Contour instruments for partially linear models with cumulant-separated treatment noise

CausalSmith*

August 25, 2026

Abstract

This paper studies estimation of the partially linear coefficient θ_0, the treatment coefficient, when the treatment innovation is independent of covariates, sub-Gaussian, and separated from Gaussianity by a fixed nonzero cumulant. The identifying resource is a zero of the treatment-innovation moment-generating function: the associated polynomial-exponential weight annihilates real shifts induced by treatment-code error, and a contour average of observable residual transforms identifies θ_0 under the stated boundary zero-freeness, nuisance zero-freeness, and positive-count conditions.

The statistical theorem is a fixed-code result. For fixed primitive constants, i.i.d. product sampling, the partially linear conditional-mean restrictions, bounded coefficient and regression ranges, sub-Gaussian treatment and outcome-noise envelopes, a nonempty fixed-code class, and the displayed current $L^1(P_X)$ treatment-code radius gate, where P_X is the covariate marginal, a finite translated-dyadic contour bank and a total Borel statistic attain matched c/n and C/n mean-squared-error bounds on the non-Gaussian spectral class. For $\gamma \in (1/2, 1)$, where γ is the probability level, the same statistic controls the generalized lower $(1-\gamma)$-quantile of absolute error by order $(\gamma n)^{-1/2}$.

The same fixed-separation rate holds on the aligned Jin–Mackey–Syrgkanis ACE comparison class. The ACE comparison records the published finite-order ACE upper guarantee and compares it with the contour upper guarantee on the common clipped-code class. The paper also gives a bounded-outcome Gaussian diagnostic and explicit sine-ratio reductions for mixture benchmarks.

The represented-data construction is a conditional reproducibility layer for the same ordinary Borel statistic. When a compiled bounded spectral adapter satisfies the full canonical build-and-compilation specification for the parameter record, transported fixed records, and base treatment-code sequence, its represented-data execution realizes the statistic used in the statistical risk statements.

1 Introduction

Partially linear models are a standard way to separate a low-dimensional structural coefficient from flexible covariate regressions. In the model studied here, the observed data are triples

*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.

$O = (X, T, Y)$, the treatment regression is $g_0(X) = E[T \mid X]$, the outcome regression is $q_0(X) = E[Y \mid X]$, and the target is the partially linear coefficient θ_0 in the conditional mean

$$E[Y \mid X, T] = q_0(X) + \theta_0\{T - g_0(X)\}.$$

Classical semiparametric theory and modern double/debiased machine learning estimate this coefficient by orthogonalizing treatment and outcome variation around nuisance regressions [Robinson, 1988, Newey, 1990, 1994, Newey and McFadden, 1994, Bickel et al., 1993, van der Vaart, 1998, van der Vaart and Wellner, 1996, Belloni et al., 2014, Chernozhukov et al., 2018]. Higher-order orthogonality and ACE methods refine this idea by exploiting additional structure in nuisance errors and treatment residuals [Robins et al., 2008, Liu et al., 2017, Newey and Robins, 2018, Liu et al., 2016, Mackey et al., 2018, Jin et al., 2025]. This paper develops a complementary fixed-separation construction based on non-Gaussian treatment innovations.

The statistical decision problem fixes the supplied clipped treatment and outcome codes at the current sample size. The risk bounds compare laws compatible with that same code pair; if the codes are obtained from an external training stage, the theorem applies after conditioning on the training output. The finite-arithmetic material supplies a conditional reproducibility clause for represented-data executions that satisfy the displayed build contract.

Four statements organize the paper. First, the contour identification theorem recovers θ_0 from an observable population contour ratio under the displayed zero-free and positive-count conditions. Second, under the theorem's fixed-code nonemptiness, sampling, and radius conditions, the fixed-code minimax theorem gives matched c/n and C/n mean-squared-error bounds on the broad non-Gaussian spectral class and the aligned Jin–Mackey–Syrgkanis ACE class. Under the same fixed-code hypotheses, the same contour statistic also has a generalized lower $(1-\gamma)$-quantile upper bound of order $(\gamma n)^{-1/2}$. Third, the ACE alignment proposition compares the published finite-order ACE upper guarantee with the contour upper guarantee on the common clipped-code class. Fourth, the bounded-outcome Gaussian diagnostic characterizes the comparison class by its implied target and fixed-code risks. Here P_X denotes the covariate marginal, r the ACE order, δ the cumulant-separation level, γ the probability level, and σ^2 the Gaussian innovation variance.

The identifying resource is a zero of the treatment-innovation moment-generating function. Let $\eta = T - g_0(X)$ denote the treatment innovation and let $M(z) = E[e^{z\eta}]$. When M has a zero z_0 of multiplicity ℓ, the polynomial-exponential weight $J_{z_0,\ell}(w) = w^{\ell-1}e^{z_0 w}$ annihilates every shifted innovation $\eta + d$. This shift invariance is useful for code residuals: if $Z_n = T - \bar{g}_n(X) = \eta + \{g_0(X) - \bar{g}_n(X)\}$, where $\bar{g}_n$ is the clipped supplied treatment code, then the same zero generates conditional moments for Z_n. Theorem 1 gives the resulting known-zero ratio, and Lemma 1 and Theorem 2 turn the ratio into an observable contour identity.

The contour identity is the paper's population mechanism. The residual transform $F_n(z) = E[e^{zZ_n}]$ factors as $M(z)H_n(z)$, where $H_n(z) = E[e^{z\{g_0(X)-\bar{g}_n(X)\}}]$ is the nuisance transform. The outcome-weighted transform $G_n(z) = E[Ye^{zZ_n}]$ decomposes into a derivative term carrying θ_0 and an analytic contamination term. On any circle where F_n is zero-free on the boundary, H_n is zero-free on the closed disk, and the enclosed zero count is positive, the normalized contour integral of G_n/F_n equals θ_0. This places the argument in the transform-based identification tradition [Feuerverger and McDunnough, 1981, Singleton, 2001, Chacko and Viceira, 2003, Carrasco et al., 2007, Carrasco, 2017, Kagan et al., 1973, Mattner, 1992, D'Haultfoeuille, 2011, Darolles et al., 2011, Hu and Shiu, 2022].

Estimator in words. The statistic splits the sample deterministically. The pilot fold scans a fixed finite bank of circles, keeps circles whose empirical residual transform has a positive boundary modulus and a positive decoded winding count, and chooses an admissible circle by the displayed lower-modulus rule with an index tie-break. The evaluation fold computes the normalized contour ratio of the outcome-weighted transform to the residual transform on that selected circle. Clipping to the fixed coefficient range makes the statistic total on every ordinary sample.

The main statistical result is fixed-separation and fixed-code. Under the non-Gaussian spectral class in Definition 3, the treatment innovation is independent of covariates, sub-Gaussian, and has a fixed nonzero $(r+1)$st cumulant bounded below by δ. Lemma 3 converts this cumulant separation and the treatment-tail scale into an explicit disk containing a transform zero. Lemma 8 constructs a finite translated-dyadic bank of circles with a positive dyadic boundary-modulus certificate. The $L^1(P_X)$ treatment-code condition in Assumption 16 keeps H_n separated from zero on the search disk, as shown in Lemma 10. Empirical transform concentration in Lemma 9 then supplies the stochastic control needed for the selected contour ratio.

The resulting estimator is the total Borel statistic $\widehat{\theta}_{\mathrm{spec},n}$ in Algorithm 1. Theorem 3 states that, for fixed primitive constants and sufficiently small current $L^1(P_X)$ treatment-code radius, this statistic attains

$$\sup_{P \in \mathcal{P}_{\mathrm{NG},n}(p;\mathrm{base})} E_P\{(\widehat{\theta}_{\mathrm{spec},n} - \theta_0(P))^2\} \le C/n,$$

while the fixed-code non-Gaussian minimax risk is bounded below by c/n. Thus the fixed-code mean-squared-error rate is $\Theta(n^{-1})$. For $\gamma \in (1/2, 1)$, the same theorem controls the generalized lower $(1-\gamma)$-quantile of absolute error by order $(\gamma n)^{-1/2}$, and it gives the parallel fixed-code result on the Jin–Mackey–Syrgkanis ACE comparison class. Theorem 4 records the corresponding sequence-level statement for experiments sharing the same primitive constants.

The comparison with ACE is stated as a comparison of upper guarantees on a common cumulant-separated domain. Lemma 5 aligns the published ACE class with the broad spectral class and its $L^s(P_X)$-restricted subclass. Proposition 1 then records the published order-r ACE generalized-quantile guarantee from Jin et al. [2025] and the spectral guarantee on the same clipped-code ACE class. Along the displayed nuisance-dominance sequences, the ratio of the contour upper guarantee to the imported finite-order ACE upper guarantee tends to zero. This is an upper-bound comparison on the common clipped-code class. The minimax lower bound used for the ACE class is class-level over all estimators on that fixed-code class, as in Theorem 3.

The Gaussian comparison characterizes the bounded-outcome Gaussian intersection in Definition 4. That class combines a nondegenerate Gaussian treatment innovation with bounded observed outcomes and the exact partially linear conditional mean. Proposition 2 shows that these restrictions force $\theta_0(P) = 0$, and consequently the corresponding fixed-code Gaussian mean-squared and generalized-quantile minimax risks are zero whenever the class intersection is nonempty. This result is a diagnostic for that simultaneous set of source assumptions.

Two explicit one-dimensional benchmarks illustrate the contour mechanism. Proposition 3 treats the symmetric mixture $\frac{1}{2}N(-1,1) + \frac{1}{2}N(1,1)$, whose transform has a known imaginary zero and whose sine-ratio estimator attains C/n risk under the stated range, tail, independence, and $L^1(P_X)$ treatment-code conditions. Lemma 7 and Theorem 5 give a Gaussian–Rademacher path and a local ACE oracle envelope, making the dependence on a shrinking fourth-cumulant magnitude explicit. These benchmarks connect the analysis to non-Gaussian identification ideas in independent component analysis and structural econometrics [Comon, 1994, Hyvärinen and

Oja, 1997, Hyvärinen et al., 2001, Lanne et al., 2017, Lee and Mesters, 2024, Hoesch et al., 2024, Reizinger et al., 2025].

The computational statement is a conditional output-correspondence theorem. Definition 17 and Algorithm 1 use the same primitive records for the ordinary Borel statistic and for represented-data execution. For any compiled bounded spectral adapter satisfying the full canonical build-and-compilation specification for the parameter record, transported fixed records, and base treatment-code sequence, the represented-execution clause in Theorem 3 states that the represented-data execution realizes the same statistic. The appendix records proof and deferred-detail information for the paper's formal statements.

The paper proceeds as follows. The setup section defines the partially linear experiment, supplied-code restrictions, law classes, quantile criterion, and minimax risks. The zero-instrument section proves the known-zero and contour-identification identities. The main fixed-separation section gives zero localization, the contour statistic, and the fixed-code minimax theorem. The comparison section aligns the spectral, ACE, and Gaussian benchmark classes. The mixture section gives explicit sine reductions and local benchmarks. The limitations and future-work section states the local-to-Gaussian adaptation agenda. The appendices collect analytic, empirical-process, lower-bound, certified-construction, verification, and deferred-proof details.

2 Related work

The partially linear model originates in semiparametric regression, where a parametric coefficient is estimated alongside a nonparametric function of covariates [Engle et al., 1986, Speckman, 1988]; Robinson [1988] gave the root-n consistent residualized estimator that remains the reference construction, Härdle and Mammen [1993] developed the associated specification testing, and Härdle et al. [2000] give the monograph treatment. Partially linear models are a central workhorse for semiparametric estimation of low-dimensional structural coefficients in the presence of high-dimensional or otherwise flexible nuisance functions. The classical formulation and efficiency theory connect the coefficient θ_0, the partially linear coefficient, to residualized treatment variation after removing the treatment regression and outcome regression [Hansen, 1982, Newey, 1990, 1994, Newey and McFadden, 1994, Bickel et al., 1993, van der Vaart, 1998, van der Vaart and Wellner, 1996]. Modern double/debiased machine learning builds on the same orthogonalization principle by allowing flexible first-stage estimates while preserving first-order stability of the target moment [Belloni et al., 2014, Chernozhukov et al., 2018]. Within that tradition, the present paper focuses on a partially linear coefficient estimated from supplied treatment and outcome codes, and studies the additional identifying information carried by non-Gaussian treatment innovations.

A closely related line of work studies higher-order orthogonality and influence-function expansions for nuisance-robust semiparametric estimation. Higher-order constructions refine the first-order orthogonal-score idea by exploiting additional smoothness or distributional structure in nuisance errors [Robins et al., 1994, van der Laan and Rubin, 2006, Robins et al., 2008, Liu et al., 2017, Newey and Robins, 2018, Liu et al., 2016, Chernozhukov et al., 2022]. For the partially linear model, Mackey et al. [2018] identify the role of non-Gaussian treatment residuals in higher-order orthogonal learning. Jin et al. [2025] give the closest finite-order comparison point: their published order-r ACE guarantee, with ACE order r, controls the $1-\gamma$ error quantile, where γ is the probability level, by a constant multiple of

$$\varepsilon_{1,n}^r \varepsilon_{2,n} + C_\theta \varepsilon_{1,n}^{r+1} + (\gamma n)^{-1/2},$$

where $\varepsilon_{1,n}$ and $\varepsilon_{2,n}$ are the treatment- and outcome-code radii, C_θ is the coefficient-range bound, and n is the sample size. The explicit prefactor $r!\,16^r \delta^{-1}$, with δ denoting the cumulant-separation level, makes the comparison sensitive to the finite order and to the amount of non-Gaussian separation.

The direct $L^1(P_X)$ gate in the contour theorem plays a different operational role from the JMS pair of $L^r(P_X)$ treatment and outcome radii. The contour gate controls the treatment-code perturbation of the residual transform and preserves the zero geometry; the ACE eligibility quantities in Definitions 8 and 10 also use the outcome-code radius and the order-r finite-difference scale. On bounded covariate-regression ranges, an $L^r(P_X)$ treatment radius at level $\varepsilon_{1,n}$ gives the corresponding $L^1(P_X)$ treatment control, so the common ACE class lies inside the contour class through Lemma 5. The contour result therefore asks the treatment code to meet the displayed small-radius gate, while the published ACE guarantee asks the treatment and outcome radii to meet the JMS eligibility inequalities.

Feature	Contour construction	JMS ACE comparison
Input accuracy	Current $L^1(P_X)$ treatment-code radius and fixed primitive constants	Current $L^r(P_X)$ treatment and outcome radii with JMS eligibility
Analytic resource	A localized zero of the treatment-innovation transform and a selected contour ratio	A finite order-r ACE expansion under cumulant separation
Risk criterion compared	Fixed-code MSE minimax bounds and a generalized-quantile upper bound for the contour statistic	Published generalized-quantile upper guarantee and class-level fixed-code MSE lower bound
Fixed-separation dependence	Constants depend on the fixed cumulant-separation level and primitive envelopes	The displayed imported bound carries the finite-order prefactor and δ^{-1} dependence

This translation gives concrete nuisance-rate regimes for the comparison in Proposition 1. If $\varepsilon_{1,n} = n^{-a}$ and $\varepsilon_{2,n} = n^{-b}$, the ACE nuisance part dominates sampling when $\min\{ra + b, (r+1)a\} < 1/2$, subject to the JMS eligibility and contour small-radius gates; in that regime the ratio of the contour upper guarantee to the imported finite-order ACE upper guarantee tends to zero on the common clipped-code class. If $\min\{ra+b, (r+1)a\} \geq 1/2$, the ACE upper guarantee and the contour upper guarantee are both governed by the sampling term up to constants and eligibility factors. These comparisons concern the available upper bounds on the aligned class.

The contour approach developed here represents the same source of identifying variation through zeros of the treatment-innovation moment-generating function and converts those zeros into instruments and contour ratios. Transform methods have a long history in econometrics and statistics, including characteristic-function estimation, empirical transform methods, and moment-generating-function arguments [Feuerverger and McDunnough, 1981, Singleton, 2001, Chacko and Viceira, 2003, Carrasco et al., 2007, Carrasco, 2017]. The classical foundation for reading distributional structure off a transform is Marcinkiewicz's theorem, that a characteristic function of the form $\exp(P)$ with P a polynomial forces $\deg P \leq 2$ and hence a Gaussian law [Marcinkiewicz, 1939]. In this paper's centered sub-Gaussian envelope, the fixed cumulant-separation condition yields the explicit zero-localization conclusion in Lemma 3: the treatment-innovation MGF has a zero inside the displayed disk. The zero-based step also connects to classical results on distributions, transforms, and completeness [Kagan et al., 1973, Mattner, 1992, D'Haultfoeuille, 2011, Darolles et al., 2011, Hu and Shiu, 2022]. In the present setting, those tools are applied to residualized treatment variation formed with a supplied treatment code, so the analytic question becomes stability of transform zeros under $L^1(P_X)$ treatment-code perturbations.

Non-Gaussianity is also a recurring identifying resource in independent component analysis, structural shock recovery, and related econometric models [Comon, 1994, Hyvärinen and Oja, 1997, Hyvärinen et al., 2001, Lanne et al., 2017, Lee and Mesters, 2024, Hoesch et al., 2024, Reizinger et al., 2025]. Relatedly, Andrews [2017] exhibits families that are L^2-complete or boundedly complete, the property that makes a vanishing transform functional informative under the relevant class. The construction here uses non-Gaussianity through localized transform-zero geometry: it localizes a transform zero and integrates the observable ratio on a selected circle around the enclosed zero structure. The paper applies that resource in a partially linear estimation problem: cumulant separation supplies a transform zero, and the contour construction converts the resulting analytic geometry into a coefficient estimator. This places the contribution between higher-order semiparametric learning and non-Gaussian identification, with the comparison to ACE made on the common cumulant-separated class where both sets of assumptions are imposed.

The fixed-separation statistical statement is expressed in minimax terms. Under fixed cumulant separation and stable supplied treatment codes, the contour estimator attains the parametric n^{-1} mean-squared-error order over the non-Gaussian spectral class, while the comparison with Jin et al. [2025] records the published ACE generalized-quantile upper guarantee on the aligned ACE class. The sequence comparison in Proposition 1 is an upper-guarantee comparison on the common clipped-code class: when the displayed ACE nuisance terms dominate $n^{-1/2}$, the ratio of the fixed-separation contour upper guarantee to the imported finite-order ACE upper guarantee tends to zero.

Finally, the finite-construction component draws on computable analysis and interval arithmetic. The relevant background treats real-number computation, constructive numerical representations, and certified interval enclosures [Weihrauch, 2000, Ko, 1991, Moore, 1966]. Those tools provide the arithmetic substrate for the contour bank and represented-data statistic, while the statistical exposition keeps the identification and risk arguments in standard econometric form.

3 Setup and assumptions

We use generic constants c and C only in later bounds, where their dependence is stated locally. All norms are taken under the law displayed in the surrounding statement, and $L^p(P_X)$ denotes integration with respect to the covariate marginal. The sample index is fixed within each finite-n experiment, while class constants and code-radius sequences are carried by the parameter record. Asymptotic comparisons in later sections vary the sample size through these records, with the order pair, tail scales, cumulant separation, and range bounds held according to the relevant theorem hypotheses.

The basic experiment is a partially linear model with deterministic supplied codes for the treatment and outcome regressions. The first definition records the numerical constants, the split convention, and the exact sub-Gaussian norm convention used throughout.

3.1 Fixed-code decision problem

The supplied treatment and outcome codes are inputs to the current finite-n decision problem. At sample size n, the estimator uses the clipped pair $(\bar{g}_n, \bar{q}_n)$, and the risks below compare laws whose current clipped code pair agrees with that same input. The data used in the risk bound are then the i.i.d. product sample from the law under evaluation, with the code pair treated as fixed for that experiment.

This convention is the one used by the minimax criteria in Definition 9 and by the fixed-code classes in Theorem 3. When the supplied codes come from an external training stage, the theorem applies through the deterministic code pair obtained after conditioning on that training output, or through any deterministic code sequence fixed before the product-law sample is drawn. The theorem's code-radius assumptions are assumptions about that current clipped pair under the covariate marginal P_X.

Definition 1. [`def:model-parameters`] (Parameter record and standing conventions) *A parameter record fixes a sample size $n \geq 1$, an order pair $r, k \in \mathbb{N}$ with*

$$k = r + 1, \qquad r \geq 2,$$

a comparison exponent $s \in [0, \infty]$ with $s \geq r$, a probability level $\gamma \in (1/2, 1)$, positive constants

$$C_\theta,\ C_g,\ C_q,\ \psi_\eta,\ \psi_\xi,\ \delta,\ \sigma > 0,$$

and two code-radius sequences $\varepsilon_1, \varepsilon_2 : \mathbb{N} \to \mathbb{R}$ that are nonnegative and nonincreasing on positive indices; their values at n are written $\varepsilon_{1,n}$ and $\varepsilon_{2,n}$.

Two conventions are used throughout: $N(m, v)$ denotes the Gaussian law on $\mathbb{R}$ with mean m and variance v, and $\mathbb{P}_n f = n^{-1} \sum_{i=1}^n f(O_i)$ denotes the empirical average of f over the sample. The sample is split into the two deterministic folds

$$I_0 = \{i : i < \lfloor n/2 \rfloor\}, \qquad I_1 = \{i : i \geq \lfloor n/2 \rfloor\},$$

written I_a for $a \in \{0, 1\}$.

For a real random variable W and $t > 0$, the Luxemburg sub-Gaussian convention is that $\|W\|_{\psi_2} \leq t$ means $E[\exp(W^2/t^2)] \leq 2$, and its conditional form $\|W\|_{\psi_2|X} \leq t$ means $E[\exp(W^2/t^2) \mid X] \leq 2$ almost surely.

The observation-level model follows the usual partially linear decomposition, with g_0 and q_0 denoting the treatment and outcome regressions and with supplied code sequences clipped to the fixed ranges in Definition 1.

Definition 2. [`def:plm-model`] (Observed-data partially linear model) *An observation is the triple $O = (X, T, Y)$, where X takes values in a measurable covariate space $\mathcal{X}$ and $T, Y \in \mathbb{R}$. A model over a parameter record consists of a probability law P for O under which T and Y are integrable, a real coefficient θ_0, measurable regression functions g_0, q_0 on the covariate space with*

$$g_0(X) = E_P[T \mid X], \qquad q_0(X) = E_P[Y \mid X]$$

P-almost surely, and deterministic measurable supplied code sequences $\widehat{g} = (\widehat{g}_n)_{n\in\mathbb{N}}$ and $\widehat{q} = (\widehat{q}_n)_{n\in\mathbb{N}}$ for g_0 and q_0. The treatment and outcome innovations are

$$\eta = T - g_0(X), \qquad \xi = Y - q_0(X) - \theta_0\eta,$$

and the clipped supplied codes at index n are

$$\bar{g}_n(x) = \min\{\max\{\widehat{g}_n(x), -C_g\}, C_g\}, \qquad \bar{q}_n(x) = \min\{\max\{\widehat{q}_n(x), -C_q\}, C_q\}.$$

The covariate marginal of P is written P_X. A model is written m, and when the law has to be displayed the coefficient is written $\theta_0(P)$.

The sampling and structural restrictions are collected next. They separate the finite-sample product experiment, the partially linear conditional mean, the bounded ranges, and the tail and cumulant conditions that define the statistical classes.

Assumption 1. [`ass:iid-sampling`] (I.i.d. sampling) *Under Definitions 1 and 2, the observations satisfy*

$$O_1, \ldots, O_n \overset{\text{iid}}{\sim} P.$$

The i.i.d. sampling condition is the standard finite-sample product-law setup for orthogonal and semiparametric procedures [Chernozhukov et al., 2018]. It fixes the probability space on which all estimators and generalized quantiles below are evaluated.

Assumption 2. [`ass:independent-treatment-noise`] (Independent treatment noise) *Under Definitions 1 and 2, the treatment innovation is independent of the covariates:*

$$\eta \perp X.$$

Independent treatment noise is the standard residual independence condition used in the cumulant-based partially linear comparison [Jin et al., 2025]. In this paper it also makes the treatment-innovation transform a law-level object separate from the covariate distribution.

Assumption 3. [`ass:outcome-mean-independence`] (Outcome mean independence) *Under Definitions 1 and 2, the outcome innovation has conditionally mean zero:*

$$E_P[\xi \mid X, T] = 0$$

almost surely.

Outcome mean independence is the standard partially linear conditional mean restriction [Robinson, 1988]. It pins down the coefficient through the conditional mean of Y given X and T.

Assumption 4. [`ass:theta-range`] (Coefficient range) *Under Definitions 1 and 2, the partially linear coefficient satisfies*

$$|\theta_0| \le C_\theta.$$

The coefficient range is the standard bounded target range condition in the JMS comparison framework [Jin et al., 2025]. It supplies a common compact parameter range for minimax risks and clipped estimators.

Assumption 5. [ass:g-range] (Treatment-regression range) *Under Definitions 1 and 2, the treatment regression satisfies*

$$\|g_0\|_\infty \le C_g.$$

The bounded treatment regression condition is standard in the same comparison class [Jin et al., 2025]. It also aligns the supplied treatment code with the clipping radius used in Definition 2.

Assumption 6. [ass:q-range] (Outcome-regression range) *Under Definitions 1 and 2, the outcome regression satisfies*

$$\|q_0\|_\infty \le C_q.$$

The bounded outcome regression condition is the standard common sub-Gaussian PLM range condition [Jin et al., 2025]. Together with the treatment and coefficient bounds, it fixes the envelope in which the supplied outcome code is compared.

Assumption 7. [ass:bounded-gaussian-outcome] (Bounded outcome) *In the setting of Definitions 1 and 2, the outcome Y satisfies*

$$|Y| \le C_q$$

almost surely.

The bounded observed-outcome restriction is the standard bounded-Y condition in the simultaneous Gaussian comparison experiment [Jin et al., 2025]. It is used only for that Gaussian intersection class.

Assumption 8. [ass:eta-subgaussian] (Treatment-noise sub-Gaussianity) *In the setting of Definitions 1 and 2, with the Luxemburg convention*

$$\|W\|_{\psi_2} = \inf\{u > 0 : E_P[\exp(W^2/u^2)] \le 2\},$$

the treatment innovation η satisfies

$$\|\eta\|_{\psi_2} \le \psi_\eta.$$

The treatment-tail condition is the standard uniform sub-Gaussian treatment-noise requirement under the exact Luxemburg convention [Jin et al., 2025]. The convention matches the concentration and transform-envelope bounds used later; background sub-Gaussian facts follow the usual conventions in Vershynin [2018] and Wainwright [2019].

Assumption 9. [ass:xi-subgaussian] (Outcome-noise sub-Gaussianity) *In the setting of Definitions 1 and 2, the outcome innovation ξ satisfies*

$$\|\xi\|_{\psi_2|X} \le \psi_\xi$$

almost surely.

The outcome-tail condition is the standard uniform conditional sub-Gaussian outcome-noise restriction [Jin et al., 2025]. It provides the conditional exponential-square control needed for empirical outcome-weighted transforms.

Assumption 10. [`ass:cumulant-separation`] (Cumulant separation) *In the setting of Definitions 1 and 2, the kth cumulant of the treatment innovation η satisfies*

$$|\kappa_k(\eta)| \geq \delta.$$

Cumulant separation is the standard Jin–Mackey–Syrgkanis nonzero cumulant condition with a uniform separation level [Jin et al., 2025]. In the contour analysis, the same lower bound supplies the non-Gaussian signal that later yields transform zeros.

The supplied-code restrictions specify how the deterministic first-stage codes enter the statistical classes. The $L^s(P_X)$ conditions give a common language for comparison with ACE, the $L^r(P_X)$ versions match the published order-r condition, the Gaussian condition defines the comparison experiment, and the $L^1(P_X)$ condition is the stability requirement used by the contour class.

Assumption 11. [`ass:treatment-code-radius`] (Treatment-code radius) *In the setting of Definitions 1 and 2, for every n,*

$$\|\bar{g}_n - g_0\|_{L^s(P_X)} \leq \varepsilon_{1,n}.$$

The $L^s(P_X)$ treatment-code radius is a standard structure-agnostic supplied-code uncertainty set [Jin et al., 2025]. It measures the current-index treatment-code error under the covariate marginal.

Assumption 12. [`ass:outcome-code-radius`] (Outcome-code radius) *In the setting of Definitions 1 and 2, for every n,*

$$\|\bar{q}_n - q_0\|_{L^s(P_X)} \leq \varepsilon_{2,n}.$$

The matching $L^s(P_X)$ outcome-code radius is the standard supplied-code uncertainty set for the outcome regression [Jin et al., 2025]. It places the outcome code and treatment code on the same comparison scale.

Assumption 13. [`ass:jms-treatment-code-radius`] (Treatment-code L^r radius) *In the notation of Definitions 1 and 2, for every $n \geq 1$,*

$$\|\bar{g}_n - g_0\|_{L^r(P_X)} \leq \varepsilon_{1,n}.$$

The $L^r(P_X)$ treatment-code radius is the standard Theorem 5.4 treatment-code uncertainty set in the JMS ACE guarantee [Jin et al., 2025]. It is stated separately because the published finite-order bound is indexed by the ACE order r.

Assumption 14. [`ass:jms-outcome-code-radius`] (Outcome-code L^r radius) *In the notation of Definitions 1 and 2, for every $n \geq 1$,*

$$\|\bar{q}_n - q_0\|_{L^r(P_X)} \leq \varepsilon_{2,n}.$$

The $L^r(P_X)$ outcome-code radius is the corresponding standard Theorem 5.4 outcome-code uncertainty set [Jin et al., 2025]. It is paired with Assumption 13 in the ACE comparison class.

Assumption 15. [`ass:gaussian-treatment-noise`] (Gaussian treatment noise) *In the notation of Definitions 1 and 2, the treatment innovation satisfies*

$$\eta \sim N(0, \sigma^2).$$

The Gaussian treatment-noise condition is the standard known-variance Gaussian experiment used in Theorem 3.2 of Jin et al. [2025]. It defines the Gaussian benchmark class against which the non-Gaussian contour class is compared.

Assumption 16. [`ass:l1-treatment-code-radius`] (Treatment-code L^1 radius) *In the notation of Definitions 1 and 2, for every $n \geq 1$,*

$$\|\bar{g}_n - g_0\|_{L^1(P_X)} \leq \varepsilon_{1,n}.$$

The $L^1(P_X)$ treatment radius is specific to this analysis. It is the stability condition that controls the code-residual perturbation of the treatment transform in the contour argument. Throughout, a *code residual* means $Z_n = T - \bar{g}_n(X)$ formed with the deterministic supplied clipped code $\bar{g}_n$. All risk statements condition on deterministic supplied clipped treatment and outcome codes and evaluate fixed-code performance at the current sample index.

We now package these restrictions into the law classes used by the main theorem and by the ACE and Gaussian comparisons. The broad spectral class keeps the cumulant-separated treatment noise and the $L^1(P_X)$ treatment-code control that are needed for the contour construction.

Definition 3. [`def:non-gaussian-class`] (Non-Gaussian class $\mathcal{P}_{\mathrm{NG},n}$) *Using the notation of Definitions 1 and 2, and with the structural conditions recorded in Assumptions 2 to 6 and 8 to 10, define*

$$\mathcal{P}_{\mathrm{NG},n} = \left\{ P : \begin{array}{l} g_0(x) = E_P[T \mid X = x], \quad q_0(x) = E_P[Y \mid X = x], \\ E_P[Y \mid X, T] = q_0(X) + \theta_0(P)\{T - g_0(X)\}, \\ \eta = T - g_0(X), \quad \xi = Y - q_0(X) - \theta_0(P)\eta, \\ \eta \perp X, \quad E_P[\xi \mid X, T] = 0, \\ |\theta_0(P)| \leq C_\theta, \quad \|g_0\|_\infty \leq C_g, \quad \|q_0\|_\infty \leq C_q, \\ \|\eta\|_{\psi_2} \leq \psi_\eta, \quad \|\xi\|_{\psi_2 | X} \leq \psi_\xi, \\ |\kappa_{r+1}(\eta)| \geq \delta, \quad \|\bar{g}_n - g_0\|_{L^1(P_X)} \leq \varepsilon_{1,n} \end{array} \right\}.$$

Thus membership in $\mathcal{P}_{\mathrm{NG},n}$ is exactly determined by the displayed law, range, tail, cumulant-separation, and treatment-code radius conditions.

The bounded-outcome Gaussian class uses the same partially linear residual functionals but imposes Gaussian treatment noise and the supplied-code restrictions at exponent s. It is the source-stated Gaussian comparison class for the diagnostic and minimax criteria below.

Definition 4. [`def:gaussian-class`]

$$\mathcal{P}^{\mathrm{JMS}}_{\mathrm{G},n} := \left\{ P : \begin{array}{l} 1 \leq n, \ E_P[\xi \mid X, T] = 0 \ P\text{-a.s.}, \ \xi \in L^1(P), \ |\theta_0| \leq C_\theta, \ |Y| \leq C_q \ P\text{-a.s.}, \\ \eta \perp X, \ \eta \sim N(0, \sigma^2), \ |g_0(x)| \leq C_g \ P_X\text{-a.e.}, \\ \|\bar{g}_n - g_0\|_{L^s(P_X)} \leq \varepsilon_{1,n}, \ \|\bar{q}_n - q_0\|_{L^s(P_X)} \leq \varepsilon_{2,n} \end{array} \right\}.$$

Here g_0, q_0, η, ξ are the residual functionals determined by P in Definition 3.

The ACE comparison class records the published cumulant-separated setup at order r. Its member conditions mirror the preceding assumptions while using the $L^r(P_X)$ code radii that enter the JMS finite-order guarantee.

Definition 5. [`def:jms-ace-class`] (JMS ACE comparison class $\mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}$) *Fix an admissible parameter record with positive sample-size coordinate and*

$$k = r + 1, \qquad 2 \le r, \qquad r \le s, \qquad \gamma \in (1/2, 1),$$

with

$$0 < C_\theta, \quad 0 < C_g, \quad 0 < C_q, \quad 0 < \psi_\eta, \quad 0 < \psi_\xi, \quad 0 < \delta, \quad 0 < \sigma,$$

and with radius sequences satisfying

$$0 \le \varepsilon_{j,b} \le \varepsilon_{j,a} \qquad (j \in \{1, 2\},\ 1 \le a \le b).$$

At any positive class index n, the JMS ACE comparison class $\mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}$ consists of all model records m whose observed-data component P is a probability law on $O = X \times \mathbb{R} \times \mathbb{R}$, carrying a real θ_0, measurable functions $g_0, q_0 : X \to \mathbb{R}$, and deterministic measurable supplied code sequences whose clipped versions are denoted $\bar{g}_\ell$ and $\bar{q}_\ell$. The coordinates T and Y are P-integrable, and the conditional-mean identities

$$g_0(X) = E_P[T \mid X], \qquad q_0(X) = E_P[Y \mid X]$$

hold P-almost surely. Writing $\eta = T - g_0(X)$ and $\xi = Y - q_0(X) - \theta_0\eta$, the member conditions in Assumptions 2 to 6 and 8 to 10 hold in the following form:

$$\xi \in L^1(P), \qquad E_P[\xi \mid X, T] = 0, \qquad \eta \perp X,$$

with the conditional-mean identity for ξ holding P-almost surely;

$$|\theta_0| \le C_\theta, \qquad |g_0(x)| \le C_g, \qquad |q_0(x)| \le C_q,$$

with the displayed bounds on g_0 and q_0 holding for P_X-almost every x;

$$\exp(\eta^2/\psi_\eta^2) \in L^1(P), \qquad \int \exp(\eta^2/\psi_\eta^2)\, dP \le 2,$$

$$\exp(\xi^2/\psi_\xi^2) \in L^1(P), \qquad E_P[\exp(\xi^2/\psi_\xi^2) \mid X] \le 2,$$

with the conditional tail bound holding P-almost surely; and

$$|\kappa_{r+1}(\eta)| \ge \delta, \qquad \|\bar{g}_n - g_0\|_{L^r(P_X)} \le \varepsilon_{1,n}, \qquad \|\bar{q}_n - q_0\|_{L^r(P_X)} \le \varepsilon_{2,n}.$$

Risk comparisons use a generalized lower quantile for absolute estimation error. Because $\gamma \in (1/2, 1)$, the controlled level $1 - \gamma$ lies below one half, so $Q_{P,1-\gamma}$ is a lower quantile of the absolute error and every quantile bound below is to be read on that scale. The definition is stated for an arbitrary real sample statistic so it can be applied both to estimator errors and to the published ACE quantile bound.

Definition 6. [`def:generalized-quantile`] (Generalized quantile $Q_{P,1-\gamma}(W)$) *In the sampling setup of Assumption 1 and the notation of Definitions 1 and 2, let P be an observed-data law, let $n \ge 1$, and let $W = W(O_1, \ldots, O_n)$ be a real-valued sample statistic. Under the product law $P^{\otimes n}$, write*

$$F_W(t) = \Pr\{W \le t\}.$$

For $\gamma \in (1/2, 1)$, the generalized lower $(1-\gamma)$-quantile of W is

$$Q_{P,1-\gamma}(W) = \inf\{t \in \mathbb{R} : F_W(t) \ge 1 - \gamma\}.$$

This is the left-continuous generalized inverse of F_W evaluated at probability level $1 - \gamma$.

The contour factorization in the next section is expressed through the covariate-level error made by the supplied treatment code. The following notation isolates that error and the corresponding outcome-contamination transform.

Definition 7. [`def:outcome-contamination`] (Treatment-code error D_n, outcome contamination b_n, and its transform B_n) *Fix a law P, an index n, and the clipped supplied treatment code $\bar{g}_n$. The treatment-code error and the outcome contamination are the covariate functions*

$$D_n(X) = g_0(X) - \bar{g}_n(X), \qquad b_n(X) = q_0(X) - \theta_0\{g_0(X) - \bar{g}_n(X)\} = q_0(X) - \theta_0 D_n(X),$$

and the contamination transform is their weighted exponential transform

$$B_n(z) = E_P\big[b_n(X)e^{zD_n(X)}\big], \qquad z \in \mathbb{C}.$$

For comparison with Jin et al. [2025], we also record the eligibility quantities that enter the published order-r ACE condition. They combine code radii, sample size, tail scales, coefficient range, and cumulant separation into the finite-order admissibility check.

Definition 8. [`def:jms-eligibility-quantities`] (JMS eligibility quantities $a_{1,n}, b_{1,n}, a_2, b_{2,n}$) *Under the notation and conditions in Assumptions 4, 5, 8 to 10, 13 and 14 and Definitions 1 and 2, the four quantities entering the Jin–Mackey–Syrgkanis order-r eligibility condition at index n are*

$$a_{1,n} = 2\log\left(\frac{6(C_g+\psi_\eta)}{\varepsilon_{1,n}}\right), \qquad b_{1,n} = \log\left(\frac{\gamma n}{9}\right), \qquad a_2 = 4(C_g+\psi_\eta),$$

$$b_{2,n} = \frac{200\min\{1, C_\theta\}\,\delta}{\max\big\{\varepsilon_{1,n},\ \varepsilon_{2,n},\ (\gamma n)^{-1/2}(\psi_\xi + C_\theta\psi_\eta)\big\}}.$$

The minimax criteria are class-indexed and fix the supplied codes at the current sample size. This keeps the decision problem focused on laws compatible with the same clipped code pair.

Definition 9. [`def:minimax-risks`] (Minimax risks $\mathfrak{R}_n^{\mathrm{NG}}, \mathfrak{R}_n^{\mathrm{G}}, \mathfrak{M}_{n,1-\gamma}^{\mathrm{G}}$) *For fixed supplied code sequences, let $\bar{g}_n$ and $\bar{q}_n$ denote their current-index clippings to $[-C_g, C_g]$ and $[-C_q, C_q]$. All infima below range over measurable estimators based on $O_1, \ldots, O_n$, and each supremum is restricted to laws in the displayed class whose clipped treatment and outcome codes at index n are $\bar{g}_n$ and $\bar{q}_n$. Given a parameter record p and supplied code sequences $\widehat{g}, \widehat{q}$, a class $\mathcal{C}$ of laws carries the notation $\mathcal{C}(p; \bar{g}_n, \bar{q}_n)$ for the subset cut out by this fixed-code restriction, and we write $\mathfrak{R}_n^{\mathrm{NG}}(p; \widehat{g}, \widehat{q})$, $\mathfrak{R}_n^{\mathrm{G}}(p; \widehat{g}, \widehat{q})$, and $\mathfrak{M}_{n,1-\gamma}^{\mathrm{G}}(p; \widehat{g}, \widehat{q})$ for the three risks defined next whenever their dependence on the parameter record and the supplied codes has to be displayed. Define the non-Gaussian minimax mean-squared risk by*

$$\mathfrak{R}_n^{\mathrm{NG}} = \inf_{\widetilde{\theta}(O_1,\ldots,O_n;\bar{g}_n,\bar{q}_n)} \sup_{P\in\mathcal{P}_{\mathrm{NG},n}} E_P\left[(\widetilde{\theta} - \theta_0(P))^2\right].$$

Define the Gaussian JMS minimax mean-squared risk by

$$\mathfrak{R}_n^{\mathrm{G}} = \inf_{\widetilde{\theta}(O_1,\ldots,O_n;\bar{g}_n,\bar{q}_n)} \sup_{P\in\mathcal{P}_{\mathrm{G},n}^{\mathrm{JMS}}} E_P\left[(\widetilde{\theta} - \theta_0(P))^2\right].$$

For $1/2 < \gamma < 1$, define the Gaussian JMS minimax generalized lower-quantile risk by

$$\mathfrak{M}_{n,1-\gamma}^{\mathrm{G}} = \inf_{\widetilde{\theta}} \sup_{P\in\mathcal{P}_{\mathrm{G},n}^{\mathrm{JMS}}} Q_{P,1-\gamma}\left(|\widetilde{\theta} - \theta_0(P)|\right).$$

The order-r ACE guarantee is invoked in later comparisons through its eligibility predicate. The predicate below is the finite-sample condition formed from the four quantities in Definition 8.

Definition 10. [`def:jms-eligibility`] (JMS eligibility condition $\mathsf{E}_n^{\mathrm{JMS}}$) *With $a_{1,n}, b_{1,n}, a_2, b_{2,n}$ defined as in Definition 8, the JMS order-r eligibility condition is*

$$\mathsf{E}_n^{\mathrm{JMS}} \quad \Longleftrightarrow \quad r \leq \min\left\{\frac{b_{1,n}}{a_{1,n}} - \frac{1}{a_{1,n}}\log\left(\frac{b_{1,n}}{a_{1,n}}\right), \frac{b_{2,n}}{a_2\log(a_2 b_{2,n})}\right\},$$

on the indices for which the displayed expressions are defined.

Finally, the common ACE comparison class is obtained by adding the $L^s(P_X)$ treatment- and outcome-code restrictions to the broad spectral class. This subclass is the domain on which the contour and ACE upper guarantees are aligned in the comparison section.

Definition 11. [`def:ace-comparison-subclass`] (ACE comparison subclass $\mathcal{P}_{\mathrm{NG},n}^{\mathrm{ACE}}$) *The ACE comparison subclass of the non-Gaussian class in Definition 3 is*

$$\mathcal{P}_{\mathrm{NG},n}^{\mathrm{ACE}} = \left\{P \in \mathcal{P}_{\mathrm{NG},n} : \|\bar{g}_n - g_0\|_{L^s(P_X)} \leq \varepsilon_{1,n},\ \|\bar{q}_n - q_0\|_{L^s(P_X)} \leq \varepsilon_{2,n}\right\}.$$

It consists of laws satisfying the broad spectral-class conditions together with the displayed supplied-code restrictions.

4 Zero instruments and contour identification

The analytic mechanism starts from a zero of the treatment-innovation transform and turns it into an instrument for the residualized treatment. In the partially linear model, the supplied treatment code enters through the residual $Z_n = T - \bar{g}_n(X)$; the point of the construction is that a transform zero of the innovation η continues to generate usable population moments after the residual has been shifted by the treatment-code error. The use of transform zeros follows a classical identification tradition for moment-generating and characteristic functions [Kagan et al., 1973, Mattner, 1992, D'Haultfoeuille, 2011, Darolles et al., 2011, Hu and Shiu, 2022], while the contour formulation below is aligned with transform-based estimation in econometrics [Feuerverger and McDunnough, 1981, Singleton, 2001, Carrasco et al., 2007, Carrasco, 2017].

The first object is the treatment-innovation transform $M(z)$ and its zero-based instrument. If z_0 is a zero of multiplicity ℓ, differentiating the exponential tilt at that zero produces the polynomial-exponential weight used throughout this section.

Definition 12. [`def:zero-instrument`] (Treatment-innovation transform M) *Let $M : \mathbb{C} \to \mathbb{C}$, let $z_0 \in \mathbb{C}$, and let $\ell \in \mathbb{N}$ satisfy $1 \leq \ell$, $M(z_0) = 0$, and analytic order ℓ for M at z_0. Define*

$$J_{z_0,\ell}(w) = w^{\ell-1}e^{z_0 w}, \qquad w \in \mathbb{C}.$$

The instrument $J_{z_0,\ell}(w)$ is indexed by the complex zero z_0 and the positive multiplicity ℓ. Its role is local in the transform variable but global in the residual shift: the same weight will be evaluated at shifted innovations and at the code residual.

Theorem 1. [`thm:known-zero-instrument`] (Zero-based instrument identities) *Fix a model-parameter record p, a model m with observed-data law P, and an integer $n \geq 1$. Let*

$$M(z) = E_P[e^{z\eta}], \qquad \eta = T - g_0(X), \qquad Z_n = T - \bar{g}_n(X), \qquad H_n(z) = E_P\left[e^{z\{g_0(X)-\bar{g}_n(X)\}}\right].$$

Suppose that:

- ***(Non-Gaussian class.)*** *The model m belongs to $\mathcal{P}_{\mathrm{NG},n}$ as specified in Definition 3.*
- ***(Zero and multiplicity.)*** *For some $z_0 \in \mathbb{C}$ and integer $\ell \geq 1$, $M(z_0) = 0$ and the analytic order of M at z_0 is ℓ.*
- ***(Instrument.)*** *$J_{z_0,\ell}$ is the zero-based instrument from Definition 12, equivalently*

$$J_{z_0,\ell}(w) = w^{\ell-1} e^{z_0 w}.$$

Then, for every $d \in \mathbb{R}$,

$$E_P[J_{z_0,\ell}(\eta + d)] = 0.$$

Moreover,

$$E_P[J_{z_0,\ell}(Z_n) \mid X] = 0 \quad P\text{-almost surely,}$$

and

$$E_P[Z_n J_{z_0,\ell}(Z_n)] = M^{(\ell)}(z_0) H_n(z_0).$$

Whenever the denominator is nonzero,

$$E_P[Z_n J_{z_0,\ell}(Z_n)] \neq 0,$$

the coefficient is identified by

$$\theta_0 = \frac{E_P[Y J_{z_0,\ell}(Z_n)]}{E_P[Z_n J_{z_0,\ell}(Z_n)]}.$$

Theorem 1 gives the moment-ratio form of the argument when a valid zero is known. The identity $E_P[J_{z_0,\ell}(\eta + d)] = 0$ applies uniformly over real shifts d, so conditioning on X converts the residualized treatment into a valid zero moment. The denominator identity identifies the relevant scale as the product of the nonzero derivative at the transform zero and the nuisance transform $H_n(z)$; when that scale is nonzero, the partially linear coefficient is the displayed population ratio.

For data-driven contour selection, the paper uses observable transforms of the code residual. The treatment-residual transform $F_n(z)$ carries the zero geometry, the outcome-weighted transform $G_n(z)$ carries the numerator information, and a contour average of G_n/F_n extracts the coefficient over all enclosed zeros at once.

Definition 13. [`def:contour-functional`] (Contour functional $\Theta_{C_j}(F_n, G_n)$) *For $Z_n = T - \bar{g}_n(X)$, define*

$$F_n(z) = E_P[e^{zZ_n}], \qquad G_n(z) = E_P[Y e^{zZ_n}], \qquad H_n(z) = E_P[e^{z\{g_0(X) - \bar{g}_n(X)\}}].$$

Let C_j be a positively oriented circle centered at the origin with positive radius, and suppose F_n is analytic on a neighborhood of the closed disk bounded by C_j. Let $N_{C_j}(F_n)$ denote the number of zeros of F_n strictly inside C_j, counted with multiplicity. When F_n is zero-free on C_j and $N_{C_j}(F_n) \geq 1$, define

$$\Theta_{C_j}(F_n, G_n) = \{N_{C_j}(F_n) 2\pi i\}^{-1} \oint_{C_j} \frac{G_n(z)}{F_n(z)}\, dz.$$

Here C_j is the positively oriented circle used for the contour integral, $N_{C_j}(F_n)$ is the enclosed zero count, and $\Theta_{C_j}(F_n, G_n)$ is the normalized contour functional. The normalization by the zero count makes the functional an average over the enclosed zero structure.

The next identity links this observable contour world to the latent treatment innovation. It factors the residual transform through the innovation transform and isolates the covariate-only contamination term that enters the outcome-weighted transform.

Lemma 1. [`lem:observable-factorization`] (Observable transform factorization) *Let m be a model with observed-data law P in the non-Gaussian class $\mathcal{P}_{\mathrm{NG},n}$ of Definition 3. For*

$$Z_n = T - \bar{g}_n(X), \qquad D_n(X) = g_0(X) - \bar{g}_n(X), \qquad \eta = T - g_0(X),$$

define, for $z \in \mathbb{C}$,

$$F_n(z) = E_P[e^{zZ_n}], \qquad M(z) = E_P[e^{z\eta}], \qquad H_n(z) = E_P[e^{zD_n(X)}],$$

and

$$G_n(z) = E_P[Ye^{zZ_n}], \qquad B_n(z) = E_P[b_n(X)e^{zD_n(X)}],$$

where $b_n(X)$ is the outcome-contamination function in Definition 7. Then, for every $z \in \mathbb{C}$,

$$F_n(z) = M(z)H_n(z)$$

and

$$G_n(z) = \theta_0(P)F_n'(z) + M(z)B_n(z).$$

The proof is deferred to Section E.

Lemma 1 is the bridge from residual contamination to contour identification. The factorization $F_n = MH_n$ states that the observable residual transform inherits zeros from the treatment-innovation transform whenever the nuisance transform stays away from zero. The second display decomposes G_n/F_n into the derivative term carrying θ_0 and a nuisance ratio; integrated around a valid contour, the latter contribution is analytic inside the contour and the former contributes through the enclosed zero count.

The population identification statement packages these ingredients into the contour ratio used later by the statistical construction. The contour must be valid for the observable residual transform and for the nuisance transform on the same disk.

Theorem 2. [`thm:exact-contour-identification`] (Exact contour identification) *Fix $n \geq 1$, a law $P \in \mathcal{P}_{\mathrm{NG},n}$ as in Definition 3, and a certified contour-bank input $\mathfrak{p}_\star$ as in Definition 17. For any bank index $j \in \{0, \ldots, J_{\mathrm{base}}\}$, let C_j be the positively oriented circle with radius ρ_j, and let*

$$F_n(z) = E_P[e^{zZ_n}], \qquad G_n(z) = E_P[Ye^{zZ_n}], \qquad H_n(z) = E_P[e^{z\{g_0(X) - \bar{g}_n(X)\}}].$$

Assume:

- ***(Residual zero-freeness.)*** *$F_n(z) \neq 0$ for every $z \in C_j$.*
- ***(Nuisance zero-freeness.)*** *$H_n(z) \neq 0$ for every z on or inside C_j.*
- ***(Positive contour count.)*** *The zero count $N_{C_j}(F_n)$, counted with multiplicity inside C_j, satisfies $N_{C_j}(F_n) \geq 1$.*

Then the contour functional of Definition 13 identifies the partially linear coefficient and is given by the normalized contour integral:

$$(\theta_0(P) : \mathbb{C}) = \Theta_{C_j}(F_n, G_n) = \{N_{C_j}(F_n)\, 2\pi i\}^{-1} \oint_{C_j} \frac{G_n(z)}{F_n(z)}\, dz.$$

Theorem 2 establishes the population contour identity on any bank circle satisfying the stated zero-free and positive-count conditions. Residual zero-freeness makes G_n/F_n well defined on the integration path; nuisance zero-freeness keeps the zeros enclosed by the residual transform aligned with the innovation-zero geometry supplied by Lemma 1; and the positive count gives a nontrivial normalization. Under those conditions, the contour average of the observable ratio is exactly the partially linear coefficient.

The section therefore reduces identification with code residuals to a geometric task: find a circle on which the observable residual transform is separated from zero, whose interior contains at least one residual-transform zero, and on whose closed disk the nuisance transform is separated from zero. The next section supplies fixed-separation conditions and a finite contour bank under which such circles are available with enough empirical stability to yield the stated fixed-code risk bounds.

5 Main fixed-separation result

The preceding section gave the population contour identity. The fixed-separation result turns that identity into a uniform statistical construction by combining three ingredients: an explicit zero-localization radius from cumulant separation, an $L^1(P_X)$ treatment-code gate that preserves the relevant zero geometry, and empirical transform control on a finite contour bank. The empirical-process component follows the same uniform-supremum logic used in semiparametric and double/debiased learning analyses [Bickel et al., 1993, van der Vaart, 1998, van der Vaart and Wellner, 1996, Chernozhukov et al., 2018], while the contour bank uses finite analytic localization of zeros of exponential transforms [Michelen and Sahasrabudhe, 2019, Eremenko and Fryntov, 2021, Dinh et al., 2021].

The first auxiliary bound fixes the Luxemburg normalization used throughout the paper. It converts the sub-Gaussian convention in Definition 1 into a concrete moment-generating-function envelope, so the constants entering the zero-localization radius remain explicit.

Lemma 2. [`lem:luxemburg-mgf-envelope`] (Luxemburg MGF envelope) *Following the Luxemburg sub-Gaussian convention in Definition 1, let $\psi > 0$ be a Luxemburg scale and let W be a real-valued measurable random variable on a probability space with law P. Assume:*

- ***(Integrability.)*** *$W \in L^1(P)$ and $\exp(W^2/\psi^2) \in L^1(P)$.*
- ***(Centeredness.)*** *$E_P W = 0$.*
- ***(Luxemburg envelope.)***

$$E_P \exp(W^2/\psi^2) \le 2.$$

Then, for every $t \in \mathbb{R}$,

$$E_P e^{tW} \le \exp(4\psi^2 t^2).$$

The proof is deferred to Section E.

Lemma 2 supplies the analytic tail envelope used to control treatment-innovation transforms on complex disks. With the exact convention fixed in Assumption 8, the bound carries the displayed factor 4, which then feeds directly into the localization constants.

Cumulant separation gives the treatment-innovation transform a zero inside an explicit disk. This step is the bridge from the population identification formulas in Theorems 1 and 2 to a finite search region.

Lemma 3. [`lem:zero-localization`] *Fix a primitive parameter tuple p and a law P for $O = (X, T, Y)$. Assume:*

- $k = r + 1$ *with* $r \geq 2$*;*
- $\eta = T - g_0(X)$ *is centered;*
- $\|\eta\|_{\psi_2} \leq \psi_\eta$ *under the convention in Assumption 8; and*
- $|\kappa_k(\eta)| \geq \delta$ *as in Assumption 10.*

Define

$$A_k = (2^{k+4}k!)^{1/(k-2)}, \qquad R_0 = A_k(\psi_\eta^2/\delta)^{1/(k-2)}, \qquad M(z) = E_P[e^{z\eta}].$$

Then there exists $z_0 \in \mathbb{C}$ such that $|z_0| \leq R_0$ and $M(z_0) = 0$.

The proof is deferred to Section E.

The radius R_0 is determined by the fixed cumulant order, the treatment-noise Luxemburg scale, and the separation level. The contour construction expands this disk to the outer radius R_1, allowing the finite bank to place candidate circles around the localized zero while retaining a buffer for perturbation control.

The numerator transform also needs a fixed envelope on the search region. The next bound records a population scale for the observable outcome-weighted transform on the disk used by the contour bank.

Lemma 4. [`lem:population-numerator-envelope`] (Population numerator envelope) *Let p be a parameter record as in Definition 1, and let m be a model with observed-data law P as in Definition 2. Write $n = p.n$, let R_1 be the outer search radius from Definition 17, and define the population numerator envelope*

$$C_G(p) = \left[64\{C_q^4 + 4C_\theta^4\psi_\eta^4 + 4\psi_\xi^4\} + 2\exp\{16R_1C_g + 16R_1^2\psi_\eta^2\}\right]^{1/2}.$$

Suppose that:

- ***(Non-Gaussian class.)*** *The model m, equivalently its law P, lies in the non-Gaussian class $\mathcal{P}_{\mathrm{NG},n}$ of Definition 3.*
- ***(Search window.)*** *The complex point $z \in \mathbb{C}$ satisfies $|z| \leq R_1$.*

Then the observable outcome-weighted transform $G_n(z) = E_P[Ye^{zZ_n}]$ from Definition 13 satisfies

$$|G_n(z)| \leq C_G(p).$$

The proof is deferred to Section E.

Lemma 4 keeps the outcome side of the contour ratio on the same fixed scale as the denominator search region. The envelope depends on the primitive bounds and R_1, so the later risk constants can be indexed by the fixed class constants.

The estimator is a finite, split-sample contour program. The pilot fold selects a circle using the empirical winding number and a denominator lower-modulus certificate; the evaluation fold computes the normalized contour ratio on that selected circle and clips the real midpoint to the parameter range. The certified-record clauses state the corresponding finite-rational implementation contract for the same statistic.

Algorithm 1 (Adaptive contour estimator $\widehat{\theta}_{\mathrm{spec},n}$)

The estimator $\widehat{\theta}_{\mathrm{spec},n}$ is the packaged adaptive contour estimator on the public domain where the current supplied treatment-regression code underlying $\bar{g}_n$ is measurable. It takes as inputs $O_1, \ldots, O_n$, the supplied treatment-regression code sequence through its current clipped values $\bar{g}_n$, the parameter record p, and fixed certified records $\mathfrak{p}_\star$ and $\mathfrak{c}_{\theta,\star}$ representing $k, \delta, \psi_\eta, R_0, R_1$, and C_θ, with

$$R_0 = A_k \left(\frac{\psi_\eta^2}{\delta}\right)^{(k-2)^{-1}}, \qquad R_1 = R_0 + 1.$$

Here p contains $n \geq 1$, $r, k \in \mathbb{N}$ with $k = r + 1$ and $r \geq 2$, $s \in [0, \infty]$ with $r \leq s$, $\gamma \in (1/2, 1)$, positive constants $C_\theta, C_g, C_q, \psi_\eta, \psi_\xi, \delta, \sigma$, and nonnegative nonincreasing sequences $\varepsilon_{1n}, \varepsilon_{2n} : \mathbb{N} \to \mathbb{R}$ on positive indices.

1. Define the deterministic folds

$$I_0 = \{i : i < \lfloor n/2 \rfloor\}, \qquad I_1 = \{i : i \geq \lfloor n/2 \rfloor\}.$$

   For each real observation coordinate x and integer $m \geq 0$, define the floor-dyadic interval

$$\mathcal{I}_m(x) = \left[2^{-(m+1)} \lfloor 2^{m+1} x \rfloor, 2^{-(m+1)} \{\lfloor 2^{m+1} x \rfloor + 1\}\right].$$

2. Form the residualized treatment values

$$Z_{n,i} = T_i - \bar{g}_n(X_i), \qquad i = 1, \ldots, n.$$

   For each fold $a \in \{0, 1\}$, derivative order p, and complex z, define

$$\widehat{F}_{a,n}^{(p)}(z) = \big(\max\{|I_a|, 1\}\big)^{-1} \sum_{i \in I_a} Z_{n,i}^p e^{z Z_{n,i}}, \qquad p \in \{0, 1, 2\},$$

   and

$$\widehat{G}_{a,n}^{(p)}(z) = \big(\max\{|I_a|, 1\}\big)^{-1} \sum_{i \in I_a} Y_i Z_{n,i}^p e^{z Z_{n,i}}, \qquad p \in \{0, 1\}.$$

   Thus an empty fold contributes the empty sum and is normalized by one.

3. For represented observation data, define

$$\widetilde{\mathfrak{s}} = \big((\mathfrak{t}_i, \mathfrak{y}_i, \mathfrak{g}_i)_{i=1}^n\big),$$

where $\mathfrak{t}_i$, $\mathfrak{y}_i$, and $\mathfrak{g}_i$ are certified-real records naming T_i, Y_i, and $\bar{g}_n(X_i)$. The full represented-data input is

$$\mathfrak{s}_\star = \big((\mathfrak{t}_i, \mathfrak{y}_i, \mathfrak{g}_i)_{i=1}^n, \mathfrak{p}_\star, \mathfrak{c}_{\theta,\star}\big).$$

For $i = 1, \dots, n$, put

$$\mathfrak{z}_i = \mathfrak{t}_i - \mathfrak{g}_i.$$

4. Invoke the contour bank determined by $\mathfrak{p}_\star$. For the jth circle, define

$$C_j = \{z_j(t) : 0 \le t \le 1\}, \qquad z_j(t) = \rho_j e^{2\pi i t}.$$

At error tolerance one, refine $\mathfrak{z}_i$, $\mathfrak{y}_i$, and ρ_j using their own executable moduli, and let $U_{Z,i}$, $U_{Y,i}$, and $U_{\rho,j}$ be the largest absolute endpoint of the corresponding rational intervals. For a canonical observation name this is the interval $\mathcal{I}_{\mu(1)}$ with the convention above; for $\mathfrak{z}_i$ it is the certified subtraction interval obtained from $\mathfrak{t}_i$ and $\mathfrak{g}_i$, whose modulus uses the two half-error input moduli. Using $e^x \le 3^{\lceil x \rceil}$ for $x \ge 0$ and $\pi < 4$, form rational upper bounds for the empirical derivative sums on each C_j.

5. For each j, compute a rational bracket of width at most $a_\star/64$ enclosing

$$\inf_{z \in C_j} |\widehat{F}_{0,n}(z)|.$$

Let m_j be the lower endpoint of this bracket. When $m_j > 0$, define

$$W_{0j}(t) = \rho_j e^{2\pi i t} \frac{\widehat{F}'_{0,n}(z_j(t))}{\widehat{F}_{0,n}(z_j(t))}.$$

Use the rational Lipschitz bound

$$|W'_{0j}(t)| \le 64 \left[\rho_j \frac{\sup |\widehat{F}'_{0,n}|}{m_j} + \rho_j^2 \left\{ \frac{\sup |\widehat{F}''_{0,n}|}{m_j} + \frac{\sup |\widehat{F}'_{0,n}|^2}{m_j^2} \right\} \right].$$

6. Compute a rectangle enclosure, with coordinate widths strictly below 1/4, for

$$\int_0^1 W_{0j}(t)\, dt = \frac{1}{2\pi i} \oint_{C_j} \frac{\widehat{F}'_{0,n}(z)}{\widehat{F}_{0,n}(z)}\, dz.$$

Define

$$N_j = \frac{1}{2\pi i} \oint_{C_j} \frac{\widehat{F}'_{0,n}(z)}{\widehat{F}_{0,n}(z)}\, dz$$

when the real coordinate contains a unique nonnegative integer and the imaginary coordinate contains zero. A circle is admissible when

$$m_j \ge a_\star/2 \qquad \text{and} \qquad N_j \ge 1.$$

Let

$$j_\star = \min \left\{ j : j \in \arg \max_{\ell:\, m_\ell \ge a_\star/2,\ N_\ell \ge 1} m_\ell \right\}.$$

If the admissible set is empty, set

$$\widehat{\theta}_{\mathrm{spec},n} = 0.$$

7. On the evaluation fold for the selected circle $C_{j_\star}$, compute a lower endpoint

$$m_{j_\star}^{(1)} \le \inf_{z \in C_{j_\star}} |\widehat{F}_{1,n}(z)|.$$

If

$$m_{j_\star}^{(1)} < a_\star/4,$$

set

$$\widehat{\theta}_{\mathrm{spec},n} = 0.$$

8. Otherwise define the evaluation-fold contour integrand

$$V_{j_\star}(t) = \rho_{j_\star} e^{2\pi i t} \frac{\widehat{G}_{1,n}(z_{j_\star}(t))}{\widehat{F}_{1,n}(z_{j_\star}(t))}.$$

Use the rational Lipschitz bound

$$|V_{j_\star}'(t)| \le 64 \left[\rho_{j_\star} \frac{\sup|\widehat{G}_{1,n}|}{m_{j_\star}^{(1)}} + \rho_{j_\star}^2 \left\{ \frac{\sup|\widehat{G}_{1,n}'|}{m_{j_\star}^{(1)}} + \frac{\sup|\widehat{G}_{1,n}| \sup|\widehat{F}_{1,n}'|}{(m_{j_\star}^{(1)})^2} \right\} \right].$$

For each rational tolerance $e > 0$ and rational Lipschitz bound L, the schedule spends one third of the tolerance on discretization and uses

$$N_{\mathrm{mesh}} = \left\lceil \frac{3L}{e} \right\rceil + 1,$$

so that $L/N_{\mathrm{mesh}} \le e/3$.

9. Enclose the unnormalized contour integral

$$\int_0^1 V_{j_\star}(t)\, dt = \frac{1}{2\pi i} \oint_{C_{j_\star}} \frac{\widehat{G}_{1,n}(z)}{\widehat{F}_{1,n}(z)}\, dz.$$

Divide the resulting rectangle by $N_{j_\star}$ to enclose

$$\frac{1}{N_{j_\star} 2\pi i} \oint_{C_{j_\star}} \frac{\widehat{G}_{1,n}(z)}{\widehat{F}_{1,n}(z)}\, dz$$

with coordinate radius at most $1/n$. Take $y \in \mathbb{Q}$ to be the real midpoint of that enclosure, so that y approximates the normalized contour integral to within $1/n$:

$$\left| y - \mathrm{Re}\left(\frac{1}{N_{j_\star} 2\pi i} \oint_{C_{j_\star}} \frac{\widehat{G}_{1,n}(z)}{\widehat{F}_{1,n}(z)}\, dz \right) \right| \le \frac{1}{n}.$$

Only rational quantities are ever formed, so the estimator returns this certified rational approximation rather than the integral itself; the $1/n$ tolerance is negligible beside the statistical error.

10. Output

$$\widehat{\theta}_{\mathrm{spec},n} = \mathrm{clip}_{C_\theta}(y) = \frac{|y + C_\theta| - |y - C_\theta|}{2} = \Pi_{[-C_\theta, C_\theta]}(y).$$

The statistic $\widehat{\theta}_{\mathrm{spec},n}$ is a total Borel rule because every branch has a displayed output, including the fallback value used when the empirical admissibility or evaluation-fold lower-modulus checks fail. The represented-data clauses give an executable correspondence for implementations satisfying the bounded contour-arithmetic specification of Definition 16; the statistical risk statements below concern the same ordinary-sample statistic.

The main theorem gives the fixed-code minimax statement under fixed cumulant separation. The fixed-code restriction keeps the supplied clipped treatment and outcome codes common across the laws in the risk criterion, matching the decision-theoretic convention in Definition 9.

Theorem 3. [`thm:adaptive-rootn-minimax`] (Fixed-code root-n minimaxity under fixed cumulant separation) *Fix $r \geq 2$ and positive constants $\delta, C_\theta, C_g, C_q, \psi_\eta, \psi_\xi$. Then there are constants $c > 0$ and $C > 0$, depending only on these fixed primitive constants, such that the following statements hold.*

For every measurable covariate space, let p_0 be a parameter record whose primitive entries $r, \delta, C_\theta, C_g, C_q, \psi_\eta, \psi_\xi$ equal the fixed constants above, and fix one experiment-wide record consisting of the certified contour-bank input and the represented range input for p_0. For every parameter record p with the same primitive entries as p_0, and for every base model satisfying the usual measurability, conditional-mean, and integrability requirements, write n for the sample size of p. Let $\widehat{\theta}_{\mathrm{spec},n}$ be the total Borel statistic of Algorithm 1, constructed from the transported fixed records and the base treatment-code sequence. Define the fixed-code non-Gaussian class

$$\mathcal{P}_{\mathrm{NG},n}(p;\mathrm{base}) = \{P : \ P \in \mathcal{P}_{\mathrm{NG},n} \text{ of Definition 3}, \ \bar{g}_n(P) = \bar{g}_n(\mathrm{base}), \ \bar{q}_n(P) = \bar{q}_n(\mathrm{base})\}.$$

Define the fixed-code JMS ACE class

$$\mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}(p;\mathrm{base}) = \{P : \ P \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n} \text{ of Definition 5}, \ \bar{g}_n(P) = \bar{g}_n(\mathrm{base}), \ \bar{q}_n(P) = \bar{q}_n(\mathrm{base})\}.$$

- ***(Measurable statistic.)*** *The statistic $\widehat{\theta}_{\mathrm{spec},n}$ is measurable.*
- ***(Represented execution.)*** *For every compiled bounded spectral adapter satisfying the full canonical build-and-compilation specification for p, the transported fixed records, and the base treatment-code sequence, the represented-data execution realizes the same $\widehat{\theta}_{\mathrm{spec},n}$ as in Algorithm 1.*
- ***(Class relations.)*** *Every law in $\mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}$ belongs to $\mathcal{P}_{\mathrm{NG},n}$. If $s = r$, then $\mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}$ coincides with the $L^s(P_X)$-restricted subclass $\mathcal{P}^{\mathrm{ACE}}_{\mathrm{NG},n}$ of Definition 11.*
- ***(Non-Gaussian fixed-code risk.)*** *If the sample law is the i.i.d. product law of Assumption 1, $\mathcal{P}_{\mathrm{NG},n}(p;\mathrm{base})$ is nonempty, $n \geq 2$, and*

  $$\varepsilon_{1,n} \leq \{4R_1 \exp(2C_g R_1)\}^{-1},$$

  where R_1 is the search radius from the contour-bank construction in Algorithm 1, then

  $$\frac{c}{n} \leq \mathfrak{R}^{\mathrm{NG}}_n \leq \sup_{P \in \mathcal{P}_{\mathrm{NG},n}(p;\mathrm{base})} E_P\left[(\widehat{\theta}_{\mathrm{spec},n} - \theta_0(P))^2\right] \leq \frac{C}{n},$$

  and

  $$\sup_{P \in \mathcal{P}_{\mathrm{NG},n}(p;\mathrm{base})} Q_{P,1-\gamma}\left(|\widehat{\theta}_{\mathrm{spec},n} - \theta_0(P)|\right) \leq \sqrt{\frac{C}{\gamma n}},$$

  with generalized lower quantiles as in Definition 6.

- ***(ACE fixed-code risk.)*** *If $\mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}(p;\mathrm{base})$ is nonempty, $n \geq 2$, and*

$$\varepsilon_{1,n} \leq \{4R_1 \exp(2C_g R_1)\}^{-1},$$

then

$$\frac{c}{n} \leq \inf_{\widetilde{\theta}} \sup_{P \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}(p;\mathrm{base})} E_P\left[(\widetilde{\theta} - \theta_0(P))^2\right] \leq \sup_{P \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}(p;\mathrm{base})} E_P\left[(\widehat{\theta}_{\mathrm{spec},n} - \theta_0(P))^2\right] \leq \frac{C}{n},$$

where the infimum is over measurable real-valued estimators of $O_1, \ldots, O_n$. Moreover,

$$\sup_{P \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}(p;\mathrm{base})} Q_{P,1-\gamma}\left(|\widehat{\theta}_{\mathrm{spec},n} - \theta_0(P)|\right) \leq \sqrt{\frac{C}{\gamma n}}.$$

Theorem 3 establishes the fixed-separation statistical guarantee. On the non-Gaussian fixed-code class, the minimax mean-squared risk is sandwiched between constants times n^{-1}, and the contour statistic attains the upper bound. The same statistic also attains the n^{-1} MSE order and the displayed generalized-quantile bound on the fixed-code JMS ACE class whenever that class is nonempty and the treatment-code radius satisfies the stated small-radius gate.

The intuition is the same as in the population contour identity, with sampling error added. Cumulant separation localizes a zero of M; the $L^1(P_X)$ radius controls the perturbation from η to Z_n, preserving a usable contour inside the finite bank; and the split empirical transforms approximate the population transforms on the selected circle at the usual $n^{-1/2}$ stochastic scale. Squaring the coefficient error gives the n^{-1} MSE order, while the lower bound records the matching parametric difficulty inside the same fixed-separation experiment.

The sequence-level statement packages the same construction for common primitive constants. It also records the bounded-outcome Gaussian diagnostic under the JMS comparison class, separating the fixed-separation non-Gaussian rate statement from the Gaussian benchmark risk calculation.

Theorem 4. [`thm:common-experiment-dichotomy`] (Common experiment dichotomy) *Fix $r \geq 2$ and positive constants $\delta, C_\theta, C_g, C_q, \psi_\eta, \psi_\xi$. For every measurable covariate space, the following hold.*

- ***(Non-Gaussian sequence.)*** *Let p_0 have these fixed values of $r, \delta, C_\theta, C_g, C_q, \psi_\eta, \psi_\xi$, and fix the primitive records $\mathfrak{p}_\star$ and $\mathfrak{c}_{\theta,\star}$ of Definition 17 and Algorithm 1. Let (p_j) be any parameter sequence with the same fixed values of $r, \delta, C_\theta, C_g, C_q, \psi_\eta, \psi_\xi$ as p_0, and with common supplied radii $\varepsilon_{1,n}$ and $\varepsilon_{2,n}$. Let $(\widehat{g}_n, \widehat{q}_n)$ be any measurable deterministic code pair, and write $(\bar{g}_n, \bar{q}_n)$ for the clipped pair. Suppose that $n_j = p_j.n \to \infty$ and, for all sufficiently large j,*

$$\mathcal{P}_{\mathrm{NG},n_j}(p_j; \bar{g}_{n_j}, \bar{q}_{n_j}) \neq \varnothing, \qquad n_j \geq 2, \qquad \varepsilon_{1,n_j} \leq \Big(4R_1(p_j)\exp\{2C_g R_1(p_j)\}\Big)^{-1}.$$

Then there are constants $c, C > 0$ such that, with B the translated-dyadic contour bank generated from p_0 and $\mathfrak{p}_\star$, for all sufficiently large j the transported primitive bank for p_j generates the same bank B, the statistic $\widehat{\theta}_{\mathrm{spec},n_j}$ of Algorithm 1 is measurable, every *compiled bounded spectral adapter satisfying the full canonical build and compilation specification is represented by the corresponding execution on the same transported primitive records, and*

$$\frac{c}{n_j} \le \mathfrak{R}^{\mathrm{NG}}_{n_j}(p_j; \widehat{g}, \widehat{q}) \le \frac{C}{n_j}.$$

Moreover, the same statistic witnesses the upper bound:

$$\sup_{P \in \mathcal{P}_{\mathrm{NG},n_j}(p_j;\bar{g}_{n_j},\bar{q}_{n_j})} \mathbb{E}_P\Big[\big(\widehat{\theta}_{\mathrm{spec},n_j} - \theta_0(P)\big)^2\Big] \le \frac{C}{n_j}.$$

- ***(Gaussian class.)*** *For every parameter record p, writing n for its sample size, every $P \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{G},n}$ in Definition 4 satisfies*

  $$\theta_0(P) = 0.$$

  For every measurable deterministic code pair $(\widehat{g}_n, \widehat{q}_n)$, whenever the clipped-code intersection of $\mathcal{P}^{\mathrm{JMS}}_{\mathrm{G},n}$ is nonempty,

  $$\mathfrak{R}^{\mathrm{G}}_{n}(p; \widehat{g}, \widehat{q}) = 0, \qquad \mathfrak{M}^{\mathrm{G}}_{n,1-\gamma}(p; \widehat{g}, \widehat{q}) = 0.$$

Theorem 4 gives two common-experiment conclusions. Along any sequence with the same fixed primitive constants and common supplied radii, the translated-dyadic contour bank stabilizes and the non-Gaussian minimax MSE remains of order n_j^{-1}, with the contour statistic attaining the upper bound. On the bounded-outcome Gaussian JMS class, the partially linear target is identically zero, so both the fixed-code MSE risk and the fixed-code generalized-quantile risk equal zero whenever the clipped-code intersection is nonempty.

The non-Gaussian sequence clause reflects the finite nature of the contour bank: fixed primitive constants determine the same search geometry for all sufficiently large indices satisfying the displayed radius gate. The Gaussian clause records a diagnostic property of the simultaneous bounded-outcome and Gaussian treatment-noise restrictions in Definition 4; under those restrictions, the target itself is pinned to zero, and the corresponding minimax risks inherit that degeneracy. Together, Theorems 3 and 4 provide the fixed-separation rate statement used for the ACE comparison in the next section.

6 Comparison with ACE and Gaussian benchmarks

The fixed-separation theorem in Theorem 3 gives a contour guarantee on the broad spectral class and on the fixed-code ACE class. This section records the comparison with the published order-r ACE guarantee of Jin et al. [2025] on the same clipped-code experiment. The comparison has two parts: first, the class relations align the cumulant-separated spectral and ACE restrictions; second, the two generalized-quantile upper bounds can be read on the common class under the same JMS eligibility condition.

The following proposition states the alignment and the resulting upper-guarantee comparison. The constant C_{spec} is fixed by the primitive constants in the contour construction, while C_γ and the estimator $\widehat{\theta}_{\mathrm{ACE},r,n}$ come from the published ACE handle. The displayed quantity $B_{\mathrm{ACE},n,\gamma}$ is the order-r generalized-quantile upper guarantee from Jin et al. [2025]. The sequence clause compares the displayed upper bounds on the common clipped-code class.

Proposition 1. [`prop:jms-ace-alignment`] (ACE and spectral alignment)[1] *Fix an ACE order $r \geq 2$ and positive constants $\delta, C_\theta, C_g, C_q, \psi_\eta, \psi_\xi$. There is a constant $C_{\text{spec}} > 0$, depending only on these fixed constants, such that the following statements hold.*

Let $\gamma \in (1/2, 1)$, let $C_\gamma > 0$ be the constant in the published order-r ACE generalized-quantile guarantee, and write $\widehat{\theta}_{\text{ACE},r,n}$ for the estimator supplied by that published ACE handle. Let p_0 be a parameter record whose $r, \delta, C_\theta, C_g, C_q, \psi_\eta, \psi_\xi$, and γ coordinates equal the displayed constants, and fix the certified primitive bank and range records associated with p_0. For any parameter record p sharing the same fixed experiment constants as p_0 and the same probability level γ, and for any supplied code sequences whose clipped versions are $\bar{g}_n$ and $\bar{q}_n$, the following hold, under the usual measurability and integrability conditions.

- ***Class relations.*** *At the sample size n of p,*

$$\mathcal{P}^{\text{JMS}}_{\text{ACE},n} \subseteq \mathcal{P}_{\text{NG},n}, \qquad \mathcal{P}^{\text{ACE}}_{\text{NG},n} \subseteq \mathcal{P}^{\text{JMS}}_{\text{ACE},n}.$$

  If the comparison exponent of p satisfies $s = r$, then the common clipped-code convention gives

$$\mathcal{P}^{\text{JMS}}_{\text{ACE},n} = \mathcal{P}^{\text{ACE}}_{\text{NG},n}.$$

  Here the classes and the comparison subclass are those of Definitions 3, 5 and 11.

- ***ACE upper guarantee.*** *If a law $P \in \mathcal{P}^{\text{JMS}}_{\text{ACE},n}$ uses the clipped supplied treatment and outcome codes $\bar{g}_n, \bar{q}_n$, if the JMS eligibility condition $\mathsf{E}^{\text{JMS}}_n$ of Definitions 8 and 10 holds, and if $\varepsilon_{1,n} > 0$ and $\varepsilon_{2,n} \geq 0$, then*

$$Q_{P,1-\gamma}\big(|\widehat{\theta}_{\text{ACE},r,n} - \theta_0(P)|\big) \leq B_{\text{ACE},n,\gamma},$$

  where $Q_{P,1-\gamma}$ is the generalized quantile in Definition 6 and

$$\begin{aligned} B_{\text{ACE},n,\gamma} = C_\gamma r! 16^r \delta^{-1} \Big[&\varepsilon_{1,n}^r \varepsilon_{2,n} + C_\theta \varepsilon_{1,n}^{r+1} \\ &+ 64(C_g + \psi_\eta)^r \{r^2(C_g + \psi_\eta) + \psi_\xi + C_\theta \psi_\eta\}(\gamma n)^{-1/2}\Big]. \end{aligned}$$

- ***Spectral upper guarantee.*** *If $\mathcal{P}^{\text{JMS}}_{\text{ACE},n}$ contains at least one law using the clipped supplied codes, $n \geq 2$, $\mathsf{E}^{\text{JMS}}_n$ holds, $\varepsilon_{1,n} > 0$, and*

$$\varepsilon_{1,n} \leq \{4R_1 \exp(2C_g R_1)\}^{-1},$$

  where R_1 is the outer contour radius of Definition 17 determined by the parameter record in force, then the certified contour statistic $\widehat{\theta}_{\text{spec},n}$ of Algorithm 1, built from the fixed primitive records, satisfies

$$\sup_{P \in \mathcal{P}^{\text{JMS}}_{\text{ACE},n}} Q_{P,1-\gamma}\big(|\widehat{\theta}_{\text{spec},n} - \theta_0(P)|\big) \leq \sqrt{\frac{C_{\text{spec}}}{\gamma n}},$$

  where the supremum ranges over laws in the published ACE class whose clipped supplied treatment and outcome codes are $\bar{g}_n$ and $\bar{q}_n$.

- ***Upper-guarantee separation along sequences.*** *For any sequence (p_n) of parameter records whose sample size at index n is n, with the same fixed experiment constants as p_0, the same radius sequences $\varepsilon_{1,\cdot}, \varepsilon_{2,\cdot}$, and $p_n.\gamma = \gamma$, suppose that eventually $n \geq 2$, $\mathsf{E}_n^{\mathrm{JMS}}$ holds, $\varepsilon_{1,n} > 0$, $\varepsilon_{2,n} \geq 0$,*

$$\varepsilon_{1,n} \leq \{4R_1 \exp(2C_g R_1)\}^{-1},$$

and the published ACE class contains a law using the clipped supplied codes. If

$$\frac{\varepsilon_{1,n}^r \varepsilon_{2,n} + C_\theta \varepsilon_{1,n}^{r+1}}{n^{-1/2}} \longrightarrow \infty,$$

then

$$\frac{\sqrt{C_{\mathrm{spec}}/(\gamma n)}}{B_{\mathrm{ACE},n,\gamma}} \longrightarrow 0.$$

Proposition 1 places the spectral and ACE guarantees on a shared statistical domain. The class relations use the fact that the ACE $L^r(P_X)$ code control implies the $L^1(P_X)$ treatment-code control needed by the spectral class, while the $L^s(P_X)$-restricted spectral subclass matches the published ACE code restrictions when $s = r$. Under JMS eligibility, the ACE bound retains its finite-order nuisance terms, and the contour bound is governed by the fixed-separation sampling term $(\gamma n)^{-1/2}$. Along the displayed nuisance-dominance sequences, the ratio of the contour upper guarantee to the imported finite-order ACE upper guarantee tends to zero. The conclusion is an upper-bound comparison on the common clipped-code class; the ACE-related lower-bound statement used in the paper is the class-minimax lower bound over all measurable estimators on the fixed-code class.

For later reference, the class relation used inside Proposition 1 is also recorded separately. This isolates the comparison-class bookkeeping from the estimator-specific upper bounds.

Lemma 5. [`lem:jms-ace-class-relations`] (ACE class relations) *Fix a measurable covariate space and a parameter record p with $r, k \in \mathbb{N}$, $k = r + 1$, $r \geq 2$, $s \in [0, \infty]$ satisfying $s \geq r$, $\gamma \in (1/2, 1)$, positive constants $C_\theta, C_g, C_q, \psi_\eta, \psi_\xi, \delta, \sigma$, and nonnegative nonincreasing code-radius sequences $\varepsilon_1, \varepsilon_2 : \mathbb{N} \to \mathbb{R}$ on positive indices. Then, for every class index n,*

- ***(JMS inclusion.)*** *For every model m, if $m \in \mathcal{P}_{\mathrm{ACE},n}^{\mathrm{JMS}}$ as defined in Definition 5, then $m \in \mathcal{P}_{\mathrm{NG},n}$ as defined in Definition 3.*

- ***(Comparison inclusion.)*** *For every model m, if $m \in \mathcal{P}_{\mathrm{NG},n}^{\mathrm{ACE}}$ as defined in Definition 11, then $m \in \mathcal{P}_{\mathrm{ACE},n}^{\mathrm{JMS}}$.*

- ***(Equal-exponent equivalence.)*** *If $s = r$, then for every model m,*

$$m \in \mathcal{P}_{\mathrm{ACE},n}^{\mathrm{JMS}} \quad \Longleftrightarrow \quad m \in \mathcal{P}_{\mathrm{NG},n}^{\mathrm{ACE}}.$$

The proof is deferred to Section E.

Lemma 5 is the set-inclusion component of the comparison. It identifies which restrictions are shared by the broad cumulant-separated class and the published ACE class, and it pins down the exact equality case under the common clipped-code convention. This makes the comparison

[1]**Formalization scope.** This result uses the published conclusion [Jin et al., 2025] (Theorem 5.4, equations (21)–(22)). The cited source proof is not formalized here: Lean verifies its use and all remaining steps. If that cited conclusion is false, the portions of this result that depend on it are not certified.

in Proposition 1 an upper-guarantee comparison on the same law class and the same clipped-code experiment.

The second benchmark is the bounded-outcome Gaussian comparison class of Definition 4. Proposition 2 gives a bounded-outcome Gaussian diagnostic for the intersection of the exact partially linear conditional mean, bounded observed outcome, and nondegenerate Gaussian treatment innovation. In every nonempty fixed-code intersection of that class, the decision-theoretic mean-squared and generalized-quantile risks are zero because the member restrictions determine the target value itself.

Proposition 2. [`prop:bounded-outcome-gaussian-degeneracy`] (Bounded Gaussian degeneracy) *Fix a measurable covariate space and a parameter record p with sample size $n \geq 1$, orders $r, k \in \mathbb{N}$ satisfying $k = r + 1$ and $r \geq 2$, exponent $s \geq r$, probability level $\gamma \in (1/2, 1)$, positive constants $C_\theta, C_g, C_q, \psi_\eta, \psi_\xi, \delta, \sigma$, and nonnegative nonincreasing code-radius sequences $\varepsilon_1, \varepsilon_2 : \mathbb{N} \to \mathbb{R}$ on positive indices. Then every model $P \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{G},n}$ satisfying Definition 4 has*

$$\theta_0(P) = 0.$$

Moreover, for any deterministic supplied code sequences $\widehat{g}, \widehat{q} : \mathbb{N} \to \mathcal{X} \to \mathbb{R}$, if there exists at least one $P \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{G},n}$ whose clipped codes at sample size n satisfy

$$\bar{g}_n(x) = \min\{\max\{\widehat{g}_n(x), -C_g\}, C_g\}, \qquad \bar{q}_n(x) = \min\{\max\{\widehat{q}_n(x), -C_q\}, C_q\}$$

for every x, then the fixed-code Gaussian minimax risks are both zero:

$$\mathfrak{R}^{\mathrm{G}}_n(p; \widehat{g}, \widehat{q}) = 0, \qquad \mathfrak{M}^{\mathrm{G}}_{n,1-\gamma}(p; \widehat{g}, \widehat{q}) = 0.$$

Proposition 2 gives a diagnostic for the source-stated Gaussian comparison class. A bounded observed outcome has bounded conditional support, while the partially linear term $\theta_0\eta$ is driven by nondegenerate Gaussian treatment noise under Definition 4; the proposition records the resulting target value and the induced zero minimax risks on each nonempty fixed-code intersection. Thus the Gaussian benchmark characterizes that simultaneous set of restrictions, whereas the fixed-separation comparison with ACE in Proposition 1 is the active finite-sample upper-guarantee comparison on the cumulant-separated class.

Together, Propositions 1 and 2 and Lemma 5 separate the two comparison roles. The ACE alignment places the published finite-order guarantee of Jin et al. [2025] and the contour guarantee on a common cumulant-separated domain, with the sequence clause identifying the regime in which the ratio of displayed upper guarantees tends to zero. The Gaussian proposition explains the bounded-outcome Gaussian class by its implied target and the corresponding decision-theoretic risks, matching the diagnostic role already summarized in Theorem 4.

7 Explicit mixture reductions and local benchmarks

The preceding comparison keeps the cumulant separation level fixed. This section records two constructive weak-non-Gaussian benchmarks that make the transform-zero mechanism explicit in one dimension. The first benchmark treats a symmetric two-component Gaussian mixture, where the relevant zero is known and the contour search collapses to a sine moment. The second follows a Gaussian–Rademacher path indexed by a mixture amplitude, so the denominator scale and the fourth cumulant can be read from the same parameter. These examples connect the contour construction to non-Gaussian identification ideas in independent-component and structural-shock settings [Comon, 1994, Hyvärinen and Oja, 1997, Hyvärinen et al., 2001, Lanne et al., 2017, Lee and Mesters, 2024, Hoesch et al., 2024, Reizinger et al., 2025].

The first construction is the clipped sine-ratio estimator $\widehat{\theta}_{\sin,n}$. It uses the known imaginary zero of the symmetric mixture transform and estimates the corresponding ratio directly from the empirical moments of $Y\sin(\pi Z_n/2)$ and $Z_n\sin(\pi Z_n/2)$.

Definition 14. [`def:sine-estimator`] (Sine-ratio estimator $\widehat{\theta}_{\sin,n}$) *Write* $\Pi_{[-C_\theta,C_\theta]}(u) = \min\{\max\{u,-C_\theta\},C_\theta\}$ *for the projection of a real number onto* $[-C_\theta,C_\theta]$. *For* $n\ge 1$ *observed triples* $O_i=(X_i,T_i,Y_i)$, *let* $\mathbb{P}_n f = n^{-1}\sum_{i=1}^n f(O_i)$ *and let* $Z_n = T-\bar g_n(X)$, *where* $\bar g_n$ *is the supplied treatment code clipped to* $[-C_g,C_g]$. *The clipped sine-ratio estimator is*

$$\widehat{\theta}_{\sin,n} = \begin{cases} \Pi_{[-C_\theta,C_\theta]}\left\{\dfrac{\mathbb{P}_n[Y\sin(\pi Z_n/2)]}{\mathbb{P}_n[Z_n\sin(\pi Z_n/2)]}\right\}, & \mathbb{P}_n[Z_n\sin(\pi Z_n/2)] \ge e^{-\pi^2/8}/4,\\ 0, & \mathbb{P}_n[Z_n\sin(\pi Z_n/2)] < e^{-\pi^2/8}/4. \end{cases}$$

The threshold in Definition 14 is tied to the population denominator scale that appears in the next proposition. In the symmetric mixture experiment, the treatment innovation has a known transform zero at $i\pi/2$, and $L^1(P_X)$ control of the supplied treatment code keeps the sine denominator bounded away from zero.

Proposition 3. [`prop:symmetric-mixture-reduction`] (Symmetric mixture reduction) *For any real constants* $C_\theta, C_g, C_q, \psi_\xi$, *there is a constant* $C>0$, *depending only on these four constants, such that the following holds. Let* p *be a parameter record whose constants* $C_\theta, C_g, C_q, \psi_\xi$ *are the ones just fixed, write* n *for its sample size, and let* P *be a law in the corresponding model. Assume the usual measurability and integrability conditions, and suppose that*

- ***(Sampling.)*** $O_1,\dots,O_n$ *are drawn according to the i.i.d. product law in Assumption 1.*
- ***(Symmetric mixture.)*** *The treatment innovation* $\eta = T - g_0(X)$ *has law*

  $$\tfrac12 N(-1,1) + \tfrac12 N(1,1).$$

- ***(Independence and ranges.)*** *The treatment innovation is independent of* X *as in Assumption 2, the outcome innovation satisfies Assumption 3, and the range and tail conditions in Assumptions 4 to 6 and 9 hold.*
- ***(Treatment code.)*** *The supplied treatment code obeys*

  $$\|\bar g_n - g_0\|_{L^1(P_X)} \le \varepsilon_{1,n}, \qquad \varepsilon_{1,n} \le \frac{1}{\pi}.$$

Then the fourth cumulant of η *is*

$$\kappa_4(\eta) = -2,$$

the treatment-innovation moment-generating function $M(z) = E_P[e^{z\eta}]$ *satisfies*

$$M(i\pi/2) = 0,$$

and, writing $Z_n = T-\bar g_n(X)$ *and* $D_n = g_0(X)-\bar g_n(X)$,

$$E_P\left[Z_n \sin\left(\frac{\pi Z_n}{2}\right)\right] = e^{-\pi^2/8} E_P\left[\cos\left(\frac{\pi D_n}{2}\right)\right],$$

with

$$\frac{e^{-\pi^2/8}}{2} \le E_P\left[Z_n \sin\left(\frac{\pi Z_n}{2}\right)\right].$$

Moreover, for the sine estimator $\widehat{\theta}_{\sin,n}$ *in Definition 14,*

$$E_P\left[\left(\widehat{\theta}_{\sin,n} - \theta_0(P)\right)^2\right] \le \frac{C}{n}.$$

Proposition 3 gives a fully explicit root-n benchmark for a symmetric non-Gaussian treatment innovation. The mixture has fourth cumulant -2, its transform vanishes at $i\pi/2$, and the sine denominator equals a positive contamination factor times $e^{-\pi^2/8}$. The resulting empirical ratio attains C/n mean-squared error under the stated range, tail, independence, and $L^1(P_X)$ treatment-code conditions.

The next elementary inequality is the algebraic device behind the sine-risk calculations. It separates numerator fluctuation from denominator fluctuation after clipping, using only a lower bound on the population denominator mean.

Lemma 6. [`lem:clipped-ratio-risk-decomposition`] (Clipped ratio risk decomposition) *Use the empirical-average notation* $\mathbb{P}_n$ *from Definition 1. Let* W *and* R *be real-valued functions of one observation, let* $\theta, C_\theta, A, \mu_W \in \mathbb{R}$, *and let* $O_1, \ldots, O_n$ *be any sample. Assume:*

- ***(Sample size.)*** $n \ge 1$.
- ***(Clipping range.)*** *The target value* θ *lies inside the clipping bound* C_θ*, in the sense that* $|\theta| \le C_\theta$.
- ***(Denominator scale.)*** *The scale* A *is strictly positive.*
- ***(Population denominator level.)*** *The denominator mean satisfies* $\mu_W \ge A/2$.

Define the clipped ratio statistic at threshold $A/4$ *by*

$$\widehat{\theta} = \begin{cases} \min\left\{\max\left\{\dfrac{\theta\,\mathbb{P}_n W + \mathbb{P}_n R}{\mathbb{P}_n W}, -C_\theta\right\}, C_\theta\right\}, & \text{if } A/4 \le \mathbb{P}_n W, \\ 0, & \text{if } \mathbb{P}_n W < A/4. \end{cases}$$

Then, pathwise on this sample,

$$(\widehat{\theta} - \theta)^2 \le \frac{16}{A^2}\left\{(\mathbb{P}_n R)^2 + C_\theta^2\left(\mathbb{P}_n(W - \mu_W)\right)^2\right\}.$$

The proof is deferred to Section E.

Lemma 6 is deterministic. In applications, W is the sine-weighted residualized treatment, R is the centered outcome-noise contribution to the sine numerator, and A is the population denominator scale. Once the sine moment gives $\mu_W \ge A/2$, the ratio error is controlled by two empirical averages.

The Gaussian–Rademacher path provides a local family approaching Gaussian treatment noise through an explicit fourth-cumulant parameter. Let η_a denote the centered Gaussian–Rademacher treatment innovation in the next statement. Its first positive transform zero, denominator scale, and cumulant magnitude are all explicit functions of a.

Lemma 7. [`lem:gaussian-rademacher-l1-benchmark`] (Gaussian–Rademacher sine benchmark) *Fix real constants $C_\theta, C_g, C_q, \psi_\xi$. There is a constant $C > 0$, depending only on these four constants, such that the following statement holds. Let p be an admissible parameter record whose constants $C_\theta, C_g, C_q, \psi_\xi$ are the ones just fixed and whose cumulant order is $k = 4$, let $a \in (0, 1]$, and let P be a law in the model m. Assume:*

- ***(Gaussian–Rademacher path.)*** *The treatment innovation $\eta = T - g_0(X)$ has the law of*
$$\eta_a = \sqrt{1 - a^2}\, G_\circ + aS,$$
where $G_\circ \sim N(0, 1)$, S is symmetric Rademacher, and $G_\circ$ and S are independent.
- ***(Sampling and model restrictions.)*** *The law satisfies Assumptions 1 to 6 and 9.*
- ***(Current treatment-code radius.)*** *For the clipped supplied treatment code $\bar{g}_n$,*
$$\int |\, \bar{g}_n(x) - g_0(x) \,|\, dP_X(x) \le \varepsilon_{1,n},$$
under the usual integrability condition for this $L^1(P_X)$ quantity.

Define
$$t_a = \frac{\pi}{2a}, \qquad A_a = a \exp\left\{ -\frac{(1 - a^2)\pi^2}{8a^2} \right\}, \qquad \Delta_a = 2a^4.$$
Then the treatment-innovation moment-generating function satisfies
$$M_a(z) = \exp\left(\frac{1 - a^2}{2} z^2 \right) \cosh(az),$$
the fourth cumulant satisfies $\kappa_4(\eta_a) = -2a^4$, and $t_a > 0$ is the first positive zero of the characteristic transform: $M_a(it_a) = 0$ and $M_a(iu) \ne 0$ for every $u \in (0, t_a)$. Moreover, for every $d \in \mathbb{R}$,
$$E_P[\sin\{t_a(\eta + d)\}] = 0.$$
Let $Z_n = T - \bar{g}_n(X)$ and $D_n = g_0(X) - \bar{g}_n(X)$. If $t_a \varepsilon_{1,n} \le 1/2$, then
$$E_P[Z_n \sin(t_a Z_n)] = A_a E_P[\cos(t_a D_n)]$$
and
$$E_P[Z_n \sin(t_a Z_n)] \ge A_a/2.$$
With $\mathbb{P}_n$ denoting empirical averaging over $O_1, \ldots, O_n$, define
$$\widehat{\theta}_{a,n} = \begin{cases} \Pi_{[-C_\theta, C_\theta]} \dfrac{\mathbb{P}_n[Y \sin(t_a Z_n)]}{\mathbb{P}_n[Z_n \sin(t_a Z_n)]}, & \mathbb{P}_n[Z_n \sin(t_a Z_n)] \ge A_a/4, \\ 0, & \textit{otherwise}, \end{cases}$$
where $\Pi_{[-C_\theta, C_\theta]}$ clips to $[-C_\theta, C_\theta]$. Its mean squared error obeys
$$E_P[(\widehat{\theta}_{a,n} - \theta_0)^2] \le C \min\left\{ 1, \frac{1}{n} a^{-2} \exp\left(\frac{(1 - a^2)\pi^2}{4a^2} \right) \right\}.$$

The same smallness condition also gives

$$\varepsilon_{1,n} \le \frac{\Delta_a^{1/4}}{2^{1/4}\pi}$$

and the equivalent cumulant-parameterized bound

$$E_P[(\widehat{\theta}_{a,n} - \theta_0)^2] \le C \min\left\{1, \frac{1}{n}\Delta_a^{-1/2} \exp\left(\frac{\pi^2\sqrt{2}}{4}\Delta_a^{-1/2}\right)\right\}.$$

The proof is deferred to Section E.

Lemma 7 shows how the explicit sine reduction behaves as the treatment-noise law moves toward Gaussianity. The first zero is $t_a = \pi/(2a)$, while the denominator scale A_a decays exponentially as a decreases. Expressing the same bound through $\Delta_a = 2a^4$ makes the dependence on fourth-cumulant magnitude explicit. The condition $t_a\varepsilon_{1,n} \le 1/2$ is exactly the treatment-code accuracy gate used by this pathwise calculation.

The section closes with a local benchmark statement. It combines the Gaussian–Rademacher path just described with a localized version of the published ACE oracle envelope from Jin et al. [2025]. The sequence $(\delta_n)_{n\in\mathbb{N}}$ supplies the shrinking cumulant threshold, and the ACE clause evaluates the published bound after replacing the fixed separation level by the local value $\Delta = \delta_n$.

Theorem 5. [`thm:local-to-gaussian-partial-benchmarks`] (Local Gaussian benchmarks) *Let $(\delta_n)_{n\in\mathbb{N}}$ be a real sequence with $\delta_n > 0$ for every n, nonincreasing in n, and $\delta_n \to 0$. Fix $C_\theta, C_g, C_q, \psi_\xi \in \mathbb{R}$. The following two assertions hold.*

- ***(Gaussian–Rademacher path.)*** *There is a constant $C > 0$, depending only on $C_\theta, C_g, C_q, \psi_\xi$, such that the following holds. Let p be any parameter record whose constants $C_\theta, C_g, C_q, \psi_\xi$ are the ones just fixed and whose orders are $k = 4$ and $r = 3$. For any $a \in (0, 1]$ and any law P in a model m, assume that the treatment noise satisfies*

$$\eta \sim \sqrt{1 - a^2}\, G_\circ + aS,$$

where $G_\circ \sim N(0, 1)$, S is symmetric Rademacher, and $G_\circ \perp S$. Assume also Assumptions 1 to 6 and 9. At the sample size n of p, assume the supplied clipped treatment code $\bar{g}_n$ satisfies

$$x \mapsto |\bar{g}_n(x) - g_0(x)| \in L^1(P_X), \qquad \int |\bar{g}_n(x) - g_0(x)|\, dP_X(x) \le \varepsilon_{1,n}.$$

Then the conclusion of Lemma 7 — the explicit mean-squared-error bound along the Gaussian–Rademacher path — holds with constant C for p, m, a.

- ***(Local ACE oracle envelope.)*** *For every published ACE handle, every $\gamma \in \mathbb{R}$, and every $C_\gamma \in \mathbb{R}$, if the handle satisfies the conclusion of Theorem 5.4 of Jin et al. [2025] at (γ, C_γ), then $C_\gamma > 0$ and the following local ACE statement holds. For every parameter record p whose probability level is γ, writing n for its sample size, if $\varepsilon_{1,n} > 0$ and $\varepsilon_{2,n} > 0$, then for every deterministic treatment code $\widehat{g}$, deterministic outcome code $\widehat{q}$, and every model, whenever*

$$\bar{g}_n = \Pi_{[-C_g,C_g]}\widehat{g}_n, \qquad \bar{q}_n = \Pi_{[-C_q,C_q]}\widehat{q}_n,$$

and, writing r and k for the orders of p and $\Delta = \delta_n$, the local ACE class conditions

$$1 \le n, \qquad |\kappa_k(\eta)| \ge \Delta,$$

hold together with Assumptions 2 to 6 and 9, the treatment-noise exponential envelope

$$o \mapsto \exp\{\eta(o)^2/\psi_\eta^2\} \in L^1(P), \qquad \int \exp\{\eta(o)^2/\psi_\eta^2\}\, dP(o) \le 2,$$

and the nuisance-radius bounds

$$\|\bar{g}_n - g_0\|_{L^r(P_X)} \le \varepsilon_{1,n}, \qquad \|\bar{q}_n - q_0\|_{L^r(P_X)} \le \varepsilon_{2,n},$$

and whenever the local JMS eligibility quantities

$$a_{1,n} = 2\log\Big(6(C_g + \psi_\eta)\varepsilon_{1,n}^{-1}\Big), \qquad b_{1,n} = \log(\gamma n/9),$$

$$a_2 = 4(C_g + \psi_\eta), \qquad b_{2,n}(\Delta) = \frac{200 \min\{1, C_\theta\}\Delta}{\max\{\varepsilon_{1,n}, \varepsilon_{2,n}, (\gamma n)^{-1/2}(\psi_\xi + C_\theta \psi_\eta)\}}$$

satisfy

$$a_{1,n} \ne 0, \qquad 0 < \frac{b_{1,n}}{a_{1,n}}, \qquad 0 < a_2 b_{2,n}(\Delta), \qquad a_2 \log\{a_2 b_{2,n}(\Delta)\} \ne 0,$$

and

$$r \le \min\left\{ \frac{b_{1,n}}{a_{1,n}} - a_{1,n}^{-1}\log\left(\frac{b_{1,n}}{a_{1,n}}\right), \frac{b_{2,n}(\Delta)}{a_2 \log\{a_2 b_{2,n}(\Delta)\}} \right\},$$

the published order-r ACE estimator obeys

$$Q_{P,1-\gamma}\Big(\Big|\widehat{\theta}_{\mathrm{ACE},r,n} - \theta_0(P)\Big|\Big) \le B_{\mathrm{ACE},n,\gamma}(\delta_n; C_\gamma),$$

where

$$B_{\mathrm{ACE},n,\gamma}(\Delta; C_\gamma) = C_\gamma\, r!\, 16^r\, \Delta^{-1} \left(\varepsilon_{1,n}^r \varepsilon_{2,n} + C_\theta \varepsilon_{1,n}^{r+1} + 64(C_g + \psi_\eta)^r \big(r^2(C_g + \psi_\eta) + \psi_\xi + C_\theta\psi_\eta\big)(\gamma n)^{-1/2} \right).$$

Theorem 5 gives two procedure-specific local upper benchmarks. The Gaussian–Rademacher clause reuses Lemma 7 for the $k = 4, r = 3$ path under the stated $L^1(P_X)$ treatment-code condition. The ACE clause specializes the published order-r generalized-quantile guarantee of Jin et al. [2025] to a local cumulant threshold $\Delta = \delta_n$, with the displayed eligibility quantities and the resulting envelope $B_{\mathrm{ACE},n,\gamma}(\delta_n; C_\gamma)$. The last algebraic clause records how an oracle upper bound that is valid under two sets of hypotheses can be summarized by the smaller of the two displayed upper inputs.

These local statements complement the fixed-separation results by making the dependence on the non-Gaussian signal explicit in tractable submodels. For the symmetric mixture, the sine denominator is bounded by a fixed positive scale and the risk is C/n. Along the Gaussian–Rademacher path, the same sine construction has a denominator scale A_a and a cumulant magnitude Δ_a, yielding the explicit exponential dependence displayed in Lemma 7. The localized ACE envelope gives the corresponding published finite-order benchmark under its stated eligibility and nuisance-radius conditions.

8 Limitations and future work

The fixed-separation results in Theorems 3 and 4 use a cumulant-separation constant that is fixed across the experiment. The local benchmarks in Theorem 5 point to a complementary triangular-array program in which the non-Gaussian signal is allowed to shrink with the sample size. This section records that agenda separately from the fixed-δ contour theorem and from the procedure-specific local upper bounds of the preceding section.

A useful starting point for that program is a finite, sample-facing object that stores the contour information available at a given library depth. The following definition names the empirical handle used to organize candidate contours, empirical winding information, contour moments, and angular variability.

Definition 15. [`def:local-gaussian-handle`] (Local Gaussian handle $\mathscr{C}_m$) *Fix a parameter tuple with positive sample size n, fixed ACE order $r \geq 2$, cumulant order $k = r+1$, $s \geq r$, $\gamma \in (1/2, 1)$, strictly positive $C_\theta, C_g, C_q, \psi_\eta, \psi_\xi, \delta, \sigma$, and nonnegative nonincreasing sequences ϵ_{1n} and ϵ_{2n} on positive indices. Work on a measurable covariate space with observations $O = (X, T, Y)$, and fix a probability law P with real functional $\theta_0(P)$, measurable conditional means $g_0(X) = E_P[T \mid X]$ and $q_0(X) = E_P[Y \mid X]$, integrable T and Y, and deterministic measurable supplied regression-code sequences for treatment and outcome.*

For each library depth d, let $\mathscr{C}_d$ be the deterministic finite library of rational circles represented by the positive rational radii $\rho = j+1$, $0 \leq j < d$. For a realized sample indexed by $\{0, \ldots, n-1\}$, form the empirical transforms $\widehat{F}_{a,n}$ and $\widehat{G}_{a,n}$ on the prespecified deterministic fold $a \in \{0, 1\}$, using indices below $n/2$ for $a = 0$ and indices at least $n/2$ for $a = 1$. For a circle $C \in \mathscr{C}_d$, $C = \{\rho e^{it} : 0 \leq t \leq 2\pi\}$, on which the empirical boundary modulus is certified positive, the candidate record consists of quantities definitionally computed from the empirical contour integrands

$$W_{a,C}(t) = \rho e^{it} \frac{\widehat{F}'_{a,n}(\rho e^{it})}{\widehat{F}_{a,n}(\rho e^{it})}, \qquad V_{a,C}(t) = \rho e^{it} \frac{\widehat{G}_{a,n}(\rho e^{it})}{\widehat{F}_{a,n}(\rho e^{it})}.$$

Specifically, the record contains a rational interval enclosure of the boundary modulus $\inf_{z \in C} |\widehat{F}_{a,n}(z)|$, a complex rational interval enclosure for $(2\pi)^{-1} \int_0^{2\pi} W_{a,C}(t)\, dt$, a complex rational interval enclosure for $(2\pi)^{-1} \int_0^{2\pi} V_{a,C}(t)\, dt$, and a rational interval enclosure of the angular variance

$$\widehat{v}_{a,C} = \frac{1}{2\pi} \int_0^{2\pi} \left| V_{a,C}(t) - \frac{1}{2\pi} \int_0^{2\pi} V_{a,C}(u)\, du \right|^2 dt.$$

Each enclosure is accompanied by its soundness certificate for the displayed empirical quantity, and the actual empirical transforms on the stated deterministic fold are the transform inputs of the schema. These records are candidate data for one realized sample, one deterministic inference fold, and one rational circle.

Definition 15 provides a concrete data structure for studying adaptive contour selection in local regimes. It keeps the deterministic library, fold-specific empirical transforms, certified boundary separation, contour moments, and angular variance in one record. For shrinking cumulant separation, such records give a way to describe what a selector observes before choosing among contour-based procedures and more classical alternatives.

The open statistical question is naturally triangular. Let (δ_n) denote shrinking positive cumulant-separation thresholds. The fixed-separation theorem controls risk when the primitive separation constant is fixed, while the local benchmarks in Theorem 5 describe particular upper envelopes when the non-Gaussian signal is weak. The remaining problem is to identify the sharp joint rate and the corresponding inference theory across the ordinary DML, finite-order ACE, and contour regimes [Chernozhukov et al., 2018, Newey and Robins, 2018, Mackey et al., 2018, Jin et al., 2025, Lee and Mesters, 2024, Hoesch et al., 2024].

Remark 1. [`oeq:local-to-gaussian-frontier`] (Local-to-Gaussian frontier) *A natural next question concerns triangular treatment-noise laws satisfying*

$$|\kappa_k(\eta_n)| \geq \delta_n \qquad \textit{with} \qquad \delta_n \downarrow 0$$

under the same supplied code sequences. The target is a data-driven selector among ordinary DML, finite-order ACE, and global-contour procedures that attains the sharp minimax mean-squared-error rate as a function of

$$(n, \varepsilon_{1,n}, \varepsilon_{2,n}, \delta_n),$$

supports uniformly valid inference across these regimes, and admits a matching local minimax lower bound. Future work could determine the sharp-rate functional, the corresponding uniform-inference criterion, and the local lower-bound construction for this frontier.

Remark 1 states the future-work target as an open research agenda rather than as an identification or risk theorem. Its rate arguments are the sample size, the two supplied-code radii, and the local cumulant-separation threshold. The agenda also includes a uniform-inference criterion and a local minimax lower-bound construction, so any completed theory would need to align estimation, coverage, selection, and lower-bound witnesses on the same triangular experiment.

This separation clarifies the scope of the paper's established results. The fixed-separation contour theorem gives the $\Theta(n^{-1})$ MSE characterization under fixed cumulant separation and the stated treatment-code stability condition. The local mixture and ACE benchmarks give constructive upper comparisons under their own displayed hypotheses. The shrinking-separation program in Remark 1 is the natural next step for a unified adaptive theory across Gaussian-adjacent and fixed non-Gaussian regimes.

Appendices

A Proofs for identification and analytic localization

This appendix collects the analytic statements that support the identification argument in Theorems 1 and 2 and Lemma 1 and the fixed-bank localization used in Theorem 3. The common theme is that zeros of moment-generating functions can be controlled by classical complex-analytic arguments and then transferred to residualized-treatment transforms. The zero and completeness perspective is standard in identification arguments based on transforms [Kagan et al., 1973, Mattner, 1992], and the finite-bank construction uses the same analytic geometry that underlies modern zero-localization bounds for moment-generating functions [Michelen and Sahasrabudhe, 2019, Eremenko and Fryntov, 2021, Dinh et al., 2021].

For Theorem 1, the analytic order of a zero of M determines which polynomial-exponential weight annihilates shifted treatment innovations. Independence of the treatment innovation from X, as imposed in Assumption 2, then turns the shifted identity into a conditional moment for the code residual Z_n. The denominator calculation is the corresponding differentiated transform identity, and the partially linear conditional-mean restriction in Assumption 3 supplies the numerator relation.

The factorization in Lemma 1 isolates the effect of treatment-code contamination. Since $Z_n = \eta + D_n(X)$, independence separates the residual transform into the treatment-innovation transform and the nuisance transform. The outcome-weighted transform has the derivative term carrying θ_0 plus the covariate-only contamination transform from Definition 7. This decomposition is the algebraic input for Theorem 2: on a contour where the residual transform is separated from zero and the nuisance transform has no zero on the closed disk, the logarithmic-derivative contribution counts the enclosed zeros and the analytic contamination contribution integrates to zero.

The remaining analytic step is uniform localization. Lemma 3 gives a disk, determined by the fixed cumulant order, sub-Gaussian scale, and separation constant, that contains at least one zero of the treatment-innovation transform. The finite-bank statement below turns that disk-level information into a deterministic list of circles, with explicit dyadic bookkeeping and a positive boundary-modulus certificate. Its role is to provide the law-blind contour geometry used later by the empirical statistic.

Lemma 8. [`lem:finite-contour-bank`] (Finite contour bank) *Let p be a primitive parameter record with $k = r+1$, $r \geq 2$, $\gamma \in (1/2, 1)$, $s \geq r$, positive constants $C_\theta, C_g, C_q, \psi_\eta, \psi_\xi, \delta, \sigma$, and nonnegative nonincreasing code-radius sequences $\varepsilon_1, \varepsilon_2$. Let $\mathfrak{p}_\star$ be a certified-bank input record for p in the sense of Definition 17, so that*

$$R_0 = A_k \left(\frac{\psi_\eta^2}{\delta}\right)^{(k-2)^{-1}}, \qquad R_1 = R_0 + 1.$$

For any sample size $n \geq 1$, any model m with observed-data law P, and any membership certificate $m \in \mathcal{P}_{\mathrm{NG},n}$ as defined in Definition 3, form the bank data B of Definition 17 for p and $\mathfrak{p}_\star$. Then $a_\star > 0$, and the components of B satisfy

$$N_{\mathrm{cert}} = 4U_\psi^2(U_R+2)^3, \qquad m_\star = N_{\mathrm{cert}} + 3, \qquad J = 2^{m_\star - 1} + 1,$$

$$d_\star = \left(\frac{1}{2}\right)^{m_\star}, \qquad h_\star = \frac{d_\star}{3}, \qquad e_\star = 8U_R U_\psi^2 (U_R+1)^2,$$

$$d_\star^{\mathrm{den}} = 3(2U_R+1), \qquad u_\star = 2e_\star + N_{\mathrm{cert}}\big(m_\star + d_\star^{\mathrm{den}}\big), \qquad a_{\star,\mathbb{Q}} = \left(\frac{1}{2}\right)^{u_\star}.$$

The bank radii are strictly increasing and lie in the search annulus:

$$R_0 < \rho_j < R_1 \quad \textit{for every } j \in \{1, \ldots, J\}.$$

Writing $C_j = \{z \in \mathbb{C} : |z| = \rho_j\}$ and $M(z) = E_P[e^{z\eta}]$ for the treatment-innovation moment-generating function, there exists an index $j \in \{1, \ldots, J\}$ such that

$$N_{C_j}(M) \geq 1 \quad \textit{and} \quad a_\star \leq |M(z)| \textit{ for every } z \in C_j.$$

Proof of Lemma 8. Let B be the bank produced from $(p, \mathfrak{p}_\star)$. The construction in Definition 17 gives

$$N_{\mathrm{cert}} = 4U_\psi^2(U_R+2)^3, \qquad m_\star = N_{\mathrm{cert}} + 3, \qquad J = 2^{m_\star - 1} + 1,$$

$$d_\star = 2^{-m_\star}, \qquad h_\star = d_\star/3,$$
$$e_\star = 8U_R U_\psi^2 (U_R+1)^2, \qquad d_\star^{\mathrm{den}} = 3(2U_R+1),$$

and

$$u_\star = 2e_\star + N_{\mathrm{cert}}(m_\star + d_\star^{\mathrm{den}}), \qquad a_\star = 2^{-u_\star}.$$

Thus $a_\star > 0$ and the displayed bookkeeping identities hold.

The radii have the explicit form

$$\rho_j = R_0 + \frac{1}{4} + j\,2^{-m_\star}, \qquad 0 \le j \le 2^{m_\star - 1}.$$

Hence $j \mapsto \rho_j$ is strictly increasing. Since

$$\frac{1}{4} + j2^{-m_\star} \le \frac{1}{4} + 2^{m_\star-1}2^{-m_\star} = \frac{3}{4} < 1,$$

and $R_1 = R_0 + 1$, each bank radius satisfies

$$R_0 < \rho_j < R_1.$$

It remains to prove the treatment-noise certificate. The error-one certified refinements and the ceiling construction give

$$\psi_\eta \le U_\psi, \qquad R_1 \le U_R.$$

Let

$$M(z) = E_P e^{z\eta}, \qquad R_{\mathrm{fac}} = R_1 + 1, \qquad R_{\mathrm{out}} = R_1 + 2.$$

The conditional-mean identity for g_0 centers η. The treatment-tail condition in $\mathcal{P}_{\mathrm{NG},n}$, together with Lemma 2, gives for every $t \in \mathbb{R}$

$$E_P e^{t\eta} \le \exp(4\psi_\eta^2 t^2).$$

The same real exponential integrability gives the entire complex moment-generating function, normalization at the origin follows from the probability law, and the complex modulus estimate yields

$$M(0) = 1, \qquad M \text{ is entire}, \qquad |M(z)| \le \exp(4\psi_\eta^2 |z|^2) \quad (z \in \mathbb{C}).$$

Let

$$N_0 = N_{\{|z|<R_{\mathrm{fac}}\}}(M)$$

be the number of zeros of M in the open disk of radius R_{fac}, counted with multiplicity. Jensen's zero-counting bound applied on R_{out} gives

$$N_0 \log\left(\frac{R_{\mathrm{out}}}{R_{\mathrm{fac}}}\right) \le 4\psi_\eta^2 R_{\mathrm{out}}^2.$$

Since $R_{\mathrm{out}} = R_{\mathrm{fac}} + 1$,

$$\frac{1}{R_{\mathrm{out}}} \le \log\left(\frac{R_{\mathrm{out}}}{R_{\mathrm{fac}}}\right).$$

Therefore

$$N_0 \le 4\psi_\eta^2 R_{\mathrm{out}}^3 \le 4U_\psi^2 (U_R+2)^3 = N_{\mathrm{cert}}.$$

List the N_0 zeros in $\{|z| < R_{\text{fac}}\}$, with multiplicity, as

$$a_1, \dots, a_{N_0}.$$

The finite Blaschke factorization on the disk of radius R_{fac} gives a holomorphic function g, zero-free in $\{|z| < R_{\text{fac}}\}$, such that

$$M(z) = Q(z)g(z), \qquad |z| \le R_{\text{fac}},$$

where

$$Q(z) = \prod_{\nu=1}^{N_0} \frac{R_{\text{fac}}(z - a_\nu)}{R_{\text{fac}}^2 - \overline{a_\nu} z}.$$

The grid spacing is $d_\star$, and the mesh identities give $d_\star > 0$, $h_\star = d_\star/3$, and $2h_\star < d_\star$. Thus one listed zero modulus $|a_\nu|$ can be within distance $< h_\star$ of at most one grid radius: two distinct grid radii are separated by at least $d_\star$, while two distances $< h_\star$ would put them less than $2h_\star$ apart. Also

$$N_0 \le N_{\text{cert}} < 2^{N_{\text{cert}}+2} + 1 = 2^{m_\star - 1} + 1 = J.$$

The pigeonhole argument for this separated dyadic grid gives an index j such that

$$h_\star \le \big||a_\nu| - \rho_j\big| \qquad (1 \le \nu \le N_0).$$

By Lemma 3, M has a zero z_0 with $|z_0| \le R_0$. Since $R_0 < \rho_j$, this zero lies strictly inside $C_j = \{|z| = \rho_j\}$. If M had a zero z on C_j, then $z \in \{|z| < R_{\text{fac}}\}$, and the factorization $M = Qg$ together with the zero-freeness of g would force $z = a_\nu$ for some ν. This would give

$$\big||a_\nu| - \rho_j\big| = 0 < h_\star,$$

contradicting the separation of the chosen grid point. Hence the contour has positive zero count:

$$N_{C_j}(M) \ge 1.$$

It remains to lower-bound $|M|$ on this circle. Put

$$A = 4\psi_\eta^2 R_{\text{fac}}^2.$$

On $|w| = R_{\text{fac}}$, the Blaschke product has $|Q(w)| = 1$, so the boundary envelope for M gives $|g(w)| \le e^A$. Also $M(0) = 1$ and $|Q(0)| \le 1$, hence $|g(0)| \ge 1$. Applying Harnack's inequality to the positive harmonic function $A - \log|g|$ gives, for every $|z| \le R_1$,

$$\exp\left\{-\left(\frac{R_{\text{fac}} + R_1}{R_{\text{fac}} - R_1} - 1\right) A\right\} |Q(z)| \le |M(z)|.$$

Since $R_{\text{fac}} = R_1 + 1$,

$$\left(\frac{R_{\text{fac}} + R_1}{R_{\text{fac}} - R_1} - 1\right) A = 2R_1 \cdot 4\psi_\eta^2 (R_1 + 1)^2 = 8\psi_\eta^2 R_1 (R_1 + 1)^2.$$

Thus, for every $|z| \le R_1$,

$$\exp\{-8\psi_\eta^2 R_1 (R_1 + 1)^2\}\, |Q(z)| \le |M(z)|.$$

Since

$$8\psi_\eta^2 R_1(R_1+1)^2 \le 8U_R U_\psi^2 (U_R+1)^2 = e_\star,$$

the elementary dyadic-exponential comparison $2^{-2E} \le e^{-E}$ for natural E, applied to $E = e_\star$, gives

$$2^{-2e_\star} \le \exp\{-8\psi_\eta^2 R_1(R_1+1)^2\}.$$

Now take $z \in C_j$. For each zero a_ν,

$$|z-a_\nu| \ge \big||z|-|a_\nu|\big| = \big|\rho_j - |a_\nu|\big| \ge h_\star.$$

Also $|z| \le R_1$, and therefore

$$R_{\mathrm{fac}} + |z| \le 2R_1 + 1 \le 2U_R + 1.$$

For each Blaschke factor,

$$\left|\frac{R_{\mathrm{fac}}(z-a_\nu)}{R_{\mathrm{fac}}^2 - \overline{a_\nu} z}\right| \ge \frac{|z-a_\nu|}{R_{\mathrm{fac}}+|z|} \ge \frac{h_\star}{2U_R+1} = \frac{d_\star}{d_\star^{\mathrm{den}}}.$$

Multiplying over the at most N_{cert} factors and using $(d_\star^{\mathrm{den}})^{-1} \ge 2^{-d_\star^{\mathrm{den}}}$ gives

$$|Q(z)| \ge 2^{-N_{\mathrm{cert}}(m_\star + d_\star^{\mathrm{den}})}.$$

Combining the last two lower bounds,

$$|M(z)| \ge 2^{-2e_\star}\, 2^{-N_{\mathrm{cert}}(m_\star + d_\star^{\mathrm{den}})} = 2^{-u_\star} = a_\star.$$

Thus for this bank index j,

$$a_\star \le |M(z)| \qquad (z \in C_j),$$

and the asserted uniform treatment-noise certificate follows. □

Lemma 8 is the population geometric certificate used by the empirical selector in Algorithm 1. The proof below records the dyadic bookkeeping and the zero-localization argument behind that certificate.

B Empirical process, stability, and lower-bound lemmas

This appendix collects the auxiliary statistical statements used by the fixed-separation contour theorem. The first two results control the analytic transforms that enter the contour statistic: one gives uniform L^2 concentration of the empirical residual and outcome transforms on a fixed disk, and the other gives a deterministic $L^1(P_X)$ stability bound for the nuisance factor. Together they supply the stochastic and population inputs behind the contour-risk bounds in Theorem 3. The concentration statement uses the same sub-Gaussian scale conventions and product-sample structure as Definition 1 and Assumption 1, in line with standard empirical-process and high-dimensional probability arguments [van der Vaart, 1998, van der Vaart and Wellner, 1996, Vershynin, 2018, Wainwright, 2019].

Lemma 9. [`lem:empirical-transform-uniform-l2`] (Uniform transform variance) *Fix real numbers $C_\theta, C_g, C_q, \psi_\eta, \psi_\xi$ and R_1. There is a constant $K > 0$ such that the following holds. Let p be a parameter record with*

$$p.C_\theta = C_\theta, \qquad p.C_g = C_g, \qquad p.C_q = C_q, \qquad p.\psi_\eta = \psi_\eta, \qquad p.\psi_\xi = \psi_\xi, \qquad R_1 = \mathrm{searchRadius}(p),$$

and sample size $n = p.n \geq 2$. Let m be a model in $\mathcal{P}_{\mathrm{NG},n}$, as in Definition 3, and let $O_1, \ldots, O_n$ be sampled under the i.i.d. product law $P^{\otimes n}$, as in Assumption 1. For each $a \in \{0,1\}$, write I_a for the corresponding deterministic inference fold and suppose $|I_a| \geq 1$. Define, for $|z| \leq R_1$,

$$\widehat{F}_{a,n}(z) = |I_a|^{-1} \sum_{i \in I_a} \exp(zZ_{n,i}), \qquad \widehat{G}_{a,n}(z) = |I_a|^{-1} \sum_{i \in I_a} Y_i \exp(zZ_{n,i}),$$

where $Z_{n,i} = T_i - \bar{g}_n(X_i)$, and let

$$F_n(z) = E_P[\exp(zZ_n)], \qquad G_n(z) = E_P[Y \exp(zZ_n)].$$

Then

$$E_{P^{\otimes n}}\left[\sup_{|z| \leq R_1} |\widehat{F}_{a,n}(z) - F_n(z)|^2\right] + E_{P^{\otimes n}}\left[\sup_{|z| \leq R_1} |\widehat{G}_{a,n}(z) - G_n(z)|^2\right] \leq \frac{K}{|I_a|}.$$

Proof of Lemma 9. Set

$$A_F = 2\exp\{8R_1C_g + 4R_1^2\psi_\eta^2\},$$

$$A_G = 64\{C_q^4 + 4C_\theta^4\psi_\eta^4 + 4\psi_\xi^4\} + 2\exp\{16R_1C_g + 16R_1^2\psi_\eta^2\},$$

and define

$$K = 4(A_F + A_G).$$

The constants A_F and A_G are strictly positive, so $K > 0$; it depends only on the displayed primitive constants and R_1.

Fix a fold for which I_a is nonempty, and let $I = I_a$, $m_I = |I|$, and $Z_n = T - \bar{g}_n(X)$. Since the contour-bank radius in Definition 17 satisfies $R_1 = R_0 + 1$ with $R_0 \geq 0$, we have $R_1 > 0$. Write $D_n(X) = g_0(X) - \bar{g}_n(X)$, so $Z_n = \eta + D_n(X)$. The range bound for g_0 in Definition 3 and the clipping of $\bar{g}_n$ in Definition 2 give $|D_n(X)| \leq 2C_g$ almost surely. Hence, for almost every observation,

$$4R_1|Z_n| \leq 4R_1|\eta| + 8R_1C_g \leq \frac{\eta^2}{\psi_\eta^2} + 4R_1^2\psi_\eta^2 + 8R_1C_g,$$

where the second inequality is the Young inequality obtained by expanding $(|\eta|/\psi_\eta - 2R_1\psi_\eta)^2 \geq 0$. Therefore

$$\{\exp(2R_1|Z_n|)\}^2 \leq \exp\{8R_1C_g + 4R_1^2\psi_\eta^2\}\exp(\eta^2/\psi_\eta^2).$$

Integrating and using the Luxemburg treatment-tail bound from Definition 3 yields

$$E_P\left[\exp\{2R_1|Z_n|\}^2\right] \leq A_F.$$

The outcome-weighted envelope is

$$E_P\left[\{|Y|\exp(2R_1|Z_n|)\}^2\right] \leq A_G.$$

Indeed, Definition 2 gives $Y = q_0(X) + \theta_0\eta + \xi$. For almost every observation, the algebraic bound

$$|q_0(X) + \theta_0\eta + \xi|^4 \le 64\{|q_0(X)|^4 + |\theta_0|^4|\eta|^4 + |\xi|^4\}$$

combines with the bounds in Definition 3, namely $|q_0(X)| \le C_q$ almost surely and $|\theta_0| \le C_\theta$, to give

$$|Y|^4 \le 64\{C_q^4 + C_\theta^4|\eta|^4 + |\xi|^4\}$$

almost surely. The Luxemburg moment bounds give $E_P|\eta|^4 \le 4\psi_\eta^4$ and $E_P|\xi|^4 \le 4\psi_\xi^4$, and hence

$$E_P|Y|^4 \le 64\{C_q^4 + 4C_\theta^4\psi_\eta^4 + 4\psi_\xi^4\}.$$

Together with

$$\{|Y|\exp(2R_1|Z_n|)\}^2 \le |Y|^4 + \{\exp(4R_1|Z_n|)\}^2,$$

and the preceding residual envelope applied at radius $2R_1$, this gives the stated bound.

We use the following coefficient-series bound. Let $J \subset \{1,\dots,n\}$ be a nonempty deterministic finite subset of sample indices, and write $m_J = |J|$. Let U, V be measurable real functions, let $R > 0$, let $C \ge 0$, and suppose

$$E_P\left[\{|U|\exp(2R|V|)\}^2\right] \le C^2.$$

Define

$$c_{J,k}(O_1,\dots,O_n) = m_J^{-1}\sum_{i\in J}\left\{\frac{U(O_i)V(O_i)^k}{k!} - E_P\frac{UV^k}{k!}\right\}.$$

Then, under the i.i.d. product law for $O_1,\dots,O_n$,

$$E\left[\sup_{|z|\le R}\left|\sum_{k=0}^{\infty} c_{J,k}z^k\right|^2\right] \le \left(\frac{2C}{\sqrt{m_J}}\right)^2.$$

The envelope gives

$$\left\|\frac{UV^k}{k!}\right\|_{L^2(P)} \le \frac{C}{(2R)^k},$$

because $x^k/k! \le e^{2Rx}/(2R)^k$ for $x \ge 0$; the hypotheses $R > 0$ and $C \ge 0$ make the displayed majorant nonnegative. Under the product sample law, centering over the block J gives

$$\|c_{J,k}\|_{L^2(P^{\otimes n})} \le \frac{C}{(2R)^k\sqrt{m_J}}.$$

The series of majorants is summable, and the centered analytic-series supremum bound yields

$$\left(\sum_{k=0}^{\infty}\frac{C}{(2R)^k\sqrt{m_J}}R^k\right)^2 = \left(\frac{2C}{\sqrt{m_J}}\right)^2.$$

The same coefficient construction represents the empirical transform errors exactly. For the same block J and for $|z| \le R$,

$$m_J^{-1}\sum_{i\in J}U(O_i)e^{zV(O_i)} - E_P[Ue^{zV}] = \sum_{k=0}^{\infty} c_{J,k}z^k,$$

with the exchange of summation, finite averaging, and integration justified by the doubled exponential envelope.

Apply the preceding bound first with the block $J = I_a$, $U \equiv 1$, $V = Z_n$, $R = R_1$, and $C = \sqrt{A_F}$. Since $A_F > 0$, this choice has $C \geq 0$, and

$$E\left[\sup_{|z|\leq R_1} |\widehat{F}_{a,n}(z) - F_n(z)|^2\right] \leq \frac{4A_F}{m_I}.$$

Apply it again with the block $J = I_a$, $U = Y$, $V = Z_n$, $R = R_1$, and $C = \sqrt{A_G}$. Since $A_G > 0$, this choice has $C \geq 0$, and

$$E\left[\sup_{|z|\leq R_1} |\widehat{G}_{a,n}(z) - G_n(z)|^2\right] \leq \frac{4A_G}{m_I}.$$

Adding the two displays yields

$$E\left[\sup_{|z|\leq R_1} |\widehat{F}_{a,n}(z) - F_n(z)|^2\right] + E\left[\sup_{|z|\leq R_1} |\widehat{G}_{a,n}(z) - G_n(z)|^2\right] \leq \frac{4(A_F + A_G)}{|I_a|} = \frac{K}{|I_a|}.$$

This proves the assertion. □

Lemma 9 is the stochastic input for the selected-contour risk proof. The proof below supplies the disk-uniform L^2 bound used after the pilot fold selects a circle.

The next statement is the population stability input. It uses the direct $L^1(P_X)$ treatment-code radius in Definition 3 to keep the nuisance transform close to one on the contour search region.

Lemma 10. [`lem:l1-nuisance-zero-free`] (Zero-free nuisance factor) *For every $P \in \mathcal{P}_{\mathrm{NG},n}$ satisfying Definition 3, the associated treatment-code discrepancy D_n obeys*

$$|D_n| \leq 2C_g \quad \textit{almost surely}, \qquad E_P|D_n| \leq \varepsilon_{1,n}.$$

Consequently, for the corresponding factor H_n and radius R_1,

$$\sup_{|z|\leq R_1} |H_n(z) - 1| \leq R_1 e^{2C_g R_1} \varepsilon_{1,n}.$$

If $\varepsilon_{1,n} \leq (4R_1 e^{2C_g R_1})^{-1}$, then $|H_n(z)| \geq 3/4$ for every $|z| \leq R_1$, and F_n and M have the same zeros in $\{|z| \leq R_1\}$, counted with multiplicity.

Proof of Lemma 10. Let

$$D_n(X) = g_0(X) - \bar{g}_n(X), \qquad H_n(z) = E_P[e^{zD_n(X)}].$$

Since $|g_0(X)| \leq C_g$ almost surely and $|\bar{g}_n(X)| \leq C_g$ by clipping,

$$|D_n(X)| \leq |g_0(X)| + |\bar{g}_n(X)| \leq 2C_g \quad \text{almost surely.}$$

The treatment-code condition in $\mathcal{P}_{\mathrm{NG},n}$ gives

$$E_P|D_n(X)| = \int |g_0(x) - \bar{g}_n(x)|\, dP_X(x) \leq \varepsilon_{1,n}.$$

For $|z| \le R_1$, the exponential remainder bound gives, with $w = zD_n(X)$,

$$|e^{zD_n(X)} - 1| \le |z|\,|D_n(X)|\,e^{|z|\,|D_n(X)|}.$$

Together with $|D_n(X)| \le 2C_g$, this yields

$$|H_n(z) - 1| \le E_P|e^{zD_n(X)} - 1| \le R_1 e^{2C_g R_1} E_P|D_n(X)| \le R_1 e^{2C_g R_1} \varepsilon_{1,n}.$$

Taking the supremum over $|z| \le R_1$ gives the displayed uniform bound.

Assume

$$\varepsilon_{1,n} \le (4R_1 e^{2C_g R_1})^{-1}.$$

Then the preceding display gives $|H_n(z) - 1| \le 1/4$ throughout the disk, and therefore

$$|H_n(z)| \ge 1 - |H_n(z) - 1| \ge 3/4 \qquad (|z| \le R_1).$$

It remains to record why the multiplicity comparison is legitimate at each point of the disk. The almost-sure bound $|D_n(X)| \le 2C_g$ implies that $E_P \exp\{tD_n(X)\} < \infty$ for every real t, so the complex MGF H_n is analytic in a neighborhood of every $|z| \le R_1$. The treatment-tail condition in $\mathcal{P}_{\mathrm{NG},n}$, under the Luxemburg convention of Definition 1, gives the same local exponential-moment condition for η: for every real t, completing the square bounds $\exp\{t\eta\}$ by a constant multiple of $\exp\{\eta^2/\psi_\eta^2\}$, which is integrable. Hence the treatment-noise MGF $M(z) = E_P[e^{z\eta}]$ is analytic in a neighborhood of every point of the disk as well.

By Lemma 1,

$$F_n(z) = M(z)H_n(z).$$

Fix $|z| \le R_1$. The preceding lower bound gives $H_n(z) \neq 0$, so the analytic order of H_n at z is zero. Applying additivity of analytic order to the displayed factorization gives

$$\mathrm{ord}_z(F_n) = \mathrm{ord}_z(M) + \mathrm{ord}_z(H_n) = \mathrm{ord}_z(M).$$

Thus F_n and M have exactly the same zeros in $\{|z| \le R_1\}$, counted with multiplicity. □

Lemma 10 is the deterministic transfer from treatment-transform zeros to observable residual-transform zeros. The proof below uses the clipping bound and the $L^1(P_X)$ radius to control the nuisance factor H_n on the search disk.

A separate tail calculation is used when Gaussian outcome innovations are embedded in lower-bound and benchmark paths. The following elementary scale bound matches the Luxemburg convention fixed in Definition 1.

Lemma 11. [`lem:gaussian-exponential-square-quarter-scale`] (Gaussian square exponential bound) *Use the Gaussian-law notation of Definition 1. For every positive scale $\psi > 0$, let $\zeta \sim N(0, (\psi/4)^2)$. Then $\exp(\zeta^2/\psi^2)$ is integrable under the law of ζ, and*

$$\mathbb{E}\exp(\zeta^2/\psi^2) \le 2.$$

Proof of Lemma 11. Let

$$v = \left(\frac{\psi}{4}\right)^2.$$

Since $\psi > 0$, this variance is positive. Under $N(0, v)$, the density identity is, for every $x \in \mathbb{R}$,

$$\frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{x^2}{2v}\right) \exp\left(\frac{x^2}{\psi^2}\right) = \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{7x^2}{\psi^2}\right),$$

because $1/(2v) = 8/\psi^2$.

Since $7/\psi^2 > 0$, the function $x \mapsto \exp\{-7x^2/\psi^2\}$ is integrable on $\mathbb{R}$. The density identity therefore transfers this integrability to $x \mapsto \exp(x^2/\psi^2)$ under $N(0, v)$, so $\exp(\zeta^2/\psi^2)$ is integrable under the law of ζ.

For the expectation, the same density identity rewrites the Gaussian-law integral as

$$\mathbb{E} \exp(\zeta^2/\psi^2) = \frac{1}{\sqrt{2\pi v}} \int_{\mathbb{R}} \exp\left(-\frac{7x^2}{\psi^2}\right) dx.$$

The Gaussian integral formula applied with coefficient $7/\psi^2$ gives

$$\mathbb{E} \exp(\zeta^2/\psi^2) = \frac{1}{\sqrt{2\pi v}} \sqrt{\frac{\pi}{7/\psi^2}}.$$

Finally,

$$\sqrt{\frac{\pi}{7/\psi^2}} \leq 2\sqrt{2\pi v},$$

since after squaring both sides this is

$$\frac{\pi\psi^2}{7} \leq 8\pi v = \frac{\pi\psi^2}{2}.$$

Therefore

$$\mathbb{E} \exp(\zeta^2/\psi^2) \leq 2.$$

□

Lemma 11 is the tail-scale normalization used in the lower-bound construction below.

The final auxiliary result supplies the decision-theoretic lower-bound path. It constructs one-dimensional submodels in which the observed treatment law, regression functions, and supplied clipped codes remain fixed while the coefficient varies locally at $n^{-1/2}$ scale. Such paths are the standard input for parametric lower bounds in semiparametric experiments [Bickel et al., 1993, van der Vaart, 1998].

Lemma 12. [`lem:non-gaussian-hard-submodel`] (One-dimensional hard submodel) *Fix $r \in \mathbb{N}$ and real constants $\delta, C_\theta, C_g, C_q, \psi_\eta, \psi_\xi$ with $\psi_\xi > 0$. There exist constants $c_0, C_0, \tau, c_{\mathrm{ACE}} \in \mathbb{R}$ such that*

$$0 < c_0, \qquad 0 < C_0, \qquad 0 < \tau \leq \psi_\xi, \qquad 0 < c_{\mathrm{ACE}}.$$

For every parameter record p whose constants $r, \delta, C_\theta, C_g, C_q, \psi_\eta, \psi_\xi$ are the ones just fixed, and every $n \geq 1$, if the non-Gaussian class in Definition 3 is nonempty, then there are a base law $P^\circ \in \mathcal{P}_{\mathrm{NG},n}$ and a family $\{P_{n,\vartheta} : |\vartheta| \leq c_0\}$ such that:

- ***(Class membership and target.)*** *For every $|\vartheta| \leq c_0$, $P_{n,\vartheta} \in \mathcal{P}_{\mathrm{NG},n}$ and $\theta_0(P_{n,\vartheta}) = \vartheta$.*
- ***(Fixed observed treatment law.)*** *For every $|\vartheta| \leq c_0$, the law of (X, T) under $P_{n,\vartheta}$ equals its law under P°.*

- ***(Fixed regressions and supplied codes.)*** *For every* $|\vartheta| \le c_0$, *the functions* g_0, q_0 *and the clipped supplied codes* $\bar{g}_n, \bar{q}_n$ *under* $P_{n,\vartheta}$ *equal those under* P°.

- ***(Gaussian outcome innovation.)*** *For every* $|\vartheta| \le c_0$, *the outcome innovation* $\xi = Y - q_0(X) - \theta_0\eta$ *under* $P_{n,\vartheta}$ *has law* $N(0, \tau^2)$ *and is independent of* (X, T).

- ***(Outcome equation.)*** *For every* $|\vartheta| \le c_0$,

$$Y = q_0(X) + \vartheta\,\eta + \xi$$

almost surely under $P_{n,\vartheta}$.

- ***(Product relative entropy.)*** *For all real* $\vartheta_0, \vartheta_1, h$ *with* $|\vartheta_0| \le c_0$, $|\vartheta_1| \le c_0$, *and*

$$|\vartheta_1 - \vartheta_0| = \frac{h}{\sqrt{n}},$$

one has, with D_{KL} *the Kullback–Leibler divergence,*

$$D_{\mathrm{KL}}\Big(P_{n,\vartheta_0}^{\otimes n} \,\Big\|\, P_{n,\vartheta_1}^{\otimes n}\Big) \le C_0 h^2.$$

Moreover, for every starting law $P^\circ \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}$ *in the class of Definition 5, there is a family* $\{P^{\mathrm{ACE}}_{n,\vartheta} : |\vartheta| \le c_0\}$ *such that:*

- ***(JMS ACE membership and target.)*** *For every* $|\vartheta| \le c_0$, $P^{\mathrm{ACE}}_{n,\vartheta} \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}$ *and* $\theta_0(P^{\mathrm{ACE}}_{n,\vartheta}) = \vartheta$.

- ***(Fixed observed treatment law.)*** *For every* $|\vartheta| \le c_0$, *the law of* (X, T) *under* $P^{\mathrm{ACE}}_{n,\vartheta}$ *equals its law under* P°.

- ***(Fixed nuisance objects.)*** *For every* $|\vartheta| \le c_0$, *the functions* g_0, q_0, *the clipped supplied codes* $\bar{g}_n, \bar{q}_n$, *and the treatment innovation* η *under* $P^{\mathrm{ACE}}_{n,\vartheta}$ *equal those under* P°, *and the treatment and outcome code-radius conditions in Definition 5 hold at* $P^{\mathrm{ACE}}_{n,\vartheta}$.

- ***(Gaussian outcome innovation.)*** *For every* $|\vartheta| \le c_0$, *the outcome innovation* $\xi = Y - q_0(X) - \theta_0\eta$ *under* $P^{\mathrm{ACE}}_{n,\vartheta}$ *has law* $N(0, \tau^2)$ *and is independent of* (X, T).

- ***(Outcome equation.)*** *For every* $|\vartheta| \le c_0$,

$$Y = q_0(X) + \vartheta\,\eta + \xi$$

almost surely under $P^{\mathrm{ACE}}_{n,\vartheta}$.

- ***(Product relative entropy.)*** *For all real* $\vartheta_0, \vartheta_1, h$ *with* $|\vartheta_0| \le c_0$, $|\vartheta_1| \le c_0$, *and*

$$|\vartheta_1 - \vartheta_0| = \frac{h}{\sqrt{n}},$$

one has

$$D_{\mathrm{KL}}\big((P^{\mathrm{ACE}}_{n,\vartheta_0})^{\otimes n} \,\big\|\, (P^{\mathrm{ACE}}_{n,\vartheta_1})^{\otimes n}\big) \le C_0 h^2.$$

Finally, for every pair of clipped-code functions $g_\star, q_\star : \mathcal{X} \to \mathbb{R}$, if there is a law $P^\circ \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}$ with $\bar{g}_n = g_\star$ and $\bar{q}_n = q_\star$, then the fixed-code JMS ACE minimax risk satisfies

$$\frac{c_{\mathrm{ACE}}}{n} \le \inf_{\widetilde{\theta}(O_1,\dots,O_n;\bar{g}_n,\bar{q}_n)} \sup_{\substack{P \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}: \\ \bar{g}_n = g_\star,\ \bar{q}_n = q_\star}} E_P\left[\{\widetilde{\theta} - \theta_0(P)\}^2\right].$$

Proof of Lemma 12. Set

$$c_0 = \begin{cases} \min\{C_\theta, 1\}, & C_\theta > 0, \\ 1, & C_\theta \le 0, \end{cases} \qquad \tau = \frac{\psi_\xi}{4}, \qquad C_0 = \max\left\{1, \frac{16\psi_\eta^2}{\psi_\xi^2}\right\}.$$

Then $c_0 > 0$, $C_0 > 0$, and $0 < \tau \le \psi_\xi$. For any admissible choice of parameters whose displayed constants agree with the fixed constants, Definition 1 gives $C_\theta > 0$, and hence $c_0 = \min\{C_\theta, 1\} \le C_\theta$.

The two-point testing bound at relative-entropy budget $C_0 c_0^2$ supplies a constant $c_{\mathrm{LC}} > 0$, depending only on that budget, such that whenever two n-sample laws have relative entropy at most $C_0 c_0^2$, every measurable sample statistic $\widehat{\alpha}$ with finite quadratic risks satisfies

$$c_{\mathrm{LC}}(\theta_1 - \theta_0)^2 \le \max\left\{E_0[(\widehat{\alpha} - \theta_0)^2], E_1[(\widehat{\alpha} - \theta_1)^2]\right\}.$$

The minimax reduction applies this bound after symmetrically clipping an arbitrary estimator to a range containing both target values. Define

$$c_{\mathrm{ACE}} = c_{\mathrm{LC}} c_0^2 > 0.$$

Fix now an admissible choice of parameters p with the displayed primitive constants and a class index $n \ge 1$. Assume first that $\mathcal{P}_{\mathrm{NG},n}$ is nonempty, and choose $P^\circ_{\mathrm{NG}} \in \mathcal{P}_{\mathrm{NG},n}$, represented by a base model m°_{NG}. For each $\vartheta \in \mathbb{R}$, define $P_{n,\vartheta}$ by preserving the joint law of (X, T) under P°_{NG}, drawing an independent

$$\xi_\star \sim N(0, \tau^2),$$

and setting

$$Y = q_0(X) + \vartheta\{T - g_0(X)\} + \xi_\star.$$

Equivalently, the path is the product of the retained (X, T)-law and the centered Gaussian innovation, transported through this affine outcome map.

Along this path,

$$\theta_0(P_{n,\vartheta}) = \vartheta, \qquad g_0(P_{n,\vartheta}) = g_0(P^\circ_{\mathrm{NG}}), \qquad q_0(P_{n,\vartheta}) = q_0(P^\circ_{\mathrm{NG}}),$$

and the clipped supplied codes satisfy

$$\bar{g}_n(P_{n,\vartheta}) = \bar{g}_n(P^\circ_{\mathrm{NG}}), \qquad \bar{q}_n(P_{n,\vartheta}) = \bar{q}_n(P^\circ_{\mathrm{NG}}).$$

The covariate-treatment marginal is also retained:

$$(P_{n,\vartheta})_{X,T} = (P^\circ_{\mathrm{NG}})_{X,T}.$$

The treatment innovation is unchanged along the path:

$$\eta_{P_{n,\vartheta}} = \eta_{P^\circ_{\mathrm{NG}}}.$$

The outcome innovation is the fresh Gaussian coordinate:

$$\xi_{P_{n,\vartheta}} = Y - q_0(X) - \vartheta\eta = \xi_\star,$$

so

$$(P_{n,\vartheta}) \circ \xi_{P_{n,\vartheta}}^{-1} = N(0,\tau^2), \qquad \xi_{P_{n,\vartheta}} \perp (X,T).$$

For $|\vartheta| \le c_0$, the coefficient range holds because $c_0 \le C_\theta$. All treatment-side assumptions, the regression ranges, the cumulant separation, and the treatment $L^1(P_X)$ code-radius condition are inherited from P°_{NG}. The conditional mean condition follows from the displayed outcome equation, the independence of $\xi_\star$ from (X,T), and its centering. Finally, Lemma 11 applied at scale ψ_ξ gives

$$E\exp\{\xi_\star^2/\psi_\xi^2\} \le 2.$$

Since $\xi_{P_{n,\vartheta}} = \xi_\star$ is independent of X, the conditional Luxemburg part of Assumption 9 is

$$\exp\{\xi_{P_{n,\vartheta}}^2/\psi_\xi^2\} \in L^1(P_{n,\vartheta}), \qquad E_{P_{n,\vartheta}}\Big[\exp\{\xi_{P_{n,\vartheta}}^2/\psi_\xi^2\} \mid X\Big] \le 2 \quad P_{n,\vartheta}\text{-a.s.}$$

Hence

$$|\vartheta| \le c_0 \quad \Longrightarrow \quad P_{n,\vartheta} \in \mathcal{P}_{\mathrm{NG},n}.$$

The same construction gives the asserted outcome equation:

$$Y = q_0(X) + \vartheta\,\eta + \xi \qquad P_{n,\vartheta}\text{-a.s.}$$

It remains, for this branch, to record the product relative-entropy bound. For the affine Gaussian path, one observation satisfies

$$D_{\mathrm{KL}}(P_{n,\vartheta_0}\|P_{n,\vartheta_1}) \le \frac{(\vartheta_1-\vartheta_0)^2\,2\psi_\eta^2}{2\tau^2}.$$

The moment bound used here is the Luxemburg consequence

$$E_{P^\circ_{\mathrm{NG}}}[\eta^2] \le 2\psi_\eta^2.$$

Tensorization over n independent observations gives

$$D_{\mathrm{KL}}\Big(P_{n,\vartheta_0}^{\otimes n}\,\Big\|\,P_{n,\vartheta_1}^{\otimes n}\Big) \le n\,\frac{(\vartheta_1-\vartheta_0)^2\,2\psi_\eta^2}{2\tau^2}.$$

If $|\vartheta_1-\vartheta_0| = h/\sqrt{n}$ and $\tau = \psi_\xi/4$, then

$$n\,\frac{(\vartheta_1-\vartheta_0)^2\,2\psi_\eta^2}{2\tau^2} = \frac{16\psi_\eta^2}{\psi_\xi^2}h^2 \le C_0h^2.$$

Thus the required entropy display holds.

Now let $P^\circ_{\mathrm{ACE}} \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}$. Define $P^{\mathrm{ACE}}_{n,\vartheta}$ by the same affine Gaussian outcome construction from this starting law:

$$Y = q_0(X) + \vartheta\{T-g_0(X)\} + \xi_\star, \qquad \xi_\star \sim N(0,\tau^2), \quad \xi_\star \perp (X,T).$$

The displayed preservation identities remain valid, and the target coordinate is the path parameter:

$$\theta_0(P^{\mathrm{ACE}}_{n,\vartheta}) = \vartheta, \qquad (P^{\mathrm{ACE}}_{n,\vartheta})_{X,T} = (P^{\circ}_{\mathrm{ACE}})_{X,T},$$
$$g_0(P^{\mathrm{ACE}}_{n,\vartheta}) = g_0(P^{\circ}_{\mathrm{ACE}}), \quad q_0(P^{\mathrm{ACE}}_{n,\vartheta}) = q_0(P^{\circ}_{\mathrm{ACE}}), \quad \eta_{P^{\mathrm{ACE}}_{n,\vartheta}} = \eta_{P^{\circ}_{\mathrm{ACE}}},$$

and

$$\bar{g}_n(P^{\mathrm{ACE}}_{n,\vartheta}) = \bar{g}_n(P^{\circ}_{\mathrm{ACE}}), \qquad \bar{q}_n(P^{\mathrm{ACE}}_{n,\vartheta}) = \bar{q}_n(P^{\circ}_{\mathrm{ACE}}).$$

Consequently the two JMS ACE code-radius inequalities are transported unchanged:

$$\|\bar{g}_n - g_0\|_{L^r(P_X)} \le \varepsilon_{1,n}, \qquad \|\bar{q}_n - q_0\|_{L^r(P_X)} \le \varepsilon_{2,n}.$$

Together with the transported treatment-side and range conditions, the coefficient bound $|\vartheta| \le c_0 \le C_\theta$, and the Gaussian conditional Luxemburg calculation above, this gives

$$|\vartheta| \le c_0 \implies P^{\mathrm{ACE}}_{n,\vartheta} \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}.$$

The Gaussian residual law, residual independence, outcome equation, and product entropy bound are the same as in the non-Gaussian branch:

$$D_{\mathrm{KL}}\big((P^{\mathrm{ACE}}_{n,\vartheta_0})^{\otimes n} \,\|\, (P^{\mathrm{ACE}}_{n,\vartheta_1})^{\otimes n}\big) \le C_0 h^2$$

whenever $|\vartheta_1 - \vartheta_0| = h/\sqrt{n}$.

It remains to prove the fixed-code ACE lower bound. Suppose a starting law $P^{\circ}_{\star} \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}$ has clipped codes $g_\star, q_\star$. Set

$$\vartheta_- = -\frac{c_0}{2\sqrt{n}}, \qquad \vartheta_+ = \frac{c_0}{2\sqrt{n}},$$

and let $P_- := P^{\mathrm{ACE}}_{n,\vartheta_-}$, $P_+ := P^{\mathrm{ACE}}_{n,\vartheta_+}$, using $P^{\circ}_{\star}$ as the starting law. Since $n \ge 1$,

$$|\vartheta_-| \le c_0, \qquad |\vartheta_+| \le c_0, \qquad |\vartheta_+ - \vartheta_-| = \frac{c_0}{\sqrt{n}}.$$

Both laws belong to the fixed-code JMS ACE class with clipped codes $g_\star, q_\star$, and the entropy bound gives

$$D_{\mathrm{KL}}(P_-^{\otimes n} \| P_+^{\otimes n}) \le C_0 c_0^2.$$

Applying the two-point lower bound with target values ϑ_- and ϑ_+ yields, for every measurable estimator whose two quadratic risks are finite,

$$\max_{\nu \in \{-,+\}} E_{P_\nu^{\otimes n}}\Big[(\widetilde{\theta} - \theta_0(P_\nu))^2\Big] \ge c_{\mathrm{LC}}(\vartheta_+ - \vartheta_-)^2 = \frac{c_{\mathrm{LC}} c_0^2}{n} = \frac{c_{\mathrm{ACE}}}{n}.$$

The minimax reduction clips arbitrary measurable estimators to $[-C_\theta, C_\theta]$, so the preceding finite-risk display applies inside the fixed-code risk. Taking the supremum over the fixed-code JMS ACE class and then the infimum over estimators gives

$$\frac{c_{\mathrm{ACE}}}{n} \le \inf_{\widetilde{\theta}} \sup_{\substack{P \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}: \\ \bar{g}_n = g_\star,\ \bar{q}_n = q_\star}} E_P\Big[\{\widetilde{\theta} - \theta_0(P)\}^2\Big].$$

□

Lemma 12 is the lower-bound ingredient for the n^{-1} minimax statements. The proof below builds the fixed-code local path and applies the relative-entropy calculation used in the testing-to-estimation step.

The ACE branch of the proof verifies the same fixed-code path inside $\mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}$, with the treatment innovation and clipped nuisance codes preserved along the path.

C Certified construction and executable correspondence

This appendix records the finite arithmetic contract behind the contour statistic in Algorithm 1. The statistical arguments in the main text use ordinary empirical transforms and contour ratios; the material here specifies a rational interval substrate, a deterministic contour-bank record, and the deterministic perturbation statement that connects a selected represented contour calculation to the population ratio. The construction follows the computable-analysis and interval-arithmetic convention that real inputs are accessed through nested rational enclosures with effective moduli [Weihrauch, 2000, Ko, 1991, Moore, 1966].

The first definition fixes the elementary arithmetic layer. It gives represented real and positive-real records, complex rectangles, guarded division, elementary functions on bounded intervals, circle-node construction, and certified mesh integration. These operations are the only numerical primitives used by the represented contour program.

Definition 16. [`def:certified-contour-arithmetic-substrate`] (Certified contour arithmetic substrate $\mathsf{CCIA}_{\mathrm{build}}$) *The bounded-build specification $\mathsf{CCIA}_{\mathrm{build}}$ consists of the following finite rational arithmetic operations and their stated enclosure properties.*

A represented real number x is given by rational closed intervals $I_m = [I_m^-, I_m^+]$ and an executable modulus $\mu : \mathbb{Q}_{>0} \to \mathbb{N}$ such that

$$x \in I_{m+1} \subseteq I_m, \qquad I^+_{\mu(e)} - I^-_{\mu(e)} \le e.$$

A positive represented real number additionally supplies a rational lower bound $p > 0$ with $p \le x$. The executable data are the interval sequence, the modulus, and the positive lower bound.

For rational intervals $A = [a_-, a_+]$ and $B = [b_-, b_+]$, interval addition, subtraction, and multiplication are

$$A + B = [a_- + b_-, a_+ + b_+], \qquad A - B = [a_- - b_+, a_+ - b_-],$$

$$AB = [\min\{a_-b_-, a_-b_+, a_+b_-, a_+b_+\}, \max\{a_-b_-, a_-b_+, a_+b_-, a_+b_+\}] .$$

A complex rectangle is $A + iB$. Complex addition, subtraction, conjugation, and multiplication use these endpoint operations together with

$$(A + iB)(C + iD) = (AC - BD) + i(AD + BC).$$

For $A = [a_-, a_+]$, define

$$\mathrm{sq}_-(A) = \begin{cases} 0, & a_- \le 0 \le a_+, \\ \min\{a_-^2, a_+^2\}, & \textit{otherwise}, \end{cases} \qquad \mathrm{sq}_+(A) = \max\{a_-^2, a_+^2\}.$$

The squared modulus of $A + iB$ is enclosed by adding the corresponding lower and upper square bounds. Rational bisection between nonnegative rational endpoints gives lower and upper square-root enclosures whose width after q bisections is at most the initial width times 2^{-q}. Division is the guarded operation

$$\frac{A + iB}{C + iD} = (A + iB)(C - iD)(C^2 + D^2)^{-1},$$

performed when the modulus routine returns a rational lower bound $m > 0$, with outward division by the positive rational interval enclosing $C^2 + D^2$.

The represented value of π *is constructed from Machin's identity*

$$\pi = 16\arctan(1/5) - 4\arctan(1/239).$$

For $0 < x < 1$*,*

$$\arctan x = \sum_{j=0}^{N}(-1)^j \frac{x^{2j+1}}{2j+1} + R_N(x), \qquad |R_N(x)| \le \frac{x^{2N+3}}{2N+3}.$$

Intersecting the first $m+1$ *requested enclosures gives rational nested enclosures. If* $e = p/q > 0$ *is in lowest terms, the choice* $N = 5q+5$ *makes the propagated Machin remainder smaller than* e*, giving an effective modulus.*

For a rational interval contained in $[-B, B]$*, with rational* $B \ge 0$*, the Taylor polynomials for* exp, sin*, and* cos *are evaluated by the preceding interval operations, with Lagrange remainders bounded by*

$$3^{\lceil B\rceil}\frac{B^{N+1}}{(N+1)!}, \qquad \frac{B^{N+1}}{(N+1)!}, \qquad \frac{B^{N+1}}{(N+1)!}.$$

For a requested rational error $e = p/q > 0$*, take*

$$N = \lceil 2B\rceil + q + 2\lceil B\rceil + 2.$$

After index $\lceil 2B\rceil$*, successive absolute Taylor terms decrease by at least a factor* $1/2$*; since* $3^{\lceil B\rceil} < 2^{2\lceil B\rceil}$ *and* $2^{-q} \le q^{-1} \le e$*, the displayed* N *gives error at most* e*. Finite intersections make the output rectangles nested. Consequently,*

$$\exp(x+iy) = \exp(x)\{\cos y + i\sin y\}$$

has a nested complex-rectangle extension with an effective modulus on every bounded input rectangle. Combining the constructed representation of π *with these trigonometric operations gives, from a positive radius representation* $\mathfrak{r}$ *and integers* $N \ge 1$ *and* $0 \le q < N$*, represented circle-node values*

$$r\exp(2\pi iq/N) \qquad \text{and} \qquad 2\pi i r\exp(2\pi iq/N).$$

Finite rectangle sums and products are primitive recursive. If a complex-valued function on $[0,1]$ *has a supplied rational Lipschitz bound* L*, then a uniform mesh with*

$$N_{\text{mesh}} = \lceil L/e\rceil + 1$$

has oscillation at most e *on each cell. Evaluating every node to a rational error budget divided by the displayed finite number of operations yields sound finite infimum, supremum, and trapezoidal-integral enclosures. The required finite computation consists of* $N_{\text{mesh}} + 1$ *node evaluations, the displayed Taylor cutoffs, square-root bisection cutoffs, and finitely many endpoint operations. Absolute value of a real interval is implemented by the square-free endpoint rule implicit in* sq_- *and* sq_+*, with the same compositional modulus property.*

A concrete implementation satisfies $\mathsf{CCIA}_{\text{build}}$ *when it implements these algorithms, exports their soundness and effective-modulus contracts, and compiles the resulting module. Later executable conclusions are conditional on a compiled implementation satisfying this bounded-build specification.*

Definition 16 supplies the finite representation interface used throughout the executable clauses. Nested intervals and moduli give a standard Type-2-style name for real values, while the endpoint operations and guarded division give sound complex-rectangle enclosures for the contour integrands. The mesh rule turns a supplied Lipschitz bound into finite enclosures for boundary moduli and contour integrals, matching the interval-analysis approach of Moore [1966] within the computable-real framework of Weihrauch [2000], Ko [1991].

The next definition specializes that arithmetic substrate to the contour bank. The primitive record stores certified names for the fixed separation, treatment-noise scale, zero-localization radius, and outer search radius. From those names and the parameter record, the bank construction produces a deterministic translated dyadic list of circles and a positive dyadic lower-modulus certificate.

Definition 17. [`def:contour-bank-handle`] (Certified contour-bank handle $\mathfrak{p}_\star$) *Let p be a parameter record with sample size $n \in \mathbb{N}$, $1 \le n$, orders $r, k \in \mathbb{N}$ satisfying $k = r + 1$ and $2 \le r$, exponent $s \in [0, \infty]$ satisfying $r \le s$, probability level $\gamma \in (1/2, 1)$, constants $C_\theta, C_g, C_q, \psi_\eta, \psi_\xi, \delta, \sigma > 0$, and sequences $\varepsilon_1, \varepsilon_2 : \mathbb{N} \to \mathbb{R}$ satisfying $\varepsilon_\ell(t) \ge 0$ for $t \ge 1$ and $\varepsilon_\ell(j) \le \varepsilon_\ell(i)$ whenever $1 \le i \le j$, for $\ell = 1, 2$. Let $\mathrm{Name}(x)$ denote a supplied certified-real record with value x, and let $\mathrm{Name}_+(x)$ denote such a record together with a rational lower-bound witness $\ell > 0$ satisfying $\ell \le x$. As part of the experiment convention, fix the supplied record*

$$\mathfrak{p}_\star = \big(k_\star, \mathfrak{d}_\star, \mathfrak{p}_{\eta,\star}, \mathfrak{r}_{0,\star}, \mathfrak{r}_{1,\star}\big),$$

where $k_\star$ is exact with $k_\star = k$ and

$$\mathfrak{d}_\star : \mathrm{Name}_+(\delta), \quad \mathfrak{p}_{\eta,\star} : \mathrm{Name}_+(\psi_\eta), \quad \mathfrak{r}_{0,\star} : \mathrm{Name}_+(R_0), \quad \mathfrak{r}_{1,\star} : \mathrm{Name}_+(R_1).$$

The value contracts are

$$R_0 = (2^{k+4}k!)^{1/(k-2)}(\psi_\eta^2/\delta)^{1/(k-2)}, \qquad R_1 = R_0 + 1.$$

Writing $A_k = (2^{k+4}k!)^{1/(k-2)}$ for the numerical factor, this reads $R_0 = A_k(\psi_\eta^2/\delta)^{1/(k-2)}$. The search radius of the record is this outer radius: $\mathrm{searchRadius}(p) = R_1$. The record $\mathfrak{p}_\star$ is the primitive certified input shared by the bank and the estimator for this parameter record. The bank data assembled below from p and $\mathfrak{p}_\star$ is written B.

Invoke the supplied moduli of $\mathfrak{p}_{\eta,\star}$ and $\mathfrak{r}_{1,\star}$ at error one. If the returned intervals have rational upper endpoints u_ψ and u_R, define

$$U_\psi = \lceil \max\{1, u_\psi\} \rceil, \qquad U_R = \lceil \max\{1, u_R\} \rceil, \qquad N_{\mathrm{cert}} = 4U_\psi^2(U_R + 2)^3.$$

Set

$$m_\star = N_{\mathrm{cert}} + 3, \qquad d_\star = 2^{-m_\star}, \qquad h_\star = d_\star/3, \qquad J_{\mathrm{base}} = 2^{m_\star - 1}.$$

The bank therefore contains

$$J = J_{\mathrm{base}} + 1$$

circles, indexed by $j \in \{0, \dots, J_{\mathrm{base}}\}$. For each $j = 0, \dots, J_{\mathrm{base}}$, use certified addition on the fixed record $\mathfrak{r}_{0,\star}$ and the rational point record for $1/4 + j2^{-m_\star}$ to form

$$\rho_j = \mathfrak{r}_{0,\star} + \mathrm{Name}(1/4 + j2^{-m_\star}).$$

The associated real-valued radius field is

$$\rho_j = (\rho_j).\,\mathrm{value} = R_0 + 1/4 + j2^{-m_\star}.$$

Define the integers

$$e_\star = 8U_R U_\psi^2 (U_R+1)^2, \qquad d_\star^{\mathrm{den}} = 3(2U_R+1),$$

$$u_\star = 2e_\star + N_{\mathrm{cert}}(m_\star + d_\star^{\mathrm{den}}).$$

Define the rational dyadic certificate and its real value by

$$a_{\star,\mathbb{Q}} = 2^{-u_\star} > 0, \qquad a_\star = (a_{\star,\mathbb{Q}} : \mathbb{R}).$$

Every output datum is obtained by the displayed finite modulus calls, rational endpoint operations, certified additions with rational point records, and primitive recursion over the displayed integer bounds. The construction is a finite certified procedure in the supplied record $\mathfrak{p}_\star$.

Definition 17 converts the analytic radius from Lemma 3 into a finite bank used by Algorithm 1. The radii fill a translated dyadic grid outside the zero-localization disk and inside the outer search region, and the certificate $a_\star$ gives the deterministic denominator scale used by the selector. Because the construction depends on fixed primitive records and finite modulus calls, the same bank can be transported across the common-experiment sequences described in Theorem 4.

The represented-data transducer in Algorithm 1 applies Definition 16 to the sample records $\mathfrak{t}_i, \mathfrak{y}_i, \mathfrak{g}_i$. It first forms certified residual records, then evaluates the pilot-fold winding integrand on each bank circle, and finally evaluates the selected contour ratio on the evaluation fold. The fallback branches are part of the total rule: a failed admissibility check or a failed evaluation-fold modulus check returns the displayed value 0, while a successful run returns the clipped midpoint of a certified real interval. Thus the ordinary statistic and the represented execution have the same branch structure and the same clipped value whenever the compiled arithmetic implementation satisfies the bounded-build contract.

The final lemma is the deterministic accuracy statement used by the statistical risk proof. It compares the evaluation-fold empirical transforms with arbitrary population maps F and G on the selected circle, assumes the selected contour identifies a real target ϑ, and turns uniform transform errors into a bound for the clipped contour statistic.

Lemma 13. [`lem:selected-contour-perturbation-bound`] (Selected contour perturbation) *Fix a parameter record p as in Definition 1, the certified primitive and range records $\mathfrak{p}_\star$ and $\mathfrak{c}_{\theta,\star}$ of Definition 17 and Algorithm 1, a supplied treatment-regression code sequence, and observations $O_1, \ldots, O_n$ with $O = (X, T, Y)$ as in Definition 2. Let B be the contour bank generated by p and $\mathfrak{p}_\star$, let $a_\star$ be its dyadic lower-modulus certificate, let C_j have radius ρ_j, and write $R_1 =$* $\mathrm{searchRadius}(p)$. *Let $F, G : \mathbb{C} \to \mathbb{C}$, let $d, e, C_G, \mu, \nu \in \mathbb{R}$, and let $\vartheta \in \mathbb{R}$ be the target value.*

Assume:

- ***(Target range.)*** *The target satisfies*

$$|\vartheta| \le C_\theta.$$

- ***(Selector event.)*** *The canonical represented input built from the clipped current code and the sample satisfies*

$$d \geq 0, \qquad e \geq 0, \qquad C_G \geq 0, \qquad \mu > 0, \qquad \nu > 0,$$

$$\mu = \frac{31a_\star}{64}, \qquad \nu = \frac{23a_\star}{64}, \qquad d \leq \mu - \nu,$$

and there are a bank index j and an integer N_j such that the certified selector of Algorithm 1 selects C_j, the pilot winding output on C_j decodes to N_j, $N_j \geq 1$, the evaluation-fold modulus enclosure on C_j has lower endpoint at least $a_\star/4$, and every $z \in C_j$ satisfies

$$\mu \leq |F(z)|, \qquad |G(z)| \leq C_G, \qquad |\widehat{F}_{1,n}(z) - F(z)| \leq d, \qquad |\widehat{G}_{1,n}(z) - G(z)| \leq e,$$

where $\widehat{F}_{1,n}$ and $\widehat{G}_{1,n}$ are the split-fold empirical transforms of Definition 15.

- ***(Contour integrability.)*** *For every bank index j selected by the canonical selector, both contour ratios*

$$z \mapsto \frac{\widehat{G}_{1,n}(z)}{\widehat{F}_{1,n}(z)} \qquad \text{and} \qquad z \mapsto \frac{G(z)}{F(z)}$$

are integrable on the circle C_j.

- ***(Exact identification.)*** *For every bank index j and every integer N, if the canonical selector selects C_j and the pilot winding output on C_j decodes to N, then*

$$\vartheta = \Re\left[\{N\,2\pi i\}^{-1} \oint_{C_j} \frac{G(z)}{F(z)}\,dz\right].$$

Then the total Borel contour statistic $\widehat{\theta}_{\mathrm{spec},n}$ of Algorithm 1, computed on the same canonicalized sample, satisfies

$$\left|\widehat{\theta}_{\mathrm{spec},n} - \vartheta\right| \leq R_1 \max\left\{\nu^{-1},\, C_G(\mu\nu)^{-1}\right\}(d+e) + \frac{1}{\max\{n,1\}}.$$

Proof of Lemma 13. On the selector event, choose the selected bank index j and the decoded integer $N \geq 1$. Write C_j for the selected circle and set

$$\widehat{T}_j = \Re\left[\{N\,2\pi i\}^{-1} \oint_{C_j} \frac{\widehat{G}_{1,n}(z)}{\widehat{F}_{1,n}(z)}\,dz\right], \qquad T_j = \Re\left[\{N\,2\pi i\}^{-1} \oint_{C_j} \frac{G(z)}{F(z)}\,dz\right].$$

The exact-identification hypothesis for the selected decoded contour gives $\vartheta = T_j$.

For each $z \in C_j$, the selector event gives

$$\mu \leq |F(z)|, \qquad |G(z)| \leq C_G, \qquad |\widehat{F}_{1,n}(z) - F(z)| \leq d, \qquad |\widehat{G}_{1,n}(z) - G(z)| \leq e.$$

Together with $d \leq \mu - \nu$, the triangle inequality yields the empirical denominator margin

$$|\widehat{F}_{1,n}(z)| \geq |F(z)| - |\widehat{F}_{1,n}(z) - F(z)| \geq \mu - d \geq \nu.$$

Since $\mu > 0$ and $\nu > 0$, both denominators are nonzero on C_j. Hence

$$\frac{\widehat{G}_{1,n}}{\widehat{F}_{1,n}} - \frac{G}{F} = \frac{\widehat{G}_{1,n} - G}{\widehat{F}_{1,n}} + \frac{G(F - \widehat{F}_{1,n})}{\widehat{F}_{1,n}F},$$

and the two terms have moduli bounded by e/ν and $C_G d/(\mu\nu)$, respectively. Thus, pointwise on C_j,

$$\left|\frac{\widehat{G}_{1,n}(z)}{\widehat{F}_{1,n}(z)} - \frac{G(z)}{F(z)}\right| \le \frac{e}{\nu} + \frac{C_G d}{\mu\nu}.$$

The contour-integrability hypothesis for the selected index permits integration of this pointwise bound around the circle of radius ρ_j. Therefore

$$\left|\oint_{C_j} \frac{\widehat{G}_{1,n}(z)}{\widehat{F}_{1,n}(z)}\,dz - \oint_{C_j} \frac{G(z)}{F(z)}\,dz\right| \le 2\pi\rho_j \left(\frac{e}{\nu} + \frac{C_G d}{\mu\nu}\right).$$

Multiplication by $\{N 2\pi i\}^{-1}$ and taking real parts give

$$|\widehat{T}_j - \vartheta| \le \frac{\rho_j}{N}\left(\frac{e}{\nu} + \frac{C_G d}{\mu\nu}\right).$$

It remains to relate this contour value to the returned statistic. For the same selected j and decoded N, the construction of $\widehat{\theta}_{\mathrm{spec},n}$ in Algorithm 1 uses the rational midpoint y of the real coordinate of the certified evaluation rectangle. That rectangle contains the normalized empirical contour value whose real part is $\widehat{T}_j$, and its real-coordinate width is at most $2/\max\{n,1\}$. Hence

$$|y - \widehat{T}_j| \le \frac{1}{\max\{n,1\}}.$$

The statistic is the clipping of this rational value to $[-C_\theta, C_\theta]$. Since $|\vartheta| \le C_\theta$, projection onto this interval is contractive relative to ϑ, so

$$|\widehat{\theta}_{\mathrm{spec},n} - \vartheta| \le |y - \vartheta| \le |y - \widehat{T}_j| + |\widehat{T}_j - \vartheta|.$$

By the bank construction in Definition 17, the selected radius satisfies $\rho_j < R_1$, and the decoded count satisfies $N \ge 1$. Using $d, e, C_G \ge 0$ and $\mu, \nu > 0$,

$$\frac{\rho_j}{N}\left(\frac{e}{\nu} + \frac{C_G d}{\mu\nu}\right) \le R_1 \max\{\nu^{-1},\, C_G(\mu\nu)^{-1}\}(d + e).$$

Combining the preceding displays proves the asserted bound. $\square$

Lemma 13 is the deterministic bridge between certified execution and the risk calculation in Theorem 3. The selector event supplies a lower bound for the population denominator and an evaluation-fold lower-modulus check; the exact-identification clause supplies the population target represented by the same selected contour. The conclusion then bounds the statistic by the uniform denominator and numerator errors on the selected circle, with the final $1/\max\{n,1\}$ term coming from the certified interval radius used when the real contour moment is converted into the clipped rational midpoint.

Combining Definitions 16 and 17 and Lemma 13 gives the executable correspondence needed by the main contour theorem. The arithmetic substrate provides sound finite enclosures, the bank handle supplies the fixed finite search geometry, and the perturbation lemma states the deterministic accuracy of the selected represented statistic on the good event used in the statistical argument.

D Verification scope and crosswalk

This appendix records how the reader-facing statistical claims in the paper align with the formal statements displayed in the main text and technical appendices. The crosswalk is organized by mathematical role: model primitives and law classes, analytic identification, fixed-separation estimation, comparison results, local benchmarks, and certified execution.

The model primitives are fixed by Definitions 1 and 2. Sampling and structural restrictions are the assumptions in Assumptions 1 to 16. The law classes and risk criteria used in the statistical comparisons are the objects in Definitions 3 to 11.

The analytic identification chain is the sequence Definitions 12 and 13, Theorems 1 and 2, and Lemma 1. These statements fix the treatment-innovation transform, the observable residual transforms, and the contour ratio that identifies the partially linear coefficient under the stated zero-free and positive-count conditions.

The fixed-separation estimation result is anchored by Lemma 3 and Theorems 3 and 4. These statements provide the zero-localization radius, the finite-bank contour statistic, the n^{-1} minimax mean-squared-error scale on the fixed-code non-Gaussian class, the corresponding generalized-quantile bound, and the Gaussian-class diagnostic.

The comparison with ACE is recorded in Propositions 1 and 2 and Lemma 5. The published order-r ACE guarantee enters through the cited Jin–Mackey–Syrgkanis result [Jin et al., 2025]; the displayed proposition states the exact imported bound and the class relations used to compare it with the contour guarantee.

The explicit mixture and local benchmark statements are Definition 14, Proposition 3, Lemma 7, and Theorem 5. The future-work formulation in Definition 15 and Remark 1 records the sample-level contour information and the triangular local-to-Gaussian adaptation target.

The certified execution layer is specified by Definitions 16 and 17 and Algorithm 1. These statements connect the ordinary Borel statistic used in the statistical risk bounds with the finite represented-data construction based on the same primitive records.

Verification scope. The theorem statements and proofs corresponding to the displayed formal environments are machine-checked in Lean 4 under their stated assumptions and citation interfaces. The ACE-alignment theorem in Proposition 1 verifies the class reduction, the application of the imported bound, and the comparison algebra conditional on the Jin–Mackey–Syrgkanis conclusion cited in its theorem-local footnote.

E Proofs of the main results

Proof of Theorem 1. Put

$$D_n(X) = g_0(X) - \bar{g}_n(X), \qquad b_n(X) = q_0(X) - \theta_0 D_n(X),$$

as in Definition 7, and write

$$J(w) = J_{z_0,\ell}(w) = w^{\ell-1} e^{z_0 w}, \qquad w \in \mathbb{C},$$

as in Definition 12. From the definitions of η and Z_n in Definitions 2 and 13,

$$Z_n = T - \bar{g}_n(X) = \eta + D_n(X).$$

The definitions of ξ in Definition 2 and b_n above also give, by algebra,

$$Y = \theta_0 Z_n + b_n(X) + \xi.$$

Membership in $\mathcal{P}_{\mathrm{NG},n}$, together with the clipping convention in Definition 2, supplies the integrability and boundedness used below. Pulled back to the observation space, Definition 3 gives

$$|g_0(X)| \le C_g, \qquad |q_0(X)| \le C_q, \qquad |\theta_0| \le C_\theta \qquad P\text{-a.s.},$$

and the clipping gives $|\bar{g}_n(X)| \le C_g$. Hence

$$|D_n(X)| \le 2C_g, \qquad |b_n(X)| \le C_q + C_\theta(2C_g) \qquad P\text{-a.s.}$$

The treatment and outcome sub-Gaussian clauses in Definition 3 imply real exponential moments of every order for η and ξ; the displayed bound on D_n then gives the corresponding real exponential moments for $Z_n = \eta + D_n(X)$. These facts justify the differentiated transforms, conditional expectations, and products appearing below.

1. Since M is analytic at z_0 and has analytic order $\ell \ge 1$ there,

   $$M^{(\alpha)}(z_0) = 0 \qquad \text{for every } 0 \le \alpha < \ell.$$

   Moreover, for each such α,

   $$M^{(\alpha)}(z_0) = E_P[\eta^\alpha e^{z_0\eta}],$$

   by differentiating the moment-generating function under the expectation. Hence, for any $d \in \mathbb{R}$,

   $$\begin{aligned} E_P[J(\eta + d)] &= E_P\Big[(\eta + d)^{\ell-1} e^{z_0(\eta+d)}\Big] \\ &= e^{z_0 d} \sum_{\alpha=0}^{\ell-1} \binom{\ell-1}{\alpha} d^{\ell-1-\alpha} E_P[\eta^\alpha e^{z_0\eta}] \\ &= e^{z_0 d} \sum_{\alpha=0}^{\ell-1} \binom{\ell-1}{\alpha} d^{\ell-1-\alpha} M^{(\alpha)}(z_0) = 0. \end{aligned}$$

   This is the shifted annihilation identity.

2. We next show the conditional identity. Let $\mathcal{A}_X$ be any measurable covariate event. Since $Z_n = \eta + D_n(X)$, the binomial expansion gives

   $$\begin{aligned} &\int_{\{X \in \mathcal{A}_X\}} J(Z_n)\, dP \\ &\quad = \sum_{\alpha=0}^{\ell-1} \binom{\ell-1}{\alpha} \int \eta^\alpha e^{z_0\eta} D_n(X)^{\ell-1-\alpha} e^{z_0 D_n(X)} \mathbf{1}_{\mathcal{A}_X}(X)\, dP. \end{aligned}$$

   The treatment-noise independence in Definition 3 factors each summand as

   $$\binom{\ell-1}{\alpha} E_P[\eta^\alpha e^{z_0\eta}]\; E_P\Big[D_n(X)^{\ell-1-\alpha} e^{z_0 D_n(X)} \mathbf{1}_{\mathcal{A}_X}(X)\Big],$$

   and the first factor is $M^{(\alpha)}(z_0) = 0$. Thus the integral over every covariate event is zero, which characterizes

   $$E_P[J(Z_n) \mid X] = 0 \qquad P\text{-a.s.}$$

3. Define, for $z \in \mathbb{C}$,
$$F_n(z) = E_P[e^{zZ_n}].$$
Since $Z_n = \eta + D_n(X)$, the same independence gives the factorization
$$F_n(z) = E_P[e^{z\eta}]\, E_P[e^{zD_n(X)}] = M(z)H_n(z),$$
with H_n as in Definition 13. Also,
$$E_P[Z_n J(Z_n)] = E_P[Z_n^\ell e^{z_0 Z_n}] = F_n^{(\ell)}(z_0).$$
Differentiating $F_n = MH_n$ ℓ times at z_0 yields
$$F_n^{(\ell)}(z_0) = \sum_{\alpha=0}^{\ell} \binom{\ell}{\alpha} M^{(\alpha)}(z_0) H_n^{(\ell-\alpha)}(z_0).$$
All terms with $\alpha < \ell$ vanish by the analytic-order calculation above, and the remaining term is
$$F_n^{(\ell)}(z_0) = M^{(\ell)}(z_0)H_n(z_0).$$
Therefore
$$E_P[Z_n J(Z_n)] = M^{(\ell)}(z_0)H_n(z_0).$$

4. Finally, using the algebraic decomposition $Y = \theta_0 Z_n + b_n(X) + \xi$,
$$E_P[YJ(Z_n)] = \theta_0 E_P[Z_n J(Z_n)] + E_P[b_n(X)J(Z_n)] + E_P[\xi J(Z_n)].$$
The second term is zero because $b_n(X)$ is covariate-measurable and the conditional identity above gives
$$E_P[b_n(X)J(Z_n)] = E_P[b_n(X)E_P[J(Z_n) \mid X]] = 0.$$
The third term is zero because $J(Z_n)$ is measurable with respect to (X, T), while Definition 3 gives $E_P[\xi \mid X, T] = 0$:
$$E_P[\xi J(Z_n)] = E_P[J(Z_n)E_P[\xi \mid X, T]] = 0.$$
Thus
$$E_P[YJ(Z_n)] = \theta_0 E_P[Z_n J(Z_n)].$$
Whenever $E_P[Z_n J(Z_n)] \neq 0$, division gives
$$\theta_0 = \frac{E_P[YJ(Z_n)]}{E_P[Z_n J(Z_n)]}.$$

□

Proof of Lemma 1. Fix $z \in \mathbb{C}$. The definitions give
$$Z_n = T - \bar{g}_n(X) = \eta + \{g_0(X) - \bar{g}_n(X)\} = \eta + D_n(X).$$
Moreover, Definitions 2 and 3 give the pointwise clipping bound $|\bar{g}_n(X)| \leq C_g$ and the treatment-regression range bound $|g_0(X)| \leq C_g$ almost surely, hence
$$|D_n(X)| = |g_0(X) - \bar{g}_n(X)| \leq |g_0(X)| + |\bar{g}_n(X)| \leq 2C_g \qquad P\text{-a.s.}$$

Together with the Luxemburg treatment-tail condition for η, this bound gives every real exponential moment of Z_n, and hence the exponential integrability used below. Since $D_n(X)$ is a function of X and η is independent of X under Definition 3,

$$\begin{aligned} F_n(z) &= E_P \exp\{z(\eta + D_n(X))\} \\ &= E_P\left[\exp(z\eta)\exp\{zD_n(X)\}\right] \\ &= E_P \exp(z\eta)\ E_P \exp\{zD_n(X)\} = M(z)H_n(z). \end{aligned}$$

For the second identity, first note the pointwise residual decomposition

$$Y = \theta_0 Z_n + b_n(X) + \xi,$$

because Definition 7 gives $b_n(X) = q_0(X) - \theta_0 D_n(X)$, while $Z_n = \eta + D_n(X)$ and $\xi = Y - q_0(X) - \theta_0\eta$. The preceding exponential-moment envelope for Z_n also gives differentiation under the expectation:

$$F_n'(z) = E_P\left[Z_n e^{zZ_n}\right].$$

Thus

$$\begin{aligned} G_n(z) &= E_P\left[Y e^{zZ_n}\right] \\ &= \theta_0 E_P\left[Z_n e^{zZ_n}\right] + E_P\left[b_n(X)e^{zZ_n}\right] + E_P\left[\xi e^{zZ_n}\right]. \end{aligned}$$

The contamination weight is bounded almost surely by

$$\begin{aligned} |b_n(X)| &= |q_0(X) - \theta_0 D_n(X)| \\ &\le |q_0(X)| + |\theta_0|\,|D_n(X)| \qquad P\text{-a.s.} \\ &\le C_q + C_\theta(2C_g), \end{aligned}$$

where the last line uses the outcome-regression and coefficient range bounds in Definition 3 together with the displayed bound on D_n. This boundedness and the same exponential envelope for Z_n give integrability of the contamination term.

The outcome-noise term is integrable as well. The ξ-tail condition in Definition 3 gives $E_P[|\xi|^2] < \infty$, and the learned-residual envelope gives

$$E_P\left[\left|e^{zZ_n}\right|^2\right] = E_P\left[e^{2\,\mathrm{Re}(z)Z_n}\right] < \infty.$$

By Cauchy–Schwarz, $E_P[|\xi e^{zZ_n}|] < \infty$. Since $Z_n = T - \bar{g}_n(X)$, the real and imaginary parts of e^{zZ_n} are measurable with respect to (X, T). Applying $E_P[\xi \mid X, T] = 0$ from Definition 3 separately to these integrable real products yields

$$E_P\left[\xi e^{zZ_n}\right] = 0.$$

Finally, independence of η and X factors the contamination term:

$$\begin{aligned} E_P\left[b_n(X)e^{zZ_n}\right] &= E_P\left[b_n(X)e^{zD_n(X)}e^{z\eta}\right] \\ &= E_P e^{z\eta}\ E_P\left[b_n(X)e^{zD_n(X)}\right] = M(z)B_n(z). \end{aligned}$$

Substituting the last three displays and the derivative identity gives

$$G_n(z) = \theta_0 F_n'(z) + M(z)B_n(z).$$

Since z was arbitrary, both identities hold for every $z \in \mathbb{C}$. □

Proof of Theorem 2. Write ρ_j for the radius of the selected positively oriented bank circle $C_j = \{z \in \mathbb{C} : |z| = \rho_j\}$. The bank construction gives $\rho_j > 0$. Put

$$D_n(X) = g_0(X) - \bar{g}_n(X), \qquad b_n(X) = q_0(X) - \theta_0(P)D_n(X),$$

and

$$B_n(z) = E_P\Big[b_n(X)e^{zD_n(X)}\Big], \qquad z \in \mathbb{C}.$$

By Lemma 10,

$$|D_n(X)| \le 2C_g \qquad P\text{-a.s.}$$

The clipping defining $\bar{g}_n$, the range conditions in $\mathcal{P}_{\mathrm{NG},n}$, and $|\theta_0(P)| \le C_\theta$ give

$$|b_n(X)| \le |q_0(X)| + |\theta_0(P)|\,|D_n(X)| \le C_q + 2C_\theta C_g \qquad P\text{-a.s.}$$

Consequently H_n and B_n are complex differentiable at every point. Indeed, on the unit neighborhood of any $z_0 \in \mathbb{C}$,

$$|D_n(X)e^{zD_n(X)}| \le 2C_g \exp\{(|z_0| + 1)2C_g\},$$

and

$$|b_n(X)D_n(X)e^{zD_n(X)}| \le (C_q + 2C_\theta C_g)(2C_g)\exp\{(|z_0| + 1)2C_g\},$$

so dominated differentiation under the integral applies to the weighted transforms. Separately, the sub-Gaussian exponential-moment condition for the treatment innovation η, together with $Z_n = \eta + D_n(X)$ and the almost-sure bound on D_n, gives $E_P e^{tZ_n} < \infty$ for every real t. The standard moment-generating-function analyticity criterion therefore gives analyticity of F_n on a neighborhood of the closed disk $\{|z| \le \rho_j\}$.

On the closed disk $\{|z| \le \rho_j\}$, the nuisance zero-freeness assumption gives $H_n(z) \neq 0$. Hence the quotient B_n/H_n is well-defined on the closed disk, differentiable in the open disk, and continuous on the closed disk. Cauchy's theorem on the circle gives

$$\oint_{C_j} \frac{B_n(z)}{H_n(z)}\,dz = 0.$$

For $z \in C_j$, Lemma 1 gives

$$F_n(z) = M(z)H_n(z), \qquad G_n(z) = \theta_0(P)F_n'(z) + M(z)B_n(z),$$

where $M(z) = E_P[e^{z\eta}]$. Since $F_n(z) \neq 0$ and $H_n(z) \neq 0$ on C_j, also $M(z) \neq 0$ on C_j. Therefore, pointwise for $z \in C_j$,

$$\frac{G_n(z)}{F_n(z)} = \theta_0(P)\frac{F_n'(z)}{F_n(z)} + \frac{B_n(z)}{H_n(z)}.$$

Integrating around C_j and using the Cauchy identity for the quotient B_n/H_n yields

$$\oint_{C_j} \frac{G_n(z)}{F_n(z)}\,dz = \theta_0(P)\oint_{C_j} \frac{F_n'(z)}{F_n(z)}\,dz.$$

The function F_n is analytic near the closed disk and zero-free on C_j, so the argument principle gives

$$\frac{1}{2\pi i}\oint_{C_j} \frac{F_n'(z)}{F_n(z)}\,dz = N_{C_j}(F_n).$$

Thus

$$\oint_{C_j} \frac{G_n(z)}{F_n(z)}\, dz = \theta_0(P)\, N_{C_j}(F_n)\, 2\pi i.$$

The positive-count assumption gives $N_{C_j}(F_n) \geq 1$, so $N_{C_j}(F_n)\, 2\pi i \neq 0$. Dividing by this nonzero factor gives

$$\theta_0(P) = \{N_{C_j}(F_n)\, 2\pi i\}^{-1} \oint_{C_j} \frac{G_n(z)}{F_n(z)}\, dz.$$

By Definition 13, the right-hand side is $\Theta_{C_j}(F_n, G_n)$. Therefore

$$(\theta_0(P) : \mathbb{C}) = \Theta_{C_j}(F_n, G_n) = \{N_{C_j}(F_n)\, 2\pi i\}^{-1} \oint_{C_j} \frac{G_n(z)}{F_n(z)}\, dz.$$

□

Proof of Lemma 2. Fix $t \in \mathbb{R}$ and put

$$\lambda = t\psi, \qquad v = \frac{W}{\psi}.$$

The square-exponential envelope first gives integrability of e^{tW}: for every outcome point,

$$tW \leq \frac{W^2}{\psi^2} + \frac{t^2\psi^2}{4},$$

because $(W/\psi - t\psi/2)^2 \geq 0$. Hence e^{tW} is dominated by $e^{t^2\psi^2/4}e^{W^2/\psi^2}$.

Consider first the case $|\lambda| \leq 1$. For every real u,

$$e^u \leq 1 + u + u^2 e^{|u|}.$$

Indeed, when $|u| \leq 1$, the Taylor-remainder bound $|e^u - 1 - u| \leq u^2$ gives $e^u - 1 - u \leq u^2 \leq u^2 e^{|u|}$. When $|u| > 1$ and $u \geq 0$, one has $u > 1$, so $e^u \leq u^2 e^{|u|} \leq 1 + u + u^2 e^{|u|}$. When $|u| > 1$ and $u < 0$, one has $u \leq -1$, whence $e^u \leq 1 \leq 1 + u + u^2 \leq 1 + u + u^2 e^{|u|}$. With $u = \lambda v$,

$$|\lambda v| \leq \frac{\lambda^2 + v^2}{2} \leq \frac{1 + v^2}{2}.$$

Also,

$$v^2 e^{(1+v^2)/2} = e^{1/2} v^2 e^{v^2/2} \leq e^{1/2} e^{v^2} \leq 2e^{v^2},$$

where the two inequalities use $v^2 \leq e^{v^2/2}$ and $e^{1/2} \leq 2$. Combining these estimates with monotonicity of the exponential gives

$$(\lambda v)^2 e^{|\lambda v|} \leq 2\lambda^2 e^{v^2}.$$

Therefore

$$e^{tW} = e^{\lambda v} \leq 1 + tW + 2\lambda^2 \exp(W^2/\psi^2).$$

Integrating and using $E_P W = 0$ and $E_P \exp(W^2/\psi^2) \leq 2$,

$$E_P e^{tW} \leq 1 + 4\lambda^2 \leq \exp(4\lambda^2) = \exp(4\psi^2 t^2).$$

It remains to treat $|\lambda| > 1$. Completing the square gives

$$e^{tW} \leq \exp(\lambda^2/4)\exp(W^2/\psi^2).$$

After integration,

$$E_P e^{tW} \leq 2\exp(\lambda^2/4).$$

Since $2 \leq e$ and $|\lambda| > 1$ implies $1 < \lambda^2$,

$$2\exp(\lambda^2/4) \leq \exp(1 + \lambda^2/4) \leq \exp(4\lambda^2) = \exp(4\psi^2 t^2).$$

The two cases prove the asserted bound for every $t \in \mathbb{R}$. □

Proof of Lemma 3. Let

$$M(z) = E_P e^{z\eta}.$$

From $k = r + 1$ and $r \geq 2$, one has $k \geq 3$. The treatment-tail assumption gives

$$E_P \exp(\eta^2/\psi_\eta^2) \leq 2,$$

together with the corresponding integrability, and the conditional-mean identity for g_0 gives

$$E_P \eta = E_P\{T - g_0(X)\} = 0.$$

Applying Lemma 2 to η gives, for every real t,

$$E_P e^{t\eta} \leq \exp(4\psi_\eta^2 t^2).$$

Consequently, for every $z \in \mathbb{C}$,

$$|M(z)| \leq E_P e^{\mathrm{Re}(z)\eta} \leq \exp\{4\psi_\eta^2 |z|^2\}.$$

The same exponential-integrability envelope gives analyticity of the complex MGF on all of $\mathbb{C}$, and normalization at the origin gives $M(0) = 1$.

Suppose, toward a contradiction, that $M(z) \neq 0$ for every $|z| \leq R_0$. Then $\log|M|$ is harmonic on the open disk $\{|z| < R_0\}$. Choose an analytic function F on this disk whose real part is $\log|M|$, and set

$$f(z) = F(z) - F(0).$$

Then $f(0) = 0$ and, for $|z| < R_0$,

$$\mathrm{Re}\, f(z) = \log|M(z)| \leq 4\psi_\eta^2 R_0^2.$$

The Borel–Carathéodory estimate on the disk of radius R_0, applied to the normalized function f, gives on the sphere $|z| = R_0/2$

$$|f(z)| \leq 8\psi_\eta^2 R_0^2.$$

Cauchy's estimate on the disk of radius $R_0/2$ therefore yields

$$|f^{(k)}(0)| \leq \frac{k!\, 8\psi_\eta^2 R_0^2}{(R_0/2)^k} = 2^{k+3} k!\, \psi_\eta^2 R_0^{2-k}.$$

Near the origin, $M(0) = 1$, so M takes values in the slit plane on a small disk and the principal logarithm $\log M$ is analytic there. The analytic function

$$z \mapsto f(z) - \log M(z)$$

has identically zero real part on that small connected disk and vanishes at 0; hence it is identically zero there. Thus

$$f^{(k)}(0) = \frac{d^k}{dz^k} \log M(z)\bigg|_{z=0}.$$

Since $\kappa_k(\eta)$ is the real part of this derivative and $|\operatorname{Re} w| \le |w|$,

$$|\kappa_k(\eta)| \le 2^{k+3} k!\, \psi_\eta^2 R_0^{2-k}.$$

By the definition

$$R_0 = (2^{k+4} k!)^{1/(k-2)} \left(\frac{\psi_\eta^2}{\delta} \right)^{1/(k-2)},$$

we have

$$R_0^{k-2} = 2^{k+4} k!\, \frac{\psi_\eta^2}{\delta}.$$

Substitution in the preceding bound gives

$$|\kappa_k(\eta)| \le 2^{k+3} k!\, \psi_\eta^2 \left(2^{k+4} k!\, \frac{\psi_\eta^2}{\delta} \right)^{-1} = \frac{\delta}{2} < \delta.$$

This contradicts the cumulant separation condition $|\kappa_k(\eta)| \ge \delta$. Therefore there exists $z_0 \in \mathbb{C}$ with $|z_0| \le R_0$ and $M(z_0) = 0$. □

Proof of Lemma 4. Put $R = R_1$, $Z_n = T - \bar{g}_n(X)$, and

$$A = 64\{C_q^4 + 4C_\theta^4 \psi_\eta^4 + 4\psi_\xi^4\} + 2\exp\{16RC_g + 16R^2\psi_\eta^2\}.$$

The radius is nonnegative, since $R = R_0 + 1$ and $R_0 \ge 0$ by its displayed construction in Definition 17.

First,

$$E_P|Y|^4 \le 64\{C_q^4 + 4C_\theta^4 \psi_\eta^4 + 4\psi_\xi^4\}.$$

Indeed, $Y = q_0(X) + \theta_0\eta + \xi$; the range bounds give $|q_0(X)| \le C_q$ and $|\theta_0| \le C_\theta$, while the Luxemburg bounds give $E_P|\eta|^4 \le 4\psi_\eta^4$ and, after integrating the conditional bound, $E_P|\xi|^4 \le 4\psi_\xi^4$. Applying the elementary fourth-power bound to the three summands gives the display.

Next,

$$E_P \exp\{8R|Z_n|\} \le 2\exp\{16RC_g + 16R^2\psi_\eta^2\}.$$

This is the residual exponential envelope applied with radius $2R$. In detail, $Z_n = \eta + D_n$, where $D_n = g_0(X) - \bar{g}_n(X)$ satisfies $|D_n| \le 2C_g$ almost surely. Hence $8R|Z_n| \le 8R|\eta| + 16RC_g$, and Young's inequality at this radius gives

$$8R|\eta| \le \eta^2/\psi_\eta^2 + 16R^2\psi_\eta^2.$$

The Luxemburg bound $E_P \exp(\eta^2/\psi_\eta^2) \le 2$ then yields the displayed exponential bound.

Therefore

$$E_P\left[\{|Y|\exp(2R|Z_n|)\}^2\right] \le A.$$

The pointwise inequality used here is

$$\{|Y|\exp(2R|Z_n|)\}^2 \le |Y|^4 + \exp(8R|Z_n|),$$

and the preceding two displays bound the two expectations.

Let

$$E(o) = |Y(o)|\exp\{2R|Z_n(o)|\}.$$

Then $E \in L^2(P)$ and $\|E\|_2 \le \sqrt{A}$. For $|z| \le R$,

$$|Y\exp(zZ_n)| = |Y|\exp\{\mathrm{Re}(zZ_n)\} \le |Y|\exp\{2R|Z_n|\} = E,$$

because $\mathrm{Re}(zZ_n) \le |z|\,|Z_n| \le 2R|Z_n|$. Hence

$$|G_n(z)| = |E_P\{Y\exp(zZ_n)\}| \le E_P E \le \|E\|_2 \le \sqrt{A}.$$

Since $\sqrt{A} = C_G(p)$, the claimed bound follows. □

Proof of Theorem 3. 1. Define the primitive search radius

$$R_{\mathrm{fix}} = A_{r+1}\left(\frac{\psi_\eta^2}{\delta}\right)^{1/(r-1)} + 1,$$

where A_{r+1} is the numerical factor in Definition 17. Let $K_{\mathrm{emp}} > 0$ be the constant supplied by Lemma 9 for $(C_\theta, C_g, C_q, \psi_\eta, \psi_\xi, R_{\mathrm{fix}})$. For $y \in \mathbb{R}$, put

$$u_{\mathrm{ceil}}(y) = \lceil \max\{1, y+1\} + 1 \rceil,$$

and define

$$\begin{gathered}
U_\psi^\circ = u_{\mathrm{ceil}}(\psi_\eta), \qquad U_R^\circ = u_{\mathrm{ceil}}(R_{\mathrm{fix}}),\\
N_{\mathrm{cert}}^\circ = 4(U_\psi^\circ)^2(U_R^\circ+2)^3, \qquad m_\star^\circ = N_{\mathrm{cert}}^\circ + 3,\\
e_\star^\circ = 8U_R^\circ(U_\psi^\circ)^2(U_R^\circ+1)^2, \qquad d_{\star,\mathrm{den}}^\circ = 3(2U_R^\circ+1),\\
u_\star^\circ = 2e_\star^\circ + N_{\mathrm{cert}}^\circ\left(m_\star^\circ + d_{\star,\mathrm{den}}^\circ\right), \qquad a_\star^\circ = 2^{-u_\star^\circ}.
\end{gathered}$$

Thus $a_\star^\circ > 0$. Set

$$C_G^\circ = \left[64\{C_q^4 + 4C_\theta^4\psi_\eta^4 + 4\psi_\xi^4\} + 2\exp\{16R_{\mathrm{fix}}C_g + 16R_{\mathrm{fix}}^2\psi_\eta^2\}\right]^{1/2},$$

$$L_{\mathrm{fix}} = R_{\mathrm{fix}}\max\left\{\left(\frac{23a_\star^\circ}{64}\right)^{-1}, C_G^\circ\left(\frac{31a_\star^\circ}{64}\frac{23a_\star^\circ}{64}\right)^{-1}\right\},$$

and

$$C_{\mathrm{fix}} = 12L_{\mathrm{fix}}^2K_{\mathrm{emp}} + \frac{1536C_\theta^2K_{\mathrm{emp}}}{(a_\star^\circ)^2} + 2.$$

Since $R_{\mathrm{fix}} > 0$, $K_{\mathrm{emp}} > 0$, and $a_\star^\circ > 0$, one has $C_{\mathrm{fix}} > 0$. The two fixed-code converse arguments below give positive constants $\mathfrak{c}_{\mathrm{NG}}$ and $\mathfrak{c}_{\mathrm{JMS}}$, depending only on the same primitive tuple: $\mathfrak{c}_{\mathrm{NG}}$ comes from the fixed-code non-Gaussian converse, and $\mathfrak{c}_{\mathrm{JMS}}$ comes from the separate fixed-code JMS ACE converse. We take

$$c = \min\{\mathfrak{c}_{\mathrm{NG}}, \mathfrak{c}_{\mathrm{JMS}}\}, \qquad C = C_{\mathrm{fix}}.$$

Then $c > 0$ and $C > 0$.

2. Fix the covariate measurable space, p_0, the fixed records, p, and the base model as in the statement. The transported-record hypothesis gives

$$p.C_\theta = C_\theta, \quad p.C_g = C_g, \quad p.C_q = C_q, \quad p.\psi_\eta = \psi_\eta, \quad p.\psi_\xi = \psi_\xi, \quad \mathrm{searchRadius}(p) = R_{\mathrm{fix}}.$$

Let $\mathfrak{p}_{\star,p}$ and $\mathfrak{c}_{\theta,\star,p}$ denote the transported bank and range records. If $a_{\star,p}$ is the lower-modulus certificate of the actual contour bank, then the width-one certified refinement containing ψ_η and R_{fix} gives

$$u_{\star,p} \le u_\star^\circ, \qquad a_\star^\circ = 2^{-u_\star^\circ} \le 2^{-u_{\star,p}} = a_{\star,p}.$$

Consequently the actual selector constant

$$C_{\mathrm{sel}}(p, \mathfrak{p}_{\star,p}) = 12L_p^2 K_{\mathrm{emp}} + \frac{1536 p.C_\theta^2 K_{\mathrm{emp}}}{a_{\star,p}^2} + 2,$$

with

$$L_p = \mathrm{searchRadius}(p) \max\left\{ \left(\frac{23a_{\star,p}}{64}\right)^{-1}, C_G(p)\left(\frac{31a_{\star,p}}{64}\frac{23a_{\star,p}}{64}\right)^{-1} \right\},$$

satisfies

$$C_{\mathrm{sel}}(p, \mathfrak{p}_{\star,p}) \le C.$$

The monotonicity is exactly the monotonicity of the bank exponent, the map $u \mapsto 2^{-u}$, and positive reciprocals.

3. The statistic is measurable. Indeed, $\widehat{\theta}_{\mathrm{spec},n}$ is obtained by running the finite rational contour program of Algorithm 1 and then clipping its raw real output:

$$\widehat{\theta}_{\mathrm{spec},n}(O_1, \ldots, O_n) = \min\left\{\max\left\{Y_{\mathrm{raw}}(O_1, \ldots, O_n), -C_\theta\right\}, C_\theta\right\}.$$

On each finite branch trace, the queried observation endpoints are floor-dyadic functions of the sample coordinates, and the remaining operations are rational arithmetic and finite comparisons. The terminal traces are countable and exhaustive, so Y_{raw} is Borel measurable; clipping by constants preserves measurability.

4. The represented execution clause follows from the canonical compilation contract. For every sample,

$$\mathrm{Rep}_{\mathrm{compiled}}\big(\mathrm{canonicalInput}_{p,\mathfrak{p}_{\star,p},\mathfrak{c}_{\theta,\star,p}}\big) = \mathrm{Ord}_{p,\mathfrak{p}_{\star,p},\mathfrak{c}_{\theta,\star,p}},$$

the represented trace equals the ordinary trace, and the represented output record certifies

$$\min\{\max\{Y_{\mathrm{raw}}, -C_\theta\}, C_\theta\}.$$

The last identity is the elementary three-case equality

$$\frac{|y + C_\theta| - |y - C_\theta|}{2} = \min\{\max\{y, -C_\theta\}, C_\theta\},$$

used with the positive range record for C_θ.

5. The class relations are those of Lemma 5: for every model m,

$$m \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n} \quad \Longrightarrow \quad m \in \mathcal{P}_{\mathrm{NG},n},$$

and, if $s = r$,

$$m \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n} \quad \Longleftrightarrow \quad m \in \mathcal{P}^{\mathrm{ACE}}_{\mathrm{NG},n}.$$

The proof uses monotonicity of $L^p(P_X)$-norms on a probability space to pass from the published $L^r(P_X)$ treatment-code condition to the $L^1(P_X)$ treatment-code condition, and when $s = r$ the comparison subclass imposes the same code-radius requirement.

6. We record the pointwise risk bound used in both fixed-code classes. Let $m \in \mathcal{P}_{\mathrm{NG},n}$, assume $n \geq 2$, and assume

$$\varepsilon_{1,n} \leq \{4R_1 \exp(2C_g R_1)\}^{-1},$$

where $R_1 = \mathrm{searchRadius}(p)$. For $a \in \{0, 1\}$ and every sample define

$$d_a = \sup_{|z| \leq R_1} |\widehat{F}_{a,n}(z) - F_n(z)|, \qquad e_a = \sup_{|z| \leq R_1} |\widehat{G}_{a,n}(z) - G_n(z)|.$$

Set

$$\mathcal{A} = \{d_0 \leq a_{\star,p}/8,\ d_1 \leq a_{\star,p}/8\}.$$

On $\mathcal{A}$, the canonical selector packet supplies exactly the hypotheses of Lemma 13. The packet is obtained as follows. First, Lemma 8 gives a bank contour C_{j_0} for the treatment-innovation transform with positive zero count and boundary modulus at least $a_{\star,p}$. By Lemma 10, the smallness condition gives $|H_n(z)| \geq 3/4$ on the search disk and preserves the zero count from the treatment-innovation transform to $F_n = MH_n$; hence C_{j_0} has positive F_n-zero count, H_n is zero-free on and inside the contour, and $|F_n(z)| \geq 3a_{\star,p}/4$ on its boundary.

The pilot-fold inequality $d_0 \leq a_{\star,p}/8$, together with the certified modulus width $a_{\star,p}/64$ and the certified winding enclosure in Algorithm 1, makes this population contour admissible: its pilot lower endpoint is at least $39a_{\star,p}/64$, and its decoded pilot integer is the positive population zero count. The selected contour maximizes the pilot lower endpoint among admissible contours, so for the selected contour C_j and decoded integer N,

$$\mu_p := \frac{31a_{\star,p}}{64} \leq |F_n(z)| \qquad (z \in C_j),$$

while the evaluation-fold inequality $d_1 \leq a_{\star,p}/8$ gives the evaluation modulus margin required by the selector and

$$\nu_p := \frac{23a_{\star,p}}{64} > 0, \qquad d_1 \leq \mu_p - \nu_p.$$

The same certified node specifications give

$$|\widehat{F}_{1,n}(z) - F_n(z)| \leq d_1, \qquad |\widehat{G}_{1,n}(z) - G_n(z)| \leq e_1 \qquad (z \in C_j),$$

and Lemma 4 gives $|G_n(z)| \leq C_G(p)$ on $|z| \leq R_1$. Continuity and the positive denominator margins give the two contour-integrability hypotheses for the empirical and population ratios. Finally, Theorem 2 applies to the selected contour, because F_n is zero-free on C_j, H_n is zero-free on and inside C_j, and the decoded integer N is the positive population zero count; therefore the selected population contour ratio has real part $\theta_0(m)$.

Applying Lemma 13 with

$$\mu_p = \frac{31a_{\star,p}}{64}, \qquad \nu_p = \frac{23a_{\star,p}}{64}$$

gives, on $\mathcal{A}$,

$$|\widehat{\theta}_{\mathrm{spec},n} - \theta_0(m)| \leq L_p(d_1 + e_1) + \frac{1}{n}.$$

Hence, on $\mathcal{A}$,

$$(\widehat{\theta}_{\mathrm{spec},n} - \theta_0(m))^2 \leq 4L_p^2(d_1^2 + e_1^2) + \frac{2}{n^2}.$$

The folds have cardinality at least $n/3$, and Lemma 9 gives

$$E(d_a^2 + e_a^2) \leq \frac{3K_{\mathrm{emp}}}{n} \qquad (a = 0, 1).$$

Markov's inequality therefore yields

$$P(\mathcal{A}^c) \leq \frac{384K_{\mathrm{emp}}}{a_{\star,p}^2 n}.$$

Since the estimator and the target both lie in $[-C_\theta, C_\theta]$,

$$(\widehat{\theta}_{\mathrm{spec},n} - \theta_0(m))^2 \leq 4C_\theta^2$$

on $\mathcal{A}^c$. Integrating the two displays gives

$$E_m\Big[(\widehat{\theta}_{\mathrm{spec},n} - \theta_0(m))^2\Big] \leq \frac{C_{\mathrm{sel}}(p, \mathfrak{p}_{\star,p})}{n} \leq \frac{C}{n}.$$

7. The quantile bound is the Markov consequence of the preceding mean-square bound. For

$$W(O_1, \ldots, O_n) = |\widehat{\theta}_{\mathrm{spec},n}(O_1, \ldots, O_n) - \theta_0(m)|, \qquad t = \sqrt{\frac{C}{\gamma n}},$$

the risk bound gives

$$P(W > t) \leq \frac{C/n}{t^2} = \gamma.$$

Equivalently,

$$P(W \leq t) \geq 1 - \gamma,$$

and the defining property of the generalized lower quantile in Definition 6 gives

$$Q_{P,1-\gamma}(W) \leq \sqrt{\frac{C}{\gamma n}}.$$

8. It remains to identify the two lower constants. Put

$$b_\theta = \min\{C_\theta, 1\}, \qquad V_{\mathrm{KL}} = \max\left\{1, \frac{16\psi_\eta^2}{\psi_\xi^2}\right\}, \qquad \tau_{\mathrm{lo}} = \frac{\psi_\xi}{4}.$$

For a nonempty fixed-code non-Gaussian class, choose m° in that class. For $n \geq 2$, define

$$\theta_- = -\frac{b_\theta}{2\sqrt{n}}, \qquad \theta_+ = \frac{b_\theta}{2\sqrt{n}}.$$

The affine Gaussian perturbations m_- and m_+ keep the law of (X,T), g_0, q_0, and the supplied clipped codes from m°, and use the outcome equation

$$Y = q_0(X) + \theta_\pm \eta + \xi_\tau, \qquad \xi_\tau \sim N(0, \tau_{\mathrm{lo}}^2),$$

with ξ_τ independent of (X,T). Since $|\theta_\pm| \leq b_\theta \leq C_\theta$, and since Lemma 11 gives the required outcome-innovation Luxemburg envelope at scale ψ_ξ, both m_- and m_+ remain in the same fixed-code non-Gaussian class. Their target separation is

$$|\theta_+ - \theta_-| = \frac{b_\theta}{\sqrt{n}}.$$

The pair $\{m_-, m_+\}$ is a one-dimensional fixed-code path of the kind supplied by Lemma 12: the observed treatment law, the regression functions, and the clipped supplied codes are held fixed while the target moves by $h/\sqrt{n}$, and the divergence along such a path is quadratic in h. Carrying out that Gaussian likelihood ratio calculation here, with $h = b_\theta$, gives

$$D_{\mathrm{KL}}\big(m_-^{\otimes n} \,\big\|\, m_+^{\otimes n}\big) \leq n\, \frac{(\theta_+ - \theta_-)^2\, 2\psi_\eta^2}{2\tau_{\mathrm{lo}}^2} = \frac{16\psi_\eta^2}{\psi_\xi^2} b_\theta^2 \leq V_{\mathrm{KL}} b_\theta^2.$$

Le Cam's two-point MSE inequality therefore supplies a constant $\lambda_{\mathrm{LC}} > 0$, depending only on $V_{\mathrm{KL}} b_\theta^2$, such that every measurable estimator $\widetilde{\theta}$ satisfies

$$\max_{v \in \{-,+\}} E_{m_v}\left[(\widetilde{\theta} - \theta_0(m_v))^2\right] \geq \lambda_{\mathrm{LC}} \frac{b_\theta^2}{n}.$$

Thus the fixed-code non-Gaussian minimax risk is at least

$$\frac{\mathfrak{c}_{\mathrm{NG}}}{n}, \qquad \mathfrak{c}_{\mathrm{NG}} = \lambda_{\mathrm{LC}} b_\theta^2 > 0.$$

This is the fixed-code non-Gaussian converse.

For a nonempty fixed-code JMS ACE class, the separate fixed-code ACE converse applies the same two-point argument starting from a law in that JMS ACE class. The affine perturbations preserve JMS ACE membership, the target values θ_- and θ_+, the law of (X,T), the nuisance functions g_0, q_0, the treatment innovation η, and both clipped supplied codes; this is the JMS ACE branch of Lemma 12. The same product relative-entropy calculation and Le Cam inequality give a constant $\mathfrak{c}_{\mathrm{JMS}} > 0$, depending only on the primitive tuple, such that

$$\frac{\mathfrak{c}_{\mathrm{JMS}}}{n} \leq \inf_{\widetilde{\theta}} \sup_{P \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}(p;\mathrm{base})} E_P\left[(\widetilde{\theta} - \theta_0(P))^2\right].$$

9. For the fixed-code non-Gaussian class, assume the hypotheses in the theorem. The lower bound from the preceding step and $c \le \mathfrak{c}_{\mathrm{NG}}$ give

$$\frac{c}{n} \le \mathfrak{R}_n^{\mathrm{NG}}.$$

The statistic $\widehat{\theta}_{\mathrm{spec},n}$ is measurable by Step 3, so it is an admissible rule in the minimax infimum, and therefore

$$\mathfrak{R}_n^{\mathrm{NG}} \le \sup_{P \in \mathcal{P}_{\mathrm{NG},n}(p;\mathrm{base})} E_P\left[(\widehat{\theta}_{\mathrm{spec},n} - \theta_0(P))^2\right].$$

For every P in this fixed-code class, the current clipped treatment code agrees with the base clipped treatment code. Since the canonical inputs in Algorithm 1 use the current clipped treatment code, the statistic assembled from the base treatment-code sequence agrees pointwise with the statistic assembled from the class member's treatment-code sequence. Applying Step 6 to the latter statistic and then using this pointwise equality gives

$$\sup_{P \in \mathcal{P}_{\mathrm{NG},n}(p;\mathrm{base})} E_P\left[(\widehat{\theta}_{\mathrm{spec},n} - \theta_0(P))^2\right] \le \frac{C}{n}.$$

The same pointwise equality reduces the absolute-error statistic in the generalized-quantile display to the statistic covered by Step 7, giving the displayed generalized-quantile bound over the same supremum.

10. For the fixed-code JMS ACE class, assume its nonemptiness, $n \ge 2$, and the same smallness condition on $\varepsilon_{1,n}$. The ACE lower bound from Step 8 and $c \le \mathfrak{c}_{\mathrm{JMS}}$ give

$$\frac{c}{n} \le \inf_{\widetilde{\theta}} \sup_{P \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}(p;\mathrm{base})} E_P\left[(\widetilde{\theta} - \theta_0(P))^2\right].$$

The estimator is measurable, hence admissible, so the infimum is bounded above by its worst-case risk. By Step 5, every JMS ACE law lies in $\mathcal{P}_{\mathrm{NG},n}$. The fixed-code restriction gives the same current clipped treatment code as the base model, so the pointwise congruence from Algorithm 1 identifies the statistic built from the base treatment-code sequence with the statistic built from the class member's treatment-code sequence. Step 6 therefore yields

$$\sup_{P \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}(p;\mathrm{base})} E_P\left[(\widehat{\theta}_{\mathrm{spec},n} - \theta_0(P))^2\right] \le \frac{C}{n}.$$

The same pointwise equality and Step 7 give the asserted generalized-quantile bound on this class as well.

□

Proof of Proposition 1. Fix the displayed constants and choose the constants c and C supplied by Theorem 3. Set

$$C_{\mathrm{spec}} = C.$$

Then $C_{\mathrm{spec}} > 0$, and its dependence is only through $r, \delta, C_\theta, C_g, C_q, \psi_\eta, \psi_\xi$.

1. For the class relations, apply Lemma 5 at the displayed parameter record and the displayed sample size n. It gives, law by law,

$$P \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n} \Longrightarrow P \in \mathcal{P}_{\mathrm{NG},n}$$

and

$$P \in \mathcal{P}^{\mathrm{ACE}}_{\mathrm{NG},n} \Longrightarrow P \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}.$$

Since these pointwise implications are exactly the membership predicates for the displayed law classes, they yield

$$\mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n} \subseteq \mathcal{P}_{\mathrm{NG},n}, \qquad \mathcal{P}^{\mathrm{ACE}}_{\mathrm{NG},n} \subseteq \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}.$$

The same result also gives, when $s = r$,

$$P \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n} \Longleftrightarrow P \in \mathcal{P}^{\mathrm{ACE}}_{\mathrm{NG},n},$$

hence the asserted equality of the two classes under the common clipped-code convention.

2. Fix a model m with observed-data law $P \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}$ using the clipped supplied codes $\bar{g}_n, \bar{q}_n$, and suppose $\mathsf{E}^{\mathrm{JMS}}_n$, $\varepsilon_{1,n} > 0$, and $\varepsilon_{2,n} \geq 0$. If $\varepsilon_{2,n} > 0$, the positive-radius part of the assumed published order-r ACE generalized-quantile guarantee, with the objects and gates stated in Proposition 1, applied with the supplied handle and the same clipped codes, gives

$$Q_{P,1-\gamma}\big(|\widehat{\theta}_{\mathrm{ACE},r,n} - \theta_0(P)|\big) \leq B_{\mathrm{ACE},n,\gamma}.$$

It remains to treat the boundary case $\varepsilon_{2,n} = 0$. For any $\chi > 0$, let $p^{(\chi)}$ be the parameter record obtained from p by replacing the whole second radius sequence by

$$\varepsilon^{(\chi)}_{2,j} = \max\{\varepsilon_{2,j}, \chi\}, \qquad j \geq 1,$$

and leaving every other coordinate unchanged. This sequence is again nonnegative and nonincreasing. Read the same observed-data law, regressions, and supplied code sequences as a model $m^{(\chi)}$ over $p^{(\chi)}$; its observed-data law is still P. If $0 < \chi < \varepsilon_{1,n}$, then the class membership is preserved: every condition in $\mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}$ is unchanged except the outcome-code radius, and at index n

$$\varepsilon^{(\chi)}_{2,n} = \chi \quad \text{with} \quad 0 \leq \varepsilon_{2,n} \leq \chi.$$

Moreover, with

$$\omega_n = (\gamma n)^{-1/2}(\psi_\xi + C_\theta \psi_\eta),$$

the denominator entering the second JMS eligibility quantity is unchanged:

$$\max\{\varepsilon_{1,n}, \max\{\varepsilon^{(\chi)}_{2,n}, \omega_n\}\} = \max\{\varepsilon_{1,n}, \max\{\varepsilon_{2,n}, \omega_n\}\},$$

because $\varepsilon_{2,n} = 0$ and $0 < \chi < \varepsilon_{1,n}$. All other eligibility quantities are unchanged by construction, so $\mathsf{E}^{\mathrm{JMS}}_n$ holds for $p^{(\chi)}$ at the same index.

Let

$$\Lambda_n = C_\gamma r! 16^r \delta^{-1} \varepsilon^r_{1,n}.$$

The assumed ACE guarantee includes $C_\gamma > 0$, and the displayed hypotheses give $\Lambda_n > 0$. For any real $\beta > B_{\mathrm{ACE},n,\gamma}$, set

$$\chi_\beta = \min\left\{\frac{\varepsilon_{1,n}}{2}, \frac{\beta - B_{\mathrm{ACE},n,\gamma}}{2\Lambda_n}\right\}.$$

Then $0 < \chi_\beta < \varepsilon_{1,n}$, and the positive-radius result applied to $p^{(\chi_\beta)}$ gives

$$Q_{P,1-\gamma}\big(|\widehat{\theta}_{\mathrm{ACE},r,n} - \theta_0(P)|\big) \le B_{\mathrm{ACE},n,\gamma} + \Lambda_n \chi_\beta.$$

By the definition of χ_β,

$$\Lambda_n \chi_\beta \le \frac{\beta - B_{\mathrm{ACE},n,\gamma}}{2},$$

so the same quantile is at most β. Since this holds for every $\beta > B_{\mathrm{ACE},n,\gamma}$, the desired boundary inequality follows:

$$Q_{P,1-\gamma}\big(|\widehat{\theta}_{\mathrm{ACE},r,n} - \theta_0(P)|\big) \le B_{\mathrm{ACE},n,\gamma}.$$

3. For the spectral guarantee, assume the displayed nonemptiness condition and choose a model m° with observed-data law $P^\circ \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}$ using the clipped supplied codes $\bar{g}_n, \bar{q}_n$. Define the fixed-code ACE class

$$\mathcal{A}_n^\star = \left\{P \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n} : \bar{g}_n(P) = \bar{g}_n,\ \bar{q}_n(P) = \bar{q}_n\right\}.$$

This is the ACE-class set associated with the base model m° in Theorem 3. The statistic computed from the supplied treatment-code sequence agrees pointwise with the statistic computed from the base model's supplied treatment-code sequence, because both constructions use the current treatment code only through the residuals

$$T_i - \bar{g}_n(X_i), \qquad i = 1, \dots, n.$$

The ACE-class part of Theorem 3, applied with the fixed primitive and range records, $n \ge 2$, and

$$\varepsilon_{1,n} \le \{4R_1 \exp(2C_g R_1)\}^{-1},$$

therefore yields

$$\sup_{P \in \mathcal{A}_n^\star} Q_{P,1-\gamma}\big(|\widehat{\theta}_{\mathrm{spec},n} - \theta_0(P)|\big) \le \sqrt{\frac{C_{\mathrm{spec}}}{\gamma n}}.$$

Since $\mathcal{A}_n^\star$ is exactly the class of laws in the statement with the clipped supplied codes $\bar{g}_n, \bar{q}_n$, this is the asserted spectral upper guarantee.

4. Finally consider a sequence (p_n) satisfying the hypotheses in the last item of the proposition. Define

$$\mathcal{L}_n = \varepsilon_{1,n}^r \varepsilon_{2,n} + C_\theta \varepsilon_{1,n}^{r+1}$$

and

$$\lambda_{\mathrm{ACE}} = C_\gamma r! 16^r \delta^{-1}.$$

The assumed ACE guarantee includes $C_\gamma > 0$, so $\lambda_{\mathrm{ACE}} > 0$. Along the sequence, the fixed-constant identities in the hypotheses rewrite the ACE bound as

$$B_{\mathrm{ACE},n,\gamma} = \lambda_{\mathrm{ACE}} \left[\mathcal{L}_n + 64(C_g + \psi_\eta)^r \{ r^2 (C_g + \psi_\eta) + \psi_\xi + C_\theta \psi_\eta \} (\gamma n)^{-1/2} \right].$$

The second summand in brackets is nonnegative, hence eventually

$$\lambda_{\mathrm{ACE}} \mathcal{L}_n \le B_{\mathrm{ACE},n,\gamma}.$$

The assumed divergence is

$$\frac{\mathcal{L}_n}{n^{-1/2}} \longrightarrow \infty.$$

For $n \ge 1$,

$$\sqrt{\frac{C_{\mathrm{spec}}}{\gamma n}} = \sqrt{\frac{C_{\mathrm{spec}}}{\gamma}}\, n^{-1/2}.$$

Once $\mathcal{L}_n > 0$ and $\lambda_{\mathrm{ACE}} \mathcal{L}_n \le B_{\mathrm{ACE},n,\gamma}$, we have

$$0 \le \frac{\sqrt{C_{\mathrm{spec}}/(\gamma n)}}{B_{\mathrm{ACE},n,\gamma}} \le \frac{\sqrt{C_{\mathrm{spec}}/\gamma}}{\lambda_{\mathrm{ACE}}} \left(\frac{\mathcal{L}_n}{n^{-1/2}} \right)^{-1}.$$

The right-hand side tends to 0, so the squeeze theorem gives

$$\frac{\sqrt{C_{\mathrm{spec}}/(\gamma n)}}{B_{\mathrm{ACE},n,\gamma}} \longrightarrow 0.$$

□

Proof of Theorem 4.

1. Apply Theorem 3 with the fixed constants

$$r, \delta, C_\theta, C_g, C_q, \psi_\eta, \psi_\xi.$$

It gives constants $c, C > 0$, depending only on these displayed constants, with the fixed-code non-Gaussian lower and upper bounds stated there. Also apply Proposition 2 on the chosen measurable covariate space. It gives, for every parameter record p, writing n for the sample size of p, and every $m \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{G},n}$,

$$\theta_0(m) = 0,$$

and it gives the two zero Gaussian minimax conclusions for every nonempty clipped-code intersection at that same sample size.

2. Fix p_0, the fixed primitive records, the sequence (p_j), the supplied measurable deterministic codes $(\widehat{g}, \widehat{q})$, and the displayed eventual non-Gaussian event. Write n_j for the sample size of p_j, and put

$$\mathfrak{p}_{\star,j} := \text{the transport of } \mathfrak{p}_\star \text{ from } p_0 \text{ to } p_j, \qquad \mathfrak{c}_{\theta,\star,j} := \text{the transport of } \mathfrak{c}_{\theta,\star} \text{ from } p_0 \text{ to } p_j.$$

For every sufficiently large j, the displayed event in the theorem supplies a base model m_j^{anc} such that

$$m_j^{\mathrm{anc}} \in \mathcal{P}_{\mathrm{NG},n_j}(p_j), \qquad \bar{g}_{m_j^{\mathrm{anc}},n_j} = \bar{g}_{n_j}, \qquad \bar{q}_{m_j^{\mathrm{anc}},n_j} = \bar{q}_{n_j},$$

together with

$$n_j \geq 2, \qquad \varepsilon_{1,n_j} \leq \Big(4R_1(p_j)\exp\{2C_g R_1(p_j)\}\Big)^{-1}.$$

Let g_j^{anc} and q_j^{anc} be the supplied code sequences carried by m_j^{anc}. The two clipped-code equalities above imply equality of the two fixed-code law sets,

$$\begin{aligned}&\{m:\ m \in \mathcal{P}_{\mathrm{NG},n_j}(p_j),\ \bar{g}_{m,n_j} = \bar{g}_{n_j},\ \bar{q}_{m,n_j} = \bar{q}_{n_j}\}\\ &\quad = \{m:\ m \in \mathcal{P}_{\mathrm{NG},n_j}(p_j),\ \bar{g}_{m,n_j} = \bar{g}_{m_j^{\mathrm{anc}},n_j},\ \bar{q}_{m,n_j} = \bar{q}_{m_j^{\mathrm{anc}},n_j}\}.\end{aligned}$$

The statistic also agrees pointwise when the supplied treatment code is replaced by g_j^{anc}:

$$\widehat{\theta}_{\mathrm{spec},n_j}(p_j, \mathfrak{p}_{\star,j}, \mathfrak{c}_{\theta,\star,j}; \widehat{g}) = \widehat{\theta}_{\mathrm{spec},n_j}(p_j, \mathfrak{p}_{\star,j}, \mathfrak{c}_{\theta,\star,j}; g_j^{\mathrm{anc}}),$$

because Algorithm 1 forms the represented observation record with the current clipped values $\bar{g}_{n_j}(X_i)$. Applying Theorem 3 to p_j with base model m_j^{anc}, and using the i.i.d. product sampling law of that base model, yields the minimax bounds

$$\frac{c}{n_j} \leq \mathfrak{R}^{\mathrm{NG}}_{n_j}(p_j; \widehat{g}, \widehat{q}) \leq \frac{C}{n_j},$$

and, after the preceding law-set and statistic equalities, the estimator upper bound

$$\sup_{P \in \mathcal{P}_{\mathrm{NG},n_j}(p_j; \bar{g}_{n_j}, \bar{q}_{n_j})} \mathbb{E}_P\Big[\big(\widehat{\theta}_{\mathrm{spec},n_j} - \theta_0(P)\big)^2\Big] \leq \frac{C}{n_j}.$$

The measurability of $\widehat{\theta}_{\mathrm{spec},n_j}$ follows from the same pointwise equality and the measurability conclusion of Theorem 3. The transported primitive bank record reuses the same certified names for $k, \delta, \psi_\eta, R_0$, and R_1. By Definition 17, the translated-dyadic contour-bank data are a deterministic function of those names and their error-one refinements. Thus, writing B for the bank generated from p_0 and $\mathfrak{p}_\star$, the bank generated from p_j and $\mathfrak{p}_{\star,j}$ is also B.

For the compiled statement, fix a compiled bounded spectral adapter satisfying the full canonical build-and-compilation specification for $(p_j, \mathfrak{p}_{\star,j}, \mathfrak{c}_{\theta,\star,j}, \widehat{g})$. The represented-data soundness implication for compiled bounded spectral adapters states that this full canonical build-and-compilation specification entails the represented-execution contract for the same parameter record, primitive bank record, range record, and treatment-code sequence. Its proof rewrites the pilot-boundary, winding-quadrature, and evaluation-quadrature entry points to the reference finite programs, identifies the compiled represented program with the ordinary finite-rational program on canonical represented inputs, and then applies the clipping certificate for $\mathfrak{c}_{\theta,\star,j}$. Therefore the compiled represented-data execution realizes

$$(p_j, \mathfrak{p}_{\star,j}, \mathfrak{c}_{\theta,\star,j}, \widehat{g})$$

with the same ordinary finite-rational result, trace, and certified clipped output required by the represented-execution clause of the theorem.

3. For the Gaussian part, fix a parameter record p, write n for its sample size, and fix $P \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{G},n}$. The first conclusion of Proposition 2 gives

$$\theta_0(P) = 0.$$

Now fix measurable deterministic codes $(\widehat{g}, \widehat{q})$ whose clipped-code Gaussian intersection at this sample size is nonempty. The second conclusion of Proposition 2, applied to this same code pair, gives

$$\mathfrak{R}_n^{\mathrm{G}}(p; \widehat{g}, \widehat{q}) = 0, \qquad \mathfrak{M}_{n,1-\gamma}^{\mathrm{G}}(p; \widehat{g}, \widehat{q}) = 0.$$

These are precisely the two Gaussian assertions.

□

Proof of Proposition 3. Set

$$t = \frac{\pi}{2}, \qquad A_{\text{sin}} = \exp\left(-\frac{\pi^2}{8}\right),$$

and

$$K_{\text{sin}} = 4(C_q + 2C_\theta C_g)^2 + 4\psi_\xi^2 + C_\theta^2(8 + 16C_g^2).$$

Choose

$$C = 1 + \frac{16}{A_{\text{sin}}^2} K_{\text{sin}}.$$

Since $A_{\text{sin}} > 0$ and $K_{\text{sin}} \geq 0$, this constant is positive and depends only on $C_\theta, C_g, C_q, \psi_\xi$. Fix p, P, and the model objects satisfying the hypotheses. Define, on the observation space,

$$Z_n = T - \bar{g}_n(X), \qquad D_n(X) = g_0(X) - \bar{g}_n(X),$$

so that $Z_n = \eta + D_n(X)$.

1. The stipulated law gives the treatment-innovation transform explicitly. Indeed, for every $z \in \mathbb{C}$,

$$\begin{aligned} M(z) &= \frac{1}{2}\exp\left(-z + \frac{z^2}{2}\right) + \frac{1}{2}\exp\left(z + \frac{z^2}{2}\right) \\ &= \exp\left(\frac{z^2}{2}\right)\cosh z. \end{aligned}$$

   Therefore

$$M(it) = \exp\left(-\frac{\pi^2}{8}\right)\cos\left(\frac{\pi}{2}\right) = 0.$$

   Near zero,

$$\frac{d}{dz}\log M(z) = z + \tanh z.$$

   The third derivative at zero of z is zero, while

$$(\tanh z)' = \operatorname{sech}^2 z, \qquad (\tanh z)'' = -2\tanh z\,\operatorname{sech}^2 z, \qquad (\tanh z)'''(0) = -2.$$

   Hence the fourth logarithmic derivative of M at zero is -2, which is the asserted fourth cumulant:

$$\kappa_4(\eta) = -2.$$

   All differentiations are justified by the Gaussian mixture's finite exponential moments.

2. The same transform computation gives the weighted characteristic moment used for the denominator. Since

$$M'(z) = \exp(z^2/2)\, z \cosh z + \exp(z^2/2) \sinh z,$$

evaluating at $z = it$ yields

$$E_P\big[\eta\, e^{it\eta}\big] = M'(it) = iA_{\sin}.$$

Using $Z_n = \eta + D_n(X)$, the independence of η and X, and the identity $M(it) = 0$,

$$\begin{aligned} E_P[Z_n e^{itZ_n}] &= E_P[\eta e^{it\eta}]\, E_P[e^{itD_n(X)}] + E_P[e^{it\eta}]\, E_P[D_n(X) e^{itD_n(X)}] \\ &= iA_{\sin}\, E_P[e^{itD_n(X)}]. \end{aligned}$$

Taking imaginary parts gives

$$E_P[Z_n \sin(tZ_n)] = A_{\sin} E_P[\cos(tD_n(X))]\,,$$

which is the displayed identity after substituting $t = \pi/2$.

3. The $L^1(P_X)$ treatment-code condition gives

$$E_P|D_n(X)| = \|\bar{g}_n - g_0\|_{L^1(P_X)} \le \varepsilon_{1,n}.$$

Since $\varepsilon_{1,n} \le 1/\pi$,

$$t\varepsilon_{1,n} \le \frac{1}{2}.$$

Using $|1 - \cos u| \le |u|$,

$$E_P[\cos(tD_n(X))] \ge 1 - tE_P|D_n(X)| \ge 1 - t\varepsilon_{1,n} \ge \frac{1}{2}.$$

Combining this with the denominator identity gives

$$\frac{A_{\sin}}{2} \le E_P[Z_n \sin(tZ_n)]\,.$$

4. Define the residual sine score

$$\mathrm{Rem}_n(O) = \{Y - \theta_0(P) Z_n\} \sin(tZ_n)$$

and the denominator sine score

$$\mathrm{Den}_n(O) = Z_n \sin(tZ_n).$$

With

$$b_n(X) = q_0(X) - \theta_0(P) D_n(X),$$

the outcome equation gives

$$Y - \theta_0(P) Z_n = b_n(X) + \xi.$$

The zero $M(it) = 0$ implies

$$E_P[\sin(t\eta)] = 0, \qquad E_P[\cos(t\eta)] = 0.$$

By independence of η and X,

$$E_P[b_n(X)\sin\{t(\eta + D_n(X))\}] = 0.$$

The conditional mean restriction for the outcome innovation gives

$$E_P[\xi \sin(tZ_n)] = E_P[E_P(\xi \mid X, T)\sin(tZ_n)] = 0.$$

Thus

$$E_P[\mathrm{Rem}_n(O)] = 0.$$

5. The mixture second moment is bounded by

$$E_P[\eta^2] \le 4.$$

The range condition for g_0, together with clipping of $\bar{g}_n$, gives

$$|D_n(X)| \le 2C_g \qquad P\text{-a.s.}$$

Since $|\sin u| \le 1$,

$$E_P[\mathrm{Den}_n(O)^2] \le E_P[Z_n^2] \le 2E_P[\eta^2] + 2E_P[D_n(X)^2] \le 8 + 16C_g^2.$$

Also,

$$|b_n(X)| \le C_q + 2C_\theta C_g.$$

The Luxemburg sub-Gaussian condition in Assumption 9 gives

$$E_P[\xi^2] \le 2\psi_\xi^2.$$

Therefore

$$E_P[\mathrm{Rem}_n(O)^2] \le 4(C_q + 2C_\theta C_g)^2 + 4\psi_\xi^2.$$

6. By Definition 14, the statistic $\widehat{\theta}_{\sin,n}$ is the clipped ratio with denominator score Den_n, remainder score Rem_n, frequency $t = \pi/2$, and threshold $A_{\sin}/4$, because

$$\mathbb{P}_n\{Y\sin(tZ_n)\} = \theta_0(P)\,\mathbb{P}_n\,\mathrm{Den}_n + \mathbb{P}_n\,\mathrm{Rem}_n\,.$$

Let

$$\mu_{\mathrm{Den}} = E_P[\mathrm{Den}_n(O)].$$

The preceding lower bound gives $\mu_{\mathrm{Den}} \ge A_{\sin}/2$. Applying Lemma 6 pathwise and then averaging under the i.i.d. product law yields

$$E_P\left[(\widehat{\theta}_{\sin,n} - \theta_0(P))^2\right] \le \frac{16}{A_{\sin}^2 n}\left\{E_P[\mathrm{Rem}_n(O)^2] + C_\theta^2 E_P[\mathrm{Den}_n(O)^2]\right\}.$$

Using the two score bounds,

$$E_P\left[(\widehat{\theta}_{\sin,n} - \theta_0(P))^2\right] \le \frac{16K_{\sin}}{A_{\sin}^2 n} \le \frac{C}{n}.$$

Together with the identities and denominator lower bound proved above, this proves all claims.

□

Proof of Lemma 6. Write

$$\bar{W} = \mathbb{P}_n W, \qquad \bar{R} = \mathbb{P}_n R, \qquad \Delta_W = \mathbb{P}_n(W - \mu_W).$$

Since $n \geq 1$,

$$\Delta_W = \bar{W} - \mu_W.$$

Consider first the branch $\bar{W} \geq A/4$. Then $\bar{W} > 0$, and the unclipped ratio obeys

$$\frac{\theta\bar{W} + \bar{R}}{\bar{W}} - \theta = \frac{\bar{R}}{\bar{W}}.$$

The projection onto $[-C_\theta, C_\theta]$ fixes θ, because $|\theta| \leq C_\theta$, and is 1-Lipschitz. Hence

$$|\widehat{\theta} - \theta| \leq \left|\frac{\bar{R}}{\bar{W}}\right| \leq \frac{4|\bar{R}|}{A}.$$

Squaring and adding the nonnegative term involving Δ_W gives

$$(\widehat{\theta} - \theta)^2 \leq \frac{16}{A^2}\bar{R}^2 \leq \frac{16}{A^2}\left\{\bar{R}^2 + C_\theta^2\Delta_W^2\right\}.$$

It remains to handle the branch $\bar{W} < A/4$, where $\widehat{\theta} = 0$. Since $\mu_W \geq A/2$,

$$\mu_W - \bar{W} \geq A/4, \qquad \text{so} \qquad A^2 \leq 16(\bar{W} - \mu_W)^2 = 16\Delta_W^2.$$

Also $|\theta| \leq C_\theta$, hence $\theta^2 \leq C_\theta^2$. Therefore

$$(\widehat{\theta} - \theta)^2 = \theta^2 \leq C_\theta^2 \leq \frac{16}{A^2}C_\theta^2\Delta_W^2 \leq \frac{16}{A^2}\left\{\bar{R}^2 + C_\theta^2\Delta_W^2\right\}.$$

The two branches give the claimed pathwise bound. □

Proof of Theorem 5. 1. Fix $C_\theta, C_g, C_q, \psi_\xi$. By Lemma 7, there is a constant $C > 0$, depending only on these four displayed constants, such that whenever an admissible parameter record has primitive entries $C_\theta, C_g, C_q, \psi_\xi$ equal to the displayed constants and cumulant order $k = 4$, every $a \in (0, 1]$ satisfying the Gaussian–Rademacher law assumption on η, the sampling and model restrictions, and the $L^1(P_X)$ treatment-code radius satisfies the full Gaussian–Rademacher path conclusion with constant C. Applying this rule to the theorem's records with $r = 3$ and $k = 4$ gives the first assertion.

2. Now fix a published ACE handle, $\gamma \in \mathbb{R}$, and $C_\gamma \in \mathbb{R}$, and assume that the handle satisfies the theorem's published ACE interface at (γ, C_γ). Unpacking that hypothesis gives first

$$C_\gamma > 0.$$

It also gives the following specialization rule: for every parameter record p whose probability-level coordinate is γ, every real $\Delta > 0$, and every pair of strictly positive current code radii

$$\varepsilon_{1,n} > 0, \qquad \varepsilon_{2,n} > 0,$$

every supplied code pair $(\widehat{g}, \widehat{q})$ and every model satisfying the clipped-code identities, the local ACE class conditions at Δ, and the local JMS eligibility inequalities at Δ obey

$$Q_{P,1-\gamma}\Big(\Big|\widehat{\theta}_{\mathrm{ACE},r,n} - \theta_0(P)\Big|\Big) \le B_{\mathrm{ACE},n,\gamma}(\Delta; C_\gamma).$$

For the theorem, take

$$\Delta = \delta_n.$$

The standing assumption $\delta_n > 0$ supplies the required positivity of Δ. Together with the displayed strict positivity conditions $\varepsilon_{1,n} > 0$ and $\varepsilon_{2,n} > 0$, the theorem's clipped-code identities, local ACE class conditions, and local JMS eligibility inequalities are exactly the premises of the specialization rule at this Δ. Substitution therefore yields

$$Q_{P,1-\gamma}\Big(\Big|\widehat{\theta}_{\mathrm{ACE},r,n} - \theta_0(P)\Big|\Big) \le B_{\mathrm{ACE},n,\gamma}(\delta_n; C_\gamma),$$

where

$$B_{\mathrm{ACE},n,\gamma}(\Delta; C_\gamma) = C_\gamma\, r!\, 16^r\, \Delta^{-1}\left(\varepsilon_{1,n}^r \varepsilon_{2,n} + C_\theta \varepsilon_{1,n}^{r+1} + 64(C_g + \psi_\eta)^r\left(r^2(C_g + \psi_\eta) + \psi_\xi + C_\theta \psi_\eta\right)(\gamma n)^{-1/2}\right).$$

□

Proof of Proposition 2. 1. Fix a model m in the displayed Gaussian class, and write P for its observed-data law. Define

$$\eta = T - g_0(X), \qquad \xi = Y - q_0(X) - \theta_0(P)\eta.$$

The conditional mean restrictions in Definitions 2 and 4 give

$$E_P[\xi \mid X, T] = 0,$$

and $q_0(X)$ and $\theta_0(P)\eta$ are measurable with respect to (X, T). Hence conditional-expectation linearity yields

$$E_P[Y \mid X, T] = q_0(X) + \theta_0(P)\eta \qquad P\text{-a.s.}$$

The bounded-outcome condition gives $-C_q \le Y \le C_q$ P-a.s. Conditioning the two inequalities on (X, T) gives

$$-C_q \le E_P[Y \mid X, T] \le C_q \qquad P\text{-a.s.},$$

and therefore

$$\big|q_0(X) + \theta_0(P)\eta\big| \le C_q \qquad P\text{-a.s.}$$

Conditioning the same bounded-outcome inequalities on X, and using

$$E_P[Y \mid X] = q_0(X),$$

gives

$$|q_0(X)| \le C_q \qquad P\text{-a.s.}$$

Combining the two displayed bounds,

$$|\theta_0(P)\eta| = \big|(q_0(X) + \theta_0(P)\eta) - q_0(X)\big| \le \big|q_0(X) + \theta_0(P)\eta\big| + |q_0(X)| \le 2C_q \qquad P\text{-a.s.}$$

2. Suppose, for contradiction, that $\theta_0(P) \neq 0$, and define

$$\Lambda_\theta = \frac{2C_q}{|\theta_0(P)|} \in \mathbb{R}.$$

The previous step implies

$$|\eta| \leq \Lambda_\theta \qquad P\text{-a.s.},$$

so

$$P\{\eta > \Lambda_\theta\} = 0.$$

The Gaussian membership in Definition 4 gives $\eta \sim N(0, \sigma^2)$, and the parameter record in Definition 1 gives $\sigma > 0$. Thus the variance is nonzero and

$$N(0, \sigma^2)\big((\Lambda_\theta, \infty)\big) = 0.$$

For a Gaussian law with positive variance, Lebesgue measure is absolutely continuous with respect to that Gaussian law. Hence the last display implies

$$\lambda\big((\Lambda_\theta, \infty)\big) = 0.$$

But

$$\lambda\big((\Lambda_\theta, \infty)\big) = \infty.$$

This contradiction proves

$$\theta_0(P) = 0.$$

3. Now fix deterministic code sequences $\widehat{g}, \widehat{q}$ and assume the fixed-code Gaussian class in the statement is inhabited. Consider the estimator

$$\widetilde{\theta}_0(O_1, \ldots, O_n; \widehat{g}_n, \widehat{q}_n) = 0.$$

For every model in that fixed-code class, the first part gives $\theta_0(P) = 0$, and hence

$$E_{P^{\otimes n}}\Big[\{\widetilde{\theta}_0(O_1, \ldots, O_n; \widehat{g}_n, \widehat{q}_n) - \theta_0(P)\}^2\Big] = 0.$$

Taking the supremum over the fixed-code class and then the infimum over all estimators in Definition 9 gives

$$\mathfrak{R}_n^{\mathrm{G}}(p; \widehat{g}, \widehat{q}) \leq 0.$$

Since this risk is nonnegative by definition,

$$\mathfrak{R}_n^{\mathrm{G}}(p; \widehat{g}, \widehat{q}) = 0.$$

4. It remains to evaluate the generalized-quantile criterion for the same zero estimator. Since $\gamma \in (1/2, 1)$,

$$0 < 1 - \gamma < 1.$$

For every model in the fixed-code Gaussian class,

$$\big|\widetilde{\theta}_0(O_1, \ldots, O_n; \widehat{g}_n, \widehat{q}_n) - \theta_0(P)\big| = 0 \qquad P^{\otimes n}\text{-a.s.}$$

Thus the law of this loss is the point mass at zero. Its distribution function is 0 for negative thresholds and 1 at every nonnegative threshold, so the generalized lower $(1-\gamma)$-quantile of Definition 6 is

$$Q_{P,1-\gamma}(0) = 0.$$

Consequently the supremum of the fixed-code quantile loss for the zero estimator is 0, and the minimax infimum is at most 0. Nonnegativity in Definition 9 gives

$$\mathfrak{M}^{\mathrm{G}}_{n,1-\gamma}(p; \widehat{g}, \widehat{q}) = 0.$$

$\square$

Proof of Lemma 5. Fix the parameter record, an index n, and a model m. For every code index j, the clipped-code discrepancies

$$x \longmapsto \bar{g}_j(x) - g_0(x), \qquad x \longmapsto \bar{q}_j(x) - q_0(x)$$

are P_X-measurable, because the supplied codes and the regressions are measurable and clipping preserves measurability. The covariate marginal P_X is a probability law, since it is the image of the observed-data probability law under the covariate map.

Assume first that $m \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}$. The positive class-index condition and the structural, range, tail, and cumulant requirements of $\mathcal{P}_{\mathrm{NG},n}$ are the corresponding member conditions in Definition 5. The treatment-code requirement in Definition 3 follows from the JMS ACE treatment radius. Put

$$f(x) = \bar{g}_n(x) - g_0(x).$$

The JMS ACE radius bound gives a finite $L^r(P_X)$ bound for f. Since P_X is a probability law and $2 \le r$, exponent monotonicity gives $f \in L^1(P_X)$ and

$$\|f\|_{L^1(P_X)} \le \|f\|_{L^r(P_X)}.$$

Together with $\|f\|_{L^r(P_X)} \le \varepsilon_{1,n}$ and the nonnegativity of $\varepsilon_{1,n}$ at the positive index n, this yields

$$\int |\bar{g}_n(x) - g_0(x)|\, dP_X(x) \le \varepsilon_{1,n},$$

with the displayed integrand integrable. Thus $m \in \mathcal{P}_{\mathrm{NG},n}$.

Next assume $m \in \mathcal{P}^{\mathrm{ACE}}_{\mathrm{NG},n}$. The inherited non-Gaussian member conditions supply the positive class-index condition and the common structural, range, tail, and cumulant requirements appearing in Definition 5. The subclass restrictions in Definition 11 give

$$\|\bar{g}_n - g_0\|_{L^s(P_X)} \le \varepsilon_{1,n}, \qquad \|\bar{q}_n - q_0\|_{L^s(P_X)} \le \varepsilon_{2,n}.$$

Because $r \le s$ and P_X is a probability law, exponent monotonicity gives

$$\|\bar{g}_n - g_0\|_{L^r(P_X)} \le \|\bar{g}_n - g_0\|_{L^s(P_X)}, \qquad \|\bar{q}_n - q_0\|_{L^r(P_X)} \le \|\bar{q}_n - q_0\|_{L^s(P_X)}.$$

These are the two JMS ACE code-radius requirements, so $m \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}$.

Finally suppose $s = r$. The preceding paragraph gives

$$m \in \mathcal{P}^{\mathrm{ACE}}_{\mathrm{NG},n} \implies m \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}.$$

Conversely, if $m \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n}$, the first inclusion gives $m \in \mathcal{P}_{\mathrm{NG},n}$, and the two JMS ACE $L^r(P_X)$ code bounds become exactly the two $L^s(P_X)$ subclass bounds in Definition 11. Hence

$$m \in \mathcal{P}^{\mathrm{JMS}}_{\mathrm{ACE},n} \quad\Longleftrightarrow\quad m \in \mathcal{P}^{\mathrm{ACE}}_{\mathrm{NG},n}.$$

$\square$

Proof of Lemma 7. Choose

$$K = 4(C_q + 2C_\theta C_g)^2 + 4\psi_\xi^2 + C_\theta^2(8 + 16C_g^2), \qquad C = 1 + 64K + 4C_\theta^2.$$

Then $C > 0$. Fix p, a, m satisfying the assumptions, and write

$$t = t_a = \frac{\pi}{2a}, \qquad A = A_a = a\exp\left\{-\frac{(1-a^2)\pi^2}{8a^2}\right\}.$$

Since $0 < a \le 1$, $t > 0$ and $A > 0$.

Independence of $G_\circ$ and S gives the product transform

$$M_a(z) = E\exp\{z\sqrt{1-a^2}G_\circ\}\, E\exp\{zaS\} = \exp\left(\frac{1-a^2}{2}z^2\right)\cosh(az).$$

The fourth logarithmic derivative at zero of this transform is the Rademacher contribution, hence

$$\kappa_4(\eta_a) = -2a^4.$$

On the imaginary axis,

$$M_a(iu) = \exp\left(-\frac{1-a^2}{2}u^2\right)\cos(au).$$

The exponential factor is nonzero. Therefore the first positive zero occurs when $au = \pi/2$, namely

$$t_a = \frac{\pi}{2a}, \qquad M_a(it_a) = 0, \qquad M_a(iu) \neq 0 \quad (0 < u < t_a).$$

For any $d \in \mathbb{R}$,

$$E_P[\sin\{t_a(\eta + d)\}] = \Im\left(e^{it_a d}E_P[e^{it_a\eta}]\right) = \Im\left(e^{it_a d}M_a(it_a)\right) = 0.$$

Let

$$Z_n = T - \bar{g}_n(X), \qquad D_n = g_0(X) - \bar{g}_n(X),$$

so that

$$Z_n = \eta + D_n(X).$$

The derivative of the transform at the first zero is

$$M_a'(it_a) = iA_a,$$

equivalently

$$E_P\left[\eta\, e^{it_a\eta}\right] = iA_a.$$

Using $Z_n = \eta + D_n(X)$, independence of η and X, and $M_a(it_a) = 0$,

$$
\begin{aligned}
E_P[Z_n e^{it_a Z_n}] &= E_P[\eta e^{it_a \eta}]\, E_P[e^{it_a D_n(X)}] + E_P[e^{it_a \eta}]\, E_P[D_n(X) e^{it_a D_n(X)}] \\
&= iA_a\, E_P[e^{it_a D_n(X)}].
\end{aligned}
$$

Taking imaginary parts gives

$$E_P[Z_n \sin(t_a Z_n)] = A_a E_P[\cos(t_a D_n)].$$

Assume now that $t_a \varepsilon_{1,n} \le 1/2$. The $L^1(P_X)$ condition gives

$$E_P|D_n(X)| \le \varepsilon_{1,n}.$$

Since $1 - \cos u \le |u|$,

$$E_P[\cos(t_a D_n)] \ge 1 - t_a E_P|D_n| \ge 1 - t_a \varepsilon_{1,n} \ge \frac{1}{2}.$$

Together with the preceding identity and $A_a > 0$, this yields

$$E_P[Z_n \sin(t_a Z_n)] \ge A_a/2.$$

Define the two scores

$$W = Z_n \sin(t_a Z_n), \qquad R = (Y - \theta_0 Z_n) \sin(t_a Z_n).$$

The zero-shift identity and the conditional mean condition imply

$$E_P[R] = 0.$$

Indeed, writing

$$Y - \theta_0 Z_n = q_0(X) - \theta_0 D_n(X) + \xi,$$

the covariate term is killed by the sine-shift identity after conditioning on X, while

$$E_P[\xi \sin(t_a Z_n)] = E_P[\sin(t_a Z_n) E_P(\xi \mid X, T)] = 0.$$

Thus

$$E_P[Y \sin(t_a Z_n)] = \theta_0\, E_P[Z_n \sin(t_a Z_n)].$$

The Gaussian–Rademacher path has the second-moment envelope

$$E_P[\eta^2] \le 4.$$

With $|g_0| \le C_g$ and $|\bar{g}_n| \le C_g$, we have

$$|D_n| \le 2C_g,$$

and therefore

$$E_P[W^2] \le 8 + 16C_g^2.$$

Similarly, the range conditions give

$$|q_0(X) - \theta_0 D_n(X)| \le C_q + 2C_\theta C_g,$$

and the conditional Luxemburg bound for ξ gives

$$E_P[\xi^2] \le 2\psi_\xi^2.$$

Consequently

$$E_P[R^2] \le 4(C_q + 2C_\theta C_g)^2 + 4\psi_\xi^2.$$

The two displayed square-moment bounds are exactly the components entering K.

Let

$$\overline{W} = \mathbb{P}_n W, \qquad \overline{R} = \mathbb{P}_n R, \qquad \mu_W = E_P W.$$

The estimator in the statement is the clipped ratio based on

$$\mathbb{P}_n[Y \sin(t_a Z_n)] = \theta_0 \overline{W} + \overline{R}$$

with threshold $A_a/4$. Since $\mu_W \ge A_a/2$, Lemma 6 gives the pathwise inequality

$$(\widehat{\theta}_{a,n} - \theta_0)^2 \le \frac{16}{A_a^2} \left\{ \overline{R}^2 + C_\theta^2 (\overline{W} - \mu_W)^2 \right\}.$$

Taking expectation under the i.i.d. product law and using $E_P R = 0$ gives

$$E_P[(\widehat{\theta}_{a,n} - \theta_0)^2] \le \frac{16}{nA_a^2} \left(E_P[R^2] + C_\theta^2 E_P[W^2] \right) \le \frac{16K}{nA_a^2}.$$

The clipping range also gives the deterministic bound

$$(\widehat{\theta}_{a,n} - \theta_0)^2 \le 4C_\theta^2.$$

Combining this with the preceding display and using

$$A_a^{-2} = a^{-2} \exp\!\left(\frac{(1-a^2)\pi^2}{4a^2} \right)$$

yields

$$E_P[(\widehat{\theta}_{a,n} - \theta_0)^2] \le C \min\!\left\{ 1, \frac{1}{n} a^{-2} \exp\!\left(\frac{(1-a^2)\pi^2}{4a^2} \right) \right\}.$$

The smallness condition is equivalent to the stated radius bound. Since

$$t_a \varepsilon_{1,n} \le \frac{1}{2} \quad \iff \quad \varepsilon_{1,n} \le \frac{a}{\pi},$$

and

$$\Delta_a = 2a^4, \qquad \Delta_a^{1/4} = 2^{1/4} a,$$

we obtain

$$\varepsilon_{1,n} \le \frac{\Delta_a^{1/4}}{2^{1/4}\pi}.$$

Finally,

$$\Delta_a^{-1/2} = (2a^4)^{-1/2} = (\sqrt{2}\, a^2)^{-1}.$$

The first risk factor is bounded by the cumulant-parametrized factor:

$$a^{-2} \exp\!\left(\frac{(1-a^2)\pi^2}{4a^2} \right) \le \Delta_a^{-1/2} \exp\!\left(\frac{\pi^2\sqrt{2}}{4} \Delta_a^{-1/2} \right).$$

Substituting this inequality into the previous risk display gives

$$E_P[(\widehat{\theta}_{a,n} - \theta_0)^2] \le C \min\left\{1, \frac{1}{n}\Delta_a^{-1/2} \exp\left(\frac{\pi^2\sqrt{2}}{4}\Delta_a^{-1/2}\right)\right\}.$$

□

References

Donald W.K. Andrews. Examples of ¡mml:math xmlns:mml=”http://www.w3.org/1998/math/mathml” id=”mml1” display=”inline” overflow=”scroll” altimg=”si1.gif”¿¡mml:msup¿¡mml:mrow¿¡mml:mi¿l¡/mml:mi¿ complete and boundedly-complete distributions. *Journal of Econometrics*, 199(2):213–220, 2017. doi: 10.1016/j.jeconom.2017.05.011.

Alexandre Belloni, Victor Chernozhukov, and Christian Hansen. Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, 81(2):608–650, 2014. doi: 10.1093/restud/rdt044.

Peter J. Bickel, Chris A. J. Klaassen, Ya'acov Ritov, and Jon A. Wellner. *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press, Baltimore, 1993. URL https://www.stat.washington.edu/jaw/RESEARCH/BOOKS/book3.html.

Marine Carrasco. Efficient estimation using the characteristic function. *Econometric Theory*, 33(2):479–526, 2017. doi: 10.1017/S0266466615000520.

Marine Carrasco, Jean-Pierre Florens, and Eric Renault. Chapter 77 linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. In James J. Heckman and Edward E. Leamer, editors, *Handbook of Econometrics*, volume 6B, pages 5633–5751. Elsevier, 2007. doi: 10.1016/S1573-4412(07)06077-1.

George Chacko and Luis M. Viceira. Spectral gmm estimation of continuous-time processes. *Journal of Econometrics*, 116(1–2):259–292, 2003. doi: 10.1016/S0304-4076(03)00109-X.

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.

Victor Chernozhukov, Juan Carlos Escanciano, Hidehiko Ichimura, Whitney K. Newey, and James M. Robins. Locally robust semiparametric estimation. *Econometrica*, 90(4):1501–1535, 2022. doi: 10.3982/ECTA16294.

Pierre Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3): 287–314, 1994. doi: 10.1016/0165-1684(94)90029-9.

Serge Darolles, Yanqin Fan, Jean-Pierre Florens, and Eric Renault. Nonparametric instrumental regression. *Econometrica*, 79(5):1541–1565, 2011. doi: 10.3982/ECTA6539.

Xavier D'Haultfoeuille. On the completeness condition in nonparametric instrumental problems. *Econometric Theory*, 27(3):460–471, 2011. doi: 10.1017/S0266466610000368.

Tien-Cuong Dinh, Subhroshekhar Ghosh, Hoang-Son Tran, and Manh-Hung Tran. Gaussian fluctuations for spin systems and point processes: near-optimal rates via quantitative marcinkiewicz's theorem. *arXiv preprint*, 2021.

Robert F. Engle, C. W. J. Granger, John Rice, and Andrew Weiss. Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association*, 81(394):310–320, 1986. doi: 10.1080/01621459.1986.10478274.

Alexandre Eremenko and Alexander Fryntov. Stability in the marcinkiewicz theorem. *Zhurnal Matematicheskoi Fiziki, Analiza, Geometrii*, 17(4):463–467, 2021. doi: 10.15407/mag17.04.463.

Andrey Feuerverger and Philip McDunnough. On the efficiency of empirical characteristic function procedures. *Journal of the Royal Statistical Society: Series B (Methodological)*, 43(1):20–27, 1981. URL https://utstat.toronto.edu/andrey/pubs/ECFefficiency.pdf.

Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054, 1982. doi: 10.2307/1912775.

Wolfgang Härdle and Enno Mammen. Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, 21(4):1926–1947, 1993. doi: 10.1214/aos/1176349403.

Lukas Hoesch, Adam Lee, and Geert Mesters. Locally robust inference for non-gaussian svar models. *Quantitative Economics*, 15(2):523–570, 2024. doi: 10.3982/QE2274.

Yingyao Hu and Ji-Liang Shiu. A simple test of completeness in a class of nonparametric specification. *Econometric Reviews*, 41(4):373–399, 2022. doi: 10.1080/07474938.2021.1957285.

Aapo Hyvärinen and Erkki Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483–1492, 1997. doi: 10.1162/neco.1997.9.7.1483.

Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent Component Analysis*. Wiley, New York, 2001. ISBN 9780471405405. doi: 10.1002/0471221317.

Wolfgang Härdle, Hua Liang, and Jiti Gao. *Partially Linear Models*. Contributions to Statistics. Physica-Verlag, Heidelberg, 2000. doi: 10.1007/978-3-642-57700-0.

Jikai Jin, Lester Mackey, and Vasilis Syrgkanis. It's hard to be normal: The impact of noise on structure-agnostic estimation. *Advances in Neural Information Processing Systems*, 2025.

A. M. Kagan, Yu. V. Linnik, and C. R. Rao. *Characterization Problems in Mathematical Statistics*. Wiley, New York, 1973. ISBN 9780471454212.

Ker-I Ko. *Complexity Theory of Real Functions*. Progress in Theoretical Computer Science. Birkhäuser, Boston, 1991. ISBN 9781468468045. doi: 10.1007/978-1-4684-6802-1.

Markku Lanne, Mika Meitz, and Pentti Saikkonen. Identification and estimation of non-gaussian structural vector autoregressions. *Journal of Econometrics*, 196(2):288–304, 2017. doi: 10.1016/j.jeconom.2016.06.002.

Adam Lee and Geert Mesters. Locally robust inference for non-gaussian linear simultaneous equations models. *Journal of Econometrics*, 240(1):105647, 2024. doi: 10.1016/j.jeconom.2023.105647.

Lin Liu, Rajarshi Mukherjee, James Robins, and Eric Tchetgen Tchetgen. Adaptive estimation of nonparametric functionals. *Journal of Machine Learning Research*, 22(99):1–66, 2016.

Lin Liu, Rajarshi Mukherjee, Whitney K. Newey, and James M. Robins. Semiparametric efficient empirical higher order influence function estimators. *arXiv preprint*, 2017.

Lester Mackey, Vasilis Syrgkanis, and Ilias Zadik. Orthogonal machine learning: Power and limitations. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 3375–3383, 2018. URL https://proceedings.mlr.press/v80/mackey18a.html.

Józef Marcinkiewicz. Sur une propriété de la loi de gauß. *Mathematische Zeitschrift*, 44(1): 612–618, 1939. doi: 10.1007/BF01210677.

Lutz Mattner. Completeness of location families, translated moments, and uniqueness of charges. *Probability Theory and Related Fields*, 92:137–149, 1992. doi: 10.1007/BF01194918.

Marcus Michelen and Julian Sahasrabudhe. Central limit theorems from the roots of probability generating functions. *Advances in Mathematics*, 358:106840, 2019. doi: 10.1016/j.aim.2019.106840.

Ramon E. Moore. *Interval Analysis*. Prentice-Hall, Englewood Cliffs, NJ, 1966. URL https://books.google.com/books?id=csQ-AAAAIAAJ.

Whitney K. Newey. Semiparametric efficiency bounds. *Journal of Applied Econometrics*, 5(2): 99–135, 1990. doi: 10.1002/jae.3950050202.

Whitney K. Newey. The asymptotic variance of semiparametric estimators. *Econometrica*, 62 (6):1349–1382, 1994. doi: 10.2307/2951752.

Whitney K. Newey and Daniel McFadden. Chapter 36 large sample estimation and hypothesis testing. In Robert F. Engle and Daniel L. McFadden, editors, *Handbook of Econometrics*, volume 4, pages 2111–2245. Elsevier, 1994. doi: 10.1016/S1573-4412(05)80005-4.

Whitney K. Newey and James M. Robins. Cross-fitting and fast remainder rates for semiparametric estimation. *arXiv preprint*, 2018.

Patrik Reizinger, Lester Mackey, Wieland Brendel, and Rahul G. Krishnan. Estimating treatment effects with independent component analysis. *arXiv preprint*, 2025.

James Robins, Lingling Li, Eric Tchetgen, and Aad W. van der Vaart. Higher order influence functions and minimax estimation of nonlinear functionals. In *Probability and Statistics: Essays in Honor of David A. Freedman*, volume 2 of *Institute of Mathematical Statistics Collections*, pages 335–421. Institute of Mathematical Statistics, 2008. doi: 10.1214/193940307000000527.

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.

P. M. Robinson. Root-n-consistent semiparametric regression. *Econometrica*, 56(4):931–954, 1988. doi: 10.2307/1912705.

Kenneth J. Singleton. Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics*, 102(1):111–141, 2001. doi: 10.1016/S0304-4076(00)00092-0.

Paul Speckman. Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(3):413–436, 1988. URL https://www.jstor.org/stable/2345705.

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.

Aad W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, Cambridge, 1998. ISBN 9780521496032.

Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes*. Springer, New York, 1996. doi: 10.1007/978-1-4757-2545-2.

Roman Vershynin. *High-Dimensional Probability*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, 2018. ISBN 9781108415194. doi: 10.1017/9781108231596.

Martin J. Wainwright. *High-Dimensional Statistics*, volume 48 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, 2019. ISBN 9781108498029. doi: 10.1017/9781108627771.

Klaus Weihrauch. *Computable Analysis*. Texts in Theoretical Computer Science. Springer, Berlin, 2000. ISBN 9783540668176. doi: 10.1007/978-3-642-56999-9.