SHARP BOUNDS ON THE CENTRAL MOMENTS OF THE INDIVIDUAL CAUSAL EFFECT FROM SEVERAL MARGINAL CENTRAL MOMENTS: TWO OPEN CELLS SETTLED BY EXACT DUAL CERTIFICATES

KOREA SUPERINTELLIGENCE LABS

Abstract. Let Y_1, Y_0 be the potential outcomes of a binary treatment, $D = Y_1 - Y_0$ the individual causal effect and $\mu^{(m)} = \mathbb{E}[(D - \mathbb{E}D)^m]$ its central moments. Hashimoto, Kawakami and Tian (arXiv:2607.04957) ask what can be said about $\mu^{(m)}$ when only the marginal central moments $\sigma_x^{(k)} = \mathbb{E}[(Y_x - \mathbb{E}Y_x)^k]$, $x \in \{0,1\}$, $k \in K$, are known: they give sharp bounds for every single order k, intersect these for several orders, prove that the intersections are not sharp, and name the sharp multi-order bounds as an open problem. We settle the two first open cells at explicit instances, over all finitely supported joint laws. For $K = \{2,3,4\}$ with $\sigma_1 = (2,6,26)$ and $\sigma_0 = (3,0,15)$ the sharp range of the variance $\mu^{(2)}$ is exactly $[1,9]$, both ends attained by three-atom laws, where the intersected bound is $[5 - 2\sqrt{6},\, 5 + 2\sqrt{6}] \approx [0.101,\, 9.899]$; for $\sigma_1 = (3,-6,36)$ and $\sigma_0 = (7,-12,76)$ the sharp upper bound of $\mu^{(3)}$ is 48, attained, where the intersected bound is ≈ 97.80. The range of $\mu^{(m)}$ is the value of a linear program in the joint law of the two centred outcomes; its dual asks for a nonnegative quartic $F(u,v) = f(u) + g(v) - \psi(u,v)$, and each bound is certified by an exact rational sum of three squares together with an atomic law carried by the zero set of F. The instances are designed from their contact points rather than solved. At both instances the order-4 Hankel matrices of all four marginals are positive definite, so none of the four is determined by the four moments prescribed for it; this is precisely what puts the instances outside the reach of the degeneracy arguments by which the source proves non-sharpness. The bounds, the attaining laws, the comparisons with the published intervals and the non-degeneracy statements are verified in Lean 4 against *Mathlib* for every finitely supported joint law over an arbitrary index type. The sharp lower bound of $\mu^{(3)}$ at the second instance is left open.

1. Introduction

The objects. A binary treatment has two potential outcomes Y_1, Y_0, the *individual causal effect* (ICE) is $D = Y_1 - Y_0$, and the shape of its distribution across individuals is summarised by the central moments

$$\mu^{(m)} = \mathbb{E}\big[(D - \mathbb{E}D)^m\big], \qquad m = 2, 3, 4, \ldots$$

The joint law of (Y_1, Y_0) is never observed, so $\mu^{(m)}$ is not identified and one asks for its sharp range under whatever is known about the marginals. The classical line of work — Heckman, Smith and Clements [6], Fan and Park [5], Frandsen and Lefgren [7], and recently Post and Van den Heuvel [3, 4] and Kawakami and Tian [2] — assumes that the two *marginal distributions* of Y_1 and Y_0 are known and bounds the distribution, or the moments, of D over all couplings; the Fréchet–Hoeffding bounds [20, 21, 22] are the basic tool. Hashimoto, Kawakami and Tian [1] weaken the information set to what summary tables of a published trial actually contain: only the *marginal central moments*

$$\sigma_x^{(k)} = \mathbb{E}\big[(Y_x - \mathbb{E}Y_x)^k\big], \qquad x \in \{0,1\},\ k \in K,$$

for a finite set of orders K, and nothing about the marginal distributions beyond these numbers. Writing $U = Y_1 - \mathbb{E}Y_1$ and $V = Y_0 - \mathbb{E}Y_0$ one has $D - \mathbb{E}D = U - V$, so the question is: over all joint laws of (U,V) with $\mathbb{E}U = \mathbb{E}V = 0$, $\mathbb{E}U^k = \sigma_1^{(k)}$ and $\mathbb{E}V^k = \sigma_0^{(k)}$ for $k \in K$, what is the exact range of $\mathbb{E}[(U-V)^m]$?

The source and its open cells. The source answers this completely for a *single* order k. Its Table 1 collects its twelve single-order theorems, with the caption "All bounds reported in this table are sharp"; for instance its Theorem 3 is

$$\Big(\sqrt{\sigma_1^{(2)}} - \sqrt{\sigma_0^{(2)}}\Big)^2 \le \mu^{(2)} \le \Big(\sqrt{\sigma_1^{(2)}} + \sqrt{\sigma_0^{(2)}}\Big)^2,$$

its Theorem 10 bounds $\mu^{(3)}$ from the fourth moments alone by $\pm\frac{\sqrt{2}}{\sqrt[4]{27}}\Big(\sqrt[4]{\sigma_1^{(4)}} + \sqrt[4]{\sigma_0^{(4)}}\Big)^3$, and its Theorems 8 and 9 show that $k = 2$ alone and $k = 3$ alone say nothing about $\mu^{(3)}$ (outside the degenerate case in which both marginals are constant). For *several* orders it intersects the single-order intervals — Theorem 7 for $m = 2$, Theorem 11 for $m = 3$, Theorem 15 for $m = 4$, Corollary 3 in general — and says of these, in its introduction, that "these bounds are generally not sharp". Two of the intersected cells are the ones settled here:

- Theorem 7(4), $m = 2$ and $K = \{2, 3, 4\}$, returns the Theorem 3 interval unchanged; the source proves ("the bounds in (3) and (4) of Theorem 7 are not sharp: we provide their counterexamples in Appendix A.2") that it is not sharp. By contrast it proves that Theorem 7(1) ($K = \{2, 3\}$) and Theorem 7(2) ($K = \{2, 4\}$) *are* sharp, so $K = \{2, 3, 4\}$ is the first combination of orders at which the third and fourth moments jointly carry information about the variance.
- Theorem 11(2), $m = 3$ with $\sigma^{(4)}$ and at least one of $\sigma^{(2)}, \sigma^{(3)}$ available, returns the Theorem 10 interval unchanged and is proved not sharp in the same appendix. A remark in its Section 6.3 adds that settings in which orders $k > m$ are available, "of which Theorem 10 is a special instance", are "very challenging".

The Conclusion states the gap as an open problem, verbatim:

> "A limitation of this study is that we do not establish sharp bounds for $\mu^{(m)}$ when multiple $(\sigma_1^{(k)}, \sigma_0^{(k)})$ pairs are jointly available. Moreover, the bounds derived for the skewness and kurtosis of the ICE are not sharp. Deriving sharp bounds in these settings appears to be substantially more challenging and remains a future research direction."

Its notion of sharpness is stated in its Section 5: a closed interval $[l, u]$ is a sharp bound "if there exist no strictly tighter closed bounds for $\mu^{(m)}$ under the given moment information", and the preamble of its Appendix A.2 makes explicit that non-attainment of an endpoint does not by itself contradict sharpness. Both bounds proved below are *attained*, so they are sharp in the source's sense and in the stronger attained sense. (Statement, section and table numbers of [1] are those of the arXiv PDF of version 1, the only version as of 2026-09-07; see the bibliography.)

Results. Everything below is stated for finitely supported joint laws of (U, V) — a finite family of atoms $(u_i, v_i) \in \mathbb{R}^2$ with weights $w_i \ge 0$ summing to one, indexed by an arbitrary finite index set — with the prescribed moments; Section 7 explains why this is the right class and what is not claimed.

- **The variance cell** (Theorem 4.2). At $\sigma_1 = (\sigma_1^{(2)}, \sigma_1^{(3)}, \sigma_1^{(4)}) = (2, 6, 26)$ and $\sigma_0 = (3, 0, 15)$ every such law has $1 \le \mu^{(2)} \le 9$, and both ends are attained by explicit three-atom laws with rational atoms and weights. The source's Theorem 7(4) gives $[(\sqrt{2} - \sqrt{3})^2, (\sqrt{2} + \sqrt{3})^2] = [5 - 2\sqrt{6}, 5 + 2\sqrt{6}]$, strictly wider at both ends (Proposition 4.3): its lower end is smaller by the factor $1/(5 - 2\sqrt{6}) \approx 9.90$ and its upper end larger by about 10%.
- **The third-moment cell** (Theorem 5.2). At $\sigma_1 = (3, -6, 36)$ and $\sigma_0 = (7, -12, 76)$ every such law has $\mu^{(3)} \le 48$, attained by an explicit three-atom law. The source's Theorem 11(2) gives $\frac{\sqrt{2}}{\sqrt[4]{27}}(\sqrt[4]{36} + \sqrt[4]{76})^3 \approx 97.80$, larger by the factor 2.04 (Proposition 5.3). The matching sharp *lower* bound at this instance is not settled here (Remark 5.5).
- **Non-degeneracy** (Proposition 6.1). At both instances, for every real a, b, $\mathbb{E}[(U^2 + aU + b)^2]$ and $\mathbb{E}[(V^2 + aV + b)^2]$ are bounded below by $4, 6$ (first instance) and $15, 45/7$ (second instance). So no monic quadratic annihilates any of the four marginals, every compatible marginal has at least three atoms, and — equivalently, in Hankel language — none of the four marginals is determined by the four moments prescribed for it. This is the precise sense in which the

instances lie outside the source's method: every counterexample of its Appendix A.2 rests on moment data that pin one of the two marginals down to at most two atoms.

cell (m, K)	instance σ_1; σ_0	sharp (this paper)	intersected bound of [1]
$(2, \{2, 3, 4\})$	$(2, 6, 26)$; $(3, 0, 15)$	$1 \le \mu^{(2)} \le 9$, both attained	$[5 - 2\sqrt{6}, 5 + 2\sqrt{6}]$, Thm. 7(4)
$(3, \{2, 3, 4\})$	$(3, -6, 36)$; $(7, -12, 76)$	$\mu^{(3)} \le 48$, attained	$\mu^{(3)} \le \frac{\sqrt{2}}{\sqrt[4]{27}}(\sqrt[4]{36} + \sqrt[4]{76})^3$, Thm. 11(2)

TABLE 1. The two cells, with $\sigma_x = (\sigma_x^{(2)}, \sigma_x^{(3)}, \sigma_x^{(4)})$. Numerically the intersected bounds are [0.101, 9.899] and 97.80 (rounded); the comparisons are proved exactly (Propositions 4.3 and 5.3).

Method. The range of $\mu^{(m)}$ is the value of a linear program in the joint law: the objective $\mathbb{E}[(U-V)^m]$ and the constraints $\mathbb{E}U^k = \sigma_1^{(k)}$, $\mathbb{E}V^k = \sigma_0^{(k)}$ are integrals of fixed polynomials. Once the marginal moments are fixed, $\mu^{(2)}$ is $\sigma_1^{(2)} + \sigma_0^{(2)} - 2\mathbb{E}[UV]$ and $\mu^{(3)}$ is $\sigma_1^{(3)} - \sigma_0^{(3)} + \mathbb{E}[3UV^2 - 3U^2V]$, so the whole question is the range of one cross-moment functional $\mathbb{E}[\psi(U, V)]$ with $\psi = uv$, respectively $\psi = 3uv^2 - 3u^2v$. The dual program asks for a polynomial

$$F(u, v) = f(u) + g(v) - \psi(u, v) \ \ge 0 \text{ on all of } \mathbb{R}^2, \qquad \deg f, \deg g \le 4,$$

and then $0 \le \mathbb{E}[F(U, V)]$ reads off as $\mathbb{E}[\psi] \le \kappa + \sum_k (f_k\sigma_1^{(k)} + g_k\sigma_0^{(k)})$, a fixed rational combination of the data. A law supported on the zero set of F turns the inequality into an equality, so a pair (nonnegative dual polynomial, atomic law on its zero set) certifies a sharp bound, and both halves are finite rational data: $F \ge 0$ is exhibited as an exact rational sum of three squares, and the atomic law is written down. This duality is classical — Isii [9], Karlin and Studden [10], Kemperman [11], Bertsimas and Popescu [8], Lasserre [12, 13] — and it is not what is new here. What is new is the answer at the two cells, and the way the instances were obtained: they were *designed* from their contact points (Section 3) rather than solved, because the optimal support of a given instance is algebraic and in general irrational, whereas the map from a rational contact configuration to (dual polynomial, instance, value) is rational.

What is new and what is not. We are careful about attribution. The single-order bounds, the intersected multi-order bounds, and the fact that the intersected bounds are not sharp are all the source's [1]; the Cauchy–Schwarz variance bound of its Theorem 3 is elementary. The moment-problem duality and the sum-of-squares certification of nonnegative bivariate quartics (Hilbert [19]) are classical. The existence of finitely atomic representing measures (Tchakaloff [14], Richter [15], Bayer and Teichmann [16]) and the theory of the truncated moment problem (Curto and Fialkow [17], Schmüdgen [18]) are used only in Section 7, to explain the generality of the statements, and are not needed for the proofs. The heuristic explanation in Remark 3.1 of why $\{2, 3, 4\}$ is the first cell with something to compute is our reading, not the source's. What we claim is: the exact sharp values at the two instances of Table 1, the two certificates and the three attaining laws, the exact comparison with the published intervals, and the non-degeneracy statements that separate these instances from the source's counterexamples. As of 2026-09-07 the source has no citing work in OpenAlex; a full-text sweep of 890,835 arXiv e-prints for the phrases "central moments of the individual/causal/treatment effect", "marginal (central) moments information/only/given" and "sharp bounds on the variance/central moments of the individual/treatment/causal effect" returned the source and four unrelated papers (a Kalman-filter passage, a robust-mechanism-design passage, a bibliography entry on estimator variances, and a lemma about an estimator's moments); the arXiv API returns exactly one paper, the source, for `"central moments" AND "individual causal effect"` and none for `"central moments" AND "individual treatment effect"` or `"sharp bounds" AND "marginal moments"`.

Verification. Every theorem and proposition below that carries a formal counterpart in Appendix A is machine-checked in Lean 4 against *Mathlib*, for every finitely supported joint law over an arbitrary index type; Section 8 records the modules and the axioms. Where the text says more than its formal counterpart — the Hankel reading of Proposition 6.1, the identification of the radical expressions of Propositions 4.3 and 5.3 with the source's formulas, the zero-set computations of Remarks 4.4 and 5.4, the linear programs of Remarks 4.5 and 5.5, and the extension to laws with infinite support in Section 7 — it says so at that place.

2. Preliminaries

2.1. **Reduction to centred variables.** Let Y_1, Y_0 have finite fourth moments and put $U = Y_1 - \mathbb{E}Y_1$, $V = Y_0 - \mathbb{E}Y_0$. Then $D - \mathbb{E}D = U - V$, so $\mu^{(m)} = \mathbb{E}[(U-V)^m]$, and $\sigma_x^{(k)}$ is the k-th moment of U (for $x = 1$) or of V (for $x = 0$). The information "$\sigma_x^{(k)}$ for $k \in K$" is thus the statement $\mathbb{E}U = \mathbb{E}V = 0$, $\mathbb{E}U^k = \sigma_1^{(k)}$, $\mathbb{E}V^k = \sigma_0^{(k)}$ for $k \in K$, with the coupling free. Throughout, $K = \{2,3,4\}$ and we write $\sigma_1 = (\sigma_1^{(2)}, \sigma_1^{(3)}, \sigma_1^{(4)})$, σ_0 likewise.

2.2. **Finitely supported joint laws.** A *finitely supported joint law* is a finite index set I together with weights $w\colon I \to \mathbb{R}$, $w_i \geq 0$, $\sum_i w_i = 1$, and atoms $(u_i, v_i) \in \mathbb{R}^2$, $i \in I$; atoms may repeat and weights may vanish, which changes nothing. Its moments are

$$\mathbb{E}U^k = \sum_i w_i u_i^k, \qquad \mathbb{E}V^k = \sum_i w_i v_i^k, \qquad \mathbb{E}[U^pV^q] = \sum_i w_i u_i^p v_i^q,$$
$$\mu^{(m)} = \mathbb{E}[(U-V)^m] = \sum_i w_i (u_i - v_i)^m.$$

For moment data $\sigma_1, \sigma_0 \in \mathbb{R}^3$ let $\mathcal{L}(\sigma_1;\sigma_0)$ be the class of finitely supported joint laws with $\mathbb{E}U = \mathbb{E}V = 0$, $\mathbb{E}U^k = \sigma_1^{(k)}$ and $\mathbb{E}V^k = \sigma_0^{(k)}$ for $k = 2,3,4$. Nothing but these nine equations is imposed: not the index set, not the number of atoms, not their positions.

2.3. **The duality lemma.** Every bound in this paper is the composition of the two elementary statements below.

Lemma 2.1. *Let (I, w, u, v) be a finitely supported joint law.*

(a) *If $F\colon \mathbb{R}^2 \to \mathbb{R}$ satisfies $F(x,y) \geq 0$ for all x, y, then $\sum_i w_i F(u_i, v_i) \geq 0$.*

(b) *For real $c_0, \dots, c_4, d_1, \dots, d_4, e_1, e_2, e_3$,*

$$\sum_i w_i \Big(c_0 + \sum_{k=1}^4 c_k u_i^k + \sum_{k=1}^4 d_k v_i^k + e_1 u_i v_i + e_2 u_i v_i^2 + e_3 u_i^2 v_i \Big)$$
$$= c_0 \sum_i w_i + \sum_{k=1}^4 c_k \mathbb{E}U^k + \sum_{k=1}^4 d_k \mathbb{E}V^k + e_1 \mathbb{E}[UV] + e_2 \mathbb{E}[UV^2] + e_3 \mathbb{E}[U^2V].$$

(c) $\mathbb{E}[(U-V)^2] = \mathbb{E}U^2 + \mathbb{E}V^2 - 2\mathbb{E}[UV]$ *and* $\mathbb{E}[(U-V)^3] = \mathbb{E}U^3 - \mathbb{E}V^3 + \big(3\mathbb{E}[UV^2] - 3\mathbb{E}[U^2V]\big)$.

Proof. (a) is a sum of products of nonnegative numbers; (b) is linearity of a finite sum; (c) is (b) applied to the expansions of $(u-v)^2$ and $(u-v)^3$. □

Corollary 2.2 (the dual principle). *Let ψ be one of uv, $3uv^2 - 3u^2v$, let $f(u) = \sum_{k=0}^4 f_k u^k$ and $g(v) = \sum_{k=1}^4 g_k v^k$ be real polynomials, and suppose $F = f(u) + g(v) - \psi(u,v) \geq 0$ on $\mathbb{R}^2$. Then for every law in $\mathcal{L}(\sigma_1;\sigma_0)$,*

$$\mathbb{E}[\psi(U,V)] \;\leq\; f_0 + \sum_{k=2}^4 \big(f_k \sigma_1^{(k)} + g_k \sigma_0^{(k)} \big).$$

If moreover a law in $\mathcal{L}(\sigma_1;\sigma_0)$ is supported on $\{F = 0\}$, it attains the bound.

Proof. By Lemma 2.1(a) and (b), $0 \leq \sum_i w_i F(u_i, v_i) = f_0 + \sum_{k\geq 1} f_k \mathbb{E}U^k + \sum_{k \geq 1} g_k \mathbb{E}V^k - \mathbb{E}[\psi]$; the first-order moments vanish and the others are the data. If all atoms lie in $\{F = 0\}$ the left side is 0. □

By Lemma 2.1(c), an upper bound $\mathbb{E}[UV] \le c$ is the lower bound $\mu^{(2)} \ge \sigma_1^{(2)} + \sigma_0^{(2)} - 2c$, a lower bound $\mathbb{E}[UV] \ge -c'$ is $\mu^{(2)} \le \sigma_1^{(2)} + \sigma_0^{(2)} + 2c'$, and an upper bound $\mathbb{E}[3UV^2 - 3U^2V] \le c$ is $\mu^{(3)} \le \sigma_1^{(3)} - \sigma_0^{(3)} + c$.

Remark 2.3 (why the dual certificate is enough). Maximising $\mathbb{E}[\psi(U,V)]$ over $\mathcal{L}(\sigma_1;\sigma_0)$ is a linear program in the law (nine linear equality constraints, positivity), and Corollary 2.2 is weak duality for it: the multipliers of the nine constraints are $(f_0, f_1, \dots, f_4, g_1, \dots, g_4)$ and dual feasibility is exactly $f(u) + g(v) - \psi(u,v) \ge 0$ on $\mathbb{R}^2$. Attainment by a law on $\{F = 0\}$ is complementary slackness. No existence theory — neither strong duality nor Tchakaloff-type representation theorems — is needed to *verify* a certificate; those enter only in Section 7.

3. Certificates by design

3.1. **Structure of the contact set.** If $F \ge 0$ and $F(P) = 0$ then P is a minimum, so $\nabla F(P) = 0$: at every atom of an attaining law,

$$f'(u_i) = \partial_u \psi(u_i, v_i), \qquad g'(v_i) = \partial_v \psi(u_i, v_i).$$

For $\psi = uv$ this says $v_i = f'(u_i)$ and $u_i = g'(v_i)$: the second coordinate of every atom is a function of the first, so the optimal coupling of the variance problem is deterministic. It also shows that two atoms in "crossed" position cannot both be zeros of one certificate with $f = g$: from $F(r,s) = F(s,r) = 0$ one gets $F(r,r) + F(s,s) = -(r-s)^2 < 0$, a contradiction. This is the cyclical-monotonicity obstruction in miniature.

3.2. **The design procedure.** Fix ψ and three rational contact points $P_i = (u_i, v_i)$, $i = 1,2,3$, with distinct u_i and distinct v_i.

(1) Let L_f be the quadratic interpolating $u_i \mapsto \partial_u \psi(P_i)$ and put $f'(u) = L_f(u) + D_f \prod_i (u - u_i)$; likewise $g'(v) = L_g(v) + D_g \prod_i (v - v_i)$. Then $\nabla F(P_i) = 0$ for every choice of the constants D_f, D_g. Integrate, with $f(0) = \kappa$ and $g(0) = 0$.
(2) The three equations $F(P_i) = 0$ are linear in (D_f, D_g, κ); solve them exactly. Require $D_f, D_g > 0$, which makes F coercive.
(3) Decide $F \ge 0$ by an exact rational sum of squares in the basis $z = (1, u, v, u^2, uv, v^2)$, as follows. Write $F = z^\mathsf{T} G z$ with G a symmetric 6×6 matrix; the 15 coefficients of F are linear conditions on the 21 entries of G. If some $G \succeq 0$ represents F then $z(P_i)^\mathsf{T} G z(P_i) = F(P_i) = 0$ forces $Gz(P_i) = 0$, which adds 18 linear conditions. In both cases below the combined system has a unique solution G, and $G \succeq 0$ is checked by an exact LDL^T factorisation, whose pivots are the coefficients of the squares. If the unique candidate is not positive semidefinite, F is not a sum of squares in this basis, and by Hilbert's theorem on ternary quartics [19] (applied to the homogenisation $t^4 F(u/t, v/t)$, nonnegative on $\mathbb{R}^3$) a nonnegative F would be; so the configuration is discarded correctly.
(4) The weights are forced: $\sum w_i = 1$, $\sum w_i u_i = 0$, $\sum w_i v_i = 0$ is a 3×3 system. Require $w_i > 0$.
(5) Read the instance σ_1, σ_0 off the moments of the witness; the bound is $\sum_i w_i \psi(P_i)$.

Every quantity is a rational number and every step is exact; no floating point and no linear-programming solver was used at any point. The two certificates below were produced by this procedure and then re-derived, from the displayed data alone, by an independent implementation of steps 3–5 (exact Gram matrix, LDL^T, moments of the witness).

Remark 3.1 (why $\{2,3,4\}$ is the first cell with something to compute). This is our heuristic reading of the source's Theorem 7(1),(2), not a theorem. Let $k_{\max}$ be the largest available order. A far atom of mass δ at position M with $\delta M^{k_{\max}}$ held fixed contributes a constant to the $k_{\max}$-th moment, $\delta M^j = (\delta M^{k_{\max}}) M^{j - k_{\max}} \to 0$ to every lower moment, and, placed on a coordinate axis, nothing to $\mathbb{E}[UV]$. So the top-order moment can be raised freely above its Cauchy–Schwarz floor without disturbing the coupling problem, and only the orders below $k_{\max}$ constrain it. With $K = \{2,3\}$ or $K = \{2,4\}$ that leaves $k = 2$ alone — the source's Theorem 3 — which is what its Theorem 7(1),(2) assert. With $K = \{2,3,4\}$ the free order is 4 and $k = 3$ is rigid, so the third moments genuinely constrain the coupling. The argument is only heuristic as stated (it sends an atom to infinity), and the sharpness of Theorem 7(1),(2) is the source's.

4. The variance cell $(m, K) = (2, \{2, 3, 4\})$

Throughout this section $\sigma_1 = (2, 6, 26)$ and $\sigma_0 = (3, 0, 15)$, and

$$F_2(u,v) = \tfrac{43}{144} + u + \tfrac{35}{36}u^2 - \tfrac{11}{36}u^3 + \tfrac{5}{144}u^4 - \tfrac{1}{2}v + \tfrac{1}{8}v^2 + \tfrac{1}{18}v^3 + \tfrac{1}{48}v^4 - uv. \tag{1}$$

Lemma 4.1. *$F_2 \geq 0$ on $\mathbb{R}^2$. Explicitly,*

$$F_2(u,v) = \tfrac{1}{6192}\big(43 + 72u - 36v - 7u^2 - 4uv - 7v^2\big)^2 + \tfrac{1}{2226024}\big(719u - 166v - 221u^2 - 28uv + 166v^2\big)^2$$
$$+ \tfrac{7}{51768}\big(2v - 6u^2 + 9uv - 2v^2\big)^2.$$

Proof. Expanding the right-hand side gives (1); this is a polynomial identity in $\mathbb{Q}[u,v]$, checked by `ring` in the formal development and by two independent exact expansions outside it. □

Theorem 4.2 (the sharp range of the variance)**.** *Let (I, w, u, v) be a finitely supported joint law with $\sum_i w_i = 1$, $w_i \geq 0$,*

$$\textstyle\sum_i w_i u_i = 0, \qquad \sum_i w_i u_i^2 = 2, \qquad \sum_i w_i u_i^3 = 6, \qquad \sum_i w_i u_i^4 = 26,$$
$$\textstyle\sum_i w_i v_i = 0, \qquad \sum_i w_i v_i^2 = 3, \qquad \sum_i w_i v_i^3 = 0, \qquad \sum_i w_i v_i^4 = 15.$$

Then

$$1 \;\leq\; \sum_i w_i (u_i - v_i)^2 \;\leq\; 9.$$

Both bounds are attained: the law with atoms $(-1,-2), (0,1), (4,3)$ and weights $\frac{2}{5}, \frac{1}{2}, \frac{1}{10}$ satisfies the hypotheses and has $\sum_i w_i(u_i - v_i)^2 = 1$; the law with atoms $(-1,2), (0,-1), (4,-3)$ and the same weights satisfies the hypotheses and has $\sum_i w_i(u_i - v_i)^2 = 9$. In the language of Section 1: for potential outcomes whose centred joint law is finitely supported with $\sigma_1 = (2,6,26)$ and $\sigma_0 = (3,0,15)$, the sharp range of the variance $\mu^{(2)}$ of the individual causal effect is exactly $[1, 9]$.

Proof. Lower bound. Corollary 2.2 with $F = F_2$ and $\psi = uv$ gives

$$\mathbb{E}[UV] \;\leq\; \tfrac{43}{144} + \tfrac{35}{36}\cdot 2 - \tfrac{11}{36}\cdot 6 + \tfrac{5}{144}\cdot 26 + \tfrac{1}{8}\cdot 3 + \tfrac{1}{48}\cdot 15 = \tfrac{43+280-264+130+54+45}{144} = \tfrac{288}{144} = 2,$$

hence $\mu^{(2)} = 2 + 3 - 2\mathbb{E}[UV] \geq 1$ by Lemma 2.1(c).

Upper bound. The reflection $v \mapsto -v$ preserves the hypotheses, because the odd moments of V that are prescribed are $\mathbb{E}V = \mathbb{E}V^3 = 0$; this is why the instance was designed with $\sigma_0^{(3)} = 0$. Applying Corollary 2.2 to $F_2(u,-v) = f(u) + g(-v) + uv \geq 0$, i.e. to the certificate for $\psi = -uv$, gives $-\mathbb{E}[UV] \leq 2$ by the same arithmetic, hence $\mu^{(2)} = 5 - 2\mathbb{E}[UV] \leq 9$.

Attainment. For the first law: $\sum w_i = \frac{2}{5} + \frac{1}{2} + \frac{1}{10} = 1$; $\sum w_i u_i = -\frac{2}{5} + \frac{4}{10} = 0$, $\sum w_i u_i^2 = \frac{2}{5} + \frac{16}{10} = 2$, $\sum w_i u_i^3 = -\frac{2}{5} + \frac{64}{10} = 6$, $\sum w_i u_i^4 = \frac{2}{5} + \frac{256}{10} = 26$; $\sum w_i v_i = -\frac{4}{5} + \frac{1}{2} + \frac{3}{10} = 0$, $\sum w_i v_i^2 = \frac{8}{5} + \frac{1}{2} + \frac{9}{10} = 3$, $\sum w_i v_i^3 = -\frac{16}{5} + \frac{1}{2} + \frac{27}{10} = 0$, $\sum w_i v_i^4 = \frac{32}{5} + \frac{1}{2} + \frac{81}{10} = 15$; and $\sum w_i(u_i - v_i)^2 = \frac{2}{5}\cdot 1 + \frac{1}{2}\cdot 1 + \frac{1}{10}\cdot 1 = 1$. The second law has the same u-atoms and weights and the negated v-atoms, so its V-moments of even order agree and those of odd order change sign, i.e. stay 0; and $\sum w_i(u_i - v_i)^2 = \frac{2}{5}\cdot 9 + \frac{1}{2}\cdot 1 + \frac{1}{10}\cdot 49 = 9$. All three atoms of the first law are zeros of F_2, and all three atoms of the second are zeros of $F_2(u,-v)$, as complementary slackness requires. □

Proposition 4.3 (comparison with Theorem 7(4) of the source)**.** $(\sqrt{2} - \sqrt{3})^2 < 1$ *and* $9 < (\sqrt{2} + \sqrt{3})^2$.

Proof. $(\sqrt{2} \mp \sqrt{3})^2 = 5 \mp 2\sqrt{6}$ and $\sqrt{6} > 2$ because $6 > 4$. □

The source's Theorem 7(4) at this instance is the interval $[(\sqrt{\sigma_1^{(2)}} - \sqrt{\sigma_0^{(2)}})^2, (\sqrt{\sigma_1^{(2)}} + \sqrt{\sigma_0^{(2)}})^2] = [5 - 2\sqrt{6},\ 5 + 2\sqrt{6}] \approx [0.1010,\ 9.8990]$; the identification of this interval with the expressions of Proposition 4.3 is the substitution $\sigma_1^{(2)} = 2$, $\sigma_0^{(2)} = 3$. So the published interval is strictly wider than the sharp one at both ends: at the bottom by the factor $1/(5 - 2\sqrt{6}) = 5 + 2\sqrt{6} \approx 9.90$, at the top by $(5 + 2\sqrt{6})/9 \approx 1.0999$.

That Theorem 7(4) is not sharp is the source's own result; what is new is the exact size of the gap at an instance its method does not reach (Section 6).

Remark 4.4 (the zero set of F_2). Outside the formal development. The Gram matrix of Lemma 4.1 has rank 3 with kernel spanned by $z(P_i)$, so $F_2(u,v)=0$ iff $z(u,v)=(1,u,v,u^2,uv,v^2)$ lies in that kernel, i.e. iff the barycentric coordinates $c(u,v)$ of (u,v) with respect to the three atoms satisfy $u^2=\sum c_iu_i^2$, $uv=\sum c_iu_iv_i$, $v^2=\sum c_iv_i^2$. The first and third are conics through the three atoms whose fourth intersection point is $(7,-6)$, at which the second condition fails ($uv-\sum c_iu_iv_i=-84$). Hence the zero set of F_2 is exactly $\{(-1,-2),(0,1),(4,3)\}$, and by the reflection the zero set of $F_2(u,-v)$ is exactly the second witness. In particular every law attaining either endpoint of Theorem 4.2 is carried by the corresponding three atoms, and since three atoms with five prescribed moments have forced weights, the two witnesses are the *only* attaining laws.

Remark 4.5 (exact linear programs). Outside the formal development, as a consistency check. Restricted to the 3×3 product support $\{-1,0,4\}\times\{-2,1,3\}$ of the first witness, the hypotheses of Theorem 4.2 force the two marginals, the coupling polytope is a 4-dimensional transportation polytope with 10 vertices, and $\mathbb{E}[UV]$ ranges over $[-\frac{7}{5},2]$ on it: the value 2 is attained at the comonotone vertex, while the other endpoint -2 needs the reflected support. Over all laws supported on the integer grid $\{-6,\dots,6\}^2$ (an exact rational simplex with 169 variables and 9 equality constraints) the maximum and minimum of $\mathbb{E}[UV]$ are 2 and -2, attained by the two witnesses and nowhere else. This illustrates that the support of the optimum is not something one can fix in advance — which is why the certificate quantifies over all finitely supported laws.

5. The third-moment cell $(m,K)=(3,\{2,3,4\})$

Throughout this section $\sigma_1=(3,-6,36)$, $\sigma_0=(7,-12,76)$, $\psi(u,v)=3uv^2-3u^2v$, and

$$F_3(u,v)=\tfrac{112}{11}+12u+\tfrac{112}{11}u^2+\tfrac{29}{11}u^3+\tfrac{4}{11}u^4-\tfrac{36}{11}v-\tfrac{32}{11}v^2+\tfrac{3}{11}v^3+\tfrac{4}{11}v^4-3uv^2+3u^2v. \tag{2}$$

Unlike the variance cell there is no correlation to bound: objective and certificate are genuinely cubic and quartic.

Lemma 5.1. *$F_3\ge 0$ on $\mathbb{R}^2$. Explicitly,*

$$\begin{aligned}F_3(u,v)&=\tfrac{1}{1232}\big(112+66u-18v+9u^2+6uv-19v^2\big)^2\\&\quad+\tfrac{1}{1900976}\big(3086u+258v+515u^2+642uv-129v^2\big)^2\\&\quad+\tfrac{42}{16973}\big(10v+8u^2-11uv-5v^2\big)^2.\end{aligned}$$

Proof. A polynomial identity in $\mathbb{Q}[u,v]$, checked as in Lemma 4.1. □

Theorem 5.2 (the sharp upper bound of the third moment)**.** *Let (I,w,u,v) be a finitely supported joint law with $\sum_i w_i=1$, $w_i\ge 0$,*

$$\begin{aligned}&\textstyle\sum_i w_iu_i=0,\qquad \sum_i w_iu_i^2=3,\qquad \sum_i w_iu_i^3=-6,\qquad \sum_i w_iu_i^4=36,\\&\textstyle\sum_i w_iv_i=0,\qquad \sum_i w_iv_i^2=7,\qquad \sum_i w_iv_i^3=-12,\qquad \sum_i w_iv_i^4=76.\end{aligned}$$

Then

$$\sum_i w_i(u_i-v_i)^3\ \le\ 48.$$

The bound is attained: the law with atoms $(-4,-2),(0,2),(2,-4)$ and weights $\frac{1}{8},\frac{5}{8},\frac{1}{4}$ satisfies the hypotheses and has $\sum_i w_i(u_i-v_i)^3=48$. In the language of Section 1: for potential outcomes whose centred joint law is finitely supported with $\sigma_1=(3,-6,36)$ and $\sigma_0=(7,-12,76)$, the sharp upper bound of the third central moment $\mu^{(3)}$ of the individual causal effect is 48.

Proof. Corollary 2.2 with $F=F_3$ gives

$$\begin{aligned}\mathbb{E}[3UV^2-3U^2V]\ &\le\ \tfrac{112}{11}+\tfrac{112}{11}\cdot 3+\tfrac{29}{11}\cdot(-6)+\tfrac{4}{11}\cdot 36-\tfrac{32}{11}\cdot 7+\tfrac{3}{11}\cdot(-12)+\tfrac{4}{11}\cdot 76\\&=\tfrac{112+336-174+144-224-36+304}{11}=\tfrac{462}{11}=42,\end{aligned}$$

hence $\mu^{(3)} = \mathbb{E}U^3 - \mathbb{E}V^3 + \mathbb{E}[3UV^2 - 3U^2V] = -6 + 12 + \mathbb{E}[\psi] \le 48$ by Lemma 2.1(c). For the witness: $\sum w_i = 1$; $\sum w_i u_i = -\frac{1}{2} + \frac{1}{2} = 0$, $\sum w_i u_i^2 = 2 + 1 = 3$, $\sum w_i u_i^3 = -8 + 2 = -6$, $\sum w_i u_i^4 = 32 + 4 = 36$; $\sum w_i v_i = -\frac{1}{4} + \frac{5}{4} - 1 = 0$, $\sum w_i v_i^2 = \frac{1}{2} + \frac{5}{2} + 4 = 7$, $\sum w_i v_i^3 = -1 + 5 - 16 = -12$, $\sum w_i v_i^4 = 2 + 10 + 64 = 76$; and $\sum w_i (u_i - v_i)^3 = \frac{1}{8}(-2)^3 + \frac{5}{8}(-2)^3 + \frac{1}{4} \cdot 6^3 = -1 - 5 + 54 = 48$. The three atoms are zeros of F_3, as complementary slackness requires. □

Proposition 5.3 (comparison with Theorem 11(2) of the source)**.** *Let $p, q, B > 0$ be real numbers with $p^4 = 36$, $q^4 = 76$ and $27B^4 = 4(p+q)^{12}$. Then $48 < B$.*

Proof. $(12/5)^4 = 20736/625 < 36$ and $(5/2)^4 = 625/16 < 76$ give $p > 12/5$ and $q > 5/2$, so $p+q > 49/10$ and $4(p+q)^{12} > 4(49/10)^{12} > 27 \cdot 48^4 = 143{,}327{,}232$; thus $B^4 > 48^4$ and $B > 48$. □

With $p = \sqrt[4]{36}$, $q = \sqrt[4]{76}$ and $B = \frac{\sqrt{2}}{\sqrt[4]{27}}(p+q)^3$ one has $27B^4 = 27 \cdot \frac{4}{27}(p+q)^{12} = 4(p+q)^{12}$, so B is the upper bound of the source's Theorem 11(2) at this instance (this identification, $(\sqrt{2}/\sqrt[4]{27})^4 = 4/27$, is the one step outside the formal statement); numerically $B \approx 97.804$, and the published bound exceeds the sharp one by the factor $B/48 \approx 2.04$. Again the non-sharpness of Theorem 11(2) is the source's; the exact value is new.

Remark 5.4 (the zero set of F_3). Outside the formal development, as in Remark 4.4: the fourth intersection of the two conics is $(-3, 1)$, where the middle condition fails (-10), so the zero set of F_3 is exactly $\{(-4, -2), (0, 2), (2, -4)\}$ and the witness is the only law attaining 48. On the 3×3 product support of the witness the coupling polytope has 10 vertices and $\mathbb{E}[\psi]$ ranges over $[-18, 42]$; over the grid $\{-6, \ldots, 6\}^2$ the exact maximum of $\mathbb{E}[\psi]$ is again 42, attained only by the witness.

Remark 5.5 (the lower bound at this instance is open). Nothing here is formalised. The reflection $(u, v) \mapsto (-u, -v)$ maps $\mathcal{L}(\sigma_1; \sigma_0)$ onto $\mathcal{L}(\bar\sigma_1; \bar\sigma_0)$, where $\bar\sigma = (\sigma^{(2)}, -\sigma^{(3)}, \sigma^{(4)})$, and sends ψ to $-\psi$; so the sharp lower bound of $\mu^{(3)}$ at $(\sigma_1; \sigma_0)$ is $\sigma_1^{(3)} - \sigma_0^{(3)}$ minus the sharp upper bound of $\mathbb{E}[\psi]$ at the *odd-negated* instance $\bar\sigma_1 = (3, 6, 36)$, $\bar\sigma_0 = (7, 12, 76)$. A two-sided answer therefore needs two certificates whose designed instances are related by odd negation. An exhaustive exact run of the procedure of Section 3.2 over all three-point contact configurations with integer coordinates in $[-6, 6]$ (about 7 processor-minutes) certified 5,618 distinct instances and found no such pair; two further exhaustive runs restricted to configurations with $\sigma_1^{(3)} = \sigma_0^{(3)} = 0$ and to configurations with identical marginals, in which the reflection alone would supply the second endpoint, certified nothing. What is known: the source's Theorem 10 gives $\mu^{(3)} \ge -B \approx -97.80$ for every law of the instance, and the exact simplex of Remark 4.5 over the grid $\{-6, \ldots, 6\}^2$ exhibits a nine-atom law of the instance with $\mathbb{E}[\psi] = -\frac{27621}{980}$, i.e. $\mu^{(3)} = -\frac{21741}{980} \approx -22.18$. So the sharp lower bound lies in $[-97.81, -22.18]$; we do not know it. The natural next lever is a four-point contact set, for which the interpolation ansatz of Section 3.2 is over-determined by three equations and a degenerate family is needed.

6. Non-degeneracy: why the source's method stops short of these instances

Every counterexample in the source's Appendix A.2 works by choosing moment data that pin one of the two marginals down to at most two atoms. The main device is an equality case of the Cauchy–Schwarz inequality, packaged as the source's Lemma 1: from $\mathbb{E}Z = 0$, $\mathbb{E}Z^3 = \sqrt{2}$ (or $-\sqrt{2}$) and $\mathbb{E}Z^4 = 3$ it deduces, by the Cauchy–Schwarz inequality $(\mathbb{E}Z^3)^2 \le \mathbb{E}Z^2\, \mathbb{E}(Z^2 - v)^2$ with $v = \mathbb{E}Z^2$, that $v = 1$ and that equality holds, hence that $Z^2 - 1 = \pm\sqrt{2}Z$ almost surely, so that Z is carried by the two roots of $z^2 \mp \sqrt{2}z - 1$ with the two masses $\frac{1}{2} - \frac{\sqrt{3}}{6}$ and $\frac{1}{2} + \frac{\sqrt{3}}{6}$. Lemma 1 supplies the counterexamples to its bounds (19), (20) and (31); there *both* marginals are two-atom laws, the coupling polytope is a segment (one free parameter x), $\mu^{(m)}$ is affine in x, and the interval is read off the two endpoints, which are the comonotone and countermonotone couplings of Fréchet–Hoeffding type. The remaining counterexamples — to its bound (24), and to its bounds (30) and (32) — use two other degeneracies: $\sigma^{(4)} = 0$, which forces a constant marginal outright, and the Cauchy–Schwarz equality $\sigma^{(4)} = (\sigma^{(2)})^2$, which forces a marginal onto $\pm\sqrt{\sigma^{(2)}}$. There it is enough that *one* of the two marginals is pinned down. In moment language, all of them need a monic quadratic Q with $\mathbb{E}[Q(Z)^2] = 0$ for at least one of the two marginals, i.e. a singular order-4 Hankel matrix $\begin{pmatrix} 1 & 0 & \sigma^{(2)} \\ 0 & \sigma^{(2)} & \sigma^{(3)} \\ \sigma^{(2)} & \sigma^{(3)} & \sigma^{(4)} \end{pmatrix}$ — for $(1, \pm\sqrt{2}, 3)$ its determinant is $1 \cdot 3 - 2 - 1 = 0$, and it is likewise 0 for $(0,0,0)$ and for $(\sigma^{(2)}, 0, (\sigma^{(2)})^2)$. The following statement says that at both instances of this paper this is impossible for *either* marginal, by a margin.

Proposition 6.1 (no monic quadratic annihilates a marginal)**.** *Let (I, w, u, v) be a finitely supported joint law and a, b real numbers.*

(i) *Under the hypotheses of Theorem 4.2, $\sum_i w_i(u_i^2 + au_i + b)^2 \geq 4$ and $\sum_i w_i(v_i^2 + av_i + b)^2 \geq 6$.*
(ii) *Under the hypotheses of Theorem 5.2, $\sum_i w_i(u_i^2 + au_i + b)^2 \geq 15$ and $\sum_i w_i(v_i^2 + av_i + b)^2 \geq \frac{45}{7}$.*

Proof. By Lemma 2.1(b), for a centred marginal X with moments m_2, m_3, m_4,

$$\mathbb{E}[(X^2 + aX + b)^2] = m_4 + 2am_3 + (a^2 + 2b)m_2 + b^2 = \left(m_4 - \tfrac{m_3^2}{m_2} - m_2^2\right) + m_2\left(a + \tfrac{m_3}{m_2}\right)^2 + (b + m_2)^2,$$

so the minimum over a, b is $m_4 - m_3^2/m_2 - m_2^2$, attained at $a = -m_3/m_2$, $b = -m_2$. For $(2, 6, 26)$ this is $26 - 18 - 4 = 4$ at $(a, b) = (-3, -2)$; for $(3, 0, 15)$ it is $15 - 0 - 9 = 6$ at $(0, -3)$; for $(3, -6, 36)$ it is $36 - 12 - 9 = 15$ at $(2, -3)$; for $(7, -12, 76)$ it is $76 - \frac{144}{7} - 49 = \frac{45}{7}$ at $(\frac{12}{7}, -7)$. (The formal proofs are the same completions of squares.) ☐

Corollary 6.2. *At either instance, every compatible marginal law of U and of V has at least three atoms.*

Proof. If X were carried by two points r, s, then $Q(x) = (x - r)(x - s)$ would give $\mathbb{E}[Q(X)^2] = 0$, contradicting Proposition 6.1. ☐

Equivalently, and outside the formal development: $m_4 - m_3^2/m_2 - m_2^2$ is the Schur complement of the Hankel matrix displayed above, i.e. $\det H/m_2$, and the four Hankel matrices have determinants $8, 18, 45, 45$ with leading minors $(1,2), (1,3), (1,3), (1,7)$; all four are positive definite. By the theory of the truncated Hamburger moment problem [17, 18] a positive definite order-4 Hankel matrix admits a continuum of representing measures, so none of the four marginals is determined by its moments and the optimisation over couplings is genuine: there is no segment to walk along, and the answer requires a dual polynomial certificate, which the source uses nowhere. Without Proposition 6.1 a reader could not tell whether the instances of Table 1 are the source's Appendix A.2 instances in disguise.

Remark 6.3 (the rigid case)*.* When both marginals are two-atom laws — singular Hankel matrices — the coupling polytope is a segment and $\mu^{(m)}$ is affine on it, so the sharp interval for *every* m at once is read off its two endpoints. This is the computation behind the source's counterexamples to its bounds (19), (20) and (31), routine by its own method, and we do not pursue it.

7. Generality, and what is not claimed

Every finitely supported joint law, over an arbitrary index type. The formal statements quantify over an arbitrary index type ι, a finite set s of indices, and three functions $w, u, v \colon \iota \to \mathbb{R}$ (Appendix A). Any finitely supported joint law of (U, V) can be written in this form by indexing its atoms, and any such datum with $w \geq 0$ on s and $\sum_s w = 1$ is a finitely supported joint law; the hypotheses are the nine moment equations and nothing else. So Theorems 4.2 and 5.2 really say "for every finitely supported joint law with these marginal moments", not "for laws on a chosen support" and not "for the law we built". The witnesses are laws in the same class, so sharpness is a statement inside the class. This is the natural class for a sharpness question: the source's own counterexamples are finitely supported, and its notion of sharpness is about the existence of compatible laws.

Laws with infinite support. Not formalised. The dual half needs no finiteness: if (U, V) is any pair with finite fourth moments and the prescribed marginal moments, then $F(U, V) \geq 0$ almost surely, $F(U, V)$ is integrable (its monomials have degree at most four), and linearity of the integral gives the same bound as Corollary 2.2; the witnesses are finitely supported, so the sharp values of Theorems 4.2 and 5.2 are also the sharp values over all joint laws with finite fourth moments. The same conclusion follows without touching integrals from Tchakaloff's theorem in the form of Bayer and Teichmann [16].

Their Theorem 2 takes a finite positive Borel measure μ on $\mathbb{R}^N$ concentrated on a measurable set A and an integer $m \geq 1$, assumes only that every monomial of degree at most m is absolutely μ-integrable — no compactness of the support and no condition on moments of order $m+1$ — and produces finitely many points of A and positive weights reproducing the integral of *every* polynomial of degree at most m. Applied with $N = 2$, $m = 4$, μ the law of (U, V) and A its support (their Remark 3), it says: a probability law on $\mathbb{R}^2$ whose moments of total degree at most four are all finite — which follows from $\mathbb{E}|U|^4, \mathbb{E}|V|^4 < \infty$ — admits a finitely supported law, carried by at most 15 points of its own support, with the same moments of total degree at most four. That is in particular the same nine data and the same $\mathbb{E}[\psi]$, so the range of $\mathbb{E}[\psi]$ over all such laws equals its range over $\mathcal{L}(\sigma_1; \sigma_0)$. A measure-theoretic version of Lemma 2.1 in the formal development (integrability hypotheses plus linearity of the integral) would widen the formal statements at no mathematical cost; it has not been done.

What is not claimed. (i) The sharp *lower* bound of $\mu^{(3)}$ at the second instance (Remark 5.5). (ii) Any closed form: each result is one instance of one cell, obtained by design; the sharp bound as a function of (σ_1, σ_0) on an open region is not addressed, although the designed map (contact points) $\mapsto$ (instance, value) is a smooth family and is the obvious handle. (iii) Any cell other than the two of Table 1; in particular nothing about $m = 4$ or about the skewness and kurtosis ratios. (iv) Anything about laws with infinite support inside the formal development.

8. Verification

Lemma 2.1, Theorems 4.2 and 5.2, and Propositions 4.3, 5.3 and 6.1 rest on declarations formally verified in Lean 4 against *Mathlib* [23, 24] (Lean 4.32.0, Mathlib at its `v4.32.0` tag), whose statements are reproduced verbatim in Appendix A. The development is three modules, 551 lines in all. The first (110 lines; 3 definitions, 1 structure, 4 theorems) defines the moments `mom`, the mixed moments `mix` and $\mu^{(m)}$ as `Finset` sums over $\mathbb{R}$ indexed by an arbitrary type, and proves Lemma 2.1(a)–(c) (`dual_bound`, `expand`, `mu_two`, `mu_three`). The second (242 lines; 5 definitions, 1 structure, 16 theorems) defines F_2, proves Lemma 4.1 by `rw` of the sum-of-squares identity (discharged by `ring`) followed by `positivity`, packages the hypotheses of Theorem 4.2 as a structure `Inst`, and proves both bounds (`mu2_ge_one`, `mu2_le_nine`), both witnesses (`attains_one`, `attains_nine`, with the instance equations for the witnesses discharged by `simp` and `norm_num`), Proposition 4.3 (`paper_interval_strictly_wider`) and Proposition 6.1(i). The third (199 lines; 4 definitions, 1 structure, 11 theorems) has the same shape for Theorem 5.2 (`mu3_le`, `attains_48`), Proposition 5.3 (`paper_bound_gt`) and Proposition 6.1(ii). Negative controls are included and machine-checked: $\mu^{(2)} \leq 8$ and $2 \leq \mu^{(2)}$ are refuted by the witnesses (`not_le_eight`, `not_two_le`), $\mu^{(3)} \leq 47$ is refuted (`not_le_47`), and the hypotheses of both instances are satisfiable (`inst_nonvacuous`), so no theorem here is vacuous. Every proof is checked by the Lean kernel; there is no compiled evaluation (`native_decide`), no `sorry`, and no axiom declared by hand. The axiom print (`#print axioms`) of every one of the 31 theorems of the three modules — the 16 statements pinned to the ledger, the two nonnegativity lemmas, the five controls and eight auxiliary steps — is `propext`, `Classical.choice`, `Quot.sound`; it was re-run read-only for this paper. Outside the formal development, and labelled as such where they occur, are the design procedure of Section 3.2, the zero-set computations of Remarks 4.4 and 5.4, the exact linear programs of Remarks 4.5, 5.4 and 5.5, the Hankel reading of Proposition 6.1, the radical identifications following Propositions 4.3 and 5.3, and Section 7. All of these were carried out in exact rational arithmetic (Python `fractions`), twice and independently: once when the instances were designed and once, from the displayed data alone, for this paper; the exhaustive contact-point search of Remark 5.5 was re-run for this paper as well. As recorded when the development was written, each module checks in about ten seconds with a peak resident set of about 7 GB (the cost of importing *Mathlib*); the exact computations take about a minute and the exhaustive search about seven processor-minutes. The development is currently in a private repository: repository reference to be added at submission.

Appendix A. The formal statements

The Lean statements corresponding to the results above, verbatim (statements only; proofs, docstrings and attributes omitted). `ι` is an arbitrary type, `s : Finset ι` a finite set of indices, `w u v : ι → ℝ` the weights and the two coordinates of the atoms, and `∑ i ∈ s, ...` the finite sum; `Fin 3` is the three-element index type and `![a, b, c]` the function on it with those values. The namespaces `LeanProblemSpec.IcemomSharp`, `.Mu2` and `.Mu3` are omitted.

Definitions and the duality lemma (Lemma 2.1).

```
def mom (s : Finset ι) (w x : ι → ℝ) (k : ℕ) : ℝ := ∑ i ∈ s, w i * x i ^ k

def mix (s : Finset ι) (w u v : ι → ℝ) (p q : ℕ) : ℝ := ∑ i ∈ s, w i * (u i ^ p * v i ^ q)

def mu (s : Finset ι) (w u v : ι → ℝ) (m : ℕ) : ℝ := ∑ i ∈ s, w i * (u i - v i) ^ m

theorem dual_bound (s : Finset ι) (w u v : ι → ℝ) (hw : ∀ i ∈ s, 0 ≤ w i)
    (F : ℝ → ℝ → ℝ) (hF : ∀ x y, 0 ≤ F x y) :
    0 ≤ ∑ i ∈ s, w i * F (u i) (v i)

theorem expand (s : Finset ι) (w u v : ι → ℝ)
    (c0 c1 c2 c3 c4 d1 d2 d3 d4 e1 e2 e3 : ℝ) :
    ∑ i ∈ s, w i * (c0 + c1 * u i + c2 * u i ^ 2 + c3 * u i ^ 3 + c4 * u i ^ 4
        + d1 * v i + d2 * v i ^ 2 + d3 * v i ^ 3 + d4 * v i ^ 4
        + e1 * (u i * v i) + e2 * (u i * v i ^ 2) + e3 * (u i ^ 2 * v i))
      = c0 * (∑ i ∈ s, w i) + c1 * mom s w u 1 + c2 * mom s w u 2 + c3 * mom s w u 3
        + c4 * mom s w u 4 + d1 * mom s w v 1 + d2 * mom s w v 2 + d3 * mom s w v 3
        + d4 * mom s w v 4 + e1 * mix s w u v 1 1 + e2 * mix s w u v 1 2
        + e3 * mix s w u v 2 1

theorem mu_two (s : Finset ι) (w u v : ι → ℝ) :
    mu s w u v 2 = mom s w u 2 + mom s w v 2 - 2 * mix s w u v 1 1

theorem mu_three (s : Finset ι) (w u v : ι → ℝ) :
    mu s w u v 3 = mom s w u 3 - mom s w v 3 + (3 * mix s w u v 1 2 - 3 * mix s w u v 2 1)
```

The variance cell (Lemma 4.1, Theorem 4.2, Propositions 4.3 and 6.1(i)).

```
noncomputable def F (u v : ℝ) : ℝ :=
  43/144 + u + 35/36 * u ^ 2 - 11/36 * u ^ 3 + 5/144 * u ^ 4
    - 1/2 * v + 1/8 * v ^ 2 + 1/18 * v ^ 3 + 1/48 * v ^ 4 - u * v

theorem F_nonneg (u v : ℝ) : 0 ≤ F u v

structure Inst (s : Finset ι) (w u v : ι → ℝ) : Prop where
  nonneg : ∀ i ∈ s, 0 ≤ w i
  total : ∑ i ∈ s, w i = 1
  u1 : mom s w u 1 = 0
  u2 : mom s w u 2 = 2
  u3 : mom s w u 3 = 6
  u4 : mom s w u 4 = 26
  v1 : mom s w v 1 = 0
  v2 : mom s w v 2 = 3
  v3 : mom s w v 3 = 0
  v4 : mom s w v 4 = 15
```

```
theorem mu2_ge_one (s : Finset ι) (w u v : ι → ℝ) (h : Inst s w u v) : 1 ≤ mu s w u v 2

theorem mu2_le_nine (s : Finset ι) (w u v : ι → ℝ) (h : Inst s w u v) : mu s w u v 2 ≤ 9

noncomputable def wt : Fin 3 → ℝ := ![2/5, 1/2, 1/10]
noncomputable def uu : Fin 3 → ℝ := ![-1, 0, 4]
noncomputable def vlo : Fin 3 → ℝ := ![-2, 1, 3]
noncomputable def vhi : Fin 3 → ℝ := ![2, -1, -3]

theorem inst_lo : Inst (Finset.univ : Finset (Fin 3)) wt uu vlo
theorem inst_hi : Inst (Finset.univ : Finset (Fin 3)) wt uu vhi

theorem attains_one : mu (Finset.univ : Finset (Fin 3)) wt uu vlo 2 = 1

theorem attains_nine : mu (Finset.univ : Finset (Fin 3)) wt uu vhi 2 = 9

theorem paper_interval_strictly_wider :
    (Real.sqrt 2 - Real.sqrt 3) ^ 2 < 1 ∧ (9 : ℝ) < (Real.sqrt 2 + Real.sqrt 3) ^ 2

theorem no_quadratic_kills_u (s : Finset ι) (w u v : ι → ℝ) (h : Inst s w u v) (a b : ℝ) :
    4 ≤ ∑ i ∈ s, w i * (u i ^ 2 + a * u i + b) ^ 2

theorem no_quadratic_kills_v (s : Finset ι) (w u v : ι → ℝ) (h : Inst s w u v) (a b : ℝ) :
    6 ≤ ∑ i ∈ s, w i * (v i ^ 2 + a * v i + b) ^ 2
```

The third-moment cell (Lemma 5.1, Theorem 5.2, Propositions 5.3 and 6.1(ii)).

```
noncomputable def F (u v : ℝ) : ℝ :=
  112/11 + 12 * u + 112/11 * u ^ 2 + 29/11 * u ^ 3 + 4/11 * u ^ 4
    - 36/11 * v - 32/11 * v ^ 2 + 3/11 * v ^ 3 + 4/11 * v ^ 4
    - 3 * (u * v ^ 2) + 3 * (u ^ 2 * v)

theorem F_nonneg (u v : ℝ) : 0 ≤ F u v

structure Inst (s : Finset ι) (w u v : ι → ℝ) : Prop where
  nonneg : ∀ i ∈ s, 0 ≤ w i
  total : ∑ i ∈ s, w i = 1
  u1 : mom s w u 1 = 0
  u2 : mom s w u 2 = 3
  u3 : mom s w u 3 = -6
  u4 : mom s w u 4 = 36
  v1 : mom s w v 1 = 0
  v2 : mom s w v 2 = 7
  v3 : mom s w v 3 = -12
  v4 : mom s w v 4 = 76

theorem mu3_le (s : Finset ι) (w u v : ι → ℝ) (h : Inst s w u v) : mu s w u v 3 ≤ 48

noncomputable def wt : Fin 3 → ℝ := ![1/8, 5/8, 1/4]
noncomputable def uu : Fin 3 → ℝ := ![-4, 0, 2]
noncomputable def vv : Fin 3 → ℝ := ![-2, 2, -4]
```

```
theorem inst_wit : Inst (Finset.univ : Finset (Fin 3)) wt uu vv

theorem attains_48 : mu (Finset.univ : Finset (Fin 3)) wt uu vv 3 = 48

theorem paper_bound_gt (p q B : ℝ) (hp : 0 < p) (hq : 0 < q)
    (hp4 : p ^ 4 = 36) (hq4 : q ^ 4 = 76) (hB : 0 < B)
    (hB4 : 27 * B ^ 4 = 4 * (p + q) ^ 12) : 48 < B

theorem no_quadratic_kills_u (s : Finset ι) (w u v : ι → ℝ) (h : Inst s w u v) (a b : ℝ) :
    15 ≤ ∑ i ∈ s, w i * (u i ^ 2 + a * u i + b) ^ 2

theorem no_quadratic_kills_v (s : Finset ι) (w u v : ι → ℝ) (h : Inst s w u v) (a b : ℝ) :
    45/7 ≤ ∑ i ∈ s, w i * (v i ^ 2 + a * v i + b) ^ 2
```

Controls (Section 8).

```
theorem Mu2.not_le_eight :
    ¬ ∀ (w u v : Fin 3 → ℝ), Inst Finset.univ w u v → mu Finset.univ w u v 2 ≤ 8

theorem Mu2.not_two_le :
    ¬ ∀ (w u v : Fin 3 → ℝ), Inst Finset.univ w u v → 2 ≤ mu Finset.univ w u v 2

theorem Mu2.inst_nonvacuous : ∃ w u v : Fin 3 → ℝ, Inst Finset.univ w u v

theorem Mu3.not_le_47 :
    ¬ ∀ (w u v : Fin 3 → ℝ), Inst Finset.univ w u v → mu Finset.univ w u v 3 ≤ 47

theorem Mu3.inst_nonvacuous : ∃ w u v : Fin 3 → ℝ, Inst Finset.univ w u v
```

References

[1] N. Hashimoto, Y. Kawakami and J. Tian, *Identification and bounding of central moments of causal effects using marginal moments information*, arXiv:2607.04957v1 [stat.ME], 6 July 2026. Read in full. Statement, section, table and equation numbers are those of the arXiv PDF of version 1 (12 pages including Appendices A–D), the only version as of 2026-09-07: Definition 1, Table 1, Theorems 3, 7, 8, 9, 10, 11 and 15, Corollary 3, Lemma 1, Section 5 (the definition of sharpness), Section 6.3 (the remark on orders $k > m$), Appendix A.2 (the counterexamples), and the Conclusion; the quotations in Section 1 are verbatim from that PDF. The e-print's LaTeX source does not compile as distributed (a missing font aborts the PDF stage), so the PDF, not a compiled `.aux`, is the authority for the numbering.

[2] Y. Kawakami and J. Tian, *Moments of causal effects*, in: Proceedings of the Forty-first Conference on Uncertainty in Artificial Intelligence (UAI 2025), Proceedings of Machine Learning Research **286**, PMLR, 2025, pp. 2011–2043; also arXiv:2505.04971 [stat.ME]. Proceedings data checked against the PMLR volume page (`proceedings.mlr.press/v286/kawakami25a.html`); the arXiv record carries no journal reference. Cited from the source's bibliography; not read for this work.

[3] R. A. J. Post, *Causal Effect Heterogeneity: Statistical formalization and analysis of the individual causal effect*, PhD thesis, Eindhoven University of Technology, Eindhoven, 2023, 191 pp., ISBN 978-90-386-5677-9. Bibliographic data checked against the TU/e research portal (award date 7 March 2023; no DOI is registered). Cited as in the source's bibliography; not read for this work.

[4] R. A. J. Post and E. R. van den Heuvel, *Beyond conditional averages: Estimating the individual causal effect distribution*, Journal of Causal Inference **13** (2025), no. 1, article 20240007. Bibliographic data checked against Crossref (doi:10.1515/jci-2024-0007); cited by the source for the Independent Effect Deviation assumption; not read for this work.

[5] Y. Fan and S. S. Park, *Sharp bounds on the distribution of treatment effects and their statistical inference*, Econometric Theory **26** (2010), no. 3, 931–951. Bibliographic data checked against Crossref (doi:10.1017/S0266466609990168); the nearest predecessor of the source's question, with the full marginal distributions known; not read for this work.

[6] J. J. Heckman, J. Smith and N. Clements, *Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts*, Review of Economic Studies **64** (1997), no. 4, 487–535. Bibliographic data checked against Crossref (doi:10.2307/2971729); not read for this work.

[7] B. R. Frandsen and L. J. Lefgren, *Partial identification of the distribution of treatment effects with an application to the Knowledge is Power Program (KIPP)*, Quantitative Economics **12** (2021), no. 1, 143–171. Bibliographic data checked against Crossref (doi:10.3982/QE1273); not read for this work.

[8] D. Bertsimas and I. Popescu, *Optimal inequalities in probability theory: A convex optimization approach*, SIAM Journal on Optimization **15** (2005), no. 3, 780–804. Bibliographic data checked against Crossref (doi:10.1137/S1052623401399903). The canonical modern reference for the technique: a moment problem is a linear program over measures whose dual asks for a polynomial majorant. Their setting is bounds on $P(X \in S)$ from joint moments; the projection onto a cross-moment with only marginal moments prescribed is not treated there.

[9] K. Isii, *On sharpness of Tchebycheff-type inequalities*, Annals of the Institute of Statistical Mathematics **14** (1962), no. 1, 185–197. Bibliographic data checked against Crossref (doi:10.1007/BF02868641); not read for this work.

[10] S. Karlin and W. J. Studden, *Tchebycheff Systems: With Applications in Analysis and Statistics*, Pure and Applied Mathematics XV, Interscience, New York, 1966. Not read for this work; no DOI, bibliographic data not checked against Crossref.

[11] J. H. B. Kemperman, *The general moment problem, a geometric approach*, Annals of Mathematical Statistics **39** (1968), no. 1, 93–122. Bibliographic data checked against Crossref (doi:10.1214/aoms/1177698508); not read for this work.

[12] J. B. Lasserre, *Global optimization with polynomials and the problem of moments*, SIAM Journal on Optimization **11** (2001), no. 3, 796–817. Bibliographic data checked against Crossref (doi:10.1137/S1052623400366802); not read for this work.

[13] J. B. Lasserre, *Moments, Positive Polynomials and Their Applications*, Series on Optimization and Its Applications, Imperial College Press, London, 2009. Bibliographic data checked against Crossref (doi:10.1142/p665); not read for this work.

[14] V. Tchakaloff, *Formules de cubatures mécaniques à coefficients non négatifs*, Bulletin des Sciences Mathématiques (2) **81** (1957), 123–134. Bibliographic data, including the last page, checked against zbMATH (Zbl 0079.13908, which gives *Bull. Sci. Math., II. Sér.* **81**, 123–134 (1957)); the reference list of [16] gives the same volume and first page. Not read for this work.

[15] H. Richter, *Parameterfreie Abschätzung und Realisierung von Erwartungswerten*, Blätter der Deutschen Gesellschaft für Versicherungsmathematik **3** (1957), no. 2, 147–162. Bibliographic data checked against Crossref (doi:10.1007/BF02808864); not read for this work.

[16] C. Bayer and J. Teichmann, *The proof of Tchakaloff's theorem*, Proceedings of the American Mathematical Society **134** (2006), no. 10, 3035–3040. Bibliographic data checked against Crossref (doi:10.1090/S0002-9939-06-08249-9); their Theorem 2, quoted in substance in Section 7, carries the same number in the published paper and in the e-print arXiv:math/0502473. Cited only for the existence of a finitely supported law with the same moments up to a given degree.

[17] R. E. Curto and L. A. Fialkow, *Solution of the truncated complex moment problem for flat data*, Memoirs of the American Mathematical Society **119** (1996), no. 568. Bibliographic data checked against Crossref (doi:10.1090/memo/0568); not read for this work.

[18] K. Schmüdgen, *The Moment Problem*, Graduate Texts in Mathematics 277, Springer, Cham, 2017. Bibliographic data checked against Crossref (doi:10.1007/978-3-319-64546-9); not read for this work.

[19] D. Hilbert, *Ueber die Darstellung definiter Formen als Summe von Formenquadraten*, Mathematische Annalen **32** (1888), no. 3, 342–350. Bibliographic data checked against Crossref (doi:10.1007/BF01443605). Cited for the fact that a nonnegative ternary quartic form is a sum of three squares of quadratic forms, which is what makes step 3 of Section 3.2 a decision procedure; not needed for the correctness of the certificates.

[20] M. Fréchet, *Sur les tableaux de corrélation dont les marges sont données*, Annales de l'Université de Lyon, Section A (3) **14** (1951), 53–77. Bibliographic data checked against zbMATH (Zbl 0045.22905, *Ann. Univ. Lyon, III. Sér., Sect. A* **14**, 53–77 (1951)). Not in Crossref; not read for this work.

[21] W. Hoeffding, *Maßstabinvariante Korrelationstheorie*, Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin **5** (1940), no. 3, 181–233. Bibliographic data checked against zbMATH (Zbl 0024.05602) and the Jahrbuch (JFM 66.0649.02), both of which give pages 181–233; some library catalogue records give 179–233 for the issue as a whole. Not in Crossref; not read for this work.

[22] R. B. Nelsen, *An Introduction to Copulas*, 2nd ed., Springer Series in Statistics, Springer, New York, 2006. Bibliographic data checked against Crossref (doi:10.1007/0-387-28678-0); cited for the Fréchet–Hoeffding bounds; not read for this work.

[23] L. de Moura and S. Ullrich, *The Lean 4 theorem prover and programming language*, in: Automated Deduction – CADE 28, Lecture Notes in Computer Science **12699**, Springer, 2021, pp. 625–635.

[24] The mathlib Community, *The Lean mathematical library*, in: Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP 2020), ACM, 2020, pp. 367–381.