))$$

and

$$\mathrm{Var}_P\Big(\widehat{M}_{Y,n} - \vartheta\widehat{M}_{D,n} \mid (X_j^T)_{j=1}^{N_n}\Big) \le \frac{4}{\varepsilon^2 n}\sum_x \frac{\widehat{p}_{x,n}^2}{q_{x,n}},$$

where Δ_Y and Δ_D are the source conditional contrasts of Theorem 1. The variance bound follows from the pointwise envelope

$$\big|H(O^S)(Y - \vartheta D)\big| \le \frac{2}{\varepsilon},$$

which uses the bounded source outcome $Y \in [0, 1]$ and the Boolean receipt $D \in \{0, 1\}$, giving $|Y - \vartheta D| \le 2$ for $\vartheta \in \Theta$, in addition to overlap; overlap and $\Theta = [-1, 1]$ alone do not entail the bound.

Proof of Lemma 11. Fix the realized target sample. The regular-cell class supplies measurable cells x_i, $i \in \{1, \ldots, k_n\}$, carrying the source covariate law, with

$$q_i = q_{x_i,n} = P_S^X\{X = x_i\} > 0.$$

Write

$$\widehat{p}_i = \widehat{p}_{x_i,n}, \qquad H_e(O^S) = \frac{\mathbf{1}\{Z=1\}}{e(X)} - \frac{\mathbf{1}\{Z=0\}}{1-e(X)}, \qquad G_\vartheta(O^S) = H_e(O^S)(Y - \vartheta D).$$

The contrast estimator, conditionally on the target sample, is the i.i.d. source average

$$\widehat{M}_{Y,n} - \vartheta \widehat{M}_{D,n} = \frac{1}{n} \sum_{r=1}^{n} \frac{\widehat{p}_{X_r,n}}{q_{X_r,n}} \, G_\vartheta(O_r^S),$$

where $\widehat{p}_{X_r,n}/q_{X_r,n}$ means $\widehat{p}_i/q_i$ on the cell $X_r = x_i$.

For each cell, the conditional cell mean equals the affine source contrast:

$$\frac{1}{q_i} \int_{\{X=x_i\}} G_\vartheta(O^S) \, dP_S = \frac{1}{q_i} \left[\int_{\{X=x_i\}} H_e(O^S) Y \, dP_S - \vartheta \int_{\{X=x_i\}} H_e(O^S) D \, dP_S \right].$$

By Lemma 7,

$$\int_{\{X=x_i\}} H_e(O^S) Y \, dP_S = q_i \Delta_Y(x_i), \qquad \int_{\{X=x_i\}} H_e(O^S) D \, dP_S = q_i \Delta_D(x_i).$$

Therefore

$$\frac{1}{q_i} \int_{\{X=x_i\}} G_\vartheta(O^S) \, dP_S = \Delta_Y(x_i) - \vartheta \Delta_D(x_i).$$

Taking the mean of the displayed i.i.d. average gives

$$\mathbb{E}_P\left[\widehat{M}_{Y,n} - \vartheta \widehat{M}_{D,n} \mid (X_j^T)_{j=1}^{N_n} \right] = \sum_i \widehat{p}_i \big(\Delta_Y(x_i) - \vartheta \Delta_D(x_i) \big).$$

For the variance, overlap gives $|H_e(O^S)| \le 1/\varepsilon$. Since $Y \in [0,1]$, $D \in \{0,1\}$, and $\vartheta \in [-1,1]$,

$$|Y - \vartheta D| \le |Y| + |\vartheta| \, |D| \le 2.$$

Thus

$$|G_\vartheta(O^S)| = |H_e(O^S)| \, |Y - \vartheta D| \le \frac{2}{\varepsilon}.$$

Let

$$F(O^S) = \frac{\widehat{p}_{X,n}}{q_{X,n}} \, G_\vartheta(O^S).$$

Because the source observations are conditionally i.i.d.,

$$\operatorname{Var}_P\left(\frac{1}{n} \sum_{r=1}^{n} F(O_r^S) \mid (X_j^T)_{j=1}^{N_n} \right) = \frac{1}{n} \operatorname{Var}_P(F(O^S)) \le \frac{1}{n} \mathbb{E}_P[F(O^S)^2].$$

Using the cell support and the envelope just proved,

$$\mathbb{E}_P[F(O^S)^2] = \sum_i \left(\frac{\widehat{p}_i}{q_i}\right)^2 \int_{\{X=x_i\}} G_\vartheta(O^S)^2 \, dP_S \le \sum_i \left(\frac{\widehat{p}_i}{q_i}\right)^2 \frac{4}{\varepsilon^2} q_i = \frac{4}{\varepsilon^2} \sum_i \frac{\widehat{p}_i^2}{q_i}.$$

Consequently,

$$\mathrm{Var}_P\left(\widehat{M}_{Y,n} - \vartheta \widehat{M}_{D,n} \mid (X_j^T)_{j=1}^{N_n}\right) \le \frac{4}{\varepsilon^2 n} \sum_i \frac{\widehat{p}_i^2}{q_i}.$$

Replacing the finite index i by the cell notation x gives the two displayed formulas in the statement. □

Lemma 11 supplies the conditional score centering and variance control for the regular-cell procedure. The known cell probabilities enter only through the weights $\widehat{p}_{x,n}/q_{x,n}$ and the resulting dispersion term, which is the regular-cell analogue of the collision-based scale used in Theorem 6.

A.1 Verification note

The formal derivations state their assumptions and conclusions explicitly. The note covers the local mathematical claims encoded in the assumptions, definitions, auxiliary lemmas, and main theorems displayed in the paper, including the identification identities, compact-range consequences, geometry-handle construction, fixed-geometry comparisons, score-inversion moment bounds, finite-cell dispersion controls, and regular-cell variance bounds. Citations to the econometric weak-instrument and ratio-inference literature, including Anderson and Rubin [1949], Fieller [1954], Moreira [2003], Andrews et al. [2006], and Mikusheva [2010], identify external scholarly context for Anderson–Rubin, Fieller, and robust-IV procedures; Gleser and Hwang [1987] and Dufour [1997] identify impossibility phenomena for nearly identified models.

For inspection, the Lean development is the module tree `CausalSmith/Stat/STAT_TransportedLateStrength` of the CausalSmith repository at commit `999467d`, built with `lake build` against the toolchain pinned in that repository's `lean-toolchain` and `lake-manifest.json`. Each displayed statement records the declaration name it was checked against, giving a direct map from theorem labels to Lean declarations.

B Proofs of the main results

Proof of Theorem 4. Fix the admissible geometry array $\mathfrak{g} = (P_{S,n}^X, w_n, e_n)_{n\ge1}$. We prove the three asserted conclusions in the order in which they appear.

1. Fix n. By Definition 8 the geometry second moment at index n is

$$\kappa_n = \int w_n(x)^2 \, dP_{S,n}^X(x),$$

and $P_{S,n}^X$ is a probability law on $\mathcal{X}$. The unit-weight hypothesis states $w_n(x) = 1$ for $P_{S,n}^X$-almost every x, hence also

$$w_n(x)^2 = 1 \qquad P_{S,n}^X\text{-a.s.}$$

Two functions that agree almost everywhere have the same integral, so

$$\kappa_n = \int w_n(x)^2 \, dP^X_{S,n}(x) = \int 1 \, dP^X_{S,n}(x) = 1.$$

For the equality of covariate laws, let $A \subseteq \mathcal{X}$ be measurable. The change-of-measure clause of Definition 8 states

$$P_{T,n}(A) = \int_A w_n(x) \, dP^X_{S,n}(x).$$

A $P^X_{S,n}$-almost-sure identity remains valid after restricting $P^X_{S,n}$ to A. Therefore the identity $w_n = 1$ gives

$$P_{T,n}(A) = \int_A 1 \, dP^X_{S,n}(x) = P^X_{S,n}(A).$$

Since the two measures agree on every measurable set,

$$P_{T,n} = P^X_{S,n}.$$

2. Fix n and let $P \in \mathcal{P}_n(\mathfrak{g})$. Write

$$w^P_n = \frac{dP_{T,n}}{dP^X_{S,n}}$$

for the transport-weight version induced by the laws of P, and write

$$\kappa^P_n = \int \left\{ w^P_n(x) \right\}^2 dP^X_{S,n}(x), \qquad \mu_n = \int w^P_n(x) \, \Delta_D(x) \, dP^X_{S,n}(x)$$

for the associated second moment and transported first-stage mean. In the notation of Definition 4, the effective strength of P at index n is

$$t_n = \frac{n\mu_n^2}{\kappa^P_n}.$$

By Definition 9, membership $P \in \mathcal{P}_n(\mathfrak{g})$ means that $P \in \mathcal{P}_n$ and that the index-n source covariate law, transport weight, and propensity of P are those of $\mathfrak{g}$. In particular the source covariate marginal of P is $P^X_{S,n}$ and

$$w^P_n(x) = w_n(x) \qquad P^X_{S,n}\text{-a.s.}$$

(the almost-sure form records that a density-ratio version is determined up to $P^X_{S,n}$-null sets). Integrating the squares of two functions that agree $P^X_{S,n}$-almost everywhere against the common law $P^X_{S,n}$ gives

$$\kappa^P_n = \int \left\{ w^P_n(x) \right\}^2 dP^X_{S,n}(x) = \int w_n(x)^2 \, dP^X_{S,n}(x) = \kappa_n.$$

By the first step $\kappa_n = 1$, so $\kappa^P_n = 1$, and substituting this into the display for t_n yields

$$t_n = \frac{n\mu_n^2}{\kappa^P_n} = \frac{n\mu_n^2}{\kappa_n} = \frac{n\mu_n^2}{1} = n\mu_n^2.$$

3. Fix $t_0 > 0$. We verify the hypotheses of Theorem 5 for the same $\mathfrak{g}$, c, ε, α, k_n, and N_n. Its target-sample-ratio hypothesis is $c > 0$ with $N_n/n \to c$; its envelope-scale hypothesis is $k_n > 0$ for every n, $k_n \to \infty$, and $k_n/\sqrt{n} \to 0$; its overlap and noncoverage hypotheses are $0 < \varepsilon < 1/2$ and $0 < \alpha < 1$; and its geometry hypothesis is admissibility of $\mathfrak{g}$ at scale k_n and overlap constant ε. These are exactly the standing hypotheses of the present statement.

   Its remaining hypothesis is the fixed-geometry slice condition: for every n and every transported array whose index-n law lies in $\mathcal{P}_n(\mathfrak{g})$, the conditions of Assumptions 3, 4 and 13 to 15 hold. This holds because Definition 9 defines $\mathcal{P}_n(\mathfrak{g})$ as a subset of $\mathcal{P}_n$, and each of those five conditions is one of the requirements listed in Definition 1; membership in $\mathcal{P}_n$ therefore supplies all five directly. Thus all hypotheses of Theorem 5 hold for $\mathfrak{g}$.

   Applying Theorem 5 at this t_0 therefore gives, with

$$L_\alpha = \sqrt{\frac{8}{\alpha\varepsilon^2}}, \qquad c_\alpha = \frac{3(1-\alpha)^2}{16}, \qquad C_\alpha = \max\left\{2,\ 4L_\alpha + \frac{8}{\varepsilon^2}\right\},$$

   the two-sided bound on the fixed-geometry minimax value of Definition 11,

$$c_\alpha \min\{1, t_0^{-1/2}\} \le V_{\mathfrak{g}}^\star(t_0) \le C_\alpha \min\{1, t_0^{-1/2}\},$$

   which is the asserted conclusion since $t_0 > 0$ was arbitrary.

$\square$

Proof of Theorem 6. 1. By Lemma 2, the finite-cell class is inhabited at every row under the carrier, ratio, overlap, and growth hypotheses. Define

$$B_c = 32(1 + c^{-1}), \qquad L_{\alpha,c} = \left(\frac{2B_c}{\alpha}\right)^{1/2}, \qquad C_0 = \max\{2, 4\sqrt{2}\, L_{\alpha,c} + 8B_c\}.$$

Since $c > 0$, $B_c > 0$; since $\alpha > 0$, $L_{\alpha,c} \ge 0$; hence

$$C_0 \ge 2 > 0.$$

Set

$$\xi_c = 4 + \frac{3}{c}, \qquad \eta_c = \xi_c^{-1/2}, \qquad \bar{\varepsilon} = \frac{1}{4} + \frac{1}{2}\eta_c.$$

Then $\xi_c > 4$, so $\sqrt{\xi_c} > 2$, $0 < \eta_c < 1/2$, and therefore

$$0 < \bar{\varepsilon} < \frac{1}{2}.$$

Moreover,

$$\bar{\varepsilon}\sqrt{\xi_c} = \frac{1}{4}\sqrt{\xi_c} + \frac{1}{2} \ge 1,$$

and consequently

$$\bar{\varepsilon}^{-1} \le \sqrt{\xi_c}, \qquad \bar{\varepsilon}^{-2} \le \xi_c = 4 + \frac{3}{c}.$$

For the regular-cell variance constant at this interior overlap level,

$$B_{\bar{\varepsilon},c} = 8(\bar{\varepsilon}^{-2} + c^{-1}),$$

we obtain

$$B_{\bar{\varepsilon},c} \le 8\left(4+\frac{3}{c}+\frac{1}{c}\right) = 32(1+c^{-1}) = B_c.$$

Thus

$$\left(\frac{2B_{\bar{\varepsilon},c}}{\alpha}\right)^{1/2} \le \left(\frac{2B_c}{\alpha}\right)^{1/2} = L_{\alpha,c}.$$

Applying Theorem 7 with

$$c_- = c_+ = 1, \qquad \varepsilon = \bar{\varepsilon}, \qquad L = L_{\alpha,c},$$

gives a regular-cell procedure C^{reg} that is honest over the unit-regular class and satisfies, for every $t_0 > 0$,

$$R_{\mathrm{reg}}(C^{\mathrm{reg}}, t_0) \le \max\{2, 4\sqrt{2}\, L_{\alpha,c} + 8B_{\bar{\varepsilon},c}\} \min\{1, t_0^{-1/2}\}.$$

Since $B_{\bar{\varepsilon},c} \le B_c$,

$$R_{\mathrm{reg}}(C^{\mathrm{reg}}, t_0) \le C_0 \min\{1, t_0^{-1/2}\}.$$

2. Define the finite-cell procedure C° by

$$C_n^\circ(s^S, s^T) = \mathrm{Inv}_{k_n,n,N_n,L_{\alpha,c}}(s^S, s^T),$$

where the inversion rule uses source mass $q_x = 1/k_n$ and propensity $e(x) = 1/2$, and where the totalized convention is

$$C_n^\circ(s^S, s^T) = \Theta \qquad \text{when } N_n < 2.$$

For $N_n \ge 2$, this is exactly

$$H_i = 2(2Z_i - 1),$$

$$\widehat{p}_{x,n} = \frac{1}{N_n}\sum_{j=1}^{N_n} \mathbf{1}\{X_j^T = x\},$$

$$\widehat{m}_{G,n}(x) = \frac{k_n}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i = x\} H_i G_i, \qquad \widehat{M}_{G,n} = \sum_{x=1}^{k_n} \widehat{p}_{x,n}\widehat{m}_{G,n}(x),$$

and

$$\widehat{K}_n = 1 + \frac{k_n}{N_n(N_n-1)}\sum_{j\ne\ell} \mathbf{1}\{X_j^T = X_\ell^T\}.$$

Hence

$$C_n^\circ = \left\{\vartheta \in \Theta : |\widehat{M}_{Y,n} - \vartheta \widehat{M}_{D,n}| \le L_{\alpha,c}\sqrt{\widehat{K}_n/n}\right\}.$$

Fix a finite-cell row $P \in \mathcal{N}_n$, and take the witnessing injected cells from Assumption 16. On those cells the source masses are $1/k_n$, and the propensity is $1/2$ source-almost surely. The same finite-cell source condition gives source covariate support on the injected cells; by the transport-domination clause in Definition 2, the target covariate law is also supported on the injected cells. Therefore, under the product two-sample law, all source and target sample covariates lie in the injected support almost surely, and the source propensities equal 1/2 at all sampled source covariates almost surely. On this event the regular-cell ambient extensions of the law's cell masses and propensity evaluate as the constants $1/k_n$ and $1/2$ on every sampled argument, so the regular-cell set C^{reg}, evaluated with the row's regular-cell design, agrees almost surely with C_n°.

The finite-to-regular bridge then gives

$$C^{\circ} \in \mathfrak{C}^{\mathrm{cell}}$$

and, for every $t_0 > 0$,

$$R_{\mathrm{cell}}(C^{\circ}, t_0) \leq R_{\mathrm{reg}}(C^{\mathrm{reg}}, t_0).$$

Combining this inequality with the bound in the previous step yields

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{N}_n:\, t_n\geq t_0} \mathbb{E}_P[\lambda(C_n^{\circ})] \leq C_0 \min\{1, t_0^{-1/2}\}.$$

3. It remains to prove the lower bound. Choose the injected cells from the carrier hypothesis and denote them by

$$\iota_{n,1}, \ldots, \iota_{n,k_n}.$$

Let $P_{\circ,n}^X$ be the discrete probability law

$$P_{\circ,n}^X = \frac{1}{k_n} \sum_{r=1}^{k_n} \delta_{\iota_{n,r}}.$$

Because $k_n > 0$,

$$P_{\circ,n}^X(\mathcal{X}) = 1, \qquad P_{\circ,n}^X\{\iota_{n,r}\} = \frac{1}{k_n}, \qquad P_{\circ,n}^X\big(\{\iota_{n,1}, \ldots, \iota_{n,k_n}\}\big) = 1.$$

Define a fixed geometry $\mathfrak{g}^{\circ}$ by

$$P_{S,n}^{X,\circ} = P_{T,n}^{\circ} = P_{\circ,n}^X, \qquad w_n(x) = 1, \qquad e_n(x) = \frac{1}{2}.$$

This geometry is admissible in the sense of Definition 8. Indeed, the source and target covariate laws are probabilities; since $0 < \varepsilon < 1/2$,

$$\varepsilon \leq \frac{1}{2} \leq 1 - \varepsilon;$$

and, using $k_n > 0$,

$$0 \leq 1 \leq 2k_n, \qquad \int 1 \, dP_{\circ,n}^X = 1, \qquad \int 1^2 \, dP_{\circ,n}^X = 1 \leq k_n.$$

For every measurable $A \subseteq \mathcal{X}$, the target-law identity is

$$P_{T,n}^{\circ}(A) = P_{\circ,n}^X(A) = \int_A 1 \, dP_{\circ,n}^X.$$

A row in the fixed slice $\mathcal{P}_n(\mathfrak{g}^{\circ})$ of Definition 9 has uniform source mass $1/k_n$ on the injected support and propensity $1/2$, so

$$\mathcal{P}_n(\mathfrak{g}^{\circ}) \subseteq \mathcal{N}_n \qquad \text{for every } n.$$

For each $t_0 > 0$, Lemma 3 supplies rows in this fixed slice with $t_n \geq t_0$ for all sufficiently large n.

4. Fix any $\widetilde{C} \in \mathfrak{C}^{\text{cell}}$. For a two-sample realization $s = (s^S, s^T)$, define the support event

$$\mathcal{S}_n(s) = \begin{cases} X_i^S \in \{\iota_{n,1}, \ldots, \iota_{n,k_n}\} \text{ for every source index } i, \\ X_j^T \in \{\iota_{n,1}, \ldots, \iota_{n,k_n}\} \text{ for every target index } j. \end{cases}$$

Construct an oracle procedure $D^{\widetilde{C}}$ on the fixed geometry by

$$D_n^{\widetilde{C}}(s, w, e) = \begin{cases} \widetilde{C}_n(s), & \mathcal{S}_n(s), \\ \Theta, & \mathcal{S}_n(s)^c. \end{cases}$$

The event $\mathcal{S}_n$ is measurable because it is built from the finite union of the measurable singleton cells, and the displayed rule is a measurable oracle rule that always returns a subset of Θ.

For every $P \in \mathcal{P}_n(\mathfrak{g}^\circ)$, both the source and target covariate marginals assign probability one to the injected support. Under the product two-sample law,

$$\mathbb{P}_P\{\mathcal{S}_n\} = 1.$$

Consequently, for every $P \in \mathcal{P}_n(\mathfrak{g}^\circ)$,

$$D_n^{\widetilde{C}} = \widetilde{C}_n \qquad P\text{-almost surely under the two-sample law,}$$

and the coverage and expected-length quantities agree in the self-contained form

$$\mathbb{P}_P\{\theta_T(P) \in D_n^{\widetilde{C}}\} = \mathbb{P}_P\{\theta_T(P) \in \widetilde{C}_n\}, \qquad \mathbb{E}_P[\lambda(D_n^{\widetilde{C}})] = \mathbb{E}_P[\lambda(\widetilde{C}_n)].$$

Since $\mathcal{P}_n(\mathfrak{g}^\circ) \subseteq \mathcal{N}_n$ and the fixed slice is eventually nonempty at strength t_0, the rowwise coverage infimum over $\mathcal{N}_n$ is bounded above by the rowwise coverage infimum over $\mathcal{P}_n(\mathfrak{g}^\circ)$ eventually. The honesty of $\widetilde{C}$ in Definition 7 therefore implies that $D^{\widetilde{C}}$ is honest on the fixed geometry in the sense of Definition 10. Similarly,

$$R_{\mathfrak{g}^\circ}(D^{\widetilde{C}}, t_0) \le R_{\text{cell}}(\widetilde{C}, t_0),$$

with the empty-row convention harmless because the fixed slice is eventually inhabited; the length bound used for the rowwise suprema is

$$\lambda(A) \le \lambda(\Theta) = 2$$

for every confidence set $A \subseteq \Theta$. Applying the lower half of Theorem 5 to $\mathfrak{g}^\circ$ gives

$$\frac{3(1-\alpha)^2}{16} \min\{1, t_0^{-1/2}\} \le \inf_{D \in \mathfrak{C}^{\text{or}}(\mathfrak{g}^\circ)} R_{\mathfrak{g}^\circ}(D, t_0) \le R_{\mathfrak{g}^\circ}(D^{\widetilde{C}}, t_0) \le R_{\text{cell}}(\widetilde{C}, t_0).$$

Taking the infimum over all $\widetilde{C} \in \mathfrak{C}^{\text{cell}}$ yields

$$\frac{3(1-\alpha)^2}{16} \min\{1, t_0^{-1/2}\} \le \inf_{\widetilde{C} \in \mathfrak{C}^{\text{cell}}} \limsup_{n\to\infty} \sup_{P \in \mathcal{N}_n : t_n \ge t_0} \mathbb{E}_P[\lambda(\widetilde{C}_n)].$$

Together with the construction of C° and the positivity of C_0, this proves the claim.

□

Proof of Theorem 1.

1. Let $P_{\mathrm{assign},S}$ be the assigned-source law on full data and assignment, let P_S be its observed-source image, and let P_S^X be the source covariate marginal. By Lemma 7, these are probability laws, the full-data marginal of $P_{\mathrm{assign},S}$ is $P_n^F(\,\cdot \mid S=1)$, and for every measurable $A \subseteq \mathcal{X}$,

$$\int_{\{O^S : X\in A\}} H(O^S)Y\,dP_S = \int_A \Delta_Y(x)\,dP_S^X(x), \qquad \int_{\{O^S : X\in A\}} H(O^S)D\,dP_S = \int_A \Delta_D(x)\,dP_S^X(x).$$

This gives the two observed-source representation conclusions in the statement.

2. For $s \in \{0,1\}$, let P_s^X be the X-marginal of $P_n^F(\,\cdot \mid S=s)$, so $P_1^X = P_S^X$ and $P_0^X = P_T$. Let $\Gamma_{Y,s}, \Gamma_{D,s} : \mathcal{X} \to \mathbb{R}$ be the measurable contrast versions carried by the model class, with their P_s^X-a.e. values fixed by

$$\int_A \Gamma_{Y,s}(x)\,dP_s^X(x) = \int_{\{X\in A\}} \{Y^Z(1) - Y^Z(0)\}\,dP_n^F(\,\cdot \mid S=s)$$

and

$$\int_A \Gamma_{D,s}(x)\,dP_s^X(x) = \int_{\{X\in A\}} \{\mathbf{1}[D(1)=1] - \mathbf{1}[D(0)=1]\}\,dP_n^F(\,\cdot \mid S=s)$$

for every measurable A. The source comparison uses the following inverse-propensity identity in the integrable cases used below: for measurable full-data functions F_0, F_1,

$$\int_{\{X\in A\}} \left[\frac{\mathbf{1}[Z=1]}{e(X)}F_1 - \frac{\mathbf{1}[Z=0]}{1-e(X)}F_0\right] dP_{\mathrm{assign},S} = \int_{\{X\in A\}} (F_1 - F_0)\,dP_n^F(\,\cdot \mid S=1).$$

Indeed, splitting the left side over $Z=1$ and $Z=0$, conditional randomization gives the slice densities $e(X)$ and $1-e(X)$ with respect to $P_n^F(\,\cdot \mid S=1)$, and overlap makes the divisions legitimate; the two density factors cancel.

For the outcome contrast, apply this identity with $F_z = Y(D(z))$, the treatment-potential outcome evaluated at the receipt induced by assignment z. Under the observation map, an assigned source record with $Z=z$ has observed coordinates $D = D(z)$ and $Y = Y(D(z))$, so the left side is the observed-source score integral from Lemma 7. The right side is

$$\int_{\{X\in A\}} \{Y(D(1)) - Y(D(0))\}\,dP_n^F(\,\cdot \mid S=1),$$

and exclusion rewrites $Y(D(z))$ as $Y^Z(z)$ $P_n^F(\,\cdot \mid S=1)$-almost surely. Applying the same inverse-propensity identity to the receipt functions $F_z = \mathbf{1}[D(z)=1]$, and using the observed coordinate $D = D(Z)$, gives

$$\int_A \Delta_Y\,dP_S^X = \int_A \Gamma_{Y,1}\,dP_S^X, \qquad \int_A \Delta_D\,dP_S^X = \int_A \Gamma_{D,1}\,dP_S^X$$

for every measurable A. The source-contrast integrability hypotheses and the boundedness supplied by Lemma 7 give the integrability needed to recover equality from equality of all measurable-set integrals, hence

$$\Delta_Y = \Gamma_{Y,1} \quad P_S^X\text{-a.e.}, \qquad \Delta_D = \Gamma_{D,1} \quad P_S^X\text{-a.e.}$$

3. Let $w = dP_T/dP_S^X$. Transport domination gives the Radon–Nikodym change of measure

$$\int w(x)g(x)\,dP_S^X(x) = \int g(x)\,dP_T(x)$$

for the integrable functions used below. Therefore,

$$\begin{aligned}\int w(x)\Delta_Y(x)\,dP_S^X(x) &= \int \Gamma_{Y,1}(x)\,dP_T(x)\\ &= \int \Gamma_{Y,0}(x)\,dP_T(x)\\ &= \int \{Y^Z(1) - Y^Z(0)\}\,dP_n^F(\,\cdot \mid S=0).\end{aligned}$$

The second equality is the outcome-contrast transport clause in $\mathcal{P}_n$, and the last equality is the defining integral identity for $\Gamma_{Y,0}$ with $A = \mathcal{X}$. On $P_n^F(\,\cdot \mid S=0)$-almost every full-data record, exclusion gives $Y^Z(z) = Y(D(z))$, and binary monotonicity gives

$$Y^Z(1) - Y^Z(0) = \big(Y(1) - Y(0)\big)\mathbf{1}\{D(1)=1,\ D(0)=0\}.$$

Thus

$$\int w(x)\Delta_Y(x)\,dP_S^X(x) = \int \big(Y(1)-Y(0)\big)\mathbf{1}\{D(1)=1,\ D(0)=0\}\,dP_n^F(\,\cdot \mid S=0) = \mu_{Y,n}.$$

4. The same change of measure and the a.e. equality $\Delta_D = \Gamma_{D,1}$ give

$$\begin{aligned}\int w(x)\Delta_D(x)\,dP_S^X(x) &= \int \Gamma_{D,1}(x)\,dP_T(x)\\ &= \int \Gamma_{D,0}(x)\,dP_T(x)\\ &= \int \{\mathbf{1}[D(1)=1] - \mathbf{1}[D(0)=1]\}\,dP_n^F(\,\cdot \mid S=0).\end{aligned}$$

The second equality is the receipt-contrast transport clause, and the last equality is the defining integral identity for $\Gamma_{D,0}$ with $A = \mathcal{X}$. Binary monotonicity converts the final integrand into the complier indicator:

$$\mathbf{1}[D(1)=1] - \mathbf{1}[D(0)=1] = \mathbf{1}\{D(1)=1,\ D(0)=0\} \quad P_n^F(\,\cdot \mid S=0)\text{-a.s.}$$

Hence

$$\int w(x)\Delta_D(x)\,dP_S^X(x) = P_n^F\{D(1)=1,\ D(0)=0 \mid S=0\} = \mu_n.$$

5. Write

$$\mathcal{E}_{\mathrm{co}} = \{D(1)=1,\ D(0)=0\}, \qquad \mu = P_n^F(\,\cdot \mid S=0).$$

Target complier positivity in the model class gives

$$0 < \mu(\mathcal{E}_{\mathrm{co}}) = \mu_n.$$

Combining the two identities just proved,

$$\frac{\mu_{Y,n}}{\mu_n} = \frac{\int (Y(1) - Y(0))\mathbf{1}_{\mathcal{E}_{\mathrm{co}}}\, d\mu}{\mu(\mathcal{E}_{\mathrm{co}})} = \mathbb{E}_{P_n^F}[Y(1) - Y(0) \mid D(1) = 1,\ D(0) = 0,\ S = 0] = \theta_T.$$

Full-data support gives $0 \le Y(0), Y(1) \le 1$ μ-almost surely, so

$$-1 \le Y(1) - Y(0) \le 1 \quad \mu\text{-a.s.}$$

Multiplying by $\mathbf{1}_{\mathcal{E}_{\mathrm{co}}}$ and integrating yields

$$-\mu(\mathcal{E}_{\mathrm{co}}) \le \int (Y(1) - Y(0))\mathbf{1}_{\mathcal{E}_{\mathrm{co}}}\, d\mu \le \mu(\mathcal{E}_{\mathrm{co}}).$$

Dividing by the positive denominator $\mu(\mathcal{E}_{\mathrm{co}})$ gives

$$-1 \le \theta_T \le 1,$$

that is, $\theta_T \in \Theta = [-1, 1]$.

□

Proof of Theorem 2. Set

$$c_0 = \frac{3(1-\alpha)^2}{16}, \qquad t_c = 1.$$

Since $0 < \alpha < 1$, $1 - \alpha > 0$, and hence $c_0 > 0$; also $t_c > 0$. Fix $\mathcal{X}$, N_n, k_n, and the hypotheses in the statement, and fix $t_0 > 0$. We prove the three asserted inequalities for this t_0.

1. By Lemma 2, for every n the finite-cell class $\mathcal{N}_n$ is inhabited. Since membership in $\mathcal{N}_n$ includes membership in $\mathcal{P}_n$, the transported-IV class $\mathcal{P}_n$ is also inhabited for every n. This supplies the nonempty rows used below in the oracle coverage infima.

2. Choose, for each n, an injected measurable carrier

$$\xi_{n,\iota} \in \mathcal{X}, \qquad \iota \in I_n := \{1, \ldots, k_n\},$$

as supplied by the measurable-carrier hypothesis. Define the probability measure

$$\nu_n := \frac{1}{k_n} \sum_{\iota \in I_n} \delta_{\xi_{n,\iota}}.$$

The identity $k_n \cdot k_n^{-1} = 1$, using $k_n > 0$, gives $\nu_n(\mathcal{X}) = 1$. Now define the deterministic geometry $\mathfrak{g}$ by

$$P_{S,n}^{X,\mathfrak{g}} = \nu_n, \qquad P_{T,n}^{X,\mathfrak{g}} = \nu_n, \qquad w_n^{\mathfrak{g}}(x) = 1, \qquad e_n^{\mathfrak{g}}(x) = \frac{1}{2}.$$

This geometry is admissible: the source and target covariate laws are probabilities; the propensity satisfies

$$\varepsilon \le \frac{1}{2} \le 1 - \varepsilon$$

because $0 < \varepsilon < 1/2$; the weight satisfies $0 \le 1 \le 2k_n$ because $k_n \ge 1$; and

$$\int w_n^{\mathfrak{g}}\, dP_{S,n}^{X,\mathfrak{g}} = 1, \qquad \int (w_n^{\mathfrak{g}})^2\, dP_{S,n}^{X,\mathfrak{g}} = 1 \le k_n.$$

Finally, for every measurable $A \subseteq \mathcal{X}$,

$$P_{T,n}^{X,\mathfrak{g}}(A) = \nu_n(A) = \int_A 1 \, d\nu_n = \int_A w_n^{\mathfrak{g}} \, dP_{S,n}^{X,\mathfrak{g}},$$

so the change-of-measure requirement holds.

3. The fixed-geometry slice determined by $\mathfrak{g}$ lies inside the finite-cell submodel. Indeed, if $P \in \mathcal{P}_n(\mathfrak{g})$, then $P \in \mathcal{P}_n$ and its source covariate marginal is ν_n, so

$$P_S^X\{\xi_{n,\iota}\} = \frac{1}{k_n} \quad \text{for every } \iota \in I_n,$$

and

$$P_S^X(\{\xi_{n,\iota} : \iota \in I_n\}) = 1.$$

The same source law is exactly the uniform pushforward on the injected carrier, and the slice also gives $e = 1/2$ source-almost surely. These are the finite-cell source clauses in Definition 2, together with the transported-IV clauses already present in Definition 9. Hence

$$P \in \mathcal{P}_n(\mathfrak{g}) \quad \Longrightarrow \quad P \in \mathcal{N}_n.$$

4. Let $\mathcal{A}^{\Theta}$ be the oracle procedure that always returns

$$\Theta = [-1, 1].$$

For every $P \in \mathcal{P}_n$, the source contrasts Δ_Y and Δ_D are measurable by construction and are integrable by Lemma 8. Applying Theorem 1 gives $\theta_T(P, n) \in \Theta$. Therefore

$$\Pr_P\{\theta_T(P, n) \in \mathcal{A}_n^{\Theta}\} = 1.$$

Since each row $\mathcal{P}_n$ is inhabited, the rowwise coverage infimum is 1 for every n, and hence

$$\liminf_{n\to\infty} \inf_{P \in \mathcal{P}_n} \Pr_P\{\theta_T(P, n) \in \mathcal{A}_n^{\Theta}\} = 1 \geq 1 - \alpha.$$

Thus the oracle-honest class over which $V^\star(t_0)$ and $V_{\mathcal{N}}^\star(t_0)$ take their infima is inhabited.

5. The unit-weight reduction in Theorem 4, applied to the geometry $\mathfrak{g}$, gives

$$\kappa_n(\mathfrak{g}) = 1, \qquad P_{T,n}^{X,\mathfrak{g}} = P_{S,n}^{X,\mathfrak{g}},$$

and, for every $P \in \mathcal{P}_n(\mathfrak{g})$,

$$t_n(P) = n \, \mu_n(P)^2.$$

The same result gives the fixed-geometry lower frontier

$$c_0 \min\{1, t_0^{-1/2}\} \leq V_{\mathfrak{g}}^\star(t_0).$$

6. We next compare the fixed-geometry value with the global oracle value. Let $\mathcal{A}$ be any globally oracle-honest procedure. Evaluating $\mathcal{A}$ on the declared weight and propensity

of $\mathfrak{g}$ gives a fixed-geometry procedure $\mathcal{A}^{\mathfrak{g}}$. For every $P \in \mathcal{P}_n(\mathfrak{g})$ and every two-sample realization s,

$$\mathcal{A}_n^{\mathfrak{g}}(s) = \mathcal{A}_n(s, w_n^{\mathfrak{g}}, e_n^{\mathfrak{g}}),$$

and the fixed-geometry slice identifies $w_n^{\mathfrak{g}}$ with an allowed source-almost-sure version of the law-induced oracle weight and identifies $e_n^{\mathfrak{g}}$ with the law-induced propensity. Hence the coverage and expected-length evaluations agree on the slice:

$$\Pr_P\{\theta_T(P,n) \in \mathcal{A}_n^{\mathfrak{g}}\} = \Pr_P\{\theta_T(P,n) \in \mathcal{A}_n\},$$

and

$$\mathbb{E}_P \lambda(\mathcal{A}_n^{\mathfrak{g}}) = \mathbb{E}_P \lambda(\mathcal{A}_n).$$

By Lemma 3, the strength-restricted fixed-geometry rows are eventually inhabited. With the empty-row convention that a coverage infimum is 1, the global coverage infimum is bounded above by the fixed-slice coverage infimum eventually; both quantities lie in $[0,1]$, so the same inequality passes to the liminf. Thus $\mathcal{A}^{\mathfrak{g}}$ is fixed-geometry honest. For risk, the rowwise supremum over the fixed slice is bounded by the rowwise supremum over $\mathcal{P}_n$:

$$\sup_{\substack{P \in \mathcal{P}_n(\mathfrak{g}) \\ t_n(P) \geq t_0}} \mathbb{E}_P \lambda(\mathcal{A}_n^{\mathfrak{g}}) \leq \sup_{\substack{P \in \mathcal{P}_n \\ t_n(P) \geq t_0}} \mathbb{E}_P \lambda(\mathcal{A}_n).$$

When the left row is empty its supremum is 0; otherwise it is nonnegative. In all cases $\lambda(\mathcal{A}_n) \leq 2$, since $\mathcal{A}_n \subseteq [-1,1]$, so the row suprema are bounded by 2 and the inequality passes to the limsup. Therefore

$$V_{\mathfrak{g}}^{\star}(t_0) \leq R((\mathcal{A}_n), t_0).$$

Taking the infimum over all globally oracle-honest $\mathcal{A}$ and using the fixed-geometry lower frontier above yields

$$c_0 \min\{1, t_0^{-1/2}\} \leq V^{\star}(t_0).$$

7. The same constructed geometry gives the finite-cell lower bound. Let $\mathcal{A}$ be any globally oracle-honest procedure. Define a fixed-geometry procedure by feeding $\mathcal{A}$ the canonical density ratio

$$x \mapsto \left(\frac{dP_{T,n}^{X,\mathfrak{g}}}{dP_{S,n}^{X,\mathfrak{g}}}(x) \right)$$

and the propensity $e_n^{\mathfrak{g}}$. On the fixed slice, the source and target covariate laws equal the geometry laws and the propensity equals $e_n^{\mathfrak{g}}$, so these inputs match the oracle inputs used to evaluate $\mathcal{A}$. Thus, for every $P \in \mathcal{P}_n(\mathfrak{g})$,

$$\mathbb{E}_P \lambda(\mathcal{A}_n^{\mathfrak{g}}) = \mathbb{E}_P \lambda(\mathcal{A}_n).$$

The same coverage comparison as above, using the eventual inhabitation in Lemma 3, shows that $\mathcal{A}^{\mathfrak{g}}$ is fixed-geometry honest. By the slice inclusion already proved,

$$\{P \in \mathcal{P}_n(\mathfrak{g}) : t_n(P) \geq t_0\} \subseteq \{P \in \mathcal{N}_n : t_n(P) \geq t_0\}.$$

Consequently, for all sufficiently large n, the single-row finite-cell risk bound gives

$$\sup_{\substack{P\in\mathcal{P}_n(\mathfrak{g})\\ t_n(P)\ge t_0}} \mathbb{E}_P\lambda(\mathcal{A}_n^{\mathfrak{g}}) \le \sup_{\substack{P\in\mathcal{N}_n\\ t_n(P)\ge t_0}} \mathbb{E}_P\lambda(\mathcal{A}_n).$$

The left rows are nonnegative, and the finite-cell rows are bounded above by 2 because confidence sets are evaluated inside $\Theta = [-1, 1]$. Passing to limsups gives

$$V_{\mathfrak{g}}^{\star}(t_0) \le R_{\mathcal{N}}((\mathcal{A}_n), t_0),$$

where the right-hand side is the finite-cell restricted oracle risk. Taking the infimum over globally oracle-honest $\mathcal{A}$ and using the fixed-geometry lower frontier above gives

$$c_0 \min\{1, t_0^{-1/2}\} \le V_{\mathcal{N}}^{\star}(t_0).$$

8. Finally suppose $0 < t_0 \le t_c = 1$. Since the exponent $-1/2$ is nonpositive,

$$1 \le t_0^{-1/2}, \qquad \min\{1, t_0^{-1/2}\} = 1.$$

Substituting this into the global lower bound proved above gives

$$c_0 \le V^{\star}(t_0).$$

The constants c_0 and t_c were chosen as functions only of (α, ε, c), completing the proof. □

Proof of Theorem 3. **1. Construction.** Choose an admissible deterministic geometry $\mathfrak{g}$ supplied by the admissible-geometry hypothesis; this choice is used below in two places: it gives eventual row nonemptiness, and in the final step it carries the fixed-geometry frontier bounds across to the oracle value. Put

$$L = L_\alpha = \left(\frac{8}{\alpha\varepsilon^2}\right)^{1/2}, \qquad C_0 = \max\left\{2, 4L + \frac{8}{\varepsilon^2}\right\}.$$

Since $\alpha > 0$ and $\varepsilon > 0$, $L > 0$, and hence $C_0 \ge 2 > 0$. Define the oracle procedure from the oracle input (w, e) by

$$C_n = \left\{\vartheta \in \Theta : \left|\widehat{A}_n - \vartheta\widehat{B}_n\right| \le L\sqrt{\widehat{\kappa}_n/n}\right\},$$

with $\widehat{A}_n$, $\widehat{B}_n$, and $\widehat{\kappa}_n$ formed from that input. The set is contained in Θ by construction. Its sample-by-parameter graph is measurable for every measurable oracle pair (w, e), because the three displayed statistics are measurable functions of the source sample and the defining inequality is Borel. If two admissible versions of w agree P_S^X-almost surely, then their source-sample evaluations agree with probability one, and the same equality holds after adjoining the independent target covariate sample. Thus the random set is a valid oracle procedure and is invariant to the density-ratio representative under the product two-sample law.

2. Pointwise receipt and scale facts. Fix $n \ge 1$ and $P \in \mathcal{P}_n$. Write

$$\mu_n = \mu_n(P), \qquad \kappa_n = \kappa_n(P), \qquad t_n = t_n(P) = \frac{n\mu_n^2}{\kappa_n}.$$

By Lemma 9, $\theta_T \in \Theta$ and μ_n equals the target complier share; its strict positivity $\mu_n > 0$ comes separately from target complier positivity in Assumption 11. The change-of-measure identity for the canonical density ratio gives

$$\int w(x)\, dP_S^X(x) = 1.$$

The weight envelope gives square-integrability of w, and Cauchy–Schwarz then yields

$$1 \leq \int w(x)^2\, dP_S^X(x) = \kappa_n,$$

so $\kappa_n > 0$.

We shall use the following pullout consequence of Lemma 7. Let q be an integrable source-observation function and let d be an integrable covariate function such that, for every measurable $A \subseteq \mathcal{X}$,

$$\int_{\{X \in A\}} q(O^S)\, dP_S = \int_A d(x)\, dP_S^X(x).$$

If w is the measurable transport weight and both weighted functions are integrable, then

$$\int w(X) q(O^S)\, dP_S = \int w(x) d(x)\, dP_S^X(x).$$

Indeed, the setwise identity says that $d(X)$ is a version of the conditional expectation of $q(O^S)$ given X; multiplying by the X-measurable factor $w(X)$ and integrating gives the display.

Let

$$G_{D,n}(O^S) = w(X) H(O^S) D.$$

Applying the preceding pullout identity to the receipt-score identity in Lemma 7, with $q(O^S) = H(O^S)D$ and $d(x) = \Delta_D(x)$, and then using Lemma 9, gives

$$\mathbb{E}_P[G_{D,n}] = \int w(x) \Delta_D(x)\, dP_S^X(x) = \mu_n.$$

Overlap gives $|H(O^S)| \leq \varepsilon^{-1}$ almost surely and $D \in \{0, 1\}$, hence

$$\mathrm{Var}_P(G_{D,n}) \leq \mathbb{E}_P[G_{D,n}^2] \leq \frac{1}{\varepsilon^2} \int w(x)^2\, dP_S^X(x) = \frac{\kappa_n}{\varepsilon^2}.$$

Therefore the source-sample average $\widehat{B}_n = n^{-1} \sum_i G_{D,n}(O_i^S)$ satisfies

$$\mathbb{E}_P[\widehat{B}_n] = \mu_n, \qquad \mathrm{Var}_P(\widehat{B}_n) \leq \frac{\kappa_n}{\varepsilon^2 n}.$$

Applying Lemma 10 gives

$$P\Big(\frac{\mu_n}{2} < |\widehat{B}_n - \mu_n|\Big) \leq \frac{4}{\varepsilon^2 t_n}.$$

The empirical scale $\widehat{\kappa}_n = n^{-1} \sum_i w(X_i)^2$ is nonnegative and has mean

$$\mathbb{E}_P[\widehat{\kappa}_n] = \kappa_n.$$

3. Expected length at a fixed law. For a realized source sample set

$$r_n = L\sqrt{\widehat{\kappa}_n/n}.$$

The acceptance set is the intersection of $\Theta = [-1, 1]$ with the inverse image of $[-r_n, r_n]$ under the affine map

$$\vartheta \mapsto \widehat{A}_n - \vartheta \widehat{B}_n.$$

Consequently, always $\lambda(C_n) \leq 2$, and on

$$\left\{|\widehat{B}_n - \mu_n| \leq \frac{\mu_n}{2}\right\}$$

one has $|\widehat{B}_n| \geq \mu_n/2$ and therefore

$$\lambda(C_n) \leq \frac{2r_n}{|\widehat{B}_n|} \leq \frac{4r_n}{\mu_n}.$$

Taking expectations and using Jensen's inequality,

$$\mathbb{E}_P[r_n] \leq L\sqrt{\frac{\mathbb{E}_P[\widehat{\kappa}_n]}{n}} = L\sqrt{\frac{\kappa_n}{n}}.$$

Together with the bad-slope bound,

$$\mathbb{E}_P[\lambda(C_n)] \leq \frac{4L}{\sqrt{t_n}} + \frac{8}{\varepsilon^2 t_n}.$$

If $t_n \geq 1$, then $t_n^{-1} \leq t_n^{-1/2}$, so

$$\mathbb{E}_P[\lambda(C_n)] \leq \left(4L + \frac{8}{\varepsilon^2}\right) t_n^{-1/2}.$$

If $0 < t_n < 1$, the bound $\lambda(C_n) \leq 2$ gives

$$\mathbb{E}_P[\lambda(C_n)] \leq 2.$$

Both cases are summarized, for $n \geq 1$, as

$$\mathbb{E}_P[\lambda(C_n)] \leq C_0 \min\{1, t_n^{-1/2}\}.$$

4. Coverage at a fixed law. The row definitions in Definitions 3 and 4 start at $n \geq 1$. For the rowwise limiting argument, extend the auxiliary error sequence to the single index $n = 0$ by setting $\delta_0 = 1$; this added row is immaterial for the limiting statements. At that auxiliary index the coverage probability is nonnegative, and

$$1 - \alpha - \delta_0 = -\alpha \leq 0.$$

Now let $n \geq 1$ and set

$$\delta_n = \frac{16k_n^2}{n}.$$

Define the centered affine score

$$\xi_{n,P}(O^S) = w(X)H(O^S)\{Y - \theta_T D\}, \qquad \bar{\xi}_{n,P} = \frac{1}{n}\sum_{i=1}^{n} \xi_{n,P}(O_i^S).$$

Apply the preceding pullout identity to the two score identities in Lemma 7. With $q(O^S) = H(O^S)Y$, $d = \Delta_Y$, and with $q(O^S) = H(O^S)D$, $d = \Delta_D$, it gives

$$\int w(X)H(O^S)Y\,dP_S = \int w(x)\Delta_Y(x)\,dP_S^X(x) = \mu_{Y,n},$$

and

$$\int w(X)H(O^S)D\,dP_S = \int w(x)\Delta_D(x)\,dP_S^X(x) = \mu_n,$$

where the last equalities use Lemma 9. Combining these identities with

$$\frac{\mu_{Y,n}(P)}{\mu_n(P)} = \theta_T$$

from Lemma 9 gives

$$\mathbb{E}_P[\xi_{n,P}] = 0, \qquad \bar{\xi}_{n,P} = \widehat{A}_n - \theta_T \widehat{B}_n.$$

Since $Y \in [0,1]$, $D \in \{0,1\}$, and $\theta_T \in [-1,1]$,

$$|Y - \theta_T D| \le 2.$$

Together with overlap,

$$\xi_{n,P}(O^S)^2 \le \frac{4w(X)^2}{\varepsilon^2} \quad \text{almost surely,}$$

and hence

$$\mathrm{Var}_P(\bar{\xi}_{n,P}) \le \frac{4\kappa_n}{\varepsilon^2 n}.$$

The envelope $0 \le w \le 2k_n$ gives

$$w(X)^4 \le 4k_n^2 w(X)^2 \quad \text{almost surely,}$$

so the empirical scale satisfies

$$\mathrm{Var}_P(\widehat{\kappa}_n) \le \frac{4k_n^2\kappa_n}{n}.$$

Chebyshev's inequality and $\kappa_n \ge 1$ yield

$$P\Big(\widehat{\kappa}_n < \frac{\kappa_n}{2}\Big) \le \frac{16k_n^2}{n\kappa_n} \le \delta_n.$$

Let

$$a_{n,P} = L\sqrt{\frac{\kappa_n}{2n}}.$$

Using $L^2 = 8/(\alpha\varepsilon^2)$, another application of Chebyshev's inequality gives

$$P\big(a_{n,P} < |\bar{\xi}_{n,P}|\big) \le \frac{4\kappa_n/(\varepsilon^2 n)}{a_{n,P}^2} = \alpha.$$

On the complement of these two bad source-sample events,

$$|\widehat{A}_n - \theta_T \widehat{B}_n| = |\bar{\xi}_{n,P}| \le L\sqrt{\frac{\kappa_n}{2n}} \le L\sqrt{\frac{\widehat{\kappa}_n}{n}},$$

and $\theta_T \in \Theta$, so $\theta_T \in C_n$. The events depend only on the source sample, and the product two-sample law has the source marginal as its first factor. The union bound therefore gives, for every $P \in \mathcal{P}_n$,

$$P(\theta_T \in C_n) \ge 1 - \alpha - \delta_n.$$

5. Oracle honesty. The sequence δ_n tends to zero because $k_n/\sqrt{n} \to 0$. By Lemma 3, applied to the chosen admissible $\mathfrak{g}$ and the fixed threshold 1, the model-class rows are eventually nonempty. The coverage probabilities lie in $[0, 1]$, and the preceding display holds uniformly over $P \in \mathcal{P}_n$. Applying the rowwise limiting bound to these ordinary nonempty row infima gives

$$1 - \alpha \le \liminf_{n\to\infty} \inf_{P \in \mathcal{P}_n} P(\theta_T(P) \in C_n).$$

The explicit score-inversion formula for C_n is the procedure being evaluated, and the representative-invariance established in the construction makes the oracle evaluation well-defined. Hence $(C_n) \in \mathfrak{C}^{\mathrm{or}}$ by Definition 3.

6. Frontier risk and oracle values. Using the same harmless extension at the single auxiliary row, define

$$\ell_n(P) = \begin{cases} 0, & n = 0, \\ \mathbb{E}_P[\lambda(C_n)], & n \ge 1. \end{cases}$$

The row loss $\ell_n(P)$ is nonnegative. At the auxiliary index $n = 0$ the displayed value is 0; on the rows $n \ge 1$, the fixed-law expected-length calculation above gives

$$\ell_n(P) \le C_0 \min\{1, t_n(P)^{-1/2}\}.$$

The score-inversion row formula identifies the strength-restricted risk row with the supremum of $\ell_n(P)$ for every $n \ge 1$. On the strength-restricted row $t_n(P) \ge t_0 > 0$, the capped inverse-root map is decreasing, so

$$\min\{1, t_n(P)^{-1/2}\} \le \min\{1, t_0^{-1/2}\}.$$

Taking the row supremum and then the limiting upper supremum gives

$$R((C_n), t_0) \le C_0 \min\{1, t_0^{-1/2}\}.$$

We next record the fixed-geometry comparison that is part of the same conclusion. Let $\mathfrak{g}'$ be any admissible deterministic geometry. A law in the fixed-geometry slice $\mathcal{P}_n(\mathfrak{g}')$ of Definition 9 is, in particular, a law in $\mathcal{P}_n$, so the model-class implication hypotheses supply the two-sample, overlap, envelope, second-moment, and degradation clauses required by Theorem 5. Therefore, for every $t_0 > 0$,

$$\frac{3(1-\alpha)^2}{16} \min\{1, t_0^{-1/2}\} \le V^\star_{\mathfrak{g}'}(t_0) \le C_0 \min\{1, t_0^{-1/2}\}.$$

It remains to pass from fixed-geometry values to the global oracle value. For every oracle-honest sequence $(E_n) \in \mathfrak{C}^{\mathrm{or}}$, Lemma 6 applied to the chosen admissible geometry $\mathfrak{g}$ gives

$$V^\star_{\mathfrak{g}}(t_0) \le R((E_n), t_0).$$

Combining this inequality with the fixed-geometry lower bound just displayed yields

$$\frac{3(1-\alpha)^2}{16}\min\{1, t_0^{-1/2}\} \le R((E_n), t_0)$$

for every $(E_n) \in \mathfrak{C}^{\mathrm{or}}$. The class over which Definition 5 takes the infimum is nonempty because the constructed score-inversion sequence is oracle honest. Taking that infimum gives

$$\frac{3(1-\alpha)^2}{16}\min\{1, t_0^{-1/2}\} \le V^\star(t_0).$$

For the upper oracle-value bound, the risk of any oracle procedure is nonnegative: each row is a supremum of nonnegative expected lengths, with value 0 under the empty-row convention in Definition 4, and $\lambda(C_n) \le 2$. Thus the infimum in Definition 5 is bounded below, and the constructed honest sequence is an admissible candidate. Consequently,

$$V^\star(t_0) \le R((C_n), t_0) \le C_0 \min\{1, t_0^{-1/2}\}.$$

This proves oracle honesty of the constructed score-inversion sequence, its frontier-risk upper bound, the two-sided oracle-value bounds, and the two-sided fixed-geometry value bounds at every admissible geometry. □

Proof of Theorem 5. Fix $t_0 > 0$. By Lemma 3, the strength-restricted fixed-geometry rows

$$\{P \in \mathcal{P}_n(\mathfrak{g}) : t_n(P) \ge t_0\}$$

are nonempty for all sufficiently large n.

1. Define
$$\rho = \frac{1-\alpha}{8}, \qquad H(t_0) = \min\left\{\frac{1}{4}, \rho\, t_0^{-1/2}\right\}.$$
Then $\rho > 0$ and $H(t_0) > 0$. By Lemma 1, for all sufficiently large n there is a least-favorable family
$$\left(Q^{\mathfrak{g}}_{n,h} : |h| \le H(t_0)\right)$$
inside $\mathcal{P}_n(\mathfrak{g})$, with $t_n(Q^{\mathfrak{g}}_{n,h}) = t_0$ and
$$\left\| Q^{\mathfrak{g},N}_{n,h} - Q^{\mathfrak{g},N}_{n,0} \right\|_{\mathrm{TV}} \le 2\rho.$$
By Definition 14, the target parameter in this family is
$$\theta_T(Q^{\mathfrak{g}}_{n,h}) = \frac{1}{2} + h.$$

2. Let $\mathscr{C}$ be any fixed-geometry oracle-honest procedure. For each n, set
$$\Gamma_n(\mathscr{C}) = \inf_{P \in \mathcal{P}_n(\mathfrak{g})} Q_P^N\{\theta_T(P) \in \mathscr{C}_n\}$$
and
$$\Lambda_n(\mathscr{C}; t_0) = \sup_{\substack{P \in \mathcal{P}_n(\mathfrak{g}) \\ t_n(P) \ge t_0}} \mathbb{E}_P\big[\lambda(\mathscr{C}_n)\big],$$

using value 0 for the empty infimum or supremum rows. Here, for a law $P \in \mathcal{P}_n$, Q_P^N denotes the joint law of the source sample of size n and the target covariate sample of size N_n generated by P; for the least-favourable laws this is the notation $Q_{n,h}^{\mathfrak{g},N}$ already in use. Since every confidence set is contained in $\Theta = [-1, 1]$,

$$0 \le \Gamma_n(\mathscr{C}) \le 1, \qquad 0 \le \Lambda_n(\mathscr{C}; t_0) \le 2.$$

For all sufficiently large n, put

$$\mathcal{J} = [1/2 - H(t_0),\ 1/2 + H(t_0)].$$

Since $H(t_0) \le 1/4$, $\mathcal{J} \subseteq \Theta$. For $u \in \mathcal{J}$, define

$$Q_{n,u} = Q^{\mathfrak{g}}_{n,u-1/2}.$$

Then $Q_{n,u} \in \mathcal{P}_n(\mathfrak{g})$, $t_n(Q_{n,u}) = t_0$, $\theta_T(Q_{n,u}) = u$, and

$$\left\| Q_{n,u}^N - Q_{n,0}^{\mathfrak{g},N} \right\|_{\mathrm{TV}} \le 2\rho.$$

Tonelli's theorem, applied to the measurable graph of $\mathscr{C}_n$, gives

$$\mathbb{E}_{Q^{\mathfrak{g}}_{n,0}} \lambda(\mathscr{C}_n) = \int_\Theta Q_{n,0}^{\mathfrak{g},N}\{u \in \mathscr{C}_n\}\, du.$$

Therefore

$$\begin{aligned} \mathbb{E}_{Q^{\mathfrak{g}}_{n,0}} \lambda(\mathscr{C}_n) &\ge \int_{\mathcal{J}} Q_{n,0}^{\mathfrak{g},N}\{u \in \mathscr{C}_n\}\, du \\ &\ge \int_{\mathcal{J}} \Big[Q_{n,u}^N\{u \in \mathscr{C}_n\} - \left\| Q_{n,u}^N - Q_{n,0}^{\mathfrak{g},N} \right\|_{\mathrm{TV}} \Big]\, du \\ &\ge 2H(t_0)\big(\Gamma_n(\mathscr{C}) - 2\rho\big). \end{aligned}$$

Since $Q^{\mathfrak{g}}_{n,0}$ belongs to the strength-restricted row, this yields

$$2H(t_0)\big(\Gamma_n(\mathscr{C}) - 2\rho\big) \le \Lambda_n(\mathscr{C}; t_0)$$

eventually in n.

3. Fixed-geometry honesty gives

$$1 - \alpha \le \liminf_{n\to\infty} \Gamma_n(\mathscr{C}),$$

because the rows are eventually nonempty. The map $x \mapsto 2H(t_0)(x - 2\rho)$ is increasing, so the preceding display implies

$$2H(t_0)(1 - \alpha - 2\rho) \le \limsup_{n\to\infty} \Lambda_n(\mathscr{C}; t_0) = R_{\mathfrak{g}}(\mathscr{C}, t_0).$$

Now

$$1 - \alpha - 2\rho = \frac{3(1-\alpha)}{4}.$$

If $\rho t_0^{-1/2} \le 1/4$, then $H(t_0) = \rho t_0^{-1/2}$, and

$$2H(t_0)(1-\alpha-2\rho) = \frac{3(1-\alpha)^2}{16}\, t_0^{-1/2} \ge \frac{3(1-\alpha)^2}{16} \min\{1, t_0^{-1/2}\}.$$

If $1/4 \le \rho t_0^{-1/2}$, then $H(t_0) = 1/4$, and

$$2H(t_0)(1-\alpha-2\rho) = \frac{3(1-\alpha)}{8} \ge \frac{3(1-\alpha)^2}{16} \ge \frac{3(1-\alpha)^2}{16} \min\{1, t_0^{-1/2}\}.$$

Thus every fixed-geometry honest procedure satisfies

$$R_{\mathfrak{g}}(\mathscr{C}, t_0) \ge c_\alpha \min\{1, t_0^{-1/2}\}, \qquad c_\alpha = \frac{3(1-\alpha)^2}{16}.$$

The full rule $\mathscr{C}_n^{\mathrm{all}} = \Theta$ is fixed-geometry honest, since $\theta_T(P) \in \Theta$ for every $P \in \mathcal{P}_n(\mathfrak{g})$ by Lemma 9. Hence the honest class over which Definition 11 takes its infimum is nonempty, and

$$c_\alpha \min\{1, t_0^{-1/2}\} \le V_{\mathfrak{g}}^\star(t_0).$$

4. For the upper bound, define

$$L_\alpha = \left(\frac{8}{\alpha\varepsilon^2}\right)^{1/2}.$$

For $n \ge 1$, let the oracle score-inversion rule of Algorithm 1, specialized to the fixed geometry, be

$$\mathscr{C}_n^{\mathrm{sc}} = \left\{ \vartheta \in \Theta : \left| \widehat{A}_n - \vartheta \widehat{B}_n \right| \le L_\alpha \sqrt{\frac{\widehat{\kappa}_n}{n}} \right\},$$

where, using the fixed geometry density-ratio and propensity inputs,

$$\widehat{A}_n = \frac{1}{n}\sum_{i=1}^n w_n(X_i) \left\{ \frac{\mathbf{1}\{Z_i = 1\}}{e_n(X_i)} - \frac{\mathbf{1}\{Z_i = 0\}}{1 - e_n(X_i)} \right\} Y_i,$$

$$\widehat{B}_n = \frac{1}{n}\sum_{i=1}^n w_n(X_i) \left\{ \frac{\mathbf{1}\{Z_i = 1\}}{e_n(X_i)} - \frac{\mathbf{1}\{Z_i = 0\}}{1 - e_n(X_i)} \right\} D_i,$$

and

$$\widehat{\kappa}_n = \frac{1}{n}\sum_{i=1}^n w_n(X_i)^2.$$

At $n = 0$, set $\mathscr{C}_0^{\mathrm{sc}} = \varnothing$. The sample-by-parameter graph required in Definition 4 is measurable by the measurability of the fixed geometry weight and propensity and by closure of measurable functions under finite sums, products, absolute values, intervals, and square roots. The fixed-geometry invariance clause is immediate because the rule is written from the deterministic geometry inputs supplied to the fixed slice.

5. Fix $n \ge 1$ and $P \in \mathcal{P}_n(\mathfrak{g})$. Write

$$\mu_n = \mu_n(P), \qquad \kappa_n = \kappa_n(P), \qquad t_n = t_n(P) = \frac{n\mu_n^2}{\kappa_n}.$$

By Lemma 9, $\theta_T(P) \in \Theta$ and μ_n agrees with the target complier share. The target-complier-positivity clause in Definition 1 gives $\mu_n > 0$. The transport weight has mean one under P_S^X, so Cauchy's inequality gives

$$1 \leq \kappa_n, \qquad \kappa_n > 0.$$

We shall use the following weighted form of the source-observation identities. Suppose that $F(O^S)$ and $d(X)$ are integrable, w_n is the fixed geometry weight, the products below are integrable, and

$$\int_{\{X \in \mathcal{A}\}} F(O^S)\, dP_S = \int_{\mathcal{A}} d(x)\, dP_S^X(x)$$

for every measurable $\mathcal{A}$. Then the defining property of conditional expectation with respect to the sigma-field generated by X, followed by multiplication by the X-measurable function $w_n(X)$, gives

$$\int w_n(X) F(O^S)\, dP_S = \int w_n(x) d(x)\, dP_S^X(x).$$

The needed integrability conditions follow here from the overlap bound, the bounded outcome and contrast statements in Lemma 7 together with the binary receipt indicator carried by the model class of Definition 1, and the fixed-geometry envelope. Taking

$$F(O^S) = \left\{ \frac{\mathbf{1}\{Z = 1\}}{e(X)} - \frac{\mathbf{1}\{Z = 0\}}{1 - e(X)} \right\} D$$

and $d = \Delta_D$, Lemma 7 supplies the unweighted set-integral identity, while Lemma 9 identifies the weighted covariate integral with μ_n. Hence

$$\mathbb{E}_P[\widehat{B}_n] = \mu_n.$$

The overlap bound gives

$$\left| \left\{ \frac{\mathbf{1}\{Z = 1\}}{e(X)} - \frac{\mathbf{1}\{Z = 0\}}{1 - e(X)} \right\} D \right| \leq \varepsilon^{-1},$$

and hence

$$\operatorname{Var}_P(\widehat{B}_n) \leq \frac{\kappa_n}{\varepsilon^2 n}.$$

Therefore Lemma 10 gives

$$P\left\{ \frac{\mu_n}{2} < |\widehat{B}_n - \mu_n| \right\} \leq \frac{4}{\varepsilon^2 t_n}.$$

6. The affine inversion set has length at most 2, and on the event

$$|\widehat{B}_n - \mu_n| \leq \frac{\mu_n}{2}$$

it has length at most

$$\frac{2L_\alpha \sqrt{\widehat{\kappa}_n / n}}{|\widehat{B}_n|} \leq \frac{4L_\alpha \sqrt{\widehat{\kappa}_n / n}}{\mu_n}.$$

Since $\widehat{\kappa}_n \geq 0$ and

$$\mathbb{E}_P[\widehat{\kappa}_n] = \kappa_n,$$

Jensen's inequality gives

$$\mathbb{E}_P\sqrt{\frac{\widehat{\kappa}_n}{n}} \le \sqrt{\frac{\kappa_n}{n}}.$$

Consequently

$$\mathbb{E}_P[\lambda(\mathscr{C}_n^{\mathrm{sc}})] \le \frac{4L_\alpha}{\sqrt{t_n}} + \frac{8}{\varepsilon^2 t_n}.$$

Together with $\lambda(\mathscr{C}_n^{\mathrm{sc}}) \le 2$, this implies

$$\mathbb{E}_P[\lambda(\mathscr{C}_n^{\mathrm{sc}})] \le \max\{2, 4L_\alpha + 8\varepsilon^{-2}\} \min\{1, t_n^{-1/2}\}.$$

The same inequality holds at $n = 0$ by the definition $\mathscr{C}_0^{\mathrm{sc}} = \varnothing$.

7. It remains to verify honesty. Define

$$\delta_n = \begin{cases} 1, & n = 0, \\ 16k_n^2/n, & n \ge 1. \end{cases}$$

Since $k_n/\sqrt{n} \to 0$, one has $\delta_n \to 0$. Fix $n \ge 1$ and $P \in \mathcal{P}_n(\mathfrak{g})$, and put

$$\widehat{S}_n = \widehat{A}_n - \theta_T(P)\widehat{B}_n.$$

Applying the weighted pullout identity from the preceding step first with

$$F(O^S) = \left\{\frac{\mathbf{1}\{Z = 1\}}{e(X)} - \frac{\mathbf{1}\{Z = 0\}}{1 - e(X)}\right\} Y$$

and $d = \Delta_Y$, and then with

$$F(O^S) = \left\{\frac{\mathbf{1}\{Z = 1\}}{e(X)} - \frac{\mathbf{1}\{Z = 0\}}{1 - e(X)}\right\} D$$

and $d = \Delta_D$, the set-integral identities in Lemma 7 and the weighted integral identities in Lemma 9 give

$$\mathbb{E}_P[\widehat{S}_n] = \mu_{Y,n} - \theta_T(P)\mu_n = 0.$$

Moreover, since $Y \in [0, 1]$, $D \in \{0, 1\}$, $\theta_T(P) \in [-1, 1]$, and the overlap bound gives $\varepsilon \le e(X) \le 1 - \varepsilon$,

$$\left|\left\{\frac{\mathbf{1}\{Z = 1\}}{e(X)} - \frac{\mathbf{1}\{Z = 0\}}{1 - e(X)}\right\}\big(Y - \theta_T(P)D\big)\right| \le \frac{2}{\varepsilon}.$$

Thus

$$\mathrm{Var}_P(\widehat{S}_n) \le \frac{4\kappa_n}{\varepsilon^2 n}.$$

Also,

$$\mathbb{E}_P[\widehat{\kappa}_n] = \kappa_n, \qquad \mathrm{Var}_P(\widehat{\kappa}_n) \le \frac{4k_n^2\kappa_n}{n},$$

because $0 \le w_n \le 2k_n$ implies

$$w_n^4 \le 4k_n^2 w_n^2.$$

Chebyshev's inequality gives

$$P\{\widehat{\kappa}_n < \kappa_n/2\} \le \frac{16k_n^2}{n\kappa_n} \le \frac{16k_n^2}{n}.$$

With

$$a_n = L_\alpha\sqrt{\frac{\kappa_n}{2n}},$$

another application of Chebyshev gives

$$P\{|\widehat{S}_n| > a_n\} \le \frac{4\kappa_n/(\varepsilon^2 n)}{L_\alpha^2\kappa_n/(2n)} = \alpha.$$

Let

$$G_n = \left\{\widehat{\kappa}_n \ge \frac{\kappa_n}{2}\right\} \cap \left\{|\widehat{S}_n| \le a_n\right\}.$$

The statistics defining the score-inversion rule and the two bad events defining G_n^c depend only on the source sample. Under the joint two-sample law Q_P^N, the source and target samples have product law, so $Q_P^N(G_n) = P(G_n)$ for these source-sample events. On G_n, for every target covariate sample,

$$\widehat{\kappa}_n \ge \frac{\kappa_n}{2} \quad \text{and} \quad |\widehat{S}_n| \le L_\alpha\sqrt{\frac{\kappa_n}{2n}} \le L_\alpha\sqrt{\frac{\widehat{\kappa}_n}{n}},$$

so $\theta_T(P) \in \mathscr{C}_n^{\mathrm{sc}}$. Hence

$$Q_P^N\{\theta_T(P) \in \mathscr{C}_n^{\mathrm{sc}}\} \ge 1 - \alpha - \delta_n.$$

At $n = 0$ the same lower bound is immediate from $\delta_0 = 1$. Since $\delta_n \to 0$ and the fixed-geometry slice is eventually nonempty, $\mathscr{C}^{\mathrm{sc}}$ is fixed-geometry oracle honest.

8. Let

$$C_\alpha = \max\{2, 4L_\alpha + 8\varepsilon^{-2}\}.$$

The pointwise bound from the preceding steps implies, for every $P \in \mathcal{P}_n(\mathfrak{g})$,

$$\mathbb{E}_P[\lambda(\mathscr{C}_n^{\mathrm{sc}})] \le C_\alpha \min\{1, t_n(P)^{-1/2}\}.$$

Taking the supremum over the row $t_n(P) \ge t_0$, then the limsup over n, gives

$$R_{\mathfrak{g}}(\mathscr{C}^{\mathrm{sc}}, t_0) \le C_\alpha \min\{1, t_0^{-1/2}\}.$$

Since $\mathscr{C}^{\mathrm{sc}}$ is fixed-geometry oracle honest,

$$V_{\mathfrak{g}}^\star(t_0) \le C_\alpha \min\{1, t_0^{-1/2}\}.$$

Combining this with the lower bound proves the two displayed inequalities. The constants c_α and C_α involve only α and ε, so the bounds are uniform over admissible deterministic geometry arrays $\mathfrak{g}$.

$\square$

Proof of Theorem 7. 1. Define

$$B_{\varepsilon,c} = 8(\varepsilon^{-2} + c^{-1}), \qquad C_0 = \max\{2, 4\sqrt{2}L + 8B_{\varepsilon,c}\}.$$

The hypotheses give $B_{\varepsilon,c} \geq 0$, $L \geq 0$, and hence $C_0 \geq 2 > 0$. By Lemma 4, the regular class is inhabited at every row n. Let $C^\star$ be the regular-cell procedure which, on a regular-cell input consisting of a two-sample draw, the selected cell design, the displayed source cell masses, and the displayed propensity values, returns the score-inversion set

$$\{\vartheta \in \Theta : |\widehat{M}_{Y,n} - \vartheta \widehat{M}_{D,n}| \leq L\sqrt{\widehat{K}_n/n}\}$$

when $N_n \geq 2$, and returns Θ when $N_n < 2$. The measurable graph and Θ-valuedness are part of this construction. For every $P \in \mathcal{N}_n^{\mathrm{reg}}$, the input obtained from the two-sample law uses the same source cell masses $q_{x,n}$ and propensity e, so $C_n^\star$ agrees almost surely with the displayed C_n^{reg}, with the stated convention for $N_n < 2$.

2. We verify regular-cell honesty for $C^\star$. Since the regular class is inhabited at every n, the row coverage convention in $\mathfrak{C}^{\mathrm{reg}}$ is the ordinary infimum over $P \in \mathcal{N}_n^{\mathrm{reg}}$. Put

$$\delta_n = \sup_{P \in \mathcal{N}_n^{\mathrm{reg}}} P\left(\widehat{K}_n < \frac{\kappa_n(P)}{2}\right),$$

where the probability is over the target covariate sample. The sample-size ratio gives $n > 0$ and $N_n \geq 2$ eventually. For those rows, write $\widehat{K}_n = 1 + \widehat{\kappa}_{U,n}$. The collision calculation gives

$$\mathbb{E}_P[\widehat{\kappa}_{U,n}] = \kappa_n(P), \qquad \mathrm{Var}_P(\widehat{\kappa}_{U,n}) \leq \frac{32k_n^2}{c_-^2 N_n}.$$

Also $\kappa_n(P) \geq 1$. If $\kappa_n(P) \leq 2$, then $\widehat{K}_n \geq 1 \geq \kappa_n(P)/2$ pointwise. If $\kappa_n(P) > 2$, Chebyshev's inequality gives

$$P\left(\widehat{K}_n < \frac{\kappa_n(P)}{2}\right) \leq \frac{128k_n^2}{c_-^2 N_n \kappa_n(P)^2} \leq \frac{128k_n^2}{c_-^2 N_n}.$$

Because $N_n/n \to c > 0$ and $k_n/\sqrt{n} \to 0$, the last bound tends to zero uniformly on the regular class; hence $\delta_n \to 0$.

Fix such an n and $P \in \mathcal{N}_n^{\mathrm{reg}}$, and set $\theta = \theta_T(P)$. Let

$$\mathsf{score}_{n,P}(s) = \widehat{M}_{Y,n}(s) - \theta\, \widehat{M}_{D,n}(s), \qquad \mathsf{smallK}_{n,P} = \left\{s : \widehat{K}_n(s) < \frac{\kappa_n(P)}{2}\right\},$$

and

$$\mathsf{largeScore}_{n,P} = \left\{s : L\sqrt{\frac{\kappa_n(P)}{2n}} < |\mathsf{score}_{n,P}(s)|\right\}.$$

The source sample lies in the witnessing finite cell support with probability one, and target domination transfers that support statement to the target sample. On the event where this support condition holds and neither $\mathsf{smallK}_{n,P}$ nor $\mathsf{largeScore}_{n,P}$ occurs, the inequality

$$|\mathsf{score}_{n,P}(s)| \leq L\sqrt{\frac{\kappa_n(P)}{2n}} \leq L\sqrt{\frac{\widehat{K}_n(s)}{n}}$$

places $\theta_T(P)$ in $C_n^\star(s)$.

It remains to bound the large-score event. The finite witnessing support gives the needed integrability of the assignment and receipt contrasts. The transported causal-range conclusions in Lemma 9 put $\theta \in \Theta$ and identify θ with the transported ratio $\mu_{Y,n}(P)/\mu_n(P)$. On the witnessing cells ξ_i, the regular-cell score-mean identity gives

$$\mathbb{E}_P\{H(O^S)(Y - \theta D) \mid X = \xi_i\} = \Delta_Y(\xi_i) - \theta\Delta_D(\xi_i),$$

and the transported ratio identity yields the centered finite-cell score equation

$$\sum_i P_T\{\xi_i\}\{\Delta_Y(\xi_i) - \theta_T(P)\Delta_D(\xi_i)\} = 0.$$

By Lemma 11, conditionally on the target sample the score has mean $\sum_x \widehat{p}_{x,n}\{\Delta_Y(x) - \theta\Delta_D(x)\}$ and conditional variance at most

$$\frac{4}{\varepsilon^2 n}\sum_x \frac{\widehat{p}_{x,n}^2}{q_{x,n}}.$$

The target-sampling and multinomial calculations give

$$\operatorname{Var}_P\left(\sum_x \widehat{p}_{x,n}\{\Delta_Y(x) - \theta\Delta_D(x)\}\right) \le \frac{4}{N_n}, \qquad \mathbb{E}_P\left[\sum_x \frac{\widehat{p}_{x,n}^2}{q_{x,n}}\right] \le \kappa_n(P) + \frac{k_n}{c_- N_n}.$$

Combining these bounds gives, uniformly for all sufficiently large n,

$$\mathbb{E}_P[\mathsf{score}_{n,P}] = 0, \qquad \operatorname{Var}_P(\mathsf{score}_{n,P}) \le B_{\varepsilon,c}\frac{\kappa_n(P)}{n}.$$

Here the last inequality uses $k_n/(c_- N_n) \le \kappa_n(P)$, $4/N_n \le 8/(cn)$, and $\kappa_n(P) \ge 1$ eventually. Chebyshev's inequality and $L^2 \ge 2B_{\varepsilon,c}/\alpha$ yield

$$P(\mathsf{largeScore}_{n,P}) \le \frac{B_{\varepsilon,c}\kappa_n(P)/n}{L^2\kappa_n(P)/(2n)} = \frac{2B_{\varepsilon,c}}{L^2} \le \alpha.$$

Thus

$$P\{\theta_T(P) \in C_n^\star\} \ge 1 - \alpha - \delta_n.$$

Taking the infimum over $P \in \mathcal{N}_n^{\text{reg}}$, then the limit inferior, gives

$$\liminf_{n\to\infty}\inf_{P\in\mathcal{N}_n^{\text{reg}}} P\{\theta_T(P) \in C_n^\star\} \ge 1 - \alpha,$$

so $C^\star \in \mathfrak{C}^{\text{reg}}$.

3. Fix $t_0 > 0$. The sample-size ratio again gives $N_n \ge 2$ eventually. For $P \in \mathcal{N}_n^{\text{reg}}$, write

$$\mu_n(P) = \text{the transported first-stage mean}, \qquad \kappa_n(P) = \text{the transport-weight second moment}, \qquad t_n(P) =$$

By Lemma 9, $\mu_n(P) > 0$, and the regular-cell mass calculation gives $\kappa_n(P) \ge 1$. The receipt score has mean $\mu_n(P)$, and the uniform moment calculation used above gives

$$\operatorname{Var}_P(\widehat{M}_{D,n}) \le B_{\varepsilon,c}\frac{\kappa_n(P)}{n}.$$

Therefore

$$P\left(\frac{\mu_n(P)}{2} < |\widehat{M}_{D,n} - \mu_n(P)|\right) \le \frac{4B_{\varepsilon,c}}{t_n(P)}.$$

Moreover the feasible dispersion proxy satisfies

$$\mathbb{E}_P[\widehat{K}_n] = 1 + \kappa_n(P) \le 2\kappa_n(P).$$

For a realized sample let $r_n = L\sqrt{\widehat{K}_n/n}$. On $\{|\widehat{M}_{D,n} - \mu_n(P)| \le \mu_n(P)/2\}$, affine inversion of $|\widehat{M}_{Y,n} - \vartheta\widehat{M}_{D,n}| \le r_n$ over Θ has length at most $4r_n/\mu_n(P)$; on the complementary event its length is at most 2. Jensen's inequality gives

$$\mathbb{E}_P[\lambda(C_n^\star)] \le \frac{4L}{\mu_n(P)}\sqrt{\frac{\mathbb{E}_P[\widehat{K}_n]}{n}} + 2\,P\left(\frac{\mu_n(P)}{2} < |\widehat{M}_{D,n} - \mu_n(P)|\right),$$

and hence

$$\mathbb{E}_P[\lambda(C_n^\star)] \le \frac{4\sqrt{2}L}{\sqrt{t_n(P)}} + \frac{8B_{\varepsilon,c}}{t_n(P)}.$$

This last display supplies the inverse-root bound when $t_n(P) \ge 1$: since $t_n(P)^{-1} \le t_n(P)^{-1/2}$,

$$\mathbb{E}_P[\lambda(C_n^\star)] \le (4\sqrt{2}L + 8B_{\varepsilon,c})t_n(P)^{-1/2} \le C_0 \min\{1, t_n(P)^{-1/2}\}.$$

When $0 < t_n(P) < 1$, the diameter bound $\lambda(C_n^\star) \le 2$ gives instead

$$\mathbb{E}_P[\lambda(C_n^\star)] \le 2 \le C_0 = C_0 \min\{1, t_n(P)^{-1/2}\}.$$

Thus for all sufficiently large rows and every regular-cell law in the row,

$$\mathbb{E}_P[\lambda(C_n^\star)] \le C_0 \min\{1, t_n(P)^{-1/2}\}.$$

Taking the supremum over $P \in \mathcal{N}_n^{\mathrm{reg}}$ with $t_n(P) \ge t_0$, then the limit superior, yields

$$\limsup_{n\to\infty} \sup_{P \in \mathcal{N}_n^{\mathrm{reg}}:\, t_n(P) \ge t_0} \mathbb{E}_P[\lambda(C_n^\star)] \le C_0 \min\{1, t_0^{-1/2}\}.$$

The finitely many rows before $N_n \ge 2$ are absorbed by the limit superior.

4. It remains to prove the lower bound. Choose the injected cells $\xi_{n,i}$, $i \in \{1, \ldots, k_n\}$, from the carrier hypothesis and define the deterministic geometry

$$\nu_n = \frac{1}{k_n}\sum_{i=1}^{k_n} \delta_{\xi_{n,i}}, \qquad P_{S,n}^X = \nu_n, \qquad P_{T,n} = \nu_n, \qquad w_n \equiv 1, \qquad e_n \equiv \frac{1}{2}.$$

This geometry is admissible: $0 < \varepsilon < 1/2$ gives overlap for e_n, the transport identity is $\int_A 1\, d\nu_n = \nu_n(A)$, the envelope bound is $0 \le 1 \le 2k_n$, and the weight-second-moment bound is

$$\int w_n^2\, dP_{S,n}^X = 1 \le k_n.$$

Its source cell masses are

$$\nu_n\{\xi_{n,i}\} = \frac{1}{k_n},$$

so

$$\frac{c_-}{k_n} \le \frac{1}{k_n} \le \frac{c_+}{k_n}$$

by $c_- \le 1 \le c_+$. Hence every law in this fixed-geometry slice belongs to $\mathcal{N}_n^{\mathrm{reg}}$.

Let $C \in \mathfrak{C}^{\mathrm{reg}}$. Build a fixed-geometry oracle procedure $\mathcal{O}$ by feeding C_n the regular-cell design $(\xi_{n,i})_{i=1}^{k_n}$, the masses $1/k_n$, and the propensity values $1/2$ whenever the realized source and target samples lie on the selected cells; set $\mathcal{O}_n = \Theta$ outside that support event. Under every law in the fixed-geometry slice, the support event has probability one, so

$$\mathcal{O}_n(s) = C_n(s) \quad \text{for two-sample almost every } s$$

after identifying the fixed-geometry input with the corresponding regular-cell input. Consequently, for every such P,

$$P\{\theta_T(P) \in \mathcal{O}_n\} = P\{\theta_T(P) \in C_n\}, \qquad \mathbb{E}_P[\lambda(\mathcal{O}_n)] = \mathbb{E}_P[\lambda(C_n)].$$

Since the fixed-geometry slice is contained in the regular class, regular-cell honesty of C, together with the preceding coverage identity and Lemma 3, makes $\mathcal{O}$ a fixed-geometry honest oracle procedure. The rowwise expected lengths are bounded by 2 because all sets lie in $\Theta = [-1, 1]$. Therefore the fixed-geometry risk of $\mathcal{O}$, written as the limiting supremum from Definition 11, is bounded by the corresponding regular-cell risk:

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n(\mathfrak{g}):\, t_n(P)\ge t_0} \mathbb{E}_P[\lambda(\mathcal{O}_n)] \le \limsup_{n\to\infty} \sup_{P\in\mathcal{N}_n^{\mathrm{reg}}:\, t_n(P)\ge t_0} \mathbb{E}_P[\lambda(C_n)].$$

Because $V_{\mathfrak{g}}^\star(t_0)$ is the infimum of this fixed-geometry risk over fixed-geometry honest oracle procedures in Definition 11,

$$V_{\mathfrak{g}}^\star(t_0) \le \limsup_{n\to\infty} \sup_{P\in\mathcal{N}_n^{\mathrm{reg}}:\, t_n(P)\ge t_0} \mathbb{E}_P[\lambda(C_n)].$$

Taking the infimum over regular-cell honest C and applying the fixed-geometry lower bound from Theorem 5 gives

$$\frac{3(1-\alpha)^2}{16} \min\{1, t_0^{-1/2}\} \le \inf_{C\in\mathfrak{C}^{\mathrm{reg}}} \limsup_{n\to\infty} \sup_{P\in\mathcal{N}_n^{\mathrm{reg}}:\, t_n(P)\ge t_0} \mathbb{E}_P[\lambda(C_n)].$$

Combining this lower bound with the construction and upper bound above proves the asserted regular-cell attainment statement.

□

References

T. W. Anderson and Herman Rubin. Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, 20(1):46–63, 1949. doi: 10.1214/aoms/1177730090. URL https://doi.org/10.1214/aoms/1177730090.

Donald W. K. Andrews, Marcelo J. Moreira, and James H. Stock. Optimal two-sided invariant similar tests for instrumental variables regression. *Econometrica*, 74(3):715–752, 2006. doi: 10.1111/j.1468-0262.2006.00680.x. URL https://doi.org/10.1111/j.1468-0262.2006.00680.x.

Isaiah Andrews, James H. Stock, and Liyang Sun. Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics*, 11:727–753, 2019. doi: 10.1146/annurev-economics-080218-025643. URL https://doi.org/10.1146/annurev-economics-080218-025643.

Joshua D. Angrist, Guido W. Imbens, and Donald B. Rubin. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455, 1996. doi: 10.1080/01621459.1996.10476902. URL https://doi.org/10.1080/01621459.1996.10476902.

Timothy B. Armstrong and Michal Kolesár. Finite-sample optimal estimation and inference on average treatment effects under unconfoundedness. *Econometrica*, 89(3):1141–1177, 2021. doi: 10.3982/ECTA16907. URL https://doi.org/10.3982/ECTA16907.

Alexander Balke and Judea Pearl. Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92(439):1171–1176, 1997. doi: 10.1080/01621459.1997.10474074. URL https://doi.org/10.1080/01621459.1997.10474074.

Elias Bareinboim and Judea Pearl. A general algorithm for deciding transportability of experimental results. *Journal of Causal Inference*, 1(1):107–134, 2013. doi: 10.1515/jci-2012-0004. URL https://doi.org/10.1515/jci-2012-0004.

Elias Bareinboim and Judea Pearl. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113(27):7345–7352, 2016. doi: 10.1073/pnas.1510507113. URL https://doi.org/10.1073/pnas.1510507113.

Marinho Bertanha and Marcelo J. Moreira. Impossible inference in econometrics: Theory and applications. *Journal of Econometrics*, 218(2):247–270, 2020. doi: 10.1016/j.jeconom.2020.04.016. URL https://doi.org/10.1016/j.jeconom.2020.04.016.

Howard S. Bloom. Accounting for no-shows in experimental evaluation designs. *Evaluation Review*, 8(2):225–246, 1984. doi: 10.1177/0193841X8400800205. URL https://doi.org/10.1177/0193841X8400800205.

John Bound, David A. Jaeger, and Regina M. Baker. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, 90(430):443–450, 1995. doi: 10.1080/01621459.1995.10476536. URL https://doi.org/10.1080/01621459.1995.10476536.

Zhongren Chen and Melody Huang. Generalizing causal effects with noncompliance: Application to deep canvassing experiments, 2025. URL https://arxiv.org/abs/2506.00149.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018. doi: 10.1111/ectj.12097. URL https://doi.org/10.1111/ectj.12097.

Jaerim Choi and Shu Shen. Two-sample instrumental-variables regression with potentially weak instruments. *The Stata Journal*, 19(3):581–597, 2019. doi: 10.1177/1536867X19874235. URL https://doi.org/10.1177/1536867X19874235.

Jaerim Choi, Jiaying Gu, and Shu Shen. Weak-instrument robust inference for two-sample instrumental variables regression. *Journal of Applied Econometrics*, 33(1):109–125, 2018. doi: 10.1002/jae.2580. URL https://doi.org/10.1002/jae.2580.

Stephen R. Cole and Elizabeth A. Stuart. Generalizing evidence from randomized clinical trials to target populations: The ACTG 320 trial. *American Journal of Epidemiology*, 172(1): 107–115, 2010. doi: 10.1093/aje/kwq084. URL https://doi.org/10.1093/aje/kwq084.

Issa J. Dahabreh, Lucia C. Petito, Sarah E. Robertson, Miguel A. Hernán, and Jon A. Steingrimsson. Toward causally interpretable meta-analysis. *Epidemiology*, 31(3):334–344, 2020a. doi: 10.1097/EDE.0000000000001177. URL https://doi.org/10.1097/EDE.0000000000001177.

Issa J. Dahabreh, Sarah E. Robertson, Jon A. Steingrimsson, Elizabeth A. Stuart, and Miguel A. Hernán. Extending inferences from a randomized trial to a new target population. *Statistics in Medicine*, 39(14):1999–2014, 2020b. doi: 10.1002/sim.8426. URL https://doi.org/10.1002/sim.8426.

Jean-Marie Dufour. Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica*, 65(6):1365–1387, 1997. doi: 10.2307/2171740. URL https://doi.org/10.2307/2171740.

E. C. Fieller. Some problems in interval estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 16(2):175–185, 1954. doi: 10.1111/j.2517-6161.1954.tb00159.x. URL https://doi.org/10.1111/j.2517-6161.1954.tb00159.x.

Constantine E. Frangakis and Donald B. Rubin. Principal stratification in causal inference. *Biometrics*, 58(1):21–29, 2002. doi: 10.1111/j.0006-341X.2002.00021.x. URL https://doi.org/10.1111/j.0006-341X.2002.00021.x.

Leon Jay Gleser and Jiunn T. Hwang. The nonexistence of $100(1-\alpha)$expected diameter in errors-in-variables and related models. *The Annals of Statistics*, 15(4):1351–1362, 1987. doi: 10.1214/aos/1176350597. URL https://doi.org/10.1214/aos/1176350597.

James J. Heckman and Edward Vytlacil. Structural equations, treatment effects, and econometric policy evaluation1. *Econometrica*, 73(3):669–738, 2005. doi: 10.1111/j.1468-0262.2005.00594.x. URL https://doi.org/10.1111/j.1468-0262.2005.00594.x.

Keisuke Hirano, Guido W. Imbens, and Geert Ridder. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 71(4):1161–1189, 2003. doi: 10.1111/1468-0262.00442. URL https://doi.org/10.1111/1468-0262.00442.

Guido W. Imbens and Joshua D. Angrist. Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475, 1994. doi: 10.2307/2951620. URL https://doi.org/10.2307/2951620.

Guido W. Imbens and Donald B. Rubin. Estimating outcome distributions for compliers in instrumental variables models. *The Review of Economic Studies*, 64(4):555–574, 1997. doi: 10.2307/2971731. URL https://doi.org/10.2307/2971731.

V. Joseph Hotz, Guido W. Imbens, and Julie H. Mortimer. Predicting the efficacy of future training programs using past experiences at other locations. *Journal of Econometrics*, 125 (1–2):241–270, 2005. doi: 10.1016/j.jeconom.2004.04.009. URL https://doi.org/10.1016/j.jeconom.2004.04.009.

Yukun Ma. Identification-robust inference for the LATE with high-dimensional covariates, 2023. URL https://arxiv.org/abs/2302.09756.

Anna Mikusheva. Robust confidence sets in the presence of weak instruments. *Journal of Econometrics*, 157(2):236–247, 2010. doi: 10.1016/j.jeconom.2009.12.003. URL https://doi.org/10.1016/j.jeconom.2009.12.003.

Marcelo J. Moreira. A conditional likelihood ratio test for structural models. *Econometrica*, 71(4):1027–1048, 2003. doi: 10.1111/1468-0262.00438. URL https://doi.org/10.1111/1468-0262.00438.

Trang Quynh Nguyen, Cyrus Ebnesajjad, Stephen R. Cole, and Elizabeth A. Stuart. Sensitivity analysis for an unobserved moderator in RCT-to-target-population generalization of treatment effects. *The Annals of Applied Statistics*, 11(1):225–247, 2017. doi: 10.1214/16-AOAS1001. URL https://doi.org/10.1214/16-AOAS1001.

Judea Pearl and Elias Bareinboim. Transportability of causal and statistical relations: A formal approach. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, pages 247–254, 2011. URL https://dl.acm.org/doi/10.5555/2900423.2900462.

Muyang Ren. Extrapolating LATE with weak IVs, 2025. URL https://arxiv.org/abs/2512.23854.

James M. Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866, 1994. doi: 10.1080/01621459.1994.10476818. URL https://doi.org/10.1080/01621459.1994.10476818.

Rachael K. Ross, Iván Díaz, Amy J. Pitts, Elizabeth A. Stuart, and Kara E. Rudolph. Transporting results from a trial to an external target population when trial participation impacts adherence. *Epidemiology*, 37(1):39–49, 2026. doi: 10.1097/EDE.0000000000001925. URL https://doi.org/10.1097/EDE.0000000000001925.

Kara E. Rudolph and Mark J. Laan. Robust estimation of encouragement design intervention effects transported across sites. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(5):1509–1525, 2016. doi: 10.1111/rssb.12213. URL https://doi.org/10.1111/rssb.12213.

Kara E. Rudolph, Nicholas T. Williams, Elizabeth A. Stuart, and Iván Díaz. Improving efficiency in transporting average treatment effects. *Biometrika*, 112(3):asaf027, 2025. doi: 10.1093/biomet/asaf027. URL https://doi.org/10.1093/biomet/asaf027.

Ezequiel Smucler, Ludovico Lanni, and David Masip. A note on the properties of the confidence set for the local average treatment effect obtained by inverting the score test, 2025. URL https://arxiv.org/abs/2506.10449.

Douglas Staiger and James H. Stock. Instrumental variables regression with weak instruments. *Econometrica*, 65(3):557–586, 1997. doi: 10.2307/2171753. URL https://doi.org/10.2307/2171753.

James H. Stock, Jonathan H. Wright, and Motohiro Yogo. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, 20(4):518–529, 2002. doi: 10.1198/073500102288618658. URL https://doi.org/10.1198/073500102288618658.

Elizabeth A. Stuart, Stephen R. Cole, Catherine P. Bradshaw, and Philip J. Leaf. The use of propensity scores to assess the generalizability of results from randomized trials. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(2):369–386, 2011. doi: 10.1111/j.1467-985X.2010.00673.x. URL https://doi.org/10.1111/j.1467-985X.2010.00673.x.

Elizabeth Tipton. Improving generalizations from experiments using propensity score subclassification. *Journal of Educational and Behavioral Statistics*, 38(3):239–266, 2013. doi: 10.3102/1076998612441947. URL https://doi.org/10.3102/1076998612441947.

Mark J. van der Laan and Daniel Rubin. Targeted maximum likelihood learning. *The International Journal of Biostatistics*, 2(1):Article 11, 2006. doi: 10.2202/1557-4679.1043. URL https://doi.org/10.2202/1557-4679.1043.

Zhenghao Zeng, Edward H. Kennedy, Lisa M. Bodnar, and Ashley I. Naimi. Efficient generalization and transportation, 2024. URL https://arxiv.org/abs/2302.00092.