On Cofactuals

Chae-Yeon Xon
Yonsei University

September 2026

Abstract

The distribution of policy gains depends on how outcomes are paired across regimes. I study a shared-component model in which policy labels fix the loadings of independent interaction terms. Under local exponential integrability, the marginal laws admit this model if and only if their Möbius-transformed cumulant generating functions are admissible; the same transforms construct the joint law within the class. A declared collection of k regimes requires only its intersection closure, containing at most $2^k - 1$ worlds. For two regimes, their two marginals and intersection marginal determine the full gain distribution. A row-space condition shows which marginal cumulants identify a declared contrast, and a Poisson construction gives necessity at every order. For several regimes, the endpoints and distinct pairwise intersections form the unique inclusion-minimal bundle for all pairwise odd and even gain cumulants, including under heterogeneous positive costs. Transform admissibility is equivalent to a hierarchy of positive-semidefinite matrices, yielding a consistent sieve and finite-sample size control under bounded or exponentially integrable unbounded outcomes. With paired observations, a joint-CGF defect records the complete departure of the gain law from its shared-component reference and bounds every cumulant error. A global positive-definiteness criterion gives the attained set of gain laws within a declared sensitivity band. In a discounted decision model, an action-gap condition preserves the optimal policy under reward and transition perturbations; a common-resolvent bound and a margin condition quantify departures. Paired and state-adjusted moment procedures assess selected restrictions implied by the model.

JEL Codes: C10, C18, C61

Keywords: counterfactuals, policy-gain distributions, structural econometrics, Möbius inversion, specification testing

Yonsei University. Email: eunixon@yonsei.ac.kr. This draft is circulated for discussion; comments and suggestions are welcome.

1. Introduction

How dispersed are the gains from a policy? To study this question, consider outcomes for the same population under a baseline and an alternative. Their pairing matters for the dispersion of individual gains. For baseline outcome $Y_\varnothing$ and policy outcome Y_1,

$$\mathrm{Var}(Y_1 - Y_\varnothing) = \mathrm{Var}(Y_1) + \mathrm{Var}(Y_\varnothing) - 2\,\mathrm{Cov}(Y_1, Y_\varnothing). \tag{1}$$

The two variances can be learned from marginal information. The covariance depends on which outcomes belong to the same unit or structural draw. Thus the same marginal laws can support different accounts of gain heterogeneity (Firpo and Ridder 2019; Gao et al. 2025).

I study a shared-component restriction on this pairing. The two-regime case is a useful place to begin. Write $Y_\varnothing = C_\varnothing$ and $Y_1 = C_\varnothing + C_1$, with the baseline component $C_\varnothing$ independent of the increment C_1. If their moment generating functions are finite near zero, the marginal cumulant generating functions satisfy

$$K_{C_1}(t) = K_1(t) - K_\varnothing(t). \tag{2}$$

The condition has a simple interpretation: the difference on the right must be the cumulant generating function of an increment law. When it is, that law gives the distribution of the gain under the maintained independence restriction. For Gaussian marginals with means $m_\varnothing, m_1$ and variances $v_\varnothing, v_1$, compatibility reduces to $v_1 \geq v_\varnothing$. The gain law is then $N(m_1 - m_\varnothing, v_1 - v_\varnothing)$, including a degenerate increment at equality. The chosen baseline fixes the direction of this restriction.[1]

With several policy switches, each subset labels a regime. I call the resulting baseline-anchored interaction terms *cofactual components*.[2] For two switches, $C_i = Y_i - Y_\varnothing$ and $C_{12} = Y_{12} - Y_1 - Y_2 + Y_\varnothing$. Möbius inversion writes each outcome as the sum of the components whose labels it contains. The policy labels fix the loadings across regimes. Independence then supplies the restriction on cross-world dependence.

Main result. Theorem 1 gives a marginal-to-joint characterization. Under local exponential integrability, the marginal laws admit an independent-component coupling if and only if every Möbius-transformed CGF is the CGF of a probability law. When the condition holds, those transforms construct the unique component laws and the unique joint law within the factorized class. Proposition 2 writes admissibility as a hierarchy

[1]Reversing the baseline applies the same construction with the regime roles exchanged and can give a different compatibility condition.

[2]The label refers to baseline-anchored Möbius interaction terms. Whether a regime is factual or counterfactual depends on the observation design and the reference case.

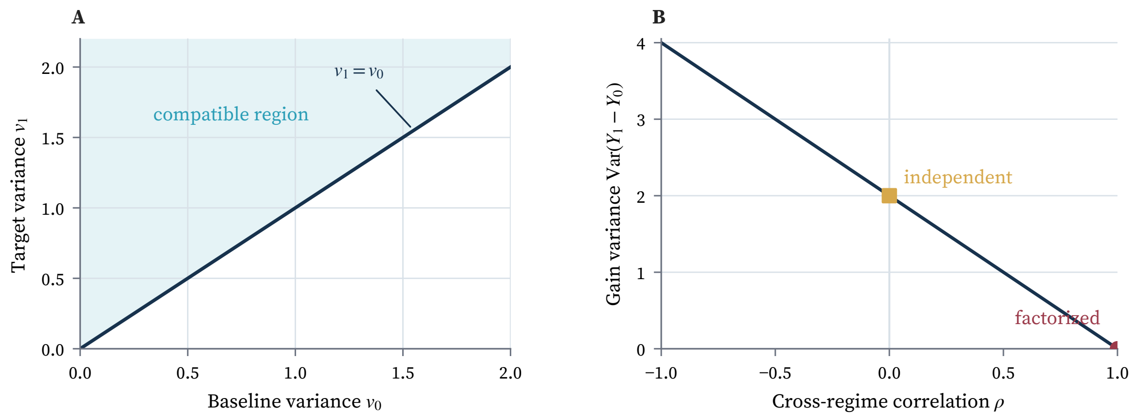

FIGURE 1. Marginal compatibility and coupling-sensitive gain dispersion

Panel A gives the factorized Gaussian compatibility region for baseline variance v_0 and target variance v_1; the diagonal gives a degenerate increment. Panel B holds both marginal variances at one and plots $\mathrm{Var}(Y_1 - Y_0) = 2(1-\rho)$ against the cross-regime correlation. The factorized equal-variance model corresponds to $\rho = 1$, while an independent pairing corresponds to $\rho = 0$. The comparison is algebraic.

of positive-semidefinite matrices, Theorem 2 gives a consistent expanding-grid check from separate samples, and Corollaries 1–2 give finite-sample size control for bounded outcomes and for unbounded outcomes with a uniform exponential-moment envelope.

The full policy cube provides all comparisons simultaneously. Theorem 3 begins with a declared collection of k regimes and uses only its intersection closure, containing at most $2^k - 1$ distinct worlds. For two regimes S and T, the marginals of S, T, and $S \cap T$ construct the joint reference law and the full distribution of $Y_S - Y_T$. Theorem 6 then treats observation choice directly: a contrast defines a loading on component cumulants, observed worlds span available incidence rows, and membership of the loading in that row space gives reconstruction. A Poisson construction gives the converse at every cumulant order. For a two-world gain, odd cumulants require exactly the endpoint marginals. An even cumulant requires two marginals under nesting and three for incomparable worlds. For several regimes, Corollary 7 shows that every identifying marginal design contains the distinct endpoints and pairwise intersections. This bundle is therefore the unique inclusion-minimal design and minimizes every positive additive acquisition cost.

For illustration, equal standard-normal marginals yield a zero gain under factorization and gain variance two under independent pairing. The marginal evidence agrees across the two constructions; the maintained coupling selects the gain distribution. More generally, Theorem 7 compares an observed paired law with the factorized reference through one joint-CGF defect. Its anti-diagonal is the exact error in the gain CGF, and its derivatives give cumulant errors together with order-specific bounds. Theorem 8 then uses global characteristic-function positive definiteness to characterize and attain every gain law consistent with fixed endpoint marginals and a declared sensitivity band.

Independent reward modules provide one economic source of the restriction. A common policy-evaluation operator propagates the modules separately through a fixed Bellman resolvent. Theorem 4 allows agents to reoptimize: a baseline action gap and a bound on reward and transition perturbations preserve the optimal policy. A common-resolvent bound measures the distance from the factorized reference, and Corollary 5 converts a distribution of action gaps into a bound on policy switching and gain-law distance. A common pre-intervention state can also account for shared observable heterogeneity, with the restriction applied within each state.[3]

The distributional result also has lower-order implications. At second order, orthogonality of components is equivalent to

$$\mathrm{Cov}(Y_S, Y_T) = \mathrm{Var}(Y_{S\cap T}). \tag{3}$$

The intersection records the components shared by the two regimes. Under orthogonality, the variance surface has nonnegative Möbius coefficients; higher cumulants satisfy corresponding intersection identities. For a general coupling, a union identity shows which component covariances enter a marginal variance contrast. The variance-error bound and reward-credit calculation trace their effect on the policy target.

Observation and assessment. Separate regime samples can supply the marginal transforms under an appropriate assignment and support design. The semidefinite sieve consistently detects fixed failures of marginal compatibility on its transform interval. Simultaneous transform bands give finite-sample control uniformly over the positive-semidefinite cone, including its boundary, under either a known outcome bound or a known exponential-moment envelope. Paired rows record how outcomes are joined and permit assessment of selected restrictions on that joint law. The paired procedures below cover covariance equalities and state-adjusted residual moments for a fixed intervention family. Their calibration uses the recorded pairing.[4]

Related literature. Chernozhukov et al. (2013) develop inference for marginal counterfactual distributions. Here those distributions are the population inputs, and the object to recover is their cross-regime joint law. Firpo and Ridder (2019) and Frandsen and Lefgren (2021) bound treatment-effect distributions over classes of marginal-compatible couplings, and Gao et al. (2025) formulates such problems as multi-marginal optimal transport. Wu and Mao (2026) use variation across experiments with transportability and rank conditions to identify joint potential-outcome distributions. The present paper studies a specified

[3] Conditional independence within the recorded state permits unconditional dependence generated by variation in that state.

[4] Examples are a simulator's common draw and an empirical pairing whose unit-level interpretation is supported by the observation design.

source of identifying information: independence of interaction components whose loadings are fixed by policy labels. Theorem 1 characterizes the marginal inputs that this restriction admits and the joint law it selects.

Compatibility of marginal objects also arises outside counterfactual analysis. Wang (2004) use Lancaster interactions to characterize completion of overlapping marginal densities by an all-positive joint density. Azrieli and Rehbeck (2026) characterize consistency of marginal menu and choice frequencies through the core of a totally monotone game and show that random utility adds no restriction on their unrestricted marginal dataset. These results concern feasibility of their respective marginal systems. The factorized class here fixes Boolean component loadings and turns admissible cumulant transforms into a unique joint law within that class.

McKay and Wolf (2023) ask which causal responses suffice for policy-rule counterfactuals in linearized structural models. Their policy-rule problem and the fixed-regime pairing problem here use different observed objects and maintained restrictions. Christensen and Connault (2023) bound counterfactuals over neighborhoods of latent distributions subject to model constraints. The defect here fixes a factorized reference from selected marginals and records how the paired gain law departs from it. The global sensitivity result applies Bochner's positive-definite representation (Berg et al. 1984) with fixed marginal axes and a constrained gain anti-diagonal; the representation theorem is imported, while the restrictions arise from the declared pair target. Escanciano and Li (2013) characterize identified sets for continuous linear functionals in a nonparametric instrumental-variable problem. The finite row-space result here operates on Boolean component-cumulant loadings and identifies which policy-world marginals recover a declared contrast.

The coordinates are classical Möbius inversion, related to Harsanyi dividends, factorial contrasts, and causal or attribution decompositions (Rota 1964; Harsanyi 1963; Dasgupta et al. 2015; Jansma 2025; Forré and Jansma 2025; Gao and Zhao 2025). Shared-source and convolution constructions provide related probabilistic representations (Schulz et al. 2021; Balan 2002); the covariance here is a Boolean meet matrix (Mattila and Haukkanen 2014). Balan (2002) constructs set-indexed independent-increment processes from convolution systems on finite semilattices. The target-closure result uses the policy intersections generated by a declared counterfactual target and ties their source transforms to recoverability from marginal world laws. Mesters and Zwiernik (2024) study identification of a mixing matrix using restrictions on component moments and cumulants.[5] Möbius inversion and product construction supply the algebra; the analysis connects them to target-specific marginal acquisition and gain-law reconstruction.

For assessment, I use covariance-testing and semiparametric methods (Shah and Peters 2020; Scheidegger et al. 2022; Chernozhukov et al. 2018, 2022) for the moments

[5]Their loading matrix is an unknown object. Here the intervention labels fix the loading matrix before the component restrictions are imposed.

implied by the coupling restriction. Amengual et al. (2022) test moment implications of independent latent components while accounting for estimated shocks. Here the intervention labels fix an invertible transform, so paired world outcomes reveal the components directly once all worlds are recorded. The appendix also studies precision when paired evaluations are costly, taking the diagnostic and required world bundle as given.

The reward-credit example carries the component law, its paired defect, and the resulting policy-gain dispersion through one calculation. Proofs, inference details, and numerical exhibits accompany the main results.

2. Cofactual coordinates

Fix a baseline and $p \geq 1$ labeled interventions. Let $\mathcal{W} = 2^{[p]}$, ordered by inclusion, and write $n = 2^p$ for the number of regimes. Each set denotes a joint regime with a fixed meaning throughout the family. We use Y_i for $Y_{\{i\}}$ and Y_{12} for $Y_{\{1,2\}}$; two-regime examples also write $Y_0 = Y_\varnothing$. History-dependent interventions require a richer index.

DEFINITION 1 (Counterfactual system). *On a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, a finite counterfactual system is a family*

$$Y = \{Y_S : S \in \mathcal{W}\},$$

where each $Y_S : \Omega \to \mathbb{R}^d$ *is measurable. The coordinate* $Y_\varnothing$ *is the reference target. When the actual regime is the baseline, it is the factual outcome; the remaining coordinates are counterfactual outcomes relative to that regime.*

The target may be a potential outcome, policy value, equilibrium moment, or welfare vector. The common ω is the unit, type, or structural draw held fixed across regimes. It specifies what it means to compare outcomes for the same underlying case. A simulation implements the resulting coupling; an empirical application gives it content through the model and observation design.

DEFINITION 2 (Cofactual component). *For* $A \subseteq [p]$*, define*

$$C_A = \sum_{B \subseteq A} (-1)^{|A|-|B|} Y_B. \tag{4}$$

For $A \neq \varnothing$*,* C_A *is the* cofactual component *associated with* A*. The coordinate* $C_\varnothing = Y_\varnothing$ *is the reference component.*

PROPOSITION 1 (Cofactual representation). *For every* $S \subseteq [p]$*,*

$$Y_S = \sum_{A \subseteq S} C_A. \tag{5}$$

The collection $\{C_A\}_{A \subseteq [p]}$ *is the unique collection with this property.*

PROOF. Substitution and a change in the order of summation give

$$\sum_{A \subseteq S} C_A = \sum_{B \subseteq S} Y_B \sum_{A : B \subseteq A \subseteq S} (-1)^{|A|-|B|}.$$

Writing $A = B \cup D$, the inner sum is $(1-1)^{|S \setminus B|}$. Only $B = S$ remains. Uniqueness follows by applying the same inversion formula to any representation of the form (5). □

Order the worlds consistently with inclusion and stack Y_S and C_A. Let

$$M_{A,B} = (-1)^{|A|-|B|}\mathbf{1}\{B \subseteq A\}, \qquad Z_{S,A} = \mathbf{1}\{A \subseteq S\}.$$

Then

$$C = MY, \qquad Y = ZC, \qquad MZ = ZM = I. \tag{6}$$

For vector targets, the stacked transform is $M \otimes I_d$, with inverse $Z \otimes I_d$. I suppress this identity factor in the scalar notation below. The original outcomes can be recovered from the components, so the change of coordinates retains the full system.

The coordinates also define a multilinear polynomial, $\mathscr{Y}_\omega(z) = \sum_A C_A(\omega) z^A$. Its values at Boolean vertices recover the regimes, and its mixed derivatives at zero recover the components. Interior values interpolate algebraically between regimes; interpreting them as randomized policies requires a randomization design. Under a degree restriction, the same representation yields compression formulas indexed by the number of policy switches in an interaction. The cumulant order describes a feature of its probability law.

3. Marginal compatibility and policy-gain distributions

The coordinate representation holds for every joint law. The question now is which marginal distributions admit a joint law with independent components. Throughout this distributional analysis, the target is scalar.

DEFINITION 3 (Cofactual factorization). *A scalar counterfactual system has a* cofactual factorization *when the family*

$$\{C_A : A \subseteq [p]\}$$

is mutually independent.

Under factorization, each C_A is a latent source shared by the worlds containing A. The root component enters every world. Independence is the substantive restriction on these sources and can be imposed conditional on a common state.

Factorization implies a marginal convolution identity with component laws $\mu_A = \mathcal{L}(C_A)$ such that

$$\mathcal{L}(Y_S) = \mathop{*}_{A \subseteq S} \mu_A. \tag{7}$$

This is a Boolean convolution system. The representation is close in spirit to set-indexed processes with independent increments (Balan 2002). Here the set index is a policy world, and the component laws will be used to recover cross-world dependence from separate world distributions.

REMARK 1 (Reference regime). *Factorization is indexed by the chosen factual root. With one intervention and square-integrable independent $C_\varnothing, C_1$, reversing the reference regime gives root $Y_1 = C_\varnothing + C_1$ and increment $-C_1$. Their covariance is $-\operatorname{Var}(C_1)$. Thus the direction of the intervention comparison belongs to the model specification. Relabeling interventions while retaining the root preserves the restriction.*

3.1. Marginal-to-joint characterization

ASSUMPTION 1 (Local exponential integrability). *The world marginal MGFs are finite on a common open interval containing zero.*

DEFINITION 4 (Admissible CGF). *A real-valued function on a neighborhood of zero is an admissible CGF when it is the cumulant generating function of a probability law with a finite MGF on that neighborhood.*

THEOREM 1 (Marginal-to-joint characterization). *Let $P = (P_S)_{S \in \mathcal{W}}$ be a family of scalar marginal laws satisfying Assumption 1, with CGFs K_S. Define*

$$H_A(t) = \sum_{B \subseteq A} (-1)^{|A|-|B|} K_B(t). \tag{8}$$

A factorized coupling of P exists if and only if every H_A is admissible on a common neighborhood of zero. Within this class, the component laws and joint law are unique. On their respective neighborhoods,

$$K_S(t) = \sum_{A \subseteq S} H_A(t), \tag{9}$$

$$K_Y^{\text{joint}}(t) = \sum_{A \subseteq [p]} H_A\left(\sum_{S \supseteq A} t_S\right). \tag{10}$$

Local transform domains. If the n marginal MGFs are finite for $|t| < \delta$, then for any compatible coupling and $\max_S |t_S| < \delta/n$, Hölder's inequality gives

$$\mathbb{E}\exp\left(\sum_S t_S Y_S\right) \le \prod_S \{\mathbb{E}\exp(n t_S Y_S)\}^{1/n} < \infty.$$

Thus every compatible joint law has a local joint MGF. A finite linear map, including M, retains this property on a smaller neighborhood.

PROOF. Suppose a factorized coupling exists. Since $Y_S = \sum_{A\subseteq S} C_A$, independence gives

$$\mathbb{E}[e^{tY_S}] = \prod_{A\subseteq S} \mathbb{E}[e^{tC_A}].$$

Taking logs gives (9); Boolean inversion gives (8). The local-domain argument above makes each component transform finite where used. Each H_A is therefore admissible.

Conversely, choose the probability law μ_A with local CGF H_A and construct the independent family $C_A \sim \mu_A$ on the finite product space. Set $Y_S = \sum_{A\subseteq S} C_A$. Its CGF is the sum in (9), which equals K_S by inversion. Local MGF uniqueness gives the marginal law P_S. The same uniqueness identifies every μ_A. Their product law is unique, and $Y = ZC$ determines its image. Finally,

$$\sum_S t_S Y_S = \sum_A C_A \sum_{S\supseteq A} t_S$$

proves (10) by independence. All identities are on a common neighborhood small enough for their arguments. This proves existence, reconstruction, and uniqueness within the stated class. □

Möbius inversion gives a candidate H_A for any marginal CGFs.[6] Admissibility asks whether that candidate is the CGF of a probability law. Normalization at zero and nonnegative second derivatives provide necessary checks. In general, the answer depends on the full local function; for Gaussian marginals it reduces to parameter restrictions.

The full restriction has a semidefinite form. Let $I = (-\delta, \delta)$ lie inside the common transform domain and define $\Lambda_A(t) = \exp\{H_A(t)\}$ for $t \in I$.

[6] For Möbius inversion on finite partially ordered sets, see Rota (1964); Chateauneuf and Jaffray (1989) develops its use for capacities and lower probabilities. Here inversion is applied to cumulant-generating functions, and admissibility supplies the probabilistic restriction.

PROPOSITION 2 (Semidefinite admissibility criterion). *The candidate H_A is admissible on I if and only if, for every integer $m \geq 1$ and every $t_1, \dots, t_m$ satisfying $t_i + t_j \in I$,*

$$\Gamma_A(t_1, \dots, t_m) = \left[\Lambda_A(t_i + t_j)\right]_{i,j=1}^{m} \succeq 0. \tag{11}$$

PROOF. The function Λ_A is continuous and $\Lambda_A(0) = 1$. Widder's representation theorem states that continuity and the matrix inequalities in (11) are equivalent to a bilateral Laplace representation by a nonnegative measure (Widder 1934; Buescu and Paixão 2018). Evaluation at zero makes its total mass one. Reflecting the integration variable gives $\Lambda_A(t) = \mathbb{E}[e^{tC_A}]$ on I for a probability law of C_A, which is equivalent to admissibility of H_A. □

The criterion replaces an existential transform condition with a hierarchy of matrices computed from the marginal CGFs. Finite grids give necessary restrictions. An expanding grid separates every fixed incompatible candidate because failure of exponential convexity has a finite witness.

For a sampling statement, suppose each world S supplies an independent sample $(Y_{S,i})_{i=1}^{N_S}$. Let $N_{\min} = \min_S N_S$ and define

$$\widehat{K}_S(t) = \log\left\{N_S^{-1}\sum_{i=1}^{N_S} e^{tY_{S,i}}\right\}, \qquad \widehat{H}_A(t) = \sum_{B\subseteq A}(-1)^{|A|-|B|}\widehat{K}_B(t), \qquad \widehat{\Lambda}_A(t) = e^{\widehat{H}_A(t)}.$$

Fix $\tau < \delta$, and let $\mathcal{T}_N = -\mathcal{T}_N \subset [-\tau/2, \tau/2]$ have cardinality m_N and mesh approaching zero. Write $\widehat{\Gamma}_{A,N} = [\widehat{\Lambda}_A(s+t)]_{s,t\in\mathcal{T}_N}$.

THEOREM 2 (Consistent marginal-compatibility sieve). *Suppose, for a deterministic $r_N \downarrow 0$,*

$$\max_{S\in\mathcal{W}}\sup_{|t|\leq\tau}|\widehat{K}_S(t) - K_S(t)| = O_p(r_N). \tag{12}$$

Choose $c_N \downarrow 0$ such that $m_N r_N / c_N \to 0$, and reject compatibility when

$$\min_{A\in\mathcal{W}}\lambda_{\min}(\widehat{\Gamma}_{A,N}) < -c_N. \tag{13}$$

Under compatible marginals, the rejection probability converges to zero. If some Λ_A has a finite witness whose matrix has a negative eigenvalue, with arguments in $[-\tau/2, \tau/2]$, the rejection probability converges to one. Condition (12) *holds with $r_N = N_{\min}^{-1/2}$ when p is fixed, sample shares stay bounded away from zero, and $\mathbb{E}[e^{2\tau|Y_S|}] < \infty$ for every world.*

Proof. See Appendix B.3.

The same matrices admit a finite-sample calibration when the outcome range is known. Let $\mathcal{U}_N = \{s + t : s, t \in \mathcal{T}_N\}$ and $L_N = |\mathcal{U}_N|$.

COROLLARY 1 (Bounded-support finite-sample calibration). *Suppose $|Y_{S,i}| \le B$ almost surely in every world, and fix $\alpha \in (0,1)$. Define*

$$\eta_{N,\alpha} = \left(e^{\tau B} - e^{-\tau B}\right)\sqrt{\frac{\log\{2|\mathcal{W}|L_N/\alpha\}}{2N_{\min}}},$$

$$\Delta_{N,\alpha} = 2^p \exp\{(2^p+1)\tau B\}\eta_{N,\alpha}. \tag{14}$$

Reject compatibility when

$$\min_{A\in\mathcal{W}} \lambda_{\min}(\widehat{\Gamma}_{A,N}) < -m_N \Delta_{N,\alpha}. \tag{15}$$

Under compatible marginals, the rejection probability is at most α for every collection of sample sizes $(N_S)_{S\in\mathcal{W}}$.

Proof. See Appendix B.4.

The same argument can use a known envelope for the second moments of the exponential transforms. This allows unbounded outcomes while retaining finite-sample control. Let $Q_N = |\mathcal{W}|L_N$.

COROLLARY 2 (Unbounded-support finite-sample calibration). *Suppose a known constant $\nu \ge 1$ satisfies*

$$\max_{S\in\mathcal{W}} \max_{t\in\mathcal{U}_N} \mathbb{E}[e^{2tY_S}] \le \nu.$$

For $\alpha \in (0,1)$, define

$$\eta^{\infty}_{N,\alpha} = \sqrt{\frac{\nu Q_N}{\alpha N_{\min}}}, \qquad \ell_{N,\alpha} = \nu^{-1/2} - \eta^{\infty}_{N,\alpha}, \qquad \overline{m}_{N,\alpha} = \nu^{1/2} + \eta^{\infty}_{N,\alpha},$$

and suppose $\ell_{N,\alpha} > 0$. Put

$$\kappa_{N,\alpha} = \max\{-\log \ell_{N,\alpha}, \log \overline{m}_{N,\alpha}, \tfrac{1}{2}\log\nu\}, \qquad \Delta^{\infty}_{N,\alpha} = 2^p e^{2^p \kappa_{N,\alpha}} \frac{\eta^{\infty}_{N,\alpha}}{\ell_{N,\alpha}}.$$

Reject compatibility when

$$\min_{A\in\mathcal{W}} \lambda_{\min}(\widehat{\Gamma}_{A,N}) < -m_N \Delta^{\infty}_{N,\alpha}. \tag{16}$$

Under compatible marginals, the rejection probability is at most α uniformly over every collection of laws satisfying the displayed envelope.

Proof. See Appendix B.5.

The sieve supplies consistency under an exponential-moment condition. Corollaries 1 and 2 add finite-sample size control, including at singular points of the positive-semidefinite cone. Their constants come from simultaneous transform bands and favor transparent coverage over local power. Studentized refinements connect to moment selection and inference on directionally differentiable maps (Andrews and Soares 2010; Fang and Santos 2019).

What the conditions provide. The fixed baseline and complete intervention family determine the invertible incidence matrix. Admissibility provides a probability law for each transformed CGF, and local exponential integrability makes these transforms finite and law-determining near zero. Independence joins the recovered component laws by their product. Degenerate components are included. Each condition thus has a separate role in the construction.

These transform conditions belong to the full-law result. The variance characterization uses finite second moments and orthogonality; the paired Wald result uses fourth moments and a nonsingular influence covariance. The latter condition concerns sampling calibration. A factorized law can be well defined even when that covariance is singular.

3.2. Target closure

The full Boolean cube supplies every cross-world comparison at once. A declared collection of regimes can require a smaller set of marginals. Fix $S_1, \dots, S_k \in \mathcal{W}$. For each nonempty $L \subseteq [k]$, let $I_L = \bigcap_{\ell \in L} S_\ell$. The distinct worlds among the I_L form the intersection closure of the target collection. A grouped-source coupling uses mutually independent $(U_J)_{\varnothing \neq J \subseteq [k]}$ and sets

$$W_L = \sum_{J \supseteq L} U_J, \qquad (Y_{S_1}, \dots, Y_{S_k}) = (W_{\{1\}}, \dots, W_{\{k\}}). \tag{17}$$

It matches the closure marginals when $W_L \overset{d}{=} Y_{I_L}$ for every nonempty L.

Suppose these closure marginals have finite MGFs on a common neighborhood of zero, and denote their CGFs by K_{I_L}. For every nonempty $J \subseteq [k]$, define

$$G_J(t) = \sum_{L : J \subseteq L \subseteq [k]} (-1)^{|L|-|J|} K_{I_L}(t). \tag{18}$$

THEOREM 3 (Target-closure characterization). *A grouped-source coupling matching the closure marginals exists if and only if every* G_J *is admissible on a common neighborhood of zero. Within this class, the source laws and the joint law of* $(Y_{S_1}, \ldots, Y_{S_k})$ *are unique, with local joint CGF*

$$K^{\mathrm{F}}_{S_1,\ldots,S_k}(t_1, \ldots, t_k) = \sum_{\varnothing \neq J \subseteq [k]} G_J\left(\sum_{j \in J} t_j\right). \tag{19}$$

Proof. See Appendix B.2.

The closure contains at most $2^k - 1$ distinct worlds, with fewer when different index sets have the same intersection. Its size depends on the declared target collection. Equation (18) also checks whether the available closure marginals support the grouped-source law. The result concerns the selected regimes and their intersections; extending the construction to additional specified world marginals invokes Theorem 1 on the enlarged family.

COROLLARY 3 (Two-world reconstruction). *Fix* $S, T \in \mathcal{W}$ *and let* $I = S \cap T$. *A grouped-source coupling for* (S, T, I) *exists if and only if* $K_S - K_I$ *and* $K_T - K_I$ *are admissible CGFs. When it exists,*

$$K^{\mathrm{F}}_{S,T}(u, v) = K_S(u) + K_T(v) - K_I(u) - K_I(v) + K_I(u + v), \tag{20}$$

$$K^{\mathrm{F}}_{G_{S,T}}(t) = K_S(t) + K_T(-t) - K_I(t) - K_I(-t). \tag{21}$$

If $T \subseteq S$, *equation* (21) *reduces to* $K^{\mathrm{F}}_{G_{S,T}} = K_S - K_T$.

PROOF. Apply Theorem 3 with $k = 2$. The three source CGFs are $K_S - K_I$, $K_T - K_I$, and K_I. Substitution in (19) gives (20); setting $(u, v) = (t, -t)$ gives (21). Under $T \subseteq S$, the intersection is T. □

3.3. Joint-law identification and policy gains

COROLLARY 4 (Marginal-to-joint identification). *Suppose Assumption 1 and the admissibility condition in Theorem 1 hold. Within the factorized class, the marginal laws identify all component laws, all cross-world joint cumulants, and the full joint law in* (10). *For regimes S, T, define*

$$d_A(S,T) = \mathbf{1}\{A \subseteq S\} - \mathbf{1}\{A \subseteq T\}. \tag{22}$$

The policy gain $G_{S,T} = Y_S - Y_T$ has local CGF

$$K_{G_{S,T}}(t) = \sum_{A \subseteq [p]} H_A\big(d_A(S,T)t\big), \tag{23}$$

which uniquely determines its probability law.

PROOF. Theorem 1 identifies the component and joint laws. Equation (5) expresses the gain as $\sum_A d_A(S,T)C_A$. Independence gives the sum of CGFs in (23). Local exponential integrability and MGF uniqueness determine the gain law. Components entering both regimes have zero coefficient and cancel. □

Equation (23) gives the economic content of the reconstruction. Within the maintained class, it determines gain probabilities, quantiles, and dispersion. For a baseline comparison $T = \varnothing$, the gain consists of the components indexed by nonempty subsets of S. Between two arbitrary regimes, components specific to S enter positively and those specific to T enter negatively. Shared components cancel.

4. Economic foundations and their scope

Independent reward modules provide one economic setting for factorization. A common decision rule propagates the modules separately; reoptimization can link the resulting components. The two analytical examples below follow this change while keeping track of the latent draw and transition rule.

4.1. Modular Bellman propagation

Let the state space be finite and let σ be a stationary policy. Write P_σ for its transition matrix and assume $\beta \in (0,1)$. For each $A \subseteq [p]$, let u_A be a random reward vector, with the collection $\{u_A\}$ mutually independent. Randomness here indexes latent types or reward coefficients across evaluations; each Bellman equation evaluates the realized reward vector. The rule σ and matrix P_σ are fixed across these draws. Suppose the reward vector under world S is

$$r_S^\sigma = \sum_{A\subseteq S} u_A. \tag{24}$$

The policy value solves

$$V_S^\sigma = r_S^\sigma + \beta P_\sigma V_S^\sigma. \tag{25}$$

PROPOSITION 3 (Modular Bellman propagation). *Under* (24)–(25)*, the value cofactual associated with A is*

$$C_A^V = (I - \beta P_\sigma)^{-1} u_A. \tag{26}$$

Hence the random vectors $\{C_A^V\}$ are mutually independent. For any fixed linear welfare functional $q^\top V_S^\sigma$, the resulting scalar counterfactual system has a cofactual factorization. If this fixed σ is optimal in every world almost surely, the same statements hold for the optimal value family.

PROOF. The Bellman resolvent gives

$$V_S^\sigma = (I - \beta P_\sigma)^{-1} \sum_{A\subseteq S} u_A = \sum_{A\subseteq S} (I - \beta P_\sigma)^{-1} u_A.$$

Uniqueness of the Möbius representation identifies the Ath value cofact with (26). Separate measurable linear images of mutually independent random vectors remain mutually independent. A fixed linear functional preserves independence across the indexed components. □

The same argument applies conditional on X when $\sigma(X)$ and $P_\sigma(X)$ depend on the common state alone and the u_A are independent given X. The value components are then conditionally independent. Choosing the rule using the realized reward modules introduces an additional dependence channel that requires its own analysis.

The common resolvent is doing the work: each value component depends on its own reward module. A change in the optimal rule, the transition matrix, or dependence among the primitive modules can change the component covariance. Several such effects may offset one another in a particular moment. To interpret a policy change that affects behavior, the analyst therefore needs the decision problem under each regime (Lucas 1976; Sargent and Stachurski 2026).

4.2. Policy stability under reoptimization

The common-policy condition has a primitive sufficient condition. Consider a finite-state, finite-action discounted decision problem with discount factor $\beta \in (0, 1)$, baseline transition kernel P_0, and world kernel P_S. Let r_0 be the baseline reward, let π_0 be its unique optimal stationary policy, and define the minimum baseline action gap

$$g = \min_x \left\{ Q_0^*(x, \pi_0(x)) - \max_{a \neq \pi_0(x)} Q_0^*(x, a) \right\} > 0. \tag{27}$$

Action gaps also control performance losses from approximate action values (Farahmand 2011). Transition perturbations form a related MDP sensitivity problem (Müller 1997). Here the gap supplies a radius in primitive space over which the same optimal policy joins the policy worlds. For world S, write

$$\varepsilon_S = \|r_S - r_0\|_\infty, \qquad \delta_S = \sup_{x,a} \sum_{x'} |P_S(x' \mid x, a) - P_0(x' \mid x, a)|, \qquad \eta_S = \varepsilon_S + \beta\delta_S \|V_0^*\|_\infty. \tag{28}$$

Let $V_S^{\pi_0}$ evaluate π_0 with (r_S, P_S), and let $\widetilde{V}_S^{\pi_0}$ evaluate the same policy with (r_S, P_0). The latter uses the common baseline resolvent.

THEOREM 4 (Policy stability and reoptimization error). *For every world S,*

$$\|Q_S^* - Q_0^*\|_\infty \leq \frac{\eta_S}{1 - \beta}. \tag{29}$$

If $\eta_S < (1 - \beta)g/2$, then $\pi_S^* = \pi_0$. For arbitrary η_S,

$$\|V_S^* - V_S^{\pi_0}\|_\infty \leq \frac{2\eta_S}{1 - \beta}, \tag{30}$$

and for two worlds S, T,

$$\|(V_S^* - V_T^*) - (V_S^{\pi_0} - V_T^{\pi_0})\|_\infty \leq \frac{2(\eta_S + \eta_T)}{1 - \beta}. \tag{31}$$

If $R_S = \|r_S\|_\infty$ and

$$\xi_S = \frac{2\eta_S}{1 - \beta} + \frac{\beta\delta_S R_S}{(1 - \beta)^2},$$

then

$$\|(V_S^* - V_T^*) - (\widetilde{V}_S^{\pi_0} - \widetilde{V}_T^{\pi_0})\|_\infty \leq \xi_S + \xi_T. \tag{32}$$

When the action-gap condition holds in a world, the first term in its ξ_S can be set to zero.

Proof. See Appendix C.1.

Let a common state X record the baseline reward and transition primitives, so $\pi_0(X)$ is fixed conditionally on X. Suppose the nonbaseline reward modules are mutually independent given X and $r_S - r_0 = \sum_{\varnothing \neq A \subseteq S} u_A$. With a common transition kernel, the action-gap condition makes the optimal values equal their fixed-$\pi_0(X)$ values, and Proposition 3 gives conditional factorization. With world-specific kernels, $\widetilde{V}_S^{\pi_0}$ supplies the common-resolvent factorized reference and (32) controls the optimized contrast around it.

The action gap can also control the fraction of units whose policy changes. For each realization of X, let $g(X)$ be the minimum baseline action gap in (27), and let $\eta_S(X)$ be the radius in (28).

COROLLARY 5 (Margin control of policy switching). *Suppose* $g(X) > 0$ *almost surely,* $\eta_S(X) \leq \bar{\eta}_S$, *and*

$$\Pr\{g(X) \leq z\} \leq Cz^{\alpha} \qquad (z \geq 0) \tag{33}$$

for constants $C < \infty$ *and* $\alpha > 0$. *Then*

$$\Pr\{\pi_S^*(X) \neq \pi_0(X)\} \leq C\left\{\frac{2\bar{\eta}_S}{1-\beta}\right\}^{\alpha}. \tag{34}$$

For two worlds S, T, *couple the optimized and fixed-policy value contrasts through the same* X *and primitive reward modules. Any scalar readout* G^* *of* $V_S^* - V_T^*$ *and the corresponding readout* G^0 *of* $V_S^{\pi_0} - V_T^{\pi_0}$ *satisfy*

$$d_{\mathrm{TV}}\{\mathcal{L}(G^*), \mathcal{L}(G^0)\} \leq C\left[\left\{\frac{2\bar{\eta}_S}{1-\beta}\right\}^{\alpha} + \left\{\frac{2\bar{\eta}_T}{1-\beta}\right\}^{\alpha}\right]. \tag{35}$$

Proof. See Appendix C.2.

4.3. Reoptimization and stochastic dependence

A one-state example shows a boundary of the common-policy condition. Let U take -1 and 1 with equal probability, and let the agent choose $a \in \{0, 1\}$. Baseline flow payoff is aU; after an intervention it is $a(U-2)$. Holding U fixed across regimes, reoptimization gives

$$Y_0 = \frac{\max\{0, U\}}{1-\beta}, \qquad Y_1 = 0, \qquad C_1 = -C_{\varnothing}.$$

It follows that

$$\mathrm{Cov}(C_{\varnothing}, C_1) = \mathrm{Var}(Y_1) - \mathrm{Var}(Y_0) = -\frac{1}{4(1-\beta)^2}. \tag{36}$$

Here the negative covariance violates the paired orthogonality equality, and the fall in marginal variance violates the compatibility inequality. The change comes from reoptimization with the same latent draw and transition rule.

5. A policy-gain illustration

Consider a one-state discounted policy-evaluation problem with $\beta = 0.95$. A persistent pre-intervention group $X \in \{-1, 1\}$ has equal probabilities. Let U, V be independent standard normals, also independent of X. A reward credit changes the persistent flow payoff from r_0 to r_1:

$$\begin{aligned} r_0 &= 2 + X + U, \\ r_1 - r_0 &= a(X) + \sigma(X)\{\rho U + \sqrt{1-\rho^2}V\}, \end{aligned} \tag{37}$$

where $a(X) = 0.5 + 0.25X$, $\sigma(-1) = 1$, $\sigma(1) = 3$, and $|\rho| < 1$. Each draw specifies persistent reward heterogeneity for the evaluation horizon. Write $Y_s = (1-\beta)V_s = r_s$ for the annuity value. The declared coupling holds (X, U, V) fixed across regimes. The credit increment is $C_1 = Y_1 - Y_0$ and the root is $C_\varnothing = Y_0$.

Policy question. The target is the dispersion of gains, $\tau = \mathrm{Var}(Y_1 - Y_0)$. Direct calculation gives

$$\tau = \mathbb{E}[\sigma(X)^2] + \mathrm{Var}\{a(X)\} = 5 + 1/16 = 5.0625. \tag{38}$$

Under $\rho = 0$, the components are independent conditional on X. Their unconditional covariance is $\mathrm{Cov}(2+X, 0.5+0.25X) = 0.25$. Consequently the unadjusted marginal variance difference is 5.5625, while the state-adjusted identity gives

$$\mathbb{E}[\mathrm{Var}(Y_1 \mid X) - \mathrm{Var}(Y_0 \mid X)] + \mathrm{Var}\{\mathbb{E}[C_1 \mid X]\} = 5.0625. \tag{39}$$

The 0.5 gap in the unadjusted calculation comes from common-state mean variation.

Sensitivity to the coupling. For general ρ, the residual component covariance is $\gamma(X) = \rho\sigma(X)$. The same proxy on the left of (39) becomes $5.0625 + 4\rho$, while the gain variance in (38) stays fixed. At $\rho = 0.2$, the proxy is 5.8625, exceeding the target by 0.8. The union identity explains the difference: it is $2\mathbb{E}[\gamma(X)]$. Thus $\theta = \mathbb{E}[R_\varnothing R_1] = 2\rho$ assesses the covariance restriction used by the proxy. Along this sensitivity path, the marginal law of Y_1 also changes with ρ.

Assessing the variance formula. A simulator records X, Y_0, Y_1 for each common latent draw. Subtraction recovers C_1, and a cross-fitted residual product estimates θ. At $\rho = 0$, the signal has conditional variances 1 and 9 in the two groups. It diagnoses the covariance correction; the economic target remains $\tau = \mathrm{Var}(Y_1 - Y_0)$.

The acquisition exercise fixes a prescribed bundle and expected budget, and the feasible-estimator experiment evaluates the resulting diagnostic.[7] Behavioral or equilibrium responses belong in the regime-specific outcome maps.

6. Moment restrictions, common states, and variance error

The full-law result can also be read through moments. Marginal variances and cross-regime cumulants give lower-order restrictions, and common-state adjustment helps locate the covariance terms relevant to a policy-contrast variance.

6.1. Intersection restrictions and policy-gain variance

Full factorization implies a sequence of finite-order restrictions.

DEFINITION 5 (Finite-order cofactual orthogonality). *Fix an integer $q \geq 2$. A counterfactual system has* cofactual orthogonality through order q *when, for every* $2 \leq r \leq q$,

$$\mathrm{Cum}(C_{A_1}, \ldots, C_{A_r}) = 0 \tag{40}$$

whenever at least two of the indices $A_1, \ldots, A_r$ differ. The case $q = 2$ is called cofactual orthogonality.

THEOREM 5 (Intersection-cumulant characterization). *Suppose all joint moments through order q exist. The system has cofactual orthogonality through order q if and only if, for every $2 \leq r \leq q$ and every $S_1, \ldots, S_r \in \mathcal{W}$,*

$$\mathrm{Cum}(Y_{S_1}, \ldots, Y_{S_r}) = \kappa_r\big(Y_{S_1 \cap \cdots \cap S_r}\big)\,. \tag{41}$$

If the joint MGF of Y exists in a neighborhood of the origin, then the system has a cofactual factorization if and only if (41) *holds for every $r \geq 2$.*

Proof. See Appendix D.1.

Shared components generate the intersection in the forward implication. The reverse implication applies Möbius inversion to each cumulant index. Thus the same restriction has a component form and a world form.

THEOREM 6 (Target-information duality). *Let $z_S^\top$ denote row S of the incidence matrix Z, let $\mathcal{O} \subseteq \mathcal{W}$ be the worlds whose marginal cumulants are observed, and let $Z_{\mathcal{O}}$ collect their incidence rows. For an integer $2 \leq r \leq q$, define*

$$h_r = \big(\kappa_r(C_A)\big)_{A \in \mathcal{W}}, \qquad q_r(b) = \big(Z^\top b\big)^{\circ r}, \tag{42}$$

[7] An acquisition rule optimized for τ would use the influence function of that policy target.

where $b \in \mathbb{R}^n$ is a world contrast and the power is coordinatewise. Under q-cofactual orthogonality,

$$\kappa_r(b^\top Y) = q_r(b)^\top h_r, \qquad k_{r,\mathcal{O}} := \big(\kappa_r(Y_S)\big)_{S\in\mathcal{O}} = Z_{\mathcal{O}} h_r. \tag{43}$$

Consequently, $q_r(b) \in \mathrm{row}(Z_{\mathcal{O}})$ is sufficient for identification from the observed marginal cumulants. Every solution of $Z_{\mathcal{O}}^\top \lambda = q_r(b)$ gives the reconstruction

$$\kappa_r(b^\top Y) = \lambda^\top k_{r,\mathcal{O}}. \tag{44}$$

For every $r \geq 2$, this row-space condition is also necessary over the factorized Poisson submodel with strictly positive component intensities. If it fails, every interior intensity vector has arbitrarily close alternatives that produce the same marginal laws in $\mathcal{O}$ and different laws for $b^\top Y$.

Proof. See Appendix D.2.

The theorem turns a declared contrast into a loading on latent component cumulants, while the observation design contributes a span of available incidence rows. For several targets and cumulant orders, one design covers the collection when its row space contains every corresponding loading $q_r(b)$. The Poisson construction makes this span condition a lower bound as well as a reconstruction rule at every order.

COROLLARY 6 (Information for a policy gain). *Let $I = S \cap T$ and suppose the required cumulants exist. Under q-cofactual orthogonality, for $2 \leq r \leq q$,*

$$\kappa_r(G_{S,T}) = \begin{cases} \kappa_r(Y_S) - \kappa_r(Y_T), & r \text{ odd}, \\ \kappa_r(Y_S) + \kappa_r(Y_T) - 2\kappa_r(Y_I), & r \text{ even}. \end{cases} \tag{45}$$

Thus endpoint marginals identify every odd gain cumulant, while the intersection marginal completes every even order. Two observed worlds are necessary and sufficient for every odd order. At even orders, two are necessary and sufficient under nesting, while three are necessary and sufficient for incomparable worlds; (S, T, I) attains each bound. These identification minima hold over the factorized Poisson submodel. Hence three world marginals are necessary and sufficient for every even cumulant of a gain between incomparable regimes.

Proof. See Appendix D.2.

COROLLARY 7 (Minimum bundle for pairwise gains). *Fix distinct regimes $S_1, \dots, S_k$ with $k \geq 2$, an odd order $r_o \geq 3$, and an even order $r_e \geq 2$. Let*

$$\mathcal{E} = \{S_i : 1 \leq i \leq k\} \cup \{S_i \cap S_j : 1 \leq i < j \leq k\}$$

collect the distinct endpoint and pairwise-intersection worlds. Under cofactual orthogonality through order $\max\{r_o, r_e\}$*, the marginal cumulants of worlds in $\mathcal{E}$ identify the r_oth and r_eth cumulants of every pairwise gain $Y_{S_i} - Y_{S_j}$. Every observation design based on world marginal laws that identifies this complete collection over the factorized class contains every world in $\mathcal{E}$. Hence $\mathcal{E}$ is the unique inclusion-minimal design. For any positive additive world costs, it is also the unique minimum-cost design.*

Proof. See Appendix D.2.

COROLLARY 8 (Meet covariance and cofactual orthogonality). *For a square-integrable scalar system, the following are equivalent:*

1. *the cofactual components are pairwise uncorrelated;*
2. *for all $S, T \in \mathcal{W}$,*

$$\mathrm{Cov}(Y_S, Y_T) = \mathrm{Var}(Y_{S \cap T}). \tag{46}$$

PROOF. Apply Theorem 5 at cumulant order two. Its component restriction is pairwise zero covariance, and its world-space restriction is equation (46). □

For the policy gain in Corollary 4, the second-order implication is

$$\mathrm{Var}(G_{S,T}) = \nu(S) + \nu(T) - 2\nu(S \cap T), \qquad \nu(S) = \mathrm{Var}(Y_S). \tag{47}$$

Pairwise orthogonality and finite second moments suffice for this formula.[8] The full-law result in Theorem 1 uses independence and transform admissibility as well.

It is worth keeping the two notions of order separate. Interaction order $|A|$ counts the policy switches in a component; cumulant order r counts the arguments of a probabilistic summary. A component involving many switches can still enter a second-order covariance.

[8]For a jointly Gaussian component vector, pairwise orthogonality implies mutual independence. Outside the Gaussian class, the variance identity characterizes second-order behavior; recovery of the full joint law uses additional restrictions.

TABLE 1. A hierarchy of cofactual restrictions

Layer	Restriction in cofactual coordinates	Implication in world coordinates
Unrestricted	none	arbitrary cross-world coupling
Second order	$\mathrm{Cov}(C_A, C_B) = 0$ for $A \neq B$	meet covariance
Order q	mixed cumulants vanish through q	intersection cumulants through q
Full factorization	$\{C_A\}$ mutually independent	CGF inversion and joint-law recovery

The rows are nested when the required moments exist. Under a local joint MGF, validity of the finite-order restrictions at every order reaches the full factorization row.

6.2. Second-order marginal compatibility

For a set function $\nu : \mathcal{W} \to \mathbb{R}$, define the meet-kernel matrix $K_\nu \in \mathbb{R}^{n\times n}$, $n = 2^p$, by

$$(K_\nu)_{S,T} = \nu(S \cap T). \tag{48}$$

Let $\lambda = M\nu$ denote its Möbius transform, so

$$\lambda_A = \sum_{B\subseteq A} (-1)^{|A|-|B|}\nu(B). \tag{49}$$

PROPOSITION 4 (Meet-kernel factorization). *For every set function ν,*

$$K_\nu = Z\ \mathrm{diag}(\lambda_A)\, Z^\top. \tag{50}$$

Hence K_ν is positive semidefinite if and only if $\lambda_A \geq 0$ for every A. When this holds,

$$\mathrm{rank}(K_\nu) = \#\{A : \lambda_A > 0\}. \tag{51}$$

PROOF. The (S, T) entry of the right side of (50) is

$$\sum_{A\subseteq S\cap T} \lambda_A = \nu(S \cap T)$$

by Möbius inversion. Since Z is invertible, congruence by Z preserves inertia. The positive-semidefinite and rank statements follow from the diagonal middle matrix. □

Intersection covariance kernels have a long history in set-indexed Gaussian process theory; for example, Ossiander and Pyke (1985) study additive Gaussian processes with covariance $\mu(A \cap B)$. Nonnegative Möbius transforms also appear in the theory of completely monotone capacities (Chateauneuf and Jaffray 1989). For a surface compatible with the orthogonal benchmark, $\lambda_A \geq 0$ and these coefficients are component variances. A general variance surface obeys the union identity in Proposition 7.

PROPOSITION 5 (Variance-surface characterization). *Let $v(S) = \mathrm{Var}(Y_S)$ for a square-integrable scalar system. Cofactual orthogonality implies*

$$\sigma_A^2 = (Mv)_A = \mathrm{Var}(C_A) \geq 0 \tag{52}$$

for every A. Conversely, if a proposed variance surface v satisfies $Mv \geq 0$, then there exists a centered Gaussian system with a cofactual factorization and $\mathrm{Var}(Y_S) = v(S)$ *for every S.*

PROOF. Under cofactual orthogonality, variance additivity gives

$$v(S) = \sum_{A \subseteq S} \mathrm{Var}(C_A),$$

so Möbius inversion gives (52). For the converse, choose independent $C_A \sim N(0, \sigma_A^2)$ and set $Y_S = \sum_{A \subseteq S} C_A$. □

The signs of the coefficients have a population interpretation. A negative coefficient rules out every orthogonal coupling with those variances; nonnegative coefficients admit a Gaussian construction with the proposed variances. For specified non-Gaussian marginal laws, compatibility depends on the full CGF condition in Theorem 1.

In separate samples, $M\widehat{v}$ estimates these population contrasts under the sampling conditions for the regime variances. Near a zero coefficient, statistical evidence depends on calibration at the boundary of the inequality cone (Andrews and Soares 2010). The paired equality tests use a different sampling experiment and calibration.

6.3. Common pre-intervention states

Shared observed heterogeneity can account for some cross-world dependence. Let X be an observed pre-intervention state in a standard Borel space, with the same interpretation in every regime. The incidence matrix remains Z. The conditional statements below hold for P_X-almost every x.

Write

$$\mu_A(x) = \mathbb{E}[C_A \mid X = x], \qquad R_A = C_A - \mu_A(X), \qquad \Omega_C(x) = \mathrm{Cov}(C \mid X = x).$$

Conditional factorization means mutual independence of the C_A given X. Conditional orthogonality means that $\Omega_C(x)$ is diagonal. Applying Theorem 5 to each conditional law gives the corresponding finite-order cumulant hierarchy.

PROPOSITION 6 (Conditional calculus). *Under conditional orthogonality and finite second moments,*

$$\operatorname{Cov}(Y_S, Y_T \mid X) = \operatorname{Var}(Y_{S\cap T} \mid X). \tag{53}$$

Consequently,

$$\begin{aligned}\operatorname{Cov}(Y_S, Y_T) = {} & \mathbb{E}[\operatorname{Var}(Y_{S\cap T} \mid X)] \\ & + \operatorname{Cov}(\mathbb{E}[Y_S \mid X], \mathbb{E}[Y_T \mid X]).\end{aligned} \tag{54}$$

If the components are conditionally independent and the conditional joint MGF is finite on a neighborhood of zero, then

$$H_A(t \mid x) = \sum_{B\subseteq A} (-1)^{|A|-|B|} K_B(t \mid x), \tag{55}$$

and the conditional version of (10) *determines the cross-world law given x. Knowledge of these conditional marginal laws and P_X therefore determines the unconditional joint law within this class.*

Proof. See Appendix D.8.

Residualization removes the covariance due to variation in conditional means. The average residual product therefore measures average within-state covariance. That average can be zero when positive and negative conditional covariances offset each other. The selected state weights determine which of these patterns a test can detect.

6.4. What marginal cumulants aggregate

There is a useful identity that holds for every cross-world coupling. Let $k_r(S \mid x) = \kappa_r(Y_S \mid X = x)$ and assume conditional moments through order r exist.

PROPOSITION 7 (Union aggregation). *For every $B \subseteq [p]$,*

$$(Mk_r)_B(x) = \sum_{A_1\cup\cdots\cup A_r = B} \operatorname{Cum}(C_{A_1}, \dots, C_{A_r} \mid X = x), \tag{56}$$

where the sum is over ordered tuples. In particular, if $v_S(x) = \operatorname{Var}(Y_S \mid X = x)$, then

$$(Mv)_B(x) = \operatorname{Var}(C_B \mid X = x) + \sum_{\substack{A\cup D = B\\ A\neq D}} \operatorname{Cov}(C_A, C_D \mid X = x). \tag{57}$$

The same identities hold unconditionally when the corresponding moments exist.

Proof. See Appendix D.9.

The off-diagonal sum explains how a marginal variance contrast differs from the corresponding component variance. Orthogonality sets this correction to zero. Under a general coupling it can have either sign. For one intervention, it is twice the covariance between the baseline and the increment, as in the reward-credit calculation.

6.5. Target-specific defect and variance error

The separate marginals and the intersection marginal define the pairwise reference law in Corollary 3. Paired observations reveal how the recorded joint law departs from that reference. Suppose the joint MGF of (Y_S, Y_T) is finite near the origin, let $I = S \cap T$, and define

$$\mathcal{D}_{S,T}(u,v) = K_{S,T}(u,v) - K_S(u) - K_T(v) + K_I(u) + K_I(v) - K_I(u+v). \tag{58}$$

THEOREM 7 (Pairwise defect identity). *Suppose the closure marginals satisfy the admissibility condition in Corollary 3. The actual gain CGF obeys*

$$K_{G_{S,T}}(t) = K^{\mathrm{F}}_{G_{S,T}}(t) + \mathcal{D}_{S,T}(t,-t). \tag{59}$$

For $r \geq 2$, let

$$\epsilon_{r,j} = \mathrm{Cum}(\underbrace{Y_S, \ldots, Y_S}_{j}, \underbrace{Y_T, \ldots, Y_T}_{r-j}) - \kappa_r(Y_I), \qquad 1 \leq j \leq r-1.$$

Whenever these cumulants exist,

$$\kappa_r(G_{S,T}) - \kappa_r^{\mathrm{F}}(G_{S,T}) = \sum_{j=1}^{r-1} \binom{r}{j} (-1)^{r-j} \epsilon_{r,j}, \tag{60}$$

and therefore

$$\left| \kappa_r(G_{S,T}) - \kappa_r^{\mathrm{F}}(G_{S,T}) \right| \leq (2^r - 2) \max_{1 \leq j \leq r-1} |\epsilon_{r,j}|. \tag{61}$$

The identity $\mathcal{D}_{S,T} \equiv 0$ on a neighborhood of the origin holds exactly when the actual pair law equals the grouped-source reference law.

Proof. See Appendix D.6.

At $r = 2$, equation (60) becomes

$$\mathrm{Var}(G_{S,T}) - \mathrm{Var}^{\mathrm{F}}(G_{S,T}) = -2\{\mathrm{Cov}(Y_S, Y_T) - \mathrm{Var}(Y_I)\}. \tag{62}$$

The paired covariance restriction is therefore the second derivative of one functional discrepancy. Higher mixed cumulants give its higher derivatives. The next result supplies a global sensitivity set in characteristic-function coordinates. Let φ_S, φ_T, and $\Phi^{\mathrm{F}}_{S,T}$ denote the endpoint characteristic functions and the bivariate characteristic function of the grouped-source reference from Corollary 3. For an even tolerance function $\rho : \mathbb{R} \to [0, 2]$, let $\mathfrak{C}_\rho$ contain the continuous functions $\Phi : \mathbb{R}^2 \to \mathbb{C}$ satisfying

$$\Phi(0,0) = 1, \qquad [\Phi(z_a - z_b)]_{a,b=1}^{m} \geq 0 \quad \text{for every finite } z_1, \ldots, z_m \in \mathbb{R}^2,$$
$$\Phi(u,0) = \varphi_S(u), \qquad \Phi(0,v) = \varphi_T(v), \qquad |\Phi(t,-t) - \Phi^{\mathrm{F}}_{S,T}(t,-t)| \leq \rho(t) \quad \text{for every } t \in \mathbb{R}.$$

THEOREM 8 (Sharp global gain-law sensitivity). *Suppose the closure marginals satisfy the admissibility condition in Corollary 3. The set of gain laws generated by couplings of the fixed endpoint marginals whose gain characteristic function lies within* ρ *of the grouped-source reference is exactly*

$$\left\{\nu : \varphi_\nu(t) = \Phi(t,-t) \textit{ for some } \Phi \in \mathfrak{C}_\rho\right\}. \tag{63}$$

Every law in this set is attained by a coupling of the endpoint marginals. Finite grids turn membership into positive-semidefinite completion restrictions with fixed axis and anti-diagonal entries.

Proof. See Appendix D.7.

Equation (63) completes the local defect calculation with a global feasibility condition. The tolerance chooses the neighborhood, while positive definiteness determines which gain transforms in that neighborhood extend to a joint law with the declared marginals.

A policy-contrast variance can be compared with the diagonal-component benchmark. Let $\Omega_C(x) = \mathrm{Cov}(C \mid X = x)$, and define

$$D(x) = \mathrm{diag}\{\Omega_C(x)\}, \qquad E(x) = \Omega_C(x) - D(x), \qquad \eta_2 = \{\mathbb{E}\|E(X)\|_F^2\}^{1/2}.$$

Here $D(x)$ retains the actual conditional component variances. For a fixed world contrast b, set $w = Z^\top b$, so $b^\top Y = w^\top C$.

PROPOSITION 8 (Variance error). *Suppose* $\mathbb{E}\|C\|^2 < \infty$ *and* $\eta_2 < \infty$. *For every fixed* $b \in \mathbb{R}^{\mathcal{W}}$,

$$\left|\mathbb{E}[\mathrm{Var}(b^\top Y \mid X)] - \mathbb{E}[w^\top D(X) w]\right| \leq \|w\|_2^2 \eta_2, \qquad w = Z^\top b. \tag{64}$$

Proof. See Appendix D.10.

For a gain $Y_S - Y_T$, b places +1 and −1 at the two regimes. The bound holds the actual conditional component variances fixed and controls the contribution of their off-diagonal covariance. A marginal variance contrast has the additional interpretation given by the union correction in (57). Applying the bound therefore requires paired component information or a stated sensitivity restriction on η_2.

For bounded functionals, a relative-entropy comparison can use the original law and a conditional product reference. The reference preserves the conditional component marginals. Its implied world marginals can change, so the comparison specifies both the reference law and the information used to bound the departure.

7. Observation and assessment

The coupling restriction specifies a class of joint laws. The observation scheme determines what can be learned about a law in that class. Table 2 distinguishes marginal samples, paired rows, and paired rows with a common state. The sampling results evaluate selected equalities for paired data; the acquisition problem adds a cost to those observations.

TABLE 2. Observation schemes and the quantities they support

Observation scheme	Information supplied	Analysis in this paper
Separate regime samples	Marginal laws or selected marginal moments	Population compatibility and coupling bounds; distributional sampling theory is a separate task.
Complete paired rows	A joint world vector for each common unit or draw	Covariance restrictions for the recorded joint law.
Common-state paired rows	Joint outcomes and a common pre-intervention state	Selected residual covariance moments with stated learner conditions.

Separate samples require their own assignment and support assumptions to identify regime marginals. Paired simulation provides the joint law specified by the simulator. Common-state adjustment uses the same pre-intervention interpretation of the state in every regime.

7.1. Observation fibers

Fix a class of candidate joint laws for the stacked world vector. All fibers in this section are taken within that class. Let Q denote one candidate law, and let

$$\mathcal{O} : Q \mapsto d \tag{65}$$

map latent laws to the observed object.

DEFINITION 6 (Observation fiber). *The fiber at d is*

$$\mathcal{F}(d) = \{Q : \mathcal{O}(Q) = d\}. \tag{66}$$

A functional is point identified at a feasible d when every law in $\mathcal{F}(d)$ gives it the same value. If this holds on each nonempty fiber, Proposition A7 writes the functional as a map of the observed object. This is a set-theoretic statement; regularity of the induced map is a further question. Changing the maintained class can change the fiber even with the observation map held fixed.

7.2. World marginals: unrestricted benchmark

Suppose the data identify every world marginal law

$$P = (P_S)_{S \in \mathcal{W}}, \qquad P_S = \mathcal{L}(Y_S).$$

Under the unrestricted cross-world class, the compatible joint laws form the coupling set

$$\Pi(P).$$

This is the standard missing-cross-world problem in potential-outcome analysis. Distributional treatment-effect functionals can remain partially identified even when the marginals are well learned (Fan and Park 2009; Firpo and Ridder 2019; Frandsen and Lefgren 2021). Gao et al. (2025) formulate the general causal partial-identification problem directly as multi-marginal optimal transport and obtain the exact identified set together with statistical guarantees.

For a bounded continuous cofactual functional $h(MY)$ on a finite product of Euclidean outcome spaces, the same benchmark takes the form

$$\left[\inf_{\pi \in \Pi(P)} \int h(My)\, d\pi(y), \quad \sup_{\pi \in \Pi(P)} \int h(My)\, d\pi(y) \right]. \tag{67}$$

The coupling set is weakly compact and convex, and the objective is continuous and affine, so the interval is attained and sharp. The Möbius transform changes the cost; the unresolved object is the cross-world coupling.

7.3. Separate-regime compatibility and its sampling scope

With compatible marginal laws, Corollary 4 leaves one joint law in the factorized fiber. The unrestricted fiber contains the coupling set above. Restricting only finitely many cumulant orders identifies the corresponding joint cumulants, with higher-order features still open. For selected moment observations, Proposition A8 in Appendix E gives the criterion in terms of feasible differences between observationally equivalent moment vectors.

At second order, the population object is $\lambda = M\nu$, with compatibility requiring $\lambda_A \geq 0$ for every A. Separate-sample variance estimates give $\widehat{\lambda} = M\widehat{\nu}$. A sample coefficient close to zero needs to be read together with its sampling uncertainty. Joint inference must account for the dependence among the estimated contrasts and the boundary of the inequality cone (Andrews and Soares 2010). The formal tests below concern paired covariance equalities. Their guarantees apply to the recorded joint law under the stated sampling conditions.

From the policy target to a diagnostic. In (1), the pairing model supplies the cross-world covariance used to calculate gain dispersion. Paired observations allow assessment of restrictions on that covariance. With a common state, residual products isolate its within-state part. This fixes the diagnostic before choosing the sampling procedure.

7.4. Testing second-order restrictions

Each paired row records outcomes for one common unit or structural draw. It is the independent sampling unit in the results below. In a simulator, the shared draw implements the specified economic coupling. A repeated-regime application requires a stable unit-level outcome map and control of carryover effects. The sampling theory takes the pairing and structural parameters as fixed. With independent simulator draws, its standard errors describe Monte Carlo variation around moments of that fixed simulator law. In an empirical paired sample, they describe sampling variation under the stated unit-level observation scheme. Population gain dispersion in equation (1) is the economic target in both cases.[9]

Transforming a row by M gives the components. Their off-diagonal covariances are the moments under assessment. The null sets these moments to zero, and the covariance of their influence vectors supplies the Wald weighting matrix. This keeps the restriction on the outcome law separate from the variance used for inference.

Let $n = 2^p$ and $q = n(n-1)/2$. For a symmetric $n \times n$ matrix H, let $\operatorname{vecoff}(H) \in \mathbb{R}^q$ stack the strict upper-triangular entries once. Given a paired draw Y, define

$$C = MY, \qquad \mu_C = \mathbb{E}[C], \qquad \Omega_C = \operatorname{Cov}(C),$$

and write

$$g_2 = \operatorname{vecoff}(\Omega_C). \tag{68}$$

[9] Estimated structural primitives add a first-stage term to the influence expansion.

Second-order cofactual orthogonality is the restriction $g_2 = 0$. With N i.i.d. paired draws, let $\widehat{\Omega}_C = N^{-1} \sum_j (C^{(j)} - \bar{C})(C^{(j)} - \bar{C})^\top$ be the sample covariance of the transformed draws $C^{(j)} = MY^{(j)}$, and set

$$\widehat{g}_2 = \operatorname{vecoff}(\widehat{\Omega}_C).$$

For $z = C - \mu_C$, define the influence vector

$$\psi_2(C) = \operatorname{vecoff}(zz^\top - \Omega_C), \qquad V_2 = \mathbb{E}[\psi_2(C)\psi_2(C)^\top]. \tag{69}$$

A sample analogue replaces μ_C and Ω_C by their empirical counterparts. All empirical covariance matrices below use denominator N.

THEOREM 9 (Cofactual Wald test). *Fix p as $N \to \infty$. Suppose $Y^{(1)}, \dots, Y^{(N)}$ are i.i.d. paired world draws, $\mathbb{E}\|Y\|^4 < \infty$, and V_2 is positive definite. Let $\widehat{V}_2$ be the empirical covariance of*

$$\widehat{\psi}_2^{(j)} = \operatorname{vecoff}\left[(C^{(j)} - \bar{C})(C^{(j)} - \bar{C})^\top - \widehat{\Omega}_C\right].$$

Under second-order cofactual orthogonality,

$$T_{2,N} = N\widehat{g}_2^\top \widehat{V}_2^{-1} \widehat{g}_2 \implies \chi^2_q. \tag{70}$$

Under a fixed alternative with $g_2 \neq 0$ and V_2 positive definite,

$$\frac{T_{2,N}}{N} \xrightarrow{p} g_2^\top V_2^{-1} g_2 > 0. \tag{71}$$

Hence the test that rejects for large values of $T_{2,N}$ is consistent.

Proof. See Appendix K.3.

This is a moment-restriction calculation (Hansen 1982; Newey and McFadden 1994) for the covariances implied by the fixed loading matrix. A rejection provides evidence against second-order orthogonality of the sampled coupling. The conclusion concerns those covariance moments. Assessment of higher-order dependence uses additional restrictions from the hierarchy.

Under independent components, the weighting matrix is a function of the component variances. A squared-correlation statistic is then available under stronger calibration conditions.

The local-power calculation and descriptive covariance projections use the same zero set as the covariance transform, with different weights on departures. Projection onto the linear meet space also differs from projection onto its positive-semidefinite covariance cone.

Scope of the calibration. The stated limit is pointwise. The law and intervention dimension are held fixed as the sample grows. The positive-definiteness condition on V_2 matters: deterministic components or redundant restrictions can give a singular influence covariance. Uniform calibration along sequences approaching singularity requires additional control. A prespecified subvector with positive-definite covariance has a chi-squared limit with degrees of freedom equal to its length; selecting the rank from the data requires its own calibration. The results condition on fixed structural parameters and a fixed latent pairing. An application with estimated primitives, serial dependence, or additional simulation error needs an influence expansion that includes those sources of uncertainty.

Orthogonality and the product shortcut. Let $U \sim N(0,1)$, $C_\varnothing = U$, and $C_1 = U^2 - 1$. Then

$$\mathrm{Cov}(C_\varnothing, C_1) = 0, \qquad \mathrm{Cum}(C_\varnothing, C_\varnothing, C_1) = 2.$$

The centered cross-product has variance $15 - 2 \cdot 3 + 1 = 10$. Multiplying the component variances gives $1 \cdot 2 = 2$. The robust statistic estimates the cross-product variance, which is the quantity needed for its standard error. The shortcut in Corollary A1 uses independence to replace it by the product. Simulations compare the two calibrations under this law and under independent and correlated laws.

7.5. State-adjusted moments and efficient inference

Fix a finite collection $\mathcal{J} = \{(A_\ell, B_\ell, h_\ell) : 1 \le \ell \le L\}$, where $A_\ell \neq B_\ell$ and each h_ℓ is a known bounded function of X. Define

$$H_\ell(X, C) = h_\ell(X) R_{A_\ell} R_{B_\ell}, \qquad \theta_\ell = \mathbb{E}[H_\ell(X, C)]. \tag{72}$$

Each instrument h_ℓ chooses a weighted average of a conditional component covariance. Conditional orthogonality sets every such moment to zero (Shah and Peters 2020; Scheidegger et al. 2022).

Use a fixed number of folds, with $\widehat{\mu}_A^{(-j)}$ trained outside observation j's fold. Set

$$\widehat{\theta}_\ell = \frac{1}{N} \sum_{j=1}^{N} h_\ell(X_j) \{C_{A_\ell, j} - \widehat{\mu}_{A_\ell}^{(-j)}(X_j)\} \{C_{B_\ell, j} - \widehat{\mu}_{B_\ell}^{(-j)}(X_j)\}. \tag{73}$$

The implementation follows the observation design. Form the selected components from each paired row, train their conditional means on the other folds, and average the held-out residual products with the declared instruments. Each instrument identifies the state-weighted covariance question being assessed. Let $\widehat{V}_X$ be the empirical covariance of the summand vectors $\widehat{H}_j$.

Conditional centering makes the target locally insensitive to errors in either conditional mean. The L^2 product rate controls bias, and L^4 consistency controls variation in the estimated signal. Under the full conditions of Theorem A1, the influence function is $H - \theta$ and the estimator has expansion (A26). The empirical signal covariance estimates $V_X = \mathrm{Var}(H)$, with a chi-squared calibration when it is positive definite. The efficiency bound is for these selected moments in the unrestricted paired-data model.

Scope of the null. The instrument collection determines which projections are tested. For example, let X be equally likely to be -1 or 1, with $\mathrm{Cov}(C_A, C_B \mid X) = \alpha X$. The instrument $h = 1$ gives $\theta = 0$ because the two states cancel. Using $h = X$ gives $\theta = \alpha$. All bounded measurable instruments together characterize conditional covariance zero; a finite collection has the power scope of its selected projections. Conditional independence adds higher-order restrictions. Simulated paired worlds follow this example.

7.6. Finite-sample evidence

Table 3 brings together the two null designs used to assess the covariance calibration. Both satisfy second-order orthogonality; the dependent design has a different cross-product variance from the independence shortcut. The comparison therefore isolates the additional distributional condition needed by that shortcut. At $N = 250$, the robust rejection frequencies also show the approximation error of the chi-squared calibration.

TABLE 3. Calibration under two second-order nulls

N	Component law	Robust Wald (%)	Product shortcut (%)
250	Independent	6.95 (0.57)	5.00 (0.49)
250	Orthogonal, dependent	6.60 (0.56)	25.25 (0.97)
1,000	Independent	4.70 (0.47)	4.75 (0.48)
1,000	Orthogonal, dependent	4.90 (0.48)	24.25 (0.96)

Two interventions and 2,000 replications per row; nominal level 5%. Parentheses give Monte Carlo standard errors in percentage points. The independent design uses standard normal components. The dependent design sets $C_\varnothing = U$ and $C_1 = (U^2 - 1)/\sqrt{2}$, with U standard normal and the other components independent standard normals. Both laws satisfy the covariance null. The product shortcut is calibrated under independence. These rows use the same rejection counts as Table A1, displayed here in percent.

The remaining simulations vary the intervention dimension and evaluate power against the two correlated alternatives. Their outputs have the scope of those moment tests.

The conditional experiments return to the sign-changing covariance example. Its unweighted projection is zero, whereas a state-weighted projection retains the dependence. The two tests therefore answer different moment questions. The reported rejection frequencies describe these particular implementations and sampling models.

8. Conclusion

A shared-component model links marginal regime laws to the distribution of policy gains. Under local exponential integrability, the Möbius-transformed marginal CGFs are admissible if and only if the independent-component representation exists. Compatible transforms construct the source laws and the joint law within the class. Exponential convexity expresses this condition as a hierarchy of positive-semidefinite matrices. Expanding empirical grids consistently detect fixed violations on a transform interval. Simultaneous transform bands control finite-sample rejection probability for bounded outcomes and for unbounded outcomes satisfying a known exponential-moment envelope.

A declared counterfactual target also determines its marginal information requirement. The intersection closure constructs the selected joint law from at most $2^k - 1$ worlds for k target regimes. For two regimes, the endpoints and their intersection determine the gain law. The row-space condition extends this idea to cumulant targets, and its Poisson converse gives an information lower bound at every order. For a family of regimes, every design that identifies all pairwise odd and even gain cumulants contains the endpoint and pairwise-intersection bundle. This yields the unique inclusion-minimal design and the minimum under every positive additive cost. Paired rows then reveal the joint-CGF defect from the marginally constructed reference. The anti-diagonal of that function gives the exact gain-law error, while its derivatives describe cumulant errors. A global positive-definiteness criterion characterizes the attained gain-law sensitivity set with fixed endpoint marginals.

The Bellman results connect the statistical restriction to a decision problem. A common evaluation operator preserves independent reward modules. A positive action gap gives a neighborhood in reward and transition space in which reoptimization preserves the same policy. A common-resolvent value bound measures the combined effect of policy and transition changes around the factorized reference, while the action-gap margin controls the fraction of switching units and the total-variation distance to the world-kernel fixed-policy law. The resulting policy-gain law concerns outcomes joined for the same unit, type, or structural draw, with separate marginals supplying the reference and paired observations supplying its assessment.

References

Dante Amengual, Gabriele Fiorentini, and Enrique Sentana. Moment Tests of Independent Components. *SERIEs*, 13:429–474, 2022. doi: 10.1007/s13209-021-00247-3.

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

Yaron Azrieli and John N. Rehbeck. Marginal stochastic choice. *Theoretical Economics*, 21(2): 373–401, 2026. doi: 10.3982/TE6021.

Raluca M. Balan. Set-Indexed Processes with Independent Increments. *Statistics & Probability Letters*, 59(4):415–424, 2002. doi: 10.1016/S0167-7152(02)00241-9.

Christian Berg, Jens Peter Reus Christensen, and Paul Ressel. *Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions*, volume 100 of *Graduate Texts in Mathematics*. Springer, New York, 1984. doi: 10.1007/978-1-4612-1128-0.

Jorge Buescu and António Paixão. Positive-Definiteness and Integral Representations for Special Functions, 2018. arXiv:1801.09537.

Alain Chateauneuf and Jean-Yves Jaffray. Some Characterizations of Lower Probabilities and Other Monotone Capacities through the Use of Möbius Inversion. *Mathematical Social Sciences*, 17(3): 263–283, 1989. doi: 10.1016/0165-4896(89)90056-5.

Victor Chernozhukov, Iván Fernández-Val, and Blaise Melly. Inference on counterfactual distributions. *Econometrica*, 81(6):2205–2268, 2013. doi: 10.3982/ECTA10582.

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.

Timothy Christensen and Benjamin Connault. Counterfactual sensitivity and robustness. *Econometrica*, 91(1):263–298, 2023. doi: 10.3982/ECTA17232.

Tirthankar Dasgupta, Natesh S. Pillai, and Donald B. Rubin. Causal Inference from 2^K Factorial Designs by Using Potential Outcomes. *Journal of the Royal Statistical Society: Series B*, 77(4): 727–753, 2015. doi: 10.1111/rssb.12085.

Juan Carlos Escanciano and Wei Li. On the Identification of Structural Linear Functionals. cemmap Working Paper CWP48/13, Centre for Microdata Methods and Practice, 2013.

Yanqin Fan and Sang Soo Park. Partial Identification of the Distribution of Treatment Effects and Its Confidence Sets. In *Nonparametric Econometric Methods*, volume 25 of *Advances in Econometrics*, pages 3–70. Emerald, 2009. doi: 10.1108/S0731-9053(2009)0000025004.

Zheng Fang and Andres Santos. Inference on directionally differentiable functions. *Review of Economic Studies*, 86(1):377–412, 2019. doi: 10.1093/restud/rdy049.

Amir-massoud Farahmand. Action-Gap Phenomenon in Reinforcement Learning. In *Advances in Neural Information Processing Systems 24*, 2011.

Sergio Firpo and Geert Ridder. Partial Identification of the Treatment Effect Distribution and Its Functionals. *Journal of Econometrics*, 213(1):210–234, 2019. doi: 10.1016/j.jeconom.2019.04.012.

Patrick Forré and Abel Jansma. Möbius Transforms and Shapley Values for Vector-Valued Functions on Weighted Directed Acyclic Multigraphs, 2025.

Brigham R. Frandsen and Lars J. Lefgren. Partial Identification of the Distribution of Treatment Effects with an Application to the Knowledge Is Power Program. *Quantitative Economics*, 12(1): 143–171, 2021. doi: 10.3982/QE1273.

Zijun Gao and Qingyuan Zhao. Counterfactual Explanability of Black-Box Prediction Models. In Biwei Huang and Mathias Drton, editors, *Proceedings of the Fourth Conference on Causal Learning and Reasoning*, volume 275 of *Proceedings of Machine Learning Research*, pages 1174–1174. PMLR, 2025.

Zijun Gao, Shu Ge, and Jian Qian. Bridging Multiple Worlds: Multi-Marginal Optimal Transport for Causal Partial-Identification Problem. In *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pages 721–729, 2025.

Lars Peter Hansen. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50(4):1029–1054, 1982.

John C. Harsanyi. A Simplified Bargaining Model for the n-Person Cooperative Game. *International Economic Review*, 4(2):194–220, 1963. doi: 10.2307/2525487.

Abel Jansma. Decomposing Interventional Causality into Synergistic, Redundant, and Unique Components. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-1930.

Robert E. Lucas, Jr. Econometric Policy Evaluation: A Critique. *Carnegie-Rochester Conference Series on Public Policy*, 1:19–46, 1976. doi: 10.1016/S0167-2231(76)80003-6.

Mika Mattila and Pentti Haukkanen. On the Positive Definiteness and Eigenvalues of Meet and Join Matrices. *Discrete Mathematics*, 326:9–19, 2014. doi: 10.1016/j.disc.2014.02.018.

Alisdair McKay and Christian K. Wolf. What can time-series regressions tell us about policy counterfactuals? *Econometrica*, 91(5):1695–1725, 2023. doi: 10.3982/ECTA21045.

Geert Mesters and Piotr Zwiernik. Non-Independent Components Analysis, 2024. URL `https://arxiv.org/abs/2206.13668v4`. arXiv:2206.13668, version 4.

Alfred Müller. How does the value function of a markov decision process depend on the transition probabilities? *Mathematics of Operations Research*, 22(4):872–885, 1997. doi: 10.1287/moor.22.4.872.

Whitney K. Newey and Daniel McFadden. 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.

Mina Ossiander and Ronald Pyke. Lévy's Brownian Motion as a Set-Indexed Process and a Related Central Limit Theorem. *Stochastic Processes and their Applications*, 21(1):133–145, 1985. doi: 10.1016/0304-4149(85)90382-5.

Gian-Carlo Rota. On the Foundations of Combinatorial Theory I: Theory of Möbius Functions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 2:340–368, 1964. doi: 10.1007/BF00531932.

Thomas J. Sargent and John Stachurski. Dynamic Programming, Volume I: Finite States, 2026. URL `https://quantecon.github.io/book-dp1/`. QuantEcon online book.

Cyrill Scheidegger, Julia Hörrmann, and Peter Bühlmann. The Weighted Generalised Covariance Measure. *Journal of Machine Learning Research*, 23(273):1–68, 2022. URL `https://jmlr.org/papers/v23/21-1328.html`.

Juliana Schulz, Christian Genest, and Mhamed Mesfioui. A Multivariate Poisson Model Based on Comonotonic Shocks. *International Statistical Review*, 89(2):323–348, 2021. doi: 10.1111/insr.12408.

Rajen D. Shah and Jonas Peters. The Hardness of Conditional Independence Testing and the Generalised Covariance Measure. *The Annals of Statistics*, 48(3):1514–1538, 2020. doi: 10.1214/19-AOS1857.

Yuchung J. Wang. Compatibility among marginal densities. *Biometrika*, 91(1):234–239, 2004. doi: 10.1093/biomet/91.1.234.

David Vernon Widder. Necessary and Sufficient Conditions for the Representation of a Function by a Doubly Infinite Laplace Integral. *Bulletin of the American Mathematical Society*, 40(4):321–326, 1934. doi: 10.1090/S0002-9904-1934-05862-2.

Peng Wu and Xiaojie Mao. The Promises of Multiple Experiments: Identifying Joint Distribution of Potential Outcomes. *Journal of the Royal Statistical Society: Series B*, 2026. doi: 10.1093/jrsssb/qkag118. Advance article, qkag118.