On Cofactuals

Online Appendix

Chae-Yeon Xon
Yonsei University

September 2026

Abstract

This appendix accompanies *On Cofactuals*. It gives supporting proofs, the full state-adjusted inference results, observation design, numerical experiments, and complementary policy summaries. Section references with an appendix letter refer to this document; numerical section references refer to the main paper.

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.

Contents

Appendix A. Coordinate representation and compression

A.1. A generating polynomial

DEFINITION A1 (World generating polynomial). *For each $\omega \in \Omega$, define*

$$\mathscr{Y}_\omega(z) = \sum_{A \subseteq [p]} C_A(\omega) z^A, \qquad z^A = \prod_{i \in A} z_i. \tag{A1}$$

PROPOSITION A1 (Evaluation and differentiation). *Let $\mathbf{1}_S \in \{0,1\}^p$ be the indicator of S. Then*

$$\mathscr{Y}_\omega(0) = Y_\varnothing(\omega), \qquad \mathscr{Y}_\omega(\mathbf{1}_S) = Y_S(\omega), \tag{A2}$$

and

$$\partial_A \mathscr{Y}_\omega(0) = C_A(\omega), \qquad \partial_A = \prod_{i \in A} \frac{\partial}{\partial z_i}. \tag{A3}$$

PROOF. At $z = 0$ only $C_\varnothing$ remains. At $z = \mathbf{1}_S$, the monomial z^A is one exactly when $A \subseteq S$, so Proposition 1 gives the second identity. Multilinearity makes the mixed derivative at the origin select the coefficient of z^A. □

Values at Boolean vertices recover regimes; derivatives at the origin recover interaction coordinates. Interior values extend the polynomial between regimes. A stochastic-intervention interpretation also specifies how the policies are randomized.

A.2. Degree and compression

DEFINITION A2 (Cofactual degree). *For an integer $0 \le m \le p$, the system has cofactual degree at most m when $C_A = 0$ almost surely for every $|A| > m$.*

PROPOSITION A2 (Exact compression). *If the system has degree at most m, then*

$$Y_S = \sum_{\substack{A \subseteq S \\ |A| \le m}} C_A. \tag{A4}$$

The worlds $\{Y_S : |S| \le m\}$ determine the entire system. Their number is

$$N_m = \sum_{k=0}^{m} \binom{p}{k}. \tag{A5}$$

PROOF. The first statement follows from Proposition 1. For $|A| \le m$, the formula for C_A uses only worlds $B \subseteq A$, and hence $|B| \le m$. The low-order worlds therefore recover all nonzero components and then every Y_S. □

A degree restriction limits the interactions present in the family. The next proposition bounds the L^q error from omitting higher-degree components. Here $|A|$ is interaction order; the cumulant index r describes a different property.

PROPOSITION A3 (Approximate compression). *Let $1 \le q \le \infty$ and suppose each $C_A \in L^q$. Define*

$$Y_S^{(m)} = \sum_{\substack{A \subseteq S \\ |A| \le m}} C_A.$$

Then

$$\|Y_S - Y_S^{(m)}\|_{L^q} \le \sum_{\substack{A \subseteq S \\ |A| > m}} \|C_A\|_{L^q}. \tag{A6}$$

PROOF. Proposition 1 gives

$$Y_S - Y_S^{(m)} = \sum_{\substack{A \subseteq S \\ |A| > m}} C_A.$$

Minkowski's inequality gives the result. □

PROPOSITION A4 (Unique multilinear extension). *Every function $f : \{0,1\}^p \to \mathbb{R}^d$ has a unique multilinear polynomial representation*

$$P(z) = \sum_{A \subseteq [p]} c_A z^A$$

that agrees with f on the Boolean vertices. Its coefficients are the Möbius transform

$$c_A = \sum_{B \subseteq A} (-1)^{|A|-|B|} f(\mathbf{1}_B).$$

PROOF. Existence follows from Proposition 1. If two multilinear polynomials agree at every Boolean vertex, their difference has zero Möbius coefficients and hence every coefficient vanishes. □

Appendix B. Marginal reconstruction and transform domains

B.1. Reconstruction proof

Section 3 of the main paper contains the full proof of Theorem 1, including the local exponential-moment argument. The construction chooses each component law from its admissible transform and then takes the finite product law. The identities below record the corresponding coordinate change for general joint laws.

B.2. Proof of Theorem 3

PROOF. Suppose the grouped-source representation exists and write G_J for the local CGF of U_J. For every nonempty $L \subseteq [k]$, independence and (17) give

$$K_{I_L}(t) = \sum_{J \supseteq L} G_J(t). \tag{A7}$$

Möbius inversion on the nonempty subsets of $[k]$, ordered by reverse inclusion, gives (18). Each G_J is therefore admissible.

Conversely, take independent variables U_J with the admissible CGFs in (18). Upper-set inversion gives (A7), so $\sum_{J \supseteq L} U_J$ has CGF K_{I_L} and hence the prescribed marginal law. Taking $L = \{j\}$ yields the selected world S_j. Local MGF uniqueness identifies every source law. Their finite product identifies the grouped-source law, and the linear map in (17) identifies its image. Finally,

$$\sum_{j=1}^{k} t_j Y_{S_j} = \sum_{\varnothing \neq J \subseteq [k]} U_J \sum_{j \in J} t_j,$$

so independence gives (19). The common local domain can be reduced so that every displayed argument remains inside the source domains. □

B.3. Proof of Theorem 2

PROOF. Equation (12), the finite Möbius sum, and boundedness of the candidate transforms on $[-\tau, \tau]$ give

$$\max_{A \in \mathcal{W}} \sup_{|t| \le \tau} |\widehat{\Lambda}_A(t) - \Lambda_A(t)| = O_p(r_N).$$

The operator norm of an $m_N \times m_N$ matrix is bounded by m_N times its largest absolute entry. Weyl's inequality therefore gives

$$\max_{A \in \mathcal{W}} \left|\lambda_{\min}(\widehat{\Gamma}_{A,N}) - \lambda_{\min}(\Gamma_{A,N})\right| = O_p(m_N r_N).$$

Under compatibility, Proposition 2 makes every population matrix positive semidefinite. Since $m_N r_N / c_N \to 0$, equation (13) then has rejection probability approaching zero.

Under the fixed alternative, choose a finite witness $u_1, \ldots, u_q$ and a unit vector a such that $a^\top [\Lambda_A(u_i + u_j)]_{i,j=1}^q a = -\eta$ for some $\eta > 0$. The vanishing mesh supplies grid points converging to the u_i. Continuity preserves a Rayleigh quotient below $-\eta/2$ for the associated $q \times q$ principal submatrix. Interlacing transfers this upper bound to the minimum eigenvalue of the full population grid matrix. Its estimated principal submatrix converges in operator norm because q is fixed. Since $c_N \to 0$, rejection follows with probability approaching one.

For the sufficient primitive condition, the class $\{y \mapsto e^{ty} : |t| \le \tau\}$ has square-integrable envelope $e^{\tau|y|}$ and one-dimensional Lipschitz indexing with an integrable local envelope under $\mathbb{E}[e^{2\tau|Y_S|}] < \infty$. The empirical MGFs are therefore uniformly root-N_S consistent. Population MGFs are continuous and bounded away from zero on the compact interval, so the logarithm preserves the rate. Fixed p and stable sample shares give (12) with $r_N = N_{\min}^{-1/2}$. □

B.4. Proof of Corollary 1

PROOF. For $|t| \le \tau$, bounded support gives $e^{tY_{S,i}} \in [e^{-\tau B}, e^{\tau B}]$. Hoeffding's inequality and a union bound over $S \in \mathcal{W}$ and $t \in \mathcal{U}_N$ show that, with probability at least $1 - \alpha$,

$$\max_{S \in \mathcal{W}} \max_{t \in \mathcal{U}_N} |\widehat{M}_S(t) - M_S(t)| \le \eta_{N,\alpha}, \tag{A8}$$

where $M_S(t) = e^{K_S(t)}$. Both empirical and population MGFs lie in $[e^{-\tau B}, e^{\tau B}]$, so the mean-value theorem for the logarithm gives

$$\max_{S,t} |\widehat{K}_S(t) - K_S(t)| \le e^{\tau B} \eta_{N,\alpha}.$$

Each Möbius sum contains at most 2^p terms. Also $|K_S(t)|$ and $|\widehat{K}_S(t)|$ are at most τB, so $|H_A(t)|$ and $|\widehat{H}_A(t)|$ are at most $2^p \tau B$. Applying the mean-value theorem to the exponential therefore yields

$$\max_{A \in \mathcal{W}} \max_{t \in \mathcal{U}_N} |\widehat{\Lambda}_A(t) - \Lambda_A(t)| \le 2^p \exp\{(2^p + 1)\tau B\} \eta_{N,\alpha} = \Delta_{N,\alpha}.$$

The operator norm of each estimated-minus-population Gram matrix is at most $m_N \Delta_{N,\alpha}$. Under compatibility, Proposition 2 makes every population Gram matrix positive semidefinite. Weyl's inequality then gives

$$\lambda_{\min}(\widehat{\Gamma}_{A,N}) \ge -m_N \Delta_{N,\alpha}$$

simultaneously for every A on the event in (A8). Rule (15) can reject only on the complementary event, whose probability is at most α. $\square$

B.5. Proof of Corollary 2

PROOF. Write $M_S(t) = \mathbb{E}[e^{tY_S}]$. The exponential-moment envelope gives $M_S(t) \le \nu^{1/2}$. Cauchy–Schwarz gives $M_S(t)M_S(-t) \ge 1$, and the same envelope at $-t$ therefore gives $M_S(t) \ge \nu^{-1/2}$. Chebyshev's inequality yields

$$\mathbb{P}\left\{|\widehat{M}_S(t) - M_S(t)| > \eta^{\infty}_{N,\alpha}\right\} \le \frac{\nu}{N_S(\eta^{\infty}_{N,\alpha})^2}.$$

A union bound over the $Q_N = |\mathcal{W}|L_N$ world–argument pairs shows that, with probability at least $1 - \alpha$,

$$\max_{S\in\mathcal{W}} \max_{t\in\mathcal{U}_N} |\widehat{M}_S(t) - M_S(t)| \le \eta^{\infty}_{N,\alpha}. \tag{A9}$$

On this event, every population and empirical MGF appearing in the grid calculation lies in $[\ell_{N,\alpha}, \overline{m}_{N,\alpha}]$. The mean-value theorem for the logarithm then gives

$$\max_{S,t} |\widehat{K}_S(t) - K_S(t)| \le \frac{\eta^{\infty}_{N,\alpha}}{\ell_{N,\alpha}},$$

and the definition of $\kappa_{N,\alpha}$ gives $|K_S(t)| \vee |\widehat{K}_S(t)| \le \kappa_{N,\alpha}$. Each Möbius sum has at most 2^p terms. Applying the mean-value theorem to the exponential therefore gives

$$\max_{A\in\mathcal{W}} \max_{t\in\mathcal{U}_N} |\widehat{\Lambda}_A(t) - \Lambda_A(t)| \le 2^p e^{2^p \kappa_{N,\alpha}} \frac{\eta^{\infty}_{N,\alpha}}{\ell_{N,\alpha}} = \Delta^{\infty}_{N,\alpha}.$$

The corresponding Gram-matrix perturbation has operator norm at most $m_N \Delta^{\infty}_{N,\alpha}$. Compatibility makes the population matrix positive semidefinite. Weyl's inequality implies that rule (16) can reject only outside event (A9). The union bound supplies the stated uniform finite-sample probability. $\square$

B.6. Characteristic functions and cumulants

For any joint law, the characteristic transforms satisfy

$$\phi_C(t) = \phi_Y(M^\top t), \qquad \phi_Y(s) = \phi_C(Z^\top s).$$

These identities follow from $C = MY$ and determine the respective laws. Wherever cumulants of order r exist, multilinearity gives the tensor identities in Proposition A5. Dividing marginal characteristic functions to recover components would require additional nonvanishing and admissibility arguments; the reconstruction theorem uses local MGFs.

Appendix C. Bellman policy stability

C.1. Proof of Theorem 4

PROOF. For every bounded v,

$$\sup_{x,a} |r_S(x,a) - r_0(x,a) + \beta\{P_S(\cdot \mid x,a) - P_0(\cdot \mid x,a)\}v| \leq \varepsilon_S + \beta\delta_S \|v\|_\infty.$$

Apply this bound at $v = V_0^*$. The optimal Bellman operators have contraction modulus β, so their fixed points satisfy

$$\|V_S^* - V_0^*\|_\infty \leq \beta\|V_S^* - V_0^*\|_\infty + \eta_S,$$

and hence $\|V_S^* - V_0^*\|_\infty \leq \eta_S/(1-\beta)$. Substitution in the action-value equations gives the same bound for $\|Q_S^* - Q_0^*\|_\infty$, proving (29). Hence, for every state x and action $a \neq \pi_0(x)$,

$$Q_S^*(x,\pi_0(x)) - Q_S^*(x,a) \geq g - \frac{2\eta_S}{1-\beta}.$$

The stated strict inequality keeps this gap positive and identifies π_0 as the optimal policy in world S.

For a general perturbation, the same fixed-point argument for policy π_0 gives $\|V_S^{\pi_0} - V_0^{\pi_0}\|_\infty \leq \eta_S/(1-\beta)$. Since $V_0^{\pi_0} = V_0^*$, the triangle inequality with the optimal-value bound proves (30). Applying this inequality to S and T and adding the two errors proves (31).

For the common-resolvent reference, the resolvent identity gives

$$V_S^{\pi_0} - \widetilde{V}_S^{\pi_0} = \beta(I - \beta P_S^{\pi_0})^{-1}(P_S^{\pi_0} - P_0^{\pi_0})\widetilde{V}_S^{\pi_0}.$$

Its norm is at most $\beta\delta_S R_S/(1-\beta)^2$. Adding this transition term to (30) gives $\|V_S^* - \widetilde{V}_S^{\pi_0}\|_\infty \leq \xi_S$. The triangle inequality across S and T proves (32). Under policy stability, $V_S^* = V_S^{\pi_0}$ and the reoptimization term drops from the bound. □

C.2. Proof of Corollary 5

PROOF. Apply Theorem 4 conditional on X. A policy change in world S requires

$$g(X) \leq \frac{2\eta_S(X)}{1-\beta} \leq \frac{2\bar{\eta}_S}{1-\beta}.$$

The margin condition gives (34). Under the stated coupling, $G^* = G^0$ whenever the optimal policies agree with $\pi_0(X)$ in both worlds. The coupling inequality for total variation and a union bound therefore give

$$
\begin{aligned}
d_{\mathrm{TV}}\{\mathcal{L}(G^*), \mathcal{L}(G^0)\} &\le \Pr\{G^* \neq G^0\} \\
&\le \Pr\{\pi_S^*(X) \neq \pi_0(X)\} + \Pr\{\pi_T^*(X) \neq \pi_0(X)\},
\end{aligned}
$$

and substituting (34) for the two worlds proves (35). □

Appendix D. Cumulant and conditional calculus

D.1. Proof of Theorem 5

PROOF. Suppose first that (40) holds through order q. Proposition 1 gives $Y_{S_j} = \sum_{A \subseteq S_j} C_A$. Multilinearity expands the left side of (41) into cumulants of cofactual components. Through order q, every term with distinct component indices vanishes. The surviving terms are $\kappa_r(C_A)$ with $A \subseteq S_j$ for every j, hence $A \subseteq S_1 \cap \cdots \cap S_r$. Applying the same argument to the single world $Y_{S_1 \cap \cdots \cap S_r}$ gives (41).

For the converse, apply the Möbius transform in each world index to the cumulant array in (41). The identity

$$
\sum_{B: A \subseteq B \subseteq D} (-1)^{|D|-|B|} = \mathbf{1}\{A = D\}
$$

leaves

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

whenever at least two indices differ. This proves the finite-order equivalence. Under a local joint MGF, validity for every r removes every mixed Taylor monomial from the joint CGF of C. The joint CGF is then the sum of its one-coordinate CGFs, so the joint MGF factorizes and the C_A are mutually independent. The reverse implication follows from independence. □

D.2. Proofs for target information

Proof of Theorem 6

PROOF. Put $c = Z^\top b$. Since $b^\top Y = c^\top C$, multilinearity of cumulants gives a sum over r component indices. Finite-order cofactual orthogonality removes every term involving more than one component, leaving

$$
\kappa_r(b^\top Y) = \sum_{A \in \mathcal{W}} c_A^r \kappa_r(C_A) = q_r(b)^\top h_r.
$$

The same argument with b equal to a world coordinate vector gives $\kappa_r(Y_S) = z_S^\top h_r$, which proves (43). If $Z_{\mathcal{O}}^\top \lambda = q_r(b)$, then $q_r(b)^\top h_r = \lambda^\top Z_{\mathcal{O}} h_r$, proving (44).

For necessity, suppose $q_r(b) \notin \operatorname{row}(Z_{\mathcal{O}})$. The fundamental theorem of linear algebra supplies $\delta \in \ker Z_{\mathcal{O}}$ with $q_r(b)^\top \delta \neq 0$. Fix any $h \in \mathbb{R}^n_{++}$. For all sufficiently small $\epsilon > 0$, both $h^+ = h + \epsilon\delta$ and $h^- = h - \epsilon\delta$ belong to $\mathbb{R}^n_{++}$. Under construction $d \in \{+, -\}$, let the components be mutually independent with $C_A^d \sim \operatorname{Poisson}(h_A^d)$. Each observed world then satisfies

$$Y_S^d \sim \operatorname{Poisson}(z_S^\top h^d),$$

and $Z_{\mathcal{O}} h^+ = Z_{\mathcal{O}} h^-$ makes the complete observed marginal laws identical. Every cumulant of a Poisson variable equals its intensity, so

$$\kappa_r(b^\top Y^d) = \sum_A \{(Z^\top b)_A\}^r h_A^d = q_r(b)^\top h^d.$$

These target cumulants differ, and ϵ can be arbitrarily small. The target laws therefore differ under observationally equivalent factorized systems in every neighborhood of h. □

Proof of Corollary 6

PROOF. For $b = e_S - e_T$, the component loading $c = Z^\top b$ has entries in $\{-1, 0, 1\}$. Hence $c^{\circ r} = c$ for odd r. For even r,

$$c^{\circ r} = c^{\circ 2} = z_S + z_T - 2z_{S \cap T}.$$

Substitution in Theorem 6 gives (45) and the stated reconstruction rules.

For the cardinality statements, let $S \neq T$. Every incidence row has coefficient one on the root component $\varnothing$, while each gain loading has root coefficient zero and is nonzero. A single observed world therefore cannot span the loading. At odd orders, $z_S - z_T$ is spanned by the two endpoint rows, establishing the minimum of two. Under nesting, the even loading equals the same oriented row difference and has the same minimum.

Now let S and T be incomparable. The even loading $q = z_S + z_T - 2z_I$ is a nonzero vector with entries in $\{0, 1\}$. Suppose two incidence rows z_U, z_V span it. Evaluation at the root gives coefficients that sum to zero, so $q = \alpha(z_U - z_V)$. Reorienting the pair if needed, the binary range of q implies $\alpha = 1$ and $z_U \geq z_V$, hence $V \subseteq U$. Choose $i \in S \setminus T$ and $j \in T \setminus S$. Since $q_{\{i\}} = q_{\{j\}} = 1$, both singletons belong to U and neither belongs to V. It follows that $q_{\{i,j\}} = 1$. Yet $\{i, j\}$ is a subset of neither S nor T, so the definition of the even loading gives $q_{\{i,j\}} = 0$, a contradiction. Thus two arbitrary world rows cannot span the even loading. The rows for S, T, I span it by (45), establishing the minimum of three. The Poisson construction in Theorem 6 converts every row-space lower bound above into an identification lower bound for the corresponding cumulant order. □

Proof of Corollary 7

PROOF. Write z_U for the incidence row of world U. For each pair $i < j$, the odd and even target loadings are

$$o_{ij} = z_{S_i} - z_{S_j}, \qquad e_{ij} = z_{S_i} + z_{S_j} - 2z_{S_i \cap S_j}.$$

Their half-sum and half-difference give

$$\frac{e_{ij} + o_{ij}}{2} = z_{S_i} - z_{S_i \cap S_j}, \qquad \frac{e_{ij} - o_{ij}}{2} = z_{S_j} - z_{S_i \cap S_j}.$$

Form a graph with vertex set $\mathcal{E}$ and the nonzero endpoint-to-intersection links. The graph is connected: every pair of endpoint vertices is joined through its intersection vertex, with a direct link when the intersection equals one endpoint, and every distinct intersection vertex is incident to an endpoint. Edge differences in a connected graph span the coefficient vectors on its vertices whose coefficients sum to zero. Since distinct incidence rows are linearly independent, the target-loading space is therefore

$$\mathcal{Q} = \left\{ \sum_{U \in \mathcal{E}} a_U z_U : \sum_{U \in \mathcal{E}} a_U = 0 \right\}, \qquad \dim(\mathcal{Q}) = |\mathcal{E}| - 1. \tag{A10}$$

The rows in $\mathcal{E}$ span this space, proving attainment. The full collection $(z_U)_{U \in \mathcal{W}}$ is a basis because the Boolean incidence matrix is invertible. Suppose an observed-world set $\mathcal{O}$ identifies every target in the corollary. Its row span then contains the two half-sums above and hence every nonzero edge difference $z_U - z_V$ in the connected graph. The expansion of this difference in the full incidence-row basis has nonzero coefficients exactly at U and V. Uniqueness of basis coordinates therefore requires $U, V \in \mathcal{O}$. Every vertex is incident to a nonzero edge, so $\mathcal{E} \subseteq \mathcal{O}$. The Poisson construction in Theorem 6 turns omission of any required target loading into two factorized systems with the same complete observed marginal laws and a different member of the target collection. Thus $\mathcal{E}$ is the unique inclusion-minimal identifying design. When every world has positive additive cost, adding any world to $\mathcal{E}$ raises cost, which gives the final statement. □

D.3. Index cancellation

The cancellation used in Theorem 5 can be written explicitly. Let

$$\Gamma_r(S_1, \dots, S_r) = \mathrm{Cum}(Y_{S_1}, \dots, Y_{S_r}).$$

If $\Gamma_r(S_1, \dots, S_r) = h_r(S_1 \cap \cdots \cap S_r)$, write the Möbius expansion

$$h_r(U) = \sum_{A \subseteq U} g_r(A).$$

Then

$$\Gamma_r(S_1, \dots, S_r) = \sum_{A \subseteq [p]} g_r(A) \prod_{j=1}^{r} \mathbf{1}\{A \subseteq S_j\}.$$

Applying the Möbius transform in each coordinate sends $\mathbf{1}\{A \subseteq S_j\}$ to a point mass at A. Hence

$$\mathrm{Cum}(C_{A_1}, \dots, C_{A_r}) = \begin{cases} g_r(A), & A_1 = \cdots = A_r = A, \\ 0, & \text{otherwise.} \end{cases}$$

Under local exponential integrability, the vanishing mixed cumulants imply factorization of the analytic joint MGF.

D.4. Tensor representation

PROPOSITION A5 (Cofactual tensor diagonalization). *For $r \geq 2$, let Γ_r^Y be the order-r world cumulant tensor and Γ_r^C the corresponding cofactual cumulant tensor. Then*

$$\Gamma_r^C = M^{\otimes r}\Gamma_r^Y, \qquad \Gamma_r^Y = Z^{\otimes r}\Gamma_r^C. \tag{A11}$$

Consequently, q-cofactual orthogonality holds exactly when, for every $2 \leq r \leq q$, the transformed tensor $M^{\otimes r}\Gamma_r^Y$ is supported on the diagonal $A_1 = \cdots = A_r$.

PROOF. Joint cumulants are multilinear in their arguments and $C = MY$. Applying the linear map M in each tensor index gives the first identity. The second follows from $Y = ZC$. Definition 5 is precisely the statement that all off-diagonal tensor entries vanish through order q. □

D.5. Measures of departure

DEFINITION A3 (Cofactual cumulant defect). *For $r \geq 2$, define*

$$\Delta_r(S_1, \dots, S_r) = \mathrm{Cum}(Y_{S_1}, \dots, Y_{S_r}) - \kappa_r(Y_{S_1 \cap \cdots \cap S_r}). \tag{A12}$$

At second order, write

$$\eth_2(Y)^2 = \sum_{S,T \in \mathcal{W}} \Delta_2(S,T)^2. \tag{A13}$$

Theorem 5 gives $\eth_2(Y) = 0$ if and only if the components are pairwise uncorrelated. Higher-order defects measure departures from the corresponding cumulant restrictions. Appendix D.10 gives bounds for second moments and for a full-law comparison.

D.6. Proof of Theorem 7

Proof. By definition, $K_{G_{S,T}}(t) = K_{S,T}(t, -t)$. Substitution of $(u, v) = (t, -t)$ in (58), together with $K_I(0) = 0$, gives (59). Multilinearity of cumulants gives

$$\kappa_r(Y_S - Y_T) = \sum_{j=0}^{r} \binom{r}{j} (-1)^{r-j} \operatorname{Cum}(\underbrace{Y_S, \ldots, Y_S}_{j}, \underbrace{Y_T, \ldots, Y_T}_{r-j}).$$

The endpoint terms are $\kappa_r(Y_S) + (-1)^r \kappa_r(Y_T)$. Since

$$\sum_{j=1}^{r-1} \binom{r}{j} (-1)^{r-j} = -\{1 + (-1)^r\},$$

adding and subtracting $\kappa_r(Y_I)$ in every mixed term and applying Corollary 6 gives (60). The triangle inequality and $\sum_{j=1}^{r-1} \binom{r}{j} = 2^r - 2$ give (61). Finally, $\mathcal{D}_{S,T} \equiv 0$ means that the actual local joint CGF equals (20); local MGF uniqueness gives equality of the pair laws. The converse follows by substitution. □

D.7. Proof of Theorem 8

Proof. Take any coupling in the declared sensitivity class and let Φ be its bivariate characteristic function. Characteristic functions are continuous, normalized, and positive definite. Their restrictions to the coordinate axes are the endpoint characteristic functions, while the characteristic function of $Y_S - Y_T$ is $\Phi(t, -t)$. The sensitivity restriction therefore places Φ in $\mathfrak{C}_\rho$ and its gain law in the set on the right side of (63).

Conversely, take $\Phi \in \mathfrak{C}_\rho$. Bochner's representation theorem on $\mathbb{R}^2$ gives a probability measure π whose characteristic function is Φ (Berg et al. 1984). The two axis restrictions make the marginals of π equal to the declared laws of Y_S and Y_T. The pushforward of π under $(y_S, y_T) \mapsto y_S - y_T$ has characteristic function $\Phi(t, -t)$, and the last defining inequality for $\mathfrak{C}_\rho$ places it inside the sensitivity band. Hence every element of the displayed set is attained by a coupling with the fixed marginals. Restricting the positive-definite kernel to finitely many arguments gives the stated semidefinite completion restrictions.

□

D.8. Proof of Proposition 6

Proof. The finite linear transform commutes with conditioning. The covariance identity follows from Corollary 8 applied to the conditional law, and the law of total covariance gives (54). Conditional independence gives the product of the component MGFs, so the proof of Theorem 1 applies at each x. Local MGF uniqueness identifies each conditional component law. Integrating the resulting conditional joint law against P_X completes the argument. □

D.9. Proof of Proposition 7

PROOF. Expand $k_r(S \mid x)$ by multilinearity into a sum over $A_1, \dots, A_r \subseteq S$. A fixed tuple appears in its Möbius transform at B with coefficient

$$\sum_{S: A_1 \cup \cdots \cup A_r \subseteq S \subseteq B} (-1)^{|B|-|S|} = \mathbf{1}\{A_1 \cup \cdots \cup A_r = B\}.$$

This proves (56). At order two, the diagonal tuple is (B, B), giving (57). □

D.10. Approximation bounds

The target determines the comparison used here. Gain variance calls for a bound on second moments; a bounded nonlinear functional can be compared across full probability laws. For the second-moment bound, set

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

Here $\operatorname{diag}\{\Omega_C(x)\}$ is the diagonal matrix with the same diagonal as $\Omega_C(x)$. For a full-law comparison, let Q be the joint law of (X, C) and define

$$Q^{\otimes}(dx, dc) = P_X(dx) \prod_A Q_A(dc_A \mid x), \qquad \eta_{\mathrm{KL}} = \mathrm{KL}(Q \| Q^{\otimes}). \tag{A14}$$

The standard Borel assumption supplies regular conditional laws. Divergence is measured using natural logarithms.

PROPOSITION A6 (Approximation bounds). *Suppose $\mathbb{E}\|C\|^2 < \infty$. Let $b \in \mathbb{R}^{\mathcal{W}}$ be fixed and put $w = Z^\top b$. If $\eta_2 < \infty$, then*

$$\left|\mathbb{E}[\operatorname{Var}(b^\top Y \mid X)] - \mathbb{E}[w^\top D(X) w]\right| \le \|w\|_2^2 \eta_2. \tag{A15}$$

Let $\widetilde{Q}$ and $\widetilde{Q}^{\otimes}$ be the images of Q and $Q^{\otimes}$ under $(x, c) \mapsto (x, Zc)$. For every measurable $f(x, y) \in [a, a + L]$, if $\eta_{\mathrm{KL}} < \infty$,

$$|\mathbb{E}_{\widetilde{Q}} f - \mathbb{E}_{\widetilde{Q}^{\otimes}} f| \le L\sqrt{\eta_{\mathrm{KL}}/2}. \tag{A16}$$

Moreover, $\eta_{\mathrm{KL}} = 0$ is equivalent to conditional factorization.

The reference keeps each conditional component law. Convolution gives its regime marginals, which can differ from the original marginals. Equation (A16) uses this product reference. Its η_{KL} requires cross-world information or a sensitivity bound. The second-moment comparison uses η_2; the full-law comparison uses the divergence condition.

Proof of Proposition A6

PROOF. The variance difference is $\mathbb{E}[w^\top E(X)w]$. Use $|w^\top E(x)w| \leq \|w\|_2^2\|E(x)\|_F$ and Cauchy–Schwarz. The map $(x,c) \mapsto (x, Zc)$ is a measurable bijection, so it preserves relative entropy. Pinsker's inequality and the bounded range of f give (A16) (Tsybakov 2009). Zero relative entropy means equality of the two laws, which is the conditional product restriction. $\square$

Appendix E. Observation fibers and moments

E.1. Identification on observation fibers

PROPOSITION A7 (Fiber criterion). *A functional $\Psi(Q)$ is point identified at a feasible d with nonempty $\mathcal{F}(d)$ if and only if Ψ is constant on $\mathcal{F}(d)$. If Ψ is constant on every fiber of $\mathcal{O}$, then there exists a function $\widetilde{\Psi}$ on the observation range such that*

$$\Psi = \widetilde{\Psi} \circ \mathcal{O}. \tag{A17}$$

The converse also holds.

PROOF. The first statement is the definition of point identification. For the global statement, define $\widetilde{\Psi}(d)$ as the common value of $\Psi(Q)$ on $\mathcal{F}(d)$. Constancy on every fiber makes the definition well defined and gives (A17). The reverse implication follows by substitution. $\square$

E.2. Linear moment observation

A more general observation design may reveal selected linear moments while leaving the complete marginal laws unrecorded. Let $Y \in \mathbb{R}^n$, $n = 2^p$, and write

$$m_r = \mathbb{E}[Y^{\otimes r}], \qquad q_r = \mathbb{E}[C^{\otimes r}] = M^{\otimes r} m_r.$$

Suppose the observed moment vector is

$$d_r = O_r m_r. \tag{A18}$$

Let $\mathfrak{M}_r$ be the feasible moment class and define its observationally silent span

$$\mathcal{N}_r = \operatorname{span}\{m - m' : m, m' \in \mathfrak{M}_r,\ O_r m = O_r m'\}. \tag{A19}$$

PROPOSITION A8 (Linear cofactual identification). *A linear cofactual functional $a^\top q_r$ is identified on $\mathfrak{M}_r$ if and only if*

$$a^\top M^{\otimes r} v = 0 \qquad \text{for every } v \in \mathcal{N}_r. \tag{A20}$$

If $\mathcal{N}_r = \ker O_r$, this is equivalent to

$$a^\top M^{\otimes r} \in \operatorname{row}(O_r). \tag{A21}$$

PROOF. Set $\ell^\top = a^\top M^{\otimes r}$. The functional takes the same value on every feasible observational equivalence class exactly when ℓ annihilates the differences spanning $\mathcal{N}_r$, which gives (A20). If $\mathcal{N}_r = \ker O_r$, then $\mathcal{N}_r^\perp = \operatorname{row}(O_r)$ by the fundamental theorem of linear algebra. □

The feasible moment class and the observation map enter the criterion separately. This is useful for an application that maintains selected cross-world moment restrictions while allowing a broader joint-law class.

If the observation design identifies every world mean, every mean cofactual is identified by linearity:

$$\mathbb{E}[C_A] = \sum_{B\subseteq A} (-1)^{|A|-|B|}\mathbb{E}[Y_B]. \tag{A22}$$

At second order,

$$\operatorname{Cov}(C_A, C_D) = \sum_{B\subseteq A}\sum_{E\subseteq D} (-1)^{|A|-|B|+|D|-|E|}\operatorname{Cov}(Y_B, Y_E). \tag{A23}$$

This makes the source of cross-world ambiguity visible before any factorization assumption is imposed.

A small example is useful. Let $Y_0, Y_1 \sim \text{Bernoulli}(1/2)$ under two candidate joint laws. Under $Y_1 = Y_0$, $C_1 = 0$ almost surely. Under $Y_1 = 1 - Y_0$, the same marginals hold while $C_1 \in \{-1, 1\}$ with equal probabilities. Thus the mean cofact is zero in both cases, while its variance is zero in the first coupling and one in the second. Factorization selects a particular form of cross-world structure; the unrestricted marginal model leaves this difference open.

Appendix F. Poset identities

For a finite poset $(\mathcal{W}, \preceq)$ with Möbius function μ, define

$$C(w) = \sum_{u\preceq w} \mu(u, w)Y(u).$$

Incidence-algebra inversion gives

$$Y(w) = \sum_{u \le w} C(u).$$

If P and Q are finite posets, then

$$\mu_{P\times Q}((x_1, x_2), (y_1, y_2)) = \mu_P(x_1, y_1)\mu_Q(x_2, y_2).$$

If $Q \subseteq P$ is order convex, the interval $[x, y]$ is the same in Q and P whenever $x, y \in Q$, so the local Möbius coefficient is preserved. The interval identity concerns the Möbius function itself. Restricting the already transformed target agrees with recomputing it on a subposet when that subposet is a lower ideal. On a filter, omitted lower-world terms alter the root. For a chain $0 < 1$ restricted to $\{1\}$, the recomputed coefficient is Y_1, while the original coefficient is $Y_1 - Y_0$. This boundary term must be retained when changing the reference system.

Appendix G. State-adjusted inference

Section 7.5 defines the selected moment vector θ, residual signals H_ℓ, and cross-fitted estimator. The theorem below states the sampling and learner conditions. The canonical-gradient and remainder calculations then explain the first-order representation and its variance.

G.1. Sampling theorem and scope

THEOREM A1 (Conditional cofactual inference). *Suppose (X_j, Y_j) are i.i.d., p and L are fixed, $\mathbb{E}\|C\|^4 < \infty$, and fold proportions are bounded away from zero. In every training fold, for each selected pair (A, B), assume*

$$\begin{aligned} &\|\widehat{\mu}_A - \mu_A\|_{L^4(P_X)} + \|\widehat{\mu}_B - \mu_B\|_{L^4(P_X)} = o_p(1), \\ &\|\widehat{\mu}_A - \mu_A\|_{L^2(P_X)}\|\widehat{\mu}_B - \mu_B\|_{L^2(P_X)} = o_p(N^{-1/2}). \end{aligned} \tag{A24}$$

In the unrestricted paired-data model, the efficient influence function for θ *is*

$$\psi_X(X, C) = H(X, C) - \theta, \qquad V_X = \mathrm{Var}\{H(X, C)\}. \tag{A25}$$

The estimator (73) *satisfies*

$$\sqrt{N}(\widehat{\theta} - \theta) = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \psi_X(X_j, C_j) + o_p(1), \tag{A26}$$

and $\widehat{V}_X \xrightarrow{p} V_X$. If V_X is positive definite, then under $\theta = 0$,

$$N\widehat{\theta}^\top \widehat{V}_X^{-1}\widehat{\theta} \Longrightarrow \chi^2_L. \tag{A27}$$

Its asymptotic variance equals the canonical-gradient bound V_X. Along regular local submodels for which the expansion and its central limit theorem hold locally uniformly, the estimator is regular and attains that bound for these selected moments.

PROOF. A perturbation of either conditional mean has first derivative zero because $\mathbb{E}[R_A \mid X] = \mathbb{E}[R_B \mid X] = 0$. Differentiating the target along a regular submodel therefore gives $\mathbb{E}[\{H - \theta\}s]$ for score s, identifying the canonical gradient. Conditional on each training sample, the residual-product expansion has bias $\mathbb{E}[h(\widehat{\mu}_A - \mu_A)(\widehat{\mu}_B - \mu_B)]$. Condition (A24) controls this bias, while the L^4 conditions control the conditional variance of the score difference. A multivariate central limit theorem and Slutsky's theorem yield the result. Appendix G gives the expansion and covariance argument. □

Conditional centering gives the residual-product score its local robustness (Kennedy 2016; Chernozhukov et al. 2018, 2022). Cross-fitting separates each evaluation fold from its fitted conditional means. The efficiency statement concerns a fixed θ in the unrestricted paired-data model. Treating factorization as known would change the tangent space and hence the relevant bound.

G.2. Canonical gradient

Consider one selected pair (A, B) and a bounded instrument h. Write $R_A = C_A - \mu_A(X)$, $R_B = C_B - \mu_B(X)$ and $\theta = \mathbb{E}[hR_AR_B]$. Along a dominated regular submodel with score $s(X, C)$, differentiate the expectation and the two conditional means. The derivative is

$$\begin{aligned}\dot{\theta} &= \mathbb{E}[hR_AR_Bs] - \mathbb{E}[h\dot{\mu}_A(X)R_B] - \mathbb{E}[h\dot{\mu}_B(X)R_A] \\ &= \mathbb{E}[(hR_AR_B - \theta)s].\end{aligned}$$

The last two terms vanish by conditional centering. In the unrestricted model, the tangent space is $L^2_0(Q)$, so this mean-zero derivative representer is the canonical gradient. Finite fourth moments and bounded h give square integrability. For a finite vector of moments the calculation applies coordinatewise, with covariance bound V_X.

G.3. Cross-fitted expansion

In an evaluation fold, condition on its training sample and put $e_A = \widehat{\mu}_A - \mu_A$ and $e_B = \widehat{\mu}_B - \mu_B$. The score difference is

$$\widehat{H} - H = -he_A(X)R_B - he_B(X)R_A + he_A(X)e_B(X). \tag{A28}$$

Its conditional expectation is $\mathbb{E}[he_A e_B]$, with absolute value bounded by $\|h\|_\infty \|e_A\|_2 \|e_B\|_2 = o_p(N^{-1/2})$. Moreover,

$$\|e_A R_B\|_2 \le \|e_A\|_4 \|R_B\|_4 = o_p(1), \qquad \|e_A e_B\|_2 \le \|e_A\|_4 \|e_B\|_4 = o_p(1).$$

The corresponding statement holds for $e_B R_A$. Each fold's centered score difference therefore contributes $o_p(1)$ on the $\sqrt{N}$ scale by conditional variance control. The fixed number of folds allows these bounds to be summed. This proves (A26).

The same L^2 control and conditional Markov inequality give $N^{-1} \sum_j \|\widehat{H}_j - H_j\|^2 = o_p(1)$. Cauchy–Schwarz then makes the empirical first and second moments of $\widehat{H}$ asymptotically equivalent to those of H. The ordinary law of large numbers gives $\widehat{V}_X \to_p V_X$. The joint central limit theorem and positive definiteness give (A27).

G.4. A finite instrument set

For integrable $\gamma_{AB}(X)$, the equalities $\mathbb{E}[h(X)\gamma_{AB}(X)] = 0$ for all bounded measurable h imply $\gamma_{AB}(X) = 0$ almost surely by taking $h = \mathrm{sign}(\gamma_{AB})$. A finite instrument set leaves the orthogonal complement of its span unrestricted. This describes the power scope of the finite-dimensional Wald test. A squared criterion such as $\mathbb{E}[\gamma_{AB}(X)^2]$ has zero first derivative at the null and requires its own null-distribution calculation. The stated procedure uses regular linear moments.

Appendix H. Observation costs for a prescribed bundle

Suppose a fixed set of regimes can be evaluated jointly for selected reference units. The analyst chooses the acquisition probabilities to improve precision for θ in (72), subject to an expected cost budget. In the reward-credit example, this target is the diagnostic covariance used to assess the gain-variance formula. Both the target and the world bundle are taken as given.

H.1. The diagnostic target and observation rule

For the moments in (72), define the required world bundle

$$\mathcal{U}_{\mathcal{J}} = \bigcup_{\ell=1}^{L} \big(\{S : S \subseteq A_\ell\} \cup \{S : S \subseteq B_\ell\} \big).$$

Observing these worlds jointly recovers every component used by the selected moments. Low-order pairs can therefore use a smaller bundle than the entire Boolean cube.

In a two-phase design, X is observed for every reference unit. The bundle is acquired with indicator D and known probability $e(X) = \mathbb{P}(D = 1 \mid X)$, with $0 < e_{\min} \le e(X) \le 1$ and D conditionally independent of the latent outcomes given X. Here e denotes an observation probability; the scalar ρ in the reward-credit example governs residual dependence. Let

$$r(X) = \mathbb{E}[H(X,C) \mid X], \qquad S_H(X) = \mathrm{Var}\{H(X,C) \mid X\}.$$

These quantities describe heterogeneity in the influence signal for the chosen target.

PROPOSITION A9 (Two-phase efficiency). *In the unrestricted two-phase model with known acquisition probability, the efficient influence function for θ is*

$$\phi_e = r(X) - \theta + \frac{D}{e(X)} \{H(X,C) - r(X)\}. \tag{A29}$$

The bound per reference unit is

$$V_e = \mathrm{Var}\{r(X)\} + \mathbb{E}\left[\frac{S_H(X)}{e(X)}\right]. \tag{A30}$$

A cross-fitted estimator with this influence function is

$$\widehat{\theta}_e = \frac{1}{N} \sum_{j=1}^{N} \left[\widehat{r}^{(-j)}(X_j) + \frac{D_j}{e(X_j)} \{\widehat{H}_j - \widehat{r}^{(-j)}(X_j)\} \right], \tag{A31}$$

under the conditions of Theorem A1 for the conditional-mean learners and $\|\widehat{r} - r\|_{L^2(P_X)} = o_P(1)$ in each fold. The term involving $\widehat{H}_j$ is evaluated only for acquired bundles. Conditional-mean and signal-mean learners use only observed training-fold data. Efficiency refers to regular local attainment under the corresponding conditions of Theorem A1.

PROOF. The observed-data score consists of a marginal-state score plus D times a conditional-outcome score. Equation (A29) represents the target derivative against both parts and lies in the observed tangent space. Conditional expectation gives the two variance terms in (A30). Known inverse acquisition weights make the bias from $\widehat{r} - r$ cancel, leaving the conditional-mean product remainder from Theorem A1. Appendix I supplies the details. □

The calculation uses the augmented inverse-probability form for two-phase observation. The cofactual target determines its complete-data signal and the bundle required to calculate that signal. This follows the influence-function approach to two-phase design (Chen and Lumley 2020).

PROPOSITION A10 (Cost-constrained acquisition). *Fix a positive semidefinite loss matrix W and put*

$$a(x) = \operatorname{tr}\{WS_H(x)\}.$$

Assume $0 < a(X) < \infty$ *almost surely,* $\mathbb{E}[a(X)] < \infty$*, and costs satisfy* $0 < c_{\min} \le c(X) \le c_{\max} < \infty$*. For a budget b with*

$$e_{\min}\mathbb{E}[c(X)] < b < \mathbb{E}[c(X)],$$

the rule minimizing $\operatorname{tr}(WV_e)$ *subject to* $\mathbb{E}[c(X)e(X)] \le b$ *and* $e_{\min} \le e \le 1$ *is*

$$e^*(x) = \min\left\{1, \max\left\{e_{\min}, \sqrt{\frac{a(x)}{\lambda c(x)}}\right\}\right\}, \tag{A32}$$

where $\lambda > 0$ *is chosen so that* $\mathbb{E}[c(X)e^*(X)] = b$.

PROOF. The design-dependent objective is $\mathbb{E}[a(X)/e(X)]$. For each x, the convex function $a(x)/u + \lambda c(x)u$ is minimized on $[e_{\min}, 1]$ by (A32). Its expected cost varies continuously from $\mathbb{E}c$ to $e_{\min}\mathbb{E}c$ as λ increases from zero to infinity. A multiplier meeting the budget therefore exists. Integrating the pointwise minimizing inequality proves optimality among all feasible rules. □

The allocation responds to conditional influence variance relative to cost. Its optimality is an oracle statement, taking $a(x)$ and the state distribution as given under an expected cost budget per reference unit. The floor $e_{\min}$ preserves observation support. An independent pilot can choose a main-sample rule; conditional on the pilot, Proposition A9 applies to the implemented probabilities under its stated conditions. Performance guarantees for learning the rule, updating it sequentially, or choosing the bundle require further analysis.

Two-state calculation. Let X have two equally likely states, with $a(-1) = 1$, $a(1) = 9$, unit cost, $b = 1/2$, and $e_{\min} = 0.1$. The optimal probabilities are 0.25 and 0.75. The design-dependent bound is 8, compared with 10 under uniform acquisition at probability 1/2. If $r(X)$ is constant, this is a 20% reduction in the full efficiency bound. The calculation can arise from conditionally independent Gaussian residuals whose product has these two conditional variances.

H.2. Interpretation of the allocation

For fixed stratum probabilities, the marginal precision benefit per unit of expected cost is $a(x)/\{c(x)e(x)^2\}$. This is a local priority score for the specified objective. Proposition A10 gives the oracle allocation; learned or sequential rules require corresponding performance guarantees.

Appendix L describes invariant mean summaries and allocates variance to intervention modules under orthogonality. Acquisition priority concerns precision for θ: a large policy contribution can coexist with precise estimation of the relevant moment.

Additional population information contracts an observation fiber. The design criterion here concerns the sampling variance of an estimator. A criterion based on identified-set reduction would also need the law of the information to be acquired.

H.3. The reward-credit observation plan

A simulator records X, Y_0, Y_1 for each latent draw. Subtraction recovers C_1, and regression on the two groups recovers the conditional means. The cross-fitted residual product estimates θ. At $\rho = 0$, its conditional variance is $\sigma(X)^2$, equal to 1 and 9 in the two groups. These are the same influence variances used in the acquisition example in Appendix H.1. A fixed expected half-sample budget yields probabilities 0.25 and 0.75 and an oracle variance bound of 8, compared with 10 under uniform acquisition. Appendix J implements the entire calculation with missing bundles masked before estimation, training on acquired observations, and known acquisition probabilities.

The allocation improves precision for the diagnostic covariance θ. Its role is to assess the formula used for the economic target τ, the variance of policy gains. Optimizing acquisition directly for τ would require its influence function. Any behavioral or equilibrium responses belong in the value map for each regime.

Appendix I. Derivations for observation design

I.1. Efficient influence function under acquisition

Only the components used by $\mathcal{J}$ are relevant here; denote their jointly observed vector by $C_{\mathcal{J}}$. In the unrestricted observed-data model the state score is $s_X(X)$, and the conditional-outcome score is $s_C(C_{\mathcal{J}}, X)$ with $\mathbb{E}[s_C \mid X] = 0$. Known acquisition probabilities give observed score $s_X + Ds_C$. Set $r(X) = \mathbb{E}[H \mid X]$. Then

$$\begin{aligned}\mathbb{E}[\phi_e\{s_X + Ds_C\}] &= \mathbb{E}[(r-\theta)s_X] + \mathbb{E}[(H-r)s_C] \\ &= \mathbb{E}[(H-\theta)(s_X + s_C)].\end{aligned}$$

This is the full-data target derivative. The function ϕ_e belongs to the closure of the observed tangent space, since $(H-r)/e$ has conditional mean zero. It is therefore the canonical gradient. Conditional variance and the orthogonality of the two displayed terms yield (A30).

For the estimator, let $u(X) = \widehat{r}(X) - r(X)$. Its score difference from the efficient signal equals

$$\left(1 - \frac{D}{e}\right) u(X) + \frac{D}{e}(\widehat{H} - H).$$

The first term has conditional expectation zero and L^2 norm tending to zero because $e \geq e_{\min}$. The second term has expectation $\mathbb{E}[\widehat{H} - H]$ and its squared norm is at most $e_{\min}^{-1}\mathbb{E}[(\widehat{H} - H)^2]$. The proof in Appendix G applies. Thus the estimator is asymptotically linear with influence function ϕ_e, and the covariance of its cross-fitted summands consistently estimates V_e. The argument uses the known design probabilities; an estimated missingness mechanism requires additional nuisance-rate conditions.

I.2. Separate-world allocation for a marginal target

There is also a marginal-only design problem. Let $v_S = \operatorname{Var}(Y_S)$ and let $\tau = a^\top M v = b^\top v$, where $b = M^\top a$ is fixed. In independent regime samples, define

$$\omega_S^2 = \operatorname{Var}\{(Y_S - \mathbb{E}Y_S)^2\}.$$

Under finite fourth moments, the within-regime sample variance has efficient influence function $(Y_S - \mathbb{E}Y_S)^2 - v_S$. The leading variance of $\widehat{\tau}$ is therefore $\sum_S b_S^2 \omega_S^2 / N_S$. Suppose $\omega_S > 0$ for every $b_S \neq 0$. If each observation costs $c_S > 0$ and $\sum_S c_S N_S = B$, the continuous oracle allocation over these relevant coordinates is

$$N_S^* = \frac{B|b_S|\omega_S/\sqrt{c_S}}{\sum_T |b_T|\omega_T\sqrt{c_T}}, \qquad \min \operatorname{Var}(\widehat{\tau}) \simeq \frac{\left(\sum_T |b_T|\omega_T\sqrt{c_T}\right)^2}{B}. \tag{A33}$$

Cauchy–Schwarz proves the lower bound, with equality at the displayed allocation. This is a cost-adjusted Neyman calculation; related factorial allocation results appear in Ravichandran et al. (2024). It optimizes estimation of a marginal variance contrast. Proposition 7 specifies its interpretation under general coupling, while paired acquisition estimates the conditional cross-world moments themselves.

Appendix J. Numerical experiments and implementation checks

The accompanying Python scripts reproduce the checks in this appendix using NumPy and SciPy. Algebraic checks cover $MZ = ZM = I$, the union identity through order three, the meet-kernel factorization and projection, and the conditional variance bound. The acquisition formula is compared with constrained numerical optimization. These checks supplement the proofs by verifying their finite-dimensional implementations.

J.1. Covariance-test size and power

For $p \in \{1, 2\}$ and $N \in \{250, 1000\}$, four component laws are simulated. The independent law has standard normal components. The orthogonal dependent law sets $C_\varnothing = U$ and $C_1 = (U^2 - 1)/\sqrt{2}$, with remaining components independent standard normals. The two alternatives set $C_1 = \rho C_\varnothing + \sqrt{1-\rho^2}\epsilon_1$ for $\rho = 0.10$ or 0.25, with all primitive normals independent. Every draw is converted to $Y = ZC$ and transformed back before estimation. Both procedures use denominator N and the chi-squared reference with $2^p(2^p - 1)/2$ degrees of freedom.

TABLE A1. Size and power of covariance tests

p	N	Component law	Robust Wald	Product shortcut
1	250	Independent	0.061 (0.005)	0.055 (0.005)
1	250	Orthogonal, dependent	0.052 (0.005)	0.382 (0.011)
1	250	Correlation 0.10	0.350 (0.011)	0.350 (0.011)
1	250	Correlation 0.25	0.980 (0.003)	0.980 (0.003)
1	1,000	Independent	0.049 (0.005)	0.046 (0.005)
1	1,000	Orthogonal, dependent	0.052 (0.005)	0.381 (0.011)
1	1,000	Correlation 0.10	0.879 (0.007)	0.881 (0.007)
1	1,000	Correlation 0.25	1.000 (0.000)	1.000 (0.000)
2	250	Independent	0.070 (0.006)	0.050 (0.005)
2	250	Orthogonal, dependent	0.066 (0.006)	0.253 (0.010)
2	250	Correlation 0.10	0.191 (0.009)	0.176 (0.009)
2	250	Correlation 0.25	0.871 (0.007)	0.869 (0.008)
2	1,000	Independent	0.047 (0.005)	0.048 (0.005)
2	1,000	Orthogonal, dependent	0.049 (0.005)	0.242 (0.010)
2	1,000	Correlation 0.10	0.665 (0.011)	0.659 (0.011)
2	1,000	Correlation 0.25	1.000 (0.000)	1.000 (0.000)

2,000 replications per row, nominal level 5%. Parentheses give Monte Carlo standard errors. The first two laws satisfy the robust covariance null. Only the independent law satisfies the shortcut's independence calibration. Every generated run is retained; singularity is an explicit error condition. The dimensions and distributional families are fixed.

The orthogonal dependent law satisfies the robust second-order null. Its standardized cross-product variance is 5; the independence shortcut uses 1. The rejection frequencies show the consequence of that calibration choice. They also show the finite-sample approximation error for the robust procedure. The evidence is specific to these fixed-dimensional designs and the reported repetition counts.

J.2. Common states and a fixed instrument collection

For the sampling illustration, let $p = 2$, $X \in \{-1, 1\}$ have equal probabilities, and

$$C = \alpha X + \epsilon, \qquad \alpha = (0.6, 0.7, -0.4, 0.2)^\top, \qquad Y = ZC.$$

Conditional errors are Gaussian with unit variances. In the null design they are independent. In the cancellation design their sole off-diagonal conditional covariance is $\mathrm{Cov}(\epsilon_1, \epsilon_2 \mid X) = 0.35X$. The unadjusted test concerns unconditional cofactual covariance; the two adjusted tests use all six component pairs, with instruments 1 or $(1, X)$. Conditional means are estimated by ordinary least squares on an intercept and X, using five-fold cross-fitting. All tests use robust empirical covariance matrices and nominal level 5%.

TABLE A2. Rejection frequencies for declared paired-world designs

	Unadjusted	Adjusted: 1	Adjusted: $(1, X)$
Conditional null	1.000 (0.000)	0.056 (0.007)	0.071 (0.008)
Sign-changing covariance	1.000 (0.000)	0.058 (0.007)	1.000 (0.000)

1,000 replications, 1,000 observations per replication; five-fold cross-fitting. Parentheses report Monte Carlo standard errors. All tests use a nominal 5% level. The unadjusted null differs from the conditional null; the reported unadjusted rejections reflect the common state. Seeds and the complete implementation are included in the source package.

At $N = 4,000$, using the same 1,000 replications and seeds, the adjusted null rejection frequencies are 0.047 and 0.050 for instruments 1 and $(1, X)$, respectively (Monte Carlo standard errors 0.007 and 0.007). These finite-sample frequencies accompany the asymptotic statements and their stated regularity conditions.

The common state induces unconditional covariance in both designs. Residualization targets zero moments in the null design. In the cancellation design, the unweighted residual moment is again zero, but the state-weighted moment equals 0.35. These instrument collections test different hypotheses under the same simulated coupling. The comparison shows which conditional pattern each test can detect.

J.3. Two-phase reward-credit experiment

The experiment uses the reward-credit model in Section 5 at $\rho = 0$, so the target $\theta = \mathbb{E}[R_\varnothing R_1]$ is zero. At $N = 1000$ and 4000, 2,000 independent replications compare uniform acquisition with the fixed oracle probabilities $(0.25, 0.75)$. The same latent draws are used for the two rules within a replication. Unacquired bundles are masked before the estimating routine receives the data. In each training fold, group-specific means and residual-product means are estimated using acquired bundles; validation-fold signals use the known design probability in (A31).

TABLE A3. Two-phase inference for the reward-credit example

N	Acquisition rule	Bound	$N\,\widehat{\mathrm{Var}}_{MC}(\widehat{\theta})$	Rejection rate
1,000	Uniform	10.0	9.96 (0.31)	0.051 (0.005)
1,000	Oracle allocation	8.0	8.32 (0.26)	0.057 (0.005)
4,000	Uniform	10.0	9.82 (0.31)	0.048 (0.005)
4,000	Oracle allocation	8.0	8.24 (0.26)	0.051 (0.005)

2,000 replications per row, nominal level 5%, five-fold fitting using acquired bundles only. Both rules acquire half of reference units in expectation. The rule is fixed at the data-generating law's oracle probabilities; conditional means and signal means are estimated. Parentheses give Monte Carlo standard errors, including the sampling uncertainty in the empirical variance. This experiment evaluates the specified two-phase design.

The table reports the analytical bound, scaled sampling variance, and rejection frequency of the feasible augmented estimator. Both sets of conditional means are estimated; the acquisition probabilities use the known variance profile of the data-generating process. Every reference unit contributes its common state. The experiment follows the estimator from acquisition through inference for a prescribed bundle. Pilot-based design learning and bundle selection would require additional experiments and theory.

The replication package contains seeds, rejection counts, Monte Carlo standard errors, numerical failure counts, and environment versions. A separate machine-readable timing log records runtime. Every generated replication is retained. The implementation raises an explicit error for a degenerate influence covariance or an empty training stratum. Finite algebra checks and simulations supplement the analytical proofs; each has the scope specified above.

Appendix K. Additional covariance results

K.1. Independence-calibrated shortcut

COROLLARY A1 (Factorized simplification). *In the sampling framework of Theorem 9, suppose the cofactual components are mutually independent and have positive variances. Then V_2 is diagonal, with the entry associated with the pair (A, B) equal to*

$$\mathrm{Var}(C_A)\,\mathrm{Var}(C_B), \qquad A < B.$$

Consequently,

$$T^{\mathrm{F}}_{2,N} = N \sum_{A<B} \frac{\widehat{\Omega}^2_{C,AB}}{\widehat{\Omega}_{C,AA}\widehat{\Omega}_{C,BB}} \implies \chi^2_q. \tag{A34}$$

Thus the factorized statistic is asymptotically the sum of squared sample cofactual correlations, multiplied by N.

Proof. See Appendix K.4.

K.2. Descriptive covariance projection

Let

$$\mathbb{K}_\cap = \{K_\nu : \nu \in \mathbb{R}^{\mathcal{W}}\}$$

be the linear space of meet-kernel matrices.

PROPOSITION A11 (Projection onto the meet-kernel space). *For a symmetric matrix $H = (H_{S,T})$, define*

$$\bar{H}_A = \frac{1}{3^{p-|A|}} \sum_{R,Q:R\cap Q=A} H_{R,Q}. \tag{A35}$$

Then the Frobenius projection of H onto $\mathbb{K}_\cap$ is

$$(\Pi_\cap H)_{S,T} = \bar{H}_{S\cap T}. \tag{A36}$$

The space $\mathbb{K}_\cap$ has dimension $n = 2^p$ and codimension

$$\frac{n(n-1)}{2} \tag{A37}$$

in the space of symmetric $n \times n$ matrices.

PROOF. The Frobenius objective separates across classes of ordered pairs with the same intersection. For a fixed A, the least-squares constant for the class $\{(R, Q) : R \cap Q = A\}$ is its arithmetic mean. Each element of A must belong to both R and Q. Each element outside A has three admissible membership patterns: in R only, in Q only, or in neither. Hence the class contains $3^{p-|A|}$ ordered pairs, which gives (A35). The map $v \mapsto K_v$ is injective because $v(A) = (K_v)_{A,A}$, so $\dim \mathbb{K}_\cap = n$. The symmetric matrix space has dimension $n(n+1)/2$, giving (A37). □

This suggests the empirical defect

$$\widehat{D}_2 = \|\widehat{\Sigma} - \Pi_\cap \widehat{\Sigma}\|_F, \tag{A38}$$

where $\widehat{\Sigma}$ is a paired-world covariance estimate. With n worlds, it imposes $n(n-1)/2$ independent linear covariance restrictions.

The same null can be inspected directly in cofactual coordinates. Since $C = MY$,

$$\Omega_C = M\Sigma M^\top. \tag{A39}$$

Cofactual orthogonality is equivalent to diagonality of Ω_C. Thus $\|\operatorname{offdiag}(M\widehat{\Sigma}M^\top)\|_F$ provides a corresponding measure of departure in cofactual coordinates. The world-space projection in (A38) and the cofactual-space transform in (A39) have the same zero set and place different weights on departures from it.

Projection and covariance validity. The projection above is onto a linear space. Positive semidefiniteness is a further constraint. For $p = 1$, projecting $H = \operatorname{diag}(4, 1)$ gives

$$\Pi_\cap H = \begin{pmatrix} 4/3 & 4/3 \\ 4/3 & 1 \end{pmatrix}, \qquad \det(\Pi_\cap H) = -4/9.$$

Thus the projection is useful for equality residuals. A valid meet-covariance estimate additionally enforces $Mv \geq 0$, for example by minimizing the same squared error over the resulting closed convex cone.

COROLLARY A2 (Consistency of the second-order diagnostic). *Let $Y^{(1)}, \ldots, Y^{(N)}$ be i.i.d. paired draws of the world vector with finite second moments, and let $\widehat{\Sigma}_N$ be the sample covariance matrix. Then*

$$\|\widehat{\Sigma}_N - \Pi_\cap \widehat{\Sigma}_N\|_F \xrightarrow{p} \|\Sigma - \Pi_\cap \Sigma\|_F.$$

The limit is zero exactly under cofactual orthogonality.

PROOF. The sample covariance converges entrywise in probability to Σ. The projection is linear and continuous. Corollary 8 identifies the zero set of the limiting norm. □

With finite fourth moments, the linear restrictions underlying (A38) can also be placed in a standard Wald or GMM statistic. The projection form is often more useful for diagnosis because it reports the residual covariance associated with each pair of policy worlds.

K.3. Proof of Theorem 9

PROOF. Write $z_j = C^{(j)} - \mu_C$. The identity

$$\widehat{\Omega}_C - \Omega_C = \frac{1}{N}\sum_{j=1}^{N}(z_j z_j^\top - \Omega_C) - (\bar{C} - \mu_C)(\bar{C} - \mu_C)^\top$$

places the mean-estimation term at order $O_p(N^{-1})$. Thus the sample covariance has the asymptotic linear representation

$$\sqrt{N}\,\mathrm{vecoff}(\widehat{\Omega}_C - \Omega_C) = \frac{1}{\sqrt{N}}\sum_{j=1}^{N}\psi_2(C^{(j)}) + o_p(1).$$

The multivariate central limit theorem gives

$$\sqrt{N}(\widehat{g}_2 - g_2) \Longrightarrow N(0, V_2).$$

The empirical covariance of the influence vectors converges in probability to V_2. Under the null $g_2 = 0$, the quadratic form in (70) therefore converges to χ^2_q. Under a fixed alternative, $\widehat{g}_2 \xrightarrow{p} g_2$ and $\widehat{V}_2 \xrightarrow{p} V_2$, which gives (71). □

K.4. Proof of Corollary A1

PROOF. Center the cofactual components. For $A \neq B$,

$$\mathrm{Var}(C_A C_B) = \mathrm{Var}(C_A)\,\mathrm{Var}(C_B)$$

by independence. Cross-products associated with distinct unordered pairs have covariance zero: disjoint pairs factor into centered first moments, while pairs sharing one index contain the centered first moment of the remaining component. Hence V_2 is diagonal. Substitution in Theorem 9 gives (A34). □

K.5. Local alternatives and normalized departures

COROLLARY A3 (Local power). *Consider a sequence of paired-world laws for which*

$$\sqrt{N}\, g_{2,N} \to h, \qquad V_{2,N} \to V_2 > 0,$$

the triangular-array central limit theorem underlying Theorem 9 holds, and $\widehat{V}_{2,N} \xrightarrow{p} V_2$*. Then*

$$T_{2,N} \Longrightarrow \chi_q^2(\lambda), \qquad \lambda = h^\top V_2^{-1} h.$$

PROOF. The triangular-array expansion underlying Theorem 9 gives $\sqrt{N}\widehat{g}_{2,N} \Longrightarrow N(h, V_2)$. Consistency of $\widehat{V}_{2,N}$ and Slutsky's theorem turn the Wald quadratic form into the squared norm of a normal vector with mean $V_2^{-1/2}h$ and identity covariance. This is $\chi_q^2(\lambda)$ with the displayed noncentrality parameter. □

The scalar

$$\delta_2^2 = g_2^\top V_2^{-1} g_2 \tag{A40}$$

provides a studentized population measure of second-order departure from factorization. Equation (71) shows that $T_{2,N}/N$ estimates this distance under fixed alternatives. The Frobenius defect in (A38) is easier to visualize; δ_2 is easier to compare statistically across samples.

K.6. Higher-order specification tests

The same idea extends along the cumulant hierarchy. Let $\kappa_C^{(r)}$ denote the symmetric order-r cumulant tensor of C. Write $\mathrm{mix}_r(\kappa_C^{(r)})$ for a vector containing one copy of each tensor entry whose indices include at least two distinct values. With $n = 2^p$ component coordinates, the number of such restrictions at order r is

$$d_r = \binom{n+r-1}{r} - n. \tag{A41}$$

The first term is the dimension of a symmetric order-r tensor; the second counts the n diagonal cumulants that remain free.

PROPOSITION A12 (Finite-order specification hierarchy). *Fix p and an integer $Q \geq 2$. Suppose paired draws are i.i.d., $\mathbb{E}\|Y\|^{2Q} < \infty$, and the asymptotic covariance matrix of the empirical mixed cofactual cumulants through order Q is nonsingular. Stack*

$$g_{\leq Q} = \left(\mathrm{mix}_2(\kappa_C^{(2)})^\top, \ldots, \mathrm{mix}_Q(\kappa_C^{(Q)})^\top\right)^\top.$$

Then Q-cofactual orthogonality is equivalent to $g_{\le Q} = 0$. The empirical cumulant vector is asymptotically normal, and the corresponding sandwich Wald statistic converges under the null to a chi-squared distribution with

$$d_{\le Q} = \sum_{r=2}^{Q} \left[\binom{n+r-1}{r} - n \right] \tag{A42}$$

degrees of freedom.

PROOF. The equivalence follows from Proposition A5. Empirical cumulants of fixed order are polynomials in empirical moments of order at most Q. Under the stated $2Q$ moment condition, the vector of these empirical moments satisfies a multivariate central limit theorem. The delta method gives asymptotic normality of the stacked empirical cumulants. A consistent sandwich covariance estimate and the usual Wald argument complete the result. □

The dimensions p and Q are fixed in Proposition A12. At second order the restriction counts for $p = 2, 3, 4$ are 6, 28, and 120. The nonsingularity condition is substantive: deterministic components or polynomial relations among the observed variables can make the influence covariance singular.

K.7. Gaussian systems

COROLLARY A4 (Gaussian meet-kernel characterization). *Let $Y = (Y_S)_{S \in W}$ be a jointly Gaussian scalar counterfactual system. The following are equivalent:*

1. *the system has a cofactual factorization;*
2. *the system has cofactual orthogonality of every order;*
3. *there is a set function ν with $M\nu \ge 0$ such that*

$$\mathrm{Cov}(Y_S, Y_T) = \nu(S \cap T)$$

for all S, T.

Under these conditions, $\nu(S) = \mathrm{Var}(Y_S)$ and $\mathrm{Var}(C_A) = (M\nu)_A$.

PROOF. For a jointly Gaussian vector, pairwise uncorrelated cofactual coordinates are mutually independent. Corollary 8 gives the equivalence between orthogonality and meet covariance, while Proposition 4 gives the nonnegative Möbius representation. □

K.8. A two-intervention illustration

Suppose independent Gaussian cofactual components have variances

$$\sigma_\varnothing^2 = 1, \qquad \sigma_1^2 = 0.4, \qquad \sigma_2^2 = 0.6, \qquad \sigma_{12}^2 = 0.2.$$

Then the four world variances are 1, 1.4, 1.6, and 2.2. The cross-world covariances follow from intersections:

$$\mathrm{Cov}(Y_1, Y_2) = 1, \qquad \mathrm{Cov}(Y_{12}, Y_1) = 1.4, \qquad \mathrm{Cov}(Y_{12}, Y_2) = 1.6.$$

The covariance matrix is encoded by four world variances. The Möbius transform recovers the four branch variances. Gaussianity then lifts this second-order description to the full joint law.

Appendix L. Complementary policy summaries

L.1. Mean outcomes and invariant summaries

Hold the population law fixed and vary the regime. The resulting mean outcome vectors describe policy geometry. Their covariance across a design over regimes uses a different variation axis from the covariance of random components across units. Let $Y_S \in L^1(\Omega; \mathbb{R}^d)$ and write

$$m_S = \mathbb{E}[Y_S], \qquad c_A = \mathbb{E}[C_A].$$

DEFINITION A4 (Cofactual span and invariant subspace). *Define*

$$\mathcal{V}_C = \mathrm{span}\{c_A : A \neq \varnothing\} \tag{A43}$$

and

$$\mathcal{I}_C = \{a \in \mathbb{R}^d : a^\top m_S \text{ is constant over } S \in \mathcal{W}\}. \tag{A44}$$

Let ν be a probability distribution on $\mathcal{W}$ with full support and define

$$\bar{m}_\nu = \sum_S \nu(S) m_S$$

and

$$B_\nu = \sum_S \nu(S)(m_S - \bar{m}_\nu)(m_S - \bar{m}_\nu)^\top. \tag{A45}$$

PROPOSITION A13 (Cofactual invariance geometry). *For every full-support design ν,*

(A46)
$$\operatorname{range}(B_\nu) = \mathcal{V}_C, \qquad \ker(B_\nu) = \mathcal{I}_C = \mathcal{V}_C^\perp,$$

and

(A47)
$$\operatorname{rank}(B_\nu) = \dim \mathcal{V}_C.$$

Equivalently,

(A48)
$$\operatorname{aff}\{m_S : S \in \mathcal{W}\} = m_\varnothing + \mathcal{V}_C.$$

PROOF. Taking expectations in Proposition 1 gives

$$m_S = c_\varnothing + \sum_{\substack{A \subseteq S \\ A \neq \varnothing}} c_A.$$

Hence every difference $m_S - m_\varnothing$ lies in $\mathcal{V}_C$. Conversely, Möbius inversion expresses each c_A, $A \neq \varnothing$, as a linear combination of the differences $m_B - m_\varnothing$. Thus

$$\operatorname{span}\{m_S - m_\varnothing : S \in \mathcal{W}\} = \mathcal{V}_C,$$

which gives (A48).

Let D be the matrix whose columns are $\sqrt{\nu(S)}(m_S - \bar{m}_\nu)$. Then $B_\nu = DD^\top$. Full support and recentering preserve the span of world differences, so $\operatorname{range}(D) = \mathcal{V}_C$. Hence $\operatorname{range}(B_\nu) = \operatorname{range}(D) = \mathcal{V}_C$. The kernel is the orthogonal complement of the range for a symmetric matrix, giving $\ker(B_\nu) = \mathcal{V}_C^\perp$. This is exactly $\mathcal{I}_C$, since $a \perp \mathcal{V}_C$ means $a^\top(m_S - m_\varnothing) = 0$ for every S. The rank statement follows. □

DEFINITION A5 (Effective cofactual dimension). *The effective cofactual dimension of the mean target is*

$$d_C = \operatorname{rank}(B_\nu) = \dim \mathcal{V}_C,$$

for any full-support ν.

Degree counts the policy switches involved in an interaction. Effective dimension counts the directions spanned by the mean targets. They describe different features of the same family.

L.2. Deterministic policy geometry

The two-state example illustrates the mean geometry just defined. One intervention changes rewards, another changes transitions, and the optimal action changes across

TABLE A4. Four policy worlds in the dynamic example

Regime	Active interventions	$V_S(0)$	$V_S(1)$	Optimal actions $(a(0), a(1))$
Baseline	$\varnothing$	11.302	13.163	$(0, 0)$
Subsidy	$\{1\}$	12.417	14.083	$(1, 0)$
Reliability	$\{2\}$	13.073	14.626	$(1, 0)$
Both	$\{1, 2\}$	14.911	16.144	$(1, 0)$

Values are computed by enumerating all four deterministic stationary policies in each regime and checking the Bellman inequalities. Each intervention changes a model primitive, and the low-state policy changes after intervention.

worlds. The value vectors are deterministic; their cofactual interaction records nonadditivity of the policy-value map.

Environment. There are two states $x \in \{0, 1\}$ and two actions $a \in \{0, 1\}$. An agent solves

$$V_S(x) = \max_{a \in \{0,1\}} \left\{ r_S(x, a) + \beta \sum_{x'} P_S(x' \mid x, a) V_S(x') \right\} \tag{A49}$$

with $\beta = 0.95$. State 1 is the high state. Action 1 is an investment. Two policy interventions are available.

Intervention 1 is an investment subsidy. The flow reward is

$$r_S(x, a) = 0.3 + 0.8x + 0.1a - \{1 - 0.35\mathbf{1}(1 \in S)\}a. \tag{A50}$$

Intervention 2 raises the probability that investment reaches or preserves the high state. When $a = 0$,

$$P_S(1 \mid 0, 0) = 0.15, \qquad P_S(1 \mid 1, 0) = 0.75.$$

With investment,

$$P_S(1 \mid 0, 1) = 0.55 + 0.30\mathbf{1}(2 \in S),$$

$$P_S(1 \mid 1, 1) = 0.92 + 0.04\mathbf{1}(2 \in S).$$

The Bellman problem is solved under all four worlds.

The cofactual value vectors are

$$c_\varnothing = (11.302, 13.163)^\top,$$
$$c_1 = (1.114, 0.921)^\top,$$
$$c_2 = (1.771, 1.463)^\top,$$
$$c_{12} = (0.724, 0.598)^\top.$$

The joint-intervention coordinate records nonadditivity of the value map. These four value vectors are deterministic, so their component distributions are degenerate. This calculation concerns policy geometry; the stochastic examples above concern cross-world dependence.

Invariant direction. Using equal weight on the four worlds gives

$$B_{\mathcal{V}} = \begin{pmatrix} 1.7144 & 1.4162 \\ 1.4162 & 1.1699 \end{pmatrix}.$$

Its eigenvalues are approximately 2.8843 and 0. Hence $d_C = 1$: the four value vectors lie on a line. The high-state optimal action remains zero in every regime, so its common Bellman equation is

$$0.2875V_S(1) - 0.2375V_S(0) = 1.1. \tag{A51}$$

This equation explains the affine restriction. A normalized null vector is approximately

$$a = (0.6369, -0.7710)^{\top},$$

and $a^{\top}V_S \approx -2.9498$ in each world.

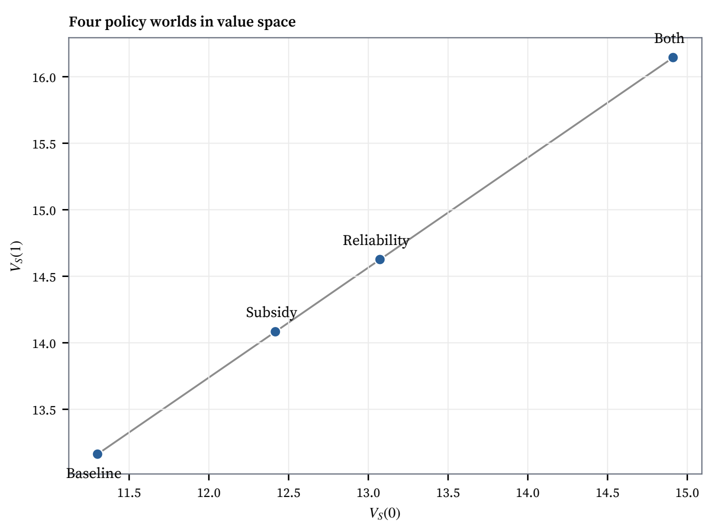

FIGURE A1. Counterfactual value geometry in the dynamic example

Each point is the value vector $(V_S(0), V_S(1))$ for one policy regime in Table A4. The four points are collinear, so the branch covariance has rank one. The orthogonal direction is invariant across the four worlds.

This is a mean-level statement built entirely from counterfactual geometry. A model with heterogeneous agents or shocks would add a distribution of V_S across units. Cofactual factorization would then impose a further structure on the cross-world heterogeneity.

L.3. Intervention variance contributions

The variance allocation uses the second-order part of the theory. Assume the system is square integrable and has second-order cofactual orthogonality, and write

$$\sigma_A^2 = \mathrm{Var}(C_A).$$

Then

$$\mathrm{Var}(Y_S) = \sum_{A \subseteq S} \sigma_A^2. \tag{A52}$$

Define the intervention variance game relative to the factual root by

$$v(S) = \mathrm{Var}(Y_S) - \mathrm{Var}(Y_\varnothing) = \sum_{\substack{A \subseteq S \\ A \neq \varnothing}} \sigma_A^2. \tag{A53}$$

PROPOSITION A14 (Cofactual variance importance). *The Shapley value of intervention i in the game* (A53) *is*

$$I_i^C = \sum_{A \ni i} \frac{\sigma_A^2}{|A|}. \tag{A54}$$

Moreover,

$$\sum_{i=1}^{p} I_i^C = \mathrm{Var}(Y_{[p]}) - \mathrm{Var}(Y_\varnothing). \tag{A55}$$

PROOF. The Harsanyi dividend of the set function v at coalition $A \neq \varnothing$ is σ_A^2 by Möbius inversion of (A53). The Shapley value allocates each dividend equally among its members, giving (A54). Summing over i allocates each σ_A^2 exactly once and yields (A55). □

Under pairwise orthogonality, (A54) is the Shapley allocation of intervention variance; higher-order dependence remains unrestricted (Shapley 1953; Owen 2014; Song et al. 2016).

References

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.

Tong Chen and Thomas Lumley. Optimal Multiwave Sampling for Regression Modelling in Two-Phase Designs. *Statistics in Medicine*, 39(30):4912–4921, 2020. doi: 10.1002/sim.8760.

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.

Edward H. Kennedy. Semiparametric Theory and Empirical Processes in Causal Inference. In *Statistical Causal Inferences and Their Applications in Public Health Research*, pages 141–167. Springer, 2016. doi: 10.1007/978-3-319-41259-7_8.

Art B. Owen. Sobol' Indices and Shapley Value. *SIAM/ASA Journal on Uncertainty Quantification*, 2 (1):245–251, 2014. doi: 10.1137/130936233.

Arun Ravichandran, Nicole E. Pashley, Brian Libgober, and Tirthankar Dasgupta. Optimal Allocation of Sample Size for Randomization-Based Inference from 2^K Factorial Designs. *Journal of Causal Inference*, 12(1):20230046, 2024. doi: 10.1515/jci-2023-0046.

Lloyd S. Shapley. A Value for n-Person Games. In Harold W. Kuhn and Albert W. Tucker, editors, *Contributions to the Theory of Games II*, pages 307–317. Princeton University Press, Princeton, 1953.

Eunhye Song, Barry L. Nelson, and Jeremy Staum. Shapley Effects for Global Sensitivity Analysis: Theory and Computation. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):1060–1083, 2016. doi: 10.1137/15M1048070.

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, New York, 2009. doi: 10.1007/b13794.